MATCHING IMAGE EDGES TO OBJECT MEMORY

Irving Biederman

MATCHING IMAGE EDGES TO OBJECT MEMORY

IRVING BIEDERMAN

STATE UNIVERSITY OF NEW YORK AT BUFFALO

ABSTRACT

A recent theory of human image understanding, Recognition-by-Components (RBC) represents an object as an arrangement of simple generalized-cone volumes. The fundamental assumption of RBC is that a particular set of these convex components, called geons [N ≤ 36], can be derived from invariant properties of edges in a 2-D image. If an arrangement of three geons can be recovered from the input, objects can be quickly recognized even when they are occluded, rotated in depth, novel, extensively degraded, or embedded in a scene.

INTRODUCTION

Humans can typically recognize an object even when it is viewed from a novel orientation, or it is a novel exemplar, or its image is extensively degraded. Moreover, most often only a single, brief fixation is all that is required to achieve quick and automatic understanding. The fundamental problem addressed by Recognition-by-Components (RBC) theory is how this is accomplished. Because a line drawing of an object can be classified as rapidly and as accurately as a full colored, textured photograph of the object, the problem can be stated as one of determining how the edges extracted from an image of an object can activate—in real time—an appropriate representation of that object in memory.

RBC assumes that an image of an object is segmented at regions of deep concavity into an arrangement of simple convex generalized cone primitives, such as cylinders, bricks, wedges, and cones, as illustrated in Fig. 1. The central assumption of the theory is that the members of a particular set (N ≤ 36) of primitives, called geons (for geometrical ions), are distinguishable on the basis of dichotomous or trichotomous contrastive properties of image edges, such as curved vs straight, parallel vs nonparallel, symmetrical vs nonsymmetrical, and orientation of edges (for defining vertices). These image properties can be determined from a general viewpoint and are highly resistant to degradation. Consequently, the geons, which are derived from these edge contrasts, themselves will be determinable under degradation and variations in viewpoint. An analysis of the representational capacity of the geons and their relations leads to the expectation that the basic level classification of most single visual entities can be achieved from an arrangement of only two or three geons.

STAGES OF PROCESSING

Figure 2 presents a schematic of the subprocesses posited by RBC. The stages are assumed to be arranged in cascade whereby partial activation (processing) at one level is sufficient to initiate activation at the next.

An early edge extraction stage, responsive to differences in surface characteristics, viz., sharp changes in luminance or texture, provides an edge-based description of the object. Edge extraction is assumed to be accomplished by a module that can proceed independently of the later stages, save for top-down influences (draw) from bottom to top in Fig. 2) that are likely to be in evidence under conditions where edge extraction is difficult or ambiguous. The only top-down route shown in Figure 2 is an effect of the nonaccidental properties (described below) on edge extraction, however it is also likely that other top-down routes, particularly from the components to edge extraction may also exist. It is possible that top-down effects of object expectancy, familiarity or scene constraints will be observed at a number of the stages, e.g., at segmentation, component definition, or matching, especially if edges are degraded. These have been
omitted from figure 2 in the interests of simplicity and because their actual paths of influence are as yet undetermined.

![Diagram of object recognition process]

**Figure 2.** RBC’s processing stages for object recognition.

Following the determination of the components, a structural description specifying the components and their relations is then matched against a like representation in memory. It is assumed that the matching of the components occurs in parallel, with no loss in capacity when matching objects with a large number of components. Partial matches are possible with the degree of match assumed to be proportional to the overlap in the componental descriptions of a representation of the image and the memorial representation.

**Parsing.** RBC assumes that a representation of the image is segmented—or parsed—into separate regions at matched deep concavities, particularly where there are discontinuities in minima of negative curvature (cusps). In general, matched concavities will arise whenever convex volumes are arbitrarily joined, a principle termed transversality. Such segmentation conforms well with human intuitions about the boundaries of object parts and does not depend on familiarity with the object. A secondary parsing principle is based on a change in nonaccidental properties of the edges, e.g., from parallel to nonparallel, as with the nosecone of a rocket.

