In: Downward Processes in the Perception Representation Mechanisms, C. Taddei-Ferretti and C. Musio (eds), World Scientific, Singapore, New Jersey, London, Hong Kong, 1998, pp 171-188.

GENERALIZATION TO NOVEL VIEWS OF FACES: PSYCHOPHYSICS AND MODELS CONCERNING THE ROLE OF BILATERAL SYMMETRY

NIKOLAUS TROJE
Max-Planck-Institut für biologische Kybernetik
Spemannstr. 38, D-72076 Tübingen, Germany

ABSTRACT

Human faces are approximately bilaterally symmetric. We study the ability to generalize to novel views of human faces focusing on the role of that symmetry. Our hypothesis is that the ability to identify mirror symmetric images is used for viewpoint generalization by approximating the symmetric view of a learned view by its mirror symmetric image. Two psychophysical experiments are performed using a same/different paradigm. Experiment 1 shows that generalization to the symmetric view is better than generalization to otherwise different views. If the symmetric view is replaced by the mirror reversed learning view, performance further increases. Experiment 2 shows that the match between the learned view and the testing image is performed directly on the level of the images. Performance drops significantly if the symmetry between the intensity patterns of learning and testing view is disturbed by an asymmetric illumination, although the symmetry between the spatial arrangement of high-level features is retained. We show that a simple image-based model can explain important aspects of the data and we show how this model can be extended towards a general algorithm for image comparison. Experimental results are discussed in terms of their relation to existing approaches to object recognition.

1. Introduction

A remarkable property of the human visual system is its ability to recognize an object that has been experienced only from a single learning image even if the testing image differs from the learned image in terms of orientation, illumination or other viewing parameters. Human faces form an object class that is of particular interest in this context (Bruce et al., 1987; Moses 1993; Troje & Bülthoff, 1996a). Faces of different people can be very similar to each other compared to objects in other object classes. Human perception, however, seems to be extremely sensitive to even small differences between faces if these differences are diagnostic for identifying the face. On the other hand, we can ignore to a large extent even remarkable differences if these differences are due to orientation, illumination or other attributes that do not concern the identity of the face.

In this paper, we focus on a property that human faces share with many other biologically relevant objects: bilateral symmetry. Bilateral symmetry with respect to a vertical axis is almost universal for any species. With the exceptions of Protozoa and a few immovable animals (e.g., Porifera and some shellfish) and some circular

symmetric species (e.g., Echinodermata), almost all living creatures are approximately bilaterally symmetric. Bilateral symmetry can also be found among plants, in particular, in those species that interact with animals (e.g., pollinators).

Not only is bilateral symmetry universal among animals, but so is the sensitivity of visual systems to symmetric patterns. Preference for symmetric patterns has been shown in a variety of different animals (Lehrer et al. 1994, Møller, 1993, 1995; Swaddle & Cuthill, 1994). Pigeons (Delius & Nowak, 1982), bees (Giurfa et al., 1996), and dolphins (Fersen et al., 1992) have been successfully trained to generalize symmetry. The animals could be trained to respond to either only symmetric or to asymmetric patterns, even if they had not seen the particular pattern before. Humans also show a high sensitivity to symmetrical patterns (e.g. Biederman & Cooper, 1991; Julez, 1971; Wagemans, 1995) as well as to slight deviations from symmetry (Barlow & Reeves, 1979). Sensitivity to bilateral symmetry with respect to a vertical axis is much higher than with respect to a horizontal axis (Corballis & Roldan, 1975; Mach 1903). A variety of different models has been proposed to describe human symmetry detection. A review of the field is provided by two special issues of the journal *Spatial Vision* (Tyler, 1994, 1995).

The ability to identify two mirror symmetric images could be used for viewpoint generalization within classes of bilaterally symmetric objects (Vetter, Poggio, & Bülthoff, 1994; Vetter & Poggio, 1994). If a bilateral symmetric object is viewed from two viewpoints with symmetric positions with respect to the symmetry plane of the object, the resulting images are more or less mirror symmetric to each other. Strictly speaking, the bilateral symmetry of an object is expressed only in the fact that the spatial arrangements of the corresponding features in the two symmetric views are mirror symmetric to each other. Usually the images themselves are also mirror symmetric to each other; however, there are situations in which this is not the case. If, for instance, the object is illuminated by a strong point light source that is located outside the plane defined by the rotation axis of the object and the observer's viewpoint, the greylevel patterns of the images can deviate significantly from mirror symmetry although the mirror symmetry between the spatial arrangements of the features is still retained (Figure 1).

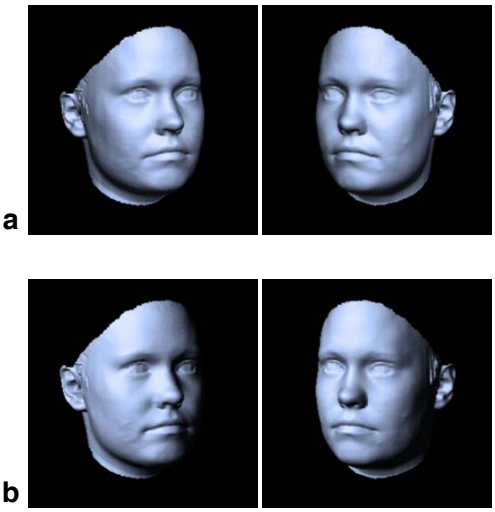

Figure 1: Symmetric views of the same face. If the face is illuminated by a light source positioned in the plane defined by the viewing position and the rotation axis of the face, the resulting images are approximately mirror symmetric to each other (a). If the light source is not within that plane, the two images are no longer mirror symmetric to each other. Only the symmetry between the spatial arrangements of the features is retained (b).

