An Exponential Pyramid-Based Model of Contour Classification in Figure-Ground Segregation

MICHAEL R. SCHEESSELE
Indiana University at South Bend
and
ZYGMUNT PIZLO
Purdue University

In a retinal image, contours belonging to a figure of interest may be intermixed with other contours (caused by occlusion, camouflage, low contrast, etc.), making difficult the identification of figure contours and thus the figure itself. In psychophysical experiments, we found that identification of a figure is facilitated by: differences in relative contour orientation, relative contour curvature, and relative contour length between contours that belong to the figure and those that do not. In short, a difference in *any* contour property seems to facilitate correct identification of a figure. A computational model, based on the exponential pyramid architecture, was constructed and model simulations of several conditions from the psychophysical experiments were performed. A critical aspect of the model is that it performs contour classification by using statistics computed from the entire image. Model simulations accounted well for the results of 11 experimental conditions, using just one free parameter. These results suggest that the human observer uses global features to make local contour classification decisions in the image and that the exponential pyramid architecture can adequately model perceptual mechanisms involved in figure-ground segregation.

Categories and Subject Descriptors: I.4.6 [**Image Processing and Computer Vision**]: Segmentation – *Edge and feature detection*; I.2.10 [**Artificial Intelligence**]: Vision and Scene Understanding – *Perceptual reasoning*; J.4 [**Computer Applications**]: Social and Behavioral Sciences - *Psychology*
General Terms: Theory, Algorithms, Experimentation
Additional Key Words and Phrases: pyramid, computational vision, figure-ground segregation, perceptual organization, Gestalt, psychophysics, contour classification

---

1. INTRODUCTION

Figure-ground segregation refers to the ability of the human visual system to: (a) group one or more regions in the retinal image that potentially correspond to an object and (b) separate this grouping from other regions which compose the background or which form other objects. (A closely related problem in computer vision is *image segmentation*, where the goal is to partition a digital image into regions that may correspond to recognizable objects.) From the demonstration of Rubin [1915/1958] to the present, understanding how figure-ground segregation works has been one of the fundamental

Authors' addresses: Michael R. Scheessele, mscheess@iusb.edu, Department of Computer and Information Sciences, Indiana University at South Bend, South Bend, IN 46634; Zygmunt Pizlo, pizlo@psych.purdue.edu, Department of Psychological Sciences, Purdue University, West Lafayette, IN 47907

challenges of perceptual psychology. The human visual system is quite effective at performing figure-ground segregation. This is certainly true when the visual scene contains just a single object in front of a homogeneous background, but it is also true for more complex images. Consider Figure 1 – it is not difficult to perceive the figure (a car) although it is partially occluded by a fence (as is a second piece of machinery) and although the background is composed of varying textures corresponding to pavement, grass, trees, bushes, and part of a building. Figure 2 may be more challenging, yet it is possible to perceive the figure (a Dalmatian), even though the region corresponding to the Dalmatian and the region corresponding to the background are both composed of black and white patches.

Figure 1. Fence partially occludes car in visual scene.

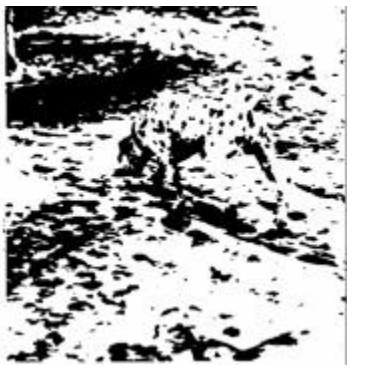


Figure 2. A Dalmatian. Image by R.C. James.

Early attempts [Rubin 1915/1958; Koffka 1935] to explain figure-ground segregation were largely descriptive; unfortunately there has been little progress since then towards an explanation [Palmer 1999]. Nevertheless, a typical presumption has been that the Gestalt 'laws of perceptual organization' [Wertheimer 1923/1958] play a role in figure-ground segregation. The Gestalt 'laws' demonstrate conditions under which distinct elements in a retinal image may be grouped together to form larger structures (e.g., perhaps even a figure of interest). *Contours* represent one possible type of 'element' in a retinal image[1]. Contours in a retinal image arise from figure boundaries in the visual scene (but may also be due to shadows, background textural elements, etc.). Understanding how pieces of image contour 'go together' to form larger contours is thus likely beneficial to understanding how a figure in the retinal image may be perceived.

There are a number of Gestalt laws of perceptual organization – several will now be described. The *law of proximity* states that, other things equal, elements near to one another in an image have a greater tendency to be grouped together than if they were farther apart. The *law of similarity* states that similar elements tend to be grouped

together. The *law of good continuation* states that contour groupings where resulting curves are smooth are preferred over those where resulting curves abruptly change direction. The *law of closure* refers to the tendency to perceive an image region enclosed by a contour as a figure, separate and somehow different from the surrounding background, even though the regions inside and outside of the contour may be of the same substance. This can occur even though the contour may be somewhat fragmented [Koffka 1935]. While original formulation of the Gestalt laws largely relied on demonstrations [Wertheimer 1923/1958; Koffka 1935], recent work has aimed at providing more rigorous quantitative explanations of these principles with regards to contour grouping (with respect to *closure*, see Elder and Zucker [1994]; with respect to *good continuation*, see Pizlo, Salach-Golyska, and Rosenfeld [1997]; Feldman [2001]; Geisler, Perry, Super, and Gallogly [2001]; Elder and Goldberg [2002]; with respect to *similarity* and *proximity*, see Elder and Goldberg [2002]). However, this recent work has tended to focus on how local pieces of contour are grouped into larger pieces of contour. With complex images, figure-ground segregation may require more than just *local* grouping of pieces of contour. It may also require that all visible contours of a figure be *classified* from those that do not belong to the figure (e.g., those belonging to the ground or to other figures). Such a requirement may entail *global* processing. Indeed, the enigmatic Gestalt idea that something about a whole guides how the parts are grouped into that whole [Wertheimer 1923/1958; Koffka 1935] anticipates this.

The task of classifying contours that belong to a figure from those that do not may be especially challenging when *occlusion* is present. Consider Figure 1 again. Notice that the fence not only interrupts the contours of the car, but it also introduces contours in the image where parts of the car would otherwise appear were it not for the fence. The task of classifying contours that belong to a figure may also be more difficult when the figure appears to be fragmented or degraded, but with no (obvious) occlusion present. One example is *camouflage*, where the camouflage agent may break up the natural contours of a figure, while simultaneously adding new contours within the boundaries of the figure. Another example is *low contrast* between parts of a figure and the background. For instance, when the background is dark and parts of a figure are covered by shadows, those parts may appear to blend in with the background, resulting in omission from the image of sections of figure contour and introduction of spurious new contours that correspond to shadow boundaries.

The occlusion literature, in particular, has long recognized the challenge posed by image contours that do not belong to a figure of interest – specifically image contours introduced by an occluder [Nakayama, Shimojo, and Silverman 1989; Kellman and Shipley 1991; Brown and Koch 1993; Grossberg 1994]. For example, a crucial component of the theory of Nakayama et al. [1989] is that the contours "intrinsic" to an occluded figure must be discriminated from those "extrinsic" to it (i.e., those belonging to the occluder). Once the image contours intrinsic to the figure are classified as such, they can be further subjected to grouping before being ultimately submitted to a recognition process. Their theory claimed that this classification of image contours into intrinsic (belonging to figure) versus extrinsic (belonging to occluder) is performed on the basis of relative *depth* of the two types of contours in the visual scene. To illustrate, assume an observer is viewing the scene depicted by Figure 1. The retinal image produced would contain contours of the occluding fence that intersect with contours of the occluded car. They point out that in nature an occluder is always closer to the observer than the figure which it occludes. They argue that such depth information is available early in vision and that it is used by the human visual system to classify the contours of the occluded figure apart from those of the occluder. (In cases where the 'visual scene' is simply a picture, they conjecture that T-junctions provide the depth information.) While it seems plausible with respect to occlusion, this theory (and other contour-based theories of the perception of partially occluded figures) cannot account for perception of degraded or fragmented figures (see Figure 2 for an example). In such cases, there is typically no depth information with which to classify contours as belonging or not to the figure. Yet, it is often possible to perceive such partially visible figures. Perhaps then there are other contour properties, in addition to depth, that enable the human visual system to classify image contours as belonging or not to a figure.

Contour classification is just one task that may be required for perception of partially visible figures. Another is contour *interpolation*. Observe that when a figure is partially visible, portions of its bounding contour are typically missing and therefore must be interpolated. Shipley and Kellman [1992] proposed that the perceptual strength of interpolation is a function of the 'support ratio' of the resulting contour. Support ratio is defined as the total length of physically present pieces of a contour divided by total length of the physically present and interpolated pieces of a contour. This idea resonates with common sense and indeed support ratio is commonly accepted (but see Singh, Hoffman, and Albert [1999], for an alternative view). However, their experiments used

square figures, where each side of a square had an identical gap in its contour [Shipley and Kellman 1992]. Does support ratio account for perceptual strength of a figure when the figure is less simple and the pattern of its contour fragmentation is random?

In this paper, we first report new psychophysical results which demonstrate that the human visual system can classify contours as belonging or not to a figure on the basis of contour properties other than depth (e.g., orientation, curvature, length). We then describe a new computational model which partitions a set of image contours into those which belong to a target figure and those which do not on the basis of a given contour property. This model is based on the exponential pyramid architecture from the computer vision literature. A pyramid facilitates both fine-to-coarse and coarse-to-fine processing of an image [Jolion and Rosenfeld 1994]. Our proposed model exploits both properties: it uses fine-to-coarse (bottom up) processing to efficiently compute image-wide statistics on various properties of image contours. It then uses coarse-to-fine (top down) processing to classify *local* image contours on the basis of *global* information derived from these statistics. Figure-ground segregation for complex images (where the figure is partially occluded or otherwise appears degraded) seems to require use of global image information, yet efficient psychologically plausible methods for extracting such information have not been demonstrated previously. The important aspect of the proposed model is that it shows how global image information can be efficiently determined and used to make local processing decisions. Finally, we report the results of model simulations of psychophysical experiments. These simulations provided a test of the psychological plausibility of our new model. Additionally, these simulations allowed evaluation of the role of support ratio in figure-ground segregation.

