

# 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.

May 10, 1994

## Abstract

An image of a face depends not only on its shape, but also on the viewing position, illumination conditions, and facial expression. Any face recognition system must overcome the changes in face appearance induced by these factors. To assess the ability of human vision to generalize across changes in illumination and pose of faces, we studied the performance of subjects in a discrimination task with either upright or inverted faces. Subjects first learned to discriminate among images of three faces, taken under fixed viewing position and illumination. They were then tested on images of the same faces taken under all combinations of four illuminations and five viewing positions. For upright faces, we found remarkably good generalization to novel conditions. For inverted faces, the generalization to novel views was significantly worse, although the performance on the training images was similar in both cases.

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 generalization 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 a single image. However, recognizing faces is a difficult problem from a computational point of view since all faces have a generally similar shape and at the same time, different images of the same face can vary considerably. An image of a given face depends not only on its three dimensional shape, but also on facial expression, age, viewing position, illumination, noise, etc. The image can be regarded as a function of all of these imaging parameters, and the task of a face recognition system, whether natural or artificial, is to recognize a face in a manner that is independent of these parameters. The basic issue we are concerned with is how the human visual system can identify a face in novel images, taken with new values of the imaging parameters. The two imaging parameters we consider are the illumination condition and viewing position.

We consider here two aspects of the problem. The first question is how well humans in fact recognize faces in novel images, that is, to what an extent novel illumination and viewing position hinder the recognition of a face. The second is the level at which the generalization of face identification to novel images takes place.

Following Moses (1993) we distinguish between three basic generalization levels: *universal*, *class based*, and *model based*. Roughly speaking, the universal level of generalization is common to all images, independent of the specific object to be recognized, and independent of other objects that have been learned by the system. For example, the extraction of edges from an image is often employed as a universal processing stage designed to reduce the effects of illumination changes. The other extreme generalization level is model-based. At this level, the processing applied to compensate for illumination and pose depends on the specific object to be recognized. It can also depend on the other objects known to the system. An example of model-based computation is recognition by alignment (Ullman, 1986; Lowe, 1986; Basri and Ullman, 1988). In this approach, the system stores a 3D model of each known object. Given an image, the model is transformed so that its projection aligns best with the target image. The effect of viewing position is compensated for in this case by using a specific object model. An intermediate level of generalization is a class-based level. At this level, the generalization process uses properties associated with certain classes of objects. For example, in the case of face images, class-level processing may include the extraction of facial features such as the eyes and the mouth. The three basic generalization levels defined here are discussed further in section 3.

The experiments described in this paper were designed to address the two basic questions posed above: how well and at what level does the human visual system generalize face identification to images taken under novel illumination and viewpoint conditions. We examined the generalization performance for two classes of objects: upright and inverted

faces. Images of inverted faces clearly have the same complexity as those of upright faces. However, the class of upright faces is more familiar to us than the class of inverted faces, and therefore easier to recognize. In the experiments we did not compare, however, the difficulty of recognition *per se*, but the ability to generalize from familiar to novel views. Differences in generalization performance between upright and inverted faces can serve to indicate the involvement of class-based processing in generalization.

Subjects first learned to recognize three images of 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. In this case, the subject learned to recognize inverted faces and was tested on 20 novel images of each of the three inverted faces.

We found that the generalization of recognition to novel views of upright faces was remarkably good (see section 2.2.2). In contrast, the subjects' performance on inverted faces was degraded when novel views were presented. That is, even after the subjects learned to recognize well specific inverted images, their generalization across illumination and viewing position was significantly worse than for the upright 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.

Several previous studies have addressed the problem of generalization of face memory to novel images taken from new viewing positions, but without changing the illumination condition (Patterson and Baddeley, 1977; Davies et al., 1978; Bruce, 1982). In these experiments a set of faces (unfamiliar or familiar) was presented briefly once to the subject (training phase). In the testing phase the subject had to determine whether a given face was shown in the training phase. Two viewing positions were used: frontal, and 3/4 profile. The results showed that the recognition of a previously seen face in a novel view was reliable. Bruce (1982) compared the recognition of familiar and unfamiliar faces in such an experiment, and found that familiar faces were recognized more quickly and accurately than unfamiliar faces. Our experiments were different in that the subjects were tested on face identification (in a three-alternative forced-choice setup), the faces were unfamiliar to the subject to begin with, but by repeated exposure at the training stage, one image of each face became familiar to the subject (This can explain the differences between our results and those of Bruce (1982) regarding the recognition of unfamiliar faces). The set of images that we considered for each face were larger (20 images compared

to two). The images in our experiments varied not only due to the pose but also due to illumination. Furthermore, our set of images were well controlled such that each parameter (e.g. viewing position or illumination) was varied independent of the others, while images of all faces were normalized to the same conditions. The extent to which viewing position, location and size of the face, illumination, background, and in many cases familiarity to the subject were controlled in (Patterson and Baddeley, 1977; Davies et al., 1978; Bruce, 1982) experiments was not made clear.

