Appeared in: Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 360-365, 1998.

Statistical Analysis of Gabor-filter Representation

Peter Kalocsai*, Hartmut Neven+ and Johannes Steffens+
*Hedco Neuroscience Bldg., University of Southern California
Los Angeles, CA 90089
+EYEmatic Interfaces Inc., 2803 Colorado Avenue
Santa Monica, CA 90404
{kalocsai, neven, steffens}@selforg.usc.edu

Abstract

A successful face recognition system calculates similarity of face images based on the activation of multiscale and multiorientation Gabor kernels, but without utilizing any statistical properties of that representation [3]. A method has been developed to weight the contribution of each element (1920 kernels) in the representation according to their power of predicting similarity of faces. The same statistical method has also been used to assess how changes in orientation (horizontal and vertical), expression, illumination and background contribute to the overall variance in the kernel activations. Weighting the elements in the representation according to their discriminative power has shown to increase recognition performance on a Caucasian and on a Japanese test image-set. It has also been demonstrated that such weighting method is particularly useful when data compression is a key requirement.

1. Introduction

Over the last couple of years it has been demonstrated that using Gabor filters as the front-end of an automated face recognition system could be highly successful. In a national competition for the best face recognition system administered by the Army Research Laboratory (FERET program) the system performed among the best ones on several tests. Although some research also suggests that the system's recognition performance shows qualitative similarities to that of humans by now means it is equivalent or better than that which leaves plenty of room for improvement.

In the current version of the system faces are represented as convolution results of the face image with 40 Gabor kernels (8 orientations x 5 scales) at 48 locations (fiducial points or also termed jets) on the face [3]. This gives a 1920 (40 x 48) element long vector keeping only the magnitude values in the representation. It intuitively makes sense that these 1920 activation values would not contribute equally to the face recognition task. In other words it is likely that there would be variability in the predictive power of the individual Gabor kernels. Consequently, an analysis that would weight the kernels according to their usefulness for recognition could potentially increase the system's performance. We might mention that statistical analysis is most likely significant part of the human face recognition system as well in that for instance several psychological phenomena seem to suggest that faces in the brain are represented according to a norm(average)-based code for the purposes of recognition. Among these phenomena the most important ones are the caricature effect, other-race effect, distinctiveness, typicality, attractiveness [1].

2. Data preparation

The most typical dimensions along which two images of the same individual could vary are due to changes in orientation (horizontal, vertical), expression, illumination, background. Our analysis focused on variations along these dimensions. Tables 1 and 2 summarizes the image-set that was collected for each individual in two databases (a Caucasian and a Japanese one). As Tables 1 and 2 shows the Caucasian database consisted images of relatively few individuals (6) over large number of conditions (243), whereas the Japanese database consisted of few conditions (14) over many individuals (101). We might mention that certainly, other possible variations could be introduced by different types of disguise, change in hairstyle, facial hair, facial accessories (e.g. glasses, earrings) which topics are not discussed here.

Conditions	# of images	levels
Horizontal Orient.	3	(0°, 15° left, 15° right)
Vertical Orient.	3	(0°, 15° up, 15° down)
Expression	3	(neutral, smiling, tired)
Illumination	3	(central, left, right)
Background	3	(white, natural, artificial)
$\prod$	243	(permutations of all levels)

Table 1. Image-set for the 6 Caucasian faces.

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)
Σ	14	(sum of all levels)

Table 2. Image-set for the 101 Japanese faces.

After the images have been collected they were run through the von der Malsburg face recognition system. The images were convolved with a set of Gabor kernels at various locations on the face. As a result of this operation each face was represented as a 40 x 48 matrix or 1920 element vector in the database. Since for example there were 243 images taken of each individual in the Caucasian database the total amount of data collected for one individual comprised a 40 x 48 x 243 matrix as indicated on Figure 1. The location of the jets on the face image is presented on Figure 2.

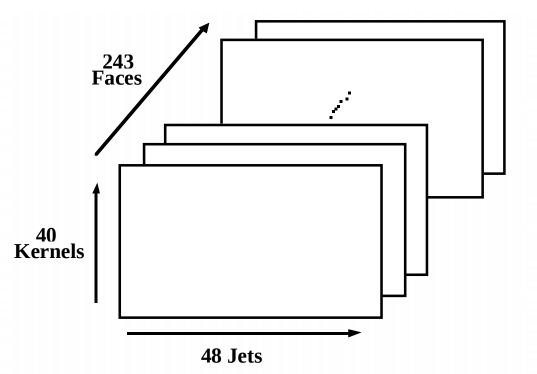

Figure 1. Data collected for one individual in the Caucasian database.

