Face Recognition on the Connection Machine CM-5

N. Petkov, P. Kruizinga, T. Lourens

Department of Mathematics and Computer Science
Rijksuniversiteit Groningen
P.O. Box 800, 9700 AV Groningen
The Netherlands

Abstract

A biologically motivated compute intensive approach to computer vision is developed and applied to the problem of face recognition. The approach is based on the use of two-dimensional Gabor functions that fit the receptive fields of simple cells in the primary visual cortex of mammals. A descriptor set that is robust against translations is extracted and used for a search in an image database. The method was applied on a database of 205 face images of 30 persons and a recognition rate of 94% was achieved. The final version of the paper will report on the results obtained by applying a set of 1024 Gabor functions on a database of 1000 face images of 150 persons and on the implementation on a Connection Machine CM-5 parallel supercomputer to be installed at our university until the end of 1992.

1 Introduction

The advent of parallel supercomputers promoted high-speed computing in the many Giga floating-point operations per second (Gflops/s, G=10^9) domain and the first Tflops/s (T=10^{12}) supercomputers are shortly expected. The awareness of the new possibilities offered by high-performance computers has led to considerable progress in computational natural and engineering sciences but at the same time left relatively untouched the research activities in the area of artificial intelligence. One possible explanation of this phenomenon might be the widely spread opinion that number crunching is less relevant in this area.

Artificial neural networks offer qualitatively new possibilities in this direction. The computation of the net inputs for multilayer feedforward neural networks, for instance, is substantially a matrix-vector multiplication. The weight corrections which are done in the learning process can also be considered as matrix operations [1-2]. New advanced learning techniques used to improve the convergence rate are based on well-known numerical techniques such as conjugate gradient. In other words, neural networks give the opportunity to give numerical formulation to non-numerical problems and, in this way, make use of supercomputer performance and the wealth of results and parallel algorithms available in the numeric computations area.

The progress, which has been achieved in the recent years in the area of artificial neural networks, has among others led to the now generally shared insight that representation and network structuring are application dependent choices that can have crucial effect on the success of this approach. With respect to mimicing the abilities of the human brain, an ensuing task and a challenge for computer scientists working in this area might be the development and verification of biologically motivated neural network models which use as a basis neurophysiologic data and give the opportunity for non-destructive exploration of the deeper brain structures. With this principal attitude in mind we approached anew the problem of computer vision and in particular the problem of automatic face recognition.

This problem has been considered to be a challenge since the very first days of computer vision. One of the first approaches to this problem was based on geometric features, such as size and relative positions of eyes, mouth, nose and chin [3-6]. Another basic technique is template matching which has reached a considerable level of sophistication [7-9]. Further approaches to face recognition use graph matching [10], Karhunen-Loewe expansion [11,12], algebraic moments [13], isodensity lines[14], etc. Connectionists approaches to the problem are described in [15-18]. We refer the reader to [19] for a comprehensive discussion of various aspects of face recognition and to [20] for a collection of recent works in this area.

Our approach is biologically motivated and based on the use of Gabor functions which have been shown to model very well the receptive profiles of the majority of simple cells in the primary visual cortex of mammals. The data obtained by projecting a two-dimensional signal (image) onto a set of Gabor functions can be interpreted as the activities of individual cells in the primary visual cortex (area V1 of the human brain). This data is then reduced to obtain a representation in a lower dimension space and use it for database storage and searching. We use an extended set of Gabor functions: 8 orientations and 8 scales give rise to a set of 64 Gabor functions and one copy of this set is centered on each point of the visual field. The compute intensiveness of the approach is due to the large resulting number of Gabor functions onto which an input image has to be projected.

The paper is organized as follows: In Section 2 we introduce the reader to two-dimensional Gabor functions and their relation to the mammalian visual system and propose a simple model for descriptor extraction. Section 3 presents our experimental setup and results on face recognition. Section 4 summarizes the approach and the results and outlines planned future work.

