Appeared in: Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 177-180, 1998.

Face similarity space as perceived by humans and artificial systems

Peter Kalocsai*, Wenyi Zhao+ and Egor Elagin±
*Hedco Neuroscience Bldg., University of Southern California, Los Angeles, CA 90089
+Center for Automation Research, University of Maryland, College Park, MD 20742
±EYEmatic Interfaces Inc., 2803 Colorado Avenue,Santa Monica, CA 90404
kalocsai@selforg.usc.edu, wyzhao@cfar.umd.edu, elagin@selforg.usc.edu

Abstract

The performance of a local feature based system, using Gabor-filters, and a global template matching based system, using a combination of PCA (Principal Component Analysis) and LDA (Linear Discriminant Analysis) was correlated with human performance on a recognition task involving 32 face images. Both systems showed qualitative similarities to human performance in that all but one of the calculated correlation coefficients were very or moderately high. The Gabor-filter model seemed to capture human performance better than the PCA-LDA model since the coefficients for this model were higher for all examined conditions. These results indicate that the preservation of local feature based representation might be necessary to achieve recognition performance similar to that of humans.

1. Introduction

In recent years several artificial systems based on a variety of computational principles have been developed for the recognition of face images. According to one type of categorization face recognition systems could be classified as belonging to one of two major types: holistic template matching based systems and geometrical local feature based systems [2]. In this paper we will correlate the performance of a representative of each of these system types with human performance.

Benchmark tests like the ones administered by the Army Research Laboratory, called the FERET test, are extremely helpful for an unbiased comparison of performance of artificial face recognition systems [5]. However, one might argue that the ultimate test of an artificial face recognition system is to compare its performance, preferably quantitatively and qualitatively as well, to the best system there is which would of course be that of humans. To make such a comparison a simple test of recognizing 32 face images was designed and administered to both humans and to the models. In the following the procedure for collecting psychophysical data will be described and then the two recognition systems and their performance will be discussed.

2. Psychophysical study

Two pictures of 16 individuals, one taken with neutral and one with angry expression, were used as stimuli in this experiment. Everything outside the face was blocked out by an oval area in order to eliminate the effect of hair and background (Figure 1). 64 subjects performed a sequential matching task on a pair of faces in which they had to judge whether the two sequentially presented images were of the same or different individuals. They were instructed to ignore differences in the expression of the faces. The stimulus sequence was the following: first a bull's eye was presented for 500 msec followed by the first face presented for 150 msec. After the first face a mask was presented for 500 msec, followed by a 150 msec presentation of the second face. Following the second face a second mask was presented for 500 msec as illustrated on Figure 1. The location of the second face on the monitor was randomized in order to eliminate iconic memory effects. Each subject performed 544 comparisons half of which were 'same' trials and the other half were 'different' trials. Subjects were instructed to ignore the intervening mask and to respond as quickly and as accurately as possible by pressing a 'same' or 'different' microswitch key after the presentation of the second face. The reaction time and error rate of the subjects was recorded.

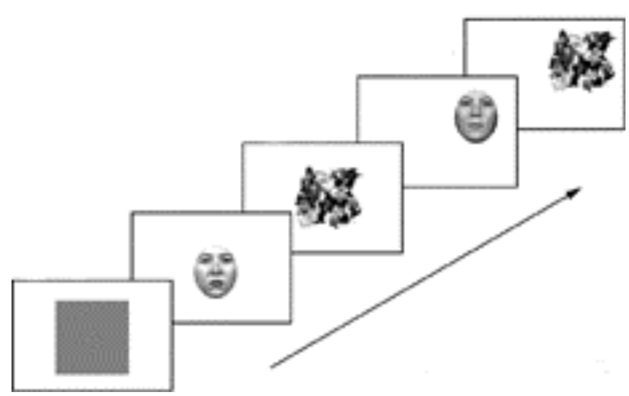

Figure 1. Stimulus sequence.

3. Brief description of the systems

In the following the two face recognition models which participated in the analysis, the Gabor-filter model and the PCA-LDA model will be described.

