Object Recognition in Dense Clutter

MARY J. BRAVO, [1] HANY FARID, [2]

Observers in recognition experiments invariably view objects against a blank background, while observers of real scenes sometimes view objects against dense clutter. In this study, we examined whether an object's background affects the information used for recognition. Our stimuli consisted of color photographs of everyday objects. The photographs were either organized as a sparse array, as is typical of a visual search experiment, or as high density clutter such as might be found in a toy chest, a handbag or a kitchen drawer. The observer's task was to locate an animal, vehicle or food target in the stimulus. We varied the information in the stimuli by convolving them with a low pass filter (blur), a high pass filter (edge) or converting them to grayscale. In two experiments, we found that the blur and edge manipulations produced a modest decrement in performance with the sparse arrangement but a severe decrement in performance with the clutter arrangement. These results indicate that the information used for recognition depends on the object's background. Thus, models of recognition that have been developed for isolated objects may not generalize to objects in dense clutter.

KEYWORDS: Object Segmentation Object Recognition Visual Search

[1] Department of Psychology, Rutgers University, Camden NJ 08102. Email: mbravo@camden.rutgers.edu; Tel: 856.225.6431; Fax: 856.225.6602

[2] Department of Computer Science and Center for Cognitive Neuroscience, Dartmouth College, Hanover NH 03755.

1 Introduction

As a fundamental task of vision, object recognition has been the focus of a great deal of research. Most of this research has involved single objects presented on blank backgrounds. Implicit in the use of such presegmented objects is the assumption that it is possible to separate the processes involved in object recognition from those involved in object segmentation. While blank backgrounds may allow these processes to be dissociated in the lab, the cluttered backgrounds of real scenes may make these processes inseparable in everyday vision (Barrow & Tenenbaum, 1981; Spelke, 1990; Dueker, Modi, & Needham, 2003; Borenstein & Ullman, 2002; Bravo & Farid, 2003). In particular, when a complex object is viewed against dense clutter, observers may need to recognize the object before they can fully segment it from its background.

If background clutter prevents object segmentation from occurring before object recognition, then the vast literature on the recognition of presegmented objects may not generalize to objects in clutter. One of the principal findings from this literature is the central role that shape plays in recognition. Compared with other attributes such as color, texture, and local features, shape is thought to provide especially potent information for recognition (Ullman, 1997). Some researchers have claimed further that the global shape information carried by low spatial frequencies plays a key role. As one example, (Bar, 2003) proposed that recognition involves top-down processes that are initiated by low spatial frequency information. Because this information is quickly relayed through the magnocellular pathway, he suggested that low spatial frequencies are used to generate a fast, initial conjecture of an object's identity. As a second, very different, example, French et al. have claimed that low spatial frequencies are particularly useful for object category learning (French, Mermillod, Quinn, Chauvin, & Mareschal, 2002). French et al. suggest that the blurry vision of infants causes them to ignore the details that distinguish objects within a category and instead attend to the global shapes that distinguish objects in different categories.

As with most recognition research, the experiments examining the role of low spatial frequencies in object recognition have used isolated objects on blank backgrounds (Olds & Engel, 1998; French et al., 2002; Fiser, Subramaniam, & Biederman, 2001; Archambault, Gosselin, & Schyns, 2000; Braje, Tjan, & Legge, 1995). The middle pair of images in Figure 1 suggests that the results from these experiments may critically depend on using a blank background. In this figure, it is easy to recognize the bottom blurry image of a beagle against a homogeneous gray field. The blur

Figure 1: The presegmented dog (bottom row) is recognizable even when the image has been degraded by blurring (middle) or edge-detection (right). The dog viewed against a cluttered background (top row) is recognizable in the original image but is more difficult to discern in the blurred and edge-detected images.

does not prevent recognition because the variously colored blobs in the image form a dog-shaped object. It is far more difficult to recognize the top blurry image of the beagle against a cluttered background. In this case, the blobs of the beagle merge with blobs in the background to form spurious groupings. This figure suggests that information used to recognize objects in isolation may differ from that used to recognize objects in clutter.

Sanocki has advanced a similar argument (Sanocki, Bowyer, Heath, & Sarkar, 1998) about theories of object recognition that emphasize the role of edges (Biederman, 1987; Marr & Nishihara, 1978; Peterson, 1994). The right pair of images in Figure 1 illustrates the problem. It is easy to recognize the beagle in the edge-extracted image when the beagle appears against a blank background. In this case, the bounding contours form a recognizable silhouette. Recognition is much more difficult with the edge-extracted image of the beagle against a cluttered background. Here, the bounding contours of the dog are masked by the numerous edges in the cluttered background and in the dog's mottled coat. It seems again that the information that is effective for the recognition of presegmented objects may not be as effective for the recognition of unsegmented objects.