In this paper, we investigate how bilateral symmetry is used in the recognition process. Our hypothesis is that the ability to identify mirror symmetric images is used for viewpoint generalization by approximating the symmetric view of a learned view by its mirror symmetric image. The hypothesis leads to the prediction that the mirror reversed image of a learned view should be recognized better than the realistic symmetric view even in cases in which mirror reversal results in an unrealistic and impossible image of the target face. Furthermore, we want to find out whether we have direct access to the bilateral symmetry of the 3D object by extracting features and using the mirror symmetry of their arrangement, or whether we are restricted to the mirror symmetry between the entire images.

We will first present two psychophysical experiments. In Experiment 1, we measured the increase in generalization performance that can be achieved by exploiting the ability to identify mirror symmetric images. In Experiment 2, we focus on decoupling the bilateral symmetry of the 3D faces from the mirror symmetry between the images yielded by symmetric views by using different lighting conditions. Following these experiments, we present a simple and straightforward model describing the perceptual distance between learning and testing views in a purely image-based framework.

2. General methods

2.1. Stimuli

The images were created using 98 surface models from a data base of 3D head models (for details of the image processing, see Troje & Bülthoff, 1996a). Images had a size of 256 x 256 pixels. The height of the faces on the screen was approximately 6 cm, subtending a region of approximately 4.5 x 4.5 degrees of visual angle at the position of the observer. The faces were rendered without using the original texture information. Instead, homogenous reflectance was assumed and an illumination model was applied. We did this for two reasons. First, we had observed in previous experiments (Troje & Bülthoff, 1996a) that most effects concerning the generalization to novel views of faces are qualitatively the same for naturally textured faces and for faces deprived of texture, but quantitatively much more pronounced for faces without texture. Second, we wanted the changes in the images due to changing illumination in Experiment 2 to be as pronounced as possible. Each face was ren-

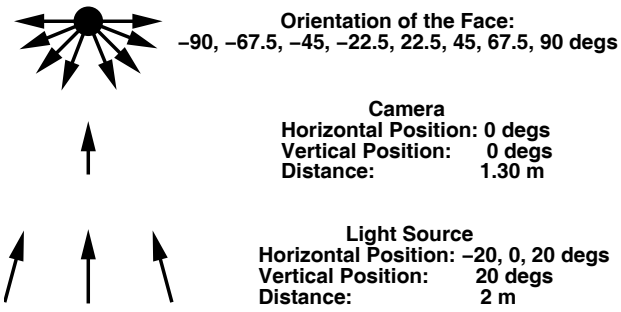

Figure 2: Each face was rendered from eight different view points and with three different light sources. This diagram illustrates the different orientations of the face and the different positions of the light source.

dered in eight different orientations using three different positions for the light source (Figure 2).

2.2. Design and Procedure

We used a same/different recognition task. Subjects were sequentially presented with two images of faces. The task was to judge whether the images showed the same person or not regardless of any change in viewing conditions. The answer was required to be given "as accurately and as quickly as possible".

Each trial was initiated by hitting the SPACE bar on a computer keyboard. A fixation cross appeared for 1000 msec on the screen. Then, the learning view was shown for 700 msec, immediately followed by a random mask. The mask was shown for 1100 msec. After that, the fixation cross appeared again for 1000 msec and finally the testing view was shown. The testing view remained on the screen until the subject responded. For each trial the subject's response and the response time was recorded.

There were four within-subject conditions in each experiment, corresponding to the combination of viewing conditions in the learning and testing images. The conditions themselves varied between the two experiments and are described in more detail for each experiment. Each subject performed 256 trials, 64 in each condition. The 96 face models appeared exactly four times each, once in each condition. Except for this constraint, the assignment of the faces to the different trials was randomized for each subject.

One half of the trials in each condition paired a face with itself (same response) and one half of the trials paired a face with another face of the same gender (different response). In each of the conditions, faces were shown equally often from either of the eight possible orientations. The presentation order of the trials within the experiment was randomized for each subject.

A total of 28 subjects participated in this study. There were 14 subjects in each of the two experiments. They were mainly undergraduate students from Tübingen University and were paid DM 15,- per hour. They were not familiar with the presented faces.

3. Experiment 1

3.1. Purpose

In the first experiment we tried to document and to quantify the viewpoint generalization advantage that can be achieved by exploiting the ability to identify mirror symmetric images. Real human faces are never perfectly bilaterally symmetric and therefore images taken from symmetric viewpoints are not perfectly mirror symmetric. In this experiment, we also tested both symmetric views and perfectly mirror symmetric images of the learned views. Since real faces always have slight asymmetries, the perfectly mirror symmetric image is an impossible and unrealistic

view of the learned face. However, if generalization is based on a virtual view derived by mirror reversing the learned image, the mirror symmetric image should be identified with the learned view better than the actual symmetric view.

3.2. Methods

All stimuli used in this experiment were rendered using a light source positioned above the camera location. Thus, images taken from symmetric viewpoints resulted in roughly mirror symmetric images. The only sources of slight asymmetries between the images are the deviations from a perfect bilateral symmetry of the faces. The four conditions were the following (Figure 2a):

- Condition A: The learning and testing images showed the faces from the same orientation.
- Condition B: The learning and testing images showed the faces from symmetric orientations.
- Condition C: The learning and testing images showed the faces from otherwise different orientations according to the following table:

Learning	Testing	Learning	Testing
+22.5	-45	+67.5	-22.5
-22.5	+45	-67.5	+22.5
+45	-90	+90	-67.5
-45	+90	-90	+67.5

  This scheme yields a mean orientation change (i.e. the angle between learning and testing view) of 112.5 degrees. This is the same mean orientation change as in conditions B and D.
- Condition D: Same as condition B but instead of the symmetric orientation the mirror symmetric image of the learning view was shown as testing view.

In each condition, in half of the trials the same face was shown, in the other half different faces were shown. Note that the distinction between conditions B and D only make sense for the trials showing the same faces.

