Image and Vision Computing 18 (2000) 273–278
www.elsevier.com/locate/imavis

Face recognition by statistical analysis of feature detectors

P. Kalocsai[a,*], C. von der Malsburg[b], J. Horn[c]

[a] U1 Consulting Group Inc., 564 Mission Street #3, San Francisco, CA 94105, USA
[b] Department of Computer Science, University of Southern California, Los Angeles, CA 90089, USA
[c] Department of Psychology, University of Southern California, Los Angeles, CA 90089, USA

Received 29 September 1998; received in revised form 30 June 1999; accepted 13 July 1999

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 the given face data [M. Lades, J.C. Vortbrüggen, J. Buhmann, J. Lange, C. von der Malsburg, R.P. Würtz, W. Konen, Distortion invariant object recognition in the dynamic link architecture, IEEE Transactions on Computers 42 (1993) 300–311]. A method has been developed to weight the contribution of each element (1920 kernels) in the representation according to its 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. It was shown on a Caucasian and a Japanese image-set that weighting the elements in the representation according to their discriminative power would increase recognition performance. It has also been demonstrated that the weighting method is particularly useful when data compression is a key requirement. The advantages of the weighting scheme were also verified by double cross-validation.

Keywords: Face recognition; Gabor-filters; Analysis of variance; Statistical analysis

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, it is by no means equivalent or better than that; there is 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 $\times$ 5 scales) at 48 locations (fiducial points, also termed jets) on the face [1]. This gives a 1920 (40×48) 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 a 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 and attractiveness [2].

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 and background. Our analysis focused on variations along these dimensions. Tables 1 and 2 summarize the image-set that was collected for each individual in two databases (a Caucasian and a Japanese one). As Tables 1 and 2 show, the Caucasian database consisted of images of relatively few

* Corresponding author. +1-3102026877; fax: +1-3102026877.
E-mail addresses: peter@u1.com (P. Kalocsai); malsburg@selforg.usc.edu (C. von der Malsburg); jhorn@rcf.usc.edu (J. Horn).

0262-8856/00/$ - see front matter
PII: S0262-8856(99)00051-7

Table 1
Image-set for the six Caucasian faces

Conditions	# of images	Levels
Horizontal orientation	3	(0°, 15° left, 15° right)
Vertical orientation	3	(0°, 15° up, 15° down)
Expression	3	(neutral, smiling, tired)
Illumination	3	(central, left, right)
Background	3	(white, natural, artificial)
Π	243	(permutation of all levels)

individuals (6) over a large number of conditions (243), whereas the Japanese database consisted of a 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.

After the images had 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 40×48 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 $40 \times 48 \times 243$ matrix. The location of the jets on the face image is presented in Fig. 1.

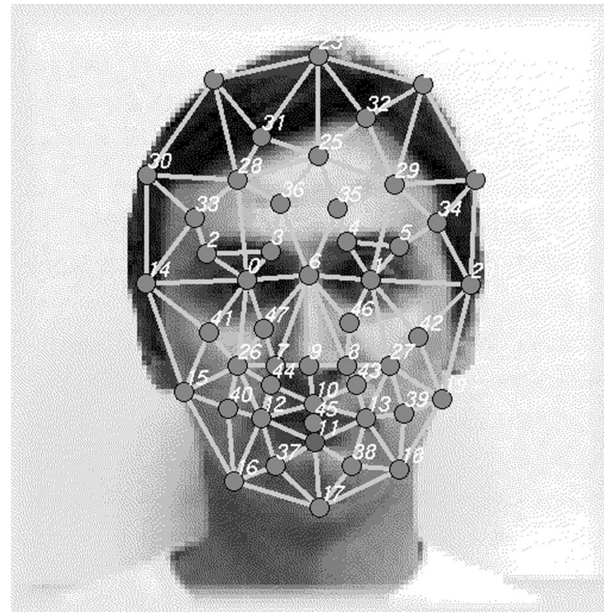

Fig. 1. Location of the 48 jets on the face image.

3. Testing and results

First the contribution of the five 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 in 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 [3]. 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 [4].

