Viewpoint dependence and face recognition

Philippe G. Schyns
Dept. of Brain and Cognitive Sciences
Massachusetts Institute of Technology
Cambridge, MA 02139
schyns@ai.mit.edu

Heinrich H. Bülthoff
Dept. of Brain and Cognitive Sciences
Massachusetts Institute of Technology
Cambridge, MA 02139
hhb@ai.mit.edu

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 research in face recognition — could help to understand algorithmic and representational issues. The current research tests whether learning only one view of a face could be sufficient to generalize recognition to other views of the same face. Computational and psychophysical research (Poggio & Vetter, 1992) showed that learning one view of a bilaterally symmetric object could be sufficient for its recognition, if this view allows the computation of a symmetric, "virtual," view. Faces are roughly bilaterally symmetric objects. Learning a side-view — which always has a symmetric view — should allow for better generalization performances than learning the frontal view. Two psychophysical experiments tested these predictions. Stimuli were views of shaded 3D models of laser-scanned faces. 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 theoretical predictions of Poggio and Vetter (1992): learning a side view allows better generalization performances than learning the frontal view.

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 internal 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 the vision community.

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. To illustrate, the single cell recordings studies of Perrett and his collaborators reported cells of the macaque superior temporal sulcus (STS) which were preferentially tuned to respond to specific views of a head (Harris & Perrett, 1991; Perrett, Mistlin & Chitty, 1989). 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. The preference for a 3/4 view — the viewpoint between the full-face and the profile view — is naturally interpreted in light of these findings as the view which elicits the highest total activity from the profile and full-face cells; an activation higher than the response of the individual cells to their preferred view.

Human psychophysics also reports a viewpoint preference compatible with view-based representations of faces. Among all views, the 3/4 view is identified faster and with greater accuracy (Bruce, Healey, Burton & Doyle, 1991; Bruce, Valentine & Baddeley, 1987; Logie, Baddeley & Woodhead, 1987, Krouse, 1981.) Researchers often invoke the higher informativity of the 3/4 view to explain the preference: The 3/4 view conjugates many shape features of a face with part of the profile.

In summary, although neurophysiological and psychological data diverge on interpretation, they converge on evidence: face recognition is viewpoint dependent. Viewpoint dependency is not particular to face stimuli, it has been reported in several studies on object recognition (see, for example, Bülthoff & Edelman, 1992; Rock & Di Vita, 1987; Tarr & Pinker, 1989.) These studies suggest 1) that objects could be represented in memory with collections of few viewer-centered 2D views and 2) that viewpoint dependence could be subsumed by the tuning curves of viewpoint-specific units in artificial and natural networks.

Poggio and Vetter (1992) showed that the recognition of a *bilaterally symmetric* object from a novel view could be achieved if *only one* non-singular 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 non-singular views from which a symmetric view can be generated. The full-face view, however, is singular. If units of a network were centered on a side-view and at the corresponding symmetric virtual view, together they could cover a larger range of the rotation of a face than a single unit centered on the full-face view. The goal of this paper is to test psychophysically if the human visual systems exploits

such information. More precisely, we will test the following implications of the virtual views idea of Poggio and Vetter:

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

Experiment 1

The first experiment is a simple control testing for possible canonical views of a face. If a particular view of a face is inherently more informative than any other view, it should be preferred in recognition. Side-views which conjugate part of the shape features and part of the profile could be *canonical* (Palmer, Rosch, & Chase, 81) in this sense. To test for canonical views, we measured recognition performances in conditions of *ideal* face learning. The conditions were ideal because subjects were equivalently exposed to each possible view of the face they could be later tested on. In these conditions of normalized learning, a canonical view should give rise to better accuracy and/or faster identification performance.

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 small rotations in depth. With familiar faces, such "easy features" could provide the shortcuts responsible of viewpoint invariant face recognition discussed in (Bruce et al, 1987). 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 shape-based face recognition.

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

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 B-spline 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). Illumination was simulated by a point light source located at the observer and a Gouraud shading model.

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 randomly selected. 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. A different target face was associated with each block. Each of the 5 viewpoints was tested twice, each time with a different target.

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.

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.

