Local figure–ground cues are valid for natural images

Charless C. Fowlkes   Department of Electrical Engineering and Computer Science, University of California at Berkeley, Berkeley, CA, USA

David R. Martin   Computer Science Department, Boston College, Boston, MA, USA

Jitendra Malik   Department of Electrical Engineering and Computer Science, University of California at Berkeley, Berkeley, CA, USA

Figure–ground organization refers to the visual perception that a contour separating two regions belongs to one of the regions. Recent studies have found neural correlates of figure–ground assignment in V2 as early as 10–25 ms after response onset, providing strong support for the role of local bottom–up processing. How much information about figure–ground assignment is available from locally computed cues? Using a large collection of natural images, in which neighboring regions were assigned a figure–ground relation by human observers, we quantified the extent to which figural regions locally tend to be smaller, more convex, and lie below ground regions. Our results suggest that these Gestalt cues are ecologically valid, and we quantify their relative power. We have also developed a simple bottom–up computational model of figure–ground assignment that takes image contours as input. Using parameters fit to natural image statistics, the model is capable of matching human-level performance when scene context limited.

Keywords: figure–ground, ecological statistics, natural scenes, perceptual organization

Introduction

In the 1920s, the Gestalt psychologists identified grouping and figure–ground as two major principles underlying the process of perceptual organization. Grouping describes the way that individual elements of a stimulus come together to form a perceptual whole. Figure–ground refers to the perception that a contour separating two regions "belongs" to one of the two regions. The figural region takes on shape imparted by the separating contour and appears closer to the viewer, whereas the ground region is seen as extending behind the figure. Both grouping and figure–ground are thought to be important in reducing the visual complexity of a scene to that of processing a small number of cohesive, nonaccidental units.

Starting with Rubin (1921), who first pointed out the significance of figure–ground organization, a long list of factors that affect figure–ground assignment have been identified. These include size, surroundedness, orientation, and contrast (Rubin, 1921), as well as symmetry (Bahnsen, 1928), parallelism (Metzger, 1953), convexity (Kanizsa & Gerbino, 1976; Metzger, 1953), meaningfulness (Peterson, 1994), and lower region (Vecera, Vogel, & Woodman, 2002).

How might these cues be computed in the brain? It is conceivable that cues such as orientation or contrast could be a function of information present in receptive fields local to a given contour, but other cues like symmetry and parallelism would seem to require long-range lateral interactions. Evidence from electrophysiology (Qiu & von der Heydt, 2005; Zhou, Friedman, & von der Heydt, 2000) suggests the existence of cells in V2 that code for contour ownership within 10–25 ms of the onset of response activity. The early availability of this signal (both in terms of time course and stage in the visual pathway) provides strong support for the role of local bottom–up cues, contrasting with the traditional Gestalt emphasis on global organization. On the other hand, studies on meaningfulness (Peterson, 1994) show that contours tend to associate with the abutting region that has a familiar shape, pointing to the integration of top–down knowledge.

Although there is little doubt that both local and global information sources play a role in figure–ground processing, a key problem is understanding their relative importance. Identifying cues, even in physiological detail, does not provide an explanation as to why they exist or how conflicting cues might be fused to yield a cohesive percept. More than 50 years ago, Egon Brunswik (Brunswik & Kamiya, 1953) suggested a solution to these concerns, namely, that the Gestalt cues reflect the statistics of the natural world in which the visual system evolved. He proposed that Gestalt cues be validated by studying the statistics of natural scenes and carried this agenda out on a limited scale.
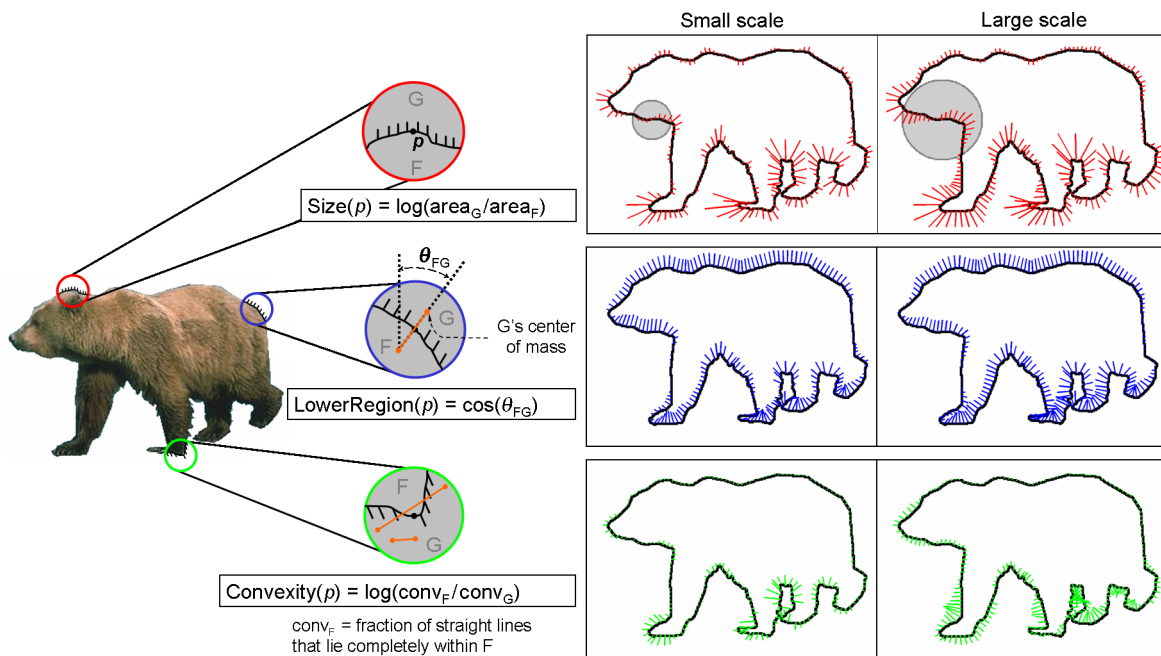
Figure 1. Formulating local cues to figure–ground assignment. Three cues are defined locally inside an analysis window centered at a contour point p. The Size cue describes the relative size of the neighboring regions. LowerRegion compares the relative locations of the center of masses of the two regions. The Convexity cue captures the relative convexity of the two neighboring regions. Convexity is defined as the probability that a line segment connecting two points in a region lies completely within the region. The six panels at the right demonstrate the information captured by each cue at two different scales. The base of each colored line segment along the boundary marks the point on the contour at which the cue was computed and points towards the predicted ground region. The length of the line indicates the relative magnitude of the cue. The cues of size, lower region, and convexity are indicated with red, blue, and green, respectively.