The recognition of inverted faces is known to be difficult (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 the research that involved the recognition of inverted faces can be found in Valentine (1988). Inverted faces were used as stimuli in memory experiments that addressed the question of whether faces are processed by a unique mechanism (Yin, 1969; Scapinello and Yarmey, 1970; Yarmey, 1971; Diamond and Carey, 1986). In these studies, face memory was compared to the memory of other objects such as dogs, landscapes and houses. 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 find out whether features or configural information is required for face recognition. Carey & Diamond (1977) proposed that the difficulty in recognizing inverted faces results from an inability to access the configural information of the facial features from inverted faces. The cue saliency in artificial inverted faces (schematic or threshold to black and white) was addressed by several investigators (Endo, 1982; Endo, 1986; Kemp et al., 1990). The results of all these experiments indicated that memory for upright and inverted faces works somewhat differently. In our case, inverted faces were used to probe specific aspects of face processing: the effects of changing illumination and viewing position on the identification of faces.

## 2 The experiments

The basic experimental paradigm was three-alternative forced-choice recognition. The subject was first trained on a set of three faces, one image of each face. She or he was then tested on 20 different images of each of the three faces, taken under all combinations of four different illumination positions and five different camera locations. 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. This assignment of orientation and triplet was balanced across subjects, so that the faces in each triplet were seen upright by one half of the subjects, and 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 (three females, five males age 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 symmetry axis of the face, the external corners of the eyes, and the bottom of the nose were located on the reference lines as shown.

Images of 18 different faces were used as stimuli. All faces were of males, without glasses, beard, mustache or other distinctive features. All faces were unfamiliar to the subjects. We used 20 different images of each face, taken under all combinations of four illumination positions (IL) and five viewing positions (VP).

All images were taken by the same camera under tightly controlled illumination and viewing position conditions. The camera (Pulnix TM-560 with Canon lens V6  $\times$  1616 – 100mm  $F1 : 1.9$ ) was attached to a robot arm (Adept I). A Symbolics Lisp Machine controlled the camera positioning. The camera was positioned at  $-34^\circ$ ,  $-17^\circ$ ,  $0^\circ$ ,  $17^\circ$