2. PSYCHOPHYSICS

In order to test whether the 'intrinsic' contours of a figure must be classified separately from contours 'extrinsic' to the figure and whether such classification can be performed with contour properties other than relative depth, we used stimuli such as the one shown in Figure 4(a), where not all fragments of a particular color (here black) belong to the target figure. (Figure 4(a) is a sample image from one of our experimental conditions.) The target figure in this particular stimulus is an upside-down, asymmetric 'U', positioned slightly left of center. The stimulus contains a number of white and black distractors. When a white distractor overlaps the black target figure, it 'erases' part of the target figure. When a black distractor overlaps the black target figure, it distorts the shape

of the figure somewhat. Note that by using white and black distractors, we minimized (although did not eliminate) the role of region information: a white point in the image could be either in the interior or exterior of the figure. Similarly, a black point could be either in the interior or exterior of the figure. Minimizing the role of region information may help in isolating the specific role (if any) of contour information.

2.1 Experiment 1: Effects of Contour Orientation and Curvature

If classification of intrinsic vs. extrinsic contours and subsequent removal of extrinsic contours is critical for perception of partially visible figures to occur, then any cue which distinguishes between these two types of contours should lead to such perception. From the "pop out" phenomenon described by Treisman [1986], we know that a straight oblique contour would pop out from a field of straight horizontal and vertical contours. Similarly, a curved contour would pop out from a field of straight horizontal and vertical contours. Thus if a target has straight horizontal and vertical contours, while distractors have either straight oblique or curved contours, then the target figure should be readily perceived.

2.1.1 Method.

Subjects. Three subjects (GW, ZP, MS), including both authors, participated. GW was naive with regards to the hypothesis being tested. All subjects had practice in the task prior to the experiment such that their performance in the task was asymptotic prior to the start of the experiment. Subjects ZP and MS were myopes and used their corrective lenses. ZP and MS were experienced as subjects in psychophysical experiments. Viewing was monocular from a distance of 82 cm. A chin-forehead rest supported a subject's head.

Stimuli. Each stimulus contained one of the five black figures shown in Figure 3 against a white background, with distractors randomly added over the entire image. Each of the figures in Figure 3 consisted of four 200x100 pixel rectangles. Two types of distractor were used – white and black. For each stimulus image, there were always an equal number of white and black distractors, and they were always the same size. This experiment consisted of six conditions characterized by distractor size (small, large) and distractor shape (diamond, circle, square). Thus, the only thing that varied between the six experimental conditions was the nature of the distractors. The total distractor area (TDA) was held constant across the six conditions. The TDA for both white and black distractors in each condition was approximately 540,000 pixels (the actual area occupied

by distractors in a stimulus image could be less, due to distractors that overlap one another, for example). Stimuli were presented on a computer monitor, with 1280x1024 resolution (39.1 cm. x 29.5 cm.). A sample stimulus from each of the six conditions is shown in Figure 4.

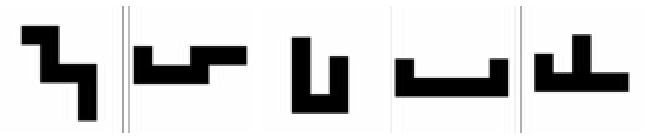


Figure 3. Target figures used in experiments ('normal' orientation).

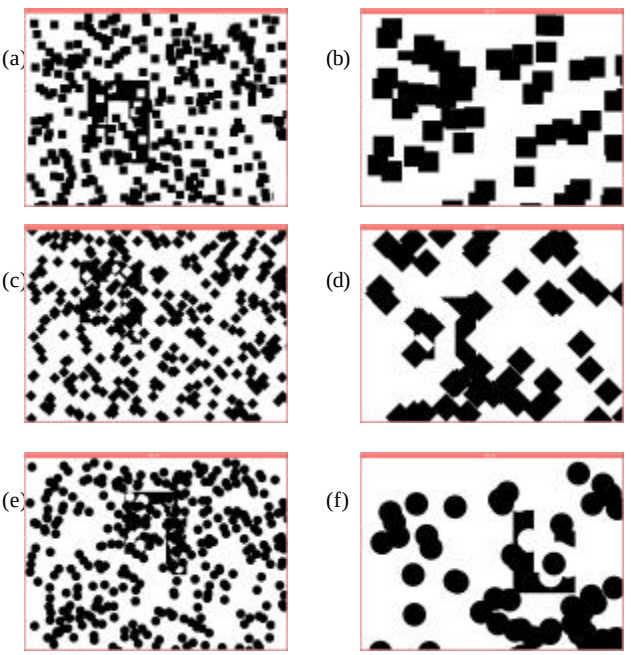


Figure 4. Sample stimuli from six experimental conditions of Experiment 1.

Procedure. The order of the six conditions was random and different for each subject. In the two square distractor conditions, the distractor contours did not differ in either orientation or curvature from the contours of the target figure. In the two diamond distractor conditions, distractor contours differed in orientation from the contours of the target figure, while in the two circle distractor conditions, distractor contours differed in curvature from the contours of the target figure.

Each experimental condition consisted of 500 experimental trials, which were preceded by 50 practice trials (subjects had the option to repeat the practice set as many times as they wished and they were informed of their performance after each such set). An identification task was used.

Each trial began with a fixation cross at the center of the screen. When the subject pressed the mouse button, a stimulus was presented for 100 msec. A stimulus contained one of the five figures shown in Figure 3 or its 180^o-rotated (upside down) version. A figure was presented amid white and black distractors, as described above. Half of the trials contained a figure in its normal orientation and the other half in the rotated orientation. Because location of the target figure in a stimulus varied randomly across trials, as did placement of the white and black distractors, the stimulus in each trial was unique. After the stimulus disappeared, the subject was shown images of the normal and the 180^o-rotated versions of the figure (without distractors) and was asked to identify which one had been presented in the stimulus. Subjects were given feedback at the end of each trial as to whether the normal or rotated version of the figure had been presented. When a 'Next Trial' button was clicked, the fixation cross reappeared, signaling the start of the next trial. Average proportion correct was used as the performance measure.

2.1.2 Results and Discussion. The results are shown in Figure 5. Performance is clearly better when either diamond or circle distractors are used than when square distractors are used. An interesting result is that while performance drops off with increasing distractor size in the case of square distractors, this decrease in performance is not as large (if present at all) when either diamond or circle distractors are used. This suggests that large distractors per se do not remove more informative parts of the figure than small ones.

The superior performance in the case of circle or diamond distractors supports the hypothesis that separation of extrinsic from intrinsic contours is critical for perception of partially visible figures, since the distractors were designed to lead to easy classification of the two types of contours. These results also show that depth cues are not required to classify those contours that are intrinsic to the figure from those that are extrinsic.

2.2 Experiment 2: Effect of Contour Length

In Experiment 1, subjects performed much better when diamond or circle, rather than square, distractors were used. Yet in the case of square distractors, subject performance was still quite good, as long as the square distractors were small. Perhaps the human visual system can distinguish the contours that belong to a figure from those that do not based on contour length. Indeed, in the small square distractor condition, there was a large difference in length of the contours of a distractor vs. those of a target figure, while in the large square distractor condition this difference was smaller. Under this hypothesis,

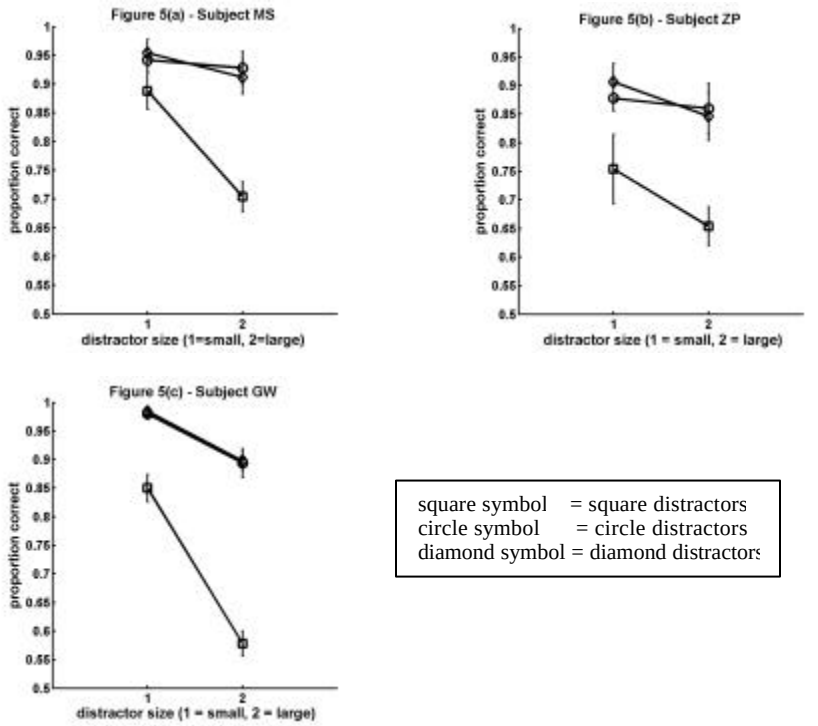


Figure 5. Results of Experiment 1. The roles of contour orientation and curvature.

performance in the square distractor conditions should be worse (and it was) than in the diamond or circle distractor conditions because *relative contour length* would be a less reliable cue than either *relative contour orientation* or *relative contour curvature* for classifying contours as intrinsic or extrinsic to the partially visible target figure. To understand why, recall that distractors are randomly placed on a stimulus image. Even in the case of small square distractors (with short contours), several white distractors could overlap the black target figure in such a way as to leave short pieces of contour of the figure. This would make classification between intrinsic and extrinsic contours on the basis of relative contour length somewhat more difficult than such classification on the basis of relative contour orientation or relative contour curvature.

2.2.1 Method.

Subjects. Three subjects participated (SG, ZP, MS), including the authors. SG was naive with regards to the hypothesis being tested. All subjects had practice in the task prior to the experiment such that their performance in the task was asymptotic prior to the start of the experiment. All subjects were myopes and used their corrective lenses. Viewing was monocular from a distance of 82 cm. A chin-forehead rest supported a subject's head.

