Journal of Experimental Psychology:
Human Perception and Performance
1998, Vol. 24, No. 1, 340-349

0096-1523/98/$3.00

Are Edges Sufficient for Object Recognition?

Thomas Sanocki, Kevin W. Bowyer, Michael D. Heath, and Sudeep Sarkar
University of South Florida

The authors argue that the concept of "edges" as used in current research on object recognition obscures the significant difficulties involved in interpreting stimulus information. Edges have sometimes been operationalized as line drawings, which can be an invalid and misleading practice. A new method for evaluating the utility of edge information, operationalized as the outputs of a local, signal-based edge extractor, is introduced. With 1-s exposures, the accuracy of identifying objects in the edge images was found to be less than half that with color photographs. Therefore, edges are far from being sufficient for object recognition. Alternative approaches to the problem of interpreting stimulus information are discussed.

As with any complex process, initial research on object recognition has been conducted within frameworks that simplify some aspects of the process. Simplification is important and necessary during initial stages of research. However, simplified frameworks must be elaborated upon or replaced on the road to a fuller understanding. In this article, we argue that the current concept of edge-based representations should be replaced because it is a simplification that obscures significant problems involved in the interpretation of stimulus information during object recognition.

We will use Biederman's (1987) recognition-by-components (RBC) model to introduce our arguments because edge-based representations are central to it and because it has been an important model of object recognition in psychology. However, similar assumptions are made in many other models in cognitive science (e.g., Bergevin & Levine, 1993; Brooks, 1981; Grimson, 1989; Hummel & Biederman, 1992; Huttenlocher & Ullman, 1990; Lowe, 1987; Stark, Eggert, & Bowyer, 1988).

A major contribution of RBC is its explanation of how representations of objects can be derived from two-dimensional images in a bottom-up manner using relatively simple, easy-to-compute information. Processing is assumed to begin as "an early edge extraction stage, responsive to differences in surface characteristics namely, luminance, texture, or color, provides a line drawing description of the object" (Biederman, 1987, p. 117). This stage determines which information (which edges) will be available for processing; all of the other main processes depend on its outputs (see Biederman, 1987, Figure 2). Given this assumption, the research process can be simplified by using line drawings or their equivalents as inputs to subsequent stages of processing (Biederman, 1987; see also Hummel & Biederman, 1992). In subsequent stages, nonaccidental properties are detected from the line drawing input and used to determine which parts (geons) are present in the object. In parallel with the detection of nonaccidental properties, a parsing process uses edge information in the line drawing to divide the stimulus into parts so that nonaccidental properties can be assigned to the correct geon.

Our concern is with the sufficiency of the processes that extract stimulus information. Can a simple, edge-based extraction process really produce information that is sufficient for recognizing objects? If an edge-based process is not sufficient, how insufficient is it? If edge information is *very* insufficient for recognition, then significant revisions may be necessary in many models of object recognition.

We focus mainly on edge-extraction processes that are local and signal based. *Local* means that decisions are made about individual edge segments by considering only information within small regions of the image. Signal based means that the process involves stimulus information but not other types of information (such as knowledge about identity or about how edges combine to form vertices). This class of edge processes is consistent with the main emphasis in RBC. It is also the primary class of edge extractor being researched in the field of computer vision. We discuss alternative approaches after presenting our experiment.

How can the sufficiency of an approach to edge extraction (or stimulus interpretation in general) be evaluated? Unfortunately, there are no well-established tools for such an evaluation (see, e.g., Heath, Sarkar, Sanocki, & Bowyer, in press-a). The most potentially relevant experiments have

Thomas Sanocki, Department of Psychology, University of South Florida; Kevin W. Bowyer, Michael D. Heath, and Sudeep Sarkar, Department of Computer Science and Engineering, University of South Florida.

This research was supported by Grant DBS-9213246 from the National Science Foundation, Grant F49620-92-J-0223 from the Air Force Office of Scientific Research, and a Sabbatical Award to Thomas Sanocki from the University of South Florida. Portions of this article were presented at the meeting of the Association for Research in Vision and Opthalmology in Sarasota, Florida, May 1995 (Sanocki, Bowyer, Adair, & Sarkar, 1995), and at the Perception and Modeling of Objects Meeting in Chicago, Illinois, November 1996.

We thank Ennio Mingolla for helpful comments on a draft of this article and Eric Sellers and Chris Woodard for their help in conducting the experiments.

Correspondence concerning this article should be addressed to Thomas Sanocki, Department of Psychology, BEH 339, University of South Florida, 4202 East Fowler Avenue, Tampa, Florida 33620-8200. Electronic mail may be sent via Internet to sanocki@chuma.cas.usf.edu.

used line drawings as an operationalization of edge-based images (e.g., Biederman & Ju, 1988). However, this practice has yielded misleading conclusions about the sufficiency of edge information.

These experiments were designed to examine the adequacy of edge-based models of object recognition in general by comparing the ability of humans to recognize objects represented as line drawings and as color photographs (e.g., Biederman & Ju, 1988). The logic of the experiments was that if edge-based representations are crucial for object recognition, then objects should be recognized as easily when represented by edge information as when represented by other types of information (i.e., by color photographs, which contain surface information such as color, texture, and relative brightness). As noted, line drawings were used as the operational definition of edge information, a practice that can be justified by the assumption identified above—that edge-extraction processes produce a line-drawing description of the stimulus. Biederman and Ju found performance to be equal with color photographs and line drawings, and they concluded that the adequacy of edge-based models was supported (see also Biederman, 1987, pp. 131–133).[1] This has been a widely cited and influential result, and it has been used to support the sufficiency of edge-based approaches. For example, Lowe (1987) cited the Biederman and Ju (1988) results in support of his edge-based model, and more recently Bergevin and Levine (1993) cited the results in support of their edge-based model. They argued that the Biederman and Ju (1988) results "suggest that, at least for generic object recognition, building coarse descriptions from single view edge maps or line drawings might be an appropriate, or even required, alternative to surface and volume reconstruction" (Bergevin & Levine, 1993, p. 19).

