Generalization to Novel Images in Upright and Inverted Faces

Yael Moses   Shimon Ullman
Shimon Edelman
Department of Applied Mathematics and Computer Science,
The Weizmann Institute of Science,
Rehovot 76100,
Israel.

November 29, 1994

Abstract

An image of a face depends not only on its shape, but also on the viewpoint, illumination conditions, and facial expression. A face recognition system must overcome the changes in face appearance induced by these factors. This paper investigate two related questions: the capacity of the human visual system to generalize the recognition of faces to novel images, and the level at which this generalization occurs. We approach this problems by comparing the identification and generalization capacity for upright and inverted faces. For upright faces, we found remarkably good generalization to novel conditions. For inverted faces, the generalization to novel views was significantly worse for both new illumination and viewpoint, although the performance on the training images was similar to the upright condition.

Our results indicate that at least some of the processes that support generalization across viewpoint and illumination are neither universal (because subjects did not generalize as easily for inverted faces as for upright ones), nor strictly object-specific (because in upright faces nearly perfect generalization was possible from a single view, by itself insufficient for building a complete object-specific model). We propose that generalization in face recognition occurs at an intermediate level that is applicable to a class of objects, and that at this level upright and inverted faces initially constitute distinct object classes.

1 Introduction

The human visual system can easily recognize the identity of a familiar face from its image. However, recognizing faces is a difficult problem from a computational point of view, because all faces have a generally similar shape and at the same time different images of the same face can vary considerably. There are several sources for the variations between images of the same face, such as changes of facial expression, age, viewpoint, illumination, noise, etc. The task of a face recognition system, whether natural or artificial, is to recognize a face in a manner that is insensitive to these variations. The basic issue we study here is how the human visual system can identify a face in novel images, in particular under changes of illumination direction and viewpoint.

We consider two aspects of the problem. The first is how well humans in fact recognize faces in novel images. The second is the level at which the generalization of face identification to novel images takes place. Recognition systems can use different types of information for overcoming image variations. We distinguish between three basic generalization levels: *universal, class based,* and *object specific.* Roughly speaking, at the universal level, no restrictive assumptions are made about the objects that may appear in the image. At the object-specific level knowledge about a specific face may be used (e.g., the three dimensional shape of a specific face); the class-based level is an intermediate level, where knowledge about faces in general (the class to which a candidate object belongs), but not about a specific face may be used (e.g., the general shape of faces). These generalization levels will be discussed further in the last section of the paper.

Face processing was previous studied by comparing the recognition of faces to other object classes such as houses, landscapes, and dogs (Valentine, 1988). One of the main problems with such a comparison is the difference in complexity between faces and other objects. To overcome this problem and to characterize the computational level at which the generalization of face recognition takes place, we compared the recognition of upright and inverted faces. From an objective standpoint, they have the same complexity, however, perceptually inverted faces are more difficult to recognize. We did not compare the difficulty of recognition *per se.* Instead, we studied the differences in the generalization of recognition from one image of a face to other images of the same face in the same orientation (upright/inverted).

It should be noted here, that one objective difference that often occurs between upright and inverted faces is that under natural conditions upright faces are illuminated from above while the faces in inverted images are illuminated from below. Since objects are often illuminated from above, Johnson *et al.* (1992) suggested this difference as a source for difficulty in recognizing inverted faces. We avoided this problem in our study by placing the illuminating at the face level. Consequencely the faces were illuminated from the same directions in the upright and inverted images.

In the experiments, subjects first learned to recognize single image of each of three distinct unfamiliar faces. Then, each subject was tested with 20 different images of each of the three faces, taken under novel illumination conditions and from novel viewpoints. The same experiment was repeated for inverted images of other faces. In this case, the subject learned to recognize images of inverted faces and was then tested on novel images of inverted faces.

We found that the recognition of novel views of upright faces was remarkably good (see section 2.2.2). In contrast, the performance on novel images of inverted faces was significantly worse than for the upright faces (although the subjects had no problem in recognizing the training images of inverted faces). Our results indicate that at least some of the processes that support generalization across viewpoint and illumination are neither universal nor strictly object-specific. They are not universal because subjects did not generalize as easily for inverted faces as for upright ones. They are not object-specific because in upright faces nearly perfect generalization was possible from a single view, by itself insufficient for building a complete object-specific model. We propose that generalization in face recognition occurs at an intermediate level that is applicable to a class of objects, in this case, a class of upright faces. A discussion of these conclusions is presented in section 3.

Before describing the experiments in details, we briefly review previous studies of face recognition from novel views and from inverted images, and relate them to our study. Several studies have addressed the problem of generalization of face memory to novel images taken from new viewpoints, but without changing the illumination conditions (Patterson and Baddeley, 1977; Davies et al., 1978; Bruce, 1982). In these experiments, a set of faces (unfamiliar or familiar) was briefly presented once to the subject during a training phase. In the testing phase, the subject determined whether a given face had appeared previously in the training phase. Two viewpoints were used: frontal, and 3/4 profile. The results showed that the recognition of a previously seen face in the novel view was reliable. Bruce (1982) compared the memory of familiar and unfamiliar faces in such an experiment, and found that familiar faces were recognized faster and more accurately than unfamiliar faces. Our experiments were different in several respects. First, our subjects were tested on face identification, in a three-alternative forced-choice setup. Second, our subjects were initially unfamiliar with the face stimuli, but one image of each face was made familiar by repeated exposure during the training stage (this can explain the differences between our results and those of Bruce (1982) regarding the recognition of unfamiliar faces). Third, the set of images that we considered for each face was larger (20 compared to two). Fourth, the images in our experiments varied not only due to pose, but also due to illumination. Finally, our set of images was precisely controlled, so that each parameter (e.g., viewpoint or illumination) was varied independently of the others, while images of all faces were normalized to the same size and location. In previous studies (Patterson and Baddeley, 1977; Davies et al., 1978; Bruce, 1982) the control over viewpoint, location, size, illumination, background, and in many cases familiarity of the faces to the subject, was not completely specified.

Recognition of inverted faces is known to be a difficult perceptual task (Köhler, 1947; Hochberg and Galper, 1967; Attneave, 1967; Yin, 1969; Scapinello and Yarmey, 1970; Yarmey, 1971; Carey and Diamond, 1977; Valentine and Bruce, 1986). A review of research concerned with the recognition of inverted faces can be found in Valentine (1988). Generally, the memory for faces was shown to be impaired when inverted images were involved (in the training or in the testing phases or in both). Inverted faces were also used in attempts to discover out whether features or configuration information are required for face recognition. Carey and Diamond (1977) proposed that the difficulty in recognizing inverted faces results from an inability to access the configuration information of the facial features from inverted faces. The cue saliency in artificial inverted faces (schematic or thresholded to black and white) was addressed by several investigators (Endo, 1982; Endo, 1986; Kemp et al., 1990). The results of these experiments indicated certain differences in the memory for upright and inverted faces. In our study, the relative difficulties of recognizing inverted faces was not of primary concern. Inverted faces were used to study certain aspects of the effects of illumination and viewpoint on face identification.