In this article, we pursue such a strategy to understand how well local bottom–up cues predict figure–ground relations in natural scenes. We study the ecological statistics of size, convexity, and lower-region cues using a large collection of natural images for which true figure–ground relations have been assigned by human observers. Measuring the frequency with which these features make the correct prediction provides an ecological validation of the corresponding Gestalt cues. We describe a simple computational model that combines these three cues to predict figure–ground relations. We find that the model performs as well as human subjects asked to make similar local judgments. We also highlight the importance of figure–ground information contained in the image luminance, which lies outside the scope of classical configural shape cues. Together, these results provide a first quantitative measurement of the relative power of local and global cues in figure–ground assignment.

Methods

Formulating cues to figure–ground

We formulate the cues of size, lower region, and convexity as functions of the boundary shape between two regions inside a local analysis window centered at a contour point p. Size(p) is defined as the log ratio of the areas of the two regions. LowerRegion(p) is defined as the cosine of the angle between the line connecting the center of masses of the two regions and the vertical direction given by the camera orientation. In contrast to measuring the angle of the boundary tangent at the center point, LowerRegion(p) incorporates information over the entire analysis window. The convexity of a single region is computed as the fraction of point pairs in a region for which a straight line connecting the two lies completely within the region. Convexity(p) is then given by the log ratio of the two region convexities.

In natural scenes, an object may appear at any distance from the viewer. As a result, a window subtending a fixed visual angle may include an entire object at a distance or cover only a small uninformative portion of a nearby object boundary. To provide an intuitive notion of context that is independent of the scale at which an object appears in a scene, we specify the analysis window radius as a percentage of the arc length of the underlying contour on which it is centered. This makes the local cues we measure approximately invariant to an object's distance from the viewer.