There are several problems with applying these results to the problems of stimulus interpretation. First, the use of line drawings implies that the problems have been solved, because an edge-based representation exists, when the solution has only been assumed. One of our goals is to illuminate the difficulty of stimulus interpretation. A second and related problem is that when edge extraction and line drawings are equated, the differences between edge extraction and line drawings are obscured. We argue that these differences are considerable and of particular significance for theories of object recognition. A third problem is that the experiments simplify stimulus interpretation by using isolated objects rather than objects in natural settings; this practice can also contribute to the overestimation of the sufficiency of the edge-based information.

In the next three sections, we elaborate on the differences between edge images and line drawings and on the importance of context. Then we apply a new method for examining the sufficiency of an edge-extraction process using an industry-standard computer vision extractor (Canny, 1986).

Edges Are Not Line Drawings

This brief review of research in computer edge extraction begins to distinguish between edge images and line drawings and also illustrates some of the difficulties of stimulus interpretation. Current thinking about edge information and object recognition has been strongly influenced by the seminal work of Marr and Hildreth (1980) on computer edge extraction. Marr and Hildreth demonstrated that edges can be detected in a local, signal-based manner by biologically plausible detectors that measure discontinuities in stimulus luminance information (*zero crossings*).

There are two important arguments for the local, signal-based approach. First, because domain-specific knowledge is not used, signal-based edge extractors have the potential to serve as a general stage of early processing for a wide variety of stimulus domains, including both manufactured and natural objects. Second, because this early stage of processing is local and signal based, it is independent of higher level processes that involve knowledge and global constraints. Marr and Hildreth's (1980) research was followed by a large amount of computational research that was also local and signal based. A review of edge-extraction research was provided by Boyer and Sarkar (1992; see also Heath et al., in press-a).

The Marr–Hildreth edge extractor produced edge segments (single pixels) that were linked to each other to form edge chains in a resultant edge image. Typically, all edge chains longer than a threshold length (possibly zero) would be retained. However, the set of edges resulting from this process is quite different from the highly structured contours (and implicit regions) present in line drawings. There are several "roadblocks" on the way to line drawings. These problems apply to all current signal-based edge extractors.

First, edge extractors fail to detect some edges (*missing edges*). Some edges may be missed entirely, and smooth contours on the object will sometimes be represented in the edge image by multiple shorter edge chains with gaps between them. At the same time, there are often edges in the edge image that do not correspond to actual contours in the objects (*false edges*). This is frequently caused by lighting effects such as highlights or shadows but can arise when intensity differences between pixels in the stimulus image are caused by object texture, surface markings, and other factors. An additional set of complexities arises because of differences in edge type. Essentially all implemented edge extractors have been formulated for *step edges*—single-step changes in pixel intensity within the field measured by the extractor. It is well known that this leads to problems with

[1] There is evidence that photographs are superior to line drawings in strenuous conditions, such as when some objects are highly similar to each other in shape (Brodie, Wallace, & Sharrat, 1991; Price & Humphreys, 1989; see also Davidoff & Ostergaard, 1988). In addition, color information can be helpful when color photographs are compared with gray-scale photographs of equal contrast (Wurm, Legge, Isenberg, & Luebker, 1993). However, this evidence can also be interpreted to support the sufficiency of edges because the performance differences under these strenuous conditions have not been that large—usually, the differences are less than 40 ms in time and only a small amount in accuracy. Thus, the improvement attributable to surface information may be a relatively small, quantitative effect that is limited to strenuous conditions rather than a large, general, or qualitative effect.

the detection of vertices, where multiple edges (steps at different orientations) join. In fact, current research efforts on local, signal-based edge extractors are considering this problem (Rothwell, Mundy, Hoffman, & Nguyen, 1995). However, increasing an algorithm's sensitivity to vertices leads to problems in detecting *roof edges*. Roof edges often occur where two surfaces with the same natural coloring join and only lighting changes mark the edges. These competing concerns define a classic trade-off with no perfect solution in a local, signal-based framework; an edge that is valid in one region of the image may have the same contrast as a false edge in another part of the image.

Therefore, substantial further processing would be necessary to transform a stimulus image into a line drawing. In addition, object edges must be discriminated from background edges. (Current edge-extraction research typically avoids this problem by arranging objects against a "clean" background, but nature is not generally so helpful.) The process of transforming an edge image into a line drawing has been attempted only for the most simple types of scenes (e.g., Olivieri, Gatti, Straforini, & Torre, 1992; Otte & Nagel, 1992). Even in these cases, the processing has been computationally intensive, and it has been necessary to use domain-specific, geometric constraints (e.g., all objects are polyhedral, so valid edge chains can be fit with straight lines). No current computer vision effort even approaches the construction of line drawings for generic objects in natural settings.

Line Drawings Were Never Edges

In contrast to edge extraction, which is typically performed on local information without higher level influences, line drawings are created by humans who have already used high-level vision to perceive the object. The crucial difference is that humans have available to them global interpretations that result from high-level vision—interpretations in terms of meaningful visual structures such as figure and ground, shadows and highlights, regions and volumes, markings, and so on. These global interpetations can resolve the many local ambiguities that plague edge extraction. (The effects of knowing the object's identity, or of artistic training, are not crucial in our argument.) One often-used method for drawing is to abstract global shape primitives such as ovals or spheres from the stimulus and then to represent the primitives in the drawing by using configurations of lines and (often) shading (see, e.g., Fitzsimmons, 1989; Zaidenberg, 1939). "Drawing's basic ingredients are strokes or marks which have a symbolic relationship with experience, not a direct, overall similarity with anything real" (Rawson, 1969, p. 1).