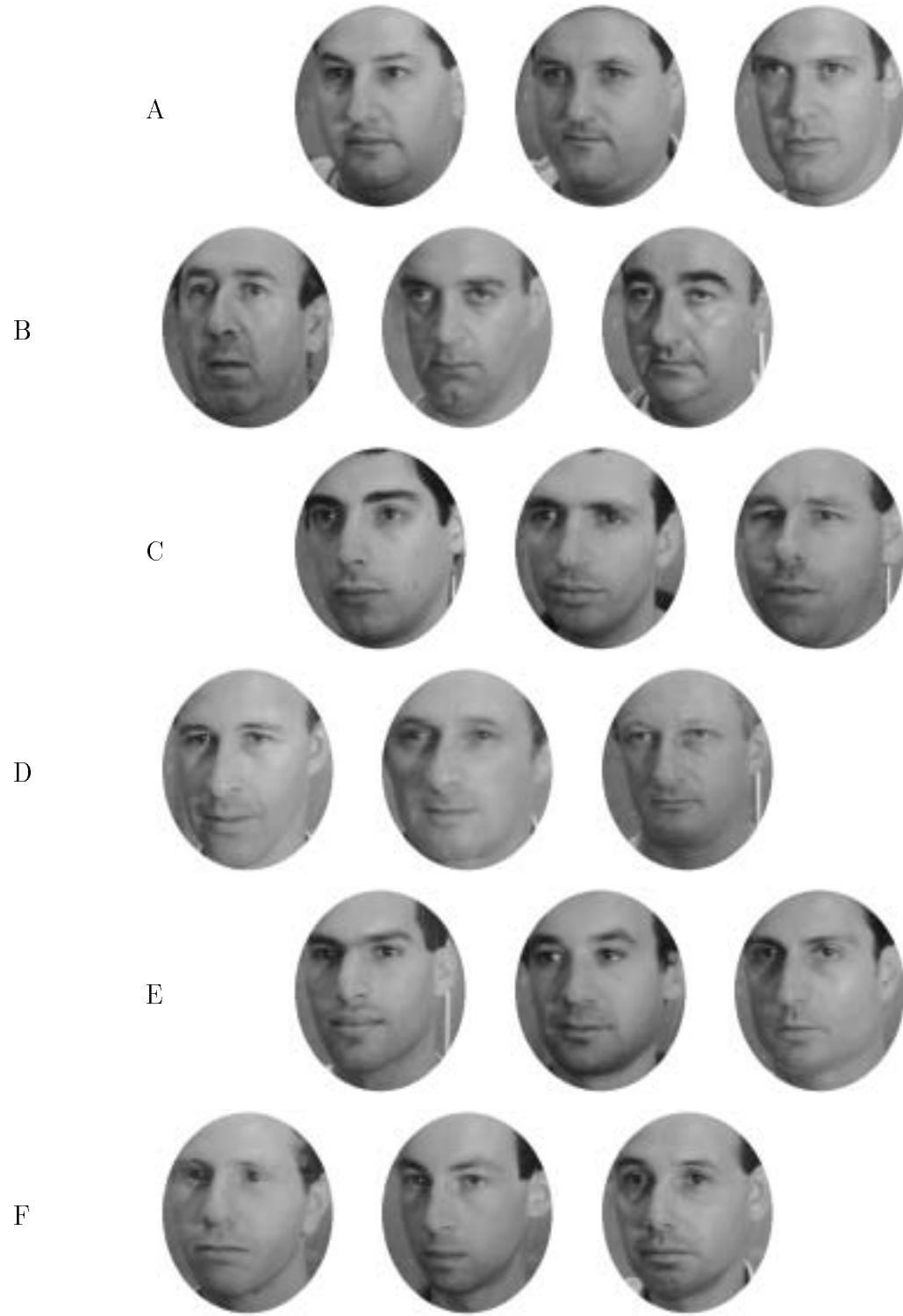


Figure 2: The triplets of images used in the experiment, one image of each face. These images were used in the training phase (VP=17 and IL =0).

and  $34^\circ$  with respect to the frontal view, in the horizontal plane. The distance of the face from the camera was fixed at about  $110\text{cm}$ . The frontal view of all faces were normalized by fixing the location of the face symmetry axis, the external corners of the eyes, and the bottom of the nose, before taking the pictures (see Figure 1). Four distinct illumination conditions were created by turning on and off three fixed light sources (see Figure 3): 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. A set of 20 images of one of the 18 faces is shown in Figure 4.

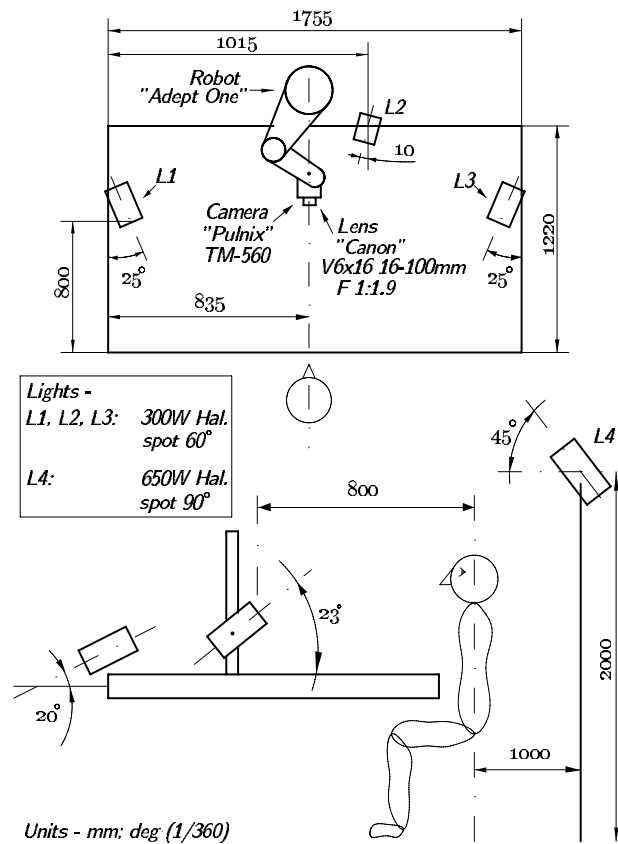


Figure 3: Imaging setup.

Each image consisted of  $512 \times 352$  pixels, 8 bits pixel. The subjects viewed the images on the screen of a Silicon Graphics Personal Iris 4D35/TG workstation, at an approximate distance of  $50\text{cm}$ . At that distance, an image subtended approximately  $6.8$  degrees of visual angle.



Figure 4: An example of 20 images of one of the faces (all combinations of five different viewing position and four different illumination).

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. We chose the triplets in such a way that the faces within a set were similar to one another. The sets were denoted by the letters A through E (see Figures 2).

