Generalizing across viewpoint changes: evaluation of a computational method with respect to the human face recognition system

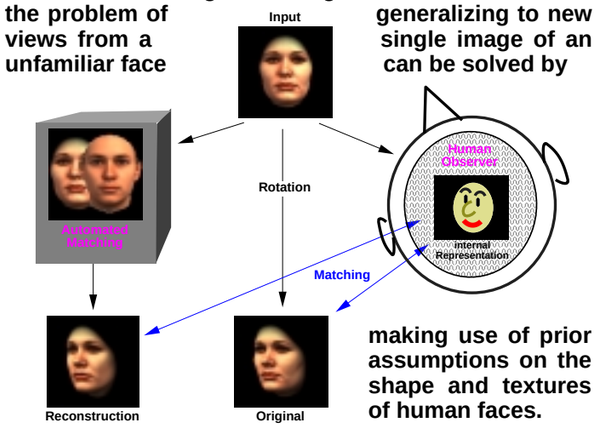
1167

Barbara Knappmeyer[1], Volker Blanz[2] and Heinrich H. Bülthoff[1]

[1]barbara.knappmeyer/heinrich.buelthoff@tuebingen.mpg.de Max−Planck−Institute for Biological Cybernetics, Tübingen, Germany [2]volker@informatik.uni−freiburg.de Albert−Ludwigs−University, Freiburg, Germany

1 PURPOSE

Blanz and Vetter (SIGGRAPH 1999) developed a computational method to estimate the 3D structure of a head from a single 2D image. It demonstrates how the problem of generalizing to new views from a single image of an unfamiliar face can be solved by



making use of prior assumptions on the shape and textures of human faces.

How does the performance of the human face recognition system relate to the performance of the reconstruction algorithm ?

2 METHOD

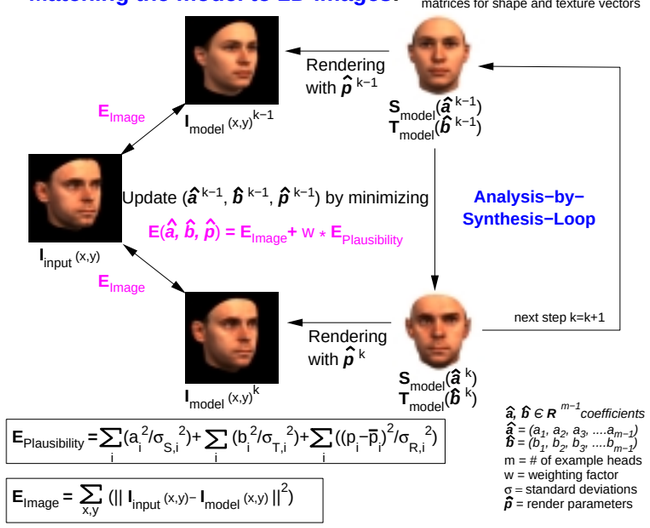
Reconstruction Algorithm

Morphable 3D face model: The morphable face model is based on laser scanned heads (Cyberware™) that are all in correspondence. It is defined as the set of new shapes and textures $\{S_{new}(\hat{a}), T_{new}(\hat{b})\}$ with:

$S_{new} = \bar{S} + \sum a_i \cdot s_i \quad T_{new} = \bar{T} + \sum b_i \cdot t_i$

$\bar{S}, \bar{T}$ = mean of shape and texture
t_i, s_i = eigenvectors of the covariance matrices for shape and texture vectors

Matching the model to 2D images:



$\hat{a}, \hat{b} \in R^{m-1}$ coefficients
$\hat{a} = (a_1, a_2, a_3, \ldots a_{m-1})$
$\hat{b} = (b_1, b_2, b_3, \ldots b_{m-1})$
m = # of example heads
w = weighting factor
σ = standard deviations
$\hat{p}$ = render parameters

$$E_{Plausibility} = \sum_i (a_i^2/\sigma_{S,i}^2) + \sum_i (b_i^2/\sigma_{T,i}^2) + \sum_i ((p_i - \bar{p}_i)^2/\sigma_{R,i}^2)$$

$$E_{Image} = \sum_{x,y} (|| I_{input}(x,y) - I_{model}(x,y) ||^2)$$

Stimuli

Reconstructions: Images rendered from 100 laser scanned heads served as input for the reconstruction algorithm. A different set of 100 laser scanned heads formed the morphable model, which was fit to best match the input images. The render parameters for the initialization of the reconstruction algorithm were known.

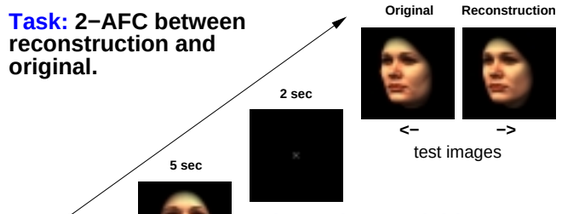
Originals: Images rendered from the same laser scanned heads, that served as input for the reconstruction algorithm.