2 The experiments

The basic experimental paradigm was a three-alternative forced-choice recognition task. The subject was first trained on a single image of each of three faces taken under identical viewing conditions. She or he was then tested on 20 images of each of the three faces, taken under all combinations of four different illumination positions and five different camera locations. Our main objective was to test the degree of generalization to new viewpoint and illumination for both upright and inverted face images. The locations of the camera and the light sources were identical for all faces. The same experiment was repeated for several sessions with a number of different triplets of faces. Some of the triplets were shown always upright, others always inverted. The assignment of orientation and triplet identity was balanced across subjects, so that the faces in each triplet were seen upright by half of the subjects, inverted by the other half. The orientation of the stimuli was fixed throughout an experimental session.

2.1 Method

2.1.1 Subjects

Eight subjects (3 females, 5 males, ages 16-35) participated in the experiment. All had normal or corrected to normal acuity, and all but one were paid for their participation. All subjects had some prior experience in psychophysical experiments.

2.1.2 Materials

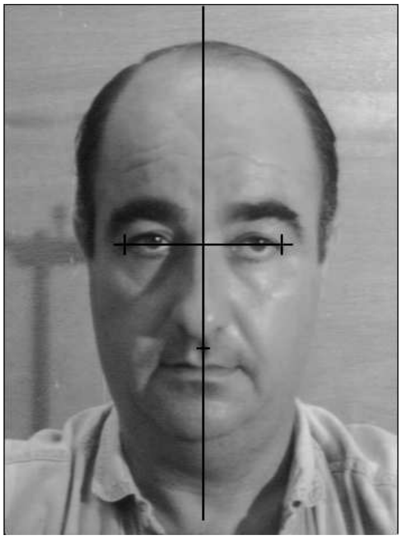


Figure 1: Each face was normalized before taking the picture so that the face's symmetry axis, the external corners of the eyes, and the bottom of the nose located on the reference lines as shown.

Twenty images (Figure 2) of each of 18 different faces were used as stimuli. All faces were of males without distinctive features (no glasses, beard, mustache, etc.). All images were taken by the same camera under precisely controlled illumination and viewpoint. The frontal view of all faces were normalized by fixing the location of the face's symmetry axis, the external corners of the eyes, and the bottom of the nose, before taking the pictures (see Figure 1).[1] The camera (Pulnix TM-560 with Canon lens $V6 \times 1616 - 100\text{mm}\ F1 : 1.9$) was attached to a robot arm (Adept I) to control its precise position. A Symbolics Lisp Machine controlled the camera positioning to: $-34°$, $-17°$, $0°$, $17°$ and $34°$ with respect to the frontal view, in the horizontal plane. The distance of the face from the camera was

[1] The position of three points on an image of a three dimensional object determines uniquely the location of the object with respect to the camera.

fixed at $110 cm$. Four distinct illumination conditions at the same height of the face were created by turning on and off three fixed light sources: left (IL=0), center (IL=1), right (IL=2) and the combination of left and right (IL=3). The subjects were asked to assume a neutral expression and to remain still. To reduce the influence of the background, the faces during the experiments were clipped by an elliptical mask that occluded most of the hair and the neck areas. Each image consisted of 512 × 352 points, 8 bits per point.

The subjects viewed the images on the screen of a Silicon Graphics Personal Iris 4D35/TG workstation, at an approximate distance of $50cm$. At that distance, an image subtended approximately 6.8 degrees of visual angle.

The 18 different faces were divided into six different sets, three faces in each set. Each set consisted of the 20 images of each of the three faces. The triplets were chosen that the faces within a set were judges by the experimenters to be similar to one another. The sets were denoted by the letters A through E (see Figures 3).

2.1.3 Procedure

An experimental session started with a training phase following by a testing phase. Any given session involved one fixed set of faces. In the training phase the subject was trained to recognize a single image of each of the three faces of the set. The same view of the face (VP=17°, and IL=0) was used for all the images of the training phase (as in Figure 3). The subject was shown repeatedly each of the three images in a pseudo-random order. In the first 15 trials of the training phase, the subject was given a graphical indication of the correct response. This indication was provided by a diagram at the bottom of the screen, showing the correct response buttons ("1", "2", or "3" a numeric). Subsequently, the indication was provided (and an audible signal was given) only if the subject's response was erroneous. Once the subject identified correctly 18 out of the last 20 appearances of the training image of each face, a special signal was sounded and the testing phase started.

In the testing phase, the subject was tested on all images of the three faces of the set. For each face, the test image included all combinations of the four different illumination locations and five different camera positions. Altogether, there were 20 × 3 different images. Each image was presented six times in the testing phase. In each trial of the experiment, the stimulus image was shown for $600msec$ followed by a mask (a jumbled face image) that remained visible until the subject responded. The subjects were required to make a three-alternative forced-choice decision regarding the identity of the displayed face image. The subjects were forewarned that different images of the same face could appear in the testing phase, and that no feedback would be given for incorrect responses. They were asked to respond as quickly and as accurately as possible. An experimental

