# Natural vision simulations - An application to face recognition

N. Petkov and T. Lourens[a]*

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

A method is proposed in which the convolutions of an input image with a set of visual receptive field functions become the subject of thresholding, orientation competition and lateral inhibition. The developed cortical filters deliver highly structured information which can be used for efficient feature extraction. The method is applied to face recognition achieving a recognition rate of 99% on a large database of face images.

## 1. INTRODUCTION

The receptive fields of simple cells in the primary visual cortex of mammals can be fitted well by Gabor functions [1], differences of offset Gaussians or other similar functions [2]. Consequently, such functions have been used as convolvers in so-called 'cortical filters' which have been the subject of intensive research in the recent years. The research carried out until now has given 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 to facilitate structuring of information. Here we propose a cortical filter model in which the output of receptive field convolvers becomes the subject of thresholding, orientation competition and lateral inhibition. The resulting cortical filters deliver highly structured information which can be used for the extraction of efficient features.

We apply the developed method to the problem of automatic face recognition. This problem, considered to be a challenge since the early days of computer vision, recently experiences a revival. Some of the previously used approaches to automatic face recognition were comparison of geometric features (such as size and relative positions of eyes, mouth, nose and chin) [3], template matching [4], graph matching [5], Karhunen-Loewe expansion [6], algebraic moments [7], isodensity lines [8], neural networks [9], etc. We refer the reader to [10] for further references. The present work is an extension of our previous work reported elsewhere [10, 11]. By applying the developed method, we achieve a recognition rate of 99% on a large database of face images.

## 2. CORTICAL FILTER MODEL

We use the following family of visual receptive field functions[2]:

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

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

*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).

[2]Only the imaginary part of a two-dimensional Gabor function is used. For argumentation see [12].

The parameters $(\xi, \eta)$, $\varphi$ and $j$ specify the centre of a receptive field in the visual field, its orientation and size ($O(\alpha^{-j})$), respectively. Fig.1 shows one such function.

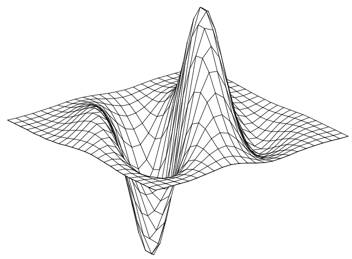


Figure 1. Visual receptive field function.

We propose to interpret the functional inner product of a two-dimensional signal (image) $s(x, y)$ with a receptive field function $g_{j,\varphi}(x - \xi, y - \eta)$

$$\tilde{s}_{j,\varphi}(\xi, \eta) = \int s(x, y) g_{j,\varphi}(x - \xi, y - \eta) dx dy \qquad (j \in \mathbf{Z},\ \varphi \in [0, 2\pi)) \tag{2}$$

as a net input to the cortical cell characterized by the function $g_{j,\varphi}(x - \xi, y - \eta)$, arizing when the visual system is presented an image $s(x, y)$. This interpretation differs from previous interpretations of this quantity as the activation of the considered cell. Note that for fixed values of $j$ and $\varphi$ and variable $(\xi, \eta)$, (2) presents the convolution (with a minus sign) of an input image $s(x, y)$ and a receptive field function $g_{j,\varphi}(x, y)$. This observation is important for the efficient computation of (2).

The convolution output (2) is next submitted to *thresholding* (here with a threshold zero)[3]:

$$a_{j,\varphi}(\xi, \eta) = \tilde{s}_{j,\varphi}(\xi, \eta) \chi(\tilde{s}_{j,\varphi}(\xi, \eta)) \ , \tag{3}$$

reflecting the fact that negative, i.e. inhibitive, input to cortical cells cannot cause them to fire. The quantities $a_{j,\varphi}(\xi, \eta)$ are, however, not considered as the actual activations either. They are rather intermediate quantities in the computation of the ultimate cell activities.

