Bilateral symmetry of human faces helps to generalize to novel views

N. F. Troje and H. H. Bülthoff

Max−Planck−Institut für biologische Kybernetik • Tübingen

1 INTRODUCTION

Recognition of a human face seen from a view symmetric to the learned view seems to be almost as easy as recognizing a face from the learned view [1]. We present two psycho−physical experiments using a SAME/DIFFERENT paradigm. They are designed to verify this observation and to examine its nature. In particular, we investigate, whether it is based on the bilateral symmetry of the 3D object or on the resulting mirror symmetry of the images.

The question can be posed in more general terms, asking whether the comparison between two views is based on:

- Extraction of viewpoint invariant scene attributes (e.g. orientation & shape of the head, position of the light source)

or

- Matching of the images using simple image transformations (e.g. translation, mirror reversal)

EXPERIMENT 1 tests the informal observation made in [1] under more controlled conditions.

- Is recognition of symmetric views really better than recognition of otherwise different views?

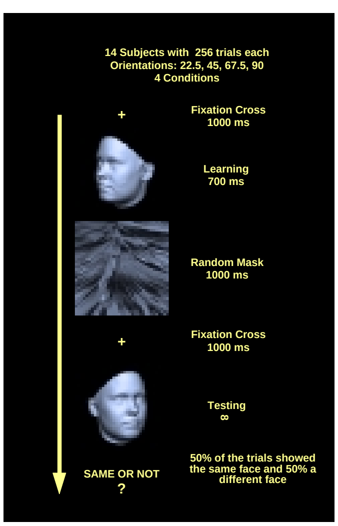
If this is the case and if the reason for it is the approximate mirror symmetry between the images, then recognition should be even better if, instead of the symmetric views, perfectly mirror symmetric images are used.

In EXPERIMENT 2 bilateral symmetry of the 3D face and mirror symmetry of the images taken from symmetric viewpoints were decoupled by using a light coming from either 20 degrees left or right of the simulated camera.

The distances between learning and testing views in the four conditions are different depending on whether they are based on a scene−based or on an image−based representation.

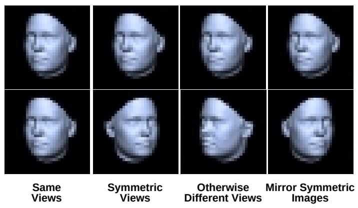
SCENE−BASED	IMAGE−BASED
nothing changed	nothing changed
pose changed	different images
pose & illumination changed	almost mirror symmetric images
pose, illumination & face changed	perfectly mirror symmetric images

2 METHODS

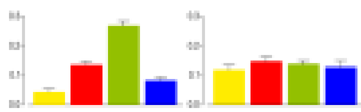


3 RESULTS

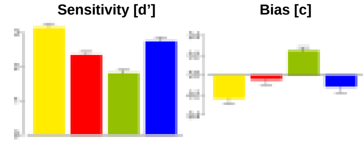
Experiment 1



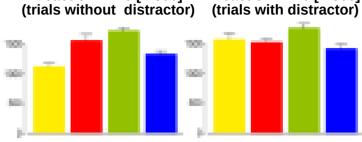
Miss Rate / False Alarm Rate



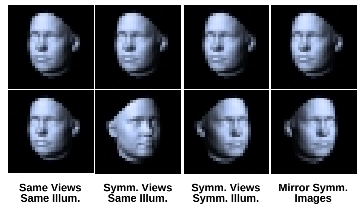
Sensitivity [d'] / Bias [c]



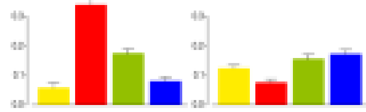
Reaction Time [msec] (trials without distractor) / Reaction Time [msec] (trials with distractor)



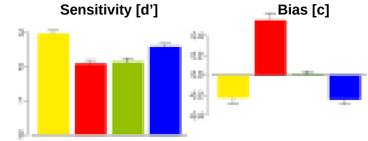
Experiment 2



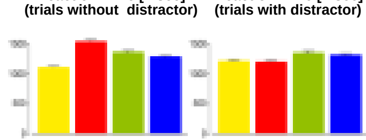
Miss Rate / False Alarm Rate



Sensitivity [d'] / Bias [c]



Reaction Time [msec] (trials without distractor) / Reaction Time [msec] (trials with distractor)



The measures for sensitivity (d') and response bias (c) are the ones used in signal detection theory [2].