Stimuli. Stimuli were like those used in Experiment 1, with the exception that all conditions in Experiment 2 used stimuli with square distractors. The primary factor manipulated in Experiment 2 was distractor size (i.e., extrinsic contour length). Three sizes of distractor were used: 40 x 40 pixels, 63 x 63 pixels, and 100 x 100 pixels. The values 40, 63, and 100 represented a linear increase of the size ratio (i.e., $63/40 \cong 100/63$). Total area used by distractors (TDA) was another factor manipulated in the experiment. Obviously, the greater the total area of the distractors used, the worse the performance should be in a figure-ground segregation task. The same TDA was used for both white and black distractors in a given stimulus image (as in Experiment 1). Three total distractor areas (TDAs) were used: 480,000, 540,000, and 600,000 pixels. These two experimental factors, distractor size and total distractor area, were orthogonal. For each total distractor area, there were three separate experimental conditions, one for each distractor size. For example, for a given total distractor area, there was a session with a large number of small distractors, a session with an intermediate number of intermediate-sized distractors, and a session with a small number of large distractors.

Procedure. There were 9 experimental conditions representing all combinations of the 3 total distractor areas and 3 distractor sizes. The order of conditions was random and different for different subjects. In all other aspects, this experiment was identical to Experiment 1.

2.2.2 Results. The results are shown in Figure 6. The abscissa represents size of the distractor, the ordinate represents average proportion correct. Individual curves represent different total distractor areas. The result of primary interest is represented by the downward slope of the curves: as distractor size increases, average proportion correct decreases. Since each curve represents constant total distractor area (many small distractors vs. fewer large distractors), the downward slope of the curves cannot be attributed to the 'amount' of the figure occluded. Instead, it may represent the effect of the relative scale (length) of intrinsic vs. extrinsic contours. The effect of the total distractor area itself is represented by the difference in heights among the curves – average proportion correct decreases with increasing total distractor area.

A 3-factor ANOVA with total distractor area, distractor size, and target figure as factors corroborates these observations. The ANOVA showed a main effect of total distractor area, ($F[2,4]=64.36$, $p=.0009$) and a main effect of distractor size, ($F[2,4]=92.95$, $p=.0004$). There was also a main effect of target figure, ($F[4,8]=34.51$, $p=.0001$) and an interaction between total distractor area and target figure ($F[8,16]=3.34$,

p=.019). There were no other significant interactions. The target figure effect is related to the fact that all subjects had worse performance on two of the target figures than on the others, and they all had better performance on one of the target figures than on the others. These differences across target figures led to large error bars in Figure 6. The total distractor area/target figure interaction may have the following explanation. For the difficult target figures, performance may have already been poor enough in the case of the smallest total distractor area that an increase in this area did not cause much more of a drop in performance. In other words, the difficult target figures showed a 'floor' effect.

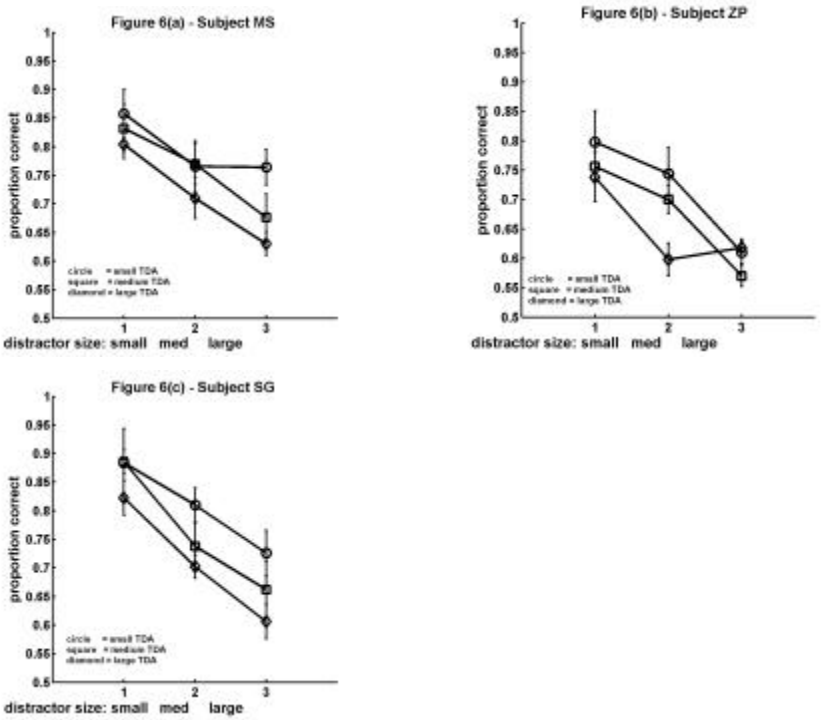


Figure 6. Results of Experiment 2. The role of contour length.

2.2.3 Discussion. In Experiment 2, two main effects were found: an effect of total distractor area and an effect of distractor size. The effect of distractor size (i.e., extrinsic contour length) is new, but the effect of total distractor area is not. Total distractor area is related to the concept of *support ratio* [Shipley and Kellman 1992]. Support ratio is the length of the visible (physically present) portion of a contour divided by total contour length. It has been shown that as the support ratio of an interpolated contour increases, perceptual strength of the interpolated contour also increases [Shipley and Kellman 1992]. When a distractor which is of the same color as the background overlaps a contour of a figure, it erases part of that contour and thus reduces its support ratio. For larger total

distractor area more contour will be removed. Therefore, support ratio decreases as total distractor area increases. As shown by the difference in heights of the three curves for each subject in Figure 6, the support ratio effect reported by Shipley and Kellman [1992] is confirmed. But the effect of distractor size found in our experiment cannot be accounted for by a support ratio explanation. Each curve in Figure 6 represents a constant total distractor area yet reflects decreasing performance with increasing distractor size. Figure 6 shows that performance can even be better for a *lower* support ratio, when the distractor is sufficiently small. As an example, the average proportion correct (for all subjects) is .79 in the small distractor/large total distractor area condition and .70 in the large distractor/small total distractor area condition. Therefore, while support ratio appears to be an important factor in the perception of partially visible figures, it is not the only factor, nor is it necessarily the most important one. Here, a more important factor may have been the distractor size (i.e., length of a distractor's contours relative to those of the figure).

Why should small distractor size lead to more success in solving this figure-ground segregation task? Before addressing this question, note that when a distractor is of the same color as the background, whenever it overlaps the boundary of the figure, it erases some of the *intrinsic* contour of the figure. On the other hand, whenever it overlaps the interior region of the figure (here, the black region inside the bounding contour of the figure), the distractor introduces contour that is *extrinsic* to the figure. Returning to our question, look at Figure 4(a). As pointed out above, small distractors are on a different scale than the larger target figure. As a result, straight line segments of extrinsic contours created by small distractors are short relative to the segments of intrinsic contours of the target figure. Large distractors, on the other hand (Figure 4(b)), are closer in scale to the target figure and thus the length of extrinsic contours created by large distractors are comparable to the length of the intrinsic contour fragments. Now, imagine the neighborhood of a given point in the image. This neighborhood may have fragment(s) of the target figure, as well as part or all of one or more distractors. The distractors add extrinsic contour to this neighborhood while the target figure adds intrinsic contour. When the extrinsic contours in a local neighborhood of the stimulus image are short relative to the intrinsic contours (i.e., small distractors are used), this difference in length of the two types of contours could be used to distinguish extrinsic from intrinsic contours. When extrinsic contours are longer (i.e., large distractors are used), it may be more difficult to distinguish extrinsic from intrinsic contours. In any case, relative contour

length would not be as reliable a property as relative contour orientation or relative contour curvature, because there would typically be overlap in the population distributions of intrinsic contour length and extrinsic contour length.

In order to choose a suitable cutoff point that separates the two distributions of contour lengths, the visual system may need to process the entire image (or a sizable part of it). This suggests that in using relative contour length to classify contours as intrinsic or extrinsic to a figure, the human visual system may need to perform *global* processing to select the cutoff length. The details of Experiment 2 pose an alternative possibility though. In the small square distractor conditions, the sides of the square distractors were 40 pixels, while the sides of a target figure (prior to occlusion) were considerably longer than this. It would have been natural then for subjects to choose a cutoff value of 41 pixels if relative contour length was the property used for contour classification. In each small square distractor experimental condition, the session was preceded by a number of practice trials. Thus, subjects "knew" beforehand about the appropriate cutoff to use for classification purposes in the experimental trials that followed. It is possible, then, that subjects did not detect a cutoff value in each experimental trial from the visual information present in the trial, but instead used a pre-determined cutoff value. A control experiment was performed to test the hypothesis that subjects have the perceptual capability to determine the cutoff value in a trial from the stimulus image itself and thus are not required to use memory.

2.2.4 Method.

Subjects. ZP and MS served as subjects in this control experiment.

Stimuli. Small square distractor stimuli and large square distractor stimuli, as before, were used. Total distractor area was held constant across distractor size. MS used TDA of 600,000 pixels, and ZP used TDA of 480,000 pixels. These values of TDA were used because they produced very similar performance in the two subjects in Experiment 2 – see Figure 6.

Procedure. Each subject ran two sessions apiece. Each session contained 500 experimental trials, preceded by 40 practice trials. Previously, a single session contained either all small square distractor trials or all large square distractor trials, but not both. Here, each session contained 250 small distractor trials and 250 large distractor trials. Individual trials were randomly assigned as either a small or large distractor trial. Thus, in the two sessions combined, each subject was tested in a total of 500 small distractor trials and 500 large distractor trials. All other details were the same as for the other

experiments. In these sessions, there was no single cutoff value that could produce efficient separation of intrinsic from extrinsic contours for both small and large distractor trials. Therefore, if a subject uses expectation/memory to determine the cutoff value, the performance in this experiment should be worse as compared to the performance in the corresponding conditions in Experiment 2 (TDA of 600,000 pixels in the case of MS and 480,000 pixels in the case of ZP), where small and large distractors were tested in separate sessions. On the other hand, if the cutoff value is computed from the visual information present in the stimulus, performance in this experiment should be the same as in Experiment 2.

2.2.5 Results and Discussion. Table 1 shows the results. If subjects simply used a pre-determined cutoff value for classification and were unable to determine a cutoff value from the stimulus image from trial to trial, performance should have been lower here than it was in Experiment 2. On the other hand, if the cutoff value was determined visually in each trial, performance should not have been lower than that in Experiment 2. Table 1 shows that the latter case was true. These results suggest then that the separation of intrinsic from extrinsic contours is done based on a cutoff value derived from *global* processing of the visual stimulus, rather than from memory.

Table 1: Comparison of performance between Experiment 2 and Control Experiment for small square distractors and for large square distractors. The numbers in parentheses represent standard errors.

