Orientation Competition in Cortical Filters - An Application to Face Recognition

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, computationally intensive approach to computer vision is developed and applied to the problem of automatic face recognition. The approach is based on the use of two-dimensional Gabor functions which model the receptive field functions of simple cells in the primary visual cortex of mammals. The convolutions of an input image with a set of antisymmetric visual receptive field functions (imaginary parts of Gabor functions) become the subject of thresholding and orientation competition. The developed cortical filters deliver highly structured information which is used for efficient feature extraction and representation in a lower dimension space. Applied to face recognition, the method gives a recognition rate of 98.5% on a large database of face images.

1 Introduction

Large scale computer simulations are nowadays a well established research tool in natural and engineering sciences such as physics, chemistry, astronomy, fluid dynamics, electrical engineering, etc. In the 1990s, proclaimed to be the decade of the brain, the insights in the microstructure of the brain provided by neurophysiological and neurobiological research together with the progress in mathematical models of artificial neural networks may open new opportunities for computational science. In the years to come large scale computer simulations may become an instrument of neuroscience, a task that they successfully fulfil for a number of years in the other branches of science mentioned above. The chances offered in this area may even turn out to be unique in certain respects, since in neuroscience the need for non-destructive exploration methods is at least as high as in the other sciences mentioned. Helping setting up computational models, which are inspired by natural information processing systems, may become an ensuing task and a challenge for computer scientists. Studying and modelling information processing in living nature and helping discover the underlying mechanisms may contribute to strengthening the exploratory role of computer science, next to its more traditional tasks of solving engineering problems and developing formalisms for empirically grown methods and practical solutions.

Neurophysiological research has delivered a number of interesting results which can serve as a starting point for computational research. For quite a time it has, for instance, been well-known that a large amount of neurons in the primary visual cortex of mammals react strongly to short oriented lines [1, 2]. More precise studies carried out in the 1980s have shown that the receptive fields of such cells can be fitted well by two-dimensional Gabor functions [3, 4], differences of offset Gaussians or other similar functions [5]. Basing on these results, one can mimic the work of the primary visual cortex by computing the activation of each individual simple cell for a given image projected on the retina. This approach, sometimes popularly referred to as 'computing cortical filters', has been the subject of intensive research in the recent years. The results obtained until now give rise to a number of open questions. Among these we consider as most important the question of how the output of cortical filters can be used to analyse images and recognize objects. A basic problem we encounter in our attempts to find an answer to this question is that of whether and how cortical filters have to interact with each other in order to facilitate structuring of information. Here we propose a cortical filter model in which the output of antisymmetric receptive field convolvers becomes the subject of thresholding and orientation competition. The resulting

cortical filters deliver highly structured information which can be used for the extraction of efficient features and image representation in a lower dimension space.

Face recognition, a problem that has been considered to be a challenge since the very first days of computer vision, recently experiences a revival. 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 [6, 7, 8]. Other basic techniques, which have reached a considerable level of sophistication, are template and graph matching [9, 10, 11, 12]. Further approaches to face recognition use Karhunen-Loewe expansion [13], algebraic moments [14], isodensity lines [15], etc. Connectionists approaches to the problem are described in [12, 16, 17, 18, 19]. We refer the reader to [20] for a comprehensive discussion of various aspects of face recognition and to [21] for a collection of works in this area.

The present work is an extension of our previous work reported in [22, 23, 24]. By using improved cortical filters, we currently achieve a recognition rate of 98.5% on a database of 205 images of 30 persons. The rest of the paper is organized as follows: In Section 2 we introduce the reader to two-dimensional Gabor functions and their relation to natural vision. In Section 3, a winner-takes-all orientation competition between different cortical filter channels is introduced as a means to improve the sensitivity to edge orientation. Section 4 outlines the transition from cortical images to a representation in a lower dimension space used for image comparison and database searching. Section 5 presents our results on face recognition. In Section 6, concluding remarks are given and ensuing questions are raised.

2 Visual receptive field functions

The *receptive field* of a visual cortical cell is the area of the visual field, typically measured in degrees, in which a stimulus can influence the response (firing rate) of the cell. The *receptive field function* specifies the effect of a small white spot (impulse) on the response of the cell as a function of the spot position. This effect can be positive (excitory) or negative (inhibitory). In the latter case, it is implicitly assumed that some other stimulus of lower intensity, e.g. constant illumination, random noise, an edge or some other structure, brings the cell at a given excitation level so that it is possible to measure the inhibitory effect. Hence, strictly speaking, one cannot consider a receptive field function, which is experimentaly determined using some background excitation, to be the impulse response of the respective cell. In order to prevent confusion, in the following we use the term *impulse pseudo-response* for such receptive field functions.