3. Testing and Results

First the contribution of the 5 tested conditions (for Caucasian faces) to the overall variance in the data will be discussed, followed by a univariate analysis of variance of kernel activations for both Caucasian and Japanese faces.

3.1. Variance produced by the conditions

For improving recognition performance it might be useful to know which conditions contributed the most to the total variance. As indicated on Table 3. the largest contributor to the overall variance was the change in horizontal orientation and the least disturbing factor was the change in expression. Note that without special treatment background produces the second largest amount of variance. However, with background suppression this condition drops to the fourth largest contributor which shows the success of the algorithm [5].

	Hori.	Vert.	Expr.	Illum.	Back.
AV	4.9	2.7	1.8	2.9	4.2
Rank	1	4	5	3	2
With Background Suppression					
	Hori.	Vert.	Expr.	Illum.	Back.
AV	5.2	2.8	2.0	3.0	2.3
Rank	1	3	5	2	4

Table 3. Ranked contribution of the different conditions to the overall variance (AV = average variance in units of normalized energy).

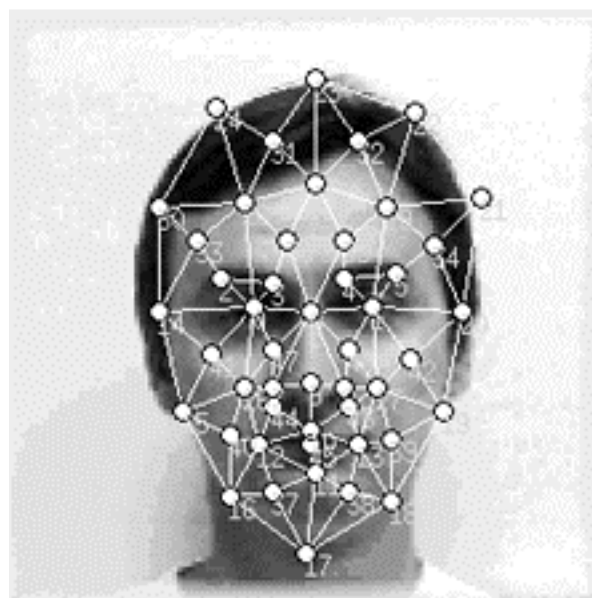

Figure 2. Location of the 48 jets on the face image.

3.2. Analysis of variance for Caucasian faces

Figure 3 shows the results for a complete between-individual one-way ANOVA calculated separately for all 1920 kernels. In this analysis 81 images (3 horizontal ori. X 3 vertical ori. X 3 expression X 3 illumination, but no background) of 6 Caucasian males were included. White areas indicate high F-values which mean high discriminability (between-individual variance is large compared to the within-individual variance). Black areas on the contrary show kernels of low discriminability and of limited power for recognition. Overall there was close to 3 magnitude difference between the kernels with highest and lowest F values. The highest F value was F(5,480) = 175 which was highly statistically significant and the lowest was F(5,480) = 0.26 which was not significant at all.