**GEONS: PRIMITIVE VOLUMES FOR OBJECT RECOGNITION**

**Registration of Nonaccidental Properties (NAPs).** Certain properties of edges in a 2-D image are taken by the visual system as strong evidence that the edges in the 3-D world contain those same properties. Five of these properties are illustrated in Figure 3 (modified from Lowe). They are termed nonaccidental because they are unlikely to be a consequence of an accidental viewpoint of edge and eye. For example, if there is a straight edge in the image (collinearity), the visual system infers that the edge producing that line in the 3-D world is also straight. The visual system ignores the possibility that the property in the image is merely a result of a (highly unlikely) accidental alignment of eye and a curved edge. Smoothly curved elements in the image (curvilinearity) are similarly inferred to arise from smoothly curved features in the 3-D world. Similar strong inferences are made about symmetry, parallelism and coterminal of edges. These NAPs are assumed to be registered simultaneously with a parsing of the image. The nonaccidental properties of the parsed regions provide critical constraints on the identity of the components.

**Principle of Non-Accidentals:** Critical information is unlikely to be a consequence of an accidental viewpoint.

**Three Space Inference from Image Features**

<table>
<thead>
<tr>
<th>2-D Relation</th>
<th>3-D Inference</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Collinearity of points or lines</td>
<td>Collinearity in 3-Space</td>
<td><img src="image1" alt="Example 1" /></td>
</tr>
<tr>
<td>2. Curvilinearity of points of arcs</td>
<td>Curvilinearity in 3-Space</td>
<td><img src="image2" alt="Example 2" /></td>
</tr>
<tr>
<td>3. Symmetry (Skew Symmetry?)</td>
<td>Symmetry in 3-Space</td>
<td><img src="image3" alt="Example 3" /></td>
</tr>
<tr>
<td>4. Parallel Curves (Over Small Visual Angles)</td>
<td>Curves are parallel in 3-Space</td>
<td><img src="image4" alt="Example 4" /></td>
</tr>
<tr>
<td>5. Vertices—two or more terminations of a common point</td>
<td>Curves terminate at a common point in 3-Space</td>
<td><img src="image5" alt="Example 5" /></td>
</tr>
</tbody>
</table>

**Figure 3.** Five nonaccidental relations (Adapted from Lowe).

**Geons from NAPs.** Each segmented region is then approximated by one of a possible set of generalized cones. A generalized cone is the volume swept out by
a cross section moving along an axis. Marr [4] showed that the contours generated by any smooth surface could be modeled by a generalized cone with a convex cross section.

The fundamental perceptual assumption of RBC is that members of this set of generalized cones, called geons, can be differentiated on the basis of dichotomous or trichotomous contrasts of NAPs. The particular set of nonaccidental properties shown in Figure 3 was emphasized because they may constitute a basis for the generation of this set of perceptually plausible components. Any primitive that is hypothesized to mediate object recognition should be rapidly identifiable and invariant over viewpoint and noise. These characteristics would be attainable if differences among components were based on contrasts in nonaccidental properties. In addition, the five relations shown in Fig. 3 reflect intuitions about significant perceptual and cognitive differences among objects.

Figure 4 illustrates the generation of the set of 36 geons from contrasts in the nonaccidental relations of 

![Diagram](https://via.placeholder.com/150)

Figure 4. An illustration of how variations in three attributes of a cross section (curved vs straight edges; constant vs expanded vs. expanded & contracted size; mirror & rotational symmetry vs. mirror symmetry vs. asymmetrical) and one of the shape of the axis (straight vs curved) can generate a set of generalized cones differing in nonaccidental relations. Constant-sized cross sections have parallel sides; expanded or expanded & contracted cross sections have sides that are not parallel. Curved vs Straight cross sections and axes are detectable through Collinearity or Curvature. The three values of cross-section Symmetry (Symmetrical under Reflection & 90° Rotation; Reflection only; or Asymmetrical) are detectable through the symmetry relation. Shown here are the neighbors of a cylinder. The full family of geons has 36 members.

four attributes of generalized cylinders. Three of the attributes specify characteristics of the cross section (its curvature, size variation, and symmetry) and one of the axis (its curvature).