Conditions	Exp. 2 proportion correct		Control exp. proportion correct	
Small Squares				
ZP	.798	(.053)	.774	(.055)
MS	.804	(.025)	.832	(.027)
Large Squares				
ZP	.610	(.019)	.676	(.030)
MS	.630	(.020)	.682	(.029)

3. MODEL DESCRIPTION

The psychophysical results suggest that in order for perception of a partially visible figure to occur, contours belonging to the figure must be distinguished from those that do not. Any contour property that facilitates such classification may be used. In addition to depth [Nakayama et al. 1989], some contour properties that may be used for such classification are relative contour orientation (Exp. 1), relative contour curvature (Exp. 1), and relative contour length (Exps. 1 & 2). Further, the contour property used for classification can be derived from global analysis of the image by the human visual system. To test this theory, we developed a computational model based on the exponential pyramid architecture. Testing the new model required three modules:

(a) an input module for extracting contours from an image,

(b) the exponential pyramid-based model that implements our theory,

(c) a template matching module, which takes the output set of contours from (b) and matches these against normally-oriented and 180^{o}-rotated templates of a figure, in order to determine the model's response to a given stimulus image.

Of these three modules, only (b) is central to the theory. The exponential pyramid detects the cutoff value of the contour property that partitions image contours into those that belong to the figure and those that do not. The input module (a) is needed to supply the exponential pyramid with the contours from an image. The template matching module (c) is needed in order to evaluate the performance of the exponential pyramid in classifying the image contours. Details of this evaluation will be given in the next section. The remainder of this section describes the main aspects of the exponential pyramid-based model's *structure* and *function*. The *input* and *output* modules (a) and (c) will also be briefly described.

3.1 Structure

The exponential pyramid architecture has been analyzed and shown to possess features useful in modeling certain Gestalt laws of perceptual organization [Rosenfeld 1990; Pizlo, Salach-Golyska, and Rosenfeld 1997]. A typical structure (the one used here) is a "non-overlapped quad-pyramid". Assume that the bottom layer of the pyramid has n processing nodes. The next layer has $n/4$ nodes, the one above that $n/16$ nodes, and so on. The top layer has only one node. Each node in a layer connects with four distinct 'children' nodes in the immediately lower layer and one 'parent' node in the immediately higher layer. Such a pyramid has $(\log_4 n) + 1$ layers. Each node in the pyramid has

limited memory and processing capability. The input image is presented to the bottom layer of the pyramid [Jolion and Rosenfeld 1994]. There are three characteristics of a pyramid:

1.) local, parallel processing: Different parts of an image can be processed simultaneously.
2.) multiscale: Different layers represent different spatial scales.
3.) hierarchical processing: Processing can go in two directions: bottom up (fine-to-coarse) and top down (coarse-to-fine).

Each node represents a receptive field, with nodes at higher layers in the pyramid having larger receptive fields. For all simulations described in the 'Testing the Model' section, a pyramid with eight layers was used.

3.2 Function

Given a set of image contours as input, the function of the exponential pyramid-based model is to determine whether there exists some contour property that partitions the set into two subsets (e.g., intrinsic and extrinsic) and, if necessary, to choose a cutoff value for that property. If such a property exists, the set of image contours can then be partitioned based on this property. For example, T-junctions can be used in classifying an image contour as belonging to either the nearer (occluding) or farther (occluded) figure in the image [Nakayama et al. 1989; Grossberg 1997]. Similarly, as the psychophysical results presented here show, differences in contour properties such as length (long vs. short), orientation (oblique vs. not oblique), and curvature (straight vs. curved) can be used in classifying an image contour as belonging to a (partially visible) figure or not. Given a small set of such candidate properties, the model analyzes statistics of these contour properties in the image and could decide which, if any, would best classify the image contours. Details of this analysis will now be illustrated with an example.

Consider the image in Figure 7(a). Assume that relative contour length is the candidate property to be used for classifying contours in this image. In order for the proposed pyramid model to perform this contour classification, it first needs to determine a cutoff value for contour length that allows classifying those contours that belong to the figure (an upside-down, asymmetric 'U') from those that do not. Bottom up (or fine-to-coarse) processing accomplishes this in the proposed pyramid model. Bottom up processing begins when all image contours are input to the bottom layer of the pyramid – see Figure 7(b). For this example, we assume a two layer pyramid: the bottom layer has

four nodes or 'receptive fields' (as shown in Figure 7(b) by dividing the image into four quadrants) and the top layer consists of just one receptive field which spans the entire image. (For all simulations, we used an eight layer pyramid. An eight layer pyramid has 128 x 128 receptive fields in the bottom layer.) Each receptive field in any layer applies the following algorithm to just those contours 'visible' to it; thus all receptive fields in a given layer may perform this processing in parallel on distinct parts of the image:

```
1.) initialize histogram for each contour property (length,
  orientation, etc.)

2.) for each contour, c, visible to this receptive field:
      - if pyramid layer ≠ top and c does not fit completely
        in this receptive field:
            - mark c to be 'passed up' to 'parent' cell
      - else
            - for each contour property (length,
              orientation, etc.):
                  - calculate contour property value for c
                  - increment bin in contour property
                    histogram corresponding to c's value

3.) for each contour property (length, orientation, etc.):
      - use histogram to compute variance of property value
      - choose maxval as the smallest value greater than
        all contour values in this receptive field
      - if pyramid layer ≠ bottom
            - choose cutoff value for this receptive field
```

The first step, initializing contour property histograms for a receptive field, involves initializing each bin in a histogram to zero when the receptive field is in the bottom layer. For all layers other than the bottom layer, each receptive field has exactly four 'child' receptive fields in the layer immediately below (see previous section on 'Structure'). Thus a receptive field in a non-bottom layer will initialize each bin in a histogram to the sum of the corresponding bins of its four 'child' receptive fields. In the second step, a receptive field considers each of the image contours 'visible' to it. In the bottom layer, this is literally true. In non-bottom layers, a receptive field receives just the image contours 'passed up' from its four 'child' receptive fields. A child passes up to its 'parent' receptive field just those image contours that do not fit completely within the child's receptive field. A child performs 'histogramming' for all other image contours in its receptive field. (Histogramming is a common technique used for image segmentation - see Horn [1986]; see Jolion and Rosenfeld [1994], for examples of histogramming in conjunction with pyramids). The third step in the algorithm tries to choose the cutoff

value for a contour property using a technique analogous to *root detection* [Jolion and Rosenfeld 1994]. This technique will now be illustrated.

(a) (b)

Figure 7. Determination of contour length property cutoff (see text for details).

Consider the lower right hand quadrant in the 'bottom layer' depicted in Figure 7(b). Most of the contours in this 'receptive field' are of the same length, so the variance of the contour length property will be quite small. As processing proceeds to the next higher layer in the pyramid, the receptive field size becomes larger (here it spans the entire image, because in this example we consider only a two layer pyramid). Now, the parent receptive field contains contours belonging to a larger polygonal figure (the target) and to the smaller squares. Since contours belonging to the target figure tend to be longer, the variance of contour length is much greater than in the child receptive field (corresponding to the lower right quadrant of Figure 7(b)). The parent receptive field (in the top layer) computes the ratio of its contour length variance to that of its child. If this ratio exceeds some threshold, it chooses the contour length cutoff to be just greater than the longest contour in the child receptive field - *maxval* in the algorithm. If this ratio exceeds the threshold for more than one child, then the minimum *maxval* is chosen as the cutoff for the parent receptive field. (We used 1.25 as the threshold for both relative contour length and relative contour orientation - this threshold was empirically determined in preliminary simulations.) In particular, note that global information is used in determining the cutoff for a contour property. In subsequent top down (or coarse-to-fine) processing, the cutoff value is used to classify image contours as belonging to the target figure or not[2]. In the section 'Testing the Model', we report simulation results which suggest that the proposed exponential pyramid-based model is psychologically plausible.

3.3 Input

The input module applies Sobel edge detection with custom contour extraction to a stimulus image to find the set of image contours. Figure 8 shows some examples. As noted previously, the input module (and the output module described next) was built to facilitate evaluation of the proposed exponential pyramid model. As such, we do not claim that the human visual system uses Sobel edge detection to recover image contours. Instead, image contours could be recovered by any of the large class of image segmentation methods [Davies 1997; Ballard and Brown 1982; Barrow and Tenenbaum 1986; Horn 1986].

3.4 Output

When two types of contour are present within the image (e.g., contours intrinsic to a figure vs. those extrinsic to it), the output of the exponential pyramid model is a classification of each image contour into one of the two types (if possible). The template matching module receives the contours that have been classified by the pyramid as being intrinsic to the target figure and matches these to normally-oriented and 180^{o}-rotated templates of the figure, to determine which template best matches the set of intrinsic contours. Again, this template matching module exists for the purpose of evaluating the performance of the proposed exponential pyramid model.

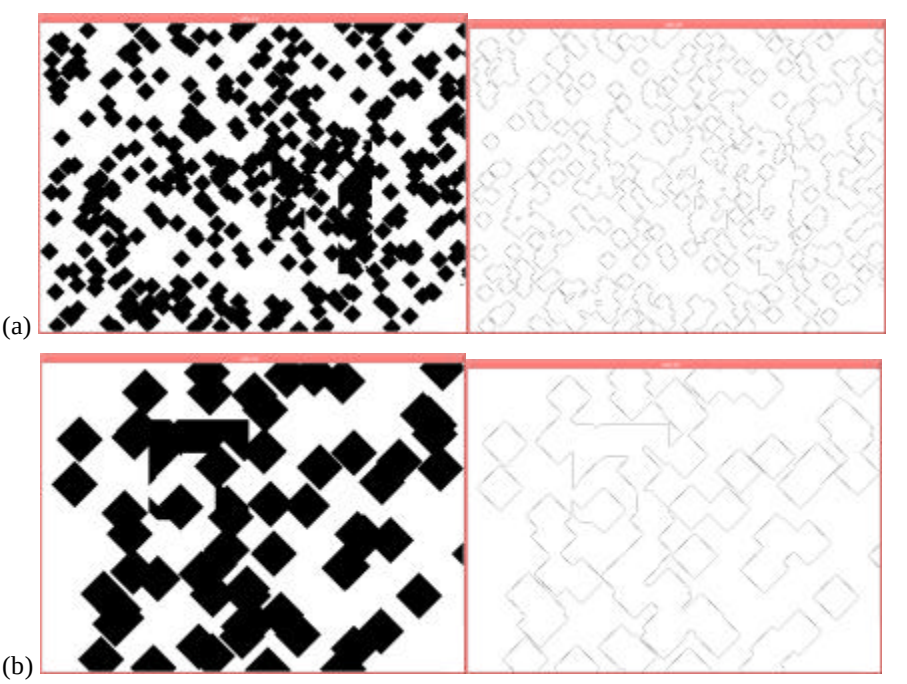


Figure 8. Left column shows original sample stimuli. Right column shows results of Sobel edge detection and custom contour extraction. (a) Small diamond distractor stimulus. (b) Large diamond distractor stimulus.