This paper reports two experiments designed to empirically confirm these observations. In these experiments, we measured how various image manipulations affect the recognition of objects presented in isolation and objects presented in clutter. Our observers searched for a category target (e.g., an animal or vehicle) in a stimulus that was composed of color photographs of everyday objects. In half the displays, the target and distractor objects were organized as a sparse array so that each of the objects was effectively presegmented. In the other half of the displays, the distractors were arranged randomly within a small area, and the target was overlaid on this background of dense clutter. This dense clutter, while more extreme than that found in most scenes, resembled the crowdedness of a toy chest or kitchen drawer. These two types of displays were then convolved with a low pass-filter (the blur condition), convolved with a high-pass filter (the edge condition), converted to grayscale (the grayscale condition) or left untouched (the none condition).

With the sparse displays, we expected that observers would rely on global shape to recognize the objects and we assumed that much of this shape information would be discernible even when the images contained only low spatial frequencies, edges, or achromatic information. Thus, we expected that observers would quickly and reliably identify the target in the sparse displays regardless of the image manipulation. With the clutter displays, we expected a different pattern of results. Because clutter can obscure the connectedness of the parts of an object, it can conceal the object's shape. As a result, recognition in clutter cannot always rely on global shape and so must tap other object attributes such as color, texture and local contrast features. Although these attributes are not obscured by clutter, they are obscured by the image manipulations we used. Thus, we expected that observers would have greater difficulty finding the targets in clutter when the displays were blurred, edge-detected or converted to grayscale.

2 Methods

2.1 Stimuli

The stimuli were generated in MatLab using photographs from the Hemera Photo Objects collection. To generate each stimulus, 7, 15 or 23 distractors were selected randomly and without replacement from 150 photographs of everyday, man-made objects (mostly tools, kitchenware and furniture). The target was selected randomly from 100 animal or 100 vehicle photographs. We used two target categories because we wanted to maximize the observer's uncertainty of the target's features. The objects depicted in the images ranged in size from a jet airplane to a beetle, but the images themselves were all initially scaled to have the same area ($16,000$ pixels). Before an image was added to a display, its area was rescaled by a random factor between 1.0 and 0.5. We varied the sizes of the items to discourage observers from attending to a particular scale. Distractor objects were randomly rotated by 0, 90, 180 or 270 degrees. We rotated the distractors to make it more difficult for observers to learn the distractor set. Target objects were not rotated; they always appeared in their upright orientation.

For the sparse stimuli, the objects were arranged in a 6×4 grid, a 4×4 grid, or a ring around fixation. For the clutter stimuli, the distractors were positioned randomly within a small area and so they often overlapped Figure 2. The targets were also positioned randomly in the clutter displays, subject to one restriction. Because the observer's task was to identify the stimulus quadrant that contained the target, the target's center of mass was always at least 50 pixels (about 1 degree of visual angle) from the vertical and horizontal midlines. The target was added last to the clutter displays, so it was never occluded. The size of the clutter displays increased as the number of distractors increased so that object density was constant. The observer's viewing distance was unrestricted.

The stimuli were manipulated using the following standard techniques:

- Grayscale: Color (RGB) images were converted to grayscale by linearly combining the three color channels using the weightings: $0.299R + 0.587G + 0.114B$.

- Blur: Color images were blurred with a Gaussian 25-tap filter: $h(x) = \exp(-x^2/(2\sigma^2))$, with $\sigma = 3.75$. Each color channel was convolved separately in both the horizontal and vertical direction with this 1-D filter. In the Fourier domain, this filter had a standard deviation of roughly 18 cycles per object.

- Edge: Grayscale images were convolved with separable lowpass and highpass filters to obtain partial horizontal, $f_x(x, y)$, and vertical, $f_y(x, y)$, derivatives. These partial derivatives were then used to compute the gradient, $\sqrt{f_x^2(x, y) + f_y^2(x, y)}$ and the results were normalized into the range $[0, 1]$. To enhance the edges, values above 0.0075 were passed through a point-wise non-linearity, $g(u) = u^{1/3}$. (Values below this threshold were set to zero.) The image was then contrast reversed so that the edges appeared black on a white background.