Consequently, line drawings differ from local, signal-based edge images in a number of ways. Line drawings may contain edges that are absent from edge images (and the actual stimuli), such as edges obscured by shadows or low contrast. Line drawings are unlikely to include false edges that are present in edge images, such as brightness discontinuities produced by highlights, texture, shadows, or surface markings. In addition, line drawings may include emphasis on subjectively important edges, such as those signaling discontinuities in depth or surface orientation in the object, while deemphasizing or excluding less important, fine-scale stimulus details.

The differences between line drawings and edge images can be appreciated by examining the two versions of the telephone used in the Biederman and Ju (1988) studies; they are reproduced in Figure 1. Edges included in the line drawing but missing in the stimulus include the inside of the handle and many of the contours above the dial. False edges would probably be produced by a (local, signal-based) edge extractor in areas with highlights such as those below the dial and near the end of the handle. There is a strong emphasis in the line drawing on discontinuities in depth and surface orientation because these are the only contours included. There are also crucial abstractions in the line drawing. For example, the feet present in the photograph are absent in the line drawing; the feet would prevent accurate extraction of the phone's vertices by an edge extractor

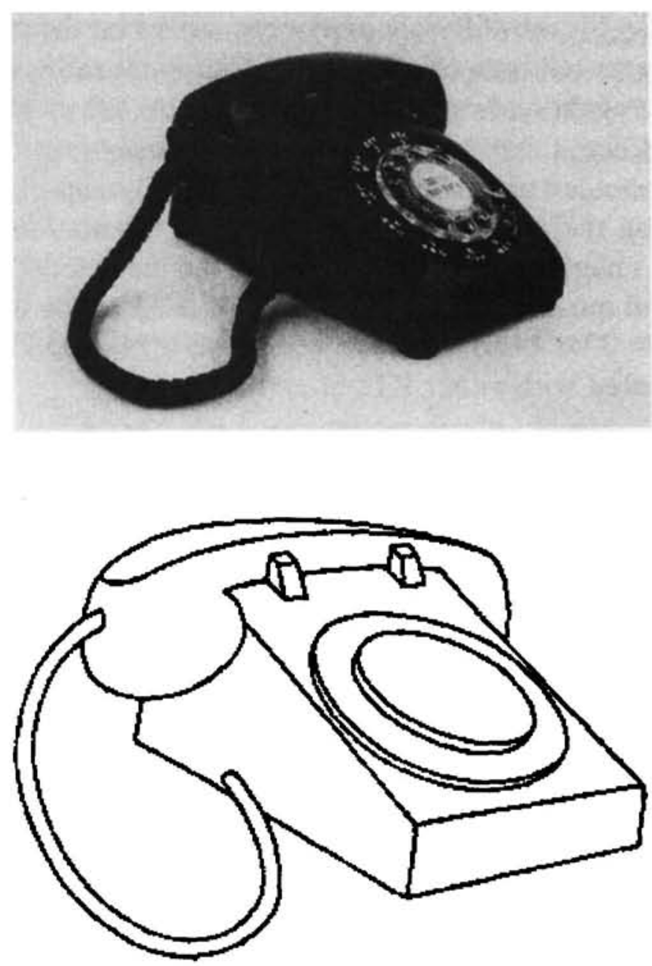

Figure 1. A photograph of a telephone and a line drawing of a telephone used by Biederman and Ju (1988). From "Surface Versus Edge-Based Determinations of Visual Recognition" by I. Biederman and G. Ju, 1988, *Cognitive Psychology, 20,* p. 42. Copyright 1988 by Academic Press. Reprinted with permission.

because their intensity is similar to that of the phone's body. A second important abstraction in the line drawing is the circular ring that represents the transparent finger-dial; researchers in edge extraction recognize transparency as being extremely difficult, if not impossible, to accurately process because transparent surfaces will produce false edges and are unlikely to produce valid edges.[2]

These points are not specific to the present example but could be made for other objects in the Biederman and Ju (1988) experiments, for the widely used Snodgrass and Vanderwart (1980) drawings, and for drawings in many other research articles on object recognition. The main point is that a line drawing is not a product of local, signal-based processing applied to a natural stimulus. A line drawing is a high-level rendering of an object in which noise is filtered out and only essential properties are abstracted and included.[3] Therefore, the ease of recognizing a line drawing is not a valid argument for the sufficiency of edge-based representations.

Object Recognition in Context

An additional limitation of experiments that compare line drawings and color photographs is that the objects are presented in isolation. This practice can be justified if low-level processes produce segmentation and edges independent of higher level influences. However, there is no known edge extractor that reliably finds all object edges or all background edges, much less finds all of each and distinguishes between them. The complexity of the segmentation problem in vision is becoming increasingly clear (e.g., Palmer & Rock, 1994a; Peterson, 1994), and there is evidence that segmentation is influenced by high-level recognition processes (e.g., Peterson & Gibson, 1994; see also Palmer & Rock, 1994b). It is possible that segmentation sometimes results from recognition, as often occurs in speech perception. The speech signal is often a continuous band of energy, with the perceived "spaces" between words being inserted by high-level perceptual processes only after the words are recognized.[4] It is likely that edge information is less sufficient when the fuller problem of recognizing objects in their natural backgrounds is considered.

Measuring the Sufficiency of an Edge-Extraction Process

In short, previous experiments with line drawings are not appropriate for examining the sufficiency of edge information in object recognition because they examined "edge images" that were created by high-level processing. A test of the sufficiency of edge information is valid only when the computation of the edges is properly controlled. In addition, the edges should be produced from objects in natural settings. In this section, we report an experiment that is consistent with these requirements.

Although computer edge-extraction routines do not produce line drawings, they remain the only method available for computing edge descriptions without high-level influences. As implied above, edge extraction has been a very large research area in computer vision for the past 20 years. Local, signal-based algorithms have been the focus, and the rationale for the approach and the underlying assumptions have been carefully thought out and debated (see, e.g., Boyer & Sarkar, 1992). Therefore, edge-extraction routines are appropriate for producing edge descriptions that are valid for our purposes.