### 2.1.3 Procedure

The experiment consisted of 24 sessions for each of the eight subjects. In each session, one fixed set of faces was used. For a given subject, half of the face sets were upright images and the other half were inverted.

An experimental session started with a training phase, in which the subject was shown repeatedly a single image of each of the three faces of the set, in a pseudorandom order. The subject was then tested on all images of the three faces. The same view of the face (viewpoint:  $VP=17^\circ$ , and illumination  $IL=0$ ) was used for all the images of the training phase (see Figure 2). 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 which appeared at the bottom of the screen showing which of the three response buttons (“1”, “2”, or “3” on the numeric keypad of the workstation) had to be pressed. 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 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. The set of all images of the three faces corresponded to all combinations of the four different illumination positions and five different camera locations of each of the faces. Altogether, there were  $20 \times 3$  different images. Each image was presented six times in the testing phase.

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.

In each trial of the experiment, the stimulus image was shown for  $600msec$ , and was 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.

An experimental 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),<sup>1</sup> followed by 360 testing trials (5 levels of  $VP \times 4$  levels of  $IL \times 3$  faces  $\times 6$  replication).

---

<sup>1</sup>If the subject did not reach this level of performance in 250 trials, the session was aborted, and was

The experiments included eight subjects, six face sets, and each subject made four passes over each of the six sets. The passes over a given set (pass number) were numbered in the analysis from 1 to 4 for each subject. In all, each subject underwent 24 experimental sessions (training followed by testing, as described above). In a given session, the faces were either all upright or all inverted in both the learning and the testing phases. For the first four subjects, the sets  $C$ ,  $B$ , and  $E$  were always upright, and sets  $A$ ,  $D$ , and  $F$  always inverted. The other four subjects saw sets  $C$ ,  $B$ , and  $E$  inverted, and sets  $A$ ,  $D$ , and  $F$  upright. The sets were numbered in the analysis for each subject according to their chronological order of appearance (set order) from one to three. The upright sets and the inverted sets were numbered separately (see appendix A).

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 normally shown to a given subject during two sessions in a row). This made it easier for the subjects, who ran up to four sessions at a time, 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). Then we analyze the generalization across viewing position and illumination in the first exposure of a subject to a set of upright and inverted faces separately (section 2.2.2). Finally, the improvement in generalization within sets and across sets with time is described (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 ( $VP=17^\circ$ , and  $IL=0$ ) in the testing phase. This criterion was satisfied in 86 out of 96 upright sessions and 76 out of 96 sessions, that is, in 162 out of the total of 192 experimental sessions (24 sessions  $\times$  8 subjects). We 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 responses (about 84% of the original volume of data). We discarded all sessions of one subject due to his general bad performance. His performance was also statistically different from that of the rest of the subjects. Finally, we averaged across the remaining subjects and sets of faces since

---

restarted from the beginning after a short break. This happened in 6 sessions, or about 3% of the total number of sessions.

there was no interaction between subjects and set of faces and the generalization to new images (see appendix B).

<i>Pass</i>	<i>Statistic</i>	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, <i>ms</i>	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, <i>ms</i>	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 (upper table) and the last pass of all sets (lower table). The five means in each row are: the grand mean over all conditions; TRAIN: the training view (VP= 17° and IL=0); VP diff: average over all new viewing position with the training illumination (VP≠ 17° and IL=0); IL diff: average over all new illumination with the training viewing position (VP=17° and IL≠ 0); VP&IL diff: averaged over all combination of new illumination and new viewing position (VP≠ 17° and IL≠ 0).

Each image was presented in a given session in the testing phase six times. The variables that we consider are the average response time (RT) and the percentage of correct responses (CR) over the six appearances of a given image in the testing phase of each session.

The mean values of RT and CR in the upright and inverted conditions averaged over all subjects in their first pass (upper table) and the last pass (lower table) of all sets are presented in Table 1.

### 2.2.2 VP and IL effects in inverted faces

As mentioned, each subject went through four passes on each set of images. We first consider the percentage of correct responses (CR) of all subjects in their first pass on each of the sets (i.e., the first exposure of each set to a subject). The upright and the inverted sets are considered separately.

The performance of all subjects on each of the 20 views of the faces are illustrated in Figure 5. The 20 views include all combinations of five viewing positions and four illumination conditions. For each view, the average CR for all subjects and upright faces

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

Table 2: Tables of percent correct (CR) on a given viewing position (VP) and illumination position (IL). Only the first session of each subject on each set (first pass number) is considered. Each entry in the table represent the average CR over all subjects of upright stimuli (left) and inverted stimuli (right). The training view was  $VP=17^\circ$  and  $IL=0$

(left) or inverted faces (right) is marked. The results for the inverted images are also summarized in Table 2.

