Scale-Invariant Contour Completion using Conditional Random Fields

Xiaofeng Ren, Charless C. Fowlkes, Jitendra Malik
Computer Science Division
University of California, Berkeley, CA 94720
{xren,fowlkes,malik}@cs.berkeley.edu

Abstract

We present a model of curvilinear grouping using piecewise linear representations of contours and a conditional random field to capture continuity and the frequency of different junction types. Potential completions are generated by building a constrained Delaunay triangulation (CDT) over the set of contours found by a local edge detector.

Maximum likelihood parameters for the model are learned from human labeled groundtruth. Using held out test data, we measure how the model, by incorporating continuity structure, improves boundary detection over the local edge detector. We also compare performance with a baseline local classifier that operates on pairs of edgels.

Both algorithms consistently dominate the low-level boundary detector at all thresholds. To our knowledge, this is the first time that curvilinear continuity has been shown quantitatively useful for a large variety of natural images. Better boundary detection has immediate application in the problem of object detection and recognition.

1. Introduction

Finding the boundaries of objects and surfaces in a scene is a problem of fundamental importance for computer vision. There is a large body of work on object recognition which relies on bottom-up boundary detection to provide information about object shape (e.g. [5, 11, 3]). Even in cases where simple intensity features are sufficient for object detection, e.g. faces, it is still desirable to incorporate bottom-up boundary detection in order to provide precise object segmentation (e.g. [25]). The availability of high quality boundary location estimates will ultimately govern whether these algorithms are successful in real-world scenes where clutter and texture abound.

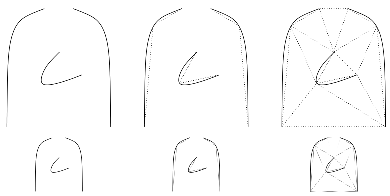
Boundaries are typically detected using some local operator. For example, the recent work by [15] trains a classifier that predicts the **p**robability of **b**oundary, Pb, at each pixel location using local brightness, texture and color gradient cues as features. Training data comes from a set of images where boundaries have been marked by human subjects. Unfortunately, this algorithm still misses many true boundaries and falsely detects others. The authors argue that detection failures are primarily due to lack of context since human observers presented with only local image patches perform no better than the classifier [16]. How can we model this missing context?

One well known source of contextual information is that provided by *curvilinear continuity*, sometimes referred to as "good continuation" or "illusory contour completion". Inspired by Wertheimer and Kanizsa, the study of this phenomenon has a long and influential tradition in psychophysics as well as neurophysiology [19]. More recently, *ecological statistics* of contours have confirmed the existence of curvilinear continuity in natural images [8] and its scale-invariant properties [21]. In computer vision there exists a rich literature on how to model curvilinear continuity (e.g. [22, 20, 18, 27]). More recent developments focus on finding *closed* salient contours [7, 28, 14].

The most basic scientific question is whether this type of additional contextual processing improves the accuracy of boundary detection. In particular, any algorithm which boosts low-contrast or illusory boundaries may also find spurious completions in heavily texture regions of a scene. Is the net effect positive or negative? This question can only be answered by quantitative measurements. *To the best of our knowledge, no such measurements have been made on a large, diverse set of real-world images.*

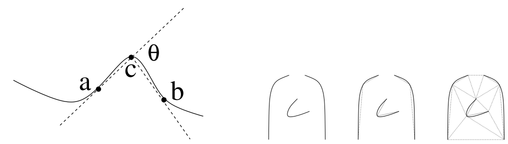
The contribution of this paper is a stochastic model of continuity and closure which provides boundary estimates that are quantifiably better at all thresholds than the input provided by a local detector. Our approach starts with locally computed Pb which is discretized into a set of piecewise linear segments. Potential completions are generated utilizing the *constrained Delaunay triangulation* as described in Section 2. We show that this scale-invariant geometric structure captures a vast majority of the true image boundaries, while reducing the complexity from hundreds of thousands of pixels to a thousand candidate line segments.

In Section 3 we develop two models of curvilinear con-



(a) (b) (c)

Figure 1: Finding potential completions: (a) an object may appear at any scale in the visual field (b) a piecewise linear curve approximates the boundaries regardless of scale. (c) *Constrained Delaunay Triangulation* connects the gaps and completes a scale-invariant representation.