2.1 Gabor functions and natural vision

The basic two-dimensional Gabor function we use as a model of the receptive field functions of visual cortical cells has the following form (L is the linear size of an $L \times L$ square part of the visual filed to be referred to in the following as 'the input image'):

$$g(x,y) = \frac{1}{\pi} e^{-(x^2+y^2)+i\pi x} \qquad (x,y \in [0,L]) \tag{1}$$

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

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

$$\begin{aligned} x' &= \ \ (x-\xi)\cos\varphi + (y-\eta)\sin\varphi \\ y' &= -(x-\xi)\sin\varphi + (y-\eta)\cos\varphi \end{aligned}$$

Fig.1 shows the real and imaginary parts of one such function. Its real part

$$\Re g_{j,\varphi}(x-\xi, y-\eta) = \frac{\alpha^{2j}}{\pi} e^{-\alpha^{2j}(x'^2+y'^2)} \cos(\pi\alpha^j x') \qquad (j \in \mathbf{Z},\ \varphi \in [0,\pi),\ \xi,\eta \in [0,L]) \tag{3}$$

is symmetric while the imaginary part

$$\Im g_{j,\varphi}(x-\xi, y-\eta) = \frac{\alpha^{2j}}{\pi} e^{-\alpha^{2j}(x'^2+y'^2)} \sin(\pi\alpha^j x') \qquad (j \in \mathbf{Z},\ \varphi \in [0,\pi),\ \xi,\eta \in [0,L]) \tag{4}$$

is antisymmetric for inversions with respect to the centre (ξ, η). We use this family of functions as models of the receptive field functions of simple visual cortical cells. As demonstrated by neurophysiological findings, there are visual cortical cells whose impulse pseudo-responses can be modelled by the real part (3) of such a Gabor function and others whose impulse pseudo-response can be modelled by the imaginary part (4). Furthermore, a number of pairs of such cells with the same receptive field and a phase difference of $\pi/2$, as if representing the two parts of a single complex Gabor function (2), have been found to be spatial neighbours in the cortex [25]. As to the parameters (ξ, η), φ and j in eqs. (2-4), they specify the centre (ξ, η) of a receptive field, its preferred orientation φ (see below) and size $(O(\alpha^{-j}))$, respectively.

The oscillations of $g_{j,\varphi}(x-\xi, y-\eta)$, which exhibit themselves in neurophysiological measurements as a number of parallel alternate excitory and inhibitory stripe zones in the receptive field, are due to the harmonic wave factor $e^{i\pi\alpha^j x'}$ with wavelength

$$\lambda_j = \frac{2}{\alpha^j} \tag{5}$$

and wavevector (spatial frequency) $\vec{k_j}$ with orientation φ, to be referred to as *the preferred orientation* of a receptive field, and magnitude

$$k_j = \pi\alpha^j. \tag{6}$$

The Gaussian factor $e^{-\alpha^{2j}(x'^2+y'^2)}$ causes the function $g_{j,\varphi}(x-\xi, y-\eta)$ to be negligible at distances from the centre (ξ, η), which are larger than a few wavelengths λ_j, and thus determines the size $O(\alpha^{-j})$ of the receptive field. The choice of taking the scaling factor in the form α^j $(j \in \mathbb{Z})$ corresponds to equidistant sampling of a logarithmic wavelength/frequency scale which reflects the logarithmic dispersion of spatial frequencies found by neurophysiological research [3, 4].

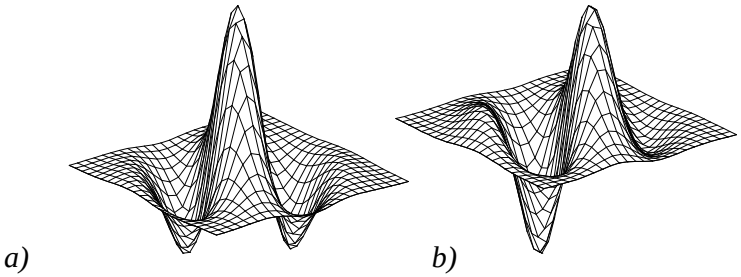

Figure 1: Real (a) and imaginary (b) part of a Gabor function.

2.2 Thresholding

The importance of the notion of impulse response in system theory is that, assuming linearity, one can compute the response of a system not only to an impulse in an arbitrary point but, by means of a weighted sum, to virtually any input stimulus. As already pointed out, the functions (3-4) do, unfortunately, *not* represent the actual impulse responses of visual cortical cells, but rather their above so-called impulse pseudo-responses in the presence of some background stimulus. Hence, the integrals[1] (the customary way to use the impulse response for computing the response to a composite signal)

$$\tilde{s}^s_{j,\varphi}(\xi, \eta) = \int s(x, y) \Re g_{j,\varphi}(x-\xi, y-\eta) dx dy \tag{7}$$

$$\tilde{s}^a_{j,\varphi}(\xi, \eta) = \int s(x, y) \Im g_{j,\varphi}(x-\xi, y-\eta) dx dy \tag{8}$$

do not represent the responses of a pair of cells characterized by the real (3) and imaginary (4) parts of $g_{j,\varphi}(x-\xi, y-\eta)$ to a two-dimensional input signal (image) $s(x, y)$ presented in the visual field. They rather represent the responses of the two cells to an input signal which is a superposition of $s(x, y)$ and the certain background stimulus which has been used to excite the cells in order to determine the receptive

[1] The superscripts ^{s} and ^{a} stand for symmetric and antisymmetric, respectively.