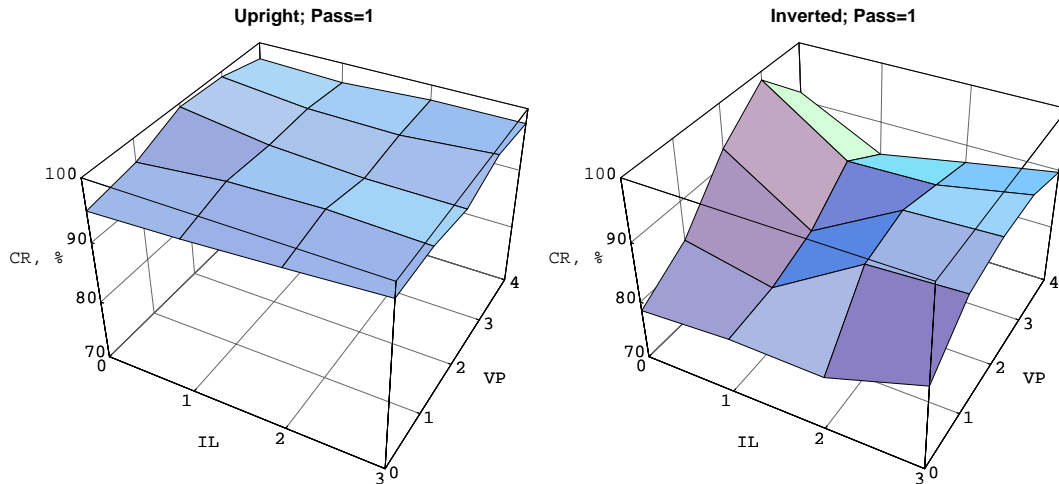


Figure 5: Plots of percent correct (CR) on a given view vs. viewing position (VP) and illumination position (IL). In these plots we consider only the first session of each subject on each set (first pass number). Each point in the graph represent the average CR over all subjects of upright stimuli (left) and inverted stimuli (right). The training view ( $VP=17^\circ$  and  $IL=0$ ) is marked by a star.

The performance of the subjects on the training images was similar for upright ( $99.1 \pm 0.5\%$  correct) and inverted ( $98.6 \pm 0.7\%$  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  $86.5 \pm 1.5\%$  for novel viewing position and  $90.1 \pm 1.6\%$  for novel illumination). For the inverted faces, the

performance for novel viewing position 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. In comparison, for inverted faces, both VP and IL had a significant effect on generalization.

### 2.2.3 Learning

In the previous section we considered only the first pass of the subjects on all sets. In this section we consider the effect of repeated exposure to a specific set of inverted faces and of repeated exposure to inverted faces in general.

We first report the effect of the global pass number, that is, the performance of a subject following repeated exposure to the same face set. We consider each of the following conditions separately: training ( $VP=17^\circ$  and  $IL=0$ ), average over new viewing positions with the training illumination ( $VP\neq 17^\circ$  and  $IL=0$ ), average over new illuminations with the training viewing position ( $VP=17^\circ$  and  $IL\neq 0$ ), and average over all combinations of new illumination and new viewing position ( $VP\neq 17^\circ$  and  $IL\neq 0$ ). Figure 6 shows the average correct rate (CR) over subjects and sets for each of the above conditions vs. the pass number. The plots are separate for upright faces (left) and inverted faces (right). In upright faces, the correct rate was very high to begin with, and there was no effect of pass number. In comparison, for inverted faces, the improvement with pass number was manifest in the increase of mean CR, and, more interestingly, in the near disappearance of the effects of VP and IL.

A distinction worth noting is between learning within a set and learning across sets. The subject is expected to perform better after having seen the same set of faces over and over again. Improvement in performance is also expected when the stimulus class (in this case of inverted faces), rather than a specific stimulus set, persists from session to session. In the following analysis, we distinguished between face-specific learning (within a set), and general learning (across sets).

To investigate the effect of learning across sets, we considered only the first pass of each subject on each set. Clearly, in this case there was no question of learning within a set. Figure 6(a) presents the mean performance of all subjects on the first pass of each set vs. the chronological order of appearance of the sets to a given subject (Set-order). The results do not indicate any learning across sets. That is, the VP and IL effects are 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 with the number of passes of each subject on the first set of faces that the subject saw. Figure 6(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.

We can therefore conclude that the subjects' learning was mostly stimulus-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, a non-specific improvement in the generalization across VP and IL for inverted face images, did not show up significantly in the data. The statistical analysis that supports these conclusions is presented in appendix D.