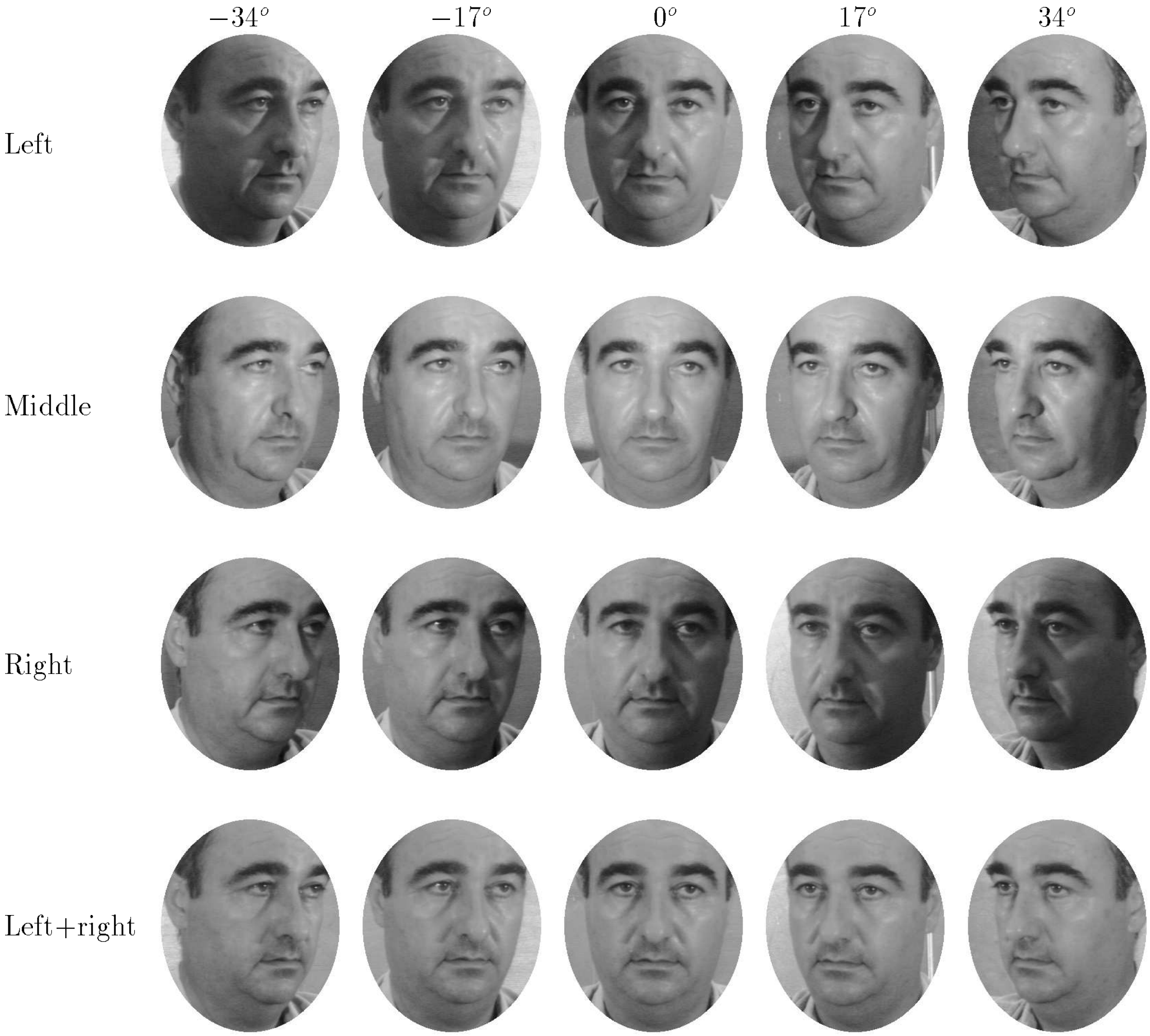


Figure 2: An example of 20 images of one of the faces (all combinations of five different viewing position and four different illumination).

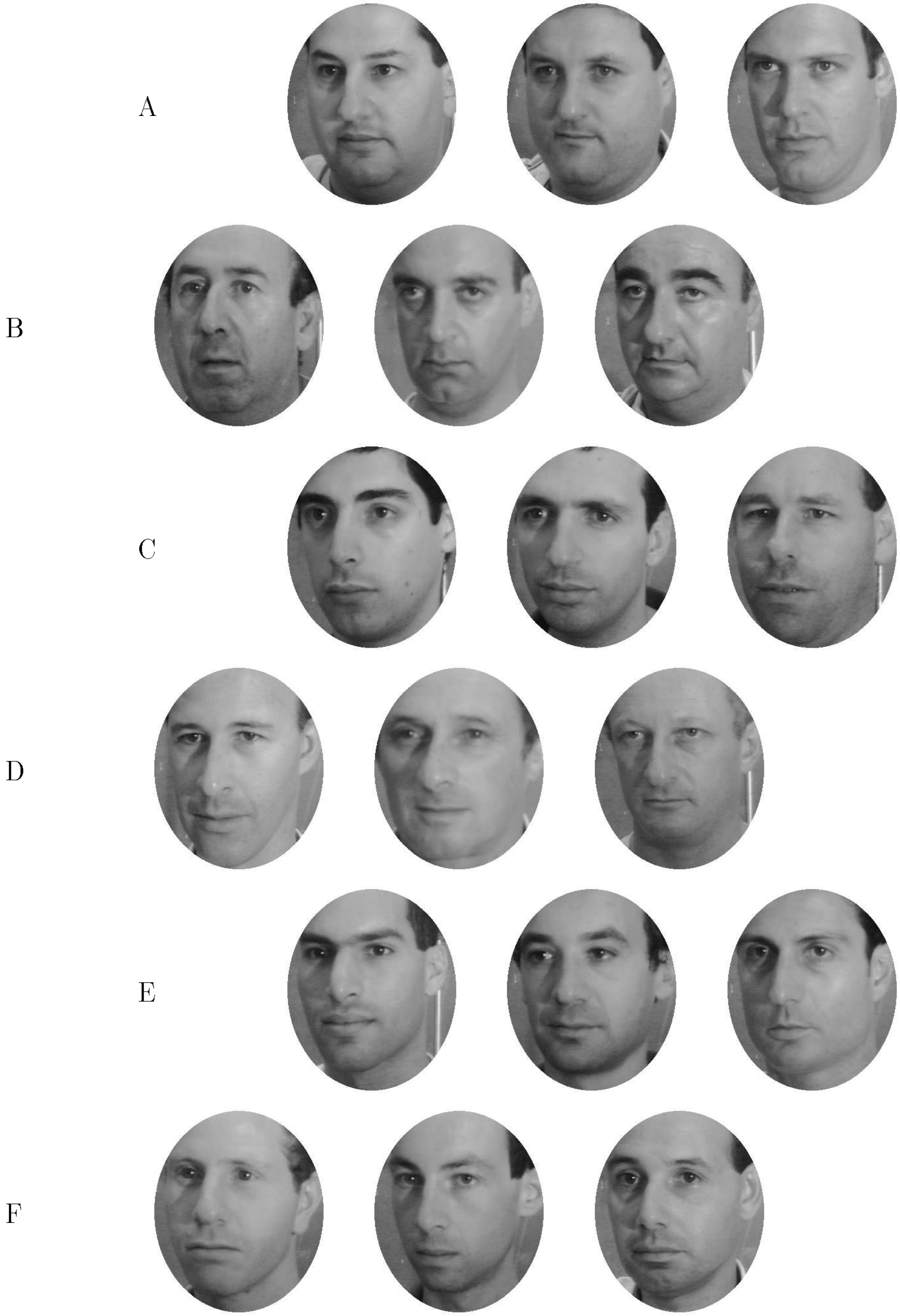


Figure 3: The triplets of images used in the experiment, one image of each face. These images were used in the training phase (VP=17 and IL =0).

session included a minimum of 105 training trials (or as many as were necessary for the subject to achieve a 90% correct rate on each face),[2] followed by 360 testing trials (5 levels of VP × 4 levels of IL × 3 faces × 6 replications).

The experiment consisted of four sessions of six sets (in all, 24 sessions) for each of the eight subjects. In a given session, the faces were either all upright or all inverted in both the learning and the testing phases. For a given subject, half of the face sets were upright images, the other half inverted. For the first four subjects, the sets C, B, E were always upright, sets A, D, F always inverted. The other four subjects saw sets C, B, and E inverted, and sets A, D, F, upright. The assignment of the set variable to a given subject and session was done in such a manner that Subject/Set combinations tended to occur in pairs (that is, the same set was shown to a given subject during two successive sessions). This made it easier for the subjects to remember what the target faces in a given session were. Otherwise, this assignment was randomized across subjects.

2.2 Results

We first present the summary statistics of the data (section 2.2.1). We then analyze separately the generalization across viewpoint and illumination in the first exposure of a subject to a set of upright and inverted faces (section 2.2.2). Finally, we analyze two learning effects: improvement in generalization within and across sets (section 2.2.3).