(a) (b)

Figure 2: Building a discrete graph. **(a)** we recursively split a line until the angle θ is below a threshold. **(b)** an illustration of the process: the input edge map, the linearization, and the Constrained Delaunay Triangulation.

tinuity on this graphical/geometric structure: a simple local model which classifies each pair of edges independently and a global model which uses a *conditional random field* [13] in order to enforce both continuity and long-range probabilistic constraints on junctions.

Quantitative evaluation is made possible by utilizing three existing image collections that have been labeled with ground-truth boundaries: a set of 30 news photos of baseball players [17], 344 images of horses [4] and the BSDS300 [1], a boundary detection benchmark based on images of natural scenes. The performance evaluation of the local and global models on these 3 datasets is presented in Section 4 along with exemplary results.

2. Scale-Invariant Completion

The starting point of our approach is the observation that piecewise linear approximations of contours can be constructed in a scale-invariant way. Using the *constrained Delaunay triangulation* to complete the piecewise linear approximations preserves this scale-invariance. Figure 1 gives an illustration. 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 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 also recently applied to image segmentation by [29].

This piecewise linear representation has many advantages over using the pixel grid: (1) by moving from 100,000 pixels to 1000 edges, it achieves high statistical and computational efficiency. (2) this discrete representation of the image is scale-invariant, independent of the image resolution. (3) By restricting completions to the edges in the graph, it partially solves the problem of having too many spurious completions. Moreover, we will show empirically that very little of the true boundary structure is lost in this discretization process.

Of particular interest to us is scale invariance. Empirical studies of the *ecological statistics* of contours [8, 21, 6] have revealed one fundamental problem in many existing models: the lack of *scale invariance*. For example, the stochastic motion model of [27] predicts that if one breaks contours into segments at high-curvature locations, the distribution of segment length should be exponential. Work by [21] suggests that the empirical distribution is *not* exponential; instead it closely follows a power law, a characteristic of *scale-invariant* phenomena. A related power law is also reported in [6]. These findings are consistent with the intuition that we see objects at arbitrary scales. Scale invariance is also crucial from a practical point of view, as it is required in many fundamental vision tasks such as object recognition.

2.1 Building a piecewise linear representation

Our discretization step starts with using Canny's hysteresis thresholding to trace the Pb contours and then recursively split them until each segment is approximately straight. Figure 2(a) shows an illustration of this linearization: for a given contour, let θ 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 θ is maximized. If θ is below a threshold, split the contour by adding a vertex at c and continue. A heuristic is added to handle T-junctions: when a vertex is very close to another line, we split this line and insert an additional vertex.

We use the TRIANGLE program [24] to produce CDTs as shown in Figure 3. The linearized edges extracted from the Pb contours become constrained edges in the triangulation which we refer to as **gradient edges** or G**-edges**. The remaining **completion edges** or C**-edges** are filled in by the CDT. Of particular interest to us is CDT's ability to complete contours across gaps in the local detector output. 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 no other vertices lie in the circumcircle and CDT will correctly com-

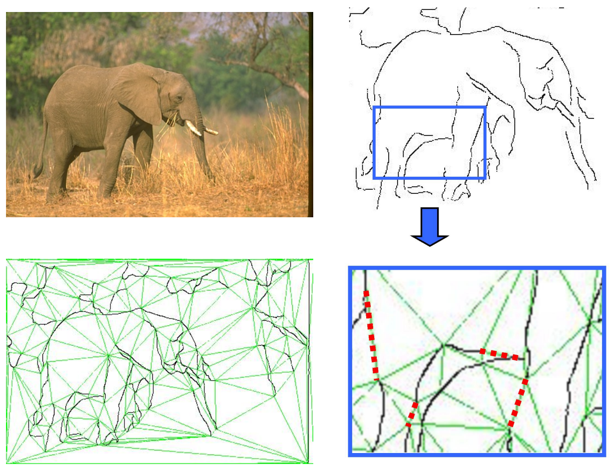


Figure 3: Examples of the CDT triangulation. G-edges (gradient edges detected by Pb) are in black and C-edges (completed by CDT) in green. Note how the CDT manages to complete gaps on the front legs of the elephant (highlighted on the inset at right). These gaps are commonly formed when an object contour passes in front of a background whose appearance (brightness/texture) is similar to that of the object.

plete the gap by connecting the two high-contrast segments. This is closely related to the ligature analysis used in [2].

