MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY
and
CENTER FOR BIOLOGICAL AND COMPUTATIONAL LEARNING

A.I. Memo No. 1432 August 1993
C.B.C.L. Paper No. 81

CONDITIONS FOR VIEWPOINT DEPENDENT FACE RECOGNITION

Philippe G. Schyns & Heinrich H. Bülthoff

Abstract

Face recognition stands out as a singular case of object recognition: although most faces are very much alike, people discriminate between many different faces with outstanding efficiency. Even though little is known about the mechanisms of face recognition, viewpoint dependence, a recurrent characteristic of many research on faces, could inform algorithms and representations. Poggio and Vetter's symmetry argument [10] predicts that learning only one view of a face may be sufficient for recognition, if this view allows the computation of a symmetric, "virtual," view. More specifically, as faces are roughly bilaterally symmetric objects, learning a side-view–which always has a symmetric view–should give rise to better generalization performances than learning the frontal view. It is also predicted that among all new views, a virtual view should be best recognized. We ran two psychophysical experiments to test these predictions. Stimuli were views of 3D models of laser-scanned faces. Only shape was available for recognition; all other face cues–texture, color, hair, etc.–were removed from the stimuli. The first experiment tested whether a particular view of a face was canonical. The second experiment tested which single views of a face give rise to best generalization performances. The results were compatible with the symmetry argument: face recognition from a single view is always better when the learned view allows the computation of a symmetric view.

This report describes research done within the Center for Biological and Computational Learning in the Department of Brain and Cognitive Sciences, and at the Artificial Intelligence Laboratory. This research is sponsored by grants from the Office of Naval Research under contracts N00014-91-J-1270 and N00014-92-J-1879; by a grant from the National Science Foundation under contract ASC-9217041 (funds provided by this award include funds from DARPA provided under the HPCC program); and by a grant from the National Institutes of Health under contract NIH 2-S07-RR07047. Additional support is provided by the North Atlantic Treaty Organization, ATR Audio and Visual Perception Research Laboratories, Mitsubishi Electric Corporation, Sumitomo Metal Industries, and Siemens AG. Support for the A.I. Laboratory's artificial intelligence research is provided by ONR contract N00014-91-J-4038. Philippe Schyns is supported by an ATIPE Fellowship from the CNRS (France). Heinrich Bülhoff is now at the Max-Planck-Institut für Biologische Kybernetik, Tübingen, Germany. The authors would like to thank Anne Coombes from the Dept. of Medical Physics, University College, London, UK, for lending the face data used in the experiments.

1 Introduction

In object recognition, it is often assumed that within-class discriminations are more difficult than between-class discriminations. For example, while people would experience no difficulty to segregate a car from a tree, it would be comparatively more complex to distinguish among brands of cars or species of trees. Researchers explain this discrepancy by the nature of the comparisons involved: within-class judgments distinguish objects comparatively more similar than between-class judgments. Face recognition stands out as a notable exception to the generality of this claim. Although most faces are very much alike–they share the same overall shape, textures and other features–people discriminate between many different faces with outstanding efficiency. Face recognition is a singular case of near perfect recognition whose underlying mechanisms are of utmost interest to computer vision and psychophysics.

Even though face recognition is well documented by psychophysical and neurophysiological studies, little is known about its algorithmic and representational characteristics. Converging evidence gathered across disciplinary boundaries report a phenomenon which could inform algorithmic and representational issues: Face recognition is viewpoint dependent.

In single cell recordings studies, Perrett and his collaborators discovered cells of the macaque superior temporal sulcus (STS) which are preferentially tuned to respond to specific views of a head [3, 6, 8, 7]. Most of the cells were *viewer-centered* responding unimodally to one view (either the frontal, the two profiles or the back views); few cells were tuned to other views of the 360 degree range. Human psychophysics also reports a viewpoint preference compatible with view-based representations of faces. Among all views, the 3/4 view–the viewpoint between the full-face and the profile views–is identified faster and with greater accuracy [1, 14, 11, 4] (see [2, 12, 13] for other evidence of viewpoint depedent object recognition). The preference for a 3/4 view is naturally interpreted in light of Perrett's findings as the view which elicits the highest total activity from the profile and full-face neurons; an activation higher than the response of the individual cells to their preferred view. In summary, neurophysiological and psychological data suggest two major constraints on representations and algorithms for face recognition: 1) faces could be represented in memory with collections of few viewer-centered 2D views and 2) viewpoint dependence could be subsumed by the tuning curves of viewpoint specific cortical cells.

