# Lateral Inhibition in Cortical Filters

N. Petkov, T. Lourens, P. Kruizinga

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

**Abstract**

This work presents explorations in the microstructure of natural vision systems based on large scale computer simulations. Similarly to previous work in this area, we compute the functional inner products of a two-dimensional input signal (image) with a set of two-dimensional Gabor functions which have been shown to fit the receptive fields of simple cells in the primary visual cortex of mammals. These inner products are then considered as net inputs to the cortical cells and used to compute the cell activations as non-linear functions. A previously used model is extended with a pixel-wise winner-takes-all competition between different Gabor filters which is introduced in order to model lateral inhibition between cortical cells. The effect of lateral inhibition is qualitatively estimated by visualization of computed cortical images and quantitatively evaluated by applying the model to a face recognition problem. Recognition rate of 97% was achieved on a database of 205 face images of 30 persons vs. 94% achieved with a previously used model.

## 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. 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 fulfill for a number of years in the other branches of science mentioned above. The chances offered by large scale computer simulations in this area may even turn 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.

Neurophysiological research has delivered a number of interesting results which can serve as a starting point for computational research. It is, for instance, well known that a large amount of neurons in the primary visual cortex of mammals react strongly to short oriented lines [1,2]. A more precise study has shown that the receptive fields of such cells can be fitted well by Gabor functions, differences or derivatives of Gaussians or other similar functions [3-5]. Basing on these results, one can mimic the function 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 research carried out until now in this area has given rise to a number of open questions. Among these we consider as most important the question of how the information delivered by the cortical filters can be effectively 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 how the cortical cells interact with each other to facilitate structuring and further analysis of information. We propose to consider the quantities computed by Gabor convolvers as actual cell activations only for impulse (spot) images at low excitation levels for which the response is substantially linear. For more complex images and higher excitation levels, the computed quantities should be considered as the net inputs to the cortical cells. The actual cell activities should be computed as non-linear functions of the net inputs, a process that has been proven to play a very important role in neural networks. Furthermore, we propose that the activities thus computed should become the subject of lateral inhibition or excitation, a mechanism which according to the results of neurobiological research has almost universal validity in natural neural networks.

A direct confirmation of the correctness of the above sketched approach can be achieved only by means of neurobiological research that would, for instance, show that the inhibitive cortical interconnections we propose in the following actually exist. It is, however, also possible to verify the plausibility of the approach by making it a part of a full object recognition system and then comparing how it scores in comparison with previously used more simple models. For this purpose, we have incorporated our model in an automatic face recognition system.

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 [7-8]. Another basic technique which has reached a considerable level of sophistication is template matching [9-11]. Further approaches to face recognition use graph matching [11-12], Karhunen-Loewe expansion [13], algebraic moments [14], isodensity lines[15], etc. Connectionists approaches to the problem are described in [11, 16-19]. We refer the reader to [20] for a collection of recent works in this area.

The present work is an extension of our previous work reported in [21-24]. By introducing lateral inhibition, we succeed to improve the recognition rate of a biologically motivated face recognition system from 94% to 97% 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 it is shown how non-linearities and lateral inhibition are made a part of the model. 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. Section 6 is a summary of the work.

## 2 Gabor functions and natural vision

The basic two-dimensional Gabor function we use in our computer simulations 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 $(\xi, \eta)$, delations parameterized by an integer number $j$ and rotations parameterized by an angle $\varphi$, one gets the following family of two-dimensional Gabor functions ($\xi$ and $\eta$ have the same domain as $x$ and $y$, respectively):

$$g_{j,\varphi}(x-\xi, y-\eta) = \frac{1}{\pi}\alpha^{2j} e^{-\alpha^{2j}(x'^2+y'^2)+i\pi\alpha^j x'} \qquad (j \in \mathbb{Z},\ \varphi \in [0,\pi)) \tag{2}$$

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

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

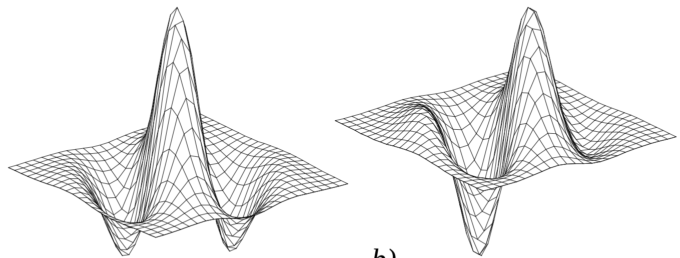
Fig.1 shows the real and imaginary parts of one such function. The oscillations of $g_{j,\varphi}(x-\xi, y-\eta)$ are due to the harmonic wave factor $e^{i\pi\alpha^j x'}$ with a wavelength