4. TESTING THE MODEL

Evaluation of the model's performance was conceptually simple – treat the model as an additional subject in an experiment. Then compare its performance to that of the human subjects from the experiment. There was one free parameter in the model: the magnitude of noise in the perceptual representation of intrinsic contour. First, model performance was compared to human performance in an experiment testing the effect of contour orientation. Then, model performance was compared to human performance in an experiment testing the effect of contour length. The model was not tested on the effect of contour curvature. Recall that the performance of human subjects was high and equally good in the contour orientation and contour curvature cases (Experiment 1). Performance was high because the separation of intrinsic from extrinsic contours was quite easy in both cases. Therefore, it was assumed that the results from the contour orientation experimental conditions would be sufficient for estimation of the free parameter in the model.

4.1 Contour Orientation

In order to compare model performance to human performance with respect to the effect of relative contour orientation, we replicated the diamond conditions from Experiment 1 with two subjects, MS and ZP. We used a total distractor area of 640,000 pixels, which is larger than that used in Experiment 1. Larger total distractor area made the task somewhat more difficult. By doing this we wanted to make sure that the subjects' performance was not too close to perfect. Both subjects used the same sets of randomly generated stimuli.

The model was applied to the same sets of stimuli used by human subjects. Two types of analysis were made for each stimulus: (i) the intrinsic/extrinsic contour classification; (ii) template matching. In the template matching part, the intrinsic contours extracted from the image were compared to both the normal orientation template and the 180^{o}-rotated orientation template of the target. For each template, the total length of intrinsic contour matched was computed. This is called the support ratio I. Then, the difference I_M between the support ratios for the normal and rotated templates was computed. A positive difference indicated the response "normal", a negative difference indicated the response "rotated".

If the model is psychologically plausible, there should be a systematic relationship between the accuracy of human response and the magnitude of I_M. High positive values of I_M, as detected by the model, should correspond to high proportions of responses "normal" by the human subject. Conversely, high negative values of I_M should correspond to high proportions of responses "rotated" by the human subject. Such a relationship is called a psychometric function. This kind of a systematic relationship would suggest that the amount of intrinsic contour detected by the visual system of the human observer, which is the input to the decisional module of the observer, is identical to the amount of intrinsic contour detected (recovered) by the model. If this is indeed the case, then the claim *that the computational methods involved in the model represent the perceptual mechanisms of the human observer* will be supported. Details of model evaluation follow.

The support ratio of a contour c is defined as:

$$I_c = d_c/D_c \qquad (1)$$

where D_c = total length of contour c and d_c = total length of *visible* (physically present) *contour* for contour c. As the support ratio of contour c increases, so does the 'perceptual strength' of contour c [Shipley and Kellman 1992].

The model estimates I_c by superimposing the template of the target at all positions on the stimulus image and computing the length of the overlap between the contours in the image that were classified as intrinsic and the contours of the template. It takes the maximum value of the overlap across positions of the template and normalizes it to D_c. For the template at normal orientation, the normalized value (support ratio) is denoted by I_N and for the template at the rotated orientation, the normalized value is denoted by I_I. The model then computes the difference I_M:

$$I_M = I_N - I_I \, . \qquad (2)$$

The decision rule which determines the model's response is given by:

$$r = \begin{cases} I_M < 0: \text{ respond "rotated"}, \\ I_M > 0: \text{ respond "normal"}, \\ I_M = 0: \text{ respond "normal" with } p = 0.5. \end{cases} \qquad (3)$$

If this model is psychologically plausible, there should be a relationship between the value of I_M, as computed by the model, and human performance. Note that I_M is what the model *detects*, as opposed to what is *actually given* in the image. Let:

$$I_M = I^* + \in_{O(M)} + \in_{I(M)} \, , \qquad (4)$$

where I_M = 'detected' difference of support ratios, I^* = 'actual' difference of support ratios, $\in_{O(M)} \sim N(0,\sigma^2_{O(M)})$, and $\in_{I(M)} \sim N(0,\sigma^2_{I(M)})$. $\in_{O(M)}$ represents error in using orientation as a classification feature. $\in_{I(M)}$ represents all other sources of model error.

If $\sigma^2_{O(M)} = \sigma^2_{I(M)} = 0$ then:

$$I_M = I^* . \quad (5)$$

This happens when classification of intrinsic vs. extrinsic contours is perfect. In such a case using decision rule (3) leads to the model's 'psychometric' function, which is a step function. In the simulations, root detection led to perfect performance in the small diamond distractor condition and nearly perfect performance (497/500 = .994 correct) in the large diamond distractor condition. Specifically, for $I_M > 0$, the model always correctly judged the target to be in its normal orientation and for $I_M < 0$, it never judged the target to be in its normal orientation. The three misses in the large diamond distractor simulation happened when $I_M = 0$. In such trials, the model randomly responded either 'normal' or 'rotated'. Evidently, it 'guessed' incorrectly in each of these three trials.

For human subjects, equation (4) takes the form:

$$I_H = I^* + \in_O + \in_I \quad (6)$$

where I_H = 'detected' difference of support ratios by the human observer, $\in_O \sim N(0,\sigma^2_O)$, and $\in_I \sim N(0,\sigma^2_I)$. $\in_O$ represents human error in using orientation as a classification feature. $\in_I$ represents error produced by noise in the perceptual representation of intrinsic contour. It seems reasonable to assume $\sigma^2_O = 0$ because it is known that human subjects are able to discriminate between lines whose orientations differ by less than a degree [Regan and Beverley 1985]; here, there is a 45^o orientation difference between the contours to be discriminated (i.e., between the intrinsic and extrinsic contours). So, (6) becomes:

$$I_H = I^* + \in_I . \quad (7)$$

From (5) and (7), we obtain:

$$I_H = I_M + \in_I \quad (7a)$$

Assume that the human observer uses a decision rule analogous to (3):

r = $I_H < 0$: respond "rotated",
$I_H > 0$: respond "normal",
$I_H = 0$: respond "normal" with p = 0.5. (8)

Using this rule leads to a psychometric function that is a sigmoid curve, rather than a step function as implied by (3). The logic for this claim is as follows. In (7), for a given I^*, I_H

is a random variable due to the error term $\in_I$, as defined in (6). The density function for I_H is therefore:

$$f(I_H) = \frac{1}{\sqrt{2p}\, s_I} e^{-\frac{1}{2}\left(\frac{I_H - I^*}{s_I}\right)^2} \qquad (9)$$

The probability of a human responding that the target is in its "normal" orientation can be written as:

$$p_H(\text{"normal"}) = p(I_H > 0) = \int_0^{+\infty} f(I_H) dI_H \qquad (10)$$

Note from (5) that I_M can be substituted for I* in (9). Therefore, when p_H("normal") is plotted against I_M, a sigmoid curve is expected. Specifically, this curve is a cumulative normal distribution with zero mean and variance σ_I^2. These psychometric functions can be determined because the model's computation of I_M and the subjects' responses p_H("normal") were obtained from the same stimuli.

Figure 9 shows the psychometric functions for each subject. (a)-(b) show the functions in the small and large diamond distractor conditions for ZP. (c)-(d) show the functions in the small and large diamond distractor conditions for MS. In each graph, the solid curve is the curve fitted to the data points by using Probit Analysis [Finney 1971]. The goodness of fit was evaluated using a χ^2 test. Large values of χ^2 indicate poor fit which might be caused by either using the wrong approximating function (here, Gaussian is assumed) or by not including factors in the model that have a substantial effect on the subject's performance. Here only one such factor is included, namely I_M. If on the other hand, the χ^2 is not very large, we are justified in claiming that I_M is the only factor (except for random Gaussian noise) which accounts for human performance in the orientation experiment. This was the case for each curve for each subject. Table 2 shows the Pearson χ^2, its p-value, the $(\hat{m}, \hat{s})$ parameter estimates, and their standard errors. Note that the estimated mean of the psychometric function was close to zero in all conditions (as compared to se). This indicates that the subjects did not have appreciable response bias towards either a "normal" or "rotated" response.

Table 2: Curve fitting and goodness of curve fit

Distractor Type/ Subject	μ	se	σ	se	χ^2	df	p
Small Diamonds							
ZP	.008	.008	.116	.009	10.779	6	.096
MS	.005	.006	.070	.005	6.034	6	.419
Large Diamonds							
ZP	.024	.009	.149	.015	1.565	7	.980
MS	-.000	.005	.062	.004	9.802	7	.200
Small and Large Combined							
ZP	.014	.006	.130	.008	9.117	7	.244
MS	.002	.004	.066	.003	11.282	7	.127

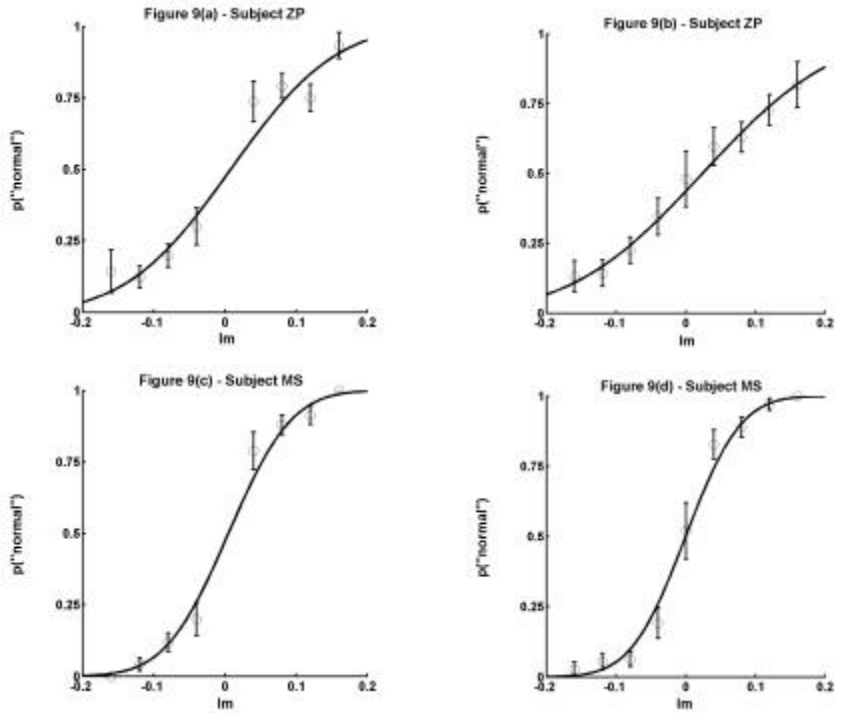


Figure 9. Psychometric functions for contour orientation conditions. (a) Small diamond distractor condition for subject ZP. (b) Large diamond distractor condition for subject ZP. (c) Small diamond distractor condition for subject MS. (d) Large diamond distractor condition for subject MS.