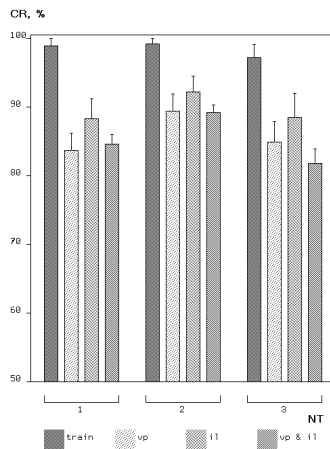


Figure 6: Plots of CR vs. number of different inverted sets that a given subject saw (set number) for inverted stimuli, the average is taken on the first exposure of a subject to each of the sets. The four bars in each group correspond, from left to right, to the following four different conditions. TRAIN: the training view (VP=  $17^\circ$  and IL=0); VP diff: average over all new viewing position with the training illumination (VP $\neq 17^\circ$  and IL=0); IL diff: average over all new illumination with the training viewing position (VP= $17^\circ$  and IL $\neq 0$ ); VP&IL diff: averaged over all combination of new illumination and new viewing position (VP $\neq 17^\circ$  and IL $\neq 0$ ).

## 2.3 Summary of results

Our results indicate that it is possible to discriminate between different inverted faces as well as between upright faces when the the same images are used in the training and in the testing stage. However, the generalization to views obtained under novel viewing position and illumination conditions is significantly worse for inverted faces than for 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, and not for inverted faces in general: following repeated exposure to the same set of

faces, the VP and IL effects almost 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 are prone to change between images are the illumination and viewing position. The first question we addressed was the ability of the human visual system to generalize across changes in illumination and viewing position. The novel views we considered were taken from five different viewing positions and under four different illumination directions. The largest angular separation between novel and familiar viewing positions was  $51^\circ$ , and the largest separation in illumination direction was about  $50^\circ$  away from the direction used in the training image. We found that for upright faces the subjects responded correctly to over 97% of the 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.

The second question we addressed concerned the level at which the human visual system generalizes the identification of faces to novel images. To address this question, we compared generalization for upright and inverted faces. Before turning to discuss the levels of processing and our experimental logic, we summarize briefly a recent result by Moses and Ullman (1992b), which indicates that nontrivial processing must be performed to solve the recognition task posed by our experiments.

Moses and Ullman (1992b) examined the applicability of simple recognition schemes to the problem of face recognition and found that they were insufficient for overcoming the differences between face images due to variations in illumination and viewing position. The simple schemes they considered stored a single image of each object as a model. The recognition process consisted of computing and comparing distances between a new image and all stored images.<sup>2</sup> The performance of these schemes was evaluated on a large database that included the face images used in the present experiment. It was found that image variations induced by illumination and pose changes can be large compared with the differences between face images of different individuals. It follows that a recognition scheme with generalization capacity similar to that of the human visual system cannot be based on relatively simple image comparisons of the types tested, and therefore some more sophisticated processing must be performed by the human visual system.

Turning back to the level at which generalization may take place, we considered upright and inverted face images as two distinct classes of objects. Previous work compared

---

<sup>2</sup>The distances computed included, for example, the  $L_2$  norm of the difference between the images, combined with compensation for shift or compensation for affine transformation of the gray-level values.

memory and processing of faces to those of other classes of objects (see Valentine 1988 for review). One of the main problems with such comparison is the difference between the complexity of images of faces and of other objects. In our work, we were able to treat inverted faces as a different class of objects, which nevertheless has the same complexity as upright faces. We compared generalization to new views of upright faces with generalization to new views of inverted faces, and found that the generalization is impaired in the latter case. This finding has implications regarding the probable level at which face processing is performed by the human visual system.

Consider first the universal processing level. At this level, the system attempts to compensate for the variability among images due to a given imaging parameter in the same manner, for all objects. An example of a universal operation widely used in computer vision is the extraction of edges from the gray-level image. A major goal of this operation is to form an intermediate representation based on image features that are relatively insensitive to illumination. The extraction of edges and lines also appears to be an important part of early visual processing in biological visual systems. Presumably, this processing step is applied uniformly to any input image, and, in particular, to upright and inverted images.

Another universal processing that may be considered, is the extraction of 3-D object shape from a single image based on shading information (Horn and Brooks, 1989).

The difference between the visual system's capacity to generalize across novel viewing position and illumination for upright and inverted faces suggests that the processing underlying this generalization is not performed at the universal level, based on either edge detection or shape from shading.