In order to establish groundtruth labels on the CDT edge for fitting our models, we need to transfer human marked boundaries from the pixel grid. This is accomplished 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.

Figure 4 shows the performance of the local boundary detector Pb as well as the performance when we assign the average underlying Pb to each CDT edge.[1] These results clearly show that moving from pixels to the CDT completion seldom gives up any boundaries found by the local measurement. The green curve documents the soft groundtruth labellings of the CDT edges (percentage matched to human marked pixels). This is the target output of the two continuity models we describe next. The gap between the asymptotic recall of Pb and the ground-truth shows that the CDT is even completing a few contours which are completely illusory (i.e. contours with no local evidence).

[1]Throughout 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, above threshold boundary points 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 [15] for more discussion).

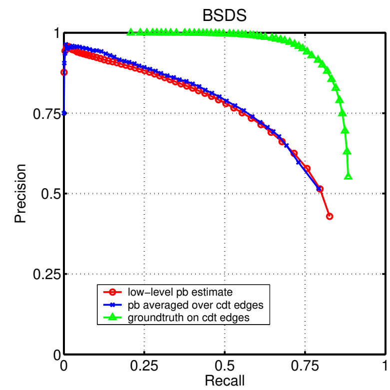


Figure 4: This Precision-Recall curve verifies that moving from pixels to the CDT completion doesn't give up any boundaries found by the local measurement. For comparison, we show the upper-bound performance given by the training data on the CDT edges. The upper bound curve has a precision near 1 even at high recall and achieves a greater asymptotic recall than the local boundary detector, indicating it is completing some gradientless gaps.

We have experimented with other alternative schemes of completion. For example, one way to complete potential gaps is to allow each vertex to connect to its k nearest neighbors, subject to visibility. This scheme adds many more completions than the CDT, without achieving any improvement in performance.

3. 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 inference about $\{X_e\}$.

We will introduce two models: one of local continuity and one of global continuity. The local continuity model considers only the local context of each e; each X_e is estimated independently in 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* [13] to build a joint probabilistic model over all edges. We estimate the marginals of X_e using loopy belief propagation.

3.1 A local continuity model

Each edge e in a CDT graph is naturally associated with a set of features including the average Pb of 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 (Figure 5(a)). Each edge can be on or off (being a true boundary or not), hence four possible labelings of this pair. However, the groundtruth 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 labelings: either both of them are on, or both are off.

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

To evaluate the local continuity model, we use the classifier to assign a new "probability of boundary" value to each edge e. Consider Figure 5(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. Recall $X_e = 1$ 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. If S_1 and S_2 are the two sets of edges connecting to e_0 at the two ends we define the new boundary operator Pb_L under the 2-edge product model to be:

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

The quantitative performance evaluation of Pb_L against Pb is shown in Section 4.

To evaluate relative contribution of each feature, we fit classifiers to subsets of features and compared their performance. The results on the baseball player dataset are shown in Figure 6. Performance is evaluated with both a precision-recall curve and a cross-entropy loss L [9].

We observe that $\overline{Pb}$ is the most useful feature with $L = 0.418$. Angle θ by itself is not as useful; however, when combined with $\overline{Pb}$, θ helps reduce the loss to 0.385. G by itself is quite informative, as 85% of the positive examples have $G = 1$ (both edges being G-edges). When the three features are combined, the loss is reduced to 0.374. We have experimented with many additional features but they are at most marginally useful, hence not included in the local model.

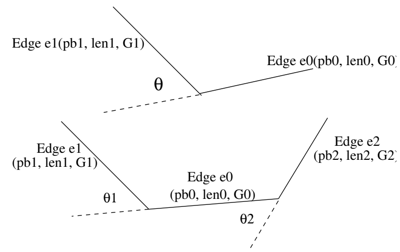


Figure 5: **(a)** A simple 2-edge model of local curvilinear continuity. each edge has an associated set of features. continuity is measured by the angle θ. **(b)** evidences of continuity come from both ends of edge e_0. The new"probability of boundary" for e_0 is the product of the 2-edge model on both pairs (e_0, e_1) and (e_0, e_2).