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

and a wavevector (spatial frequency) of orientation $\varphi$ and magnitude

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

The Gaussian factor $e^{-\alpha^{2j}(x'^2+y'^2)}$ 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 $\alpha^j$ ($j \in \mathbb{Z}$) corresponds to equidistant sampling of a logarithmic wavelength/frequency scale that corresponds to the logarithmic dispersion of spatial frequencies found by neurophysiological research [3, 4].



*a)* *b)*

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

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

$$\tilde{s}_{j,\varphi}(\xi, \eta) = \int s(x, y) g^*_{j,\varphi}(x-\xi, y-\eta) dx dy \tag{5}$$

may be considered as the amount of a harmonic wave with wavelength $\lambda_j$ and wavevector orientation $\varphi$ in a surrounding of linear size $\lambda_j$ centered on a point with coordinates $(\xi, \eta)$. In this way, equation (5) represents local spectral analysis which is embedded in global spatial coordinates $(\xi, \eta)$. The coefficient $\frac{1}{\pi}\alpha^{2j}$ in front of the exponent in (2) is a normalization factor which is chosen in such a way that for an input signal $s(x, y) = e^{i\pi\alpha^j x'}$ with magnitude one the quantity computed in (5) has also magnitude one, $|\tilde{s}_{j,\varphi}(\xi, \eta)| = 1$.

In the following we use (5) as the basis of our approach of mimicing the function of the primary visual cortex, assuming that the values $\tilde{s}_{j,\varphi}(\xi, \eta)$ delivered for the various values of the parameters $j, \varphi, \xi$ and $\eta$ correspond to the net inputs to individual cortical cells when the visual system is presented an image $s(x, y)$. Note that the computed quantities $\tilde{s}_{j,\varphi}(\xi, \eta)$ comprise more data than the original image $s(x, y)$. This is 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 this data expansion is used to effectively recognize an object. What one is certain about is that this data expansion is actually carried out in the brain as confirmed by the fact that the visual information is transferred from the retina to the primary visual cortex via $10^6$ fibers of the optic nerve but in the primary visual cortex it is encoded by $10^8 - 10^9$ simple cells (100-1000 times expansion at cortical level [25]). 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.*

# 3 Non-linearity and lateral inhibition

Note that the quantities $\tilde{s}_{j,\varphi}(\xi, \eta)$ computed in (5) depend lineraly on the input image $s(x, y)$ and are complex. Because of the latter fact, there has been critique against Gabor functions, since the relevance of using separately the real and complex parts has been confirmed by neurophysiological data until now only in twelve cases of pairs of neighbouring cortical cells [6]. For the present we circumvent this controversy and introduce non-linearity by considering the quantities (5) as net inputs to the cortical cells whereby the activity $a_{j,\varphi}(\xi, \eta)$ of a cell with receptive field centered on a point with coordinates

$\xi$ and $\eta$ and characterized by main wavelength $\lambda_j$ and wavevector orientation $\varphi$ is determined as the magnitude of the complex quantity $\tilde{s}_{j,\varphi}(\xi,\eta)$ computed in (5):

$$a_{j,\varphi}(\xi,\eta) = \sqrt{(\Re(\tilde{s}_{j,\varphi}(\xi,\eta)))^2 + (\Im(\tilde{s}_{j,\varphi}(\xi,\eta)))^2} \tag{6}$$

For fixed $j$ and $\varphi$ and variable $\xi$ and $\eta$, $a_{j,\varphi}(\xi,\eta)$ is a two-dimensional non-negative function to be referred to as a *cortical image*. Fig.2 shows a few computed cortical images which were obtained for fixed $j$ ($j=-3$, $\lambda_{-3} \approx 6$ pixels for $\alpha = \sqrt{2}$) and different wavevector orientations $\varphi$ ($\varphi_i = i\pi/8, i = 0\ldots7$) when a white triangle on a black background was taken as an input image (top-left image in Fig.2). Note that the differences between the computed cortical images are not large. In particular, the same edge is enhanced in more than one cortical image (see e.g. the horizontal edge in the first, second and last cortical images). This seems to be in contrast with psychological and neurophysiological experiments that confirm high orientational sensitivity of the visual system of mammals.