3.2. Analysis of variance for Caucasian faces

Fig. 2 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. $\times$ 3 vertical ori. $\times$ 3 expression $\times$ 3 illumination, but no background) of six 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 three orders of magnitude difference between the kernels with the 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 Fig. 2 indicates, the hairline area with the forehead and eye regions provided the highest F scores, while 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

Table 2
Image-set for the 101 Japanese faces

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

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

	Horizontal	Vertical	Expression	Illumination	Background
AV	4.9	2.7	1.8	2.9	4.2
Rank	1	4	5	3	2
With background suppression					
AV	5.2	2.8	2.0	3.0	2.3
Rank	1	3	5	2	4

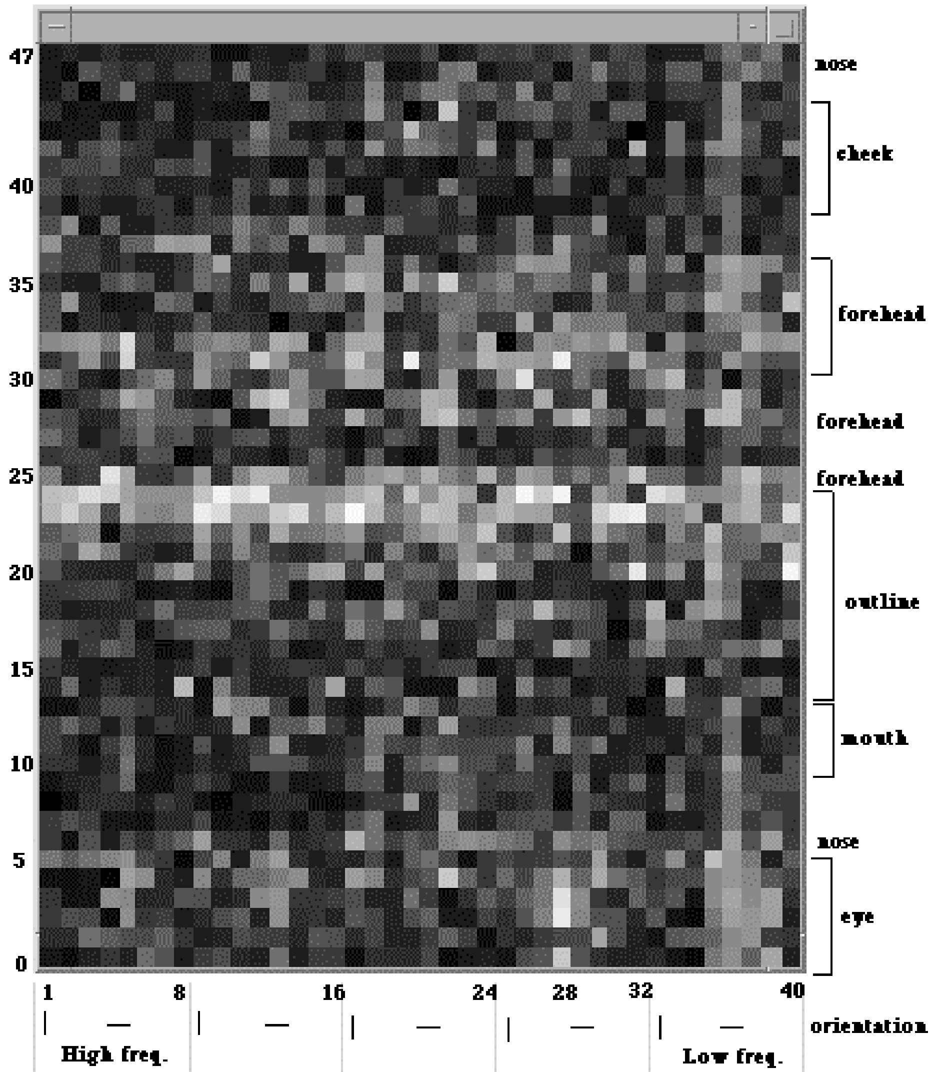

Fig. 2. Significance map (F map) of the 1920 kernels for Caucasian faces. Horizontal axis represents kernel size and orientation, vertical axis indicates kernel location. White and black areas show high and low discriminative power, respectively.