Does the side-view preference phenomenon reveal something substantial about face representation and recognition? Poggio and Vetter [10] showed that the recognition of a *bilaterally symmetric* object from a novel view could be achieved if only one nonsingular view of the object is known. If perception "assumes" symmetry, it could generate a symmetric "virtual" view from the only known view, or exploit equivalent information. A face is approximately bilaterally symmetric. Side-views of a face, before occlusion becomes too critical, are nonsingular views from which a symmetric view can be generated. The full-face view, however, is singular. In an RBF network [9], if units were centered on a side-view and its symmetric, together they could cover a larger range of the rotation of a face than a single unit centered on the full-face view. The aims of this paper is to test the psychophysical reality of the symmetry argument for face recognition. More precisely, we will test the following claims:

- The side-view preference results from an *interaction* of the learned view of a face and recognition of other views.
- Nonsingular views of a face–views from which a symmetric 2D view can be generated–give rise to better generalization performances than singular views.
- Virtual views are generalized better than the other novel views of a face.

2 Experiment 1

The first experiment is a simple control. If a particular view of a face is inherently more informative than any other view, it should always be preferred in recognition. Side-views which conjugate part of the shape features and part of the profile could be *canonical* [5] in this sense. To test for canonical views, we trained subjects on all views of different faces before a testing stage on all views. If all views are experienced equivalently during learning, a canonical view should give rise to better accuracy and/or faster identification performances.

2.1 Methods

The psychophysics of face recognition must control the subject's familiarity with the stimuli as well as the type of information available for the task. Features such as hair color, hairstyle, texture or color of the skin, type and size of eyebrows are invariant under rotation in depth. With familiar faces, such shortcuts could lead to the type of viewpoint invariant face recognition discussed in [14]. To control familiarity and information, all faces were unknown to subjects prior to the experiment, and faces were presented as grey-level images of 3D shape models. That is, obvious viewpoint invariant features were removed from the stimuli and we only tested shaped-based face recognition.

2.1.1 Subjects

11 subjects (age group 18-30) with normal or corrected vision, volonteered their time to participate to the experiment.

Testing view	-36	-18	0	18	36
Hit rate	.96	.91	.91	.91	.86
False alarm	.14	.18	.05	.05	.14
d′	2.83	2.26	2.98	2.98	2.16

Table 1: Hit rate, false alarm and d' for different views of the stimuli in Experiment 1.

2.1.2 Stimuli

Experiment 1 and 2 used the same set of stimuli. Stimuli were 256 grey-level views of 3D face models presented on the monitor of a Silicon Graphics workstation. There were 15 different face models; face data were laser-scanned three-dimensional coordinates of real faces. Each face was reconstructed by approximating the face data with a bicubic BSpline surface. Stimuli were views of each face at -36, -18, 0, 18, 36 degrees of rotation in depth (0 degree is the frontal view, see Figure 1). Faces were illuminated by a point light source located at the observer, shaded with a Gouraud shading model.

2.1.3 Procedure

The experiment was decomposed into ten blocks. A block consisted of a learning stage and a testing stage. In the learning stage, subjects had to learn a particular face (the target face). The target face rotated on the screen, once clockwise, once counterclockwise–or vice versa, depending on a random selection. The apparent rotation was produced by showing the five views of the target face in rapid succession (100 ms/view, for a total of 1 sec/face). The learning stage was immediately followed by a testing stage. Test items were two views in the same orientation, presented one at a time–orientations were selected randomly. One view was a view of the target face, and the other, a view of the distractor face. For each view, subjects had to indicate whether or not it was a view of the target face by pressing the appropriate response-key on the computer keyboard. The experiment was completed after 10 blocks as just described. A different target face was associated with each block. Each of the 5 viewpoints was tested twice, each time with a different target.