2.2.1 Preliminary analysis

Altogether, the experiment yielded nearly $70,000$ responses. The data from a session were included in the subsequent analysis if the following criterion was satisfied: the subject had to identify correctly 5 out of 6 appearances of each of the three training images in the testing phase (VP=17°, and IL=0). This criterion was satisfied in 86 out of 96 upright sessions and in 76 out of 96 inverted sessions. The other sessions were omitted from the analysis since we were interested in the generalization process of recognizing correctly novel face images after learning to recognize a single image of the same face. In sessions were this criterion was not satisfied, the subjects were unable to recognize well enough the training image. A separate analysis of the sessions that were dropped here, showed the same effects of generalization as will be described below. We also discarded records of trials in which response times were shorter than $250msec$ or longer than $3sec$ (these constituted 1.5% of the total number of trials). The final data set included $57,976$

[2] If the subject did not reach this level of performance in 250 trials, the session was aborted, and was restarted from the beginning after a short break. This happened in 6 sessions, or about 3% of the total number of sessions.

responses (about 84% of the original volume of data; we discarded all sessions of one subject due to his generally poor performance, that was statistically different from the rest of the subjects).

Pass	*Statistic*	Up/Inv	Mean	Train	VP diff.	IL diff.	VP&IL diff.
1	CR, %	upright	97.3± 0.2	99.1± 0.5	97.0± 0.5	97.8± 0.5	97.1± 0.3
		inverted	87.2± 0.6	98.6± 0.7	86.5± 1.5	90.1± 1.6	85.9± 0.8
	RT, *ms*	upright	904± 6	860± 22	916± 13	900± 15	905± 8
		inverted	1034± 7	940± 25	1066± 17	1000± 15	1040± 9
4	CR, %	upright	97.5± 0.2	97.4± 0.8	97.7± 0.4	96.5± 0.7	97.6± 0.2
		inverted	94.6± 0.4	99.1± 0.5	95.0± 0.9	95.0± 1.0	94.1± 0.6
	RT, *ms*	upright	832± 6	812± 27	823± 12	825± 16	834± 8
		inverted	909± 6	862± 21	916± 12	887± 14	916± 8

Table 1: Means and standard errors of the mean of correct rate (CR), and response time (RT) averaged over all subjects and first pass of all sets (pass-number = 1) and the last pass of all sets (passs-number = 4). The five means in each row are: the grand mean over all conditions; TRAIN: the training view (VP= 17^o and IL=0); VP diff: average over all new viewing position with the training illumination (VP$\neq$ 17^o and IL=0); IL diff: average over all new illumination with the training viewing position (VP=17^o and IL$\neq$ 0); VP&IL diff: averaged over all combination of new illumination and new viewing position (VP$\neq$ 17^o and IL$\neq$ 0).

The variables that we consider are the percentage of correct rate responses (CR) and the average response time (RT) over the six appearances of a given image in the testing phase of each session. Since there was no interaction between subjects, set of faces, and the generalization to new images (see appendix B), we averaged separately for upright and inverted faces across subjects and sets of faces for each image condition. The averaged values of CR and RT in the upright and inverted conditions over all sets and subjects on their first and last exposure to each set are presented in Table 1. A correlation analysis revealed no speed-accuracy tradeoff (the correlation between CR and RT was never positive). Our subsequent analysis, presented below, concentrated on the CR data since this is the parameter that the subjects were trained to rich the high performance on; the results concerning response times are summarized in Appendix E.

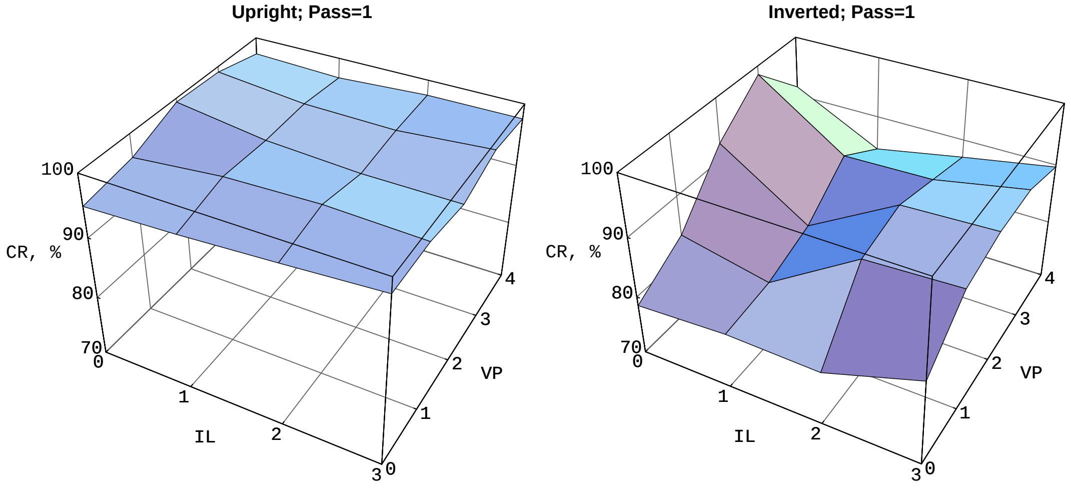
2.2.2 VP and IL effects for inverted faces

As mentioned above the same set of images was repeated four times for each subject. We begin by considering the first pass of each of the sets for all subjects. The averaged performance of all subjects is plotted in Figure 4, separately for upright and inverted

VP IL	-34	-17	0	17	34
0	78	83	93	**99**	91
1	79	80	83	89	84
2	79	90	91	89	87
3	84	90	92	92	90

Table 2: Each entry represents the average correct response (CR) over all subjects and sets for a given viewing position (VP) and illumination position (IL) for inverted faces. Only the first pass number is considered. The training view was VP= 17^o and IL=0

sets. Each point in the 2D graphs represents the average percentage of correct responses (CR) over all subjects for one of the 20 views of a face (4 illuminations × 5 viewpoints). The results for the inverted images are also summarized in Table 2.



a

b

Figure 4: Plots of percent correct (CR) for all viewing positions (VP) and illumination positions (IL). We consider only the first session of each subject on each set. Each point in the graph represents the average CR over all subjects for upright stimuli (a) and inverted stimuli (b). The training view (VP= 17^o and IL=0) is marked by a star.

We considered sessions in which the subjects had similar performance on the training images for upright (99.1 ± 0.5% correct) and inverted (98.6 + 0.7% correct) faces. The generalization for novel views of upright faces was remarkably good (above 97% correct). The generalization for novel views of inverted faces was considerably worse (average CR was 86.5 ± 1.5% for novel viewpoint and 90.1 ± 1.6% for novel illumination direction). For the inverted faces, the performance for novel viewpoint decreased monotonically with the misorientation relative to the training view. Statistical analysis (reported in appendices B and C) revealed no effects of either VP or IL for upright faces, for inverted faces both VP and IL had a significant effect on generalization: performance decreased.

2.2.3 Learning

In the previous section we considered only the first session (first pass) of all subjects on each of the sets. In this section we consider the effect of learning. In the following analysis, we distinguished between face-specific learning (within a set), and general learning (across sets). Learning within a set is due to better performance of the subject after having seen the same set of faces over and over again. In comparison, learning across sets is due to improvement in performance when the subject becomes more familiar with the task or the class (in this case of inverted faces), rather than with a specific stimulus set.