3.1. Gabor-filter based system

One of the systems which was tested on the same image-set that was used in the psychophysical experiment was the one proposed by von der Malsburg and his colleges [4]. This system represents faces as convolution results of a face image with a bank of multiscale and multiorientation kernels at different locations on the image. The locations for filtering were at the vertices of a 9 x 7 lattice which covered the face. At each of these locations the images were convolved with 40 Gabor-filters (8 orientations x 5 scales). The convolution of the image $I(\vec{x})$, with a bank of Gabor-filters is expressed as

$$(WI)(\vec{k},\vec{x}_0) = \int \psi_{\vec{k}}(\vec{x}_0 - \vec{x}) I(\vec{x}) d^2x = \psi_{\vec{k}} * I. \qquad (1)$$

The filters form a self-similar family of Gabor functions which are also known under the name of "Morlet wavelets" in the literature and have the general form

$$\psi_{\vec{k}}(\vec{x}) = c_{k,\sigma} \exp\left(-\frac{\vec{k}^2\vec{x}^2}{2\sigma^2}\right)\exp\left(j\vec{k}\vec{x}\right). \qquad (2)$$

Where $c_{k,\sigma}$ is a constant and $\vec{k}$ controls the size of the Gaussian window and the frequency and orientation of the kernel. The constant parameter σ assures that the ratio of the wavelength and the window size is such that in all cases the shape of the Gabor kernels are similar, and resemble the simple cell receptive field profiles found in V1 [3]. To give an idea of how this system performs on pairs of face images four different-individual pairs are compared on Figure 2 with the lattice positioned over the faces. The reference image is on the left side and test image is on the right. Two pairs are presented with angry expression in the upper half and two pairs with neutral expression in the lower half (Figure 2). Form top to bottom the respective similarity values for the two images are 93, 88, 92, 86 where 100 would indicate a perfect match.

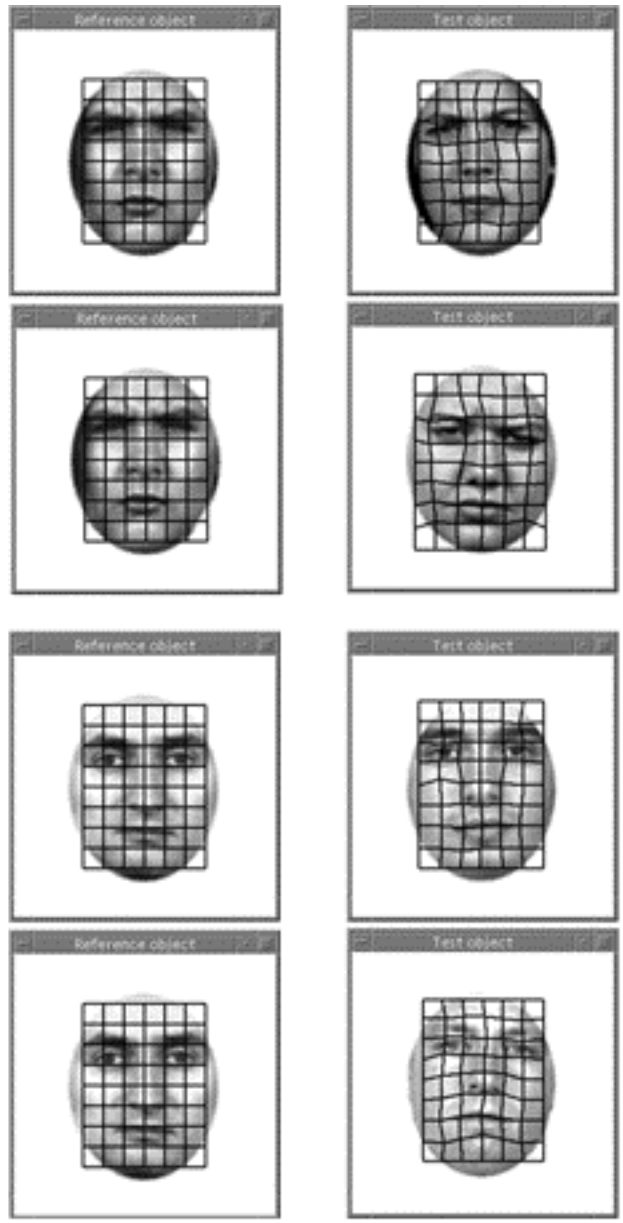

Figure 2. Face pairs used in the experiment. The similarity of the above different-person face pairs is 93, 88, 92, 86 from top to bottom (100 is perfect match).

3.2. Principal Component Analysis

Although both LDA and PCA have been suggested for face recognition before [6], there are some indications that the combination of the two methods might result in a superior performance. One serious drawback of LDA is that it is very much tuned to the specific training set. A completely new test-set could cause serious problems for a purely LDA based method. On the other hand pure PCA has limited use when there is large within-class variance in the image data. A combination of the two systems could potentially reduce both of these negative effects.

