Recognizing novel deforming objects

Lewis L. Chuang
MPI for Biological Cybernetics

Quoc C. Vuong
MPI for Biological Cybernetics

Ian M. Thornton
Department of Psychology,
University of Wales Swansea

Heinrich H. Bülthoff
MPI for Biological Cybernetics

Abstract

Computer graphics have come to play an integral role in research on human perception. Critically, it provide a high level of control over the visual attributes of the test stimuli. This control allows relevant factors of human perception to be independently controlled and systematically studied. Computer graphics techniques can also be used to reduce the number of parameters that needs to be varied. Although this reduces the search space, no study to date has tested the extent to which observers are sensitive to manipulations within a low-dimensional parameter space for object recognition. As a first step in that direction, we provide an experiment that used a method to create animations of novel 3D objects with minimal a priori constraints. Our findings revealed a perceptual sensitivity to dynamic information during the recognition and learning of novel objects.


Keywords: motion; psychophysics; object recognition; animation.

1 Introduction

In numerous studies, investigators are concerned with human observers’ ability to learn and identify unfamiliar objects across varying viewing conditions (for a recent review, see Palmeri and Gauthier, 2004). An experimental method that is widely used in such research involves the systematic manipulation of visual attributes of the stimuli. By measuring human behavioral response, one is able to make inferences concerning the relationship between the physical stimulus and its corresponding representation.

In this light, the recent growth of computer graphics techniques offers the psychophysical researcher a unique opportunity to control visual attributes relevant for human perception. For example, these techniques are often used to create artificial novel objects that observers have never experienced. As a consequence, researchers can be confident that behavioral effects are due to their manipulations rather than to observers’ prior experiences. Of particular interest here is the application of these techniques to create dynamic novel 3D objects. As we will discuss in the following sections, using dynamic objects in psychophysical research presents an interesting challenge for the researcher, a challenge for which recent computer graphics and animation techniques are well-suited. Our aim in this paper is to demonstrate how dynamic novel objects can be generated with a small set of parameters that is, nonetheless, sufficient to produce significant effects on human behavior.

The paper is organized as follows. In Section 2, we present an unresolved issue in the area of visual object recognition research: Whether observers are sensitive to dynamic information when learning to recognize objects, independent of stimulus familiarity and/or structural information. To investigate this, in Section 3, we describe possible methods to create novel 3D shapes involving complex non-rigid deformations. These methods allow for the experimental manipulation of dynamic information, independently of shape information. Importantly, the generation of our stimuli was reducible to the specification of only a small number of parameters and easily created with commercial software i.e. 3D StudioMax v.7. Finally, in Sections 5 and 6, we present and discuss human-perception data that is consistent with past findings.

2 Dynamic object recognition

There has been a recent interest in the processing of dynamic information in visual object recognition. To date, the evidence suggests that human observers are sensitive to dynamic attributes of an object. For example, it is generally accepted that observing a human face in motion is beneficial for learning and identifying people (e.g. Knappmeyer, Thornton, and Bülthoff, 2003; Lander and Chuang, 2005). However, sensitivity to dynamic information with highly familiar objects e.g. faces, could be stimulus-specific. For example, dynamic information in faces is usually relevant for social communication and could be highly learnt as a result.

To demonstrate the general relevance of dynamic information to object learning and recognition, Stone (Stone, 1999) trained participants on novel rigidly rotating objects. During a study phase, these objects followed a fixed pattern of rotation. He found that recognition performance of the trained objects was impaired at a later testing stage if their studied rotation pattern was reversed (i.e. the frame sequence was played in reverse order from training). This finding has also been replicated using novel objects with different geometry and using a different recognition task (Liu and Cooper, 2003). We will use this behaviorally robust “reversal effect” to validate our choice of stimulus generation (Section 3).
From the above findings, it is still not clear whether motion benefits recognition because it is a source of identity-specific information, or whether it contributes to the salience of shape information (e.g. recovery of 3D shape; Ullman, 1979). To address this question, we directly manipulated dynamic information — that is, how the 3D shape changes over time— independently of shape information. This was done by using a relatively straightforward and flexible computer graphics method. The next section focuses on this point.

3 Computer graphics and Stimuli generation

3.1 Parameterized models

It is sometimes assumed that complex shapes can be decomposed into independent parameters. Such parameters can be theorized either as the object’s geometric component parts (Biederman, 1987) or measured features (Cutzu and Edelman, 1998).