Generalization performance to the symmetric view is much better than generalizing to otherwise different views. Performance gets even better if, instead of symmetric views, mirror symmetric images are used.

The same/different decision is based on image similarity rather than on a scene−based representation. The testing view in the fourth condition in both Experiment 1 and 2 is an impossible view of the face seen in the learning view. Nevertheless, it is considered to be more similar to the learning view than the realistic symmetric view.

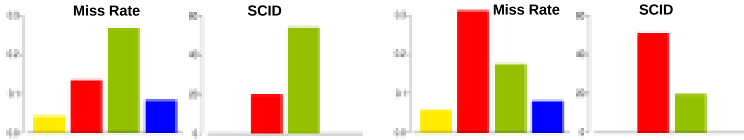
4 SYMMETRY CORRECTED IMAGE DISTANCE

The generalization performance seems to be mediated by a measure of image similarity that is virtually insensitive to mirror reversal. A very simple approach towards formulating such a "Symmetry Corrected Image Distance" (SCID) between two images A and B is taking the minimum of the Euclidian distance in pixel space D(A,B) and the Euclidian distance between one image and the mirror reversed other image. The mirror reversal itself might also account for some cost c_{sy}.

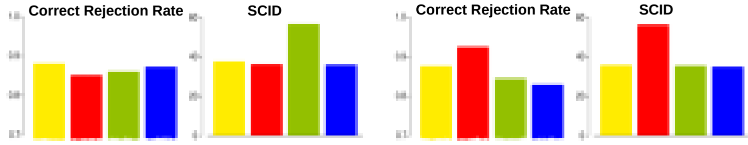
$$\text{SCID} = \text{MIN}\big(D(A,B),\ c_{sy} + D(A, sy(B))\big) \qquad D(A,B) = \sqrt{\frac{\sum (A_i - B_i)^2}{n}}$$

We do not expect this distance measure to be the one human perception is using. It does, however, model some parts of the data very well.

Miss Rate / SCID / Miss Rate / SCID



Correct Rejection Rate / SCID / Correct Rejection Rate / SCID



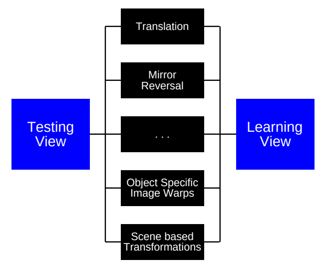
5 DISCUSSION

- The SAME/DIFFERENT performance can be better described in an image−based than in a scene−based framework.
- We are much more sensitive to asymmetries between the images than to inconsistencies between the 3D shapes.

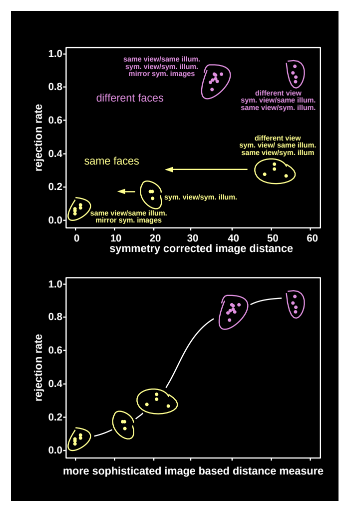
Human perception achieves invariance not only to mirror reversal but also to other image operations. Invariances can also be graded. An image operation might be "inexpensive" but not completely free of charge. A more realistic measure of image distance might be very similar to the proposed SCID, but it should take more image operations into account, each associated with a certain cost.

In such a model, the cost of an operation would be determined not only by its computational effort but also by top−down processes mediating knowledge about the whole object class. Mirror reversal should be cheap for objects that are known to be bilaterally symmetric (like faces) but not for nonsymmetric objects (e.g. letters).

The model could even account for complex nonlinear image transformations, such as the image deformations corresponding to a rotation in depth [3].



The difference between such an improved distance measure and the simple SCID measure would reduce the distance between images taken from the same face, but not from symmetric viewpoints.



6 REFERENCES

[1] Troje, N.F. and Bülthoff, H.H. (1996) Vision Research 36:1761−1771.

[2] Macmillan, N.A. and Creelman, C.D. (1991) Detection Theory. Cambridge University Press.

[3] Vetter, T. (1996) Synthesis of novel views from a single face image. MPI−Memo #26.