The two experimental conditions – small and large diamond distractors – produced similar psychometric functions. This was expected because the values of I_M computed by the model were very similar in the two conditions. In other words, these two conditions were of equal difficulty for the model. If the model is psychologically plausible, these two conditions should also be of equal difficulty for the subjects, and in fact they were (see Figure 5). Therefore, we estimated the psychometric function from both conditions taken together, to obtain a more reliable estimate of the variance, σ_I^2, of the function.

The next question is whether this model would generalize to other types of stimuli, where the property for separating intrinsic from extrinsic contours is not contour orientation. In fact, one might argue that the simulations presented so far have not provided a strong test of the model simply because the separation of intrinsic from extrinsic contours was quite easy and the *detected* support ratio I_M was usually identical to the *actual* support ratio I*. In other words, is the actual support ratio, as represented by I*, the whole story?

The results of Experiment 2 indicate otherwise (see Figure 6). Recall that in this experiment, the length of the contour was the property to be used in classification of

Table 3: Actual support ratio (I*) is approximately constant when Total Distractor Area (TDA) is constant. Results for square distractors.

Condition	average I*	se
small TDA		
small squares	.1176	.0149
medium squares	.1228	.0145
large squares	.1196	.0170
medium TDA		
small squares	.1020	.0122
medium squares	.1072	.0099
large squares	.1024	.0132
large TDA		
small squares	.0876	.0106
medium squares	.0940	.0105
large squares	.0886	.0113

intrinsic vs. extrinsic contours. In this experiment, as distractor size increased performance of subjects in the identification task decreased (by contrast, this was not so for diamond distractors). This was true even though total distractor area (TDA) was held constant across distractor size conditions. Since TDA determines the actual support ratio I*, actual support ratio was also held constant across distractor size conditions. Table 3 illustrates this. It follows that the systematic effect of distractor size found in Experiment 2 cannot be accounted for by the *actual* support ratio I*. The question now is whether the support ratio I_M, as *detected* by our model, can account for these psychophysical results. To answer this question we performed a simulation experiment in which the model was

applied to the stimuli that were used in Experiment 2. Next, we derive a prediction for the parameters of the psychometric function in the case of stimuli from Experiment 2.

4.2 Contour Length

For the size simulations, the appropriate model (analogous to (4)) is:

$$I_M = I^* + \in_{S(M)} + \in_{I(M)} \ , \qquad (11)$$

where I_M, I^*, and $\in_{I(M)}$ are as before, $\in_{S(M)} \sim N(0,\sigma^2_{S(M)})$. $\in_{S(M)}$ represents model error in using contour length as a classification property. From casual inspection of images such as Figure 4(a)-(b), it is expected that $\sigma^2_{S(M)} > 0$. As with the orientation simulations, $\sigma^2_{I(M)} = 0$. Thus (11) becomes:

$$I_M = I^* + \in_{S(M)} \qquad (12)$$

and

$$I_M \neq I^* \ . \qquad (13)$$

For human subjects, the relation analogous to (11) takes the form:

$$I_H = I^* + \in_S + \in_I \ , \qquad (14)$$

where I_H, I^*, and $\in_I$ are as before, $\in_S \sim N(0,\sigma^2_S)$. $\in_S$ represents the error produced by the human visual system in using contour length as a classification property. From (12) and (14), we obtain:

$$I_H = I_M + (\in_S - \in_{S(M)}) + \in_I . \qquad (15)$$

If the model classifies contours exactly the same way as the human visual system does (in each and every trial), then:

$$\in_S = \in_{S(M)} . \qquad (16)$$

From (15) and (16) we obtain:

$$I_H = I_M + \in_I . \qquad (17)$$

Note that this equation is identical to equation (7a) for orientation. Therefore, if we plot the proportion p_H("normal") of the human observer against I_M, we should obtain one psychometric function defined by equation (10) for all 9 conditions in the contour length experiment. Recall that the psychometric function in equation (10) was estimated from the results of an experiment where a different property, namely contour orientation, was used for contour classification. Experiment 2, the contour length experiment, involved a total of 9 conditions: there were three levels of the distractor size (discussed above) and three levels of the total distractor area. If the pyramid model were psychologically plausible, it would be able to account for psychophysical results from all 9 different experimental conditions in the contour length experiment (Exp. 2) using no free

parameters (the variance σ_I^2 was estimated from the orientation data). If, however, the model were inadequate, there would be no reason to expect that the psychometric function defined in equation (10) would provide a good fit to all nine conditions of Experiment 2.

Subject MS repeated the nine conditions of Experiment 2, using the same sets of stimuli that were used by subject ZP in that experiment. Thus, the same nine sets of stimuli were used for each subject and the model to facilitate direct comparison between the model and each subject. The results of MS from the repetition of Experiment 2 were very similar to his original results from Experiment 2.

Simulations were carried out in the same manner as for contour orientation. Here, two simulations (one producing I_M, the other producing I*) were performed for each of the nine experimental conditions. The first simulation produced classification of intrinsic and extrinsic contours based on root detection. In the second simulation, no classification of intrinsic vs. extrinsic contours was performed – all image contours were passed directly to the template matching module. Thus, the first simulation for a given condition produced *detected* support ratio (in the form of I_M), while the second simulation produced *actual* support ratio (in the form of I*). While the first set of simulations facilitated testing the model proposed here, the second set of simulations permitted testing the predictive power of actual support ratio [Shipley and Kellman 1992]. The contour orientation simulations did not allow an independent test of I_M and I* because the populations of intrinsic and extrinsic contours were so clearly separable that I_M = I* (detected support ratio was equal to actual support ratio).

We fitted one psychometric function to the results pooled from all 9 conditions of Experiment 2 for both I_M and I*. Figure 10 depicts the relationship among estimated σ's for these psychometric functions as well as for that of the combined orientation conditions. It is clear that for both subjects, there is not a significant difference between the σ's for the contour orientation psychometric function and the contour length psychometric function for the I_M case. The σ for the contour length psychometric function for I* is, however, significantly different (by a factor of 2) from the other two σ's. The fact that there is no significant difference between the σ for contour orientation and contour length in the case of I_M suggests that the curve estimated from contour orientation data can account for the psychophysical results from each of the nine individual conditions in the contour length experiment (Exp. 2). This conjecture will be evaluated by analyzing the slopes of the psychometric functions.

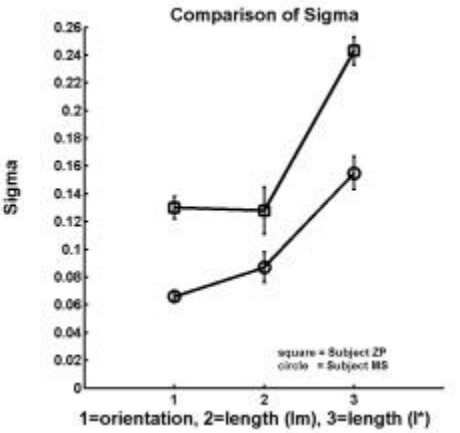


Figure 10. Comparison of σ's for contour orientation (diamond distractors), contour length (square distractors – I_M simulations), contour length (square distractors – I* simulations).

We fitted psychometric functions to the individual 9 conditions of Experiment 2. The estimated σ's are shown in Figure 11. Each graph in this figure shows estimated σ for the curve fitted to the combined diamond distractor conditions (solid horizontal line), the estimated σ for the curve fitted to the combined square distractor conditions (dashed horizontal line), and the estimated σ for the curves individually fitted for each of the nine square distractor conditions (nine data points). There is one graph each for I_M and for I* for each subject. The σ for combined length (square distractor) conditions and σ for combined orientation (diamond distractor) conditions are very similar in the case of I_M. However, this is not true in the case of I*: the solid and dashed lines for each subject are spaced farther apart. This result was already shown in Figure 10. The new aspect is the set of σ's from the individual conditions. The data points representing the σ's fall closer to both lines in the case of I_M. This fact illustrates that the psychometric function estimated in the orientation experiment can indeed account well for the results from the individual conditions in the length experiment when I_M (but not I*) is used as the independent variable. Note that the data points in the case of I* are not scattered randomly. Instead, they show a systematic pattern. The nine conditions represent three triplets. Each triplet represents a specific total distractor area: triplets {1,2,3}, {4,5,6}, and {7,8,9} represent small, medium, and large total distractor area, respectively. The σ's for I* across the triplets of conditions are similar. This means that I* can account for differences in performance across different TDAs. This is not surprising because I* is in fact closely related to TDA (as shown in Table 3). But σ's for I* within triplets are not

similar. Instead, the σ's are positively correlated with the distractor size: with the exception of the last triplet of points for subject ZP (i.e., the data points for conditions 7,8,9), each triplet of points is monotonically increasing with condition. This systematic change of σ across experimental conditions represents differences in difficulty across the conditions that are not accounted for by the psychometric function estimated in the orientation experiment. In other words, if this function were substituted for the set of nine individually fitted curves for I*, information pertaining to this systematic relationship would be lost.