field functions as shown in Fig.1. The fact that, unless a background stimulus is present, the responses as modelled by the functions shown in Fig.1 are 'half-way rectified'[2] [26], suggests that the quantities computed in (7-8) should be considered as the *net inputs* to visual cortical cells, with the actual responses $r^s_{j,\varphi}(\xi,\eta)$ and $r^a_{j,\varphi}(\xi,\eta)$ being computed as follows by *thresholding* (here with a threshold zero)[3]:

$$r^s_{j,\varphi}(\xi,\eta) = \tilde{s}^s_{j,\varphi}(\xi,\eta)\chi(\tilde{s}^s_{j,\varphi}(\xi,\eta)) \ , \tag{9}$$

$$r^a_{j,\varphi}(\xi,\eta) = \tilde{s}^a_{j,\varphi}(\xi,\eta)\chi(\tilde{s}^a_{j,\varphi}(\xi,\eta)) \ . \tag{10}$$

Note that the complex quantity $\tilde{s}_{j,\varphi}(\xi,\eta) = \tilde{s}^s_{j,\varphi}(\xi,\eta) + i\tilde{s}^a_{j,\varphi}(\xi,\eta)$ can be consudered as the amount of a harmonic wave $e^{-i\pi\alpha^j x'}$ with wavelength λ_j and wavevector orientation φ in a surrounding of linear size $O(\lambda_j)$ centered on a point of the visual field with coordinates (ξ,η). In this way, equations (7-8) represent local spectral analysis which is embedded in global spatial coordinates (ξ,η). (The coefficient α^{2j}/π in (2) is a normalization factor which is chosen in such a way that, for a harmonic input signal $s(x,y) = e^{-i\pi\alpha^j x'}$ with magnitude one, the quantity $\tilde{s}_{j,\varphi}(\xi,\eta)$ also has magnitude one, $|\tilde{s}_{j,\varphi}(\xi,\eta)| = 1$.)

We use the above scheme as the basis for mimicing the function of the primary visual cortex, assuming that the values $r^s_{j,\varphi}(\xi,\eta)$ and $r^a_{j,\varphi}(\xi,\eta)$ delivered for the various values of the parameters j, φ, ξ and η correspond to the responses of individual cortical cells when the visual system is presented an image $s(x,y)$. For fixed j and φ and variable ξ and η, $r^s_{j,\varphi}(\xi,\eta)$ and $r^a_{j,\varphi}(\xi,\eta)$ are two-dimensional non-negative functions to be referred to as *cortical images*. Fig.2 shows a simple input image $s(x,y)$ for which Fig.3 and Fig.4 show the computed cortical images obtained for one fixed value of j (corresponding to $\lambda_j \approx L/40$ where L is the image size) and 16 different wavevector orientations $\varphi_i = 2\pi i/16,\ i = 0\ \ldots\ 15$.

Figure 2: A simple input image.

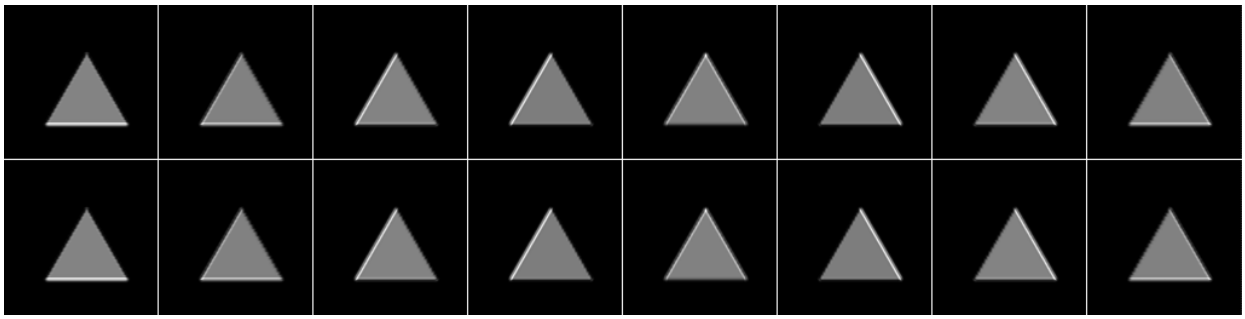

Figure 3: Cortical images $r^s_{j,\varphi}(\xi,\eta)$ computed with symmetric receptive field functions $\Re g_{j,\varphi}(x-\xi, y-\eta)$ of the same size $\lambda_j \approx L/40$ and different orientations φ. The images in the first and second row correspond, left to right, to $\varphi_i = 2\pi i/16,\ i = 0\ \ldots\ 7$, and $\varphi_i = 2\pi i/16,\ i = 8\ \ldots\ 15$, respectively.

Note that the computed cortical images $r^s_{j,\varphi}(\xi,\eta)$ and $r^a_{j,\varphi}(\xi,\eta)$ comprise more data than the original image $s(x,y)$: the use of eight values of j and 16 different orientations, for instance, leads to a growth of the amount of data by a factor of 128. This seems to be in contrast with traditional approaches to computer vision where the amount of data is reduced at each stage of a hierarchical image analysis process. At present, one cannot say how such a data expansion can be used to recognize an object. It is, however, certain that such a data expansion is actually carried out in the brain as inferred from the fact that the visual information is transferred from the retina to the primary visual cortex via 10^6 fibres of the optic nerve and in the primary visual cortex it is encoded by $10^8 - 10^9$ simple cells (100-1000 times

[2] Fig.7a illustrates the half-way rectification of an antisymmetric receptive field function. Compare Fig.1b.