First we obtain a linear projection which maps the input image *x* into the face-subspace *y*. Then *y* is mapped into the classification space *z* With PCA one can map input images to a face-subspace which could be further mapped into a classification space by LDA. Based on some distance criterion then recognition is performed on this classification space.

$$y = \phi x \qquad (3)$$

$$z = W_y^T y \qquad (4)$$

where ϕ is the PCA transform and W_y is the best linear transform on PCA feature space. After this the classification is performed on the classification space based on a distance measure criterion. For a more detailed discussion the reader is referred to [7].

4. Analysis and results

After all the human and model data have been collected a separate analysis of 'different' and 'same' trials was carried out. A 'different' trial would show similarity between faces of two different individuals. In this analysis only matches with the same facial expression (neutral or angry) were considered since we were interested in performance difference caused by differences in identity.

'Same' trials refer to two images of the same individual, but with different expressions. Quite the contrary to what we had for different trials, for same trials only image pairs with different expressions were analyzed since matching an image with itself would always provide a perfect match.

The results can be seen on Figure 3-6 and Table 1. The Gabor-filter based system correlated very highly (r = .91) with human error on different trials. It also correlated highly, but negatively with error for same trials (r = -.79). One might expect such a change of sign because for 'same' trials higher similarity would make it easier to recognize that the two images belong to the same individual. In other words human subjects would make less mistake on highly similar same-pairs. For the PCA-DLA method we received similar, but somewhat lower correlations. For different trials the correlation was r = .45 and for same trials it was r = -.43. For reaction time all the correlations were lower for both model types, but they still followed the general pattern that was observed for error rates.

	Gabor	PCA-DLA
Different trials		
Error	.91	.45
RT	.73	.33
Same trials		
Error	-.79	-.43
RT	-.21	.06

Table 1. Correlation coefficients between the two models and human data.

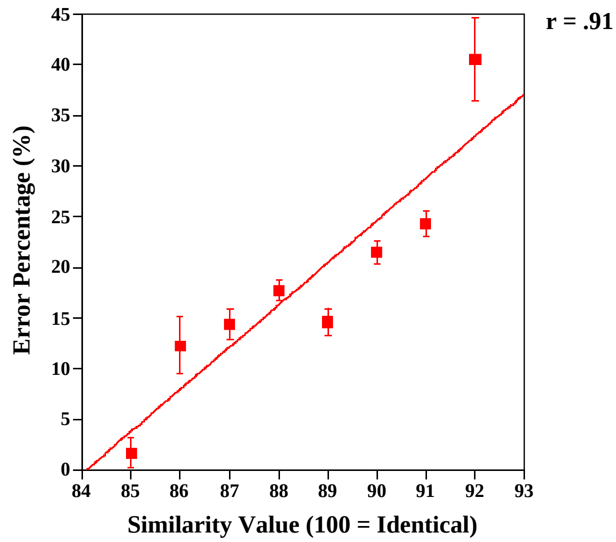

Figure 3. Correlation of the Gabor-filter model with human error on different trials.

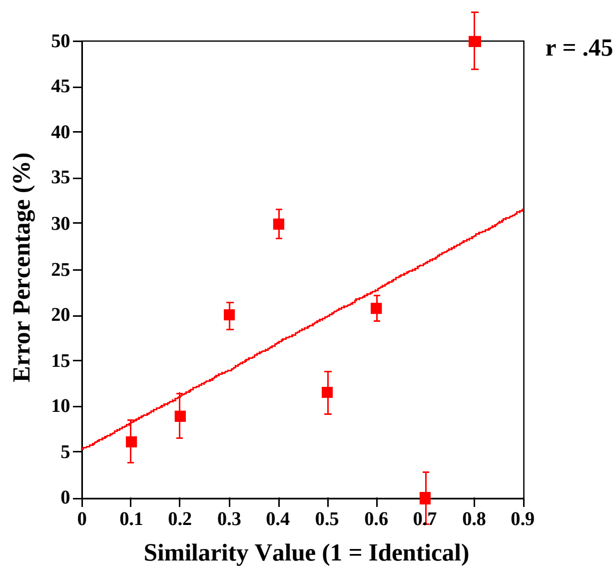

Figure 4. Correlation of the PCA-LDA model with human error on different trials.

The only exception to this rule was the correlation between human RTs and the PCA-LDA model on same trials, which was not negative as one might have expected it.

