An Introduction to Object Recognition*

Jeffrey C. Liter and Heinrich H. Bülthoff

Max-Planck-Institut für biologische Kybernetik, Spemannstraße 38, D-72076 Tübingen, Germany

Z. Naturforsch. **53c,** 610–621 (1998); received April 16, 1998

Form Perception, Categorization, Psychophysics, Object Recognition, Shape Representation

In this report we present a general introduction to object recognition. We begin with brief discussions of the terminology used in the object recognition literature and the psychophysical tasks that are used to investigate object recognition. We then discuss models of shape representation. We dispense with the idea that shape representations are like the 3-D models used in computer aided design and explore instead models of shape representation that are based on feature descriptions. As these descriptions encode only the features that are visible from a particular viewpoint, they are generally viewpoint-specific. We discuss various means of achieving viewpoint-invariant recognition using such descriptions, including reliance on diagnostic features visible from a wide range of viewpoints, storage of multiple descriptions for each object, and the use of transformation mechanisms. Finally, we discuss how differences in viewpoint dependence that are often observed for within-category and between-category recognition tasks could be due to differences in the types of features that are naturally available to distinguish among different objects in these tasks.

Introduction

Humans possess a remarkable ability to determine what things are simply by looking at them. We do this very quickly and very accurately. Nevertheless, it has proven especially difficult to build computer systems with this ability, underscoring the complexity of this task. As in many other domains, humans and (today's) computers have complementary visual processing skills. For example, although there exist artificial vision systems capable of detecting flaws in complex manufactured products that would go completely unnoticed by most human observers, we have been unable to build artificial systems that possess any three year old's ability to distinguish her own toys from her friend's.

Perhaps one reason for this difference is that a machine can know in advance exactly what visual features to look for in order to identify a defective product. This is because we can simply build the machine to look for these features. In contrast, we know much less about what features are useful for identifying everyday objects. Furthermore, besides knowing what features are useful for detecting a defective product, we can be sure that these features will always be available in a controlled manufacturing environment. For example, we can arrange it so that the size of the image to be processed is always the same, and we can control viewing conditions such as viewing angle and lighting. This is certainly not the case in the natural environment. The size in which an object is imaged on the retina might never be the same in more than one instance, and conditions such as lighting are always changing. Perhaps most important, it is not true that the same features will always be available to recognize an object in the natural environment. The same object can appear in many different orientations with respect to the viewer, and the features that are visible in different views of the object will never be exactly the same.

As a simple example of some of the many difficulties encountered in visual processing, consider the office scene depicted in Fig. 1.

Although there are many objects in the scene, we have little difficulty determining what they are and how they can be used. Most would agree, for example, that all but one of the chairs depicted in the scene are of the same make, despite the fact

* This communication is a contribution to the workshop on "Natural Organisms, Artificial Organisms, and Their Brains" at the Zentrum für interdisziplinäre Forschung (ZiF) in Bielefeld (Germany) on March 8–12, 1998.

Reprint requests to Prof. H. H. Bülthoff.
Fax: (07071) 601-616.
E-mail: heinrich.buelthoff@tuebingen.mpg.de.