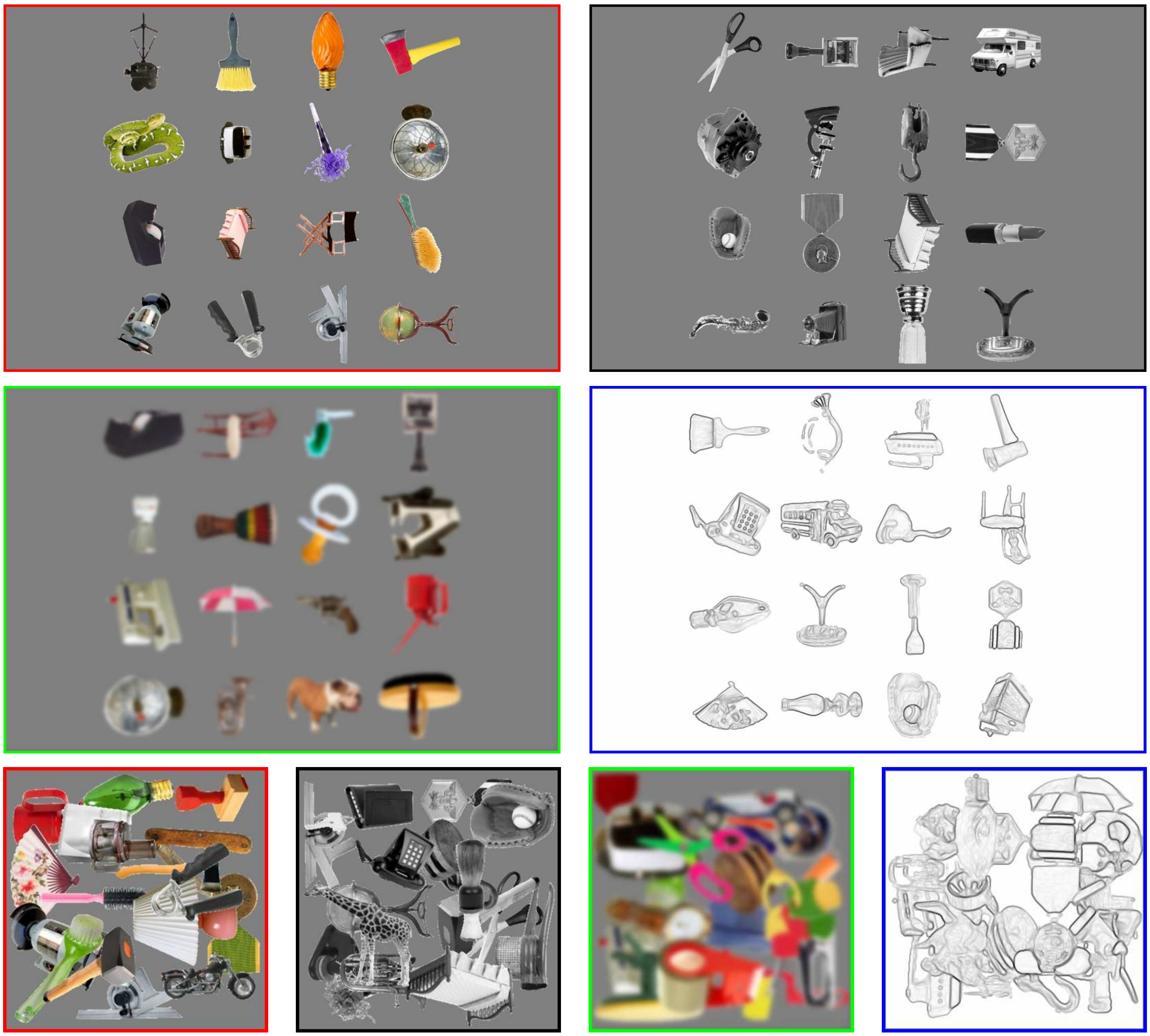

Figure 2: Examples of sparse and clutter stimuli and the four image manipulations: none, grayscale, blur and edge.

2.2 Procedure

The experiment was run on Apple PowerBook computers using MatLab and PsychToolbox routines (Brainard, 1997; Pelli, 1997). In all, there were 48 conditions: two object arrangements (sparse and clutter); two target categories (animal and vehicle); three levels of object number (8,16 and 24) and four image manipulations (edge, blur, grayscale and none). The arrangement varied across observers. Object number, target category, and image manipulation varied within observers and within blocks of trials.

The observer initiated the first trial of each block. The stimulus remained on until the observer responded by pressing one of four keys to indicate the display quadrant that contained the target. Auditory feedback was given after incorrect responses and the next stimulus was presented after a one second delay. After running a block of 15 practice trials, each observer ran 12 blocks of 48 trials in a single 50-minute session.