2 Gabor functions for computer vision

Our approach is biologically motivated in that it mimics the image transforms which take place in the mammalian visual cortex. It is well known from neurophysiologic research that a large amount of neurons in the primary visual cortex react strongly to short oriented lines [21,22].

A more precise study shows that the receptive field of such a cell and actually of the overwhelming majority of the simple cells in the primary cortex can be fit very well by a two-dimensional Gabor function [23-26]. Figure 1 shows the receptive fields of a few such cells with the corresponding Gabor functions.

The basic two-dimensional Gabor function has the following form:

$$g(x,y) = \frac{1}{\pi} e^{-(x^2+y^2)+i\pi x} \tag{1}$$

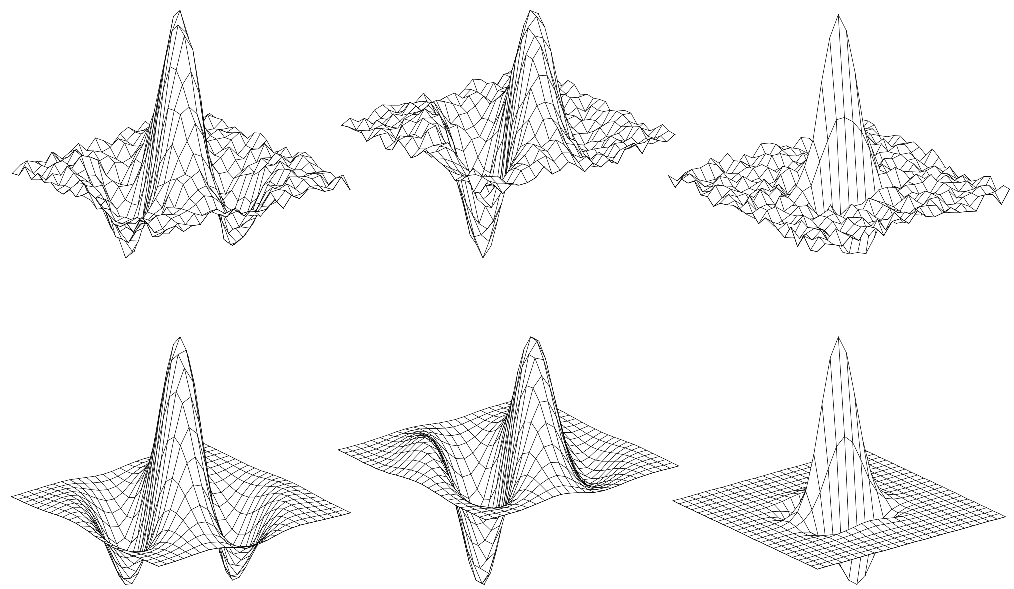
By means of translations parameterized by a pair (ξ, η), delations parameterized by an integer j and rotations parameterized by an angle φ, one gets the following family of two-dimensional Gabor functions (ξ and η have the same domain as x and y, respectively):

$$g_{j,\varphi}^{(\gamma)}(x-\xi, y-\eta) = \frac{\sqrt{\gamma}}{\pi}\alpha^{2j} e^{-\alpha^{2j}(\gamma^2 x'^2 + y'^2)+i\pi\alpha^j x'} \qquad (j \in \mathbb{Z}, \varphi \in [0,\pi)) \tag{2}$$

$$x' = (x-\xi)\cos\varphi + (y-\eta)\sin\varphi$$

$$y' = -(x-\xi)\sin\varphi + (y-\eta)\cos\varphi$$

where γ is an additional parameter called the *spatial aspect ratio*. Fig. 1b shows the real and imaginary parts of two Gabor functions with different scales and orientations. By comparing Fig. 1a and Fig. 1b one can see that both the real and imaginary part of a Gabor function are represented by cells of the visual cortex. According to neurophysiologic data, such cells are even adjacent [27].



a) b)

Figure 1: Receptive fields of simple cells in the primary visual cortex (a) and the respective Gabor functions which approximate them (b) (adapted from [26]).