0939–5075/98/0700–0610 $ 06.00

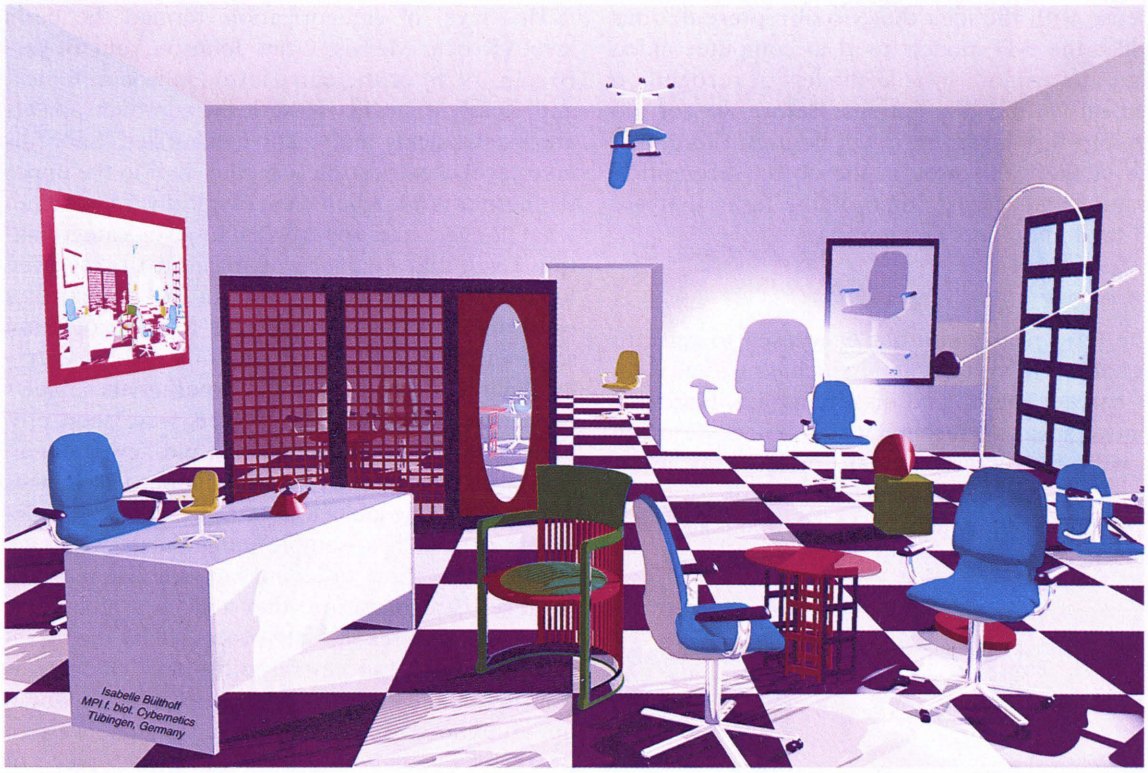

Fig. 1. An office scene illustrating some of the many difficulties encountered in visual processing.

that they appear in different orientations, under different illuminations, and in different sizes in the image. In particular, notice that the bounding contour of the chair in the lower right corner is identical to the bounding contour of the shadow on the back wall. Clearly no one would attempt to sit in the chair projected on the wall. Likewise, no one would attempt to sit in the chair atop the desk, though it's image size is identical to that of the chair seen through the door on the back wall. Another difficulty that is apparent upon viewing this scene is that objects must be segregated from the background before they can be recognized. Although we will only consider recognition of isolated objects in the present chapter, the reader should be aware that segregating figure from ground is not a trivial problem. We leave it as an exercise for the reader to identify some of the other problems one is likely to encounter in interpreting this scene.

Another reason why it has been difficult to make progress in understanding visual processing is that our subjective impressions tell us very little about how it is done. For example, although it seems that we can recognize objects equally well from any viewing angle, psychophysical evidence suggests otherwise. Palmer, Rosch, Chase (1981) studied the time required to name objects seen from different viewpoints and found that certain "canonical" views were named more quickly than other non-canonical views. This difference in naming time occurred despite the fact that subjects handled and visually inspected the objects prior to participating in the naming task. Blanz, Vetter, Bülthoff, and Tarr (1995) demonstrated that different observers agree to a large extent on what views of an object are canonical. Observers in their study selected canonical views by rotating computer-simulated, 3-D objects with a space ball.

Throughout this chapter we will present experiments relevant to the question of what visual features are used to perform various tasks, and we will discuss several models of how these features are organized into visual representations. We will

dispense with the idea that visual representations are like the 3-D models used in computer aided design and explore instead models of recognition based on feature descriptions. Before we get too deep into this, however, it will be useful to define some of the terms used in the object recognition literature and discuss some of the tasks that are used to study object recognition.

Terminology

The term *recognition* has been used to refer to many different visual abilities, including identification, categorization, and discrimination. Normally when we speak of recognizing an object we mean that we have successfully categorized it as an instance of a particular object class.[1] For example, upon viewing the objects in Fig. 2, one is likely to conclude first that they are chairs.

Fig. 2. Computer-simulated chairs used in research by Blanz *et al.* (1995).

[1]Notice that for face recognition, a widely studied area of object recognition, the term recognition refers not to the classification of the object as a face, but to a determination of whether the face is known or unknown. This underscores the fact that face recognition is an inherently subordinate-level classification task. Readers interested in learning more about face recognition research should see Bruce (1988).

This level of categorization, termed the basic level (Rosch, Mervis, Gray, Johnson, and Boyes-Braem, 1976) or the entry level (Jolicoeur, Gluck, and Kosslyn, 1984), is the level at which objects are most quickly and easily categorized. Subordinate-level classification (e.g., the chair in the upper left corner is a kitchen chair) typically takes somewhat longer, and superordinate-level categorization (e.g., that's a piece of furniture) takes even longer (Jolicoeur *et al.*, 1984). In the context of a psychophysical experiment, the term recognition sometimes means something other than entry-level categorization. In some experiments subjects must decide whether test objects were seen previously in the experiment (e.g., old-new recognition), or subjects must decide whether two images depict the same object (e.g., same-different judgments or match-to-sample judgments). Although conceptually these tasks are not equivalent to entry-level categorization, they tell us a great deal about the features that are processed by the visual system. Clearly, successful entry-level categorization is not the end of visual processing. It would not go unnoticed, for example, if someone were to break into your house and replace each piece of furniture in the den with a different piece having the same name. Likewise, we have no difficulty distinguishing the different objects in Fig. 2, although each would be immediately classified as a chair. Visual processing and visual memory go much deeper than entry-level categorization. These other tasks help us to understand visual processing at these deeper levels.

Tasks

Researchers use many different experimental tasks to investigate object recognition, some of which are summarized in Table I.

These tasks can be divided into explicit and implicit tasks. Explicit tasks require the subject to make comparisons among two or more objects presented during the experiment. In a same-different task, for example, subjects view images of two objects, either simultaneously or in sequence, and decide whether they depict the same object. In a match-to-sample task, the same target object is presented for recognition more than once among different distractor objects. Finally, in an old-new recognition task, subjects study many

Table I. A summary of some of the tasks used to study object recognition.

Explicit tasks		
Same-different judgments	Match-to-sample judgments	Old-new judgments
Decide whether two images presented at the same time or in sequence depict the same object.	Recognize the same target object each time it is shown in a list of objects.	Recognize each of several target objects in a list of objects.