2.3 Observers

Eighteen observers participated in this experiment. The observers were recruited from the Introductory Psychology subject pool at Rutgers University in Camden. All observers reported having normal color vision and normal or corrected-to-normal acuity.

3 Results and Discussion

This experiment tested the prediction that image manipulations like blurring and edge-detection would have a modest effect on search with presegmented objects and a severe effect on search in clutter. Figure 3 shows the search times and error rates for the sparse and clutter conditions. (Only correct trials were used to calculate search times.) Each graph has four functions corresponding to the edge, blur, grayscale and none image manipulations. Consider first the results from the sparse condition in which the objects were effectively presegmented (top row). In this condition, the separations between the four error rate functions and between the four response time functions are relatively small. The edge condition ("x") was the most difficult, but even in this condition, observers had an error rate of less than 10%. This pattern of results differs from that observed for the clutter condition (bottom row). In the graphs for this condition, the error

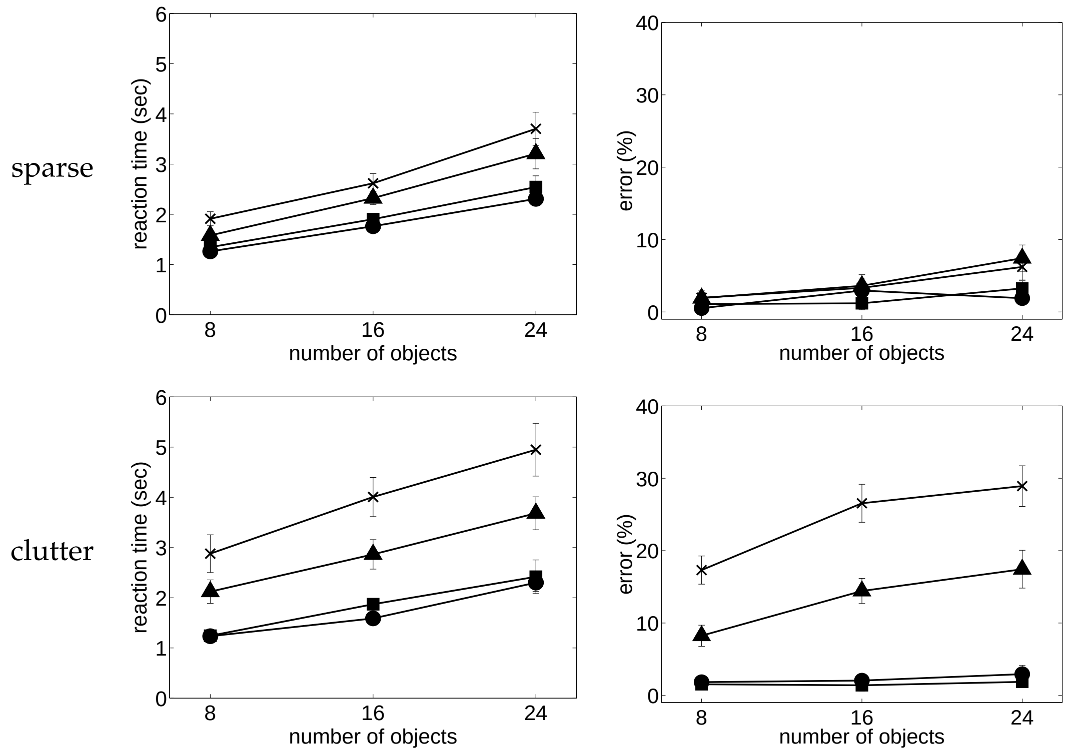

Figure 3: Results from experiment one. The top row corresponds to the sparse arrangement condition, the bottom row to the clutter arrangement condition. For each type of arrangement, there were four image manipulations: color (circle); gray (square); blur(triangle); edge (x). Note that with cluttered stimuli, the blur and edge conditions pose a considerable challenge for observers.

rate and response time functions are widely splayed. Although the grayscale manipulation (filled squares) had little effect on performance, the response times for the edge and blur manipulations were very slow, and observers often failed to locate the target in these displays. This difference in the pattern of results for the two arrangements was confirmed by an ANOVA. There was a significant interaction between stimulus arrangement and image manipulation for both mean response times ($F(3, 144) = 12.32$, $p = 0.000004$) and error rates ($F(3, 144) = 67.6$, $p < 0.000001$).