As Figure 3 indicates the hair area the forehead and eye

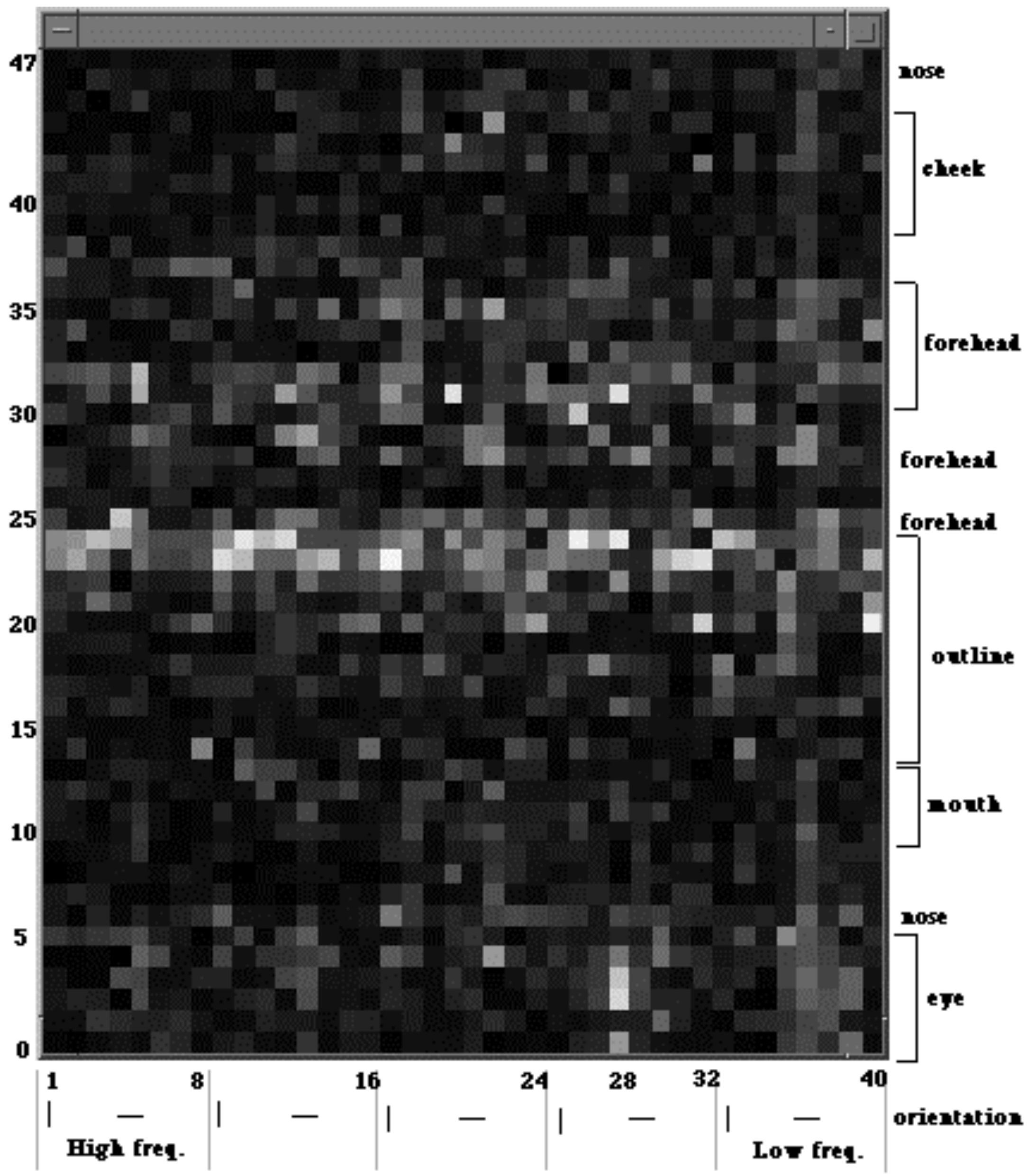

Figure 3. Significance map (F map) of the 1920 kernels for Caucasian faces. Horizontal axis represents kernel size and orientation, vertical axis indicates kernel location. White and black areas show high and low discriminative power respectively.