3.3. Results

We ran ANOVAs on both the error rates and the response times. In addition to the factor coding for the four viewing conditions, we introduced a second factor with two levels indicating whether a trial showed images from the same face or from different faces. The ANOVA for the error rate revealed a reliable main effect for the viewing conditions ($F_{3,39} = 14.45$, $p < 0.01$) and no main effect for the same/different conditions ($F_{1,13} < 1$). In addition, there was a significant interaction between the two factors ($F_{3,39} = 13.09$, $p < 0.01$). For the response times there was an effect for the viewing condition ($F_{3,39} = 11.40$, $p < 0.01$), an effect for the same/different condition ($F_{1,13} = 8.271$, $p < 0.05$), but only a marginal effect for their

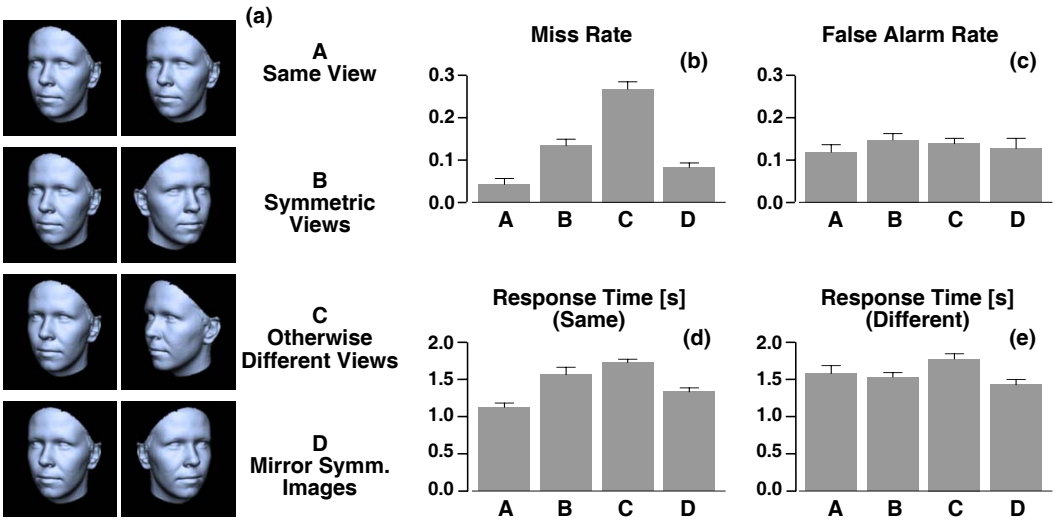

Figure 3: Results of Experiment 1. (a) Conditions. (b) Miss rates. (c) False alarm rates. (d) Response times for the trials showing the same face in learning and testing images. (e) Response times for the trials showing different faces in learning and testing images. Error bars indicate the normalized standard error of the mean.

interaction ($F_{3,39} = 2.655$, $p = 0.06$).

In order to be able to make post hoc comparisons, we also ran separate ANOVAs for the miss rate and the false alarm rate by using either only the trials showing same faces or only the trials showing different faces. The effect of the viewing conditions on the miss rate is significant ($F_{3,39} = 34.9$, $p < 0.01$), the effect on the false alarm is not significant ($F_{3,39} = 0.4$). Similarly, we ran separate ANOVAs for the response times using either only the trials showing same faces or only the trials showing different faces. If same faces were shown, the effect of the viewing conditions was significant ($F_{3,39} = 11.3$). If different faces were shown, there was no significant effect ($F_{3,39} = 2.5$, $p > 0.05$). For post hoc comparisons, we used Tukey's Honestly Significant Difference (HSD). The critical difference for the miss rates was $\overline{d}_T(p=0.05) = 0.0632$ and the critical distance for the response times in the same-face trials was $\overline{d}_T(p=0.05) = 297$ ms.

The results are presented graphically in Figure 3. Error rates and response times are plotted separately for the trials showing the same faces and the trials showing different faces. Figures 3b and c show how miss rates (i.e. the error rate of the trials using same faces) and false alarm rates (trials using different faces) depend on the four conditions. Miss rates were more strongly affected. False alarm rates were more or less constant. Identification of symmetric views was worse than identification of same views ($\overline{d}_{A,B} = 0.0915$, $p<0.05$), but still much better than of otherwise different views ($\overline{d}_{B,C} = 0.1340$, $p<0.05$). Performance in the condition showing mirror symmetric images was better than in the condition showing symmetric views ($\overline{d}_{B,D} = 0.058$). The difference is a little bit smaller than the critical value of the HSD. Since twelve of the fourteen subjects had a smaller error rate in condition D than in condition B, we nevertheless take this difference as reliable. A paired t-test

on the miss rates of conditions B and D yielded a value of $t = 2.49$ ($p < 0.05$).

In Figures 3d and e response times are presented. The pattern of the response times for trials with same faces was very similar to that for the miss rates. The response times for the trials showing different faces differed slightly from the corresponding false alarm rates. Response times were longest in the condition showing completely different views.

3.4. Discussion

Generalization performance to symmetric views of a face is much better than generalization to otherwise different views. There is still a difference, however, between the performance in conditions A (same orientation) and B (symmetric orientation). This difference decreases if we use mirror reversed images instead of symmetric views. We conclude that mirror reversal is perceptually "inexpensive" and causes few additional errors. The difference in performance between conditions A and B is most likely due to deviations from perfect bilateral symmetry in the faces.

Note that the elimination of asymmetries in condition D was not done by eliminating the asymmetries in the 3D head model but by flipping the images. Thus, the heads seen in the learning and the testing image were no longer identical (even if "same" faces were used) but were mirror symmetric copies of each other. Nevertheless, a view of this modified head is treated as being more similar to the learned view than a symmetric view of the identical head. The similarity between the images themselves seems to be more important than the similarity between the 3D objects shown in the images.

