Representing familiar and novel objects by similarities to reference shapes

Sharon Duvdevani-Bar
Dept. of Applied Math. & CS
The Weizmann Institute of Science
Rehovot 76100, Israel
sharon@wisdom.weizmann.ac.il

Shimon Edelman
Center for Biological & Computational Learning
Dept. of Brain and Cognitive Science, MIT E25-201
Cambridge, MA 02142, USA
edelman@ai.mit.edu

A visual system faced with the problem of object recognition must overcome the variability in the object's appearance caused by factors such as illumination and pose, to be able to identify different views of the same shape as such. Because *recognition* calls for remembering an object that has been seen before (albeit from another angle, or under a different illumination), it is amenable to a memory-based approach, according to which representative views of the object are stored and subsequently interpolated between to recognize its novel views (Poggio and Edelman, 1990). In comparison, *categorization* can be described as making sense of an object with which the system has had no prior encounter; traditionally, it has been assumed that this task can only be approached by representing objects in terms of generic parts and their spatial relationships (Biederman, 1987).

A recent theory (Edelman et al., 1996) offered a unified approach to recognition and categorization, according to which both familiar and novel objects are represented by their similarities to a chosen set of prototypes, or reference shapes. We now present a connectionist implementation of this approach, which addresses the needs of superordinate and basic-level categorization, and of identification of specific instances of familiar categories.

The implemented model consists of ten classifiers, each tuned to a 3D object chosen at random from a commercially available computer graphics database. The classifiers were realized by radial basis function networks (Poggio and Edelman, 1990), and were trained to recognize their assigned shapes, using as few as 15 stored views per object, chosen according to a canonical vector quantization algorithm. Test stimuli (images of novel or familiar objects) were then represented by the vectors of responses of the ten classifiers (which may be considered as points in a 10-dimensional shape space).

We evaluated the model's performance on three tasks: (1) *recognition* of novel views of the prototypes (reference objects), (2) *categorization* of views of novel objects belonging to categories of which as least one member was available as a prototype, and (3) *representation* of radically novel objects and discrimination among such objects. In all the experiments, the test set consisted of 169 views per object, taken around the canonical orientation, over a range of $\pm 60°$ in azimuth and elevation, at 10° increments.

In the recognition task, the Winner-Take-All algorithm, according to which the identity of the stimulus is determined by the label of the classifier that responds the strongest, resulted in an error rate of 10%. We then trained a second-stage classifier to map the 10-dimensional vectors of the first-level classifier responses into vectors of the same size, in which the single proper element (signifying the identity of the stimulus) was set to 1, and the others to 0. This modification reduced the error rate to 6%.

In the categorization task, we tested the ability of the model to assign views of 20 novel objects to their proper categories (e.g., both cow and giraffe had to be labeled as quadrupeds). The category label for a view of a test object was determined by a k-nearest-neighbor (Duda and Hart, 1973) majority vote in the 10-dimensional shape space, resulting in a misclassification rate of 17%. A 2D rendition of the shape-space arrangement of the views of the objects (obtained with multidimensional scaling) revealed a satisfactory clustering of the stimuli by category. This was also true of the representation task, in which the model had to discriminate among views of ten radically novel objects, represented in the same 10-dimensional shape space. The discrimination error rate in this experiment was 10%. Confining the test views to a smaller range around the canonical view reduced both the misclassification and the discrimination errors considerably.

In summary, the implemented model supports representation of and discrimination among shapes radically different from the reference ones, while bypassing the computationally problematic need for the extraction of parts and determination of their spatial relationships in stimulus images. Description by similarities to prototypes thus offers a viable approach to a range of visual categorization tasks, whose appeal, moreover, extends to other aspects of representation such as veridicality and biological relevance (Edelman, 1997).

References

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

Duda, R. O. and Hart, P. E. (1973). *Pattern classification and scene analysis*. Wiley, New York.

Edelman, S. (1997). Representation is representation of similarity. Behavioral and Brain Sciences, to appear.

Edelman, S., Cutzu, F., and Duvdevani-Bar, S. (1996). Similarity to reference shapes as a basis for shape representation. In Cottrell, G. W., editor, *Proceedings of 18th Annual Conf. of the Cognitive Science Society*, pages 260–265, San Diego, CA.

Poggio, T. and Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature*, 343:263–266.