Consider next the class-based generalization level. By "class" we mean a large (possibly infinite) collection of objects, which is restricted and does not include all possible 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, the class-based level of processing can consist of several levels of processing, depending on the hierarchy of classes that includes the object in question. At the class-based level, the processing depends on the class to which the object in the image is assumed to belong. For example, if the object is assumed to be a face, a class-level processing stage 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).

Our experiments revealed differences in generalization performance between upright and inverted faces suggesting the two generalization is not entirely universal in nature. The result also show substantial recognition capacity from a single (upright) face image.

Together, these results suggest that at least part of the generalization is performed at a class-based level. Class-based processing could be used in several manners. For example, the three dimensional shape of the object (a face in this case) may 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 bilaterally symmetry of a face can be used, for example, for dealing with the viewing position, as described e.g. in Moses and Ullman (1992a) .

Consider, finally, the model-based level of generalization. At the model-based level, the processing of the image depends on the candidate model for the matching, and may also depend on other objects previously seen by the system. Our present results do not lead to a definite conclusion regarding the role of model-based processing in face recognition. Nevertheless, certain model-based approaches, discussed below, can be ruled out based on our data.

One general approach to overcome the differences between images due to the viewing position parameter, uses multiple images of a given face as a model. There are two different ways in which multiple images can be used in a model-based system, 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 will be close to one of the images in the set, considered independently. The interdependent approach is to use 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). Since, in our experiments, only a single image was available to the system in the learning phase, such model-based approaches that relies on several images of the same face can be ruled out with respect to our experiments results.

In conclusion, our results show remarkably good generalization to novel views of upright faces, along with reduced generalization capacity for inverted faces. We suggest that the difference in the generalization performance in the two cases lies at the class level, and that the visual system regards upright and inverted faces as belonging to distinct classes of objects. To substantiate further the claim that class-based processing plays an important role in generalization across pose and illumination changes, the experiments reported here should be repeated with other classes of objects. Specifically, generalization over controlled viewing position, illumination, and other imaging parameters should be compared for different classes of objects, including synthetic objects that do not belong to any familiar class. Furthermore, it would be interesting to determine whether performance for a given class improves with repeated exposure to objects from that 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.

<i>Variable</i>	<i>Levels</i>	<i>Remarks</i>
<i>Invert</i>	0, 1	0=upright, or ↑; 0=inverted or ↓
<i>VP</i>	-34, 17, 0, 17, 34	training: VP=17°
<i>IL</i>	0, 1, 2, 3	IL=0 for left, IL=1 for center, IL=2 for right and IL=3 for left and right together. The training: IL=1
<i>Set</i>	A, B, C, D, E, F	each set consisted of images of 3 faces
<i>Subject</i>	EST,OR1,TAL,ARN JUD,NUR,MOR,OR2	Sets [A,D,F]↑, [B,C,E]↓. Sets [A,D,F]↓, [B,C,E]↑.
<i>Pass-number</i>	1, 2, 3, 4	repetition of each Subject×Set combination
<i>Session</i>	[1..24]	counted separately for each Subject
<i>Set-order</i>	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 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; see section 3).

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*.<sup>3</sup> 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 viewing position and illumination, we performed a 3-way ( $VP \times IL \times Invert$ ) 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 5).

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 viewing position 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  $VP=17^\circ$  (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

---

<sup>3</sup>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).

Duncan Grouping		Mean	N	Subject
----- Invert=0 -----				
	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 ( $VP \times IL \times Pass-number \times Set-order$ ) analysis of covariance (with *Pass-number* and *Set-order* treated as continuous variables).

The results are summarized in Figure 6 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)

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

To find out how the inversion of the stimuli affected the response time RT across changes in viewing position and illumination, we performed a 3-way ( $VP \times IL \times Invert$ ) 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 viewing position  $VP$ , and this dependence was affected by the inversion of the faces (see the  $VP * Invert$  interaction in Table 7, top, and Figure 7).