Another aspect of the results points in the same direction. The diagrams in Figure 3 show that the differences in performance between the four conditions are mainly due to differences in the miss rates. The false alarm rates are very similar for all conditions. This can be explained if one assumes that subjects match the images rather than the 3D shapes. Images of different people are always different, even if they are shown from the same viewpoint. If discrimination were based on higher order features (e.g., the shape of the nose or the distance between the eyes), then discrimination between different faces shown from the same or from symmetric viewpoints would be expected to be better than if the faces were shown from otherwise different views. Seen from similar viewpoints, differences between the features and their metric relation could be detected more reliably.

4. Experiment 2

4.1. Purpose

Matching the learning and testing images requires some kind of mental transformation between them. What is the nature of this transformation? We consider two possibilities. The transformation could be based on the extraction of parameters

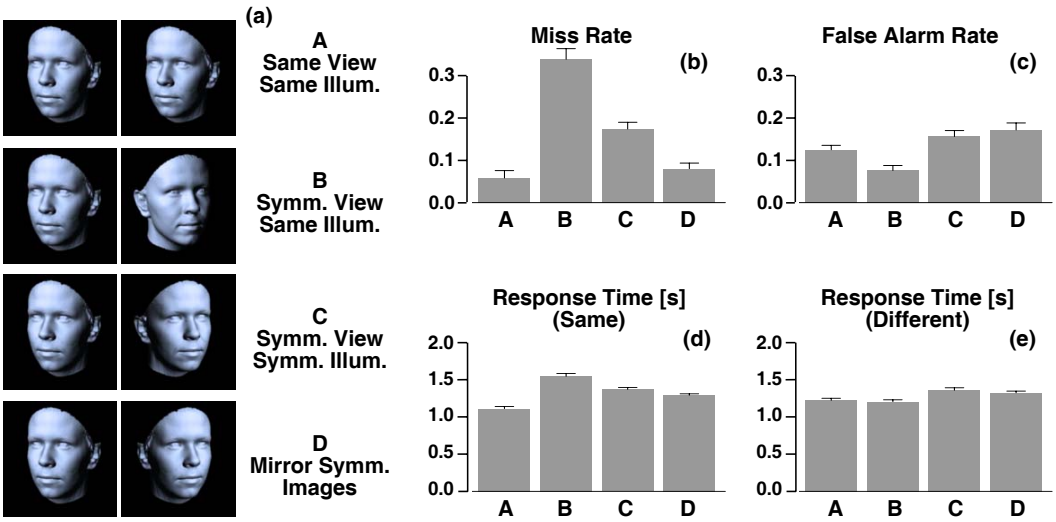

Figure 4: Results of Experiment 2. (a) Conditions. (b) Miss rates. (c) False alarm rates. (d) Response times for the trials showing the same face in learning and testing images. (e) Response times for the trials showing different faces in learning and testing images. Error bars indicate the normalized standard error of the mean.

describing the 3D scene (including information about the 3D shape of the object and illumination conditions) and a subsequent transformation of these parameters. Alternatively, it could be a simple image transformation.

A recognition system based on the extraction of scene-attributes would yield a more flexible recognition system. Information about the 3D scene is inherently viewpoint invariant. However, much computational effort is needed to extract this information. Image transformations, on the other hand, might be performed much faster and with only very basic or even without any knowledge about the content and significance of the image. Some basic knowledge (e.g., about the object class) might be needed to restrict the operation to appropriate situations. The symmetry operation, for instance, might help to generalize to the symmetric view of a bilaterally symmetric object, but it is otherwise of limited value.

With Experiment 2 we want to find out on what perceptual level symmetry information is processed. Do we only use the mirror symmetry between two images or can we extract the bilateral symmetry of the 3D face from the image and use it for recognition? In Experiment 1, stimuli were generated simulating a light source above the location of the camera. Consequently, the symmetric views always resulted in more or less mirror symmetric images. In Experiment 2, however, we used images that were rendered simulating a light source that was no longer positioned above the camera, but was instead 20 degrees either to the right or to the left of it. This allowed us to decouple the bilateral symmetry of the 3D face from the mirror symmetry between the images taken from symmetric viewpoints. Images taken from symmetric viewpoints but with a fixed position of the light source are no longer mirror symmetric to each other on the level of the pixel intensities. The symmetry between the spatial arrangements of the features in the face, however, is still retained.

4.2. Methods

As in the previous experiment the eight different orientations in which faces were shown were completely counterbalanced. In addition, the relative light source position was counterbalanced. In half of the trials the learning view was rendered such that the light came from the side towards which the face was looking and in the other half of the trials it came from the contralateral side. The orientation and the illumination in the testing view was determined by the following four conditions (Figure 4a):

- Condition A: The learning and testing images showed the faces in the same orientation and with the same illumination.
- Condition B: The learning and testing images showed the faces in symmetric orientations with the position of the light source fixed at one side.
- Condition C: Both the orientation of the face and the position of the light source were changed to their symmetric positions.
- Condition D: The mirror symmetric image of the learning view was shown as the testing view.

4.3. Results

As in Experiment 1, we calculated 4 x 2 ANOVAs modelling the error rates and the response times. The first factor accounted for the four symmetry conditions and the second was introduced to indicate whether a trial showed images from the same face or from different faces (cf. section 3.3). The symmetry conditions had a significant effect on error rates ($F_{3,39} = 16.01$, $p < 0.01$) and response times ($F_{3,39} = 17.87$, $p < 0.01$). The same/different conditions had no effect on error rates ($F_{1,13} = 1.22$) but did affect response time ($F_{1,13} = 6.117$, $p < 0.05$). The interaction between the two factors significantly affected both the error rate ($F_{3,39} = 34.13$, $p < 0.01$) and the response time ($F_{3,39} = 18.65$, $p < 0.01$).