The oscillations of $g_{j,\varphi}(x-\xi, y-\eta)$ are due to the harmonic factor $e^{i\pi\alpha^j x'}$ with wavelength

$$\lambda_j = \frac{2}{\alpha^j} \tag{3}$$

and frequency

$$f_j = \pi\alpha^j . \tag{4}$$

The Gaussian factor $e^{-\alpha^{2j}(\gamma^2 x'^2 + y'^2)}$ causes the function $g_{j,\varphi}(x-\xi, y-\eta)$ to be negligible for $|x-\xi| > \lambda_j$. The choice of taking the scaling factor in the form α^j $(j \in \mathbf{Z})$ corresponds to equidistant sampling of a logarithmic wavelength/frequency scale. that corresponds to the logarithmic dispersion of frequencies found by neurophysiologic research [23-26]. Further neurophysiologic data show that the receptive fields of the simple cortical cells have constant form that can be modeled by taking a constant spatial aspect ratio with the following value [23,24,26]:

$$\gamma = 2 \tag{5}$$

In the following, we use this spatial aspect ratio, write concisely $g_{j,\varphi}(x-\xi, y-\eta)$ and mean $g^{(2)}_{j,\varphi}(x-\xi, y-\eta)$.

The projection of a two-dimensional signal (image) $s(x,y)$ on a Gabor function $g_{j,\varphi}(x-\xi, y-\eta)$

$$\tilde{s}_{j,\varphi}(\xi,\eta) = \int s(x,y) g^*_{j,\varphi}(x-\xi, y-\eta) dx dy \tag{6}$$

may be considered as the amount of a harmonic wave with wavelength $\lambda_j = \frac{2}{\alpha^j}$ and orientation φ in a surrounding of linear size λ_j of a point (ξ, η). In this way, equation (6) represents local spectral analysis with frequency j (logarithmic magnitude) and orientation φ that is embedded in global spatial coordinates (ξ, η). The fact that the receptive fields of a large number of cells in the visual cortex can be almost perfectly fitted by members of the family of Gabor functions is striking, in that the visual cortex seems, in a way, to make a local spectral analysis embedded in global spatial coordinates.

In the following we use (6) as the basis of our approach, assuming that the values $\tilde{s}_{j,\varphi}(\xi,\eta)$ delivered for the various values of the parameters j, φ, ξ and η correspond to the activities of individual cortical cells when the visual system is presented an image $s(x,y)$.

Note that for fixed j and variable ξ and η the above quantities present a function $\tilde{s}_{j,\varphi}(\xi,\eta)$ which can be computed as convolution of the signal $s(x,y)$ with a function $g_{j,\varphi}(x-\xi, y-\eta)$. We use this fact for the efficient computation of these quantities using fast Fourier transform (FFT). The coefficient $\frac{\sqrt{\gamma}}{\pi}\alpha^{2j}$ in front of the exponent in (2) is a normalization factor that is chosen in such a way that

$$|\int e^{i\pi\alpha^j x'} g^*_{j,\varphi}(x-\xi, y-\eta)dxdy| = 1, \tag{7}$$

i.e. for the signal $s(x,y) = e^{i\pi\alpha^j x'}$ with magnitude 1, which is the harmonic factor in $g_{j,\varphi}(x-\xi, y-\eta)$, the above normalization delivers a quantity of magnitude one, $|\tilde{s}_{j,\varphi}(\xi,\eta)| = 1$.