Figure 1 provides a graphical description of each cue computation and shows the cue response along the
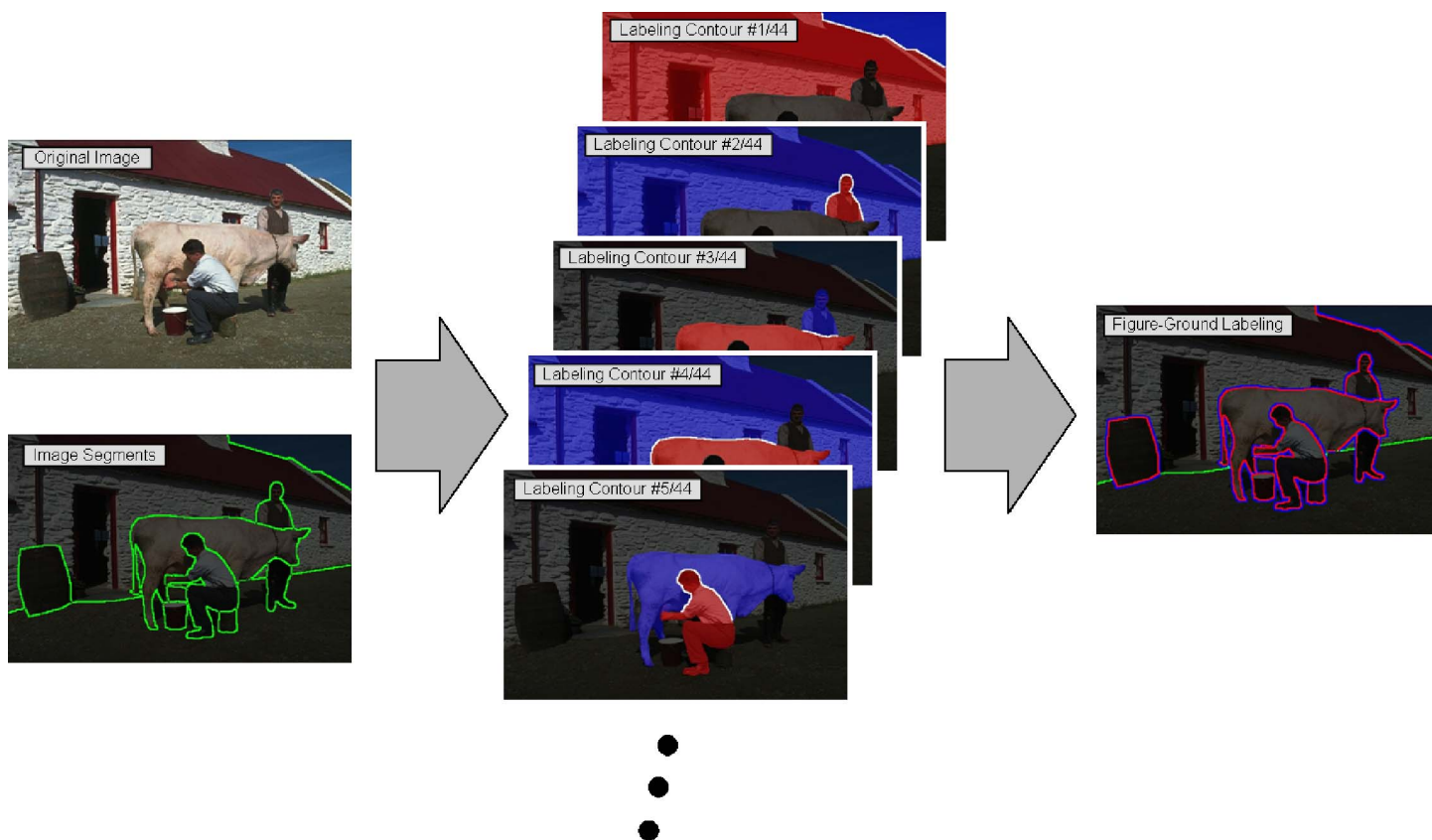
Figure 2. Acquiring figure–ground labels. Human subjects labeled each contour in an image, indicating to which region it "belongs". Starting from a segmentation of the original image (left), subjects were presented with a sequence of highlighted contours corresponding to each pair of neighboring regions (center). The subject indicated which of the two regions was the figural element. The reported figural region is displayed here with a red tint, ground with a blue tint. Subjects also had the option of attributing a boundary to a change in surface albedo or a discontinuity in the surface normal. Such a boundary, exemplified by the corner between the building and earth, marked in green, was seen as belonging to both segments. Once all the contours had been labeled, the subject was presented with the final labeling (right) and given the opportunity to fix any mistakes.

boundary of a test figure at two different scales of analysis. These local cues are not always in agreement with our global percept. Along the top of the bear's nose, all three cues correctly predict the bear-shaped segment as figural, whereas at the bottom of the bear's leading foot, Size and Convexity correctly indicate the foot as figural, but LowerRegion gives a contradictory response. At a small scale, Convexity suggests that the space between the bear's legs is figural but reverses at a larger scale.

Acquiring ground-truth labels

To understand how often each cue provides the correct prediction, we compiled ground-truth figural assignments for contours in a collection of 200 images depicting a wide variety of indoor and outdoor scenes containing manmade and natural objects, including humans and other animals. These 200 images were chosen at random from the set of 1,000 hand segmented images in the Berkeley Segmentation DataSet (Martin, Fowlkes, Tal, & Malik, 2001). A segmentation of each image was selected randomly from the set of five available color segmentations to provide ground-truth contour locations. The images and annotations are available online (http://www.cs.berkeley.edu/projects/vision/grouping/segbench).

Human subjects were asked to indicate, for each pair of abutting segments in an image, which region was figural and which was ground. Figure 2 shows a typical example of the labeling process where a subject has assigned each contour in turn yielding a complete labeling of the image. Subjects also had the option of indicating that a given contour was the result of a change in albedo or surface normal and hence "belonged" to both neighboring regions.