As a basis for post hoc comparison, we also ran separate ANOVAs on the miss rate and on the false alarm rate as well as on the response times for the trials using same and the trials using different faces. All effects were significant ($F_{3,39} = 35.9$, $p < 0.01$, for the miss rates; $F_{3,39} = 7.4$, $p < 0.01$, for the false alarm rates; $F_{3,39} = 30.6$, $p < 0.01$, for the response times when using same faces; $F_{3,39} = 5.491$, $p < 0.01$, for the response times when using different faces).

For post hoc comparisons Turkey's Honestly Significant Difference (HSD) was calculated, revealing a value of $\overline{d}_T(p=0.05) = 0.081$ for the miss rates, a value of $\overline{d}_T(p=0.05) = 0.059$ for the false alarm rates, a value of $\overline{d}_T(p=0.05) = 123$ ms for the response times when using same faces, and a value of $\overline{d}_T(p=0.05) = 120$ ms for the response times when using different faces.

Figures 4b and c show miss and false alarm rates. As in Experiment 1, the miss rate varied to a much greater degree across the four symmetry conditions than did the false alarm rate. The miss rate was much higher when only the orientation of the

face changed (condition B) than when both orientation and illumination changed (condition C) ($\bar{d}_{B,C} = 0.165$, $p<0.05$). Showing mirror symmetric images (condition D) instead of symmetric viewing conditions (condition C) further lowered the miss rate ($\bar{d}_{B,C} = 0.094$, $p<0.05$) to a value that was statistically indistinguishable from the miss rate yielded when using identical viewing conditions (condition A) ($\bar{d}_{B,C} = 0.0224$). The false alarm rates, although less affected by the symmetry condition, showed almost the opposite pattern. The false alarm rate in condition B, in which we recorded the highest miss rate, was significantly lower than in the other conditions.

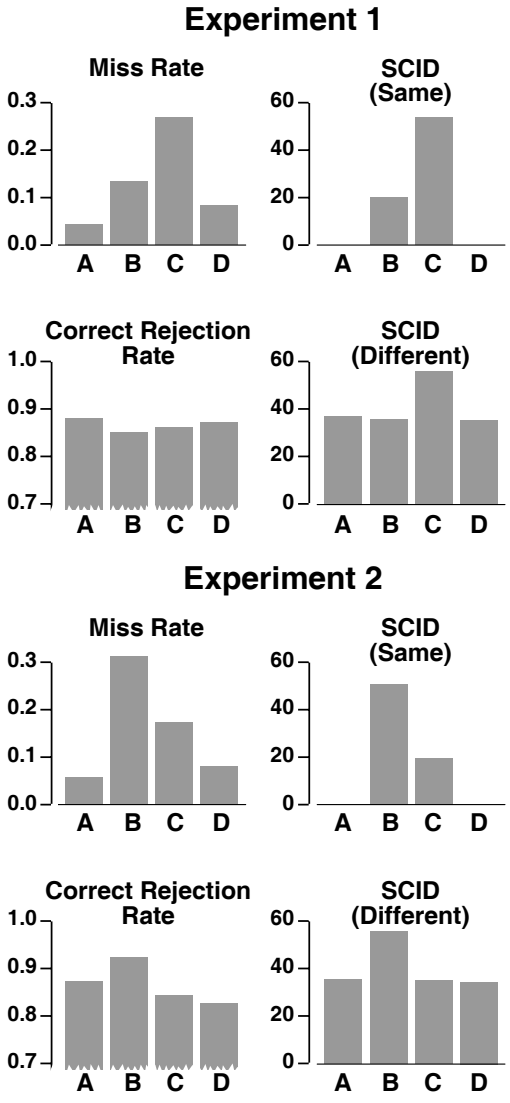

Figure 5: On the left side, error rates from both experiments are plotted. Instead of the false alarm rates, the correct rejection rates (CR = 1-FA) are shown. On the right side, the corresponding mean SCIDs are plotted. The SCIDs were

Figures 4d and e present the response times. As in Experiment 1, response times for the trials presenting the same face in learning and testing image followed the same pattern as the miss rates. The response times corresponding to the trials showing different faces were much more constant but still reflected the pattern of the false alarm rates.

4.4. Discussion

The distance between learning and testing views in the four conditions can be described either in terms of higher order attributes of the scene depicted in the image, or in terms of an image similarity. In condition A, in which the images were presented with identical viewing conditions, the distance is smallest in both cases. In terms of a scene-based description, the distance is larger in condition B with one attribute changing (the orientation of the face), but still smaller than in condition C in which two attributes change (the orientation of the face and the location of the light source). The changes between learning and testing view in condition D can also be described in terms of scene parameters. Here are not only the

orientation of the head and the position of the illumination different, but also the face itself. The face presented in the testing image is no longer identical to the one seen in the learning image, but it is the mirror reversed version of it.

The order of the distances in conditions B, C and D is reversed if we use an image-based distance measure that is insensitive to mirror reversal. The distance between learning and testing image in condition D, in which the images are perfectly mirror symmetric, is as small as in condition A. In condition C, in which the images deviate only slightly from mirror symmetry, the distance is still small, and in condition B, it is greatest.

The data clearly show a pattern that is consistent with an image-based distance rather than a distance based on scene-attributes. The difference in performance between conditions A (identical viewing conditions) and condition C (symmetrical orientation and symmetrical illumination) is most likely due to slight asymmetries between the resulting images. The symmetry operation itself causes only a very small increase in error rates.

5. Modelling the image distance

5.1. "Symmetry corrected image distance"

The experiments showed that human performance is strongly influenced by an image-based distance measure that is almost insensitive to mirror reversal. In this section, we explicitly formulate such a distance measure that is very simple and straightforward, but nevertheless accounts for some important aspects of our data.