Note that the functions $\tilde{s}_{j,\varphi}(\xi,\eta)$ comprise more data than the original image $s(x,y)$. This is in contrast to traditional approaches to computer vision where the amount of data is reduced at each stage of a hierarchical image analysis process. At present, neither we nor neurobiologists seem to know how all this information is used to effectively recognize an object. What one is certain about is that this data expansion is actually carried out in the brain. What one may wish to do on a computer is simulate this expansion, make hypotheses about the further processing stages and verify them. As we shall see below, even a very naive model delivers startling results. Let us consider the following quantities:

$$S_{j,\varphi} = \int \mid \tilde{s}_{j,\varphi}(\xi,\eta) \mid d\xi d\eta, \qquad j \in \mathbb{Z}, \varphi \in [0,\pi). \tag{8}$$

Each of them presents the cumulative activity of all cells with the same orientation φ and frequency center j independently of their positions (ξ,η) in the visual field. The naive premise is that cells doing similar things (in this case cells with identical receptive fields but responsible for different areas of the visual field) might contribute in a similar way to quantities computed at higher stages. Each of the quantities (8) might, for instance, correspond to the activity of a corresponding higher abstraction level cell that receives activating stimuli from all lower level cells with the same orientation φ and frequency center j. We have to admit that we are not aware of any neurobiologic evidence that confirms this hypothesis. Computing the quantities $S_{j,\varphi}$ according to (8) might make sense for one reason: they are not sensitive to the particular position of an object in the visual field, property which we refer to as *translational invariance*. More precisely, if $s(x,y)$ and $s'(x,y)$ are two images such that

$$s'(x,y) = s(x-t_1, y-t_2) \tag{9}$$

i.e. $s'(x,y)$ is produced by shifting $s(x,y)$ by a constant vector (t_1,t_2), one can easily show that for the respective quantities $S_{j,\varphi}$ and $S'_{j,\varphi}$ holds

$$S'_{j,\varphi} = S_{j,\varphi}. \tag{10}$$

Let us now represent two images $s(x,y)$ and $w(x,y)$ by the respective sets of quantities $S_{j,\varphi}$ and $W_{j,\varphi}(j \in \mathbb{Z}, \varphi \in [0,\pi))$ according to (8) and define the *dissimilarity* of the two images as follows

$$D_{s,w} = \sum_{j,\varphi} \mid S_{j,\varphi} - W_{j,\varphi} \mid . \tag{11}$$

The above defined dissimilarity is a non-negative quantity. It is zero for two identical images and for any two images which differ only by a translation as defined by (9). The relations (8) and (11) are the basis of our approach to automatic face recognition. The quantities (8), to be referred to in the following as the *descriptors*, are computed for all images in a database and for each new input image. The descriptor set of an input image is then used for a best match search in the database to find the prestored image for which the dissimilarity (11) is minimal.

3 Implementation and Results

We applied the above developed approach to the problem of face recognition. For this purpose, a database of face images has been built and this database is still being extended. The results reported below refer to the time when the database comprised 205 different face images of 30 persons. Several images were taken of each person, with the exact number varying from 5 to 9. The individual images of each person exhibit differences in facial expressions and/or orientations. Similarly, the individual images of one person show small deviations in size (a tolerance of approximately 5-10%) due to the fact that the distance between a subject and the camera was not controlled to keep it exactly constant. (From person to person, there are size deviations of up to 20%). Illumination was strived to be constant from session to session but no special effort was given to achieve exactly the same illumination conditions. Deviations in the illumination were due to changes in the position of the lamps and by sun light coming through the windows. The face pictures are stored as graylevel images with spatial resolution of 500×400 pixels and 8-bit quantization (256 graylevels).

Discretization is necessary for the practical computation of of the quantities $\tilde{s}_{j,\varphi}(\xi, \eta)$ according to (6) and we use the following one:

$$x, \xi = 1, 2, \ldots, 500 \tag{12}$$

$$y, \eta = 1, 2, \ldots, 400 \tag{13}$$

$$\varphi = \varphi_k = k\frac{\pi}{8}, \qquad k = 0, 1, \ldots, 7 \tag{14}$$

$$j = -1, -2, \ldots, -8 \tag{15}$$

The basic scaling factor α was taken as follows:

$$\alpha = \sqrt{2}. \tag{16}$$

This choice of α and the range of the parameter j allow for covering a wavelength domain that ranges from 2 to 32 pixels with logarithmic dispersion of the wavelength averages of the basic Gabor functions (see (3)).