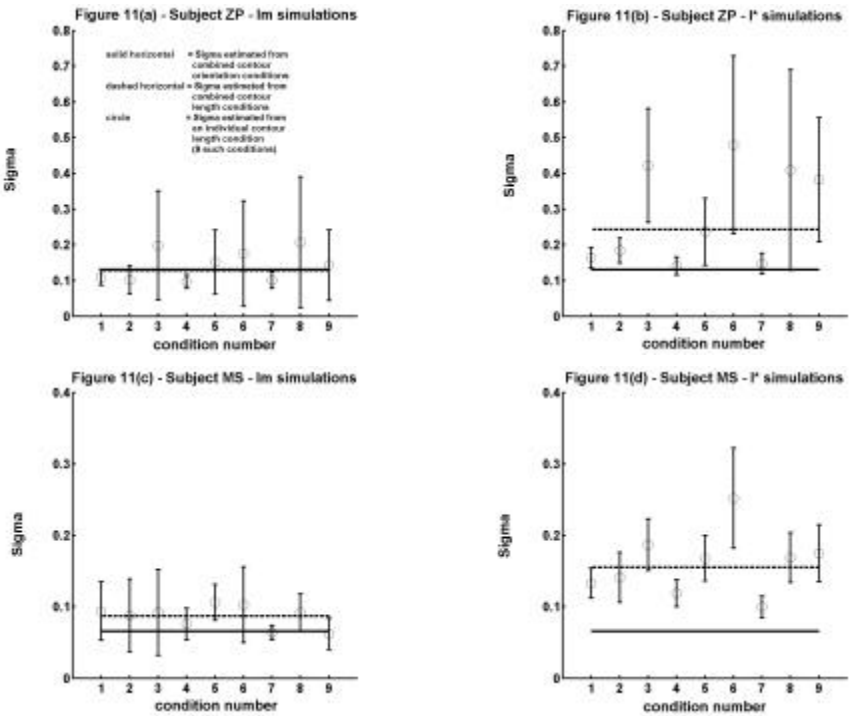


Figure 11. Comparison of 'overall' contour orientation and contour length σ's to the 'individual' σ's from each contour length condition.

Figure 12(a) shows the I_M histograms for each of the nine individual conditions. The top row represents small TDA, the middle row represents medium TDA, and the bottom row represents large TDA. Figure 12(b) shows the corresponding I* histograms for the nine conditions. Note that all I* histograms for a level of TDA are similar to one another. The I_M histograms, on the other hand, do vary substantially across conditions for a given level of TDA. The conditions that were easy for subjects led to many large absolute values of I_M. Those that were difficult for subjects led to many small absolute values of I_M. This observation led to the following analysis.

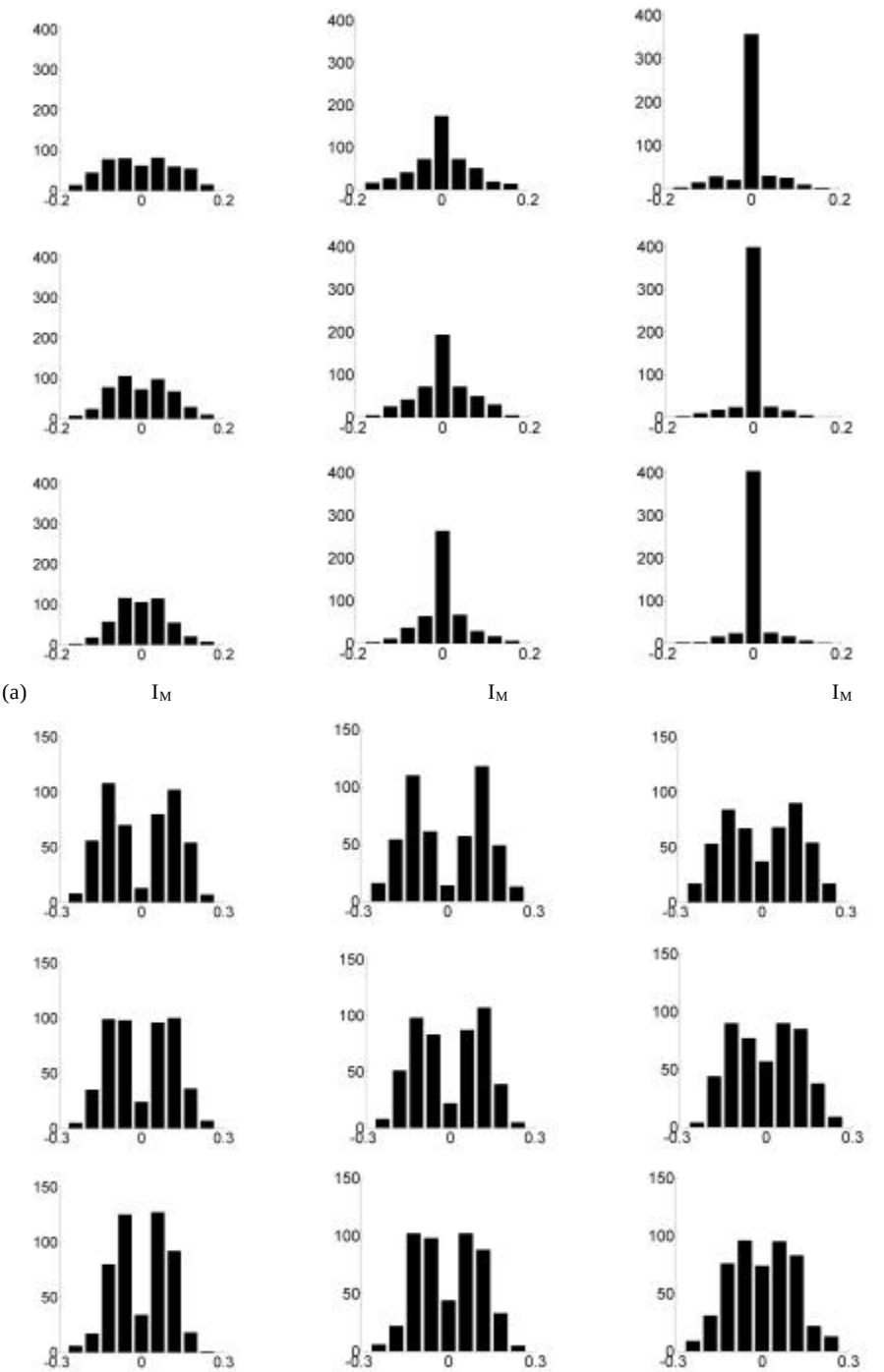


Figure 12. (a) Frequency histograms for I_M simulations. (b) Frequency histograms for I* simulations. Total distractor area (TDA) increases going from top to bottom rows. Distractor size increases going from left to right columns.

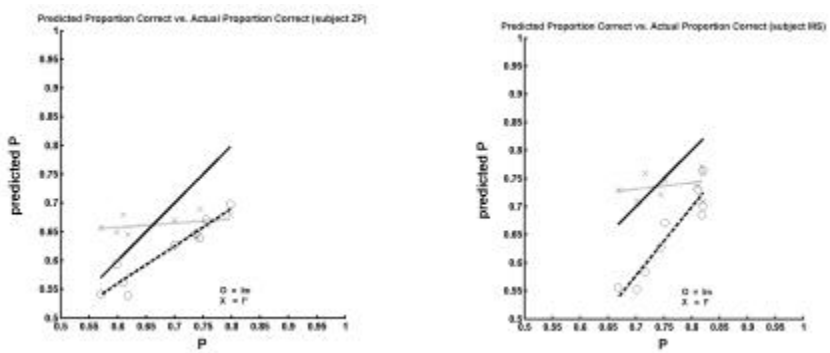
According to model predictions, one psychometric function (equation 10) can account for all nine conditions. The differences in the difficulty would be entirely attributable to the absolute magnitude of I_M as detected by the model (see histograms). If that is the case, then the proportion correct for each of the nine sessions can be predicted from the histograms of I_M and the noise inherent in the perceptual representation of intrinsic contour as estimated by σ_I. This prediction is made as follows. From the 'overall' fitted curve $P(I_M)$, the proportion of correct responses, p, for any value of I_M can be determined by taking $p = P(I_M)$ when $I_M > 0$, and $p = 1 - P(I_M)$ when $I_M < 0$. Assume a vector of such proportions, $\mathbf{p}$. Next, assume a vector of $\mathbf{n_i}$, where i = experimental condition 1,…,9 and where each element of $\mathbf{n_i}$ is the number of trials falling in a particular interval of I_M (see histograms). Then, the overall predicted proportion correct for a given experimental condition **i** is given by:

$$\mathbf{p_i)} = (\mathbf{p} \cdot \mathbf{n_i}) / (\mathbf{n_i} \cdot \mathbf{1}) \ , \qquad \mathbf{i = 1, \ldots, 9} \tag{18}$$

where **1** is a column vector whose elements are all ones. If the model is psychologically plausible, then $\mathbf{p_i)}$ computed from (18) should be equal to the proportion correct from the corresponding condition **i** in the psychophysical experiment. So, if estimated proportion correct $\mathbf{p_i)}$ is plotted against actual proportion correct from the experiment, the data points should be on a diagonal. If, however, the model has no relationship to perceptual mechanisms, then the data points should fall on a horizontal line.

Formula (18) was applied to both the I_M and I* cases, and the results are shown in Figure 13. The solid line is the diagonal (slope of +1). The o's represent the data points which correspond to predicted proportion correct based on histograms of I_M for each of the nine conditions for the subject. The x's are the data points corresponding to the predicted proportion correct where the prediction was based on the histograms of I*. A regression line was fitted to the points produced by I_M and to those produced by I*. As Figure 13 shows, the slope of the regression line fitted for the I* predictions is not much different from horizontal. A statistical analysis confirmed that the slope for the I* line was not significantly different than zero for either subject (ZP: slope = .072, se = .073; MS: slope = .115, se = .142). This means that I* has no predictive value across the experimental conditions used in this experiment. On the other hand, the slope of the regression line fitted for the I_M predictions is not close to zero. In fact, this slope is much greater than zero for both subjects. For subject MS, this slope was close to one (slope = 1.226, se = .176). For subject ZP, the slope for the I_M line was not close to 1 (slope = .653, se = .084). However, from looking at this graph, it is clear that for the more difficult

experimental conditions, there was a floor effect in the case of ZP (performance close to chance), which would likely have affected the slope of the resulting fitted regression line. To test this, the two data points corresponding to the two conditions where ZP had proportion correct less than .6 in the psychophysical experiment were removed and the regression was performed again. The slope increased to .785 (se = .072), suggesting that the slope of the relation between observed p and predicted p may indeed be one for both subjects. These results clearly show that the model that produces I_M is a much better model of human performance than the one which produces I*. The fact that the regression line for I_M for each subject is shifted down relative to the diagonal indicates that there is yet some other factor besides detected support ratio. This issue will be revisited in the General Discussion section. From this graph, however, it is clear that *detected* support ratio is a more important factor than *actual* support ratio.