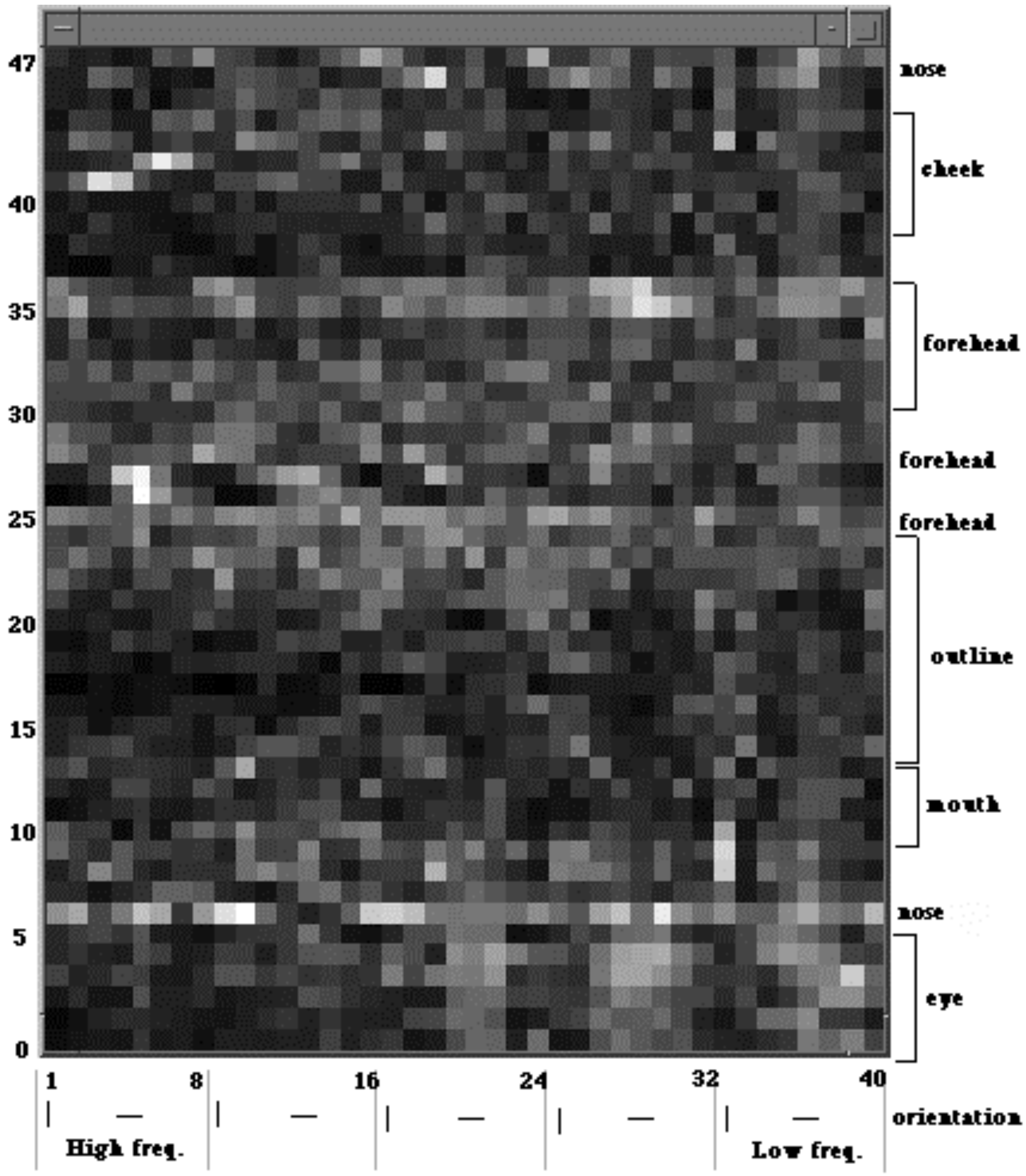

Figure 4. Significance map (F map) of the 1920 kernels for Japanese faces. Horizontal axis represents kernel size and orientation, vertical axis indicates kernel location. White and black areas show high and low discriminative power respectively.

regions provided the highest F scores and the mouth, nose, cheek and lower part of the outline region gave the lowest ones. For easier understanding some of the jet locations and kernel numbers are indicated on Figure 3, but for full reference on jet locations the reader is referred to Figure 2. The fact that the hair and forehead region scores so high is in accordance with some neurophysiological observations with monkeys where responses of most face sensitive cells showed dependence on the size of the forehead region of the test faces [7]. The low discriminability of the outline locations might be indicative that these fiducial points mainly involved in segmenting out the face from its background and are of limited use for recognition itself. This of course does not mean that these locations could not have a very significant role in finding a face in an image. It only shows that once the face has been found the role of these points is significantly reduced.

On average there does not seem to be a big difference among the predictive power of different frequency levels, although on average horizontal kernels seem to have somewhat higher F-values than vertical ones especially in the eye region. This result again seem to be in accordance with the barcode theory for face recognition according to which certain horizontal low frequency kernels would contain useful information about face images [6]. Next lets examine the F-map for Japanese faces.

3.3. Analysis of variance for Japanese faces

Using the same univariate method as above 1414 Japanese face images (101 individuals X 14 conditions) were analyzed to derive the most discriminative kernels for Japanese faces. The different conditions are indicated in Table 2.

The results of the analyses are shown on Figure 4. One might notice immediate differences and also some similarities between the results of the above two studies. For Japanese faces the nose region looks quite informative. Again the forehead and the eye regions are quite important although mainly the lower frequencies are informative for the eyes. The region between the nose and mouth is also important although it was not very significant for Caucasian faces. Strangely enough, we get some discriminative power in the cheek as well although the lower part of the outline is less important just as it was for the Caucasian faces. The different frequency channels again seem to have equal importance. The whole map for this analysis is much brighter which only means that now the F values are closer to each other. In fact for these faces the highest F value was $F(100,1313) = 25$ and the lowest was $F(100,1313) = 1.5$.

3.4. Recognition test of the weight matrices

Once the weight matrices were obtained it has been tested whether weighting of the kernel activation values would increase recognition performance. The weight of a kernel was equivalent to the F value for that kernel from the ANOVA (for the Japanese matrix the cube of those F values were taken). For the Caucasian matrix two images of 325 individuals from the FERET database were used for testing [4]. The second image for each individual contained some modification to the first one (e.g. change in background, lighting, expression). Without the weight matrix the number of failures in the matching process was 10, but with the weight matrix failures dropped to 8 (if the same individual's face was not the best match then it was counted as failure).