In the present experiment, we measured the accuracy with which participants identified objects represented as local, signal-based edges and as color photographs. The objects appeared in natural contexts and in isolation. We used the Canny (1986) edge extractor, which has served as a standard of comparison in the field of computer vision, to produce edges. After reporting the results of the experiment, we discuss recent results indicating that the Canny extractor compares very well with a number of more recent and more sophisticated edge extractors.

Method

The method is summarized here; additional details are reported in Appendix A. The test stimuli were generated from 16 target objects photographed in natural settings. From each photograph, four stimulus images were produced, one for each combination of image type (full color vs. edge images) and background type (the objects were shown either in context or with the background erased). An example of one object in each of the four conditions is presented in Figure 2. The edge extractor detected luminance and color edges by applying the Canny (1986) algorithm modified to process color images (Lee & Cok, 1991). The Canny algorithm had three parameters—one for the spatial scale of the operator and one upper and one lower threshold. The parameters were adjusted manually for each image.

In the experiment, each image was presented for a full second, without a mask. The participants, 68 college students, were divided into four groups that saw different, counterbalanced sets of stimuli. Each target object was seen only once by a given participant. The participants were instructed to write down the name of the object in the center of the image. The 16 test trials were preceded by 4 practice trials with additional stimuli. Percentages of correct identifications were measured.

---

[2] Other examples of abstracting stimulus information include the letters and numbers on the dial of the photograph but not the line drawing and the transformation of the cord. Although the characters would produce edges because of their high contrast, it is extremely difficult to extract edges that would be recognizable to humans. All edge extractors would have a difficult time even finding the coils, much less abstracting them to a smooth solid.

[3] We do not dispute the claim that line drawings can be easy to recognize (but see Cavanagh, 1991). Line drawings are useful representations of objects, and in at least some cases they may be more similar to high-level internal representations of objects than color photographs are. Line drawings will continue to be a useful stimulus for researchers studying intermediate- and high-level vision, including ourselves. Also, we note that line drawings may be useful for examining the more restricted question of whether human object recognition involves color and surface information in addition to contours and boundaries.

[4] As an example, listen to a speaker of an unfamiliar language; distinct word boundaries are difficult to perceive.

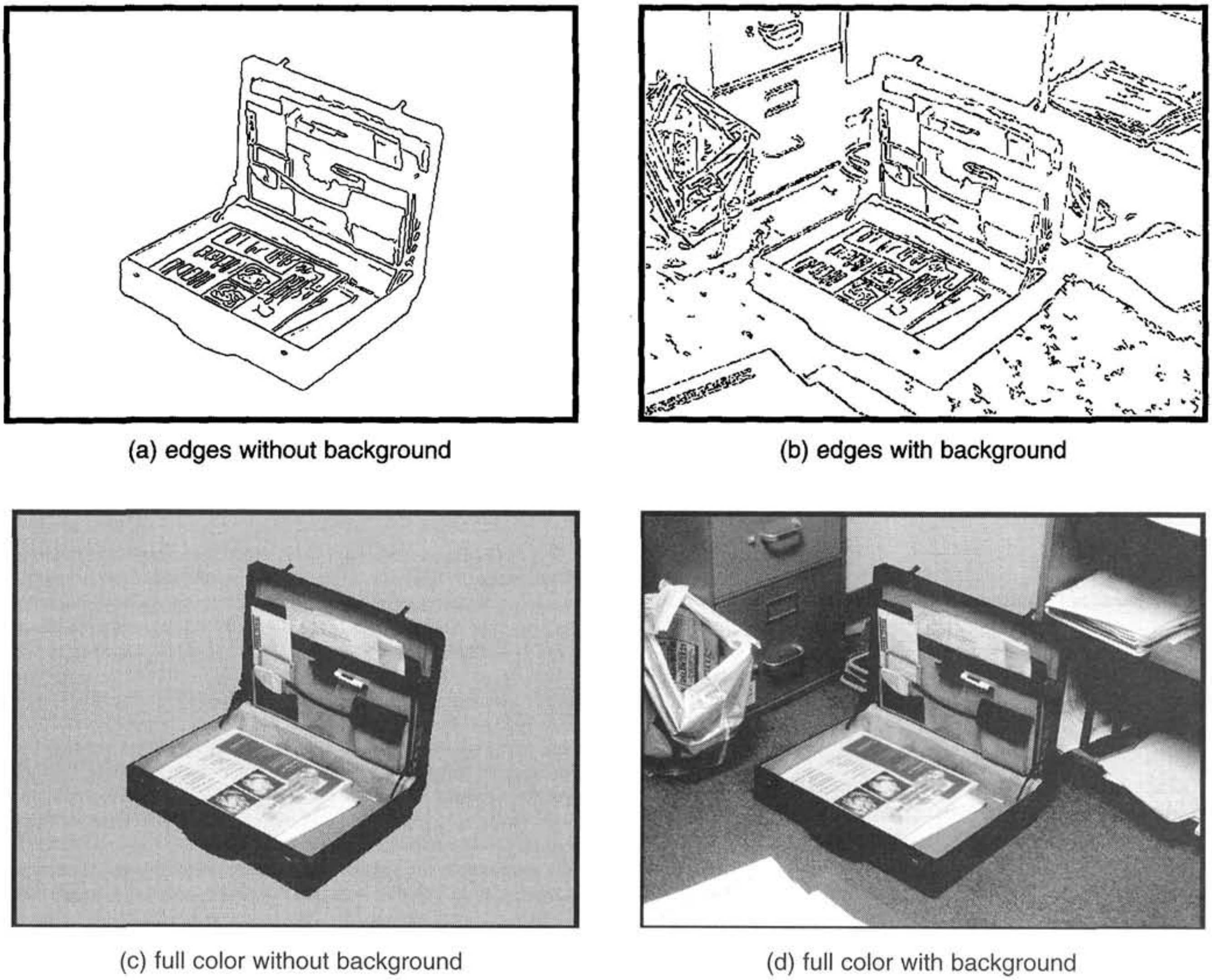

Figure 2. An example of the briefcase in each of the four image conditions. Although (c) and (d) appear in shades of gray, they were presented to participants as color slides.

