Human sensitivity to face statistics computed on V1 similarity

Peter Kalocsai
Department of Psychology, University of California, Los Angeles
CA 90095, USA

Abstract

In a biologically motivated recognition system we represent face images as convolution values with a set of multiscale and multiorientation Gabor wavelets (a simple model of V1). Based on their discriminative power on recognition tasks various face images were reconstructed from the Gabor wavelet representation to test whether the computed statistics had any psychophysical relevance. The result of our experiments indicate that human performance is sensitive to statistical information derived from our recognition system. The study provides evidence that higher face recognition areas, such as FFA in humans, potentially not only preserve, but also compute upon earlier V1 similarity space.

Keywords: Face recognition, V1, FFA, Gabor wavelet, reconstruction

---

1. Introduction

It has been known for some time that the receptive field properties of V1 simple cells can be modeled well with Gabor wavelets [3]. It has also been demonstrated that the performance of a recognition system, with multiscale and multiorientation Gabor kernels as features (similar to a V1 representation), correlates highly with human performance on face recognition tasks [5]. We hypothesized earlier that the reason for such a high correlation is that higher areas involved in face processing: FFA (Fusiform Face Area) seem to retain at least parts of the earlier V1 similarity space [2]. Taking this one step further in this study we would like to investigate whether human face recognition is also sensitive to statistics computed upon that V1 similarity.

It is well accepted that an important goal of the human visual system is to extract statistics of natural images [1,7]. However, the type of statistics being computed and used at various stages in the visual hierarchy is much less understood. The goal of this study is to see if we can relate statistics computed on a physical face space (on our Gabor wavelet representation of faces) to the psychological face space as revealed by psychophysical experimentation.

2. ANOVA on Gabor wavelet representation

A univariate analysis of variance was applied to the Gabor wavelet representation of 1414 Japanese face images (101 individuals X 14 conditions) to derive the discriminative power of all wavelets individually for the recognition of these faces. The different conditions of the face images are indicated in Table 1.

Conditions	# of images	Levels
Horizontal Orient.	5	(0°, 10° and 20° left/right)
Vertical Orient.	4	(5° and 10° up/down)
Expression	3	(neutral, surprised, tired)
Illumination	2	(light, dark)
	14	(sum of all levels)

Table 1. Image-set for one Japanese individual.

The result of the ANOVA showed that the forehead and the eye regions were quite important. The region between the nose and mouth also seemed to be rather informative. The different frequency channels overall seemed to have equal importance. The range of the F values was about one magnitude. The highest F value was $F(100,1313) = 25$ and the lowest was $F(100,1313) = 1.5$. The reader is referred to [4] for more details on the analysis. After the analysis all 1920 Gabor wavelets (48 locations X 8 orientations X 5 scales) in the representation were ranked according to their discriminative power. This ranking was used in the next step for creating various reconstructed face images, by the method in [9], for the psychological experiments.

3. Method

To investigate whether humans would be sensitive to statistics computed upon the above V1 type representation 5 reconstructed images were created: a full, two half and two hybrid ones. The applied procedure for reconstructing these images was the following: for the full reconstructed images we took a face image from the database (Fig. 1 left) and reconstructed it from all the 1920 kernel activation values. Note that the reconstruction will be far from perfect since kernels at only 48 locations on the face were used with only eight orientations and five scales (Fig. 1 right).