The edge and blur filters remove information from the stimulus, and so one might expect that these image manipulations would have an especially strong impact on a difficult task. But the clutter condition was not more difficult than the presegmented condition; performance on the unfiltered images (the none condition) was very similar for the two arrangements. This suggests that the differential effect of blurring and edge detection arises from a qualitative difference in the processing of the two conditions. As we will explain in the general discussion, we think this qualitative difference corresponds to different types of information being used to recognize objects in isolation and objects in clutter.

Although we found a clear effect of the blur and edge conditions on performance, we found no effect with the gray condition. On the surface, this is a surprising result; observes might be expected to rely heavily on color in the clutter displays because color can be used both for segmentation and recognition. But two aspects of our displays may have reduced the effect of color. First, many of the objects in our displays had multiple parts with different colors. For such objects, color discontinuities accentuate the boundaries within objects as well as the boundaries between objects. Thus, color information might have resulted in an over-segmentation of the stimulus. And second, color may play a smaller role in the recognition of animals and vehicles than it does in the recognition of other categories like food. Given that the choice of target category may affect the information used for recognition, we thought it important to repeat the first experiment with a new target category.

4 Experiment 2

The first experiment examined how the recognition process is affected by the target's background. It is likely that this process will also be affected by the target's category. For example, there is evidence that for both humans and monkeys, color plays a larger role in the recognition of some categories than others, although the size of this effect seems to depend on the observer's task (Wurm, Legge, Isenberg, & Luebker, 1993; Tanaka, Weiskopf, & Williams, 2001; Rossion & Pourtois, 2004; Santos, Hauser, & Spelke, 2001). Foods, it has been observed, often have generic shapes but characteristic colors; oranges are reliably and distinctively orange. In contrast, vehicles often have distinctive shapes but variable color; an Audi has a uniquely elegant form and although it should only be available in black, it comes in a variety of colors. Thus, color may be important for the recognition of foods but not vehicles. If so, one might expect an effect of the gray condition with food targets even though none was found for animal and vehicle targets. If the choice of target category affects the information used for recognition, then it is important to test the robustness of our results with different target categories. In this second experiment, we again measured the effect of various image manipulations on search in sparse and cluttered stimuli, but we varied the target category (vehicle, animal or food) across observers. Unlike the previous experiment, the different categories were not intermixed within blocks of trials.

4.1 Methods

This second experiment was similar to the first experiment with the following exceptions. A third target set of 100 food items was added to the sets of 100 animal and 100 vehicles. Instead of intermixing target categories, each observer searched for targets belonging to a single category. In all, there were 72 conditions (two arrangements, three target types, three levels of distractor number, and four image manipulations). Two variables (the number of distractors and the image manipulation), were varied within observers. The other two variables (arrangement and target type) were varied across observers. Fifty-four observers participated in this second experiment, with nine observers assigned to each of the six conditions that varied across observers.

4.2 Results and Discussion

In this second experiment, we tested whether the results of our first experiment generalize across target categories. The results for the three target categories are displayed in Figure 4.2. Again, the relevant comparison is between the spread of the data for the clutter and sparse conditions. We expected, as before, that these graphs would be noticeably closer in the sparse condition than in the clutter condition. For each target type, this interaction between arrangement and image manipulation was significant for at least one performance measure. (The significance levels for the analysis on response times were $F[3, 48] = 4.07$, $p = 0.0117$, $F[3, 48] = 6.22$, $p = 0.00117$, $F[3, 48] = 0.31$, $p = 0.81$ for vehicles, animals and food, respectively. The significance levels for the error rates were $F[3, 48] = 631$, $p < 0.000001$, $F[3, 48] = 38.44$, $p < 0.000001$, and $F[3, 48] = 27$, $p = 0.0009$ for vehicles, animals and food, respectively.) These results suggest that our findings from experiment one apply to at least three important object categories.

It is important to note, however, that the pattern of results varied across the three target types: there was a significant three-way interaction between target type, arrangement and image manipulation for both the response time functions (($F[6, 144] = 2.56$, $p = 0.022$) and error rate functions ($F[6, 144] = 3.50$, $p = 0.003$)). As Figure 4.2 shows, the effect of arrangement was much less compelling for the food targets than it was for the vehicle and animal targets. Before conduct-