The figure–ground labeling was carried out by 10 subjects who were naive to the experimental purpose. Each of the 200 segmented images was presented at a resolution of 481 × 321 pixels on a typical computer monitor. Subjects were not asked to label contours whose length was less than 2% of the image diagonal. There were no time constraints imposed on the labeling task.
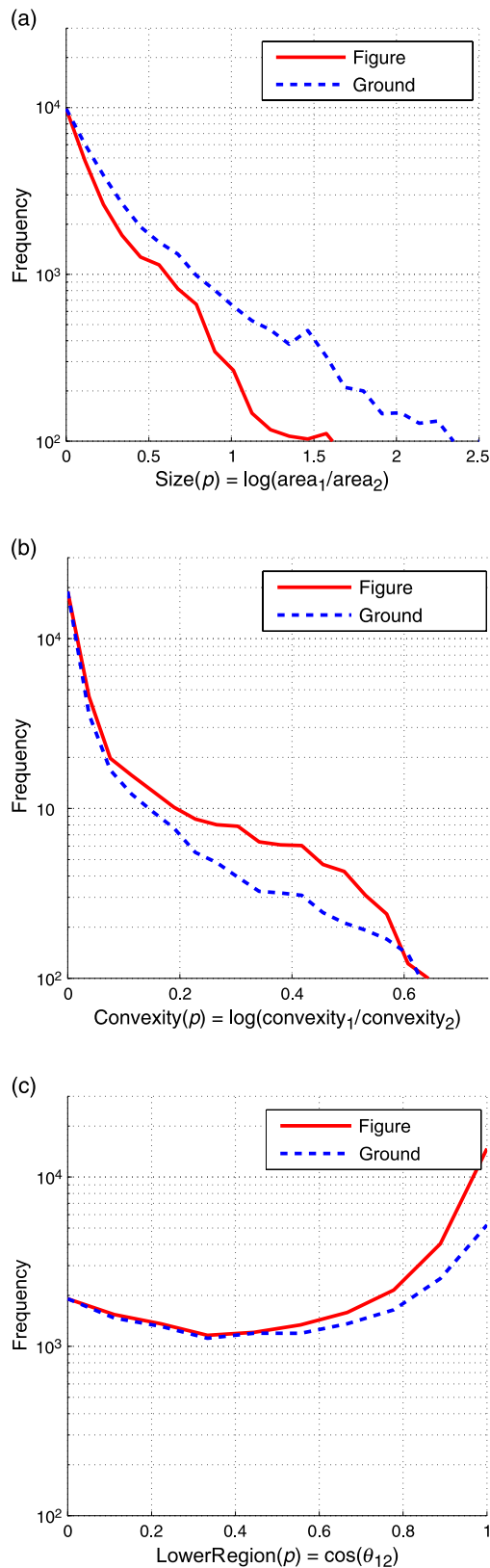
(a)

(b)

(c)

Figure 3. The statistics of local figure–ground cues in natural scenes. Each histogram shows the empirical distributions of Size(p), LowerRegion(p), and Convexity(p) for 50,000 points sampled from human-labeled contours in 200 natural images computed over a window with radius $r = 5\%$ contour length.

Each contour in the dataset was labeled by two different human observers. A consistency check of the human labels shows that observers agreed on the figure–ground/albedo labeling for 83.9% of the contour points sampled. Of the remaining 16.1%, 12.3% involved one subject marking figure–ground and the other marking albedo, whereas 3.7% had conflicting figure–ground assignments.

For use in the following experiments, 10% of the contour points were sampled from the 400 labelings, for a total of 285,000 points. We did not utilize those 40,000 points that were labeled as lying on albedo boundaries. Points that were within two thirds of the largest analysis window radius from a junction between three or more segments or an image border were also excluded, leaving 50,000 points for analysis at all scales. On this restricted set of points, humans agreed on the labeling 96% of the time. Inconsistently labeled points were included in our analysis with the label chosen randomly, establishing an upper bound of 96% on classification accuracy.

Results

Figure 3 shows the empirical distribution of cue responses at a single scale ($r = 5\%$ contour length) for 50,000 points sampled from the human-labeled boundaries. We plot only distributions for positive values of each cue. Because every boundary point contributes two values of equal magnitude and opposite sign, the distributions of negative values are identical with the roles of figure and ground reversed. Note that the marginal distribution of contour orientations is not uniform. The greater prevalence of horizontal (LowerRegion = 1) and vertical (LowerRegion = 0) boundaries is consistent with previous results on the statistics of brightness edges in natural images (Switkes, Mayer, & Sloan, 1978).

These histograms show that figural regions in natural scenes tend to be smaller, more convex, and lie below the ground regions. For example, when the sizes of the two regions are the same, Size(p) = log(Area$_1$/Area$_2$) = 0, they are equally likely to be figure. When one region is larger, Size(p) > 0, it is more common that the larger region is ground. All three cues uniformly differentiate figure and ground on average, in agreement with psychophysical demonstrations of the corresponding Gestalt cues (Kanizsa & Gerbino, 1976; Metzger, 1953; Rubin, 1921; Vecera et al., 2002). At 5% contour length, we estimate the mutual information (Cover & Thomas, 1991) between each cue and the true label to be 0.047, 0.075, and 0.018 bits for Size, LowerRegion, and Convexity, respectively.