Experiment 2

The results of Experiment 1 showed that no view was preferred over others for the task of recognition. The symmetry argument predicts that viewpoint preference could arise from an interaction between the view learned (whether it is a singular or a non-singular 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 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.

Methods

Subjects 30 subjects volunteered their time to participated to Experiment 2. They were randomly assigned to condition with the constraint that 10 subjects be assigned to the singular view learning condition and the remaining subjects be equivalently distributed in the remaining 4 conditions.

Figure 1: This figure illustrates the stimuli used in Experiment 1 and 2. From the left to the right, the pictures show the -36, -18, 0, 18, and 36 degree 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 bicubic B-spline 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.

Stimuli Stimuli were identical to those of Experiment 1: 5 views of 15 face models for a total of 45 stimuli.

Procedure Subjects were randomly assigned to one of five training conditions: the -36, -18, 0, 18, or 36 degree view. For example, subjects in group -36 only saw one view of a target face during training: the -36 degree 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 training and a testing stage. Here, however, subjects learned only one view of the target face. This view was presented for 1 second, immediately followed by a testing stage which also consisted of two successive view: one view of the target face and a view of a distractor face (both in the same orientation). In two out of the ten blocks, the testing view was the same as the training view. The remaining four pairs of two blocks were each assigned to a different testing view. A different target face was associated with each block, and each possible viewpoint was tested twice, each time with a different target. That is, subjects only saw one view of a particular face during training, and only one testing view of the same face during the experiment. With this design, we could test how the generalization performance to novel views depends on the particular orientation of the training view.

Results and Discussion

To analyze the data, we collapsed learning conditions according to their excentricity from the full-face view. That is, the -36 and 36 groups were pooled together, as were the -18 and 18 group. We then had a total of three groups, with ten subjects per group.

A two-way ANOVA was run to test for a dependence between the training view and recognition performances as measured by percent correct recognition of the target face. The results showed a main effect of training 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 training view and testing view ($F(16,100) = 2.03, p = .017$). The absence of a significant effect of testing view is not surprising. As already demonstrated in Experiment 1, no single view, by itself, stands out in recognition, so we did not expect a main effect of testing view here. Table 2 illustrates the overall recognition performances as a function of training view. The

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.

data reveal a strong interaction between the learned view of a face and generalization to other views of the same face. To elucidate further this interaction, we contrasted recognition performance in training condition 0 (the frontal view) to all other training conditions. The contrast revealed a significant difference in recognition performances between condition 0 and all the other training conditions ($F(1,1) = 14.55, p < .001$) and this comparison also interacted with the testing views ($F(1,4) = 4.28, p < .01$). A second orthogonal test showed no significant difference between training 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 for the different testing views as a function of training condition reveals an interesting trend. The symmetry 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 at the training view, and on the corresponding symmetric virtual view. Although further evidence is required to confirm the trend, a generalization curve of this form characterizes the group which learned the 36 degree 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 characterizes subjects who learned the 0 degree view (full-face). Since the full-face view is singular, it does not contain enough information for the generalization to novel views of the face. The intermediary group (the 18 degree group) displays a response profile in-between the two extremes.

To summarize, these experiments on face recognition are compatible with the predictions of the symmetry argument. Experiment 1 showed that no single view was canonical.