2.2 Results and Discussion

To test for a viewpoint preference in recognition, we compared the mean percentage of correct recognition of the target in the 5 testing conditions. A one-way ANOVA revealed no significant effect of viewpoint ($F(4,40) = .31, p = .87$, ns.). Table 1 shows the hit rate, false alarm rate and d' for the identification of the stimuli in Experiment 1.

Although subjects responded almost equivalently well to all views, it could still be argued that some views are correctly identified faster than others. A one-way ANOVA showed no effect of viewpoint ($F(4,40) = .78, p = .54$, ns.) on reaction time for correct identification. Average reaction time across all views was 811 ms. These results suggest that there is no viewpoint preference in face recognition when all views are experienced during learning. Thus, viewpoint dependent face recognition cannot simply be attributed to a recognition preference for certain views over others.

3 Experiment 2

The results of Experiment 1 showed that no view is canonical. Poggio and Vetter's symmetry argument predicts that viewpoint preference could arise from an interaction between the view learned (whether it is a singular or a nonsingular view) and the recognition stage. The aim of the second experiment is to test this prediction and to understand further the nature of the interaction. In a learning stage, distinct groups of subjects learned each a different view of the faces. All subjects were then tested on all views of the faces. We expected differences in performance between subjects who were in the singular view group from those who were in the other groups.

3.1 Methods

3.1.1 Subjects

30 subjects volunteered their time to participated to Experiment 2. They were randomly assigned to condition.

3.1.2 Stimuli

Stimuli were identical to those of Experiment 1: 5 views of 15 face models.

3.1.3 Procedure

Subjects were randomly assigned to one of five training condition: the -36, -18, 0, 18, or 36 degree view. For example, subjects in group -36 only saw one view of a target face during learning: the -36 view (see Figure 1). The procedure of Experiment 2 was very similar to the one of Experiment 1. The experiment was segmented into 10 blocks. A block was composed of a learning and a testing stage. Here, however, subjects learned only one view of the target face. The view was presented for 1 second, immediately followed by a testing stage. The testing stage also consisted of two successive views in the same orientation: one view of the testing face and a view of a distractor face. In 2 out of the 10 testing blocks, the testing view was the same as the learning view. The remaining 4 pairs of 2 blocks were each assign to a different testing view. A different target was associated with each block, and each possible viewpoint was tested twice, each time with a different target. During the experiment, subjects only saw one view of a particular face in learning, and only one testing view of the same face. With this design, we could test how changing the learning view affected recognition performances.