When the contrasts in the generating functions are translated into image features it is apparent that the geons have a larger set of distinctive nonaccidental image features than the four that might be expected from a direct mapping of the contrasts in the generating function. Figure 5 shows some of the nonaccidental contrasts distinguishing a brick from a cylinder. The silhouette of a brick contains a series of six vertices, which alternate between Ls and Arrows, and an internal Y vertex, as illustrated in figure 5. The vertices of the silhouette of the cylinder, by contrast, alternate between a pair of Ls and a pair of tangent Ys. The internal Y vertex is not present in the cylinder (or any geon with a curved cross section). These differences in image features would be available from a general viewpoint and thus could provide, along with other contrasting image features, a basis for discriminating a brick from a cylinder.

![Diagram](https://via.placeholder.com/150)

Figure 5. Some nonaccidental differences between a brick and a cylinder.

GEON RELATIONS

Although the components themselves are the focus of this paper, as noted previously (see Fig. 2) the arrangement of primitives must be specified for the representation of a particular object. The representation of an object would thus be a structural description [8] that expressed the relations among the components. A suggested (minimal) set of relations among any pair of geons is described in Biederman [1]. The relations include:

a) Verticality: whether Geon A is above, below, or to the side of Geon B (defined for approx. 80% of objects);
b) Relative Size: whether Geon A is much larger than, much smaller than, or approximately equal in size to Geon B,
c) Centering: whether the point of attachment is centered or off-centered on a geon's surface, and

d) Surface Size at Join: whether a geon is joined at a large or small surface (defined for each geon separately).
This conservative set yields 57.6 possible combinations of relations that can hold for a pair of geons. Like the components themselves, the relations are nonaccidental in that they can be determined from a general viewpoint, are preserved in the 2-D image, and are categorical, requiring the discrimination of only two or three levels.

**REPRESENTATIONAL CAPACITY OF RBC: THREE GEONS SUFFICE**

People know about 30,000 readily distinguishable object models\(^1\). Thirty thousand objects would require that the six-year-old, who not only demonstrates a competence for visual recognition that is comparable to that of an adult but can also name most of the objects in his or her environment, learn an average of 13.5 objects per day from birth or about one per waking hour. There are 74,649 possible two-geon objects (36\(^2\) possible geon combinations X 57.6 possible relations) and 154 million possible three-geon objects\(^1\) (74,649 (number of two-geon objects) X 36 X 57.6 (number of combinations)). Our six-year-old learner would have to learn 4,395 objects every waking hour to this or her life to master all possible three-geon objects.

The extraordinary disparity between the representational power of two or three geons and the number of objects in an individual's object vocabulary means that there is an extremely high degree of redundancy in the filling of the 154 million cells in the component-relation space. Even if our estimate of the number of objects in an individual's visual vocabulary was too low by a factor of five so that 150,000 objects were actually known, we would still be using less than one tenth of one percent of the possible combinations of three components (i.e., 99.9% redundancy).

There is a remarkable consequence of this redundancy if we assume that objects are distributed homogeneously throughout the object space. A sparse, homogeneous occupation of the space means that, on average, it will be rare for an object to have a neighbor with the same or similar structural descriptions (viz., geons and relations). Because the space was generated by considering only the number of possible two- or three-geon objects, we have actually derived a constraint on the estimate of the likely number of geons per object that will be sufficient for unambiguous identification. If objects were distributed homogeneously among combinations of relations and components, then only two of three components, in their specified relations, would be sufficient to unambiguously represent most objects\(^1\).

**EMPIRICAL SUPPORT FOR RBC**

According to the RBC hypothesis, the preferred input for accessing object recognition is that of the volumetric components. In most cases, only three appropriately arranged geons would be all that is required to uniquely specify an object. Rapid object recognition should then be possible. Neither the full complement of an object's geons, nor its texture, nor its color, nor the full bounding contour (or envelope or outline) of the object need be present for rapid identification. The problem of recognizing tens of thousands of possible objects from an infinity of possible images becomes, in each case, just a simple task of identifying the arrangement of a few from a limited set of geons.