The lighting conditions and head orientations of the two test images were always identical to each other, but different from the input image. The edges of the heads were faded out. 84 randomly chosen heads of the 100 were used for the experiment. The rest were used for training trials.

Task

Question: Which one of the two faces better resembles the previously seen face?

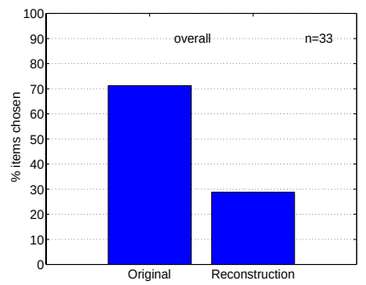
Task: 2−AFC between reconstruction and original.



Independent variables:
head orientation of input image: 0°, 30° , 60°, 90°
head orientation of test image: −90° (left profile), −60°, −30°, 0° (full face), 30°, 60°. 90° (right profile).

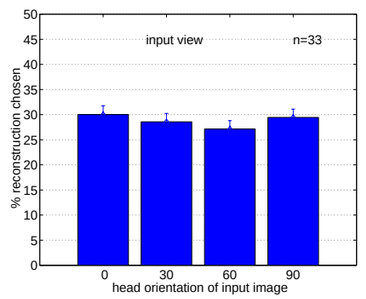
3 RESULTS

Performance

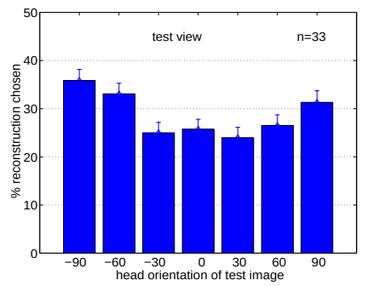


Reconstruction chosen: 28.8%

Original chosen: 71.2%



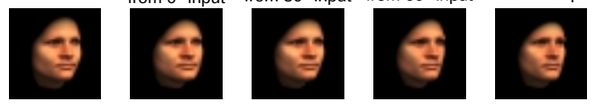
There is a no main effect of input view. ($F_{3,96} = 0.67$, $p = 0.574$)



There is a main effect of test view. ($F_{6,192} = 4.44$, $p < 0.001$)

Item Analysis

"Best" item: 19/33 subjects (58.6%) chose the reconstruction



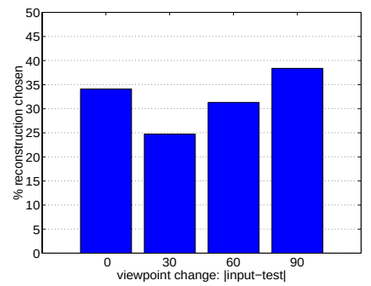
# Ss, who chose the reconstruction:	6/10	6/9	1/7	6/7

"Worst" item: none of 33 subjects (0%) chose the reconstruction

# Ss, who chose the reconstruction:	0/7	0/7	0/10	0/9

The 'best' reconstruction is preferred at some conditions (above chance level), but this is not consistent throughout conditions.

The reconstruction is more often chosen in the 'same viewpoint' condition than when there is a viewpoint change of 30° or 60°.



4 DISCUSSION

- **The reconstruction algorithm seems to preserve crucial features of the individual faces. Distinguishing between reconstruction and original is at least harder than distinguishing between two different heads, because the "error" rate (i.e. reconstructions chosen) is slightly higher than Troje & Bülthoff (1996) found in a standard old/new task (overall error rate 24.4 %) with a 30 times shorter presentation time (165 msec).**
- **A possible explanation for the fact that there is no main effect of the input view might be that the performance of the algorithm drops at the same rate as human performance. Assuming that performance in this task reflects the difference between human performance and performance of the reconstruction algorithm, the error rate should increase, when the human performance drops faster than the algorithm performance or vice versa.**
- **For the 'same view' conditions the amount of reconstruction chosen increases although humans are usually best at this condition. This might be due to the fact that the mean pixel error is lowest at this condition (Blanz, 2000).**

5 LITERATURE

Blanz, V. (2000). Automatische Rekonstruktion der dreidimensionalen Form von Gesichtern aus einem Einzelbild. *Unpublished Dissertation*, Tübingen, Germany

Blanz, V. & Vetter, T. (1999). A morphable model for the synthesis of 3D faces. *Computer Graphics Proceedings SIGGRAPH'99,* pp. 187−194.

Troje, N. F. & Bülthoff, H. H. (1996). Face recognition under varying poses: the role of texture and shape. *Vision Research,* Vol. 36, No. 12, pp. 1761−1771.