[3] The step function $\chi(.)$, which is used in (9-10), is defined as follows: $\chi(z) = 1$ for $z > 0$, $\chi(z) = 0$ for $z \leq 0$.

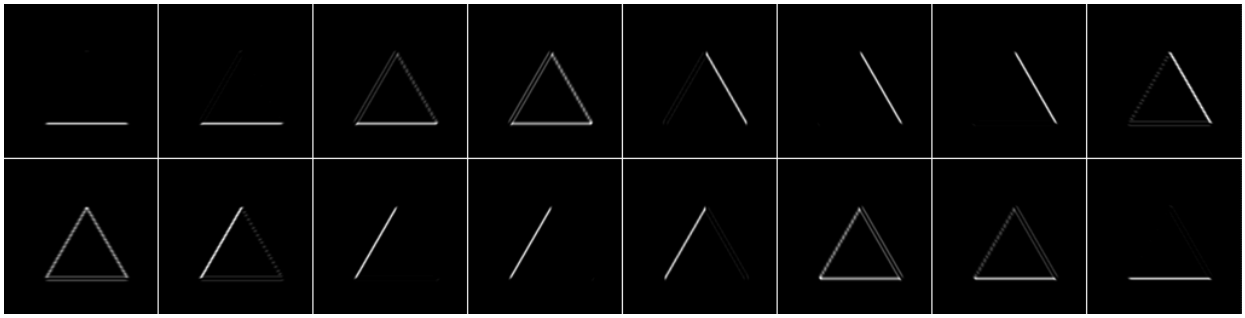

Figure 4: Cortical images $r^a_{j,\varphi}(\xi,\eta)$ computed with antisymmetric receptive field functions $\Im g_{j,\varphi}(x-\xi,y-\eta)$ of the same size $\lambda_j \approx L/40$ and different orientations φ. The images in the first and second row correspond, left to right, to $\varphi_i = 2\pi i/16,\ i = 0\ \ldots\ 7$, and $\varphi_i = 2\pi i/16,\ i = 8\ \ldots\ 15$, respectively.

expansion at cortical level [27]). We propose to simulate this expansion on a computer, make hypotheses about the further processing stages and evaluate the plausibility of a model by applying it to an object recognition problem.

2.3 Using only antisymmetric receptive field functions

In this study we use only the imaginary part (4) of the complex Gabor functions (2). This means, we are concerned only with those simple visual cortical cells which are characterised by antisymmetric receptive field functions. This seems to be sufficient for processing edges formed by the transition from dark to white areas: A cell of the considered type would react strongly to an intensity transition in its receptive field provided that the transition edge is parallel to the stripes of the receptive field function. In contrast, cells with symmetric receptive field functions, such as those modelled by the real part of a Gabor function, give response which is smaller than the zero current response[4] when their receptive fields are centered precisely on the edge. Furthermore, in this case the response is independent of the preferred orientation φ of the symmetric receptive field function. For such a cell to show sensitivity to intensity transitions, in particular to the orientation of the transition edge, the center of its receptive field has to be shifted from the edge at approximately half the radius of the receptive field. Hence, antisymmetric receptive field functions give a better indication on edge positions. As illustrated by the comparison of counterpart cortical images in Fig.3 and Fig.4, the edges enhanced in symmetric receptive field function cortical images (Fig.3) are also enhanced in antisymmetric receptive field function cortical images (Fig.4). As already mentioned, the latter images have the advantage of giving a better indication of edge positions. Therefore, as far as edges are concerned, one can use only antisymmetric receptive field functions without loss of information. One has to note that symmetric receptive field functions can be used for line detection [28] and, in combination with antisymmetric functions, for texture boundary detection [29]. These problems fall out of the scope of the present study.

3 Orientation competition

Proceeding now with the antisymmetric receptive field function cortical images only (Fig.4), we first note that the differences between cortical images, which correspond to neighbouring orientations of the respective receptive field functions, are not really large. In particular, the same edge is enhanced in more than one cortical image as illustrated, for instance, by the horizontal edge in the first three and the last two cortical images in Fig.4 which correspond to orientations 0, $\frac{1}{16}2\pi$, $\frac{2}{16}2\pi$, $\frac{3}{16}2\pi$, $\frac{14}{16}2\pi$ and $\frac{15}{16}2\pi$, respectively. The reason for this effect is given by the fact that, although the antisymmetric receptive field functions are sensitive to the orientation of an edge, this sensitivity is not sharply tuned. Fig.5a shows the computed response (10) of such a cell to an edge which passes through the centre of its receptive field as a function of the angle between the normal to the edge and the preferred orientation φ of the receptive field function. Strongest response is achieved when this angle is zero. Note, however, that for an angle of $\pm 30°$ the response is still considerable (60% of the maximum). With respect to the non-linear

[4] The so-called zero-current response is the response of a cell to a constant intensity exposure. It is illustrated by the gray level in the inside of the triangle in the cortical images shown in Fig.3.

trimming of the ultimate neuron response as modelled by sigmoid functions (not considered here), it would be difficult to determine the precise orientation of an edge basing only on the responses computed above. This would be in contrast with psychological and neurophysiological experiments that confirm high orientation sensitivity of the visual system of mammals.