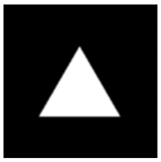


Figure 2: Input image (top-left corner) and computed non-interacting cortical images.

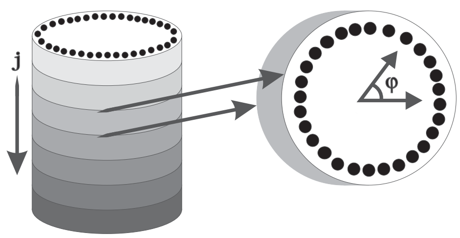



Figure 3: Schematical representation of a cortical cells whose receptive fields are centered on the same point of the visual field.

The not very good results achieved in this way made us think about introducing a mechanism which would improve the orientational sensitivity of the method. We propose to solve the problem by a winner-takes-all competition between all quantities $a_{j,\varphi}(\xi,\eta)$ with the same values of $\xi$, $\eta$ and $j$ but with different values of $\varphi$. This corresponds to inhibitive interconnections between the respective cortical cells. This decision is in part biologically motivated, since it is known from neurophysiological research that in the primary visual cortex of mammals the simple cells are organized in columns and are strongly interconnected [2]. Fig.3 shows schematically one such column of cells with receptive fields centered on point ($\xi$, $\eta$) of the visual field. The cells with the same size of the receptive field (same $j$) are represented as lying in the same cylindrical cross section of the column. The cells in one such cross section correspond to different wavevector orientations (different $\varphi$) and we introduce inhibitive interconnections among these cells. Of course, Fig.3 is only an illustration and the actual arrangement of cells corresponding to different values $j$ and $\varphi$ in one column does not have to follow this scheme. The winner-takes-all competition can be modelled in the following way

$$\tilde{a}_{j,\varphi}(\xi,\eta) = a_{j,\varphi}(\xi,\eta) \quad \text{if} \quad a_{j,\varphi}(\xi,\eta) = max\{a_{j,\phi}(\xi,\eta) \;\; | \;\; \forall\ \phi\} \tag{7}$$

$$\tilde{a}_{j,\varphi}(\xi,\eta) = 0 \qquad\qquad\quad \text{if} \quad a_{j,\varphi}(\xi,\eta) < max\{a_{j,\phi}(\xi,\eta) \;\; | \;\; \forall\ \phi\} \tag{8}$$

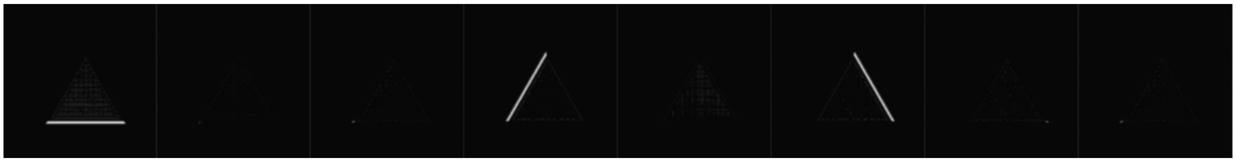


Figure 4: Cortical images computed with the involvement of lateral inhibition.

whereby the quantities $\tilde{a}_{j,\varphi}(\xi,\eta)$ should be considered as the new cortical cell activities after the competition is completed. Fig.4 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 interpreted as decomposition of a geometric object into edge lines. In this way, the computation of cortical filters delivers more structured information than a traditional edge detector such as a Laplacian operator.

# 4 Application to object recognition

In order to quantitatively evaluate the plausibility of the above presented model, we make it a part of an object recognition system. Since we have no hints from neorophysiological research about how cortical images are used in the process of object recognition, we have to make hypothesis about the further representation and processing of visual information. For this purpose we introduce the following quantities:

$$A_{j,\varphi} = \int \tilde{a}_{j,\varphi}(\xi,\eta) d\xi d\eta, \qquad j \in \mathbb{Z},\ \varphi \in [0,\pi). \tag{9}$$