Results and Discussion

The percentages of correct identification in each condition are shown (with standard errors) in Table 1. Overall performance averaged only 55.5% with edge images and 90.2% with color images. Performance was especially low for edge images of objects in context; for these images, the level of performance was 46% of the level obtained with color images. In the statistical analysis, there were main effects of image type and of background condition and an interaction.[5] These results indicate that the edge information was far from being sufficient for object identification.

Table 1
Percentage Correct and Standard Error in Each Condition

	Object in context		Object in isolation	
Condition	%	*SE*	%	*SE*
Full color	90.6	1.8	89.8	2.1
Edge	41.2	3.3	69.8	3.2

The extremely low performance with edge images of objects in context may be related to figure–ground segregation processes. When a background is included, the number of edges that must be considered in the segregation process is greatly increased. Because segregation requires considering different groupings of edges as figure and ground, the increased number of edges may greatly increase complexity. Segregation may benefit from color and surface information because it could constrain possible groupings. Also, if recognition processes overlap with segregation processes (Peterson & Gibson, 1994), the complexity of recognition may also be greatly increased in the background condition, especially with edges.

[5] For image type (color vs. edge), $F(1, 64) = 246.34$, $p < .001$; for context (background vs. isolation), $F(1, 64) = 58.10$, $p < .001$; and for their interaction, $F(1, 64) = 71.47$, $p < .001$.

The lack of any background effect in the color conditions could be due to a ceiling effect with color images. Background might have detrimental effects (or possibly facilitatory effects) if the difficulty of the identification task was increased.[6]

The lower performance with edge images was not restricted to a few objects. The data for individual objects are reported and discussed in Appendix B.

Color edges versus intensity edges. The edges used in the present experiment were defined in terms of intensity and color information in the red-green-blue color space. This is in contrast to the situation with most edge detectors, which are based solely on luminance information. An earlier experiment (first reported by Sanocki, Bowyer, Adair, & Sarkar, 1995) was conducted with the same method and the same color images but with edge images defined by the luminance-based Canny (1986) detector. The results from that study are shown in Table 2. The pattern of results is the same as in the present experiment and reinforces the conclusion that local edge information is insufficient for object recognition. The data for individual objects were also similar between experiments, as reported in Appendix B. The similarity of results between the experiments suggests that luminance-based edges carry most of the information used by humans in object recognition; the additional color-based edges used in the present experiment did not contribute to performance.

Comparing Canny's (1986) edge extractor to other types of edge extractors. It is possible that the low levels of performance obtained with the edge images were due to our choice of the Canny (1986) edge-extraction routine. Many other edge extractors have been proposed recently (see, e.g., Boyer & Sarkar, 1992; Heath et al., in press-a), including extractors with more sophisticated computational mechanisms. In two recent studies, we compared the Canny extractor to a total of six other extractors using subjective ratings of the goodness of the edge images they produced. We found the ratings to be highly reliable. More important, we found that when its parameters were adapted for each image (as done here), the Canny extractor performed as well as any of the extractors tested, including more sophisticated detectors. Heath et al. (in press-a) found that the Canny extractor produced better images than three other extractors: the Nalwa and Binford (1986) surface-fitting approach, the Sarkar and Boyer (1991) descendent of the Marr–Hildreth zero-crossing approach, and the traditional Sobel detector. More recently, Heath et al. (in press-a) found the Canny edge extractor to produce edges as well as or better than several newer edge extractors: the Bergholm (1987) detector, which focuses large-scale edges down to finer scales; the Iverson and Zucker (1995) detector, which uses logical constraints after Canny-like linear filtering in order to reduce the likelihood of false edges; and the Rothwell et al. (1995) extractor, which uses spatially adaptive processing after Canny-like filtering to improve performance in areas of junctions. In summary, it appears that the present results are representative of the entire class of local, signal-based edge extractors.

Table 2
Percentage Correct and Standard Error in Each Condition of Sanocki, Bowyer, Adair, and Sarkar (1995)

	Object in context		Object in isolation	
Condition	%	*SE*	%	*SE*
Full color	90.4	3.0	90.8	2.9
Edge	45.7	4.4	69.0	4.1

Alternative Approaches to Stimulus Interpretation

Given the insufficiency of local, signal-based edge extractors, it is important to consider alternative approaches to the problem of interpreting stimulus information.[7,8] Of particular interest is the use of constraints based on the geometry of volumetric objects. Such constraints have been incorporated into recent intermediate-level models of contour formation (e.g., Finkel & Sajda, 1994; Grossberg & Mingolla, 1985a, 1985b; Kellman & Shipley, 1991; for a review see Lesher, 1995). The constraints involve information about properties such as the continuity of edge chains, the strength of

[6] Congruent background information has been found to facilitate object identification in some studies (e.g., Biederman, 1981; Boyce, Pollatsek, & Rayner, 1989). However, when the effects of a meaningful background are assessed relative to those of a meaningless background control, there is a cost for processing background information that is combined with facilitative effects associated with the meaning of the context (Boyce et al. 1989). In the present background conditions, costs were likely to be present because there was information to process. However, facilitatory effects would be unlikely because we avoided including whole objects in the background. The (partial) background objects would have been difficult to identify, especially in the edge condition, so the context would have little or no meaning.

[7] One alternative approach is to assume that knowledge about object identity influences edge processing. Peterson and Gibson (1994) gathered considerable evidence indicating that object knowledge influences figure–ground segregation. This may be a promising alternative approach.

[8] Although Biederman has not emphasized the difficulties of interpreting stimulus information during object recognition (to our knowledge), he has been aware for some time of the insufficiency of local edges based on luminance discontinuities (I. Biederman, personal communication, November 29, 1995). For example, Hummel and Biederman (1992) specified that the edges used in object recognition were those defined by discontinuities in surface orientation and depth. However, they did not specify how such edges should be extracted and left the problem of stimulus interpretation unsolved. (In fact, this assumption seems inconsistent with the idea of recognizing objects from two-dimensional images, a primary characteristic of Biederman's [1987] original model, because depth information and reconstruction of surface orientation would seem to be required to detect the discontinuities specified.)