To further gauge the relative power of these three cues, we framed the problem of figure–ground assignment as a discriminative classification task: "With what accuracy can a cue predict the correct figure–ground labeling?"
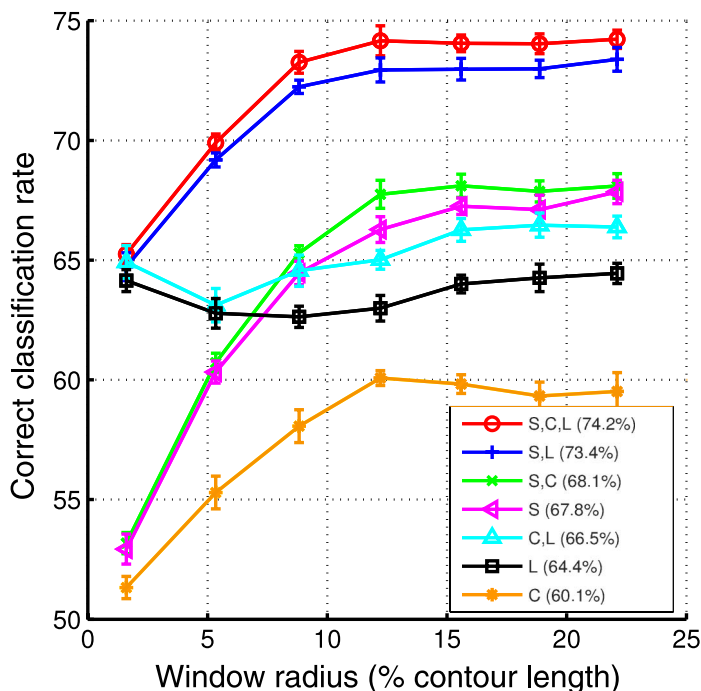
Figure 4. Quantifying the relative power of local figure–ground cues in natural scenes. The power of individual cues and cue combinations is quantified by measuring the correct classification rate, plotted here as a function of window radius. Multiple cues are combined using logistic regression fit to training data. The error bars show 1 *SD* measured over held-out data during 10-fold cross-validation. The legend gives the highest classification rate achieved for each combination of cues. The analysis window radius is measured relative to the length of the contour being analyzed to make it (approximately) invariant to an object's distance from the camera.

For individual cues, it is clear from Figure 3 that the optimal strategy is to always report the smaller, more convex, or lower region as figure. To combine multiple cues, we fit a logistic function,

$$P(\text{figure}|c(p)) = \frac{1}{1 + e^{-\beta^T c(p)}}, \qquad (1)$$

which takes a linear combination of the cue responses at point *p*, arranged into vector $c(p)$ along with a constant offset, and applies a sigmoidal nonlinearity. The classifier outputs a value in [0, 1] that is an estimate of the likelihood that a segment is figural. In the classification setting, we declare a segment to be figure if this likelihood is greater than 0.5. The model parameters β were fit using iteratively reweighted least squares to maximize the training data likelihood (Hastie, Tibshirani, & Friedman, 2001). We also considered models that attempted to

exploit nonlinear interactions between the cues, such as logistic regression with quadratic terms and nonparametric density estimation, but found no significant gains in performance over the simple linear model.

Figure 4 shows the correct classification rate as a function of the analysis window radius for different combinations of cues. Values in the legend give the best classification rate achieved for each combination of cues. The performance figures suggest that all three cues are predictive of figure–ground, with Size being the most powerful, followed by LowerRegion and Convexity. Combining LowerRegion and the Size cues yields better performance, indicating that independent information is available in each. The addition of Convexity when Size is already in use yields smaller performance gains because these two cues are closely related: A locally smaller region tends to be locally convex.

We found that increasing context past 25% contour length did not further improve the model performance. In fact, computing the relative Size, Convexity, and LowerRegion at the level of whole segments (100% context) yielded lower correct classification rates of 56.9%, 55.4%, and 59.5%, respectively. One explanation for the worse performance of global Size and Convexity is that natural scenes typically involve many interacting objects and surfaces. Object A may occlude object B, creating a contour whose local convexity cue is consistent with the figure–ground layout. However, the global convexity of the region composing A may well be affected by its relation to other objects C, D, E, and so forth, in a manner that is largely independent of the figure–ground relation between A and B.

At the most informative window radius, our combined model achieved a 74% correct classification rate, falling short of the human labeling consistency (96%). This is likely due to several sources of information absent from our local model that could have been exploited by human subjects viewing a whole image during labeling. First, integration of local noisy measurements along a contour should yield a consistent label for the entire contour. Our feed-forward approach does not assume that grouping of contours has taken place before figure–ground binding begins. Second, we exclude junctions from our analysis. Junctions embody important information about the depth ordering of regions; however, they are quite difficult to detect locally in natural scenes (McDermott, 2004). Third, human subjects have access to important nonlocal and high-level cues such as symmetry (Bahnsen, 1928), parallelism (Metzger, 1953), and familiarity (Peterson, 1994; Rubin, 1921), which we have not considered here. Lastly, our model only utilizes the shape or configuration of the abutting regions, with no regard to the luminance content associated with each one. This ignores important local photometric evidence such as terminators signaling occlusion (von der Heydt & Peterhans, 1989) and cues to three-dimensional geometry such as texture, shading, and familiarity.
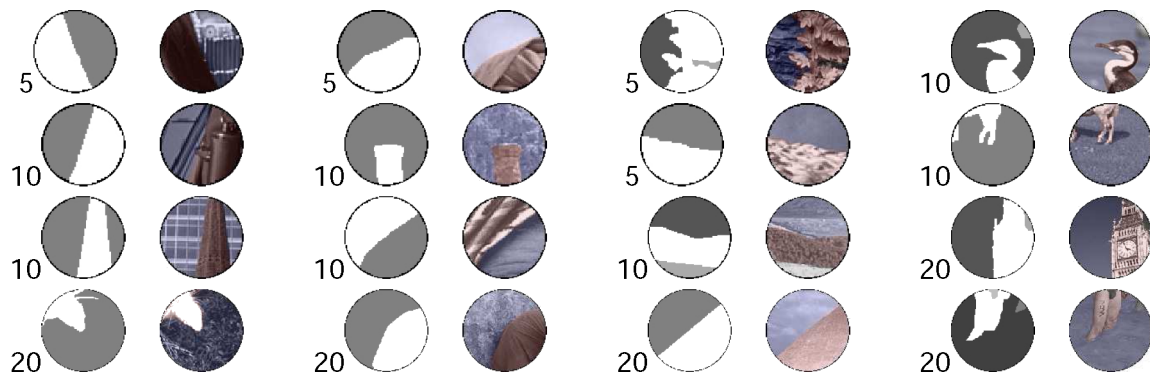
Figure 5. Subjects made figure–ground judgments for local stimuli, like those shown, consisting of a cropped disc depicting either region shape (configuration) or image luminance (configuration + content). In the luminance condition, the two regions on either side of the contour were distinguished by red and blue tints. The color assignments were randomized over trials, but in this figure, the white/red tinted segments indicate which region was figural according to the ground-truth labels. Numbers indicate the window radius for each patch as a percentage of the contour length.