(a) (b)

Figure 13. Predictions by I_M and I* of performance in Experiment 2. (a) Predictions for subject ZP. (b) Predictions for subject MS.

To summarize, the simulation results presented above show that our model of the perception of partially visible figures is psychologically plausible. Using root detection-based classification (I_M simulations), it provided a good account of the psychophysical results. The relationship between model and human results was much stronger than when classification was not used (I* simulations). This implies that *detected* support ratio is a more important factor than *actual* support ratio [Shipley and Kellman 1992] in the perception of partially visible figures. We would like to emphasize the fact that the root detection-based classification, which is a critical element in our model, involves *global* processing of the image. Only after the entire image has been analyzed does the model obtain a cutoff value for a given contour property that can then be used to reliably classify those image contours belonging to the figure. A purely local based analysis is unlikely to produce the required classification. The strong account of the psychophysical

data provided by this model suggests that humans also use this kind of global processing in classifying contours. This conclusion is consistent with the Gestalt idea that something about the whole may guide how the parts are grouped into that whole [Wertheimer 1923/1958; Koffka 1935].

5. GENERAL DISCUSSION

Figure-ground segregation for complex images where the figure is only partially visible seems to require that image contours first be classified as either belonging or not to the figure. Nakayama et al. [1989] proposed this already in the context of partially occluded figures. However, unlike their theory, we propose that depth cues are not *necessary* to facilitate such classification. Rather, *any* property of contours can potentially be used to classify the contours in an image as either belonging or not to a partially visible figure. The psychophysical results presented here show that *relative contour orientation* (Exp. 1), *relative contour curvature* (Exp. 1), and *relative contour length* (Exps. 1-2) are other contour properties, in addition to depth, that may be used in the classification of image contours. Further, our psychophysical results, especially those of the contour length (square distractor) conditions, suggest that a human observer *globally* processes an image (or a large part of it) to find the intrinsic contours. A new model that incorporates these principles was formulated. In particular, this new model is efficient in globally processing an image, because a 'receptive field' in a given layer of the model may process its distinct portion of the image in parallel with all other receptive fields at that layer. Computer simulations using this new model accounted well for the results of several psychophysical experimental conditions presented here. Using just one free parameter, the new model accounted for the results across 11 experimental conditions.

One potential objection to this proposed theory is that perhaps classification of image contours *per se* is unnecessary. That is, perhaps the visual system can group just the contours that belong to the partially visible figure without relying on an explicit prior classification stage. The contour length simulations of Experiment 2 provide strong evidence against this argument, however. Two sets of simulations of Experiment 2 were performed – the I_M simulations, which used the exponential pyramid model to perform the intrinsic/extrinsic classification, and the I* simulations, where image contours were passed, unclassified, directly to the template matching module. As shown, the I_M simulations provided the superior account of the human psychophysical results of Experiment 2. These simulations also showed that support ratio, as formulated by Shipley

and Kellman [1992], does not provide a complete explanation for the perception of partially visible figures. *Detected* support ratio, as defined in this paper, provides a more complete explanation.

While the contour length I_M simulations provided a much better account than the I* simulations of the results of Experiment 2, consider again Figure 13. Note for each subject that the fitted I_M 'prediction' line is shifted down from the diagonal (that represents human performance). This suggests the presence of a secondary factor. This secondary factor could be related to some role of region information in the perception of partially visible figures. Indeed, it has recently been shown that region or surface information can play a role in this task [Scheessele and Perez 2003; Yin, Kellman, and Shipley 1997; Tse and Albert 1998; Grossberg 1997].

FOOTNOTES

[1] The retinal image also provides *region* information, which arises from figure (object) surfaces and from the background in the visual scene. Here, we focus only on the contribution of contour information to figure-ground segregation. Additionally, non-retinal sources of information, such as memory, are not considered here (but see Peterson [1999]; Palmer and Rock [1994a], [1994b]; Vecera and O'Reilly [2000]).

[2] The use of a bottom up stage to compute the statistics of an image and a top down stage for segmentation resembles Bouman and Liu's [1991] algorithm.

REFERENCES

BALLARD, D. H., AND BROWN, C. M. 1982. *Computer vision.* Englewood Cliffs, NJ: Prentice-Hall, Inc.

BARROW, H. G., AND TENENBAUM, J. M. 1986. Computational approaches to vision. In K. R. BOFF, L. KAUFMANN, AND J. P. THOMAS (Eds.), *Handbook of human perception and performance.* (pp. 38-1 – 38-70). New York: Wiley.

BOUMAN, C., AND LIU, B. 1991. Multiple resolution segmentation of textured images. *IEEE Transactions on Pattern Analysis & Machine Intelligence, 13,* 99-113.

BROWN, J. M., AND KOCH, C. 1993. Influences of closure, occlusion, and size on the perception of fragmented pictures. *Perception & Psychophysics, 53,* 436-442.

DAVIES, E. R. 1997. *Machine vision: Theory, algorithms, practicalities, 2nd ed.* San Diego, CA: Academic Press.

ELDER, J., AND ZUCKER, S. 1994. A measure of closure. *Vision Research, 34,* 3361-3369.

ELDER, J. H., AND GOLDBERG, R. M. 2002. Ecological statistics of Gestalt laws for the perceptual organization of contours. *Journal of Vision, 2,* 324-353.

FELDMAN, J. 2001. Bayesian contour integration. *Perception & Psychophysics, 63,* 1171-1182.

FINNEY, D. J. 1971. *Probit analysis.* Cambridge: Cambridge University Press.

GEISLER, W. S., PERRY, J. S., SUPER, B. J., AND GALLOGLY, D. P. 2001. Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research, 41,* 711-724.

GROSSBERG, S. 1994. 3-D vision and figure-ground separation by visual cortex. *Perception & Psychophysics, 55,* 48-120.

GROSSBERG, S. 1997. Cortical dynamics of three-dimensional figure-ground perception of two-dimensional pictures. *Psychological Review, 104,* 618-658.

HORN, B. K. P. 1986. *Robot vision.* Cambridge, MA: MIT Press.

JOLION, J. M., AND ROSENFELD, A. 1994. *A pyramidal framework for early vision.* Dordrecht, The Netherlands: Kluwer Academic Publishers.

KELLMAN, P. J., AND SHIPLEY, T. F. 1991. A theory of visual interpolation in object perception. *Cognitive Psychology, 23,* 141-221.

KOFFKA, K. 1935. *Principles of Gestalt psychology.* New York: Harcourt Brace.

NAKAYAMA, K., SHIMOJO, S., AND SILVERMAN, G. H. 1989. Stereoscopic depth: Its relation to image segmentation, grouping, and the recognition of occluded objects. *Perception, 18,* 55-68.

PALMER, S., AND ROCK, I. 1994a. On the nature and order of organizational processing: A reply to Peterson. *Psychonomic Bulletin & Review, 1,* 515-519.

PALMER, S., AND ROCK, I. 1994b. Rethinking perceptual organization: the role of uniform connectedness. *Psychonomic Bulletin & Review, 1,* 29-55.

PALMER, S. E. 1999. *Vision science: Photons to phenomenology.* Cambridge, MA: MIT Press.

PETERSON, M. A. 1999. What's in a stage name? Comment on Vecera and O'Reilly (1998). *Journal of Experimental Psychology: Human Perception and Performance, 25,* 276-286.

PIZLO, Z., SALACH-GOLYSKA, M., AND ROSENFELD, A. 1997. Curve detection in a noisy image. *Vision Research, 37,* 1217-1241.

REGAN, D., AND BEVERLEY, K. I. 1985. Postadaptation orientation discrimination. *Journal of the Optical Society of America (A), 2,* 147-155.

ROSENFELD, A. 1990. Pyramid algorithms for efficient vision. In C. BLAKEMORE (Ed.), *Vision: Coding and efficiency* (pp. 423-430). Cambridge, Great Britain: Cambridge University Press.

RUBIN, E. 1958. Figure and ground. In D. C. BEARDSLEE AND M. WERTHEIMER (Eds.), Trans., M. Wertheimer, *Readings in Perception* (pp. 194-203). Princeton, NJ: D. Van Nostrand Company, Inc. [An abridged translation by M. Wertheimer of pp. 35-101 of Rubin, E., *Visuell wahrgenommene Figuren* (translated by P. Collett into German from the Danish *Synsoplevede Figurer*, Copenhagen: Gyldendalske, 1915). Copenhagen: Gyldendalske, 1921.]

SCHEESSELE, M. R., AND PEREZ, T. M. 2003. Effect of region information on perception of partially occluded figures [Abstract]. *Journal of Vision* (to appear).

SHIPLEY, T. F., AND KELLMAN, P. J. 1992. Strength of visual interpolation depends on the ratio of physically specified to total edge length. *Perception & Psychophysics, 52,* 97-106.

SINGH, M., HOFFMAN, D. D., AND ALBERT, M. K. 1999. Contour completion and relative depth: Petter's rule and support ratio. *Psychological Science, 10,* 423-428.

TREISMAN, A. 1986. Properties, parts, and objects. In K. R. BOFF, L. KAUFMANN, AND J. P. THOMAS (Eds.), *Handbook of human perception and performance.* New York: Wiley.

TSE, P. U., AND ALBERT, M. K. 1998. Amodal completion in the absence of image tangent discontinuities. *Perception, 27,* 455-464.

VECERA, S. P., AND O'REILLY, R. C. 2000. Graded effects in hierarchical figure-ground organization: Reply to Peterson (1999). *Journal of Experimental Psychology: Human Perception and Performance, 26,* 1221-1231.

WERTHEIMER, M. 1958. Principles of Perceptual Organization. In D. C. BEARDSLEE AND M. WERTHEIMER (Eds.), Trans., M. Wertheimer, *Readings in Perception* (pp. 115-135). Princeton, NJ: D. Van Nostrand Company, Inc. [An abridged translation by M. Wertheimer of Wertheimer, M., Untersuchungen zur Lehre von der Gestalt, II., *Psychol. Forsch., 4,* pp. 301-350, 1923.]

YIN, C., KELLMAN, P. J., AND SHIPLEY, T. F. 1997. Surface completion complements boundary interpolation in the visual integration of partly occluded objects. *Perception, 26,* 1459-1479.