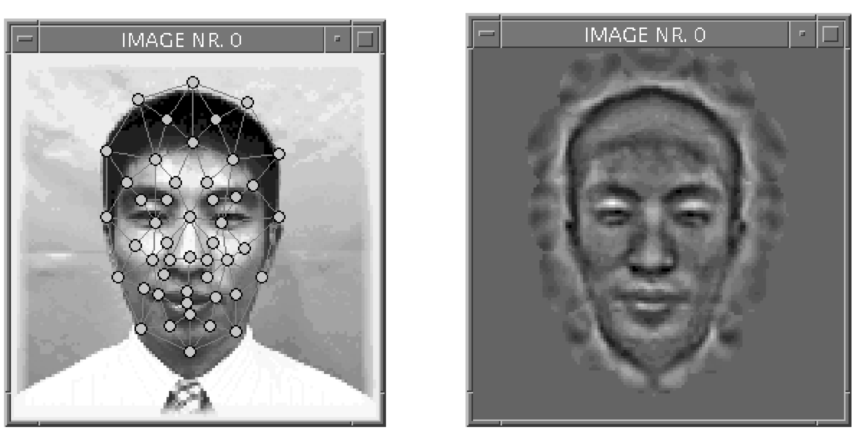

Fig. 1. Left: An arbitrarily chosen image with the kernel locations indicated on the face. Right: Reconstruction of the image on the left from the 1920 Gabor kernel activation values.

Next we created two half reconstructed images: one from the more discriminative (stronger) and one from the less discriminative (weaker) half of the kernels. For an example of these reconstructions see the lower two face images in Fig. 2 where the strong half reconstructed image is on the left side and the weak one is on the right. Similarly, two hybrid reconstructions were also created: one from the stronger half of the wavelets for the target image and from the weaker half for a randomly chosen other image, and another from the weaker half of the wavelets for the target image and the stronger half of the other image. However, the location of the fiducial points on the face for both of these reconstructions came from the target image. The target image is the one on the top in the image triplets as shown in Fig. 3. Take a look at the lower two face images in Fig. 3 for an example of these hybrid reconstructions. The strong reconstructed image is on the right side and the weak one is on the left.

All reconstructed images were adjusted for mean square error so that the stronger reconstructions were always physically less similar to the fully reconstructed image than were their weak counterparts. This adjustment, of course, works against the hypothesis of the study.