Such parameterization models are useful because they can form the theoretical basis of how humans represent objects. In addition, they can be implemented for the generation of stimuli for direct testing. For example, Cutzu and Edelman (1998) used 70 dimensions to create novel animal-like 3D shapes. They found that the configural arrangements of stimuli generated within this parameter space corresponded to the perceptual judgments of human observers. For example, observers reproduced the same configuration in time. This correspondence supported the possibility that representations of objects can be defined and organized in terms of their similarity to one another in perceptual measurement space.

There is no doubt that parameterized models offer great flexibility and experimental control for the purposes of stimulus generation. Furthermore, activation in neural regions associated with object recognition is sensitive to manipulations in multidimensional models (Edelman, 2002). The challenge, however, is that these models usually require some a priori definition of the parameters that are to be manipulated. “Good” models are models whose parameters define all the relevant perceptual dimensions. However, the purpose of conducting psychophysical research is to uncover these same relevant dimensions. For dynamic objects, this challenge is compounded by the fact that there is no consensus as to what such perceptually relevant dimensions might be. That said, whatever these parameters might be, they should be capable of exerting a broad influence over the object’s spatio-temporal properties to capture the perceptually relevant aspects of both the shape of an object and how that shape might change over time.

From a practical point of view, it is important to be able to vary objects with a small number of parameters and yet produce complex deforming shapes. This would also prevent the possible exclusion of attributes that could be relevant for human perception. Fortunately, there are several techniques to create complex shapes from a small number of parameters e.g. supershapes (Giels, Beirinckx, and Bastiaens, 2003) and physics-based modeling (Faloutsos, van de Panne, and Terzopoulos, 1997). For present purposes, we chose a relatively simple method— free form modeling. The following section provides a description of this and related techniques that provide a viable solution to the challenge posed above.

3.2 Free form modeling

One generic class of methods to smoothly and efficiently deform objects is free form deformation (FFD). The underlying idea for FFD is quite simple. An object is embedded inside a geometric space, typically a cube or cylinder. By warping that space using a few control parameters, the embedded object can be smoothly deformed in a systematic and predictable manner. FFD can be applied to different graphical objects such as points, polygons, and surfaces.

Barr (1984) originally formulated the deformation problem in terms of a geometric mapping of 3D spaces. This mapping provided simple controls to bend, twist, and rigidly rotate objects, for examples. More complicated deformations can be achieved by composing different mapping functions. Furthermore, the mapping functions can vary over time, producing the desired dynamic objects.

Sederberg and Parry (1986) later generalized Barr’s (1984) space-warping approach, and introduced the term “free form deformation”. They embedded objects into a lattice of grid points arranged in a cube. By manipulating isolated grid points, the local space inside the grid is deformed and this deformation induces transformations of the underlying graphics primitives within that space. For example, vertices of a surface would be directly mapped to new positions. By manipulating sets of grid points global transformations of the embedded graphics primitives can similarly be achieved. The power of FFD and similar methods is that they provide the user with a small number of parameters that can be systematically manipulated in space to model arbitrary shapes, and in time to animate these shapes.

Figure 1.3.1: Illustrations of a global twist deformation and a local free form deformation of a torus.

These state of the art deformation methods are standard tools in a number of leading commercial software, such as
3D Studio Max v.7. They have also been applied to diverse areas in computer graphics such as object and character animation, image analysis, and surgical simulations. For a review of FFD and other deformation methods, see Gibson and Mirtich (1997).

4 Stimuli generation

Here, we used a method similar to FFD to generate novel amoebas that deformed over time. As illustrated in the Figure 2.1.1a, we first embedded a sphere inside a cube, which defined the spatial limits of the deformation space. To reduce the number of parameters that we would need to vary, we deformed the cube with sinusoidal functions along each coordinate axis (for a review of similar methods, see Phillips, 2004). Thus, the parameters were the amplitude and phase of the functions for each axis. We further reduced the number of parameters by fixing the amplitudes of the sine waves along the axes. The summation of these phase-shifted sinusoids deformed the sphere thereby generating an amoeba. To generate different amoebas, we randomly initialized the phase of the three sinusoids. Figure 2.1.1b provides an example of the applied transformation.

Figure 2.1.1: Comparison between (a) the original sphere and (b) the transformed sphere. Dark lines describe the perimeter of the surrounding cube pre-/post-transformation.