We also note that Biederman (1987) suggested in his Footnote 1 that there may be feedback from the process of detecting nonaccidental properties to the process of edge extraction, although he did not specify how this might work. This later suggestion is consistent with the idea of using geometric constraints in edge processing, which we discuss in this section.

neighboring edges, whether or not edges combine to form vertices, and whether or not vertices combine to form regions or volumes. In models of contour formation, the constraints are typically applied in a constraint-propagation process that determines a global "best fit" and requires multiple cycles to complete.

The use of such constraints might be sufficient for solving the problems of interpreting stimulus information. However, evaluating approaches such as this one with our methods will have to await implementations that can be applied to complex images. In addition, several important issues are raised by the incorporation of such constraints.

First, note that in intermediate-level approaches to stimulus interpretation, the goal of the computations is not to identify edges per se, but to find combinations of edges that form regions, volumes, or some other intermediate-level structure. Therefore, individual edges are not important in the representations that result. Furthermore, other basic types of input information may be useful, such as that about regions or surfaces. Surface information can provide complementary constraints on contour formation (e.g., Kellman & Shipley, 1991) and object recognition (Hummel & Stankiewicz, 1995). Thus, the problem of stimulus interpretation becomes one not of extracting edges but of combining input information of various types with constraints based on geometric properties of regions and volumes to arrive at region-based or volumetric interpretations. The concept of edges per se is relatively unimportant.

A second issue is raised by the fact that constraints based on the geometry of volumes are domain specific. Such constraints can be helpful in the interpretation of plain volumetric objects but may interfere with the recognition of other objects, such as decorated objects or wire furniture and ferns. As noted, domain-specific constraints have been eschewed in computer edge-extraction research because they limit generality. Therefore, if constraints are used, limitations on generality may have to be dealt with.

A further issue raised by the incorporation of domain-specific constraints is that distinctions between lower level and higher level processing begin to disappear. In particular, geometric constraints would be used in both the stimulus interpretation process and the object recognition process. Theorists might try to combine the two processes or perhaps develop a principled distinction between the constraints used in early and late processing.

In conclusion, the use of intermediate-level knowledge is a promising and perhaps necessary development. However, it also appears that such a development may require substantial revision of existing models of object recognition.

Conclusions

The general goal of this article has been to bring to light the significant difficulties involved in the interpretation of stimulus information during object recognition. Previous theoretical assumptions and experimental practices emphasizing edges have obscured the importance and difficulty of this problem. Our results with local, signal-based edge images indicate that edge information is far from being sufficient for object recognition. The results call into question psychological and computer vision models that use local edge extractors as their only low-level process. The results should spur further development of models of object recognition, which in general have either ignored or sidestepped the central problems of interpreting stimulus information.

References

Bergevin, A., & Levine, M. D. (1993). Generic object recognition: Building and matching coarse descriptions from line drawings. *IEEE Transactions on Pattern Recognition and Machine Intelligence, 15,* 19–36.

Bergholm, F. (1987). Edge focusing. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 9,* 726–741.

Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 213–233). Hillsdale, NJ: Erlbaum.

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

Biederman, I., & Ju, G. (1988). Surface versus edge-based determinations of visual recognition. *Cognitive Psychology, 20,* 38–64.

Boyce, S. J., Pollatsek, A., & Rayner, K. (1989). Effect of background information on object identification. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 556–566.

Boyer, K. L., & Sarkar, S. (1992). Assessing the state of the art in edge detection: 1992. *Proceedings of the Society of Photo-Optical Instrumentation Engineers,* 353–362.

Bradley, J. (1994). *XV.* Sharware program available from ftp://ftp.cis.upenn.edu/pub/xv/.

Brodie, E. E., Wallace, A. M., & Sharrat, B. (1991). Effects of surface characteristics and style of production on naming and verification of pictorial stimuli. *American Journal of Psychology, 104,* 517–545.

Brooks, R. A. (1981). Symbolic reasoning among 3-D models and 2-D images. *Artificial Intelligence, 17,* 285–348.

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 8*(6), 679–698.

Cavanagh, P. (1991). What's up in top-down processing? In A. Gorea (Ed.), *Representations in vision: Trends and tacit assumptions in vision research* (pp. 295–304). Cambridge, England: Cambridge University Press.

Davidoff, J. B., & Ostergaard, A. L. (1988). The role of colour in categorical judgments. *Quarterly Journal of Experimental Psychology, 40A,* 533–544.

Finkel, L. H., & Sajda, P. (1994). Constructing visual perception. *American Scientist, 82,* 224–237.

Fitzsimmons, C. (1989). *Serious drawing.* Englewood Cliffs, NJ: Prentice Hall.

Grimson, W. E. L. (1989). On the recognition of parameterized 2-D objects. *International Journal of Computer Vision, 2,* 353–372.

Grossberg, S., & Mingolla, E. (1985a). Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review, 92,* 172–211.

Grossberg, S., & Mingolla, E. (1985b). Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations. *Perception & Psychophysics, 38,* 141–171.

Heath, M., Sarkar, S., Sanocki, T., & Bowyer, K. (in press-a). Comparison of edge detectors: A methodology and initial study. *Computer Vision and Image Understanding.*

Heath, M., Sarkar, S., Sanocki, T., & Bowyer, K. (in press-b). A robust method for assessing the relative performance of edge detection algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence.*

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review, 99,* 480–517.

Hummel, J. E., & Stankiewicz, B. J. (1995, November). *Structural descriptions and coordinates: A hybrid theory of object recognition.* Paper presented at the annual meeting of the Psychonomic Society, Los Angeles, CA.

Huttenlocher, D. P., & Ullman, S. (1990). Recognizing solid objects by alignment with an image. *International Journal of Computer Vision, 5,* 195–212.