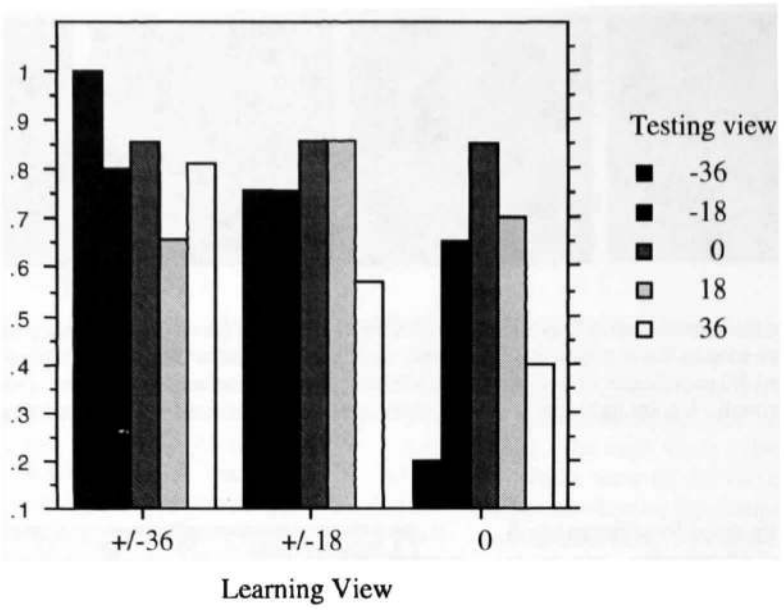

Figure 2: This figure illustrates the results of Experiment 2. The different learning conditions are grouped as a function of degrees of rotation from the full-face view. The histograms illustrate the hit rate for the different testing views. As predicted by the symmetry argument, the inverted U curve 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.

The second experiment showed that face recognition could be achieved over a larger viewpoint range, if a a non-singular view was used for training. 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.

Our results, however, concern only shape-based face recognition which by no means tests face recognition in an ecologically valid condition. Easy features such as hairdo, eyebrows and other typical visual cues of faces are invariant under small rotations in depth. The scope of our results should therefore be limited to the processes we are testing — namely, shape-based face recognition in conditions of constant illumination.

However, to the extend that our results reproduce under specific learning conditions the well-known phenomenon of viewpoint dependency observed with real face pictures (Krouse, 81), one can question the impact of viewpoint independent features for the recognition of real faces. Our first experiment shows that shape-based face recognition might be viewpoint independent under ideal learning conditions. Further research might be necessary to analyze the generality of this claim in ecologically valid situations of recognition; situations in which people must discriminate a face from many other faces they know, in various illumination conditions, and with (or without) "easy features."

References

Bruce, V., Healey, P., Burton, M. & Doyle, T. (1991). Recognizing facial surfaces. *Perception*, 20, 755-769.

Bruce, V., Valentine, T & Baddeley, A. (1987). The basis for the 3/4 view advantage in face recognition. *Applied Cognitive Psychology*. 1, 109-120.

Bülthoff, H.H. & Edelman, S. (1992). Psychophysical support for a $2-D$ view interpolation theory of object recognition. *Proceedings of the Royal Academy of Science*, 89, 60-64.

Harris, M.H. & Perrett, D.I (1991). Visual processing of face in temporal cortex: Physiological evidence for a modular organization and possible anatomical correlates. *Cognitive Neuroscience*, 3(1), 9-24.

Krouse F.L. (1981). Effect of pose, pose change,and delay on face recognition performances. *Journal of Applied Psychology*, 66, 651-654.

Logie, RH., Baddeley, A. & Woodhead M.M. (1987). Face recognition, pose and ecological validity. *Applied Cognitive Psychology*, 1, 53-69.

Palmer, S. E., Rosch, E. and Chase, P. (1981). Canonical perspective and the perception of objects. In J. Long and A. Baddeley (Ed.) *Attention and Performances*, 10, 135-151. Hilldsale, NJ.

Perrett, D.I., Mistlin, A.J. & Chitty, A. J. (1989). Visual neurones responsive to faces. *Trends in Neuroscience*, 10, 358-364.

Poggio, T. & Vetter, T. (1992). Recognition and structure from one 2D model view: Observations on prototypes, object classes and symmetries. MIT AI Memo, 1347. Cambridge: Massachussetts Institute of Technology.

Rock I., & Di Vita, J. (1987). A case of viewer-centered object perception. *Cognitive Psychology*, 19, 280-293.

Tarr, M., & Pinker, S. (1990). When does human object recognition use a viewer-centered reference frame? *Psychological Science*, 1, 253-256.

Acknowledgments

The authors would like to thank Tomaso Poggio and Thomas Vetter for fruitful discussions, and Annes Coombes from the Dept. of Medical Physics, University College, London, UK, for lending the face data used in the experiments. 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 now at the Dept. of Psychology, University of Montreal, Canada. Heinrich Bülthoff is now at the Max-Planck-Institut für Biologische Kybernetik, Tübingen, Germany.