Implicit tasks	
Object naming	Object possibility decisions
Name each object in a list of objects as quickly as possible.	Decide whether each object in a list of objects can exist in the 3-D environment.

target objects and then attempt to identify these objects among a set of distractor objects. In all of these tasks, attributes of the target objects such as the viewpoints from which they are seen or their projected sizes are often changed from study to test.

A defining characteristic of explicit recognition tasks is that they require the subject to refer back to specific, previously studied objects to perform the task. Implicit recognition tasks such as object naming or decisions of object possibility (see below) do not require subjects to refer back to previously studied objects. These tasks can be performed solely on the basis of the information presented at the time of test. In an object naming experiment (e.g., Palmer *et al.* (1981)) subjects respond as quickly as possible to each test object by calling out the name of the object or pressing a key corresponding to its name. Although naming an object requires some prior visual experience with objects from the class, it does not require that a particular exemplar of the class be recalled. In an object-decision experiment (e.g., Schacter, Cooper, Delaney, Peterson and Tharan, 1991; Williams and Tarr, 1996) subjects decide whether each test image depicts an object that can exist in the 3-D environment. (Examples of "possible" and "impossible" objects like those used in experiments by Schacter *et al.* (1991) and Williams and Tarr (1996) are shown in Fig. 3F).

In most experiments using an implicit task, subjects view the same objects more than once during the experiment. For example, subjects might name the same set of objects two or more times in different blocks of the experiment. In object-decision experiments subjects often perform very different tasks each time the objects are seen. In Schacter *et al.* (1991) experiments, for example, subjects reported whether the objects faced mostly to the left or mostly to the right the first time they were shown, and they decided whether the objects were possible or impossible the next time they were shown. Although object naming and object possibility decisions can be performed without reference to particular objects seen previously during the experiment, performance is often different for repeated objects (even if subjects cannot accurately report having seen the objects before). For this reason, these tasks are often called *priming* tasks. As in explicit tasks, it is common to vary attributes of the repeated objects from study to test, such as the viewpoint from which they are seen, their projected size, or their position in the visual field.[2]

Objects

The choice of stimuli to be used in an object recognition experiment is of special importance.

Some researchers use familiar, everyday objects, arguing that this allows for a more direct investigation of the visual processing that we engage in most, namely, entry-level categorization. These researchers have typically used artists' renderings of objects such as those found in Snodgrass and Vanderwart (1980), but more recently, with advances in desktop computer graphics, some researchers have begun to use realistically shaded objects such as the chairs shown in Fig. 2. Examples of shaded, line drawing, and silhouette objects such as those used in experiments by Biederman

[2] Some researchers have argued that explicit and implicit recognition tasks might involve very different processing of visual information. Specifically, it has been argued that explicit tasks are more sensitive to episodic attributes of stimulus items such as the particular size and orientation in which they are shown. These issues are beyond the scope of the present chapter, but the interested reader is directed to papers by Biederman and Cooper (1992) and Schacter (1987).

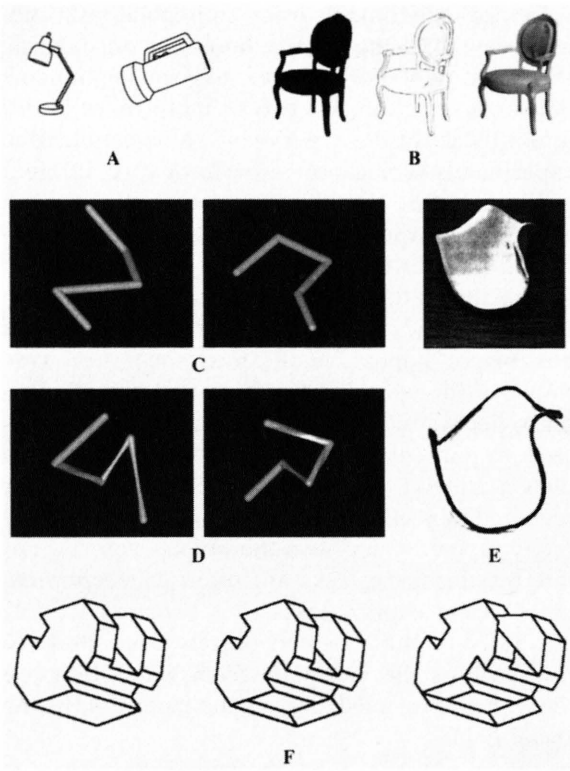

Fig. 3. Some of the objects used in recent object recognition experiments.

and Cooper (1991) and Hayward (1996) are shown in panels A and B of Fig. 3. We will have more to say about these particular objects in later sections.

A serious problem with the use of familiar objects is that it is especially difficult to control the subject's prior exposure to the objects. Although it is unlikely that subjects would have already seen the particular renderings used in an experiment, there remains the fact that they have seen similar objects.[3] Furthermore, it is difficult to control precisely the types of features that are available to distinguish different objects. These concerns have prompted many researchers to use unfamiliar objects. Examples of some of the unfamiliar object classes that have been studied are shown in panels C–F of Fig. 3. We will say more about these particular object classes in later sections.

[3] One must also be concerned with the potential influence of verbal and semantic representations when using familiar objects to study object recognition.