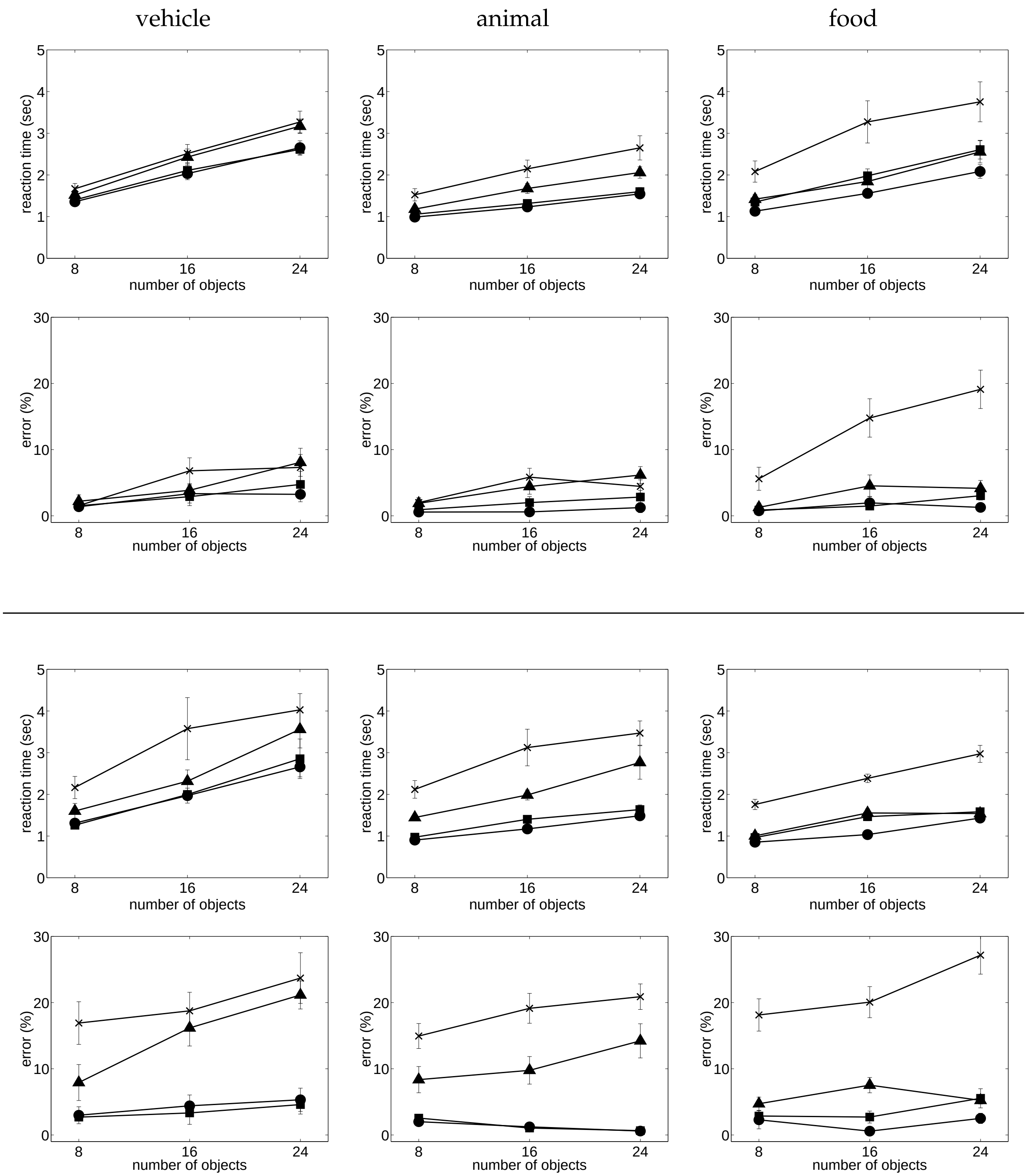

Figure 4: Results from experiment two. The top two rows corresponds to the sparse arrangement condition, and the bottom two rows to the clutter arrangement condition. For each type of arrangement, there were four image manipulations: color (circle); gray (square); blur(triangle); edge (x).

ing this experiment, we thought that, unlike animal and vehicle targets, search for food targets would rely heavily on color. But the most striking difference across target types was not with the grayscale condition, but was instead with the edge condition under the sparse arrangement. Unlike the animal and vehicle targets, the food targets in the edge condition were difficult for observers regardless of the stimulus arrangement. Clearly, edge information alone is inadequate for the recognition of food, even when the food object is presegmented. This finding is discussed further below.

5 General Discussion

Understanding object recognition is one of the greatest challenges for vision scientists. In teasing apart this difficult problem, many researchers have tried to identify the information that is most useful for recognition. Some researchers, impressed by the salience of line drawings, have emphasized edges. Other researchers, inspired by the nature of early visual processing, have examined spatial frequency. Still other researchers, motivated by the ready intelligibility of black and white images, have questioned whether color plays a role. The current study adds to the existing evidence that the information used for recognition depends on the object category being recognized. The current study makes a unique contribution by showing that the information used for recognition also depends on the object's background. We discuss these two effects below.