Iverson, L. A., & Zucker, S. W. (1995). Logical/linear operators for image curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 17,* 982–996.

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

Lee, H., & Cok, D. R. (1991). Detecting boundaries in a vector field. *IEEE Transactions on Signal Processing, 39,* 1181–1194.

Lesher, G. W. (1995). Illusory contours: Toward a neurally based perceptual theory. *Psychonomic Bulletin & Review, 2,* 279–321.

Lowe, D. G. (1987). Three dimensional object recognition from single two dimensional images. *Artificial Intelligence, 31,* 355–395.

Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of Royal Society of London, 207B,* 187–217.

Nalwa, V. S., & Binford, T. O. (1986). On detecting edges. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 8,* 679–714.

Olivieri, P., Gatti, M., Straforini, M., & Torre, V. (1992). A method for the 3D reconstruction of indoor scenes from monocular images. In G. Sandini (Ed.), *Computer vision—European Conference on Computer Vision '92* (pp. 697–700). Berlin: Springer-Verlag.

Otte, M., & Nagel, H. H. (1992). Extraction of line drawings from gray value images by non-local analysis of edge elementstructures. *Computer vision—European Conference on Computer Vision '92* (pp. 687–695). Berlin: Springer-Verlag.

Palmer, S., & Rock, I. (1994a). Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin and Review, 1,* 29–55.

Palmer, S., & Rock, I. (1994b). On the nature and order of organizational processing: A reply to Peterson. *Psychonomic Bulletin and Review, 1,* 515–519.

Peterson, M. A. (1994). The proper placement of uniform connectedness. *Psychonomic Bulletin and Review, 1,* 509–514.

Peterson, M. A., & Gibson, B. S. (1994). Must shape recognition follow figure-ground organization? An assumption in peril. *Psychological Science. 5,* 253–259.

Price, C. J., & Humphreys, G. W. (1989). The effects of surface detail on object categorization and naming. *Quarterly Journal of Experimental Psychology, 41A,* 797–828.

Rawson, P. (1969). *Drawing.* London: Oxford University Press.

Rothwell, C. A., Mundy, J. L., Hoffman, W., & Nguyen, V. D. (1995). Driving vision by topology. In *Proceedings of the International Symposium on Computer Vision* (pp. 395–400). Coral Gables, FL: IEEE Computer Society Press.

Sanocki, T., Bowyer, K. W., Adair, J., & Sarkar, S. (1995). Are real edges sufficient for object recognition? *Investigative Opthalmology & Visual Science* [Abstract], *36,* 3890.

Sarkar, S., & Boyer, K. L. (1991). Optimal infinite impulse response zero crossing based edge detectors. *Computer Vision, Graphics, and Image Processing: Image Understanding, 54,* 224–243.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 174–215.

Stark, L., Eggert, D., & Bowyer, K. W. (1988). Aspect graphs and non-linear optimization in 3-D object recognition. In *Second International Conference on Computer Vision* (pp. 501–507). Coral Gables, FL: IEEE Computer Society Press.

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

Zaidenberg, A. (1939). *Anyone can draw.* Cleveland, OH: World Publishing.

Appendix A

Details of the Method

Stimuli

The objects were photographed in their natural settings with a high-quality 35-mm camera with a 50-mm lens (this focal length is similar to that of human observers). We chose objects by walking around our home and office environments and looking for typical, easy-to-name target objects in a natural context that was somewhat cluttered. We arranged the settings somewhat but never imported objects from other settings, never created a new setting, and never moved the target object more than a meter. Existing light sources were used (and not moved). The goal was to have an unoccluded target object centered in the image, amidst portions of other objects. The target object was usually the only complete object in the photograph, occupied most of the center of the image, and was photographed at an intuitively typical orientation.

The photographs were developed and transferred to a Kodak CD-ROM. Sixteen photographs were chosen because they had good image quality and conformed to the requirement of having a single easy-to-identify object in the center. The images were transferred to a Sun workstation, where the *XV* program (Bradley, 1994) was used to crop the images somewhat and, in some cases, increase brightness and enhance contrast. This resulted in a set of 16 base images.

From each base image four slides were created that corresponded to the four conditions (see Figure 2). The full-color/full-background slides were created by transferring the base images (a gray-scale example is presented as d in Figure 2) to slide film. To create the full-color/no-background slides, the backgrounds were erased by hand and replaced with a light gray (e.g., c in Figure 2); then the images were transferred to slide film. For edge slides, edges were extracted with a Canny edge detector (Canny, 1986) that was modified to process color images using the vector gradient magnitude method (Lee & Cok, 1991). The full-background edge slides were the results of running the detector on the base images (e.g., b in Figure 2), and the no-background edge slides were the result of running the detector on the no-background images (a in Figure 2). Again, the images were transferred to slide film. For each condition, the film was developed into slides. The edges appeared black on white, as in a line drawing. One color demonstration slide and four practice edge-image slides were also created and used.

The Canny (1986) program had three parameters that were adjusted to be optimal for each image. These were (a) the size of the operator, (b) a lower edge-strength threshold, and (c) an upper edge-strength threshold. A total of 12 combinations of parameter values were considered for each image, and the experimenter selected the one that seemed to best represent the object—that is, to include contours important to the object while having a minimal number of noise edges.

Procedure and Design

The participants were 68 college students who participated in a class exercise and received course credit. There were four groups with between 15 and 18 participants each. Each group saw each of the 16 objects once (four slides from each condition). No object was seen more than once by a participant. The participants were tested in a dimly lit classroom and were seated in an arc approximately 6.4 m from a screen. The slides were presented by a projector with a timer and had maximum visual angles of 14° in width.

A trial began with a blank (darkened) screen; then the experimenter said "ready" and initiated the stimulus presentation. After 1s, the timing mechanism automatically advanced the projector to the next position, which produced a dark screen. The participants wrote down the name of the object they thought they perceived in the center of the screen. Participants were told that there would be a single object in the center of the screen and that they should write down their response if they had an idea of what the object was. However, participants were instructed to leave the response blank if they had no idea of what the object was. Responses were scored correct if they corresponded to the object names generated by the experimenters or if they were synonyms of those names. Borderline reponses were evaluated by two judges who were naive to the stimulus condition.