Visual Representations

Three-dimensional models

Given the wide-spread use of computer aided design (CAD) programs in recent years, one might be tempted to believe that visual representations are like the three-dimensional (3-D) models these programs use to represent objects. There are certainly some important advantages to such a representation. For example, a single model is sufficient to represent any given object, so that little memory is needed to represent the object, and it is easy to visualize what the object looks like from any viewpoint in space by simply rotating the model appropriately. For this reason, some theories of object recognition do rely on CAD-like 3-D models. Marr and Nishihara (1978) (see also Marr, 1982), for example, proposed a modeling scheme in which objects are represented in visual memory as hierarchical arrangements of generalized cylinders. A cylinder coincident with the main axis of the object forms the first level of the hierarchy, and the locations and orientations of cylinders in the next level are specified in 3-D coordinates relative to this cylinder. Each of the cylinders in this level can then serve as a reference for defining the locations and orientations of cylinders in the next level of the hierarchy. Because the positions of the parts are defined relative to other parts of the object, the description of an object's shape will be the same regardless of the viewpoint from which it is seen. In principle, then, only one model is needed to fully represent an object's shape.

The primary problem with the use of CAD-like models for visual representation is that models of this sort are very difficult to construct from the information available in 2-D images. Any 2-D image is always consistent with infinitely many 3-D interpretations. We just can't know what's going on behind the surfaces that happen to be visible in the image. Perhaps a more compelling reason to question the use of CAD-like visual representations by the human visual system is that there exists no solid psychophysical evidence that would lead one to believe that CAD-like models are used for object recognition. For example, because these models are completely viewpoint invariant, it should be possible using these models to recognize objects equally well from any viewpoint. Palmer *et al.* (1981) naming experiments suggest that this

is not true for familiar objects, and the data we present below show that it is certainly not true for unfamiliar, computer-generated objects.

Feature descriptions

An alternative method of representing visual information for recognition is to store descriptions of objects that consist only of the features that are visible from a particular viewpoint in space. An object can then be recognized if the feature description derived from an image of the object matches sufficiently well a feature description stored in memory. This approach avoids many of the problems associated with 3-D reconstruction, but it presents it's own problems.

It should be made clear that feature descriptions are not equivalent to "templates" or pictures in the head. Such template models of visual representation are too inflexible to serve for general purpose object recognition, as pictures of the same object taken from different viewing angles or from different distances can look very different. Thus, to recognize an object from arbitrary viewpoints using picture-like descriptions, far too many descriptions would have to be stored.

In contrast to pictures in the head, feature descriptions encode only some of the features present in an image. Determining exactly what these features should be is not an easy question and is a topic of ongoing research. Ideally, features should be easy to extract from images, stable over changes in viewing conditions such as lighting and viewpoint, and unique to particular objects. These criteria, unfortunately, are difficult to meet, and to the extent they are not met, recognition performance will suffer. For example, if the features extracted from one view of an object cannot be identified in another view of the object, then it might be necessary to store more than one feature description of the object to recognize it from these different viewpoints. We present data relevant to this issue in the next sections.

Viewpoint Dependence

Perhaps the most serious problem with the use of feature-based descriptions for object recognition is that different features are likely to be visible in different views of the same object. If only one description of the object were stored in memory, then it would be difficult to recognize the object from viewpoints in which different features were visible. This viewpoint dependence might not be apparent for objects that we are familiar with – as the visual system could store multiple descriptions of familiar objects following extensive experience with different views of these objects – but it should be apparent for unfamiliar objects that have been seen from only one or a few viewpoints. This has in fact been demonstrated in many object recognition experiments using unfamiliar objects.

Rock, DiVita, and Barbeito (1981) (see also Rock and DiVita, 1987), for example, studied recognition of objects that they constructed by bending wires into various 3-D shapes. An object similar to those used by Rock *et al.* is shown in the bottom of panel E in Fig. 3. Subjects studied several of these objects, each from a single 3-D viewpoint, and then attempted to recognize them among other similarly constructed objects. Recognition was quite good for objects that were shown from the studied viewpoint, but recognition was very poor for objects shown from novel viewpoints. Humphrey and Khan (1992) reported comparable findings for recognition of unfamiliar 3-D objects composed of differently shaped clay parts. Subjects studied 39 of these objects, each seen from a single viewpoint, and then attempted to recognize them in an old-new recognition task among similarly constructed distractor objects. Target objects were recognized easily if they were tested from the studied viewpoint, but they were recognized quite poorly if they were tested from novel viewpoints.

Similarly, Bülthoff and Edelman (1992) (see also Edelman and Bülthoff, 1992) observed viewpoint-dependent recognition of computer-generated, "paper-clip" objects composed of several cylinders connected end-to-end. Two such objects are shown in panel C of Fig. 3. Subjects studied one of these objects from two different viewpoints (a randomly selected "frontal" view and a view rotated +75° in depth about the vertical axis) and then attempted to recognize it from specific novel viewpoints in a match-to-sample task among similarly constructed distractor objects. Novel viewpoints were generated by (1) rotating the object about the vertical axis to views that were between the studied views (the INTER condition), (2) rotating the object

about the vertical axis in the opposite direction (the EXTRA condition), or (3) rotating the object about the horizontal axis (the ORTHO condition). Overall, recognition was better for views near the studied views, but performance was best in the INTER condition, somewhat worse in the EXTRA condition, and worst in the ORTHO condition. The significance of these findings will be discussed below.