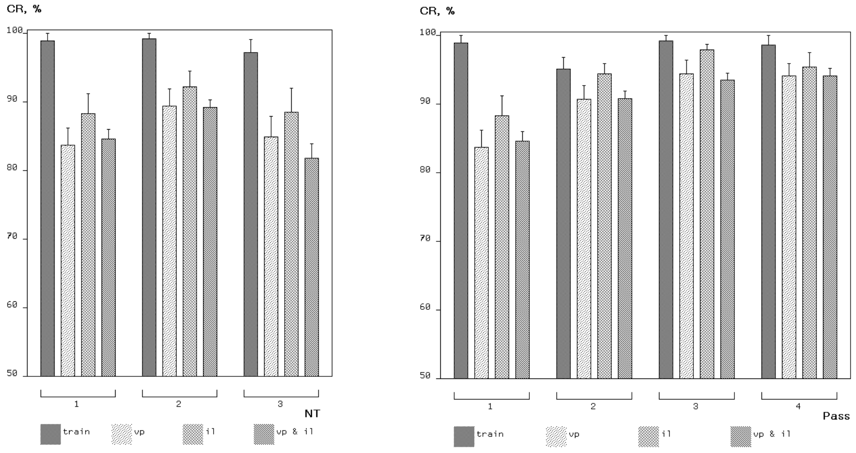
To study the effect of learning across sets, we considered only the first exposure (first pass) of each subject to each set, since this precludes of learning within a set. Figure 5(a) presents the mean performance of all subjects on the first pass of each set vs. the presentation order of appearance of the sets to a given subject (Set-order). The results do not indicate any learning across sets (see Appendix D). That is, the VP and IL effects were not reduced due to repeated exposure to sets of inverted faces.

To study the effect of learning within a set, we considered the change in performance of each subject with the number of repeated exposure (pass-number) to the first set of faces that the subject saw. Figure 5(b) presents the mean performance of all subjects and all sets vs. pass number. The improvement with the number of passes is manifest in the increase of mean CR and in the near disappearance of the effects of VP and IL (see Appendix D).

We can therefore conclude that the learning exhibited by the subjects was mostly set-specific. For each set of faces, this learning was apparent in the improvement of generalization performance with repeated exposure to that set. Learning across sets (that is, learning to perform better on the entire class of inverted faces) did not show up significantly in the data. The statistical analysis supporting these conclusions is presented in appendix D.

2.3 Summary of results

Our results indicate that subjects could be trained to discriminate between different faces in both upright and inverted faces when the same images are used in the training and in



a b

Figure 5: (a) Learning across sets: plots of CR vs. number of different inverted sets that a given subject saw (set number) for inverted stimuli, the average is taken on the first exposure of a subject to each of the sets. (b) Learning witnin a set: plots of CR vs. pass number, the average is taken on the first set that each subject saw. The four bars in each group correspond, from left to right, to the following four different conditions. TRAIN: the training view (VP= 17^o and IL=0); VP diff: average over all new viewing position with the training illumination (VP$\neq$ 17^o and IL=0); IL diff: average over all new illumination with the training viewing position (VP=17^o and IL$\neq$ 0); VP&IL diff: averaged over all combination of new illumination and new viewing position (VP$\neq$ 17^o and IL$\neq$ 0).

the testing stage. However, the generalization to views obtained under novel viewpoint and illumination conditions is significantly worse for inverted faces compared with upright faces.

We also found that the subjects were capable of non-supervised learning (improvement in generalization for novel views). This learning was specific for the face sets they saw, rather than the class of inverted faces in general: following repeated exposure to the same set of faces, the VP and IL effects nearly vanished, only to reappear when a new set of faces was introduced.

3 Conclusions and Discussion

The ability to generalize the recognition of a given face to novel images is a fundamental issue in face perception. Two natural parameters that vary between images are the illumination and viewpoint. The first question addressed in this paper was how well humans can recognize faces in novel images. The novel views we considered were taken from five different viewpoints and under four different illumination directions. The largest angular separation between novel and familiar viewpoints was 51°, and the largest separation in illumination direction was about 50° away from the direction used in the training image. We found that for upright faces the subjects that were trained to recognize well the training image, responded correctly to over 97% of the test stimuli. We conclude that the human visual system generalizes well the identification of upright faces to novel images within the range of viewpoint and illumination changes we tested.

How this impressive invariance for large discrepancy in viewpoint and illumination obtained? The comparison between upright and inverted faces allows us to draw a number of plausible conclusions regarding the probable level at which the human visual system overcomes variations between images of the same face. In the discussion below, we draw a distinction between the universal, class and object levels of achieving invariant recognition.

Consider first the universal level. At this level, the system attempts to compensate for the variability among images of a the same face in the same manner for all objects. An example of a universal process of this kind, widely used in computer vision, is the extraction of intensity edges from the image. A major goal of this process is to form an intermediate representation based on image features that are relatively insensitive to illumination. Similarly, for biological visual systems, there is physiological evidence for neurons that are sensitive to both orientation and spatial frequency in the primary visual cortex (Hubel and Wiesel, 1962; Hubel and Wiesel, 1968). This stage of processing was modeled as the application of a set of local filters to the incoming image (Daugman, 1984; Mercelja, 1980; Marr and Hildreth, 1980; Pollen and Ronner, 1983). Presumably, this processing is applied uniformly to any input image, and, in particular, to upright as well as inverted images of faces. Differences between the processing of upright and inverted faces are therefore unlikely to arise from this level of processing. More generally, the difference between the performance of the visual system in generalizing across viewpoint and illumination changes in upright and inverted faces suggests that universal processing, such as edge detection, is not sufficient by itself to compensate for these image variations.

The extreme opposite to universal processing is the object-specific approach. Here, when an image is compared with stored object model, processes that depend on the particular object in question are utilized. These processes can therefore specifically deal with the effects of viewpoint and illumination for the object in question. For example, for familiar objects the system may store detailed 3-D models acquired through past experience. Such a model can then be used in compensating for viewpoint and illumination effects of the object in question, but not to compensate for viewpoint and illumination effects in general (for other objects). Some of the object-specific approaches overcome the differences between images due to changes in viewpoint by using multiple images of a given object as reference. There are two different ways in which multiple images can be used: the independent and the interdependent approaches. The independent approach is straightforward: the system stores a sufficiently large set of images, so that each novel face image is bound to be close to one of the images in the set, considered independently. The interdependent approach uses several images of the same face together, to extract (either directly or indirectly) information about the three dimensional shape of the face (Fischler and Bolles, 1981; Ullman, 1986; Huttenlocher and Ullman, 1987; Lowe, 1987; Basri and Ullman, 1988; Grimson, 1990; Poggio and Edelman, 1990). Because in our experiments only a single image was available to the system in the learning phase, object-specific approaches that rely on several images of the same face cannot account for our experimental results. Another method for constructing the 3-D shape of the object is by using the shading information from a single image(Horn and Brooks, 1989). In this approach, computing the shape of the object is performed at the universal level (that is, no specific assumptions on faces in general or specific faces are used to compute the 3-D shape). This approach is therefore also insufficient to explain the differences between the generalization of upright and inverted faces.