Fig. 2, but for a full reference on jet locations the reader is referred to Fig. 1. The fact that the hairline and forehead region score so highly is in accordance with some neurophysiological observations with monkeys where responses of most face sensitive cells showed dependence on the shape of the hairline and size of the forehead region of the test faces [5]. The low discriminability of the outline locations might indicate that these fiducial points are 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 have no 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 powers of different frequency levels, although on average horizontal kernels seem to have somewhat higher F-values than vertical ones, especially in the eye region. These results 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 let us 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 $\times$ 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 Fig. 3. 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

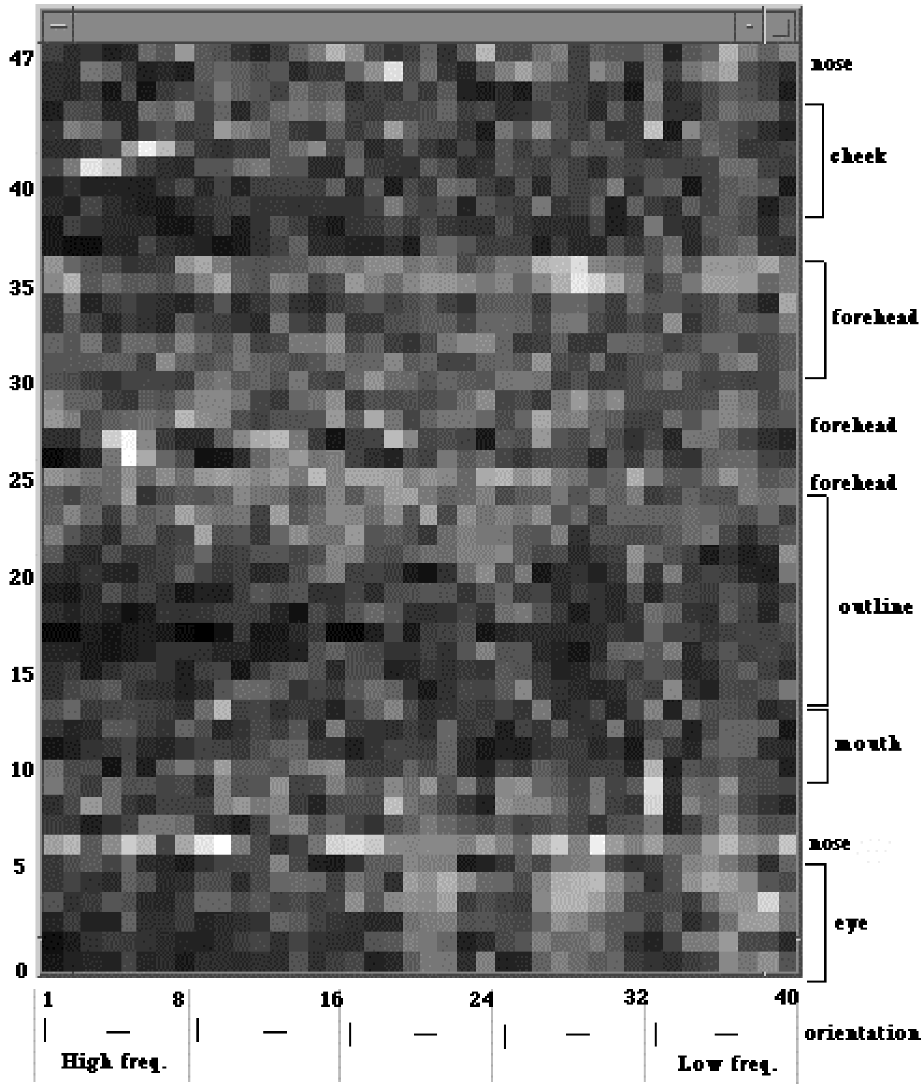

Fig. 3. Significance map (F map) of the 1920 kernels for Japanese faces. Horizontal axis represents kernel size and orientation, vertical axis indicates kernel location. White and black areas show high and low discriminative power, respectively.

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 (the highest was $F(100, 1313) = 25$, the lowest was $F(100, 1313) = 1.5$). Although some of the observed differences between the significance matrices of Caucasian and Japanese faces are probably caused by the peculiarities of the two datasets, it is likely that a significant portion of that difference is indeed caused by differences in the statistics of Caucasian and Japanese faces. Even though the individual conditions in the two databases are different, let us not forget that both contain variations over horizontal, vertical orientation, expression and illumination.