In order to increase the orientation sensitivity of the cortical filters, new quantities $b_{j,\varphi}(\xi, \eta)$ are computed next as follows:

$$b_{j,\varphi}(\xi, \eta) = a_{j,\varphi}(\xi, \eta) \quad \text{if} \quad a_{j,\varphi}(\xi, \eta) = max\{a_{j,\phi}(\xi, \eta) \ \mid \ \forall\ \phi \in [0, 2\pi)\} \tag{4}$$

$$b_{j,\varphi}(\xi, \eta) = 0 \qquad\qquad\quad \text{if} \quad a_{j,\varphi}(\xi, \eta) < max\{a_{j,\phi}(\xi, \eta) \ \mid \ \forall\ \phi \in [0, 2\pi)\}. \tag{5}$$

A quantity $b_{j,\varphi}(\xi, \eta)$ is computed from the set $\{a_{j,\phi}(\xi, \eta) \ \mid \ \forall\ \phi \in [0, 2\pi)\}$ of quantities with the same arguments $(\xi, \eta)$ and parameter $j$ in a winner-takes-all *orientation competition* among all possible orientations $\phi \in [0, 2\pi)$. Making this orientation competition a part of the model is justified by the high orientational sensitivity of natural vision systems as proven by psychophysical and neurophysiological experiments.

The ultimate quantities $c_{j,\varphi}(\xi, \eta)$, which are proposed as a representation of the actual cortical cell activities, are computed from the quantities $b_{j,\varphi}(\xi, \eta)$ by a *lateral inhibition* mechanism, in

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

which a large quantity $b_{j,\varphi}(\xi,\eta)$ suppresses all smaller similar quantities which are labeled by the same receptive field size parameter $j$, but have an opposite value $-\varphi$ of the orientation parameter and correspond to neighbouring positions within a distance $\lambda_j = 2\alpha^{-j}$ (wavelength of the harmonic factor in the receptive field function) along a line with orientation $\varphi$ which passes through the point $(\xi,\eta)$:

$$c_{j,\varphi}(\xi,\eta) = b_{j,\varphi}(\xi,\eta) \quad \text{if} \quad b_{j,\varphi}(\xi,\eta) > max\{b_{j,-\varphi}(\xi+\nu\lambda_j\cos\varphi, \eta+\nu\lambda_j\sin\varphi) \mid \forall\, \nu \in (-1,1)\} \tag{6}$$

$$c_{j,\varphi}(\xi,\eta) = 0 \quad \text{if} \quad b_{j,\varphi}(\xi,\eta) \leq max\{b_{j,-\varphi}(\xi+\nu\lambda_j\cos\varphi, \eta+\nu\lambda_j\sin\varphi) \mid \forall\, \nu \in (-1,1)\} \tag{7}$$

Without this lateral inhibition mechanism, a dark-to-light transition edge with gradient orientation $\varphi$, which is enhanced in the intermediate representation $b_{j,\varphi}(\xi,\eta)$ (fixed value of $\varphi$), is also enhanced in $b_{j,-\varphi}(\xi,\eta)$ which corresponds to the opposite direction. The enhancement takes the form of so called shadow lines which enclose the actual edge line [12]. The proposed lateral inhibition mechanism aleviates this redundancy by suppressing the shadow lines and contributes to the structuring of information. For fixed values $j$ and $\varphi$ and variable values of $\xi$ and $\eta$, the quantities $c_{j,\varphi}(\xi,\eta)$ present a non-negative two-dimensional function to be referred to as *a cortical image*. Fig.2 shows a synthetic input image $s(x,y)$. For this input image, Fig.3a shows the cortical images computed for one fixed value of $j$ ($j=-3$) and eight different orientations $\varphi_i = 2\pi i/8, i = 0\ldots7$. For comparison, Fig.3b shows the respective images based on the absolute values of the convolution quantities computed by (2). Note that each of the edge lines in the input image is enhanced in a distinct cortical image.

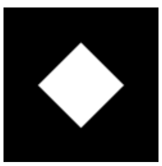


Figure 2. A simple input image.