In between the universal and object-specific levels lie the class-based level. At this level, the processing depends on the class to which the object in the image is assumed to belong. For example, class-level processing may include the extraction of facial features such as the location of the eyes, mouth and nose (Kanade, 1977; Craw et al., 1987; Yuille et al., 1989). Such a process is applicable to face images in general, but not to other objects. In general, classification can be hierarchical, that is, a given class of objects can belong to a more inclusive class as well. For example, the face of an individual belongs to the class of human faces, which, in turn, belongs to the class of animal faces, which belongs to the class of approximately bilaterally symmetric objects. Therefore, class-based processing can consist of several levels of processing.

The use of such class-based processing means that the system uses general properties of faces to compensate for the effects of illumination and viewing conditions of an individual face. Our experiments revealed differences in generalization performance between upright and inverted faces. This strongly suggests that generalization is not entirely universal in nature. Although universal process may have a contribution, class-based processes are likely to play an important role in compensating for both illumination an viewpoint. Our results also show substantial recognition ability from a single (upright) face image. This again is consistent with the use of class-based processing. The use of general face property could help to explain how generalization is obtained on the basis of a single novel image, and why this generalization is more effective for familiar faces.

Class-based processing can be used in several manners. For example, the 3D shape of the object can be easier to recover from a single image if one assumes that the object is indeed a face (the recovery of general shape from shading from a single image is impossible without assumptions on the light source position, the reflectance properties, etc.). Moreover, because faces belong to the class of bilaterally symmetric objects, the symmetry assumption can be used, for example, for dealing with the viewpoint variation, as described in Moses and Ullman (1992a) .

In conclusion, our results show a remarkably good generalization from a single image to novel views of upright faces, along with reduced generalization performance for inverted faces. We suggest that the difference in generalization performance is related to class-level processing, and that for the visual system upright and inverted faces are different classes of objects. In other words, the visual system uses general face properties to compensate for new viewing conditions of a specific face. To investigate further the possibility that class-based processing plays an important role in generalization across pose and illumination changes, the experiments reported here could be repeated with other classes of objects. Specifically, generalization over controlled viewpoint, illumination, and other imaging parameters could be compared for different classes of objects. Furthermore, it would be interesting to determine whether performance for a given class improves with repeated exposure to different objects from this class. As an example, consider the class of inverted faces, which are rarely seen in daily life. Our experiments revealed virtually no improvement in the generalization process for one set of inverted faces after repeated exposure to another set of inverted faces. It is possible, however, that a longer exposure to inverted faces would result in the learning of inverted faces as a class, leading to improved generalization from a single view of an inverted unfamiliar face.

Appendix A: the independent variables

The independent variables that were involved in the analysis are listed in Table 3.

Variable	*Levels*	*Remarks*
Invert	0, 1	0=upright, or ↑; 0=inverted or ↓
VP	-34, 17, 0, 17, 34	Training: VP=17°
IL	0, 1, 2, 3	*IL*=0 for left, *IL*=1 for center, *IL*=2 for right, and *IL*=3 for left and right together. Training: *IL*=0
Set	A, B, C, D, E, F	Each set consisted of images of 3 faces
Subject	EST,OR1,TAL,ARN JUD,NUR,MOR,OR2	Sets [A,D,F]↑, [B,C,E]↓. Sets [A,D,F]↓, [B,C,E]↑.
Pass-number	1, 2, 3, 4	The number of times a *Subject* was exposed to a given *Set*.
Session	[1..24]	counted separately for each Subject
Set-order	1, 2, 3	The sets were numbered in the analysis for each subject according to their chronological order of appearance. The upright sets and the inverted sets were numbered separately.

Table 3: The independent variables involved in the analysis.

Appendix B: effects of *Subject* and *Set*

We first tested the interaction of the variables *Subject* and *Set* with the effects of *VP* and *IL*, to determine whether the influence of subject and stimulus variability (*Set*) would have to be taken into account explicitly in the subsequent analyses. To that end, we performed a mixed-model GLM (General Linear Models) analysis, in which the effects of *VP*, *IL*, *Subject*, *Set*, and all the two-way interactions were tested, with *Subject* and *Set* declared as random effects. The analysis was carried out separately for upright and inverted conditions, and also separately for each value of *Pass-number* (because the performance changed with *Pass-number*, and the rate of this change differed among subjects).

The results yielded interactions between *Subject*, *Set*, and the effects of *VP* and *IL*, in both orientation conditions, for most of the values of *Pass-number*. A look at the data showed, however, that the source of these interactions may have been the poor performance of a single subject, ARN (see Table 4); this subject was also responsible for 18 out of the 30 sessions that were omitted from the analysis because of the lack of learning of the training configuration). Indeed, without this subject, there was virtually no interaction of *Subject* and *Set* with *VP* and *IL*.[3] Consequently, in all further analyses, we used only the data from the seven remaining subjects, and treated the variation over the *Subject* and *Set* degrees of freedom as error terms.

Appendix C: effects of *VP* and *IL*, and their interaction with *Invert*

To find out how the inversion of the stimuli affected generalization across changes in viewpoint and illumination, we performed a 3-way (*VP*×*IL*×*Invert*) GLM analysis of variance. The analysis was done separately for the first (*Pass-number*=1) and the last (*Pass-number*=4) exposure of a subject to a set. All the main effects and all the two-way interactions were significant (see Table 5). The prominence of the *VP***Invert* and *IL***Invert* interactions clearly demonstrates that generalization across *VP* and *IL* depended strongly on whether the stimuli faces were inverted or not (see also Figure 4).

We next carried out four separate GLM analyses: for the upright and the inverted conditions, for *Pass-number*=1 and *Pass-number*=4. For upright stimuli at *Pass-number*=1, the illumination *IL* had no effect on CR ($F < 1$), and there was a marginal effect of *VP* ($F(4,1120) = 2.1$, $p < 0.07$). No effects of *IL* or *VP* remained for upright stimuli at *Pass-number*=4. In contrast, for inverted stimuli at *Pass-number*=1 we found strong main effects of *IL* ($F(3,940) = 5.4$, $p < 0.0012$) and of *VP* ($F(4,940) = 10.3$, $p < 0.0001$), and no interaction; at *Pass-number*=4 both these effects were reduced but still present (*IL*: $F(3,1120) = 3.4$, $p < 0.02$; *VP*: $F(4,1120) = 5.7$, $p < 0.0002$).