Each of them represents the cumulative activity of all cells with the same wavevector orientation $\varphi$ and main spatial frequency $\pi\alpha^j$, independently of their positions $(\xi,\eta)$ in the visual field. Our 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 cell activities computed at higher stages. Each of the quantities (9) 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 receptive field form, size and orientation. We have to admit that we are not aware of neurobiological evidence that would confirm this hypothesis. Computing the quantities $A_{j,\varphi}$ according to (9) might however make sense for one reason: they are not sensitive to the particular position of an object in the visual field, a property which we refer to as *translational invariance*. Let us now represent two images $s(x,y)$ and $s'(x,y)$ by the respective sets of quantities $A_{j,\varphi}$ and $A'_{j,\varphi} (j \in \mathbb{Z}, \varphi \in [0,\pi))$ according to (9) and define the *dissimilarity* of the two images as follows

$$D_{s,s'} = \sum_{j,\varphi} \mid A_{j,\varphi} - A'_{j,\varphi} \mid . \tag{10}$$

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. The relations (5-10) are the basis of our approach to automatic face recognition. The quantities (9), 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 (10) is minimal.

# 5 Results of face recognition experiments

We applied the above developed approach to a database of 205 different face images of 30 persons. Details on the database can be found in [21,23]. The images are of size 500 × 400 pixels and this discretization applies also for the computed cortical images. Such images are computed for the following discrete values of the Gabor convolver parameters $j = -1, -2, \ldots, -8$; $\varphi_i = i\pi/8$, $i = 0, 1, \ldots, 7$. The basic scaling factor $\alpha$ is taken to be $\alpha = \sqrt{2}$. This choice of $\alpha$ and the range of the parameter $j$ allow for covering a wavelength domain that ranges from 2.8 to 32 pixels with logarithmic dispersion of the main wavelengths of the respective Gabor functions (see (3)).

Note that for fixed $j$ and $\varphi$ and variable $\xi$ and $\eta$, $\tilde{s}_{j,\varphi}(\xi, \eta)$ in (5) can be computed as convolution of the signal $s(x, y)$ with a Gabor function $g_{j,\varphi}(x, y)$. We use this fact for the efficient computation of these quantities 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 99% of the used computing time. After computing a cortical image $a_{j,\varphi}(\xi, \eta)$ for a given input image $s(x, y)$, it is reduced to a single number $A_{j,\varphi}$ according to (9). 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. 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 198 out of 205 images the search was successful as illustrated by Fig.5. The model failed in 7 cases, four of which are given in Fig.6. This gives a recognition rate of approximately 97% which as far as the error rate is concerned is an improvement by a factor of two compared to our previously used approach which did not include lateral inhibition [21-23].

Figure 5: 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.

## 6 Summary

In this paper we have demonstrated how computer simulations can be used to explore the mechanisms of natural vision. The biological relevance of our model is based on 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, our approach is an attempt to guess what might be happening in the

Figure 6: Examples of failure of the model: the best matches (second row) correspond to different persons.

further form analysis structures of the visual cortex. For this part, we have no neurophysiological and neurobiological data to build on and, therefore, we rely only on general principles such as non-linearity, lateral inhibition and translational invariance. In spite of a number of shortcomings of the model, we achieve a recognition rate of 97% on a relatively large image database. Further work on the model which is in progress will be reported elsewhere [26].

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

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

[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] 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*

[12] 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

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

[21] N. Petkov, T. Lourens and P.Kruizinga: "Computationally intensive approach to face recognition", Comp. Sc. Notes, CS9207, Dept. of Computer Science, University of Groningen, December 1992.

[22] N. Petkov, T. Lourens, P. Kruizinga: "Computationally intensive approach to face recognition", *12th Benelux Meeting on Systems and Control*, March 3-5, 1993, Houffalize, Belgium, *Abstracts and Lecture Notes* (Mons: Faculte Polytechnique de Mons, 1993) pp.27.

[23] N. Petkov, P. Kruizinga, T. Lourens: "Biologically motivated approach to face recognition", *Proc. International Workshop on Artificial Neural Networks*, June 9-11, 1993, Sitges (Barcelona), Spain (Amsterdam: Elsevier Sci. Publ., in print).

[24] N. Petkov, T. Lourens, P. Kruizinga: "Large scale natural vision simulations", *High Performance Computing and Networking 93 Conference*, 17-19 May, Amsterdam.

[25] 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.

[26] N. Petkov, T. Lourens, P. Kruizinga: "Natural vision simulations on the Connection Machine CM-5", in preparation.