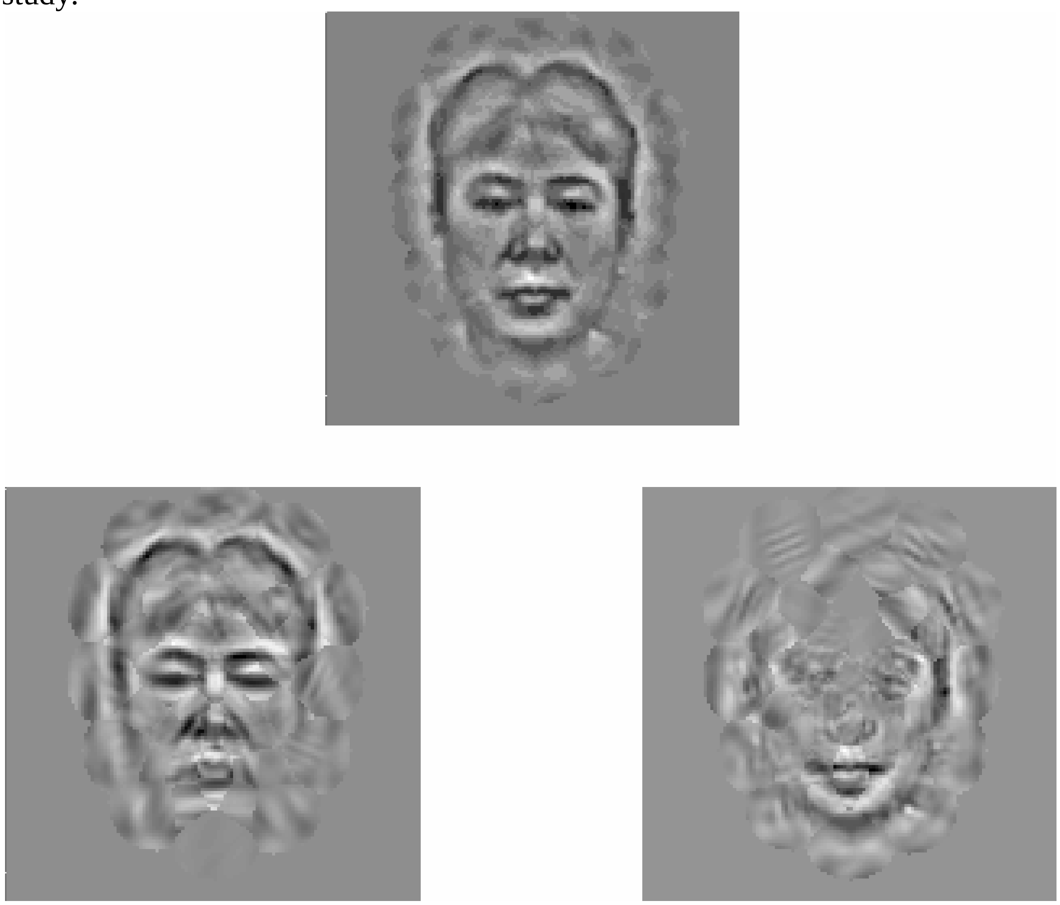

Fig. 2. Stimulus for the half reconstructed experiment. On top is the fully reconstructed image. On the bottom left side is the image reconstructed from the more discriminative (stronger) half of the kernels. On the bottom right side is the image reconstructed from the less discriminative (weaker) half.

Subjects performed two forced choice tasks in which they compared the two half reconstructed and the two hybrid reconstructed images to the fully reconstructed version.

The trials in both experiment had the same structure. On a calibrated monitor subjects viewed image triplets where the fully reconstructed image was shown in the upper half of the screen and the two partial reconstructions were presented below, left and right in the lower part. The left and right location of the strong/weak images was balanced in the experiments. Take a look at Figs. 2 and 3 to see a trial in the half and hybrid reconstructed experiment respectively. In both experiments subjects were to choose which of the two partially reconstructed images at the bottom, the one on the left or the one on the right, was more similar to the fully reconstructed version on the top. Responses of 20 subjects were collected with half of them doing the half reconstructed experiment first and the hybrid experiment next. The other half of the subjects completed the experiments in the opposite order. Subjects were graduate or undergraduate students of the University of California, Los Angeles with normal or corrected to normal vision. Each experiment consisted of 60 trials with 30 experimental and 30 control ones in random order. In the control trials the two partial reconstructions at the bottom had no relation at all to the fully reconstructed image on the top. In these trials the image on the top was that of a randomly chosen other person. Control trials were introduced to account for other possible confounding effects of strong and weak reconstructions that were not related to the identity of faces.