Viewpoint Independence

The extreme viewpoint dependence observed in the studies just discussed seems to be at odds with our phenomenal experience that objects are easy to recognize regardless of the viewpoint from which they are seen. We will discuss two ways in which viewpoint-invariant recognition can be achieved with more or less viewpoint-dependent feature descriptions: (1) store multiple descriptions of objects, each specific to a different viewpoint, and (2) encode features that are more easily recognized over a wide range of views.

Multiple descriptions

A series of experiments by Tarr (1995) lend clear support to the hypothesis that the visual system stores multiple descriptions of the same object. Subjects in his study learned names for three objects, each of which was shown repeatedly from four specific viewpoints. The objects were composed of blocks connected face-to-face like the objects used by Shepard and Cooper in their studies of mental rotation (Shepard and Cooper, 1982). Tarr found that the time needed to name novel views of the objects presented in later blocks of the experiment varied with the 3-D angular distance to the *nearest* studied view. This suggests that viewpoint-specific descriptions were stored for each of the studied views, and that novel views were recognized by mentally rotating the test object to the nearest studied view.

Bülthoff and Edelman's (1992) experiments suggest another way in which multiple stored descriptions of an object can be used to recognize entirely novel views of the object. Ullman and Basri (1991) showed that it is possible, in principle, to recognize certain novel views of an object if the 2-D image locations of a small number of the features visible in the view can be expressed as a linear combination of the image locations of corresponding features in two or more known views of the object. For example, if the two known views differ by a rotation about the vertical axis, then it is possible, in principle, to recognize any other view of the object that is generated by a rotation about the vertical axis. On the contrary, views generated by rotating the object about other axes, for example, the horizontal axis, cannot be recognized in this way. This is precisely the pattern of recognition performance that was observed in Bülthoff and Edelman's experiments using INTER, EXTRA, and ORTHO conditions.

Better feature descriptions

Although the visual system can store multiple viewpoint-dependent descriptions of the same object, it would be a serious problem if too many descriptions had to be stored to recognize everyday objects as well as we do from arbitrary viewpoints. The extreme viewpoint dependence observed in the experiments discussed above seems to suggest that feature descriptions are highly viewpoint specific, so that many descriptions would be needed to recognize an object from arbitrary viewpoints. It is important, however, to consider what kinds of features were available to discriminate the objects used in those studies. For example, different paper clip objects in Bülthoff and Edelman's (1992) experiments could be discriminated only by the 3-D connection angles between the cylinders. Experiments by Sklar, Bülthoff, Edelman, and Basri (1993) indicated that subjects recognize these objects on the basis of the 2-D projected connection angles, not the 3-D angles. The 2-D angles change considerably if the object is rotated in depth, perhaps preventing accurate recognition.

Recent experiments by Liter (1996) and Farah, Rochlin, and Klein (1994) indicate that recognition of rotated objects can be enhanced if targets and distractors can be discriminated on the basis of features that are more easily identified in rotated views. Liter (1996) studied objects similar to Bülthoff and Edelman's paper clip objects, except that three different part shapes were used in making the objects. Examples of two of these objects are presented in panel D of Fig. 3. These two objects can be discriminated by the relative positions

of the differently shaped parts. The object on the left, for example, has two constricted cylinders on one of its ends, whereas the object on the right does not. Subjects studied six of these objects, each seen from a single viewpoint, and then recognized them among an equal number of distractor objects in an old-new recognition task. As in Bülthoff and Edelman's studies, recognition declined with rotation in depth when the objects differed only by the connection angles between the parts. However, recognition of rotated objects was enhanced if the order in which the different part shapes were connected was different in different objects (as in Fig. 3D). This experiment demonstrates that the range of viewpoints over which a single studied view is useful for recognition depends on the features that are available to distinguish different objects. Apparently, the different part shapes could be identified from a wider range of viewpoints than could the connection angles.

Similarly, Farah *et al.* (1994) found that adding 3-D surface features to bent-wire objects like those studied by Rock and DiVita reliably enhanced recognition of rotated objects. Objects with surface features were made by bending oval clay disks into shapes like potato chips. Wire objects having no surface features were made by tracing the edge of each bent disk with a wax-covered string. Examples of both types of object are shown in panel E of Fig. 3. Subjects viewed two surface objects or two wire objects in sequence and decided whether they were the same or different. When the same object was shown in both intervals, the viewpoint from which it was seen was sometimes different. Different views of surface objects were more easily matched than different views of wire objects, suggesting that the additional features available in the surface objects were more easily recognized in rotated views.

Entry-Level Recognition

The studies of Liter and Farah *et al.* indicate that viewpoint dependence in object recognition depends critically on the types of features that are available to distinguish different objects. The unfamiliar objects used in the studies discussed above were very similar to one another, as is typically the case when one must distinguish among objects in the same entry-level class. To discriminate among similar objects, one must rely on rather precise features. Often, these features are difficult to identify in different views of the same object. Objects in different entry-level classes typically differ by more complex features that are likely to be identifiable from a wider range of viewpoints. Tversky and Hemenway (1984), for example, found that parts are especially useful for distinguishing objects in different entry-level classes. They found that objects in different basic-level classes are typically composed of different parts, whereas objects in the same entry-level class are typically composed of the same parts.