The convolutions (6) were computed by applying FFT. In spite of the efficiency of FFT, the convolution computation is quite compute intensive. On a 17 Mflops/s workstation, half a minute is required to compute the convolution of an image with one of the Gabor functions . For the set of 64 Gabor functions used this amounts to half an hour exclusive (non-timesharing) use of the workstation. This explains the relatively rough angle discretization (eight orientations) and the limited amount of (eight) basic frequencies that were used.

After computing a convolved image $\tilde{s}_{j,\varphi}(\xi, \eta)$ for a given input image $s(x, y)$ and a given Gabor function $g_{j,\varphi}(x, y)$, the convolved image is reduced to a single number $S_{j,\varphi}$ according to (8). In this way 64 numbers (descriptors) are computed for each input image, one number for each of the 64 basic Gabor functions, and only this information is used to represent the image for database searching.

To obtain statistics on the recognition rate, we applied the above approach to all images in the database, considering each image in turn as an input image and the rest as prestored images. For each image the first four nearest neighbours were determined. For 192 out of 205 images the search was successful as illustrated by Fig. 2. The model failed in 13 cases as illustrated by Fig. 3. This gives a recognition rate of approximately 94%.

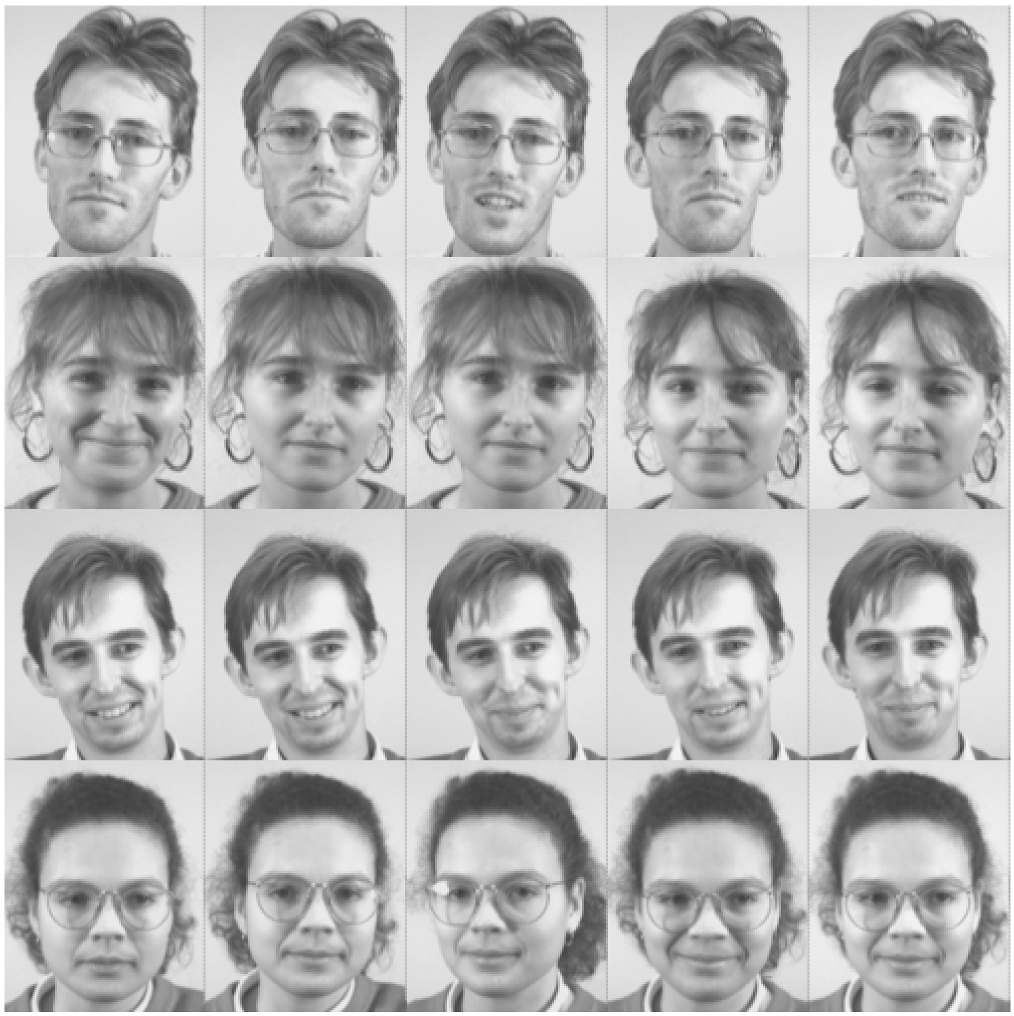


Figure 2: Examples of successful matches: the leftmost image in each row is a test image for which best match search is done in the whole image database; the images right to it are the first four best matches.

Figure 3: Examples of failure of the model.

4 Summary and Future Plans

In this paper we have shown how a biologically motivated model can be used for automatic face recognition. The biological relevance refers to the use of Gabor functions that fit the receptive fields of the overwhelming number of simple cells in the primary visual cortex of mammals. In the rest of its part, the approach is an attempt to guess what might be happening in the further form analysis structures of the visual cortex. In this case, we have no neurobiologic data to build on and, therefore, we rely only on general principles such as achieving robustness for image translations. Besides this uncertainty in the biological relevance of the final processing stages, our model comprises a certain simplification of the earlier stages: only 8 orientations and 8 basic frequencies are used. In spite of these shortcomings, we achieve a recognition rate of 94% on a database of 205 face images of 30 persons, a result which is startling with respect to previous work in this area and the short time we have been dealing with the problem until now.