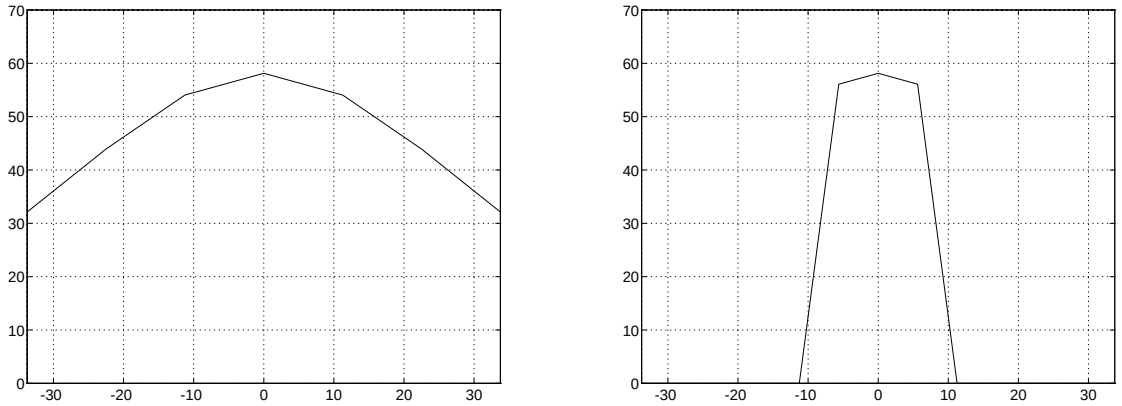

Figure 5: (a) Computed response of a cell with an antisymmetric receptive field function to an edge which passes through the centre of the receptive field as a function of the angle between the normal to the edge and the preferred orientation of the function. (b) The introduction of a winner-takes-all competition between cells corresponding to different orientations leads to a sharper tuning of the cell response to the edge orientation. (Responses are shown in relative units.)

We propose to overcome this problem by computing new quantities as follows: $q^a_{j,\varphi}(\xi,\eta)$

$$q^a_{j,\varphi}(\xi,\eta) = r^a_{j,\varphi}(\xi,\eta) \quad \text{if} \quad r^a_{j,\varphi}(\xi,\eta) = max\{r^a_{j,\phi}(\xi,\eta) \;\mid\; \forall\ \phi\} \tag{11}$$

$$q^a_{j,\varphi}(\xi,\eta) = 0 \quad \text{if} \quad r^a_{j,\varphi}(\xi,\eta) < max\{r^a_{j,\phi}(\xi,\eta) \;\mid\; \forall\ \phi\} \tag{12}$$

in a *winner-takes-all competition* between all quantities $r^a_{j,\varphi}(\xi,\eta)$ with the same values of ξ, η and j but with different values of φ. The effect of this orientation competition on the orientation sensitivity is illustrated by Fig.5b; the larger the number of discrete orientations used, the sharper the tuning.

Fig.6 shows the cortical images which correspond to these new quantities. This scheme obviously better discriminates among different orientations. Note that each edge line is enhanced in a different cortical image so that the processing can be considered as decomposition of a geometric object into edge lines. In this way, the computation of such cortical filters delivers more structured information than a traditional edge detector such as a Laplacian operator.

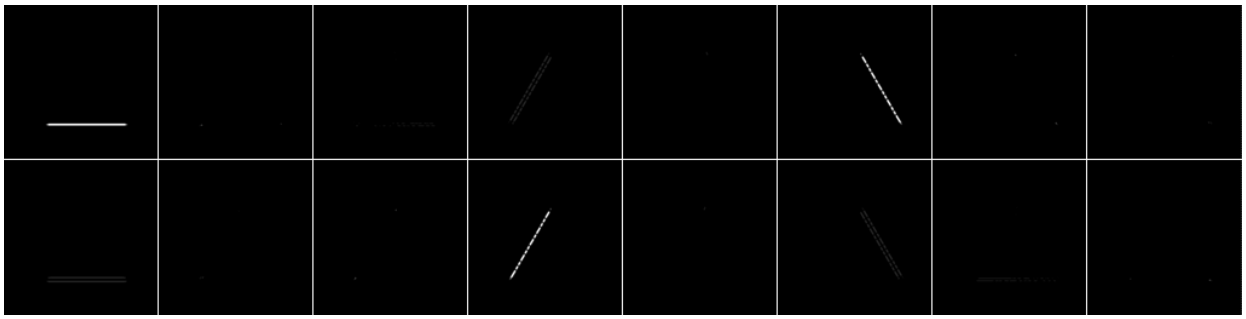

Figure 6: Cortical images computed with the involvement of orientation competition. The images in the first and second row correspond, left to right, to $\varphi_i = 2\pi i/16,\ i = 0\ \ldots\ 7$, and $\varphi_i = 2\pi i/16,\ i = 8\ \ldots\ 15$, respectively.