**Overview of Experiments**

Several object naming reaction time experiments have provided support for the general assumption of the RBC hypothesis, although none have provided tests of the specific set of components proposed by RBC. A general summary of many of the experiments can be found in Biederman\(^1\). In all experiments, subjects named or quickly verified briefly presented pictures of common objects.

**Surface vs. Edge-Based Descriptions.** That an edge-based description may provide a sufficient account of object recognition was supported by experiments showing that simple line drawings constructed to convey only the object's major geons, were identified as rapidly as full colored, detailed, textured slides of the same objects\(^9\). There was no evidence, moreover, that an object with a characteristic color, such as a fork, mushroom, or camera, would derive any additional advantage when shown as a color slide compared to an object without a diagnostic color, such as a chair, mittens, or pen.

This failure to find a color-diagnosticity effect, when combined with the finding that simple line drawings could be identified so rapidly as to approach the naming speed of fully detailed, textured, colored photographic slides, supports the premise that the earliest access to a mental representation of an object can be modeled as a matching of an edged based representation of a few simple components. Such edged-based descriptions are thus sufficient for primal access.

**Consequences of Three Geon Sufficiency**

**Partial Objects and the Effects of Complexity.** Complex objects, defined as those requiring six or more components to appear complete, as the airplane and the penguin in figure 6, could be identified perfectly from only two or three of their geons, as long as subjects were not stressed to respond quickly\(^9\). Under speed stress and with brief (100 msec) exposures, both naming reaction times (RTs) and errors increased with the removal of additional components from the complete versions. But even under these conditions, complex objects with less than half their components were accurately named on 75 percent of the trials. Importantly, for the complete versions of the objects,
complex objects were identified more rapidly than simple objects (those requiring only two or three components to appear complete, as the goblet and flashlight in figure 6).

![Image of objects](image)

Figure 6. Some partial objects. Note that with three components, all can be uniquely specified.

**Objects requiring surface descriptors.** Some objects, such as a broom or a zebra, require both a volumetric and a surface texture description. Such objects were identified less accurately at brief exposures and with longer naming times than controls that were closely matched in volumetric structure and bounding contour. For example, a shovel and a horse served as controls for the broom and zebra, respectively. If the texture region was functioning as another component, then performance should have been facilitated through the additional geon, in that complex objects can be identified more rapidly than simple ones. Alternatively, the detailed processing required to specify the texture field might not be completed in a brief exposure duration so such objects might prove to be less recognizable. The results supported the latter alternative and are consistent with the previously reported secondary status of surface features.

**The effect of an inappropriate geon.** A consequence of the three-geon rule is that the addition of a fourth but inappropriate geon (middle column of figure 7) should not result in reduced recognition speed. The three appropriate geons (left column, Fig. 7) will be sufficient to activate the object's representation and unless the inappropriate geon results in the activation of a competing object, no interference should occur, even though that same geon would facilitate the recognition speed of an object when it was appropriate. (This prediction assumes that there is no bottom-up inhibition from geons to objects.) This expectation has been recently confirmed.

![Image of geons](image)

Figure 7. Left. Three-geon versions of complex objects requiring six or nine geons to look complete (left). Right. Four-geon versions of the same objects. Middle. Versions where a fourth—and inappropriate—geon from another object has been added to the three-geon version. Recognition error rates and naming reaction times were equivalent for the 3 and 3+1 versions but error rates and RTs were lower for the 4 geon versions.

**An extension to scene perception.** The mystery about the perception of scenes is that the exposure duration required to have an accurate perception of an integrated real-world scene is not much longer than what is typically required to perceive individual objects. The recognition of a visual array as a scene requires not only
the identification of the various entities but also a semantic specification of the interactions among the object and an overall semantic specific of the arrangement.

However, the perception of a scene is not, in general, derived from an initial identification of the individual objects comprising that scene. That is, in general we do not first identify a stove, refrigerator, and coffee cup, in specified physical relations and then come to a conclusion that we are looking at a kitchen.