Our future work on this problem will focus on:

(i) *Increasing the number of Gabor functions used*: A larger number of Gabor functions is needed to improve the sensitivity of the model and to explore the possibilities for rotation and scaling compensation. We intend to proceed with 32 orientations and 32 wavelengths which amount to 1024 Gabor functions.

(ii) *Improving the model in its higher stages*: The current model is an oversimplification of the processes taking place in the higher form-analysis structures of the visual cortex. (Actually, we cannot even say that, since in this part our model is heuristic and is not supported by any neurobiologic data). We intend to improve the model by introducing local sensitivity by decomposing an input image into parts and applying the model to each individual part. This, however, will also be heuristics and we see a reasonable solution to the problem only in getting new neurobiologic data that can confirm or reject our hypothesis and give us hints on the ways to go.

(iii) *Parallel supercomputer implementation*: The computational intensiveness of the approach has become inhibitive for further investigations of the model. The use of 64 Gabor convolvers requires 30 minutes per image and the use of 1024 functions requires 8 hours per image (on a 17 Mflops/s workstation). This means one year computing time on a database of 1000 images, a delay which is unacceptable with respect to the fact that we are interested in experimenting with the model and have every day a new idea how it may be changed. The use of a supercomputer is inevitable and we already port our programs to a Connection Machine CM-5 which will be installed at our university until the end of the year. (Currently, the Connection Machines CM-2 and CM-5 of the University of Wuppertal are used for code transfer).

References

[1] N. Petkov: "Systolic simulation of multilayer, feedforward neural networks", *Proc. Int. Conf. on Parallel Processing in Neural Systems and Computers*, Düsseldorf, 1990, ed. by R. Eckmiller, G. Hartmann and G. Hauske (Amsterdam: North-Holland, 1990) pp. 303-306.

[2] N. Petkov: *Systolic Parallel Processing* (Amsterdam: North-Holland, Elsevier Sci. Publ., 1992).

[3] W.W. Bledsoe: "Man-machine facial recognition", Technical Report PRI:22, Panoramic Research Inc., (Paolo Alto, CA, 1966).

[4] A.J. Goldstein, L.D. Harmon, and A.B. Lesk: "Identification of human faces", In *Proc. IEEE*, Vol. 59 (1971) pp. 748.