Appendix B

Results for Individual Objects and Correlations With Luminance Edge Extractor

The results for each object in each condition are shown in Figure B1; the objects are ordered from left to right in order of increasing overall performance. There are several significant aspects of these data. First, note that performance in the full-color conditions was at or near the ceiling for most of the objects. The group means are less than 100% because two objects were difficult to recognize.[B1]

Second, in the edge conditions, the lower levels of performance were not restricted to a few objects. As can be seen for the edge-background condition (second panel from top in Figure B1), performance drops fairly smoothly from 100% to zero. Similary, in the edge-isolation condition (bottom panel of Figure B1) there is a smooth drop from 100% to 44%, and there is one item at the zero level. The statistical effects reported earlier in the participants' analysis were also highly reliable in an analysis that treated items as a random variable.[B2]

Third, some edge images were easy to recognize. Our examination of the stimuli suggests that the easy-to-recognize images were those in which most of the edges corresponded to actual object edges and vertices—that is, those with few false edges and many veridical edges. These were also the least complex scenes.

The present data were highly correlated with the results for individual objects in the previous study with the luminance edge extractor (Sanocki, Bowyer, Adair, & Sarkar, 1995). The correlations were .81 in the background condition and in the isolation condition. These compare with correlations of .85 and .94 for color images in those two conditions, respectively. The images in the later two conditions were identical between experiments, although the correlations may have been reduced because of the smaller range of performance. The high correlations between the results of the two experiments provide additional evidence that the recognition process worked mainly on luminance-based edges; the addition of color edges in the present experiment did not increase performance.

[B1] A CD case was difficult in both color conditions, and the sprinkler was difficult in isolation. Inspection of individual responses indicated that the CD case was confused with many other objects having the same flat-rectangle shape (e.g., magazine, box of crayons). The sprinkler had a rather unusual shape and was confused with a variety of objects (e.g., motorboat motor, statue) when presented in isolation.

[B2] For image type (color vs. edge), $F(1, 15) = 54.57, p < .001$; for context (background vs. isolation), $F(1, 15) = 14.64, p < .01$; for the interaction, $F(1, 15) = 21.53, p < .001$.

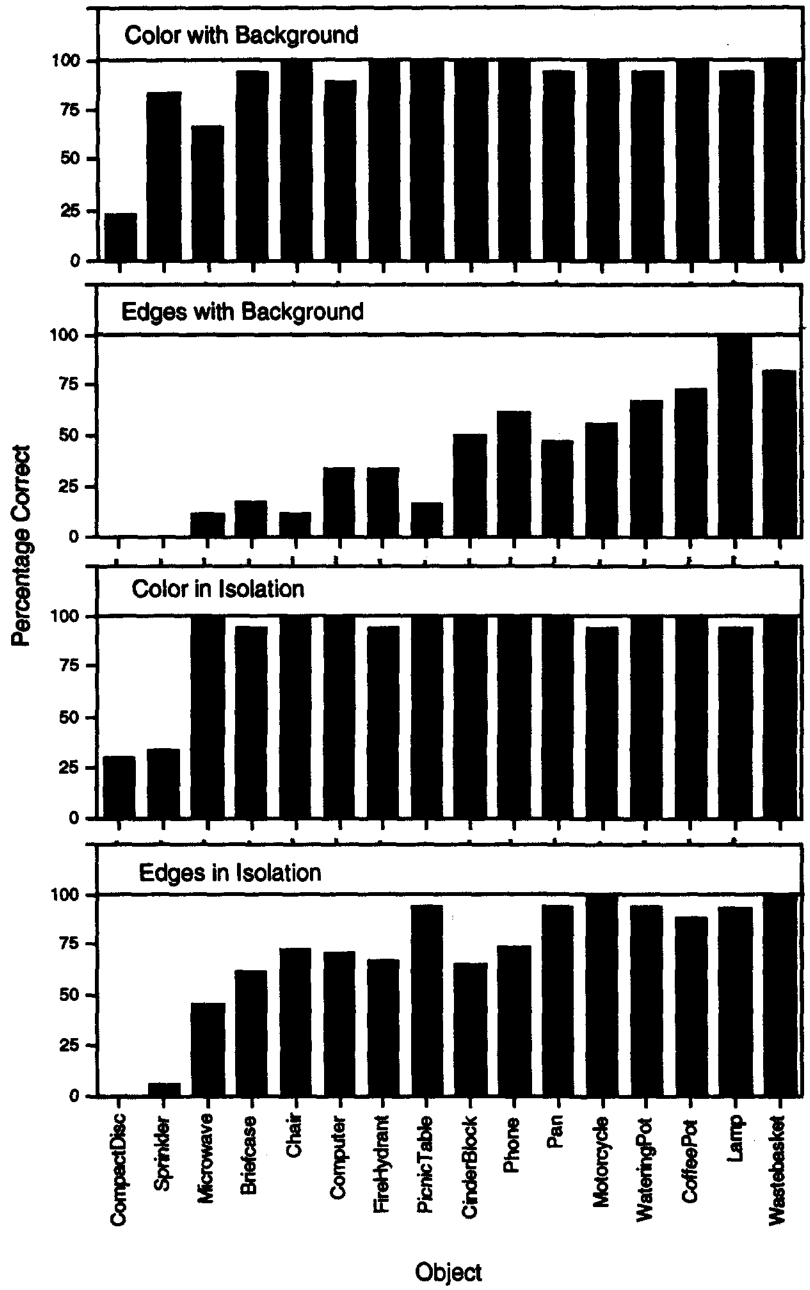

Figure B1. Percentage correct for each object in each of the four conditions.

Received July 27, 1995
Revision received September 16, 1996
Accepted December 5, 1996 ■