A direct impression of the effects of viewpoint and illumination on generalization performance may be obtained by considering the means of CR for the different values of *VP* and *IL*. At *Pass-number*=4, the adjusted marginal mean correct rate for *VP*=17° (the training viewpoint) was CR=96.0%, and for *VP*=0 it was CR=90.7% (difference significant at $p < 0.0001$; most of the other differences between the marginal means of CR were also significant). For *IL*=0 (the training illumination), the marginal mean

[3] The only marginally significant interactions were: *IL***Subject* for *Invert*=0, *Pass-number*=2 ($F(3,1091) = 1.93$, $p < 0.02$); *IL***Set* for *Invert*=1, *Pass-number*=1 and *Pass-number*=2 ($F(12,855) = 1.90$, $p < 0.03$, and $F(12,1032) = 2.39$, $p < 0.005$, respectively).

```
----------------------------------  Invert=0----------------------------------
               Duncan Grouping              Mean      N   Subject

                             A           99.3264    720   NUR
                             A
                     B       A           99.2130    720   EST
                     B       A
                     B       A           99.1204    720   OR2
                     B       A
                     B       A           98.3460    660   OR1
                     B
                     B                   98.1296    720   TAL

                             C           95.5370    540   JUD

                             D           93.8500    600   MOR

                             E           77.2722    300   ARN

----------------------------------  Invert=1 ----------------------------------
               Duncan Grouping              Mean      N   Subject

                             A            94.968    660   EST
                             A
                             A            94.942    660   OR1
                             A
                     B       A            93.181    720   NUR
                     B
                     B                    91.566    660   MOR
                     B
                     B                    91.250    600   OR2

                             C            88.503    540   TAL
                             C
                             C            87.272    540   JUD

                             D            73.389    180   ARN
```

Table 4: Results of Duncan's Multiple Range test of the differences between mean values of CR for the eight subjects, in the upright and the inverted conditions. Note the great difference between the performance of ARN and that of other subjects (in response time, ARN was ranked third in both conditions). ARN's data were subsequently omitted from the analysis.

correct rate was CR=95.8%, compared to CR=92.4% for *IL*=1 (difference significant at $p < 0.004$; differences among other levels of *IL* were not significant).

Appendix D: learning

Despite there being no feedback indicating incorrect responses during the testing stage of each experimental session, the subjects' performance improved with repeated exposure This improvement with *Pass-number* was manifest in the increase of mean CR, and, more interestingly, in the diminution of the effects of *VP* and *IL* (see the description of the effect of *Pass-number* in section 3). To determine whether this improvement was a non-specific practice effect, or an indication of stimulus-specific learning, we obtained a quantitative measure of the relative importance of *Set-order* and *Pass-number* by a four-way (*VP*×*IL*×*Pass-number*×*Set-order*) analysis of covariance (with *Pass-number* and *Set-order* treated as continuous variables).

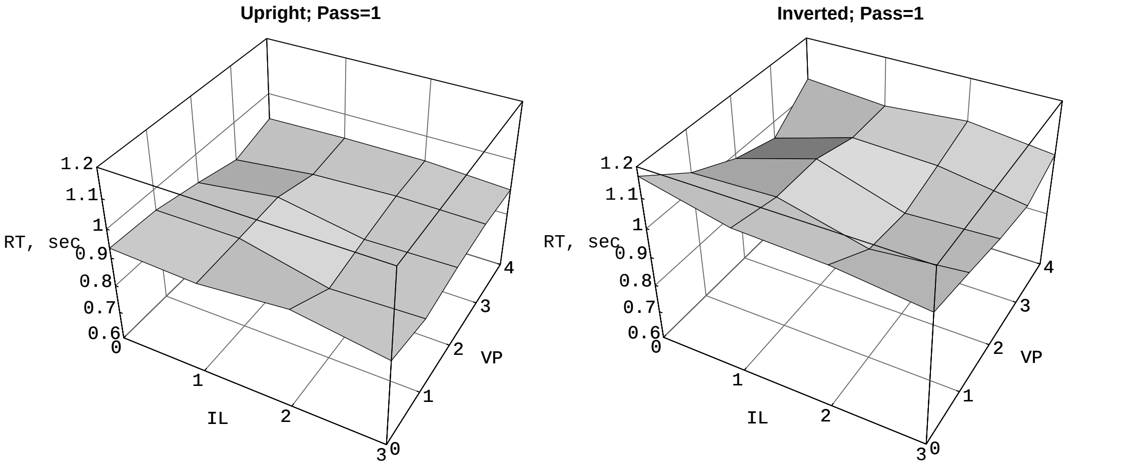
The results are summarized in Figure 5 and in Table 6. In the upright condition, we found no effects of interest. In the inverted condition, the analysis showed, as expected, all the effects of *VP* and *IL* we saw before. In addition, there was a significant main effect of *Pass-number* ($F(1,4343) = 29.37$, $p < 0.0001$), but not of *Set-order* ($F < 1$). Interestingly, familiarity (that is, *Pass-number*) affected generalization: the interaction of *VP* with *Pass-number* was marginally significant (at $p < 0.11$).

Appendix E: analysis of response time (RT)

Because the main conclusions made in this paper are based on the analysis of the CR data, our primary concern was to assure that the subjects did not trade recognition rate for response time. Indeed, we found no positive correlation between CR and RT in any of the sessions or conditions. We found difference in mean RT between upright and inverted conditions (see Table 1). Because we found no session by session speed-accuracy tradeoff, this difference in RT between upright and inverted faces does not change our interpretation of the CR results. In fact, inverted faces took longer to respond to *and* were more difficult to recognize — the opposite of what happens when speed is traded off for accuracy.

Effects of *VP*, *IL*, and *Invert*

To find out how the inversion of the stimuli affected the response time RT across changes in viewpoint and illumination, we performed a 3-way (*VP*×*IL*×*Invert*) GLM analysis of variance. As in the section on CR, the analysis was done separately for *Pass-number*=1, and for *Pass-number*=4. At *Pass-number*=1, the main effect on RT of illumination *IL* was weak, and its interactions with the other variables were not significant (see Table 7). In comparison, RT did depend strongly on viewpoint *VP*, and this dependence was affected by the inversion of the faces (see the *VP***Invert* interaction in Table 7, top, and Figure 6).



Upright; Pass=1

Inverted; Pass=1

Figure 6: Plots of RT vs. *VP* and *IL*, for upright stimuli (left) and inverted stimuli (right). The data in this plot are for *Pass-number*=1.

```
-------------------PASS NUMBER=1------------------------
 Source                        DF    F Value     Pr > F

  VP                            4      13.88     0.0001
  IL                            3       6.30     0.0003
  VP* IL                       12       1.80     0.0425
  Invert                        1     258.99     0.0001
  VP* Invert                    4       7.66     0.0001
  IL* Invert                    3       4.66     0.0030
  VP* IL* Invert               12       1.04     0.4106

-------------------PASS NUMBER=4------------------------
 Source                        DF    F Value     Pr > F

  VP                            4       4.77     0.0008
  IL                            3       2.26     0.0793
  VP* IL                       12       0.77     0.6846
  Invert                        1      36.51     0.0001
  VP* Invert                    4       4.82     0.0007
  IL* Invert                    3       3.44     0.0162
  VP* IL* Invert               12       0.45     0.9428
```

Table 5: Results of GLM analyses of variance that tested the effects of *VP*, *IL*, and *Invert* on CR for Pass-number=1 and Pass-number=4. The number of error DFs was 2060 and 2240, respectively, in the two cases. Note the diminishing influence of *Invert* on the effects of *VP* and *IL* at Pass-number=4, compared to Pass-number=1.