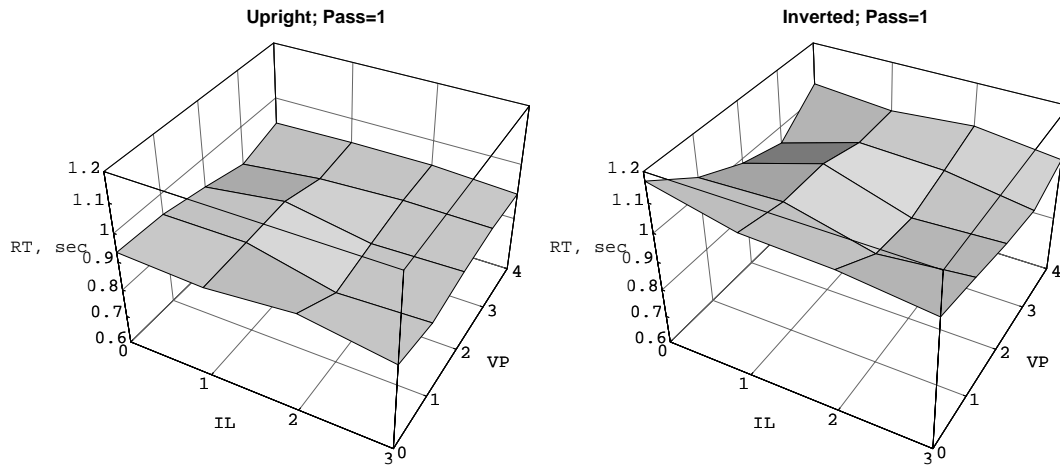


Figure 7: Plots of RT vs.  $VP$  and  $IL$ , for upright stimuli (left) and inverted stimuli (right). The data in this plot are for  $Pass\text{-}number=1$ .

### Effect of $Pass\text{-}number$

Analysis of variance showed that at  $Pass\text{-}number=4$ , the viewing position  $VP$  still significantly affected the response times (see Table 7, bottom). A plot of RT vs.  $VP$  and  $IL$  at  $Pass\text{-}number=4$  appears in Figure 8.

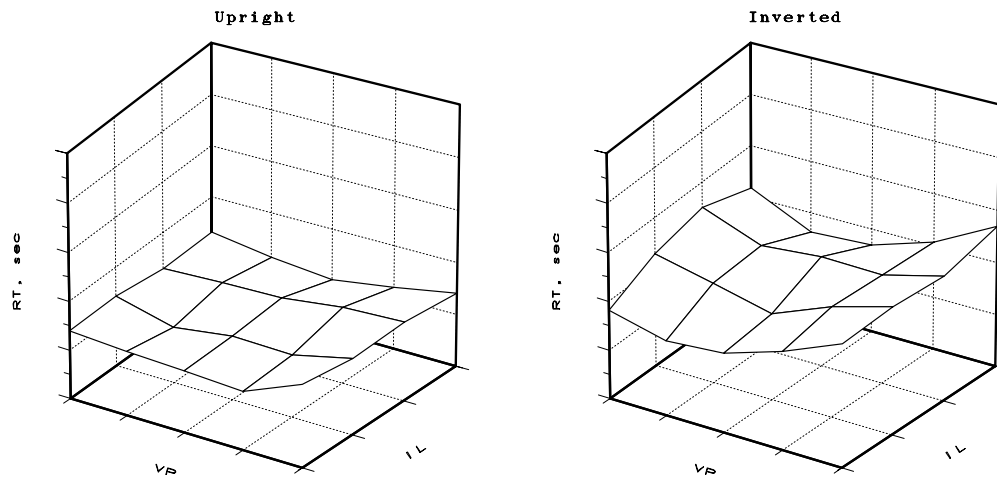


Figure 8: Plots of RT vs.  $VP$  and  $IL$ , for upright stimuli (left) and inverted stimuli (right). The data in this plot are for  $Pass-number=4$  (compare with Figure 7).

-----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 Time=1 and Time=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 Time=4, compared to Time=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 viewing position *VP* on RT is still very strong at *Pass-number=4*.

## Acknowledgments

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 Visual 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 piecemeal to configurational representation of faces. *Science*, 195:312–314.
- Craw, I., Ellis, H., and Lishman, J. (1987). Automatic extraction of face-features. *Pattern Recognition Letters*, 5:183–187.
- Davies, G. M., Ellis, H., and Shepherd, J. W. (1978). Face recognition accuracy as a function of mode of representation. *Journal of Applied Psychology*, 92:507–523.
- Diamond, R. and Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal Experimental Psychology (Gen.)*, 115:107–117.
- 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.
- 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.
- 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.
- Kanade, T. (1977). *Computer recognition of human faces*. Birkhauser Verlag. Basel and 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. (1986). *Perceptual organization and visual recognition*. Kluwer Academic Publishers, Boston, MA.
- Lowe, D. G. (1987). Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence*, 31:355–395.
- Moses, Y. (1993). *Face recognition: generalization to novel images*. PhD thesis, Weizmann Institute of Science.
- Moses, Y. and Ullman, S. (1992a). Limitation of non-model-based recognition schemes. In Sandini, G., editor, *Proc. ECCV-92*, pages 820–828. Springer-Verlag.
- Moses, Y. and Ullman, S. (1992b). Non-neligibile paramaters for face recognition. In *Proc. 9th Israeli AICV Conference*, pages 265–283.
- 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.
- 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.
- 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.