We also considered several variants such as: a second-layer classifier 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; and a full 4-way classification on each pair of edges. The simplest 2-edge product model described above performs as well as any of these variants.

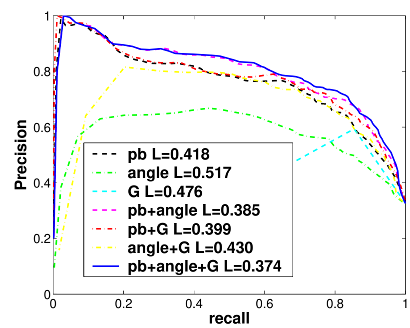


Figure 6: Evaluating combinations of features with precision-recall curves and cross-entropy loss L on the set of baseball player images. $\overline{Pb}$ is the most powerful single feature. The addition of continuity θ significantly improves the performance.

3.2 A global continuity model

In order to capture longer range dependencies of boundary presence we also consider a global probabilistic model built on top of the CDT. We utilize the same measurements as in the local-model (Pb,G,θ) along with the frequency with which junctions of different degrees appear. To integrate these features across the image, we use a conditional random field (CRF) model as first introduced by [13].

Unlike the MRF models traditionally used in vision which model the joint distribution of image measurements and hidden labels, a CRF focus directly on the conditional distribution of labels given the observations. One key ad-

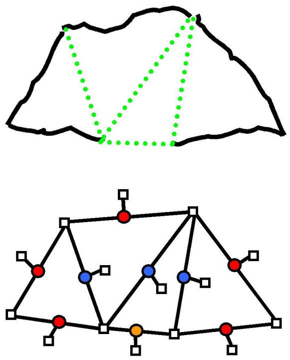


Figure 7: The factor graph representing our conditional random field model for global continuity. The CDT provides both the geometric and graphical structure for our model. Each edge in the CDT corresponds to a variable X_e (represented by circles) which is 1 if pixels corresponding to edge e constitute a true boundary. Potential functions (squares) consist of a singleton potential for each edge encodes the underlying Pb measurement and continuity/closure potentials at the vertices whose values are dependent on both the angles between incoming segments and the numbers of C- and G-edges entering a junction.

vantage 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 proposed as a method for image segmentation by[12, 23, 10], however, the focus is still on pixel-level labeling which is tied to the resolution of the pixel grid. Moving to scale free, mid-level tokens is our key ingredient to handling objects at a continuum of scales. For example, the "multi-scale" approach of [10] 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 scale-invariant a simple random field would suffices.

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

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

The ϕ potential function captures the extent to which the image evidence I supports the presence of a boundary under edge e. ψ describes the continuity conditions at a junction between contour segments. Θ is the collection of model parameters.

Our edge potential is given by

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

where Pb_e is the average Pb recorded over the pixels corresponding to edge e. The vertex potential is given by

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

where $\deg_g$ is the number of G-edges at vertex V for which $X_e = 1$ and similarly $\deg_c$ is the number of C-edges which are turned on.

(deg_g, deg_c) determines the "junction type" for every possible assignment of X_v: for example, a total degree of 1 $((deg_g + deg_c) = 1)$ corresponds to a line ending, 2 a continuation, and 3 a T-junction. The model associates a seperate weight $\alpha_{i,j}$ to each possible type. When the total degree of a vertex is 2, γ weights the continuity of the two edges.

θ is the angle formed by the two edges turned on, and f is a smooth function which is symmetric around $\theta = 0$, falling off as $\theta \to \pi$. If the angle between the two edges is close to 0, they form a good continuation, $\gamma f(\theta)$ is large and they are more likely to both be turned on.