In these experiments, we examined how search for a category target was affected by the image manipulations of edge-detection, blurring, and conversion to grayscale. In each display, one target was located among 7-23 distractors. Because the distractors far outnumbered the target, we initially assumed that the target category would have little effect on the pattern of results. This assumption turned out to be wrong: there was a significant interaction between target type and image manipulation on both response times and error rates. The most striking example of this interaction was the differential effect of edge detection on food targets compared with animal and vehicle targets. With food targets, edge detection greatly impaired performance even when the objects were presegmented so as to reveal the object's bounding contours. Bounding contours can provide salient shape information, and for many objects, shape information may be sufficient for recognition. But foods often have generic shapes and it may be their color and texture that is distinctive. When only one of these distinctive cues was missing, as in the blur and grayscale conditions, observers could still perform the task reasonably well. But when both cues were missing, as in the edge condition, performance was poor. For simplicity, the remainder of the discussion focuses on the average effects of the image manipulations, but it is important to keep in mind that these effects do depend on target category.

The central focus of this study was on how an object's background can affect the information used for recognition. To examine this, we used two types of stimulus arrangements. In one arrangement, the target and distractors were presegmented in a sparse array. In the other arrangement, these objects were randomly arranged as dense clutter. As mentioned above, we used the image manipulations edge, blur and grayscale to examine the importance of different forms of information for recognition. We found that the grayscale condition had little effect on performance regardless of the arrangement. This result differs from experiments in which observers named isolated objects (Wurm et al., 1993; Tanaka et al., 2001; Rossion & Pourtois, 2004), but it is consistent with a study, more similar to our own, in which observers detected members of an object category in natural images (Delorme, Richard, & Fabre-Thorpe, 2000). In contrast to the grayscale condition, the edge and blur conditions produced a severe effect on search with the clutter arrays but only a modest effect on search with the sparse arrays. This is, of course, the result we expected based on our observations of Figure 1. We take the interaction between image manipulation and stimulus arrangement as evidence that different types of information are used to recognize objects in isolation and object in dense clutter.

With isolated objects, shape plays a central role in the recognition of many object categories. Some researchers even use the terms object recognition and shape recognition interchangeably. Although researchers disagree on the nature of the shape representation, the bounding contours of the object are clearly an important component. Because our manipulations leave this contour largely unperturbed, these manipulations do not severely impair the recognition of presegmented animals and vehicles.

With objects in clutter, shape often plays a much smaller role in recognition. In clutter, objects overlay one another, and there may be insufficient visual cues to indicate which parts of the image are connected, and so belong to the same object, and which are simply adjacent. Thus, with cluttered stimuli, observers may be unable to delineate the bounding contours of an object without first recognizing the object. Because recognition in clutter cannot always rely on global shape, this process may rely more on local information, such as image fragments (Ullman, Sali, &

Vidal-Naquet, 2001; Lowe, 2000; Burl, Weber, & Perona, 1998). These fragments are thought to combine local texture and contour information in a way that gives them sufficient complexity to be diagnostic of a small set of objects. Because blurring and edge-detection significantly degrade the distinctive information in these fragments, these manipulations impair recognition in clutter.

We motivated this research with the claim that objects are infrequently viewed against blank backgrounds. But clearly, our clutter stimuli represent an opposite extreme that is also unusual in everyday life. Objects most often appear against familiar structured backgrounds, and such backgrounds may actually assist recognition. Experimenters have shown that observers use their knowledge of scene schemas to bias the recognition of ambiguous stimuli in favor of objects that are likely to appear in a given scene location (for a review, see (Bar, 2004)). Nonetheless, there are everyday search tasks that involve objects that are densely and randomly arranged. Examples of such search tasks include looking for a bottle opener in a kitchen drawer, looking for car keys in a handbag or looking for a doll in a toy chest. In these situations, the extreme background clutter may prevent the segmentation of whole objects, and thus prevent shape-based recognition.

Researchers in computer vision have long struggled with the problem of clutter and many computational models of object recognition are based on local fragments rather than global shape. In contrast, this idea has received relatively little attention in human vision. This is true even though the perception literature abounds with demonstrations showing that object recognition may occur before object segmentation. The most famous demonstrations of this include the Dalmatian dog image by R. C. James (Gregory, 1973) pg. 14), the cow image by Kundel and Nodine (Kundel & Nodine, 1983), and many of the classic reversible figures, including Rubin's face-vase image (Peterson, 1994). But all of these demonstrations involve either greatly impoverished images or carefully constructed drawings, and so many researchers may have assumed that ordinary scenes contain sufficient information for bottom-up object segmentation. There are situations, however, when the density of the clutter may be so great that object recognition cannot rely on having access to presegmented objects. For a complete understanding of recognition it is important that research in human vision address the problem of clutter.