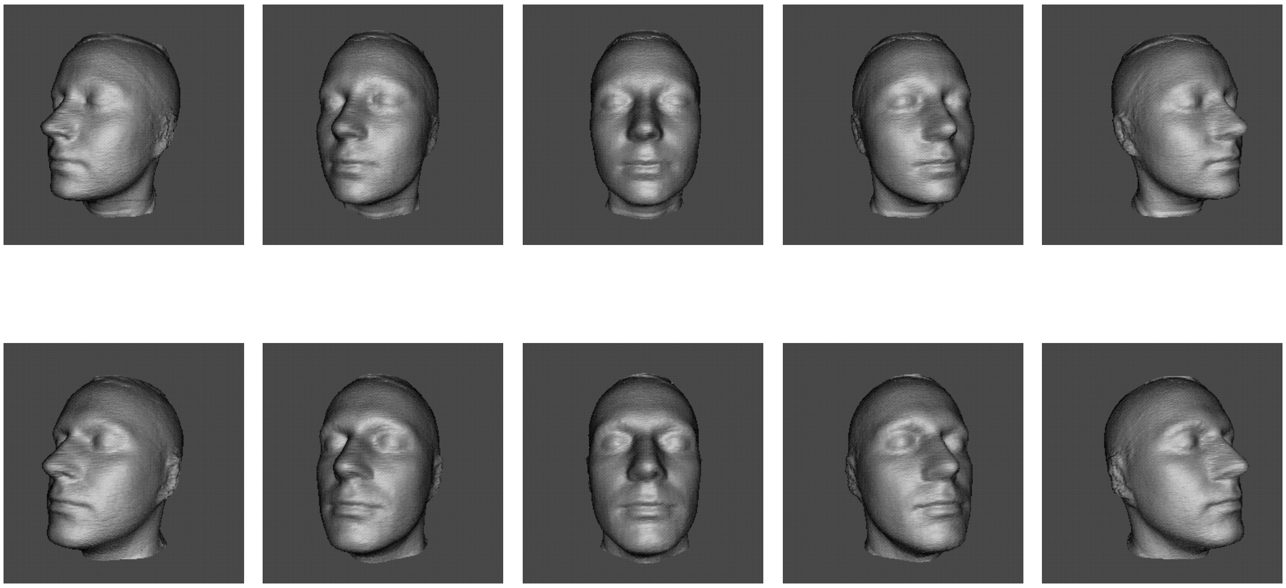

Figure 1: This figure illustrates the stimuli used in Experiment 1 and 2. The top pictures and the bottom pictures represent two different faces. From the left to the right, both sets of pictures show the -36, -18, 0, 18, and 36 views used in the experiments. The views were computed from 3D face models reconstructed by approximating laser-scanned 3D coordinates of real faces with a BSpline surface. All textural, color, and hair cues were removed from the stimuli. A point light source located at the observer illuminated the Gouraud shaded surface of the faces.

Learning view	-36	-18	0	18	36
Hit rate	.86	.78	.56	.72	.78
False alarm	.12	.18	.19	.18	.06
d'	2.26	1.68	1.03	1.5	2.26

Table 2: Hit rate, false alarm and d' for different views of the stimuli in Experiment 2.

3.2 Results and Discussion

A two-way ANOVA was run to test for a dependence between the view learned and recognition performances as measured by percent correct recognition of the target face. The results showed a main effect of learning view ($F(4,25) = 4.17, p = .01$), no main effect of testing view ($F(4,16) = 1.87, p = .13$, ns.) and a significant interaction of learning view and testing view ($F(16,100) = 2.03, p = .017$). The absence of significant effect of testing view comes as no surprise. As shown in Experiment 1, no single view, by itself, stands out in recognition. Table 2 illustrates the overall recognition performances as a function of learning view.

The data reveal a strong interaction between the learned view of a face and generalization to other views of the same face. To understand further this interaction, we contrasted recognition performance in learning condition 0 (the frontal view) to all other learning conditions. The contrast revealed a significant difference in recognition performances between condition 0 and all the other learning conditions ($F(1,1) = 14.55, p < .001$) and this comparaison also interacted with the testing views ($F(1,4) = 4.28, p < .01$). A second orthogonal test showed no significant difference between learning conditions -18 and 18 contrasted to -36 and 36 ($F(1,1) = 1.06, p = .31$). Figure 2 illustrates the interaction.

In Figure 2, the hit rate to the different testing views as a function of learning condition reveals an interesting trend. The symmetry argument argument predicts that a U shaped generalization curve should describe the response profiles to the different testing views. The peaks of the curve, should be roughly located on the learned view, and on its symmetric view. Although further evidence are required to confirm the trend, a generalization curve of this form characterizes the group which learned the 36 view.

An inverted U shape characterizes bad generalization performances– a sharp decrease of performance with increasing rotation in depth from the learned view. Such a response profile distinguishes subjects who learned the 0–full-face–view. Since the full-face view is singular, a second view could not be computed from the frontal view. New views of the faces were not recognized with high accuracy. The intermediary group (the 18 group) displays a response profile in-between the two extremes.