In the following simple model the "Symmetry Corrected Image Distance" (SCID) between two images A and B is defined as the minimum of the Euclidian distance in pixel space between the two images and the Euclidian distance between one of the images and the mirror reversed version of the other image. The mirroring operation itself might also contribute a small additional error which is accounted for by the factor c_{sy} :

$$\mathrm{SCID} = \mathrm{Min}(\mathrm{D}(A, B), c_{sy} + \mathrm{D}(A, \mathrm{sy}(B)))$$

$$\mathrm{D} = \sqrt{\frac{\sum (A_i - B_i)^2}{n}} \qquad (1)$$

A_i and B_i denote the intensities of the ith pixel in image A and B, respectively.

In Figure 5 the data from Experiments 1 and 2 are compared with the corresponding mean SCIDs calculated from the same pairs of images that were shown to the subjects. Instead of plotting the false alarm rates, the correct rejection rates (CR = 1 - FA) are plotted. Note that in order to keep the scale comparable, the ordinate ranges from 0.7 to 1.0. The calculation of the SCIDs was performed by assuming zero cost for the flipping operation ($c_{sy} = 0$).

The patterns of the diagrams showing the miss rate and the corresponding

SCIDs are very similar. One could easily account for the differences by considering a constant offset (c_{err} , even if the images are identical, we still make some errors) and by assuming that the cost of the symmetry operation (c_{sy}) is small, but still larger than zero. A linear model describing the miss rate (Miss) as a function of the transition from one image to the other by using the SCID is then formulated by

$$\text{Miss} = \lambda(c_{err} + \text{SCID}) \qquad (2)$$

λ is a scaling factor. By fitting such a linear model, the parameters $\lambda = 0.0045$, $c_{err} = 11.11$, and $c_{sy} = 4.90$ are yielded. The regression is highly significant ($r = 0.985$, $F_{2,9} = 146.8$, $p < 0.01$).

Although this linear relation seems to be striking, it cannot account for the entire behaviour of the subjects in these experiments. It models the miss rates but not the correct rejection rates, making it an incomplete model of the subject's decision. Nevertheless, it illustrates the general idea of an image distance corrected for symmetry.

5.2. A more general distance measure

The SCID as formulated above is based on a simple Euclidian distance between two images in pixel space. In addition it allows for a "shortcut". If a part of the distance can be accounted for by the symmetry operation, only the remaining distance is evaluated (Fig. 6). It is easy and straightforward to formulate an image-based distance measure that includes additional shortcuts in an analogous way, each corresponding to a particular image operation. Such shortcuts can be introduced for operations such as translation or scaling as well as for changes in overall lightness or contrast of the image. They could even be formulated for much more complex operations such as the image deformations that are due to rotation in depth (Vetter & Poggio, 1996), the deformations that result from changes in facial expression (Beymer & Poggio, 1996), or the changes corresponding to a change in illumination (Belhumeur & Kriegman, 1996; Hallinan, 1994). These latter operations are highly model specific. The operation corresponding to the rotation in depth of a face could not be applied to other objects. Applying it to a face requires specific knowledge about the shape of faces in general and the arrangement of their features. This is the point at which top-down processing comes into play. We must first classify an

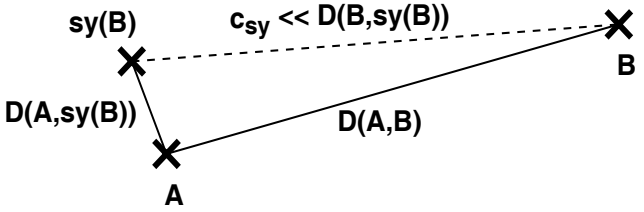

Figure 6: Symmetry corrected image distance. There are two ways to get from image B to image A. Each way is associated with a certain cost. The image distance between A and B is defined by the cost along the most inexpensive way. The cost to get from one point in the image space to another is generally defined by the Euclidian distance between the two points. However, there is a shortcut. The cost for the distance between two mirror symmetric images is much lower than the Euclidian distance between these images.

object as belonging to a certain object class before we can use sophisticated algorithms for individual identification.

Analogous to a small cost for the symmetry operation (c_{sy} in Eq. 1) that contributes to the distance measure, each of the other image operations may have an associated cost function that is determined by the computational effort and the time needed to perform the operation and by top-down knowledge about the whole object class. In addition the parameters determining the details of an operation can influence the cost function. Mirroring along a vertical axis seems to be less expensive than mirroring along a horizontal or an oblique axis. A large translation might be more expensive than a small translation.

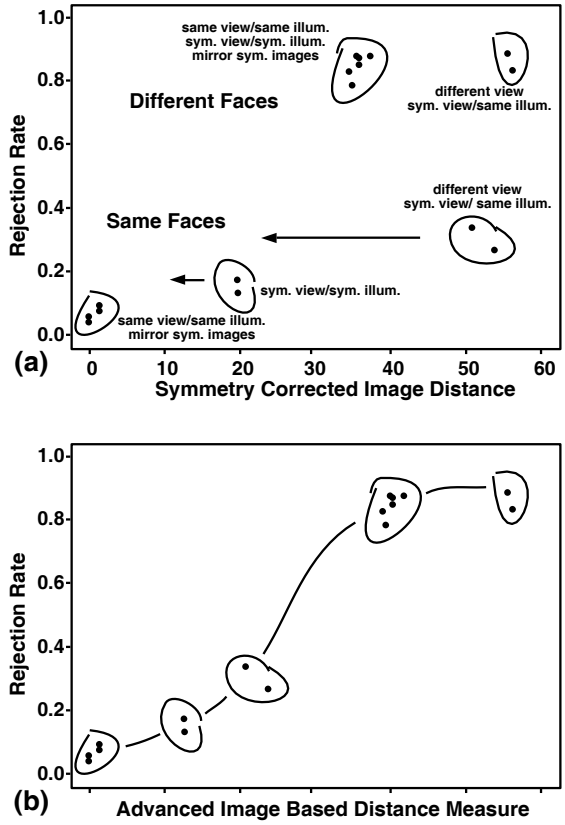

Figure 7: Rejection rates and the corresponding SCIDs (a). The arrows indicate which data will be changed if a distance measure was used, that includes additional image operations. Such an improvement would yield a distance measure that is monotonically related to the subjects response behaviour (b).