Some demonstrations and experiments suggest a possible basis for understanding rapid scene recognition. Mezzanotte showed that a readily interpretable scene could be constructed from arrangements of single geons that just preserved the overall aspect ratio of the object, such as those shown in Figure 8. In these kinds of scenes, none of the entities, when shown in isolation, could be identified as anything other than a simple volumetric body, e.g., a brick. Most important, Mezzanotte found that such settings were sufficient to cause interference effects on the identification speed of intact objects that were inappropriate to the setting.

I propose that often quick understanding of a scene is mediated by the perception of geon clusters. A geon cluster is an arrangement of geons from different objects that preserve the relative size and aspect ratio and relations of the largest visible geon of each object. In such cases, the individual geon will be insufficient to allow identification of the object. However, just as an arrangement of two or three geons almost always allows identification of an object, an arrangement of two or more geons from different objects may produce a recognizable combination. The cluster acts very much as a large object. Figure 9 shows two examples. If this account is true, fast scene perception should only be possible in scenes where such familiar object clusters are present. This account awaits empirical test.

**Contour Deletion.** That the contours specified by RBC may provide a necessary account of object recognition was supported by a demonstration that degradation (contour deletion), if applied at the regions that prevented recovery of the components, rendered an object unidentifiable. Figure 10 shows some sample stimuli from that experiment. Those in the rightmost column, from which contour was removed to prevent recovery of the components were, essentially, unrecognizable, even with as long as five seconds of viewing. An equal amount of contour was removed from the control stimuli, shown in the middle column, but in midsegment so as to allow filling in from collinearity or smooth curvature. The disruption from the contour deletion on these objects was considerably smaller and they could be identified at near perfect accuracy at long exposures.

Objects with midsegment deletion (those in the middle column of Fig. 10) did show a considerable loss in speed and accuracy of identification when compared to the intact objects shown in the left column. Additional research has established that the processes by which contour is recovered through collinearity or smooth curvature is surprisingly slow and dependent on

---

Figure 8. Two of Mezzanotte's scenes.

Figure 9. Possible geon clusters for figure 8.
an image actually being present (as opposed to being done in memory). The effect of the contour deletion is object centered, being directly related to the proportion of an object's contour that is deleted and independent of the actual size of the gaps in terms of degrees of visual angle. These characteristics of contour deletion allowed us to explore the arrangement of stages in the next investigation.

Perceiving degraded vs. partial objects. The model of RBC illustrated in figure 2 can be partitioned into two critical stages: a) those processes leading to

![Diagram of degraded objects]

Figure 10. Example of five stimulus objects in the experiment on the perception of degraded objects. The left column shows the original intact versions. The middle column shows the recoverable versions. These contours have been deleted in regions where they can be replaced through collinearity or smooth curvature. The right column shows the nonrecoverable versions. These contours have been deleted at regions of concavity so that collinearity or smooth curvature of the segments bridges the concavity. In addition, vertices have been altered, e.g., from Ys to Ls.

and including the determination of the geons, and b) those processes involved in matching an arrangement of geons to memory.

Consider figure 11 which shows, for some sample objects, one version in which whole components are deleted so that only three (of six or nine) of the components remain (as in Figs. 6 and 7) and another version in which the same amount of contour is removed, but in midsegment distributed over all of the object's components (similar to those shown in the middle column in Fig. 10). In objects missing components, the components cannot be added prior to recognition. Logically, one would have to know what object was being recognized to know what parts to add. With the midsegment deletion, components can be determined from processes employing collinearity or smooth curvature.

The two methods for removing contour may thus be affecting different stages. Deleting contour in midsegment affects processes prior to and including those involved in the determination of the components (Fig. 2). The removal of

![Diagram of midsegment deletion and component (geon) deletion]

Figure 11. Sample versions of objects with midsegment deletion and component (geon) deletion.

whole components (the partial object procedure) is assumed to affect the matching stage, reducing the number of common components between the image and
the representation and increasing the number of distinctive components in the representation.