[5] T. Kanade: "Picture processing by computer complex and recognition of human faces", Technical Report, Kyoto University, Dept. of Information Science, 1973.

[6] Y. Kaya and K. Kobayashi: "A basic study on human face recognition", in S. Watanabe (ed.) *Frontiers of Pattern Recognition* (1972) pp. 265.

[7] D.J. Burr: "Elastic matching of line drawings", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 3 (1981) No. 6, pp. 708-713.

[8] J. Buhmann, J. Lange, and C. von der Malsburg: "Distortion invariant object recognition by matching hierarchically labeled graphs", *Proceedings of IJCNN'89* (1989) pp. 151-159.

[9] A.L. Yuille: "Deformable templates for face recognition", *Journal of Cognitive Neuroscience*, Vol.3 (1991) No.1, pp. 59-70.

[10] B.S. Manjunath, R. Chellappa, and C. von der Malsburg: "A feature based approach to face recognition", *Proc. 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* Champaign, Illinois, June 1992, pp. 373-378

[11] M. Turk and A. Pentland: "Eigenfaces for recognition", Technical Report 154, MIT Media Lab Vision and Modelling Group, 1990.

[12] M. Turk and A. Pentland: "Face recognition using eigenfaces", *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Maui, Hawaii, June 1991, pp. 586-591.

[13] Zi-Quan Hong: "Algebraic feature extraction of image for recognition" *Pattern Recognition* Vol. 24 (1991) No.3, pp. 211-219.

[14] O. Nakamura, S. Mathur, and T. Minami: "Identification of human faces based on isodensity maps", *Pattern Recognition*, Vol. (1991) No.3, pp.263-272.

[15] M. Lades, J.C. Vorbrüggen, J. Buhmann, J. Lange, C. von der Malsburg, R.P. Würtz, and W. Konen: "Distortion invariant object recognition in the dynamic link architecture", 1991 (preprint).

[16] T. Kohonen: *Self-Organization and Associative Memory*, (New York: Springer Verlag, 1989).

[17] A. Fuchs and H. Haken: "Pattern recognition and associative memory as dynamical processes in a synergetic system II". *Biological Cybernetics*, Vol. 60 (1988), pp. 107-109.

[18] G. Cottrell and M. Fleming: "Face recognition using unsupervised feature extraction", *Proceedings of the International Neural Network Conference*, 1990.

[19] V. Bruce and M. Burton: "Computer recognition of faces". in *Handbook of Research on Face Processing*, A.W. Young and H.D. Ellis (eds.), (Amsterdam: Elsevier Sci.Publ., 1989) pp. 487-506.

[20] A.W. Young, and H.D. Ellis (eds.): *Handbook of Research on Face Processing*, (Amsterdam: Elsevier Sci. Publ., 1989).

[21] D. Hubel and T. Wiesel: "Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex", *J. Physiol.(London)*,1962, vol. 160, pp. 106-154.

[22] D. Hubel and T. Wiesel: "Sequence regularity and geometry of orientation columns in the monkey striate cortex", *J. Comput.Neurol.*, Vol. 158 (1974) pp. 267-293.

[23] J. Jones and L. Palmer: "An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex", *Journal of Neurophysiology*, Vol.58 (1987) pp. 1233-1258.

[24] J. Daugman: "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters", *J. Opt.Soc.Amer.*, Vol.2 (1985) No. 7, pp. 1160-1169.

[25] J. Daugman: "Two-dimensional spectral analysis of cortical receptive field profiles", *Vis.Res.*, Vol. 20 (1980) pp. 847-856.

[26] J.G. Daugman: "Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression", *IEEE Trans. on Acoustics, Speech and Signal Processing*, Vol.36 (1988) No. 7, pp. 1169-1179.

[27] D. Pollen and S. Ronner: "Phase relationships between adjacent simple cells in the visual cortex", *Science*, Vol. 212 (1981) pp. 1409-1411.