Acknowledgments

We thank J. William Whitlow for performing the statistical analysis of our data. This work was supported by NSF grant SBR-9729015 (Bravo), and an Alfred P. Sloan Fellowship, NSF CAREER award IIS-99-83806, and a departmental NSF grant EIA-98-02068 (Farid).

6 References

Archambault, A., Gosselin, F., & Schyns, P. (2000). A natural bias for the basic level? *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*. Lawrence Erlbaum and Associates.

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, *69*, 600–609.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*, 617–629.

Barrow, H., & Tenenbaum, J. (1981). Computational vision. *Proceedings of the IEEE*, *69*, 572–595.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115–147.

Borenstein, E., & Ullman, S. (2002). Class-specific, top-down segmentation. In A. Heyden, G. Sparr, M. Nielsen, & P. Johansen (Eds.), *European Conference on Computer Vision (ECCV)*. Springer-Verlag.

Brainard, D. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.

Braje, W., Tjan, B., & Legge, G. (1995). Human efficiency for recognizing and detecting low-pass filtered objects. *Vision Research*, *35*, 2955–2966.

Bravo, M., & Farid, H. (2003). Object segmentation by top-down processes. *Visual Cognition*, *10*, 471–491.

Burl, M., Weber, M., & Perona, P. (1998). A probabilistic approach to object recognition using local photometry and global geometry. *European Conference on Computer Vision (ECCV)* (pp. 628–641).

Delorme, A., Richard, G., & Fabre-Thorpe, M. (2000). Ultra-rapid categorization of natural scenes does not rely on colour cues: A study of monkeys and humans. *Vision Research*, *40*, 2187–2200.

Dueker, G., Modi, A., & Needham, A. (2003). 4.5-month-old infants' learning, retention and use of object boundary information. *Infant Behavior and Development*, *26*, 588–605.

Fiser, J., Subramaniam, S., & Biederman, I. (2001). Size tuning in the absence of spatial frequency tuning in object recognition. *Vision Research*, *41*, 1931–1950.

French, R., Mermillod, M., Quinn, P., Chauvin, A., & Mareschal, D. (2002). The importance of starting blurry: simulating improved basic-level category learning in infants due to weak visual acuity. *Proceedings of the 24th Annual Conference of the Cognitive Science Society*. Lawrence Erlbaum and Associates.

Gregory, R. L. (1973). *The intelligent eye*. McGraw-Hill.

Kundel, H., & Nodine, C. (1983). A visual concept shapes image perception. *Radiology*, *146*, 363–368.

Lowe, D. (2000). Towards a computational model for object recognition in it cortex. *First IEEE International Workshop on Biologically Motivated Computer Vision*. Cambridge University Press.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society, London B*, *200*, 269–294.

Olds, E., & Engel, S. (1998). Linearity across spatial frequency in object recognition. *Vision Research*, *38*, 2109–2118.

Pelli, D. (1997). The videotoolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

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

Rossion, B., & Pourtois, G. (2004). Revisiting snodgrass and vanderwart's obect pictorial set: the role of surface detail in basic-level object recognition. *Perception*, *33*, 217–236.

Sanocki, T., Bowyer, K., Heath, M., & Sarkar, S. (1998). Are edges sufficient for object recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 340–349.

Santos, L., Hauser, M., & Spelke, E. (2001). Recognition and categorization of biologically significant objects by rhesus monkeys (macaca mulatta): the domain of food. *Cognition*, *82*, 127–155.

Spelke, E. (1990). Principles of object perception. *Cognitive Science*, *14*, 29–56.

Tanaka, J., Weiskopf, D., & Williams, P. (2001). The role of color in high-level vision. *Trends in Cognitive Sciences*, *5*, 211–215.

Ullman, S. (1997). *High-level vision: Object recognition and visual cognition*. Bradford/MIT Press.

Ullman, S., Sali, E., & Vidal-Naquet, M. (2001). A fragment based approach to object representation and classification. In C. Arcelli, L. Cordella, & G. S. di Baja (Eds.), *Visual Form 2001, 4th International Workshop on Visual Form, IWVF-4, Capri, Italy*.

Wurm, L., Legge, G. E., Isenberg, L., & Luebker, A. (1993). Color improves object recognition in normal and low vision. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 899–911.