Our model has the collection of parameters $\Theta = \{\alpha, \beta, \gamma\}$. We fit Θ by maximizing the log likelihood. Since the likelihood is log-linear in the parameters, taking a derivative always yields a difference of two expectations. For example, the derivative with respect to the continuation parameter γ for a single training image/ground truth labeling, (I, X) is:

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

where we use 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 which is the expectation under the model given our current setting of the parameters. When the model expectations match those observed in the training data, we have found the maximum likelihood setting of the parameters. Until we reach that point, we take a small step in the gradient direction. Parameters typically converge after a few hundred iterations.

Unfortunately, computing the expectations of our features with respect to model parameters is intractable. 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). We approximate the edge and vertex degree expectations using loopy belief propagation[26]. Loopy belief propagation has been applied to many vision problems before, typically on the pixel-grid. The connectivity of the CDT graphs we study is more complicated than a regular grid. However, in our experiments belief propagation appears to converge quickly (< 10 iterations) to a reasonable solution.

We find that the parameters learned from groundtruth boundary data match our intuition well. For example, the weight $\alpha_{1,0}$ is much 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.

Once the model has been trained, we use belief propagation to estimate the marginal distributions $\{X_e\}$ on the edges of the CDT and then project these down to the pixel grid where it can then be compared to both the local model and the raw Pb.

4. Results: Is curvilinear continuity useful?

We have described two different algorithms, each which outputs 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 [17] split into 15 training and 15 testing images, images of horses [4] split into 172 training and 172 test images, and the Berkeley Segmentation Dataset [1] (BSDS) 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 [15]. Figure 8 shows the precision-recall curves for each data set. 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 Figure 9. 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.

The remaining gap in performance between our model and the upper bound may best be bridged by 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.

5. Conclusion

We have described two stochastic models of curvilinear continuity which have a verifiably favorable impact on the problem of boundary detection. The local model of curvilinear continuity, though quite simple, yields a significant performance gain. The global model, by making long-range inference 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 and the constrained Delaunay triangulation. This provides a scale-invariant geometric representations of images which tends to complete gaps in the low-level edge map. Moving 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 quantitative comparisons between algorithms. Many of the latest exciting results in object recognition and stereo reconstruction could not have occurred without the firm grounding in quantitative evaluation. Perceptual organization should be held to the same high standards.

Acknowledgments. This research was partially supported by ONR grant N00014-01-1-0890 (MURI).

References

[1] Berkeley segmentation dataset, 2002. http://www.cs.berkeley.edu/projects/vision/bsds.

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

[3] S. Belongie, J. Malik, and J. Puzicha. Matching shapes. In *ICCV*, volume 1, pages 454–461, July 2001.

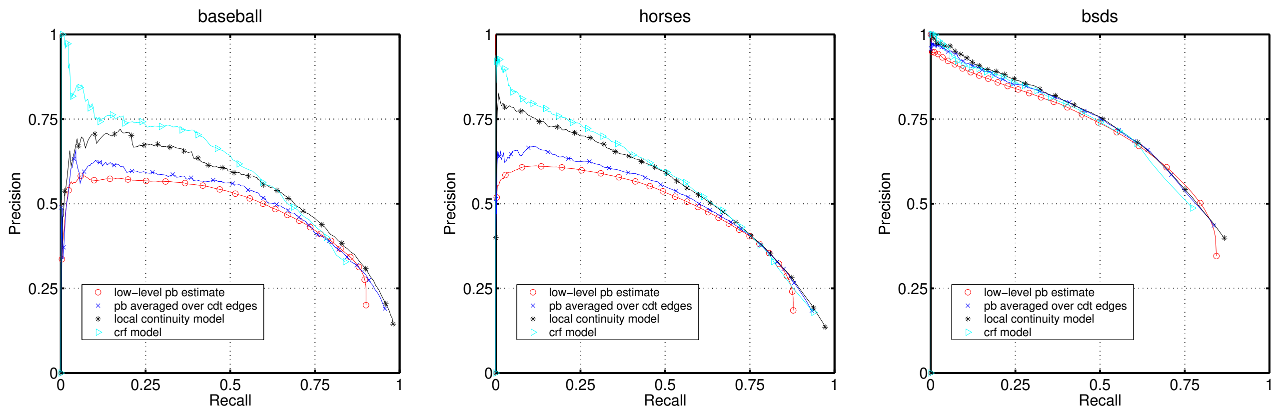


Figure 8: Pixel-based precision-recall evaluations comparing the local classifier (Pb_L), global CRF (Pb_G) and raw Pb. Both techniques improve boundary detection on all three datasets and the overall ordering of the curves is generally preserved.