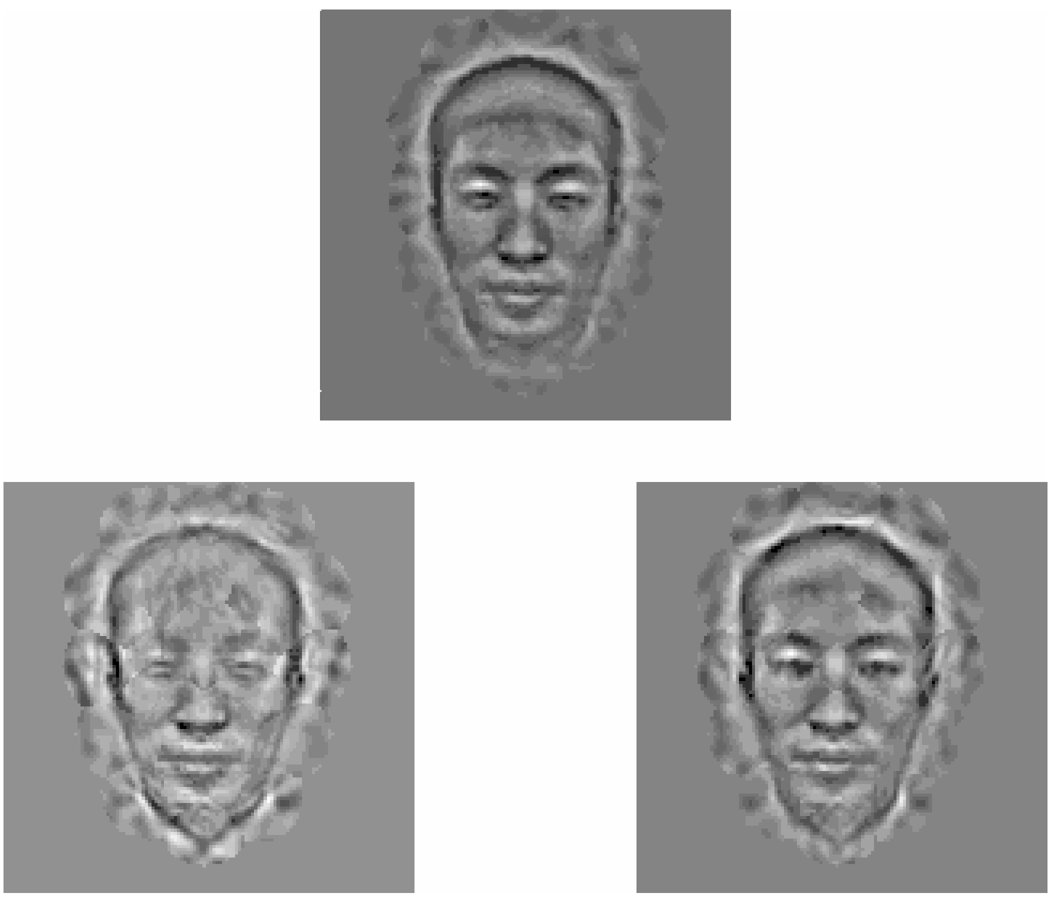

Fig. 3. A trial for the hybrid experiment. On top is the original full reconstructed image. One the bottom left side is the weak hybrid reconstruction and on the right side is the strong hybrid one.

4. Results

Two-tailed t-tests for dependent samples revealed a highly significant difference between experimental and control trials for both the half and hybrid experiment. In both experiments subjects were much more likely to choose the strong version of the two partially reconstructed images as more similar to the fully reconstructed one on the top in experimental trials, but not in control ones. For the half reconstructed experiment the difference between experimental and control trials was highly significant with $t = 6.15$ and $p < .0001$ (Fig. 4 left side). The experiment with hybrid reconstructed images gave similar results with $t = 4.91$ and $p < .0001$ (Fig. 4 right side). In both experiments subject chose the strong reconstructed images in a bit over 70% of the time in experimental trials. However, their choice on control trials was not distinguishable from the 50% chance level. Two additional t-tests revealed that in both experiments the responses on control trials were not statistically different from chance at the $p < .01$ level. The corresponding p values were $p = .63$ for the half experiment and $p = .045$ for the hybrid one.