3.4. Recognition test of the weight matrices

Once the weight matrices were obtained, the hypothesis that weighting the kernel activation values would increase recognition performance was tested. 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 [7]. The second image for each individual contained some modification to the first one (e.g. change in background, lighting and expression). Without the weight matrix the number of failures in the matching process was 10, but with the weight matrix failures dropped to eight (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,

Table 4
Number of failures with different weight matrices

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

whereas with weights failures dropped to 147, which is a 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 nine, 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 a better result than that 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' were 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 has 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.

Of course, by extending on the idea of finding a set of weights that work better for different races one could also find a unique weighting scheme for every individual in the database [8]. However, to arrive at such a unique, but also efficient weight matrix for every individual, one would also need significant amount of examples for every face in the set, which in many cases is not available.

Selecting the kernels with the highest discriminability could be particularly useful when a 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, the recognition rate was still at 93%; with 40 kernels it was 90%. When only 10 highly informative kernels were kept

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

	High discriminative power (%)	Low discriminative power (%)
All 1920 kernels	96	96
100 kernels	93	83
40 kernels	90	74
10 kernels	73	32

in the representation the recognition rate was still up to 73%, whereas if we chose low informative kernels the recognition rate dropped to 32%. This shows the usefulness of statistical analysis of kernel activation values for data compression.

The double cross-validation technique was used to access the validity of the weight matrices. The Japanese face database was randomly split into two almost equal parts. Images of 51 individuals were in one group and the remaining 50 individuals were in the other. Then the recognition performance of the system was tested on both image sets without weights, with weights obtained from analyzing all images and also with weights obtained from analyzing the same and also the other database. Even when weights were obtained from the other image set, performance still showed significant improvement compared with the no weight condition (Fig. 4).

4. Conclusions

A statistical method has been suggested for the analysis of face representation by which it is possible to 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. Compression by a factor of 48 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 of the phenomenon called the 'other race effect' was found. It

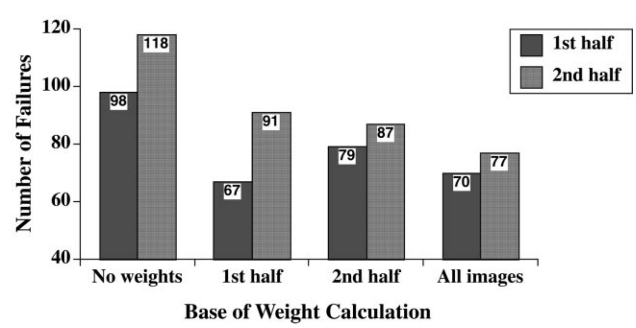

Fig. 4. Performance of the face recognition system without weights and with three different sets of weights.

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 statistical analysis described also allowed for the 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.

References

[1] M. Lades, J.C. Vortbrüggen, J. Buhmann, J. Lange, C. von der Malsburg, R.P. Würtz, W. Konen, Distortion invariant object recognition in the dynamic link architecture, IEEE Transactions on Computers 42 (1993) 300–311.

[2] V. Bruce, M. Burton, Processing Images of Faces, Ablex, Norwood, NJ, 1992.

[3] E. Elagin, J. Steffens, H. Neven, Automatic real-time pose estimation system for human faces based on bunch graph matching technology, in: Proceedings of the Third International Conference on Automatic Face and Gesture Recognition, Nara, Japan, pp. 136–141.

[4] J. Steffens, E. Elagin, H. Neven, PersonSpotter—fast and robust system for human detection, tracking and recognition, In: Proceedings of the Third International Conference on Automatic Face and Gesture Recognition, Nara, Japan, pp. 136–141.

[5] M.P. Young, S. Yamane, Sparse population coding of faces in the inferotemporal cortex, Science 256 (1992) 1327–1331.

[6] R.J. Watt, Visual Processing: Computational, Psychophysical and Cognitive Research, Lawrence Erlbaum, Hillslade, NJ, 1988.

[7] P.J. Phillips, H. Moon, P. Rauss, S.A. Rizvi, The FERET evaluation methodology for face-recognition algorithms, In: Proceedings of Computer Vision and Pattern Recognition, Puerto Rico, 1997, pp. 137–143.

[8] B. Duc, S. Fischer, J. Bigun, Face authentication with sparse grid Gabor information, IEEE Proceedings of ICASSP, Munich, Germany, Vol. 4, pp. 3053–3056.