The two stages can be regarded as being arranged in cascade, with an earlier geon determination stage relaying activation on the object matching stage. Figure 12 shows the expected activation functions from the two procedures for deleting contour. Deleting contour in midsegment results in an initial slow growth in activation as the relatively slow processes for smooth continuation are required to restore the deleted contours. Once the restoration is complete there is a rapid growth in activation at the object representation stage. By contrast, there is an initial rapid activation of the components from the partial objects which, however,

\[ a_n(t) = a_n(0) \left[ 1 - \exp \left( -\frac{t}{\tau} \right) \right] \]

\[ a_n(t) = a_n(0) \left[ 1 - \exp \left( -\frac{t}{\tau} \right) \right] \]

\[ a_n(t) = a_n(0) \left[ 1 - \exp \left( -\frac{t}{\tau} \right) \right] \]

Figure 12. Illustration of a cascade of an earlier geon activation stage and a subsequent object representation stage. The activation of the geons causes activation of the object representations. Component deletion results in a lower asymptote at the geon activation stage because missing components never get activated prior to object activation.

asympotes below the activation level of the midsegment deleted objects. The reason for this is that the missing components have activation levels of zero. Once the filling-in is completed for the objects with midsegment deletion, the complete complement of an object's components are available, providing a better match to the object's representation than is possible with a partial object that had only a few of its components. The net effect is to produce a crossover interaction over exposure duration which produces a similar effect on the next stage, activation of the representation of the object.

This prediction was supported from the results of an experiment\(^6\) which studied the naming speed and accuracy of six- and nine-component objects undergoing these two types of contour deletion. At brief exposure durations (e.g., 65 msec) performance with partial objects was better than objects with the same amount of contour removed in midsegment both for errors and RTs. At longer exposure durations (200 msec), the RTs reversed, with the midsegment deletion now faster than the partial objects.

CONCLUSIONS

1. By deriving the geons from simple contrasts in NAPs, RBC solves the classic problem of how a single 3D world can be accurately perceived from the infinity of possible worlds that could have projected any single image. We will only rarely be victimized by accidents and, moreover, those cases can be readily correctable by a slight variation in viewpoint. Also, RBC allows, for the first time, an understanding of how pattern recognition can be accomplished in real time. We need only recover two or three geons from an image to achieve recognition.

2. By stipulating that parsing be performed at cusps, the recovered primitives will be simple (=convex) and therefore robust under noise. With the parsing assumption, RBC provides an account of the heretofore unspecified role of the classic Gestalt organizational phenomena in pattern recognition, particularly those concerned with Pragnanz or Good Figure. Given that the geons are fitting simpleparsed parts of an object, the constraints toward regularization characterize not the complete object but the object's components.

3. The characterization of object perception provided by RBC bears a close resemblance to some current views as to how speech is perceived. In both cases, the ease by which we are able to code tens of thousands of words or objects is solved by mapping that input onto a modest number of primitives—55 phonemes or 36 geons—and then employing a representational system that can code and access free combinations of these primitives. In both cases, the specific set of primitives is derived from dichotomous (or trichotomous) contrasts of a small number (less than ten) of independent, invariant characteristics of the input. The ease by which we are able to perceive words or objects may thus derive less from a capacity for judging continuous physical variation than it does from a perceptual system designed to represent the free combination of a modest number of
REFERENCES


[5] Biederman, I., & Beiring, E. Perceiving objects with inappropriate parts. Image Interpretation Laboratory (IIL) TR, State University of Buffalo.


Doctoral Dissertation, Department of Psychology, State University of New York at Buffalo.


ACKNOWLEDGEMENTS

This research was supported by the Air Force Office of Scientific Research (Grant 86-0106). I would like to express my deep appreciation to Thomas W. Blickle, Ginny Ju, H. John Hilton, Elizabeth A. Beiring, and Deborah Gagnon for their invaluable contributions to the research described in this article.

Correspondence about this paper should be addressed to Irving Biederman, Department of Psychology, State University of New York at Buffalo, Park Hall, Amherst, New York 14260.