[4] E. Borenstein and S. Ullman. Class-specific, top-down segmentation. In *ECCV*, volume 2, pages 109–124, 2002.

[5] G. Borgefors. Hierarchical chamfer matching: A parametric edge matching algorithm. *IEEE Trans. PAMI*, 10(6):849–865, 1988.

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

[7] J. Elder and S. Zucker. Computing contour closures. In *ECCV*, volume I, pages 399–412, 1996.

[8] W. S. Geisler, J. S. Perry, B. J. Super, and D. P. Gallogly. Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research*, 41:711–724, 2001.

[9] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: data mining, inference and prediction*. Springer-Verlag, 2001.

[10] X. He, R. Zemel, and M. Carreira-Perpinan. Multiscale conditional random fields for image labelling. In *CVPR*, volume 2, pages 695–702, 2004.

[11] D. Huttenlocher, G. Klanderman, and W. Rucklidge. Comparing images using the Hausdorff distance. *IEEE Trans. PAMI*, 15(9):850–863, 1993.

[12] S. Kumar and M. Hebert. Discriminative random fields: A discriminative framework for contextual interaction in classification. In *ICCV*, pages 1150–1159, 2003.

[13] J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proc. 18th International Conf. on Machine Learning*, 2001.

[14] S. Mahamud, L. Williams, K. Thornber, and K. Xu. Segmentation of multiple salient closed contours from real images. *IEEE Trans. PAMI*, 25(4):433–444, 2003.

[15] D. Martin, C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using brightness and texture. In *Advances in Neural Information Processing Systems 15*, 2002.

[16] D. Martin, C. Fowlkes, L. Walker, and J. Malik. Local boundary detection in natural images: Matching human and machine performance. In *European Conference on Visual Perception [Perception, 32 supp, p. 55]*, 2003.

[17] G. Mori, X. Ren, A. Efros, and J. Malik. Recovering human body configurations: Combining segmentation and recognition. In *CVPR*, volume 2, pages 326–333, 2004.

[18] D. Mumford. Elastica and computer vision. In C. Bajaj, editor, *Algebraic Geometry and Its Applications*, pages 491–506. Springer Verlag, 1994.

[19] S. Palmer. *Vision Science: Photons to Phenomenology*. MIT Press, 1999.

[20] P. Parent and S. Zucker. Trace inference, curvature consistency, and curve detection. *IEEE Trans. PAMI*, 11(8):823–39, 1989.

[21] X. Ren and J. Malik. A probabilistic multi-scale model for contour completion based on image statistics. In *ECCV*, volume 1, pages 312–327, 2002.

[22] A. Shashua and S. Ullman. Structural saliency: the detection of globally salient structures using a locally connected network. In *ICCV*, pages 321–7, 1988.

[23] N. Shental, A. Zomet, T. Hertz, and Y. Weiss. Learning and inferring image segmentations with the gbp typical cut algorithm. In *ICCV*, pages 1243–1250, 2003.

[24] J. Shewchuk. Triangle: Engineering a 2d quality mesh generator and delaunay triangulator. In *First Workshop on Applied Computational Geometry*, pages 124–133, 1996.

[25] Z. Tu, X. Chen, A. Yuille, and S. Zhu. Image parsing: segmentation, detection, and recognition. In *ICCV*, pages 18–25, 2003.

[26] Y. Weiss. Correctness of local probability propagation in graphical models with loops. *Neural Computation*, pages 1–41, 2000.

[27] L. Williams and D. Jacobs. Stochastic completion fields: a neural model of illusory contour shape and salience. In *ICCV*, pages 408–15, 1995.

[28] L. Williams and K. Thornber. A comparison of measures for detecting natural shapes in cluttered backgrounds. *Int'l. Journal of Computer Vision*, 34(2/3):81–96, 1999.

[29] Q. Wu and Y. Yu. Two-level image segmentation based on region and edge integration. In *Proc. DICTA*, pages 957–966, 2003.