Quantifying the role of local luminance cues

To better understand the role of these unmodeled information sources, a second group of subjects was presented with circular patches extracted from the set of labeled images and asked to indicate which of the two neighboring regions they judged to be figural. Two conditions were used: one in which subjects saw the cropped grayscale image patch and one in which subjects were presented with only the corresponding cropped segment map where each region was filled in with a constant gray level (see Figure 5). These two conditions deprived the observer of global context or both context and luminance content, respectively.

Local patches were displayed through a physical aperture placed in front of a computer monitor. This was done to prevent the aperture from interfering with the perceived shape of the regions being viewed, instead giving the impression that they extended behind the aperture. The aperture size was fixed with respect to the subject (subtending 7 degrees of arc) and image patches scaled to the same presentation size. The subject's head was stabilized with a chin rest 75 cm from the monitor. Exposure times were not limited, but subjects usually spent 1–2 s per patch.

Eight hundred contour points for use in the local patch display experiment were randomly sampled with the same exclusion criteria as used in collecting statistics (described above). RMS contrast for the image patches varied widely from 0.2 to 1.1, depending on the scene. The distribution of contrast over the patches in our dataset was quite consistent with that reported by Mante, Frazor, Bonin, Geisler, and Carandini (2005). Figure–ground was not significantly correlated with brightness. The average brightness of the ground segment was greater than that of the figural segment in 51.5% of the patches sampled.

Each subject labeled all 800 patches, 200 at each of four levels of context: $r = 2.5\%$, 5%, 10%, and 20% contour length. Of the eight subjects, four were presented with

image luminance patches; and four, with segment-only displays. None of the subjects presented with luminance patches had previously seen the images used. For the segment-only display, subjects indicated which side was figural (black or white). In the luminance display, the image patch was overlaid with a red and blue tint to unambiguously specify the contour location.

The gray level or tint assignment was randomized over trials. Subjects showed little bias, choosing the white region in 51.7% of the segment-only trials and choosing the blue tinted segment in 56.0% of the luminance trials. Subjects also showed no significant bias toward the brighter segment in the luminance display, assigning it as figure in 50.5% of the trials.

The resulting local classification performance of human subjects is presented in Figure 6, along with the performance of our local configural model on this subset of patches. We found that, in combination, LowerRegion, Convexity, and Size cues approach human-level performance when only boundary shape information was available. At 20% contour length, human subjects in the configuration-only condition averaged 69% correct classification, whereas the model achieved 68%. Furthermore, labels assigned by human subjects for a given patch agreed quite closely with those of the model. On average, the model prediction matched the subject's response, both correct and incorrect, for 79% of the 800 patches classified. For comparison, pairs of human subjects averaged 75% agreement on the patch labels in the configuration-only condition. Tables 1, 2, 3, and 4 in the Appendix document the performance and level of agreement for individual subjects in both conditions.