In [23] and [24], we propose to consider these or similar quantities as the actual activities of simple visual cortical cells. This interpretation was in part biologically motivated, since it is known from neurophysiological research that these cells are organized in columns and hypercolumns where they are are

strongly interconnected [2]. One can find a similar implicit interpretation in [28] where the activities are computed in a dynamical process modelled by a differential equation which includes an orientation competition term. Here, we refrain from giving an interpretation of the above computed quantities. The reason is that, if there were such a competition and the quantities $q^a_{j,\varphi}(\xi,\eta)$ represented the activities of simple visual cortical cells, than the measured receptive field functions would have a form which is quite different from the one found experimentally. Fig.7 illustrates the difference. One could still speculate that the proposed quantities represent the activities of cells in a deeper cortical layer. The verification of such a hypothesis would be a problem for neurophysiological rather than computer science research. For the needs of the present study it is sufficient that the developed cortical filters do deliver highly structured information which, as shown below, can successfully be used for object recognition.

Figure 7: Actual (a) and computed 'would-be-in-the-presence-of-orientation-competition' (b) impulse response of a simple visual cortical cell.

4 Lower dimension space representation

Next we extract from the cortical images $q^a_{j,\varphi}(\xi,\eta)$, which are obtained by applying orientation competition, a set of features to be used for object recognition. We consider the following features:

$$Q^a_{j,\varphi} = \int q^a_{j,\varphi}(\xi,\eta)d\xi d\eta, \qquad j \in \mathbf{Z},\ \varphi \in [0,2\pi). \tag{13}$$

By means of eq. (13), each cortical image is reduced to a single number, the energy in the respective cortical channel. Note that, as far as a single object on a background of constant intensity is concerned, the proposed features do not depend on the particular position of the object in the visual filed, a property which we refer to as *translational invariance*. One has to admit that, while the obtained features give information about the strength and size of edges of particular orientation, this simple method of feature extraction leads to a complete loss of information about the spatial distribution of the edges, an evident drawback which we are well aware of. In spite of this deficiency, the method is still capable of providing features which give the opportunity to discriminate successfully between objects of considerable complexity such as faces.

Fig.8 shows a plot of the quantity $Q^a_{j,\varphi}$ for one fixed value of j (corresponding to $\lambda_j \approx L/40$) as a function of the preferred orientation φ, $\varphi \in [0,2\pi)$. The values of $Q^a_{j,\varphi}$ for $\varphi_i = 2\pi i/16, i = 0\ \ldots\ 16$, are the energies of the respective images in Fig.6. The plotted function exhibits three very clear dominant maxima which can be considered as the three edge lines in the original input image. (The small peaks between the high ones are due to so-called shadow lines which can be suppressed by lateral inhibition [24].) Within the class of convex polygons, this restricts the choice of possible objects to a triangle. For comparison, plots of similar integral features $R^s_{j,\varphi}$ and $R^a_{j,\varphi}$, which are based on the cortical representations $r^s_{j,\varphi}(\xi,\eta)$ and $r^a_{j,\varphi}(\xi,\eta)$ computed by thresholding the output of symmetic and antisymmetric Gabor convolvers, are shown in Fig.9.

The developed method is very robust for translations: if the triangle of Fig.2 is shifted in the input image, virtually the same plot as the one shown in Fig.8 will be obtained. Translations on an unsteady background lead to differences which turn out to be sufficiently small if the shifted object of importance occupies most of the image area. A rotation of the object would lead to a circular shift of the plot, and scaling leads to a similar plot, however, obtained for a different value of the receptive field size parameter j. Finally, if a triangle with edges of unequal lengths is taken, there will be a change in the strengths and relative positions of the maxima which can easily be compensated for by dynamic programming [30].

We used the above developed method for the recognition of simple geometric objects such as convex polygons. Representations such as the one shown in Fig.8 were computed for images of reference convex

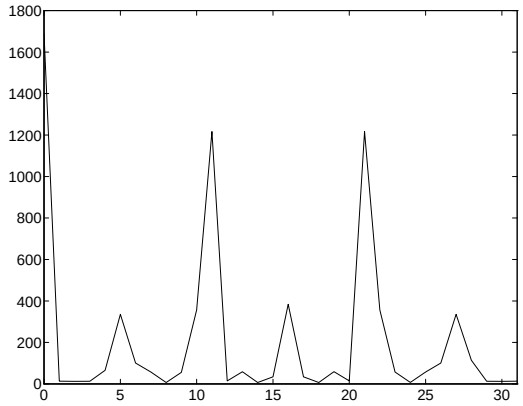

Figure 8: Lower-dimension space representations obtained by applying a global reduction operation to the cortical images computed with the involvement of orientation inhibition (Fig.6). The three high peaks correspond to the three dark-to-light transitions (edges) in the input image.

polygons with a different number of edges. Then a large number of test images of similar polygons were generated whereby position, orientation, size and relation in the size of the edges were generated randomly. For each of the test images, the above proposed lower-dimension representation was computed and compared to the representations of the reference images using multiresolution analysis and dynamic programming as sketched above. The classification was correct in all of a very large number of trials.