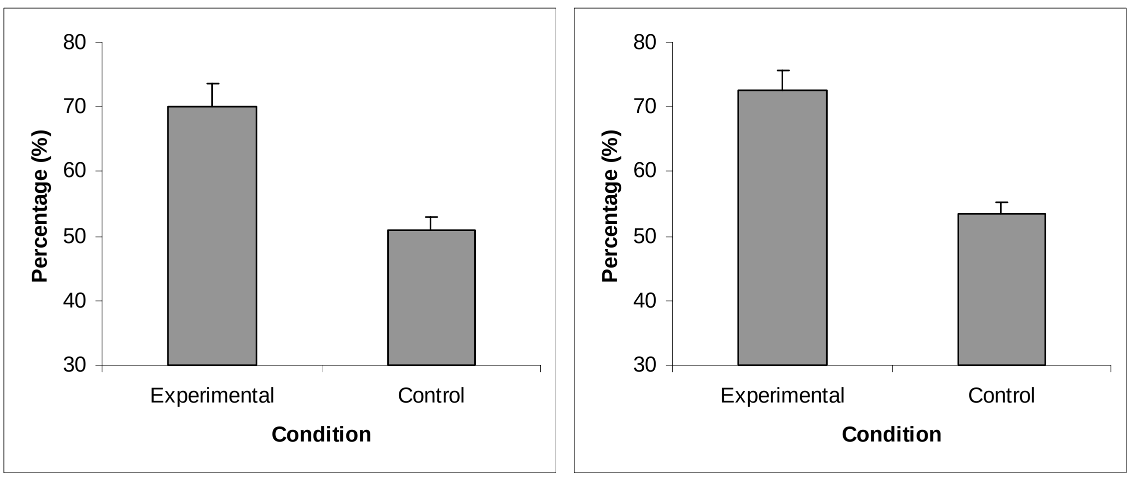

Fig. 4. Choice of strong reconstructed images in percentages for the experimental and control trials in the half reconstruction experiment (left) and the hybrid experiment (right).

5. Conclusions

The above results indicate that human face recognition is sensitive to statistical information derived from a biologically inspired artificial face recognition system which mimics V1 simple cell similarity. The study provides evidence that higher face recognition areas in humans, such as FFA probably not only preserve, but also compute upon an earlier V1 similarity space. Since choosing stronger reconstructions on control trials was not significantly different from chance we are safe to conclude that the observed result was entirely due to the effect that the various reconstructions induced upon our subjects ability to judge identity.

There had been numerous successful attempts to relate the physical and psychological face space to one another (e.g. [6, 8, 10]). Most of these approaches however used some form of principal component analysis on raw image data without much effort for trying to compute statistics upon a biologically plausible feature space. In this study the connection between physical and psychological face space was demonstrated upon a biologically derived feature space: the Gabor wavelet representation.

References

[1] H. B. Barlow, Possible principles underlying the transformations of sensory images, in: W. A. Rosenblith (Ed.), Sensory Communication, MIT Press, 1961, pp. 217-234.

[2] I. Biederman, P. Kalocsai, Neurocomputational Basis of Object and Face Recognition, Philosophical Transactions of the Royal Society: Biological Sciences, 352, (1997) 1203-1219.

[3] J. P. Jones, L. A. Palmer, The two-dimensional spatial structure of simple receptive fields in cat striate cortex, Journal of Neurophysiology, 58(6) (1987) 1187-1211.

[4] P. Kalocsai, C. von der Malsburg, J. Horn, Face recognition by statistical analysis of feature detectors, Image and Vision Computing. 18 (2000) 273-278.

[5] P. Kalocsai, W. Zhao, E. Elagin, Face Similarity Space as Perceived by Humans and Artificial Systems, Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 1998, pp. 177-180.

[6] D. A. Leopold, A. J. O'Toole, T. Vetter, V. Blanz, Prototype-referenced shape encoding revealed by high-level aftereffects. Nature Neuroscience, 4(1) (2001) 89-94.

[7] B. A. Olshausen, D. J. Field, Emergence of simple-cell receptive field properties by learning a sparse code of natural images, Nature, 381, (1996) 607-609.

[8] A. J. O'Toole, K. A. Deffenbacher, D. Valentin, H. Abdi, Structural aspects of face recognition and the other-race effect, Memory & Cognition, 22(2) (1994) 208-224.

[9] M. Pötzsch, T. Maurer, L. Wiskott, C. von der Malsburg, Reconstruction from Graphs Labeled with Responses of Gabor Filters, Proceedings of the ICANN, Springer Verlag, Berlin, Heidelberg, New York, 1996.

[10] G. Rhodes, S. Carey, G. Byatt, F. Proffitt, Coding spatial variations in faces and simple shapes: a test of two models, Vision Research, 38, (1998) 2307-2321.

Peter Kalocsai has received his Master of Computer Science degree in 1997 and his Ph.D. in Cognitive Psychology in 1998, both from the University of Southern California. He received the Ph.D. for his work on "Face Recognition with Gabor-filter Representation". Currently, he is a postdoctoral fellow at UCLA. His main research area is the computational modeling of vision, in particular that of object and face recognition. His ultimate goal is to build biologically motivated recognition systems by combining various machine learning and computer vision techniques with intricate knowledge of the mammalian visual system.