To animate these objects, we varied the phase of the sinusoidal functions over time. That is, the initial phases were offset by a sinusoidal function to ensure smooth animation. Each amoeba created in this manner deformed with a characteristic pattern because the initial phase was randomly determined. We created 4 such amoebas. For each stimulus, we also calculated how much the distance of each vertex deviated from the base sphere over time. The mean standard deviation was 2.75 for the 4 amoebas (range: 2.63-3.00). Thus our method does not introduce large deviations from a sphere that could be used as a cue to object identity. Finally, we created video sequences of the 4 amoebas for the experiment. Each sequence consisted of 100 frames and was presented at 25 frames per second.

In effect, we coupled complex changes in shape to three parameters, i.e. the initial phase of the three sinusoidal functions. In the next section, we present data to suggest that observers are sensitive to this coupling in non-trivial ways by replicating a robust finding of “rotation reversal” (Stone, 1999).

5 Psychophysical procedures

We used an old/new recognition task to test whether observers represent the dynamics of objects. In our task, observers first learned to identify two non-rigidly deforming amoebas across 102 trials. They then discriminated old targets that were seen during the learning phase from new distractors that had not been seen before, in a test phase comprising 352 trials. Twelve observers were tested.

The experiment consisted of a learning phase followed by a testing phase. Throughout the experiment, two amoebas randomly served as targets and two as distractors. During the learning phase, the targets were presented as video sequences. These sequences were always played in a forward serial order. Furthermore, only 75 frames of the 100 total were presented on a given trial. The starting frame varied from trial to trial so that presentations of the targets would vary slightly across trials. Figure 2.2.1a illustrates this.

The two targets were presented individually on each learning trial. Participants were instructed to press the key assigned to that target as quickly and as accurately as possible. To ensure learning, audio feedback was provided for incorrect responses.

During the test phase, participants were presented with video sequences of both targets and distractors. They decided whether the sequence on a given trial was old (i.e. one of the two targets) or new (i.e. one of the two distractors) as quickly and as accurately as possible. Each sequence in the test phase was 40 frames in length and was extracted between frames 26 to 75 from the original 100 frames. This was done to ensure that every frame shown during the test trials was presented an equal number of times during the learning trials.

Importantly, each video sequence was presented either in the forward serial order shown during the learning phase or in the reverse order (see Figure 2.2.1b for details). With this manipulation, the shape information of each target was
preserved, while the dynamic information was altered. This manipulation was introduced to test if observers could learn such information for non-rigid object motion. Note again, that shape information (e.g. the set of views) was left unaffected by this reversal manipulation.

6 Results and Discussion

Performance on the test trials was assessed by recognition accuracy. Figure 2.3.1 illustrates the participants’ mean accuracy in the different conditions. A 2 Stimulus Type (target, distractor) x 2 Sequence Order (forward, reverse) within-subjects analysis of variance was conducted on the accuracy data. This analysis revealed a significant main effect for Stimulus Type (F(1,11)=9.88, p<0.05) and Sequence Order (F(1,11)=32.41, p<0.001). Importantly, the interaction between these factors was also significant (F(1,11)=34.48, p<0.001). Participants were less accurate on the old/new recognition task when the video sequence was played in reverse frame order for targets but not for distractors.

![Figure 2.3.1: Participants’ recognition accuracy. Error bars represent the standard error of the mean.](image)

These findings suggest that observers recognized non-rigidly deforming objects, not only on the basis of their 2D shapes, but also from the specific way in which their shapes deformed over time. Recall that the reversal manipulation does not affect the 3D shape or the set of 2D appearances. Rather, the reversal manipulation had a specific effect on the dynamic cues. This is consistent with previous findings and show that dynamic information can be integral to an object’s representation (Liu and Cooper, 2003; Stone, 1999).

Speculatively, sensitivity to the way that an object’s shape changes over time could also contribute to how observers are able to recognize objects despite changing viewing conditions. Moreover, objects tend to transform predictably across time or move in a characteristic fashion. For example, familiar individuals can be identified simply from their gait pattern in the absence of form cues (Cutting and Kozlowski, 1977).

7 Conclusions

To summarize, this paper highlights the interaction between computer graphics and psychophysical research. The increased sophistication in computer graphics allows researchers to create novel stimuli that extend the theoretical questions regarding human perception that can be raised. In turn, the consistency of findings with past research lends validity to the computer graphics methods used. This is particularly important for research in dynamic object recognition, in which the relevant variables are largely unknown. Part of the problem in this regard is that there has been a scarcity of simple methods to systematically create dynamic objects. It is hoped that the current work provides a demonstration of how such a problem could be adequately rectified.

8 Reference List