Source	DF	F Value	Pr > F
VP	4	10.31	0.0001
IL	3	4.62	0.0031
VP*IL	12	3.19	0.0001
Set-order	1	0.23	0.6349
Set-order*VP	4	1.37	0.2426
Set-order*IL	3	1.21	0.3038
Pass-number	1	29.37	0.0001
Pass-number*VP	4	1.88	0.1115
Pass-number*IL	3	1.50	0.2115
Set-order*Pass-number	1	3.20	0.0737

Table 6: Results of the analysis of covariance that tested the influence of learning on the effects of *VP* and *IL*. Only the inverted condition is shown (in the upright condition there were no significant effects). The number of error DFs was 4343.

```
-------------------Pass-number=1------------------------
 Source                    DF   F Value     Pr > F

 VP                         4      8.34     0.0001
 IL                         3      2.92     0.0329
 VP*IL                     12      1.70     0.0605
 Invert                     1    199.90     0.0001
 VP*Invert                  4      2.64     0.0322
 IL*Invert                  3      0.31     0.8180
 VP*IL*Invert              12      0.42     0.9549

-------------------Pass-number=4------------------------
 Source                    DF   F Value     Pr > F

 VP                         4      5.56     0.0002
 IL                         3      1.29     0.2750
 VP*IL                     12      0.24     0.9966
 Invert                     1     84.72     0.0001
 VP*Invert                  4      0.97     0.4233
 IL*Invert                  3      0.96     0.4119
 VP*IL*Invert              12      0.31     0.9887
```

Table 7: Results of GLM analyses of variance that tested the effects of *VP*, *IL*, and *Invert* on RT for *Pass-number*=1 and *Pass-number*=4. The number of error DFs was 2060 and 2240, respectively, in the two cases. The interaction of *Invert* with the effect of *VP*, present at *Pass-number*=1, disappeared at *Pass-number*=4. Note that the effect of viewpoint *VP* on RT is still very strong at *Pass-number*=4.

Acknowledgements

This research was supported in part by The Israel Science Foundation administered by The Israel Academy of Sciences and Humanities. We are grateful to Eli Okon and Oded Smikt for technical assistance in setting up the face image acquisition system, to Edna Schechtman for statistical advice, and to Dov Sagi for comments on a draft of this report. We also thank the eighteen people who were patient enough to have long series of their snapshots taken, and the subjects of the psychophysical experiments for their time.

References

Attneave, F. (1967). Criteria for tenable theory of form perception. In Wathen-Dunn, W., editor, *Models for the Perception of Speech and Visulal Form*. M.I.T. press.

Basri, R. and Ullman, S. (1988). The alignment of objects with smooth surfaces. In *Proceedings of the 2nd International Conference on Computer Vision*, pages 482–488, Tarpon Springs, FL. IEEE, Washington, DC.

Bruce, V. (1982). Changing faces: visual and non visual coding processes in face recognition. *British Journal of Psychology*, 73:105–116.

Carey, S. and Diamond, R. (1977). From piecmeal to configurational representation of faces. *Science*, 195:312–314.

Craw, I., Ellis, H., and Lishman, J. (1987). Automatic extraction of face-features. *Patter Recognition Letters*, 5:183–187.

Daugman, J. G. (1984). Spatial visual channels in the fourier plane. *Vision Res.*, 24(9):891–910.

Davies, G. M., Ellis, H., and Shephered, J. W. (1978). Face recognition accuracy as a function of mode of representation. *Journal of Applied Psychology*, 92:507–523.

Endo, M. (1982). Cue saliency in upside down faces. *Tohoku Osychologia Folia*, 4:116–122.

Endo, M. (1986). Perception of upside-down faces: an analysis from the viewpoint of cue saliency. In Ellis, H., Jeeves, M., Newcombe, F., and Young, A., editors, *Aspects of Face Processing*, pages 53–58. Martnus Nijhoff Publishers.

Fawcett, R., Zisserman, A., and Brady, J. (1994). Extracting structure from an affine view of a 3d point set with one or two bilateral symmetries. *Image and Vision Computing*, 12(9):615–622.

Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24:381–395.

Grimson, W. E. L. (1990). *Model-Based Vision.* MIT Press, Cambridge, MA.

Hochberg, J. and Galper, R. E. (1967). Recognition of faces: an exploratory study. *Psychonomic Science*, 9:619–620.

Horn, B. K. P. and Brooks, M. (1989). *Seeing shape from shading.* MIT Press, Cambridge, Mass.

Hubel, D. and Wiesel, T. (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *Journal of Physiology (London)*, pages 106–154.

Hubel, D. and Wiesel, T. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, pages 215–243.

Huttenlocher, D. P. and Ullman, S. (1987). Object recognition using alignment. In *Proceedings of the 1st International Conference on Computer Vision*, pages 102–111, London, England. IEEE, Washington, DC.

Johnston, A., Hill, H., and Carman, N. (1992). Recognition faces: effects of lighting direction, inversion and brightness reversal. *Perception*, 21:365–375.

Kanade, T. (1977). *Computer recognition of human faces.* Birkhauser Verlag. Basel ans Stuttgart.

Kemp, R., McManus, I., and Pigott, T. (1990). Sensitivity to the displacement of facial features in negative and inverted images. *Perception*, 19:531–543.

Köhler, W. (1947). *Dynamics in psychology psychology.* Liveright, New York.

Lowe, D. G. (1987). Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence*, 31:355–395.

Marr, D. and Hildreth, E. (1980). Theory of edge detection. *Proc. R. Soc. Lond. B*, 207:187–217.

Mercelja, S. (1980). Mathematical description of the responses of simple cortical cells. *J. Opical Soc.*, 70:1297–1300.

Moses, Y. and Ullman, S. (1992). Limitation of non-model-based recognition schemes. In Sandini, G., editor, *Proc. ECCV-92*, pages 820–828. Springer-Verlag.

Patterson, K. and Baddeley, A. (1977). When face recognition fails. *Journal of Experimental Psychology: Human Learning and Memory*, 3:406–417.

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

Pollen, D. and Ronner, S. (1983). Visual cortical neurons as localized spatial frequency filters. *IEEE Transactions on System, Man and Cybernetics, SMC-13*, pages 907–916.

Scapinello, K. F. and Yarmey, A. (1970). The role of familiarity and orientation in immediate and delayed recognition of pictorial stimuli. *Psychonomic Science*, 21:329–331.

Ullman, S. (1986). An approach to object recognition: Aligning pictorial descriptions. A.I. Memo No. 931, Artificial Intelligence Laboratory, Massachusetts Institute of Technology.

Valentine, T. (1988). Upside-down faces: a review of the effect of inversion upon face recognition. *British Journal of Psychology*, 79:471–491.

Valentine, T. and Bruce, V. (1986). The effect of race, inversion and encoding activity upon face recognition. *Acta Psychologica*, 61:259–273.

Vetter, T., Poggio, T., and Bulthoff, H. (1994). The importance of symmetry and virtual views in three dimensional object recognition. *Current Biology*, pages 18–23.

Yarmey, A. (1971). Recognition memory for familiar "public" faces: effects of orientation and delay. *Psychonomic Science*, 24:286–288.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, 81:141–145.

Yuille, A. L., Cohen, D., and Hallian, P. (1989). Feature extraction from faces using deformable templates. In *Proc. CVPR-89*, San Diego, CA.