Human subjects did make good use of information contained in the image luminance that was not captured by the configural cues. At 20% of the object contour length, access to luminance content decreased the number of errors by more than a factor of 2 over the configuration-only presentation. Performance on luminance patches also
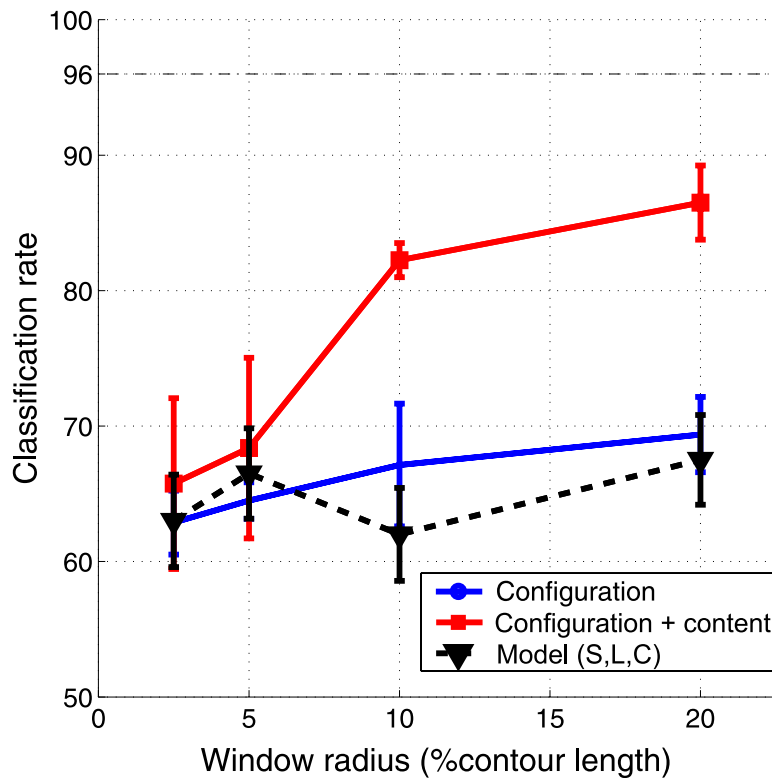
Figure 6. Quantifying the importance of context and content. The correct classification rate and standard deviation across subjects (*n* = 4 subjects in each condition) are plotted as a function of context. We also plot the classification performance of our computational model (S, L, and C) on the same set of local windows, with whiskers marking 1 *SD* of the sample proportion. The grid line at 0.96 indicates the level of global labeling consistency in the ground truth figure–ground assignments.

improved significantly ($p < 2 \times 10^{-4}$ for all subjects) as the window radius increased from 5% to 20% contour length.

Figure 5 shows individual image patches for which the difference in human classification rate without and with luminance information was particularly large. For each of these patches, more than two subjects responded correctly to the luminance content presentation, but more than two responded incorrectly to the configuration-only presentation. The jump in human performance when luminance content is available can be explained by additional local cues exemplified in the patches shown. These include terminators created by occlusion of background texture (first column); three-dimensional shape information available from shadows, shading, and highlights (second column); and recognition of familiar materials or objects based on texture and other internal markings (third and fourth columns, respectively).

Discussion

Taken together, our results provide a quantification of the relative amounts of information about figure–ground

assignment provided by local boundary configuration, local luminance content, and global scene context in natural scenes. In particular, our simple bottom–up model appears to sufficiently capture much of the figure–ground information available from local boundary shape.

Surprisingly, the gap that remains between human performance on local configurations and whole scenes appears to be bridged in large part by exploiting information contained in the local image luminance content rather than global reasoning. Although restricting context prevents the utilization of global configural cues such as parallelism or symmetry, it seems evident from the patches shown in Figure 5a that "high-level" familiarity or meaningfulness can still function locally alongside generic "low-level" cues such as texture, shading, and terminators.

As seen in Figure 1, local figure–ground assignments along a given contour are by no means consistent. It is interesting to consider how pooling local measurements might improve the classification rate by propagating information outwards from zones of high certainty (e.g., Zhaoping, 2005). One difficulty is knowing which local estimates to pool because detecting and grouping contour elements are themselves difficult tasks in natural images. A preliminary study (Ren, Fowlkes, & Malik, 2006) suggests that for those contours that can be detected

locally, pooling measurements yields small but noticeable gains in performance (approximately 5%).

The study of the statistics of natural stimuli has become an increasingly prominent theme in understanding sensory information processing. For example, natural image statistics provide an elegant explanation of the localized receptive fields found in primary visual cortex in terms of optimal coding strategies (Atick & Redlich, 1992; Olshausen & Field, 1996; Ruderman, 1994). The findings described here are more closely related to a smaller body of work, starting with Brunswik and Kamiya (1953), which examines the joint statistics of ground-truth percept and scene measurements pursued in the context of grouping by similarity (Fowlkes, Martin, & Malik, 2003) and contour completion (Elder & Goldberg, 2002; Geisler, Perry, Super, & Gallogly, 2001; Ren & Malik, 2002).

Figure–ground organization has a long history in the field of psychology, where the focus has largely been on identifying which cues impact perception. Our results provide a novel perspective on these findings, offering an explanation as to why such cues exist. An organism that exploits size, lower region, or convexity as a cue to infer figure–ground would have an obvious advantage, more often correctly grasping nearby objects and navigating through gaps rather than colliding with obstacles. Visual theorists (Brunswik & Kamiya, 1953; Gibson, 1979) have sought justification for particular cues in the physical and statistical regularities of the "external world". With the recent availability of large collections of digitized images and the development of statistical learning techniques, such theories are now amenable to direct experimental verification.