5. Conclusions

The performance of a local feature based (Gabor filter) and a global template matching based (PCA-DLA) method was correlated to human performance on a face recognition

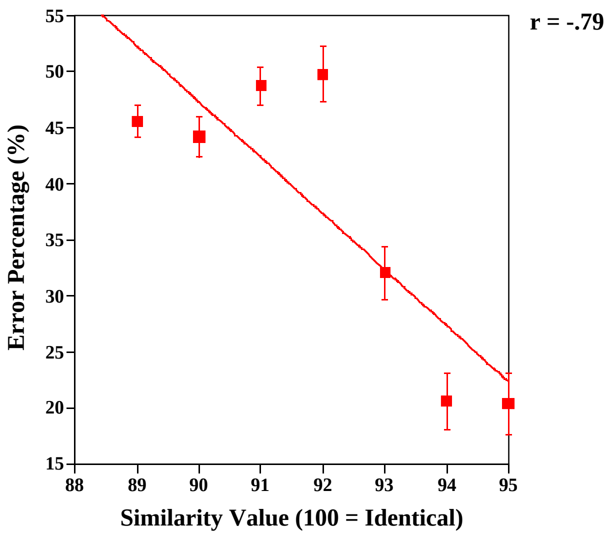

Figure 5. Correlation of the Gabor-filter model with human error on same trials.

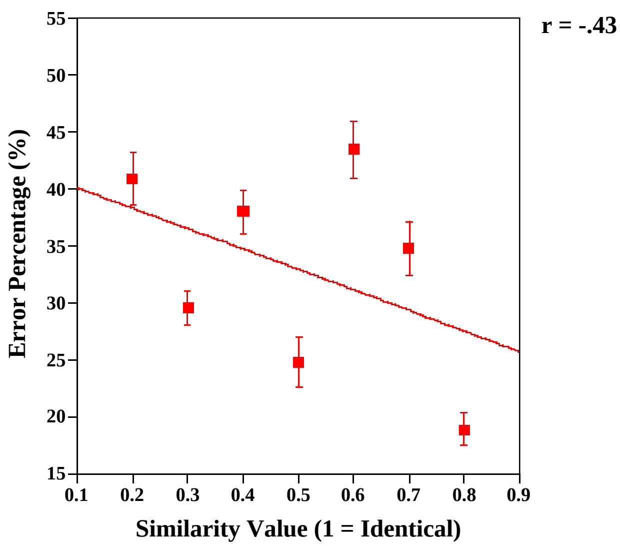

Figure 6. Correlation of the PCA-LDA model with human error on same trials.

task. Both systems showed relatively high correlations especially with human error, although in absolute value the correlation coefficients were higher for the Gabor-filter based method for all conditions. This higher correlation might indicate that the preservation of local features is necessary for human-like face recognition. In general, one could argue that to the degree that these model's performance correlates with that of human subjects their representation also spans a similar face space. From the obtained correlation values it is likely that a local feature based representation is closer to the representation employed by the human visual system for faces than a representation based on global template matching.

6. Future plans

The study at this stage provides a quantitative comparison of the performance of two face recognition systems based on two different principles. Certainly, there have been numerous other methods suggested for this task. Due to time and space limitations we are not able to review all of them here, although a separate analysis of LDA (Linear Discriminant Analysis), LFA (Local Feature Analysis) and also ICA (Independent Component Analysis) [1] methods is currently under way.

Acknowledgments

The authors wish to thank Irving Biederman for his support and invaluable help related to this project. This research was supported by ARO NVESD grant DAAH04-94-G-0065.

References

[1] A. J. Bell and T. J. Sejnowski. Edges are the 'independent components' of natural scenes. *Advances in Neural Information Processing Systems 9,* MIT Press.

[2] R. Chellappa, C. L. Wilson and S. Sirohey. Human and Machine Recognition of Faces. A survey. *Proceedings of the IEEE,* Vol. 83:705-740, 1995.

[3] R. L. DeValois and K. K. DeValois. *Spatial vision.* Oxford Press, 1988.

[4] M. Lades, J. C. Vortbrüggen, J. Buhmann J. Lange, C. von der Malsburg, R. P. Würtz, and W. Konen. Distortion Invariant Object Recognition in the Dynamic Link Architecture. *IEEE Transactions on Computers,* 42:300-311, 1993.

[5] P. J. Phillips, H. Moon, P. Rauss, and S. A. Rizvi. The FERET Evaluation Methodology for Face-Recognition Algorithms. *Proceedings of Computer Vision and Pattern Recognition,* Puerto Rico, 137-143, 1997.

[6] M. Turk and A. Pentland. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience,* Vol. 3:72-86, 1991.

[7] W. Zhao, R Chellappa and A. Krishnaswamy. Discriminant Analysis of Principal Components foe Face Recognition. Submitted to FG'98.