For the Japanese matrix the same 1414 faces were tested that were used to obtain the matrix. Without any weighting the number of failures was 216 in this test, whereas with weights failures dropped to 147 which is significant improvement.

The matrices were also tested on the other race's dataset in order to see whether we find any validation for the 'other race effect' a well known phenomenon in human face recognition. When the Japanese matrix was tested on Caucasian faces the number of failures was 9 which is some improvement compared to the condition without weights, but it is not as good as the result with the Caucasian matrix. When the Caucasian matrix was applied to the Japanese faces the same effect was observed in the opposite direction. The number of failures was 192 which is better result then what was achieved without weights, but not nearly as good as the result with the Japanese matrix. Since in both ways the weight matrices worked better for their own races and they did not perform as well for the other race we consider that as at least partial validation of the 'other race effect'. Although, certainly, other factors such as the size of the databases and the difference in the conditions could have also contributed to this effect.

To test whether the weight matrices really contain important information for recognition 'anti-matrices' was created with reciprocal weights. If the weight matrix method really helps recognition then by taking the reciprocal of the weights we should do even worse than without any weights. Indeed, on the Japanese database with using reciprocal Japanese weights the number of failures have risen to 439. For Caucasian faces the number of failures have risen to 19 with reciprocal weights. For a summary of these results see Table 4.

	Japanese	Caucasian
Without weights	216	10
With Japanese weights	147	9
With Caucasian weights	192	8
With reciprocal weights	439	19

Table 4. Number of failures with different weight matrices.

Selecting the kernels with the highest discriminability could be particularly useful when compact representation is required. When kernels with high and low discriminability are selected from the two ends of the distribution of the Caucasian matrix not only did the high discriminative kernels produce better recognition rates than low ones, but their performance also degraded much more gracefully (Table 5). With only 100 high discriminative kernels recognition rate was still at 93%, with 40 kernels it was 90%. When only 10 high informative kernels were kept in the representation recognition rate was still up to 73%, whereas if we choose low informative kernels the recognition rate dropped to 32%. This shows the usefulness of statistical analysis of kernel activation values for data compression.

	High discrim.	Low discrim.
All 1920 kernels	96%	96%
100 kernels	93%	83%
40 kernels	90%	74%
10 kernels	73%	32%

Table 5. Degradation of recognition performance for kernels with high and low discriminative power.

4. Conclusions

A statistical method has been suggested for the analysis of face representation by which it is possible the weight the contribution of individual elements in the representation scheme by their predictive power. The weighting method was shown to increase recognition performance and in particular was shown to give good recognition results even with highly compressed data. A factor of 48 compression only decreased performance by 6%. Without the using the weighting method the decrease would have been around 22%.

By testing two different race datasets some validation to the phenomenon called 'other race effect' was found. It was also shown that by taking the reciprocal of the weights performance would even drop far below the result achieved by not using any weights.

The described statistical analysis also allowed for comparison of how changes in orientation, expression, illumination and background contribute to variance in the representation. In decreasing order changes due to horizontal orientation, illumination, vertical orientation, background and expression were the most disturbing factors for recognition at least for our Caucasian dataset.

5. Future plans

As demonstrated for background suppression the given framework gives an excellent testing device for algorithms developed to dampen the effect of variables that are irrelevant for face identification. In particular, the ability of different pose-estimation solutions to reduce within-individual variance is currently being tested [2].

It is also possible that different scaling of the weight matrix might result in different level of improvement for identification. In order to find the weight-matrix most suitable for improving face recognition the exploration of several versions would be required. As a next step we propose a multivariate analysis of variance in order to account for the possible correlation of kernel activation values in the representation.

Acknowledgments

The authors wish to thank Christoph von der Malsburg, John Horn and Irving Biederman for many invaluable discussions related to this project. This research was supported by ARO NVESD grant DAAH04-94-G-0065.

References

[1] V. Bruce, M. Burton. *Processing Images of Faces.* Ablex Publishing Corporation, Norwood, NJ, 1992.

[2] E. Elagin, J. Steffens, and H. Neven. Automatic Real-Time Pose Estimation System for Human Faces Based on Bunch Graph Matching Technology. Submitted to FG'98.

[3] 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.

[4] 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.

[5] J. Steffens, E. Elagin, and H. Neven. PersonSpotter - Fast and Robust System for Human Detection, Tracking and Recognition. Submitted to FG'98.

[6] R. J. Watt. *Visual Processing: Computational, psychophysical and cognitive research*. Lawrence Erlbaum Associates, Hillslade, NJ, 1988.

[7] M. P. Young and S. Yamane. Sparse Population Coding of Faces in the Inferotemporal Cortex. *Science,* 256:1327-1331, May 1992.