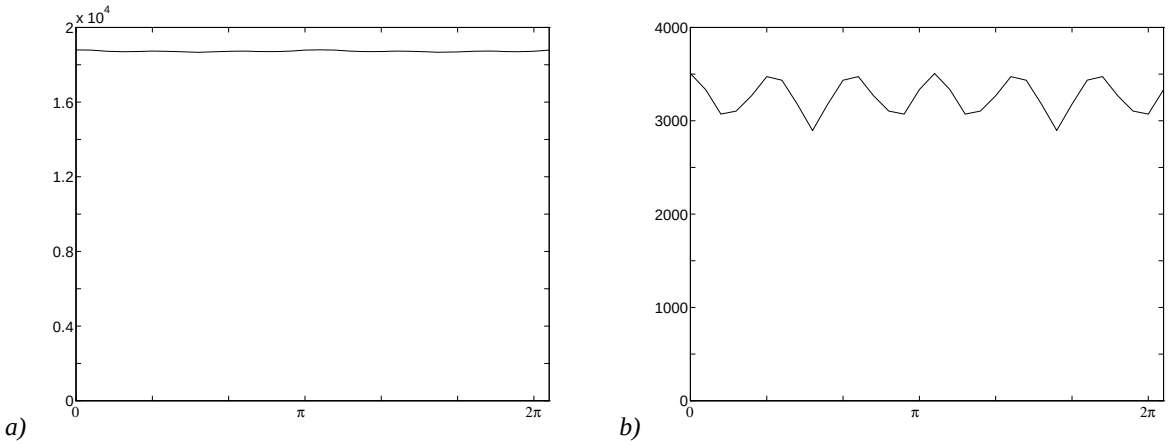

Figure 9: Lower-dimension space representations obtained by applying a global reduction operation to the cortical images shown in Fig.3 (a) and Fig.4 (b). From these plots, it would be difficult to infer the presence of a triangle in the input image.

5 Face recognition experiments

5.1 Face images database

The fact that the developed method performs well in discriminating between simple geometrical objects encouraged us to apply it to a more complex object recognition problem, automatic 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 pictures were taken of each person, with the exact number varying from 5 to 9. All pictures of the same person were taken in one session, though the entire database was created in several sessions. For two of the test persons, two series of images were taken with an interval of more than a month between the sessions, to observe the influence of different haircuts. The individual images of each

person exhibit differences in facial expressions and/or orientations. We deliberately kept these differences relatively small to gain understanding of the model and be able to make incremental changes in it and observe the effects. The underlying idea was that a model that does not compensate for small changes in an image would not function for large changes either. 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%. Pictures have been made of both males and females, adults and children, including persons with glasses, beards and moustaches. When the pictures were taken, the subjects were sitting in front of a blank shield which was used to achieve a constant background. Illumination was strived to be constant from session to session but no special effort was given to achieve exactly the same conditions. Deviations in the illumination were due to changes in the position of the lamps and by sun light coming through the windows. The contrast of all images was not exactly the same due to deviations in the camera settings from session to session. The face pictures are stored as graylevel images with spatial resolution of 500×400 pixels and 8-bit quantization (256 graylevels). Fig.10 and Fig.11 give a notion of the images which comprise the database.

5.2 Computational method

The database images were extended at the boundaries to the size of 512×512 to make the linear size a power of two which is useful for FFT computations. This resolution applies also for the computed cortical images. Such images were computed for eight different orientations in each of eight different scales. Hence, 64 cortical images were computed for each of the 205 database images.

Note that for fixed values of the parameters j and φ and variable values of ξ and η, the discrete function $\tilde{s}^a_{j,\varphi}(\xi,\eta)$ in (8) can be computed as convolution of the input signal $s(x,y)$ with a receptive field function $\Im g_{j,\varphi}(x,y)$. We use this fact for the efficient computation of $\tilde{s}^a_{j,\varphi}(\xi,\eta)$ using a fast Fourier transform (FFT) algorithm. In spite of the computational efficiency of this algorithm, the convolution computation is quite intensive and comprises more than 80% of the used computing time.

After computing a cortical image $q^a_{j,\varphi}(\xi,\eta)$ for a given input image $s(x,y)$ according to (8) and (12), the cortical image is reduced to a single number $Q^a_{j,\varphi}$ according to (13). In this way 64 numbers (features), one number for each of the 64 antisymmetric receptive field functions, are computed for each input image and stored as an 8×8 matix $[Q^a_{j,\varphi_i}]$, $j = -1 \ \ldots \ -8$; $\varphi_i = 2\pi i/8$, $i = 0 \ \ldots \ 7$, which is used to represent the image for database searching. The comparison between two images of the database is done by comparing their respective feature matrices. The normal Euclidean distance was used to determine the dissimilarity between two feature matrices. In principle, compensation for rotation and size difference of one of two images to be compared with respect to the other can be done by circular shifts along the columns and rows of the respective feature matrix. We did not make use of that possibility, since the number of Gabor functions used allows for an angular resolution of 45° and scale resolution of 1.4 (40% difference in size) wheres the orientation and scale differences of the database images do not exceed 30° and 1.2 (20% difference in size), respectively.

5.3 Results

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. Only the best match was used to determine whether the search was successful (delivering an image of the same person) or not (delivering an image of another person). For 202 out of 205 images the search was successful as illustrated by Fig.10. The model failed in only three cases which are shown in Fig.11. This gives a recognition rate of approximately 98.5%.