Appendix A

The following tables document the agreement between the model and individual human subjects on local patches. Tables 1 and 2 show the correct classification rates for each level of context with configuration only or configuration + content displays, respectively. Tables 3 and 4 document the level of agreement on labels assigned by subjects and the model. The values indicate the percentage of

r	1	2	3	4	M
2.5%	65	65	63	60	63
5%	64	64	67	64	67
10%	63	64	72	71	62
20%	72	67	68	72	68

Table 1. Average correct classification rate (%) on the configural patch display for human subjects (1–4) and the local model (M) at each window radius.

r	5	6	7	8
2.5%	57	70	67	70
5%	60	75	72	68
10%	81	84	82	82
20%	83	89	87	88

Table 2. Average correct classification rate (%) on the configuration + content patch display for human subjects (5–8) at each window radius.

Subject	2	3	4	M
1	75	79	74	80
2		80	69	77
3			74	85
4				74

Table 3. Labeling agreement (%) between human subjects (1–4) and the local model (M) for configural patch displays.

Subject	6	7	8	M
5	74	72	71	67
6		77	78	65
7			76	67
8				65

Table 4. Labeling agreement (%) between human subjects (5–8) and the local model (M) for configuration + content patch displays.

patches for which both labelers selected the same figure–ground assignment, regardless of whether or not it was correct.

Acknowledgments

Corresponding author: Charless C. Fowlkes.
Email: fowlkes@cs.berkeley.edu.
Address: Computer Science Division, Soda Hall, University of California at Berkeley, Berkeley, CA 94720.

References

Atick, J., & Redlich, A. (1992). What does the retina know about natural scenes? *Neural Computation, 4,* 196–210.

Bahnsen, P. (1928). Eine untersuchung über symmetrie und asymmetrie bei visuellen wahrnehmungen. *Zeitschrift für Psychologie, 108,* 129–154.

Brunswik, E., & Kamiya, J. (1953). Ecological cue-validity of "proximity" and of other gestalt factors. *American Journal of Psychology, 66,* 20–32. [PubMed]

Cover, T., & Thomas, J. (1991). *Elements of information theory*. New York: Wiley.

Elder, J. H., & Goldberg, R. M. (2002). Ecological statistics of Gestalt laws for the perceptual organization of contours. *Journal of Vision, 2*(4):5, 324–353, http://journalofvision.org/2/4/5/, doi:10.1167/2.4.5. [PubMed] [Article]

Fowlkes, C., Martin, D., & Malik, J. (2003). Learning affinity functions for image segmentation: Combining patch-based and gradient-based approaches. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2,* 54–61.

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. [PubMed]

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning: Data mining, inference, and prediction*. New York: Springer-Verlag.

Kanizsa, G., & Gerbino, W. (1976). Convexity and symmetry in figure–ground organization. In M. Henle (Ed.), *Vision and artifact* (pp. 25–32). New York: Springer.

Mante, V., Frazor, R. A., Bonin, V., Geisler, W. S., & Carandini, M. (2005). Independence of luminance and contrast in natural scenes and in the early visual system. *Nature Neuroscience, 8,* 1690–1697. [PubMed]

Martin, D., Fowlkes, C. C., Tal, D., & Malik, J. (2001). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. *Proceedings of the IEEE International Conference on Computer Vision, 2,* 416–425.

McDermott, J. (2004). Psychophysics with junctions in real images. *Perception, 33,* 1101–1127. [PubMed]

Metzger, F. (1953). *Gesetze des Sehens*. Frankfurt-am-Main: Waldemar Kramer.

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature, 381,* 607–609. [PubMed]

Peterson, M. A. (1994). Object recognition processes can and do operate before figure–ground organization. *Current Directions in Psychological Science, 3,* 105–111.

Qiu, F. T., & von der Heydt, R. (2005). Figure and ground in the visual cortex: V2 combines stereoscopic cues with gestalt rules. *Neuron, 47,* 155–166. [PubMed] [Article]

Ren, X., Fowlkes, C., & Malik, J. (2006). Figure/Ground assignment in natural images. *Proceedings of the European Conference on Computer Vision, 2,* 614–627.

Ren, X., & Malik, J. (2002). A probabilistic multi-scale model for contour completion based on image statistics. *Proceedings of the European Conference on Computer Vision, 1,* 312–327.

Rubin, E. (1921). *Visuell wahrgenommene figuren*. Kobenhaven: Glydenalske Boghandel.

Ruderman, D. (1994). The statistics of natural images. *Network, 5,* 517–548.

Switkes, E., Mayer, M. J., & Sloan, J. A. (1978). Spatial frequency analysis of the visual environment: Anisotropy and the carpentered environment hypothesis. *Vision Research, 18,* 1393–1399. [PubMed]

Vecera, S. P., Vogel, E. K., & Woodman, G. F. (2002). Lower region: A new cue for figure-ground assignment. *Journal of Experimental Psychology: General, 131,* 194–205. [PubMed]

von der Heydt, R., & Peterhans, E. (1989). Mechanisms of contour perception in monkey visual cortex: I. Lines of pattern discontinuity. *Journal of Neuroscience, 9,* 1731–1748. [PubMed] [Article]

Zhaoping, L. (2005). Border ownership from intracortical interactions in visual area V2. *Neuron, 47,* 143–153. [PubMed] [Article]

Zhou, H., Friedman, H. S., & von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *Journal of Neuroscience, 20,* 6594–6611. [PubMed] [Article]