Parts

Observations such as those made by Tversky and Hemenway have prompted some theorists to propose part-based theories of object recognition (Biederman, 1987; Hoffman and Richards, 1984; Marr and Nishihara, 1978). These theories fall into two categories: (1) primitive-based theories, which rely on a specific predefined set of parts, and (2) boundary-based theories, which define rules for locating part boundaries rather than specifying candidate parts in advance. Marr and Nishihara's (1978) theory is primitive-based because it represents objects as a hierarchical arrangement of cylinders. Biederman's (1987) Recognition-by-Components (RBC) theory (see also Hummel and Biederman, 1992) is also a primitive-based theory, as it represents objects with a set of 36 geometric shapes termed "geons." Unlike the 3-D models proposed in Marr and Nishihara's theory, geon descriptions are not based on a 3-D reconstruction of the object. Rather, the description consists only of a specification of which geons are visible in the image and the gross 2-D spatial relationships among them, for example, geon *A* is above geon *B*, or geon *A* is to the side of geon *B*. Because these descriptions encode only the parts of an object that are visible from a particular viewpoint (Biederman and Gerhardstein, 1993), multiple descriptions of an object will have to be stored if the object is to be recognized from sufficiently different viewpoints. Nevertheless, if parts *are* more easily identified in rotated views, then viewpoint generalization should be better when targets and distractors differ by the makeup of their parts.

Hoffman and Richards (1984) agreed that objects are represented by descriptions of their parts,

but they argued that candidate parts do not have to be defined in advance. Rather, they argued that one needs only to define rules for locating parts in images. They argued that proposals for particular sets of parts are ad hoc, arbitrary, and have never been demonstrated to be adequate for representing natural objects (see also Kurbat (1994) for a discussion of the generality of Biederman's RBC theory). Hoffman and Richards' scheme for locating part boundaries derives from the transversality regularity, which states that "When two arbitrarily shaped surfaces are made to interpenetrate they always meet in a contour of concave discontinuity of their tangent planes" (p. 69). Such 3-D discontinuities produce concomitant discontinuities in the object's projected silhouette, so that one can infer 3-D part boundaries on the basis of simple image information.

The majority of the experiments investigating the role of parts in object recognition have used priming tasks such as object naming. Priming is a useful phenomenon for investigating what features are used in object recognition because priming is believed to occur only if the same visual features are processed in the same way during study and test (see, e.g., Roediger, Weldon, and Challis (1989) and Schacter (1990) for discussions of the nature of perceptual priming).

As a simple example of visual priming, consider a series of experiments by Bartram (1974) in which subjects repeatedly named photographs of the same set of 12 objects in eight blocks of trials. There were eight exemplars of each object class (e.g., eight different chairs) so that sometimes subjects saw a different exemplar of each class in every block (the different-exemplar condition), and sometimes subjects saw the same exemplar in every block. When the same exemplar was shown, sometimes it was seen from the same viewpoint in every block (the same-view condition), and sometimes it was seen from a different viewpoint in every block (the different-view condition). In all cases, the time to name the objects decreased as the experiment progressed, indicating that having named the objects in prior blocks facilitated or primed naming them in later blocks.

Priming in the different-exemplar condition probably was not based on repeated processing of the same visual representations of the objects, as the different exemplars would have shared few visual features. Instead, priming in this condition was most likely mediated by a modality-free conceptual or semantic representation. Moreover, one can be sure that priming in this condition was not due to subjects simply responding more quickly as the experiment progressed (i.e., task learning), as altogether new objects presented in each block were named no more quickly than objects in the first block of trials. The important question for our purposes is whether priming in the same- and different-view conditions was also based on repeated use of the same non-visual representations. Two results indicate that much of the priming in these conditions was, in fact, visual. First, priming was greater in both of these conditions compared to the different-exemplar condition (with priming greatest for the same-view condition). Second, priming in the same- and different-view conditions did not transfer to the different-exemplar condition. In the transfer experiments, subjects named the same exemplars of each object class repeatedly during the first six blocks and then named new exemplars in blocks seven and eight. The new exemplars were named almost as slowly as the new objects presented in blocks seven and eight. This suggests that priming was based on visual representations specific to the objects shown during the first six blocks. Had the priming from blocks one through six been based on a representation that was not specific to the visual form of the viewed object, subjects should have named new exemplars much faster in blocks seven and eight.

Biederman and his colleagues have conducted a number of experiments to investigate whether priming in object naming tasks depends on whether the same parts are visible each time the object is named. Biederman and Cooper (1991), for example, had subjects name degraded line drawings of objects in which half of the contour had been deleted. Two versions of each drawing were created such that the contour missing from one drawing was present in the other (complementary) drawing. The deleted contour either formed complete parts of the object so that the two drawings of the object contained different parts (the complementary component condition), or the deleted contour did not form complete parts (the complementary contour condition). Biederman and Cooper argued that, although half of the contour was deleted in the complementary

contour condition, unlike in the complementary component condition, the same parts were visible in both drawings. Subjects named these contour-deleted drawings in two experimental blocks. As in Bartram's (1974) experiments, subjects named identical drawings faster in the second block. Complementary contour drawings showing the same parts as the drawings named in the first block were also named faster in the second block, but complementary component drawings showing different parts were not.