Extending the SCID by incorporating additional image operations would not alter the distance between the image pairs in our experiments if they were identical or approximately mirror symmetric. It would also not change very much in cases in which learning and testing images show different faces. They might be able to compensate for orientation and illumination differences but the difference between the faces themselves would still cause a large distance. The main effect of introducing other image transformations would occur for images showing the same face in different viewing conditions as in condition C in Experiment 1 and in condition B in Experiment 2.

Figure 7a replots the rejection rates (miss rates and correct rejection rates) of all conditions in both experiments against the SCID as formulated above. As already mentioned, the SCID cannot account for the entire behaviour of the subjects. The SCID corresponding to pairs of different faces shown under identical viewing conditions is smaller than the SCID corresponding to pairs of images showing the same face in different viewing conditions. The rejection rates are nevertheless higher if different faces are shown. Figure 7b shows how the diagram might change if instead of the SCID a

more sophisticated images distance based on additional image transformations as discussed above were used. The relation between this hypothetical distance measure and the rejection rate now becomes a monotonic function and could thus account for the subject's decision.

6. General discussion

Human face recognition is to a large extent stable against changes concerning the viewing conditions. At first sight, this seems to support theories of object recognition (including face recognition) that are based on the extraction of high-level features that are invariant to the particular viewing condition. On the other hand, evidence has been reported that objects are represented in terms of two-dimensional views (Bülthoff & Edelman, 1992, 1993).

The experiments presented in this paper exhibit both aspects. In general, subjects are able to identify two images of the same person even if the viewing conditions are very different and they can discriminate images of different persons even if the viewing conditions are identical. Nevertheless, errors are made and the error rate is strongly influenced by image-based parameters.

We showed that the mirror reversed image of a learned view of a face is in fact used as an approximation of the symmetric view of the face. Two views are treated as if they are taken from the same face when the images are mirror symmetric to each other. Bilateral symmetry of the 3D face as expressed by the mirror symmetry between the spatial arrangements of the features in the images does not appear to be exploited. Information processing as expressed in these experiments seems to be mainly image-based.

This makes the whole notion of "generalization" to a novel view or a novel illumination questionable. This notion suggests that generalization to a new instance of a scene-attribute is accomplished and can be measured independently of other attributes. Recognition seems, however, not to be based on the extraction of scene-attributes but rather on an image-based comparison between the learned and the tested instance of an object.

If we nevertheless try to describe recognition performance as being affected by different scene-attributes, we have to be aware of very prominent interactions between these attributes. The situation in Experiment 2 provides an example. Generalization to a new view causes a decrease in recognition performance (compare conditions A and B). The same is true if subjects have to generalize to a new illumination (Braje et al., 1996; Troje & Bülthoff, 1996b). If both attributes are changed, the effects of the changes in both attributes do not add but rather partly cancel each other. If the orientation of the face had already been changed, then a related change in illumination could lead to an increase in performance (compare conditions B and C). Subjects do not generalize to new instances of scene-attributes. They compare images. The way this comparison is accomplished, however, reflects an adaptation

to the requirements of recognizing objects under changing viewing conditions.

We outlined a way to formulate such an image distance. The distance between two images was defined by the most cost efficient way to transform one image into the other one. Mirror reversal was treated as a shortcut that contributed little to the distance measure.

Do we treat mirror reversal as being so inexpensive only when we know that we are dealing with a bilaterally symmetric object? Is the cost associated with this operation only so small because we already know that we are confronted with faces? Or do we take into the bargain the false identification of two asymmetric objects that are mirror symmetric to each other? The results of our experiments do not provide an answer to these questions. The likelihood that we came into a situation in which such a false identification could happen is, however, so small that we probably could easily afford the assumption that two mirror symmetric images show the same object from symmetric views. The only case in which this becomes a problem is a very modern one compared with the time scale relevant for the evolution of our cognitive system. The Latin alphabet has some letters such as b and d or p and q that would be confused by our model. In fact, young children confuse these letters more frequently than others.

Different models of object recognition have been developed in the past. How do our findings relate to these approaches? Ullman (1989) classified current and past models of object recognition into three major groups: (1) invariant properties methods, (2) parts decomposition methods, and (3) alignment methods. This classification scheme focuses on the way objects are represented and how these representations are matched. Invariant properties methods are based on a representation of the object in terms of higher order features. The features should fulfil the following criteria: (a) they can be derived from the image, (b) they are to a large extent independent of the viewing conditions, and (c) they are diagnostic, that is, they are shared by all views of the object but not by views of other objects. Use of such features would be ideal for solving the recognition task, but in practice they are not easy to find. Parts decomposition methods cope with this problem by decomposing objects into generic parts that are so simple that it is easier to find invariants for each of them. Alignment methods, finally, are based on pictorial descriptions. The basic idea is to compensate for the transformations separating the viewed object and the corresponding stored model and then compare them.

Our data can best be interpreted within the framework of an image-based alignment approach (Ullman, 1989; see also Poggio & Edelman, 1990). They are not consistent with invariant properties methods (e.g. Pitts & McCulloch, 1947) or with parts decomposition methods (e.g. Biederman, 1985). These descriptions assume the extraction of features (or parts) such as eyes, nose and mouth. An easily derivable property that is invariant with respect to symmetric views (even with nonsymmetric illumination) would be metric information about the relationship of the locations of such features. However, such information seems not to be used.