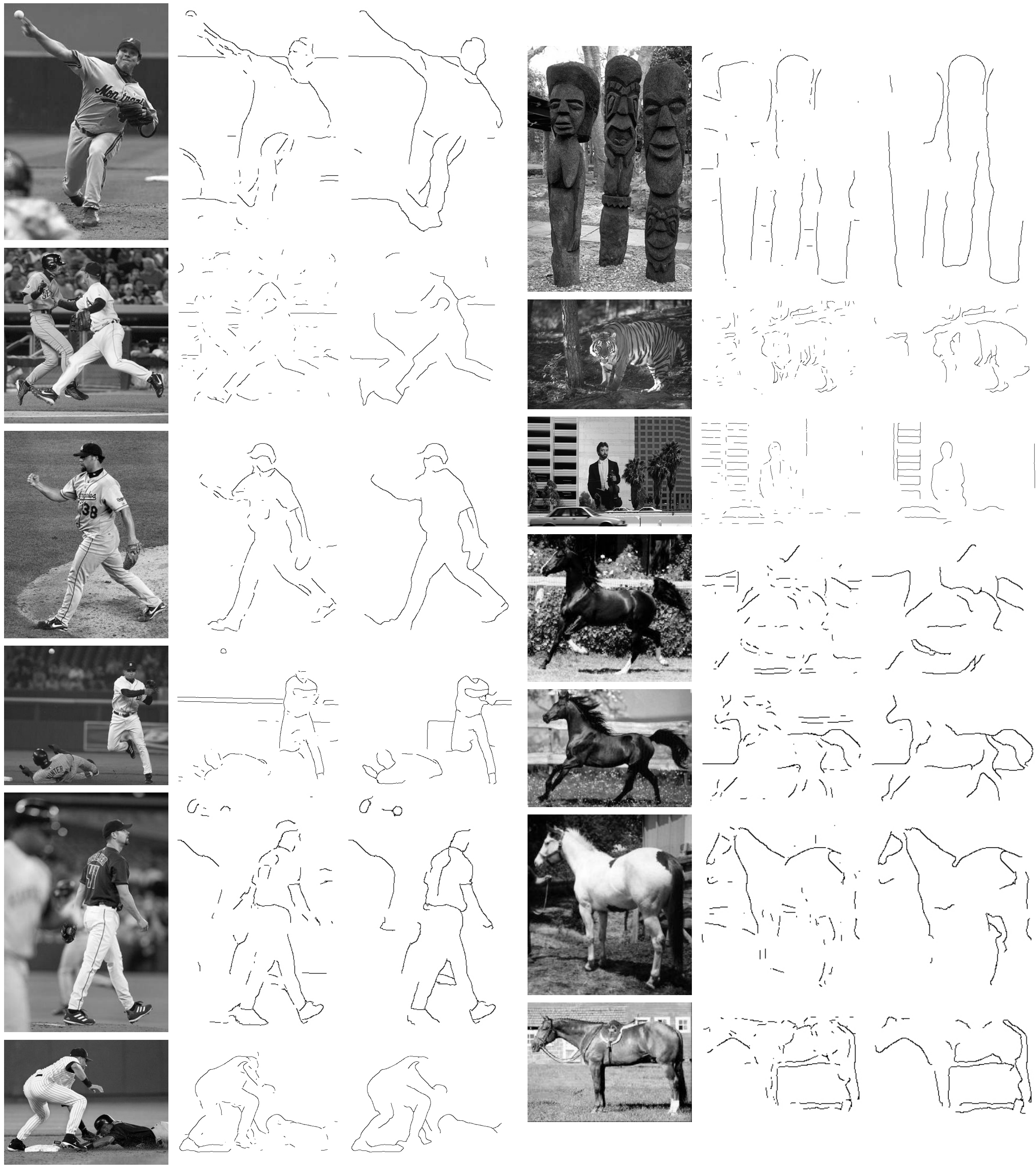


Figure 9: Example results on the three data sets. The two columns of edge maps show the local boundary detector Pb output and the CRF model respectively. The algorithms outputs have been thresholded at a level which yields 2000 boundary pixels for the baseball/BSDS images and 1000 pixels for the smaller horse images.