In a similar series of experiments, Biederman and Gerhardstein (1993) had subjects name intact line drawings of objects. The objects were either shown from the same viewpoint in both blocks, or they were shown rotated about the vertical axis in the second block. Regardless of whether the objects were rotated, subjects named them faster in block two relative to different exemplars of the same objects or entirely new objects. Furthermore, priming was slightly greater if the same parts were visible in the rotated views than if different parts were visible.

Although the visible-parts explanation of name priming seems at first to be a good one, several recent experiments cast doubt on whether it is a complete explanation. Srinivas (1995), for example, found that priming depended on the visible parts only sometimes. In many cases, priming diminished reliably with rotation in depth even when the same parts were visible in the study and test views. This suggests that priming might be based on processing of features other than parts.

Further evidence against the visible-parts explanation comes from experiments by Cave and Kosslyn (1993). Their subjects named line drawings of objects that had been broken into pieces either at natural part boundaries (in some sense preserving the visible parts) or unnatural part boundaries. Naming time was longer for fragmented as opposed to whole objects, but it was not slower for unnatural than for natural part breaks. Similarly, naming was equally slow for unnatural and natural part breaks when the fragments were moved to different image locations, creating scrambled images. These results suggest that recognition depends on the proper spatial arrangement of visual features, but that these features are not necessarily whole parts.

The bounding contour

An interesting series of experiments by Hayward (1996) suggests that a very simple source of visual information, namely, the shape of an object's bounding contour, can be used to predict the magnitude of priming in object naming tasks. Subjects in Hayward's experiments named shaded images of familiar objects in the first block of the experiment and then named rotated versions of the objects in the second block. The objects in the second block were displayed with shading information as in block one, or they were displayed as black and white silhouettes. The magnitude of priming was the same for both shaded objects and silhouettes. This result argues against the part-based explanation of priming, as recovering an object's parts from a silhouette is very difficult, and is likely to result in a part description that is much different from the description that would be obtained from a shaded image.

Hayward found further evidence in favor of the bounding-contour explanation in another experiment in which fully shaded objects in block two were rotated by a small amount from block one (so that the visible components were the same in both blocks) or by 180 degrees. The visible parts in the 180 degree condition were often very dissimilar from those in block one (because the back of the object was shown rather than the front), but the bounding contour was very similar. Aside from slight distortions due to perspective projection, the bounding contour in the 180 degree condition was simply a mirror reflection of the bounding contour seen in block one. The bounding contours of the objects rotated less than 180 degrees were much more dissimilar. Priming in this experiment was greater for objects rotated 180 degrees, suggesting that the shape of the silhouette might be an important source of information for initially accessing visual information about objects. The bounding-contour explanation of priming can also explain Biederman and Cooper's (1991) findings with complementary-contour and complementary-component line drawings, as the bounding contours of complementary-contour drawings were more similar to one another than were the bounding contours of complementary-component drawings.

Summary

The evidence presented here indicates that visual representations for object recognition consist of viewpoint-specific descriptions of a wide variety of different features. Some of these features are relatively easy to identify in rotated views, whereas others can be identified from only a limited range of viewpoints. We have shown that recognition performance can be explained in different situations by considering which of these features are used to distinguish the objects of interest. Features such as the visible components or the shape of the bounding contour are often sufficient to distinguish objects in different entry-level categories. As these features often remain unchanged over a wide range of viewpoints, entry-level categorization is largely insensitive to changes in viewpoint. To distinguish different objects in the same entry-level category, it is often necessary to rely on more precise features such as the connection angles between the objects' parts or the image locations of salient edges and vertices. These features can appear very different in different views of the same object, making recognition performance with objects from the same entry-level class sensitive to changes in viewpoint. Finally, we showed how the limitations imposed by relying on viewpoint-specific descriptions can be overcome by storing multiple descriptions of the same object, or by implementing transformation mechanisms.

Although we have not attempted to discuss neuropsychological and neurophysiological research in the present chapter, these approaches are likely to have a greater and greater impact on object recognition research in the future. Researchers in these fields examine recognition performance in brain injured populations (e.g., Farah, 1990) and in nonhuman primate populations (e.g., Logothetis, Pauls, Bülthoff, and Poggio, 1994). Recent advances in brain imaging technology such as functional magnetic resonance imaging will also contribute significantly in the near future to our understanding of the brain mechanisms that contribute to object recognition.

A great deal of work remains to be done. We need a firm computational definition of what constitutes a visual feature. We need to determine whether there are rules specifying which views of an object should be stored in memory. This could be based entirely on the frequency with which different views are experienced, or it could be based on geometric constraints. Some views of an object might simply be more informative than others. Finally, we must explore further how the visual system recognizes novel views of objects using only a small set of stored descriptions.

Bartram D. J. (1974), The role of visual and semantic codes in object naming. Cognit. Psych. **6**, 325–356.

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

Biederman I. and Cooper E. E. (1991), Priming contour-deleted images: Evidence for intermediate representations in visual object recognition. Cognit. Psych. **23**, 393–419.

Biederman I. and Cooper E. E. (1992), Size invariance in visual object priming. J. Exp. Psychol.: Human Perception and Performance **18**, 121–133.