To summarize, these experiments on face recognition

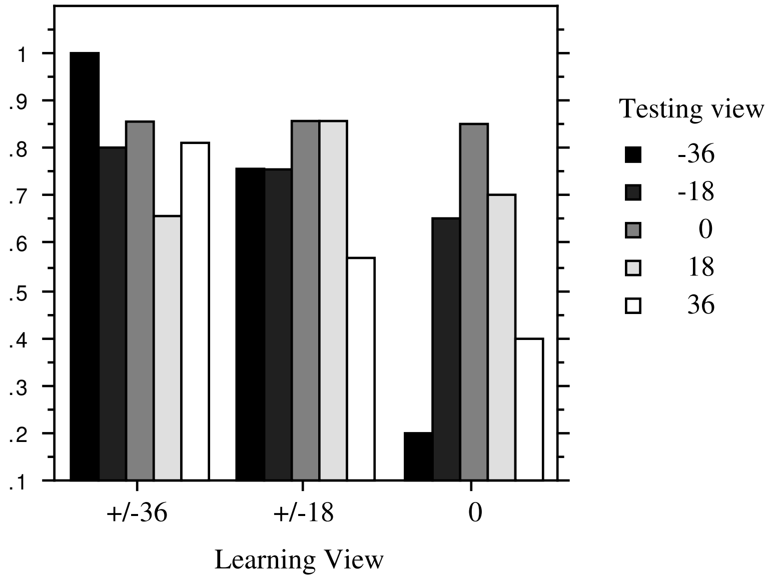

Figure 2: This figure illustrates the results of Experiment 2. The different learning conditions are grouped as a function of degrees of rotation from the full-face view. The histograms illustrate the hit rate to the different testing views. As predicted by the symmetry argument, the inverted U curve indictating poor generalization performances for the singular full-face view tends to turn into a U shaped generalization curve as the degree of rotation of the learned view increases.

are compatible with the predictions of the symmetry argument. Experiment 1 showed that no single view was canonical. The second experiment showed that face recognition could be achieved from a non-singular view. These data suggest that a side-view should be preferred over a full-face view because a side-view allows better face encoding and recognition.

References

[1] V. Bruce, P. Healey, M. Burton, and T. Doyle. Recognising facial surfaces. *Perception*, 20:755–769, 1991.

[2] H. H. Bülthoff and S. Edelman. Psychophysical support for a 2D interpolation theory of object recognition. *Proceedings of the National Academy of Science*, 89:60–64, 1992.

[3] M. H. Harries and D. I. Perrett. Visual processing of faces un temporal cortex: Physiological evidence for a modular organization and possible anatomical correlates. *Cognitive neuroscience*, 3(1):9–24, 1991.

[4] F. L. Krouse. Effect of pose, pose change, and delay on face recognition performance. *Journal of Applied Psychology*, 66:651–654, 1981.

[5] S. E. Palmer, E. Rosch, and P. Chase. Canonical perspective and the perception of objects. In J. Long and A. Baddeley, editors, *Attention and Performance IX*, pages 135–151. Erlbaum, Hillsdale, NJ, 1981.

[6] D. I. Perrett, A. J. Mistlin, and A. J. Chitty. Visual neurones responsive to faces. *Trends in Neurosciences*, 10:358–364, 1989.

[7] D. I. Perrett, M. W. Oram, M. H. Harries, R. Bevan, P. J. Benson, and S. Thomas. Viewer-centered and object-centered coding of heads in the macaque temporal cortex. *Experimental Brain Research*, 86:159–173, 1991.

[8] D. I. Perrett, P. A. J. Smith, D. D. Potter, A. J. Milstin, A. S. Head, A. D. Milner, and M. A. Jeeves. Visual cells in the temproal cortex sensitive to face view and gaze direction. *Proceedings of the Royal Society of London*, B223:293–317, 1985.

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

[10] T. Poggio and T. Vetter. Recognition and structure from one 2d model view: Observations on prototypes, object classes and symmetries. A.I. Memo No. 1347, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1992.

[11] A. Baddeley R. H. Logie and Muriel M. Woodhead. Face recognition, pose and ecological validity. *Applied Cognitive Psychology*, 1:53–69, 1987.

[12] I. Rock and J. DiVita. A case of viewer-centered object perception. *Cognitive Psychology*, 19:280–293, 1987.

[13] M. Tarr and S. Pinker. Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, 21:233–282, 1989.

[14] T. Valentine V. Bruce and A. Baddeley. The basis for the 3/4 view advantage in face recognition. *Applied Cognitive Psychology*, 1:109–120, 1987.