5.4 Computational requirements

The described method is quite computationally intensive: 40 minutes time are needed to compute the feature matrix of one 512×512 image on an 18 Mflops/s state-of-the-art workstation. As already mentioned, currently we do not apply rotation and scale compensation due to the relatively low number of receptive field functions used. We estimate that at least 32 orientations and 32 scales have to be used, in order to be able to efficiently compensate for real rotation and scale variations. This amounts to 1024 receptive field functions to be applied as convolver kernels to each of the images in the database which is

Figure 10: Examples of successful matches: each image in the first row is a test (input) image for which best match search is done in the rest of the image database; the images in the second row are the respective best matches returned by the system.

currently being extended to over 1000 images. The time to solve this problem on the mentioned state-of-the-art workstation is estimated to exceed one year. High performance computing is evidently relevant in this case and we ported the developed programs to the Connection Machine CM-5. The CM-5 which is installed at the University of Groningen currently has 16 processing nodes (four vector units per node), 512 Mbytes of main memory and a 9.4 Gbytes parallel disk store. Its peak performance is 2 Gflops/s and we achieve more than 1 Gflops/s for the computationally intensive part (FFT) of our problem. For this performance to be achieved, the database images are processed in batches of 64, one image per vector unit, and the computations on one image are carried out sequentially within one vector unit. The computational power and the parallel I/O system, which gives a throughput of nearly 18 Mbytes/s, allows us to achieve an acceleration of a factor of more than 200 compared to the above mentioned workstation. The computing time, which is needed for large experiments and model verifications, can thus be reduced to a few days.

6 Concluding remarks

In this paper we have demonstrated how computer simulations can be used to explore the mechanisms of natural vision. The amount of computations necessary compare with those characteristic of numerical simulations in fluid dynamics, molecular dynamics, semiconductor device design, quantum chromodynamics, etc. The results are very encouraging: the recognition rate of the developed method amounts to 98.5% which, as far as the error rate of 1.5% is concerned, is an improvement of a factor of four with respect to the performance of a previously used model without orientation competition which gave a recognition rate of 94% (error rate 6%) [22]. ¿From the comparison of the information delivered by the two models for simple geometrical figures, we infer that the improvement is due to the use of asymmetric receptive field functions and the introduction of orientation competition. An important ensuing question is whether orientation competition actually exists in the primary visual cortex. We raise this question, but consider that the answer has to be given by neurophysiological research.

In this study, we started with results from neurophysiological research and our simulations resulted in a question which can be ansered by further neurophysiological research. The moral might be that computer science and neuroscience might profit from each other.

Figure 11: Examples of failure of the model: the best matches (second row) correspond to different persons.

7 Acknowledgements

Most of the computations were carried out on the Connection Machine CM-5 of the University of Groningen, the investments in which were partly supported by the Netherlands Computer Science Research Foundation (SION) and the Netherlands Organization for Scientific Research (NWO).

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] D.G. Stork and H.R. Wilson: "Do Gabor functions provide appropriate descriptions of visual cortical receptive fields", *J. Opt. Soc. Am. A*, Vol. 7 (1990) No.8, pp.1362-1373.

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

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

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

[9] 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.

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

[11] 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

[12] M. Lades, J.C. Vorbrüggen, J. Buhmann, J. Lange, C. von der Malsburg, R.P. Würtz, W. Konen: "Distortion invariant object recognition in the dynamic link architecture", to appear in *IEEE Trans. on Computers*

[13] 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.

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

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

[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] H. Boattour, F. Fogelman Soulié and E. Viennet: "Solving the human face recognition task using neural nets", *Proceedings of the ICANN-92, Brighton, September 1992*, pp.1595-1598.

[20] 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.

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

[22] N. Petkov, P. Kruizinga and T. Lourens: "Biologically Motivated Approach to Face Recognition", *Proc. International Workshop on Artificial Neural Networks*, June 9-11, 1993, Sitges (Barcelona), Spain (Berlin: Springer Verlag, 1993) pp.68-77

[23] N. Petkov, T. Lourens and P. Kruizinga: "Lateral inhibition in cortical filters", Proc. of *Int. Conf. on Digital Signal Processing and Int. Conf. on Computer Applications to Engineering Systems*, July 14-16, 1993, Nicosia, Cyprus, pp.122-129.

[24] N. Petkov and T. Lourens: "Human Visual System Simulations - An Application to Face Recognition", in H. Dedieu (ed.) Proc. 1993 European Conf. on *Circuit Theory and Design*, Aug. 30 - Sept. 3, 1993, Davos, Switzerland (Amsterdam: Elsevier Sci. Publ., 1993, in print)

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

[26] S. Marcelja: "Mathematical description of the response of simple cortical cells", *J. Opt. Soc. Am.*, Vol. 70 (1980) pp. 1297-1300

[27] M. Connoly and D. van Essen: "The representation of the visual field in parvocellular and magnocellular layers in the lateral geniculate nucleus in the macaque monkey", *J. Comput. Neurol.*, Vol.226 (1984) pp. 544-564.

[28] B.S. Manjunath and R. Chellappa: "A unified approach to boundary perception: edges, textures, and illusory contours", *IEEE Transactions on neural networks*, Vol.4 (1993) No.3, pp.96-108.

[29] J. Bigün: "Gabor phase in boundary tracking and region segregation", *Proc. of the Int. Conf. on Digital Signal Processing and II Int. Conf. of Computer Applications to Engineering Systems*, Nicosia, Cyprus, July 14-16, 1993,. pp.229-234.

[30] N. Petkov: *Systolic Parallel Processing*, Ch.9, (Amsterdam: North Holland, 1993)