Biederman I. and Gerhardstein P. C. (1993), Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. J. Exp. Psychol.: Human Perception and Performance **19**, 1162–1182.

Blanz V., Vetter T., Bülthoff H. H. and Tarr M. J. (1995), What object attributes determine canonical views? Perception **24** (Supplement), 119c.

Bruce V. (1988), Recognising Faces. Lawrence Erlbaum Associates, Hillsdale, NJ.

Bülthoff H. H. and Edelman S. (1992), Psychophysical support for a two-dimensional view interpolation theory of object recognition. Proc. Natl. Acad. Sci. USA **89**, 60–64.

Cave C. B. and Kosslyn S. M. (1993), The role of parts and spatial relations in object identification. Perception **22**, 229–248.

Edelman S. and Bülthoff H. H. (1992), Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. Vis. Res. **32**, 2385–2400.

Farah M. J. (1990), Visual Agnosia: Disorders of Object Recognition and What They Tell Us about Normal Vision. MIT Press, Cambridge, MA.

Farah M. J., Rochlin R. and Klein K. L. (1994), Orientation invariance and geometric primitives in shape recognition. Cognit. Sci. **18**, 325–344.

Hayward W. G. (1998), Effects of outline shape in object recognition. J. Exp. Psychol.: Human Perception and Performance, in press.

Hoffman D. D. and Richards W. A. (1984), Parts of recognition. Cognition **18**, 65–96.

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

Humphrey G. K. and Khan S. C. (1992), Recognizing novel views of three-dimensional objects. Canad. J. Psychol. **46**, 170–190.

Jolicoeur P., Gluck M. A. and Kosslyn S. M. (1984), From pictures to words: Making the connection. Cognit. Psychol. **16**, 243–275.

Kurbat M. A. (1994), Structural description theories: Is RBC/JIM a general-purpose theory of human entry-level object recognition? Perception **23**, 1339–1368.

Liter J. C. and Bülthoff H. H. (1996), The contribution of distinct component configurations to object recognition across changes of view. Invest. Ophthalm. Vis. Sci. **37** (Supplement), 177.

Logothetis N. K., Pauls J., Bülthoff H. H. and Poggio T. (1994), View-dependent object recognition by monkeys. Curr. Biol. **4**, 401–414.

Marr D. (1982), Vision. Freeman, San Francisco.

Marr D. and Nishihara H. K. (1978), Representation and recognition of the spatial organization of three dimensional shapes. Proc.R. Soc. London B **200**, 269–294.

Palmer S. E., Rosch E. and Chase P. (1981), Canonical perspective and the perception of objects. In: Attention and Performance IX (J. Long, A. Baddeley, eds.). Lawrence Erlbaum Associates, Hillsdale, NJ, 135–151.

Rock I. and DiVita J. (1987), A case of viewer-centered object perception. Cognit. Psychol. **19**, 280–293.

Rock I., DiVita J. and Barbeito R. (1981), The effect on form perception of change of orientation in the third dimension. J. Exp. Psychol.: Human Perception and Performance **7**, 719–732.

Roediger H. L., Weldon M. S. and Challis B. H. (1989), Explaining dissociations between implicit and explicit measures of retention: A processing account. In: Varieties of Memory and Consciousness: Essays in Honour or Endel Tulving (H. L. Roediger, F. I. M. Craik, eds.). Lawrence Erlbaum Associates, Hillsdale, NJ, 3–41.

Rosch E., Mervis C. B., Gray W., Johnson D. and Boyes-Braem, P. (1976), Basic objects in natural categories. Cognit. Psychol. **8**, 382–439.

Schacter D. L. (1987), Implicit memory: History and current status. J. Exp. Psychol.: Learning, Memory, and Cognition **13**, 501–518.

Schacter D. L. (1990), Perceptual representation systems and implicit memory: Toward a resolution of the multiple memory systems debate. Ann. N.Y. Acad. Sci. **608**, 543–571.

Schacter D. L., Cooper L. A., Delaney S. M., Peterson M. A. and Tharan M. (1991), Implicit memory for possible and impossible objects: Constraints on the construction of structural descriptions. J. Exp. Psychol.: Learning, Memory, and Cognition **17**, 3–19.

Shepard R. N. and Cooper L. A. (1982), Mental Images and Their Transformations. MIT Press, Cambridge, MA.

Sklar E., Bülthoff H. H., Edelman S. and Basri R. (1993), Generalization of object recognition across stimulus rotation and deformation. Invest. Ophthal. Vis. Sci. **34** (Supplement), 1081.

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

Srinivas K. (1995), Representation of rotated objects in explicit and implicit memory. J. Exp. Psychol.: Learning, Memory, and Cognition **21**, 1019–1036.

Tarr M. J. (1995), Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. Psych. Bull. Rev. **2**, 55–82.

Tversky B. and Hemenway K. (1984), Objects, parts, and categories. J. Exp. Psychol.: General **113**, 169–193.

Ullman S. and Basri R. (1991), Recognition by linear combinations of models. IEEE Transactions on Pattern Analysis and Machine Intelligence **13**, 992–1005.

Williams P. and Tarr M. J. (1997), Structural processing and Implicit memory for possible and impossible figures. J. Exp. Psychol.: Learning, Memory, and Cognition **23**(6), 1344–1361.