Descriptions based on either invariant properties or on parts decomposition should not change when lighting changes. The distance between the learning and the testing image in conditions B and C in Experiment 2 should be about the same. The subjects response, however, indicated that this was not the case.

We are aware that the present results and the conclusions drawn from them might be restricted to the same/different paradigm that we used in these experiments. Learning and testing views were shown immediately one after the other with only a 2 second interval between them. Only short term episodic memory is needed to perform this task. The visual representations used to perform other tasks might be organized in a completely different way. The face of a well known friend might well be represented using invariant properties or models of the entire 3D structure, and it might be worthwhile to run experiments similar to the ones presented here but with different recognition paradigms that address different kinds of memory.

References

Barlow, H.B. and B.C. Reeves (1979) "The versatility and absolute efficiency of detecting mirror symmetry in random dot displays", *Vision Research* 19:783-793.

Belhumeur, P. and D. Kriegman (1996) "What is the set of images of an object under all possible lighting conditions?", in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 270-277.

Beymer, D. and T. Poggio (1996) "Image representation for visual learning", *Science* 272:1905-1909.

Biederman, I. (1985) "Human image understanding: Recent research and a theory", *Computer Vision, Graphics and Image Processing* 32:29-73.

Biederman, I. and E.E. Cooper (1991) "Evidence for complete translational and reflectional invariance in visual object priming", *Perception* 20:585-593.

Braje W.L., D. Kersten, M.J. Tarr and N. Troje (1996) "Illumination and shadows influence face recognition", *Investigative Ophtalmology and Visual Science* 37:S176.

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

Bülthoff, H.H. and S. Edelmann (1992) "Psychophysical evidence for a two-dimensional view interpolation theory of object recognition", *Proc. Natl. Acad. Sci. USA* 89:60-64.

Bülthoff, H.H. and S. Edelmann (1993) "Evaluating object recognition theories by computer graphics psychophysics", in: *Exploring Brain Functions: Models in Neuroscience*, T.A. Poggio and D.A. Glaser, eds., John Wiley and Sons Ltd.

Corballis, M.C. and C.E. Roldan (1975) "Detection of symmetry as a function of angular orientation", *Journal of Experimental Psychology: Human Perception and Performance* 1:221-230.

Delius, J.D. and B. Novak (1982) "Visual symmetry recognition in pigeons", *Psy-*

chol. Res. 44:199-212.
Fersen, L. von, C.S. Manos, B. Galdowski and H. Roitblat (1992) "Dolphin detection and conceptualization of symmetry", in: *Marine Mammal Sensory Systems*, J. Thomas, ed., New York: Plenum Press, pp. 753-762.
Giurfa, M., B. Eichmann and R. Menzel (1996) "Symmetry perception in an insect", *Nature* 382:458-461.
Hallinan, P.W. (1994) "A low-dimensional representation of human faces for arbitrary lighting conditions", in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 995-999.
Julez, B. (1971) *Foundations of Cyclopean Perception.* Chicago: University of Chicago Press.
Lehrer, M., G.A. Horridge, S.W. Zhang and R. Gadagkar (1994) "Shape vision in bees: Innate preference for flowerlike patterns", *Phil. Trans. R. Soc. Lond. B* 347:123-137.
Mach, E. (1903) *Analyse der Empfindungen*, Jena: Fischer.
Møller, A.P. (1993) "Female preference for apparently symmetrical male sexual ornaments in the barn swallow Hirundo rustica", *Behav. Ecol. Sociobiol.* 32:371-376.
Møller, A.P. (1995) "Bumblebee preference for symmetrical flowers", *Proc. natn. Acad. Sci. U.S.A.* 92:2288-2292.
Moses, Y. (1993) "Face recognition: generalization to novel images", *Applied Math. and Computer Science, The Weizmann Institute of Science, Israel, Ph.D. Thesis.*
Pitts, W. and W.S. McCulloch (1947) "How we know universals: The perception of auditory and visual forms", *Bulletin of Mathematical Biophysics* 9:127-147.
Poggio, T. and S. Edelman (1990) "A network that learns to recognize three-dimensional objects", *Nature* 343:263-266.
Swaddle, J.P. and I.C. Cuthill (1994) "Preference for symmetric males by female zebra finches", *Nature* 367:165-166.
Troje, N. and H.H. Bülthoff (1996a) "Face recognition under varying pose: The role of texture and shape", *Vision Research* 36:1761-1771.
Troje, N. and H.H. Bülthoff (1996b) "What is the basis for good performance to symmetric views of faces?", *Investigative Ophtalmology and Visual Science* 37:S194.
Tyler, C.W. (Ed.). (1994) "The Perception of Symmetry, Part I: Theoretical Aspects [Special Issue]", *Spatial Vision* 8 (4).
Tyler, C.W. (Ed.). (1995) "The Perception of Symmetry, Part II: Empirical Aspects [Special Issue]", *Spatial Vision* 9 (1).
Ullman, S. (1989). Aligning pictorial descriptions: an approach to object recognition", *Cognition* 32*:*193-254.
Vetter, T. and T. Poggio (1994) "Symmetric 3D objects are an easy case for 2D object recognition", *Spatial Vision* 8:443-453.
Vetter T., T. Poggio and H.H. Bülthoff (1994) "The importance of symmetry and

virtual views in three-dimensional object recognition", *Current Biology* 4:18-23.
Vetter T., and T. Poggio (1996) "Image Synthesis from a Single Example Image", in: B. Buxton and R. Cippola, eds., *Computer Vision -- ECCV'96, Lecture Notes in Computer Science* 1064, Cambridge UK: Springer, pp. 652-659.
Wagemans, J. (1995) "Detection of visual symmetries", *Spatial Vision* 9:9-32.