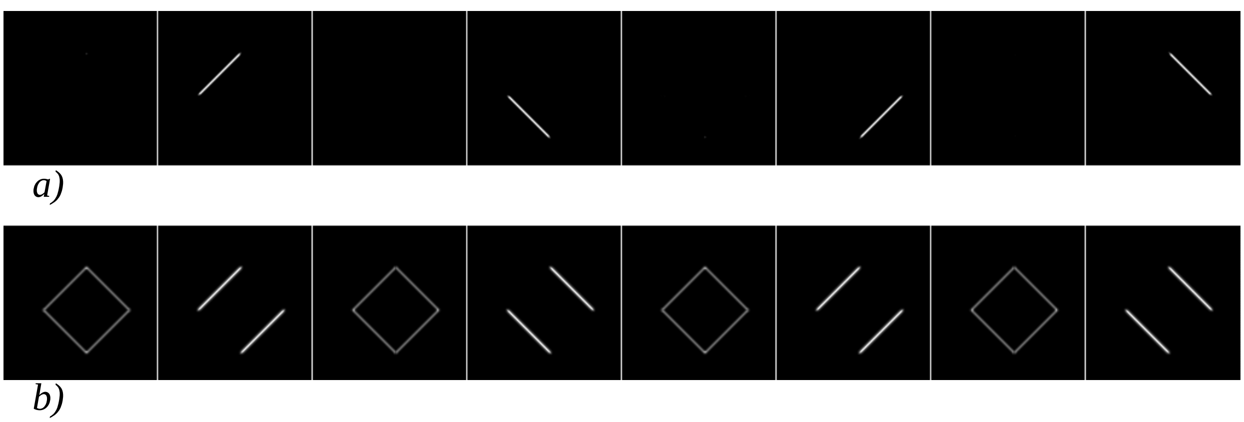



Figure 3. Images obtained by applying convolutions ($a$) and the proposed cortical filters ($b$).

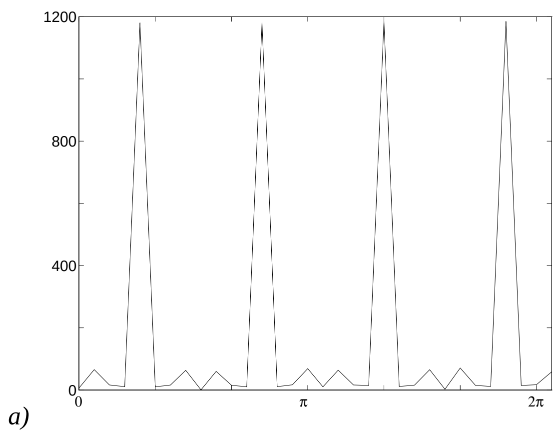
## 3. FEATURE EXTRACTION

Next we extract from the cortical images thus obtained a set of features to be used for object recognition. We consider the following features:

$$C_{j,\varphi} = \int c_{j,\varphi}(\xi,\eta)d\xi d\eta, \quad j \in \mathbf{Z},\ \varphi \in [0, 2\pi). \tag{8}$$

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, a property which we refer to as *translational invariance*. Fig.4a shows a plot of the features $C_{j,\varphi}$ for one fixed value of $j$ ($j = -3$) and different values of $\varphi$, $\varphi \in [0, 2\pi)$. The values of $C_{j,\varphi}$ for $\varphi_i = 2\pi i/8, i = 0 \ldots 7$, are the energies of the respective images in Fig.3a. The plot exhibits four very clear peaks which can be interpreted as four dominant edge lines in the original input image. Within the class of polygons, this restricts the choice of possible objects to a polygon with four edges. For comparison, a plot of similar integral features based on the absolute values of the convolution outputs (2) is shown in Fig.4b.

The developed method is very robust for translations: if the square in the input image is shifted, it would produce virtually the same plot as the one shown in Fig.4a. 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 square would lead to a circular shift of the plot, and scaling leads to the same plot, however obtained for a different value of the receptive field size parameter $j$. Finally, if a polygon with four 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 [13].



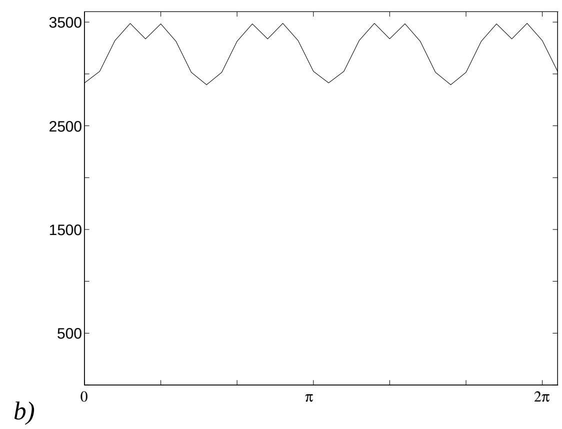




Figure 4. Image representation in a lower-dimension feature space.

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.4a were computed for 14 images of reference convex polygons with three through 16 edges. Then a large number of test images of similar polygons were generated whereby positition, 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 the 1000 trials.

## 4. APPLICATION TO FACE RECOGNITION

An interesting question is whether the method can be applied to more complex objects. For this purpose, we used the method for automatic face recognition. A database of 300 different face images of 40 persons was constructed. Technical details of the database can be found in [10]. For each of the face images, a lower-dimension representation was computed according to eqs. (6-7) and based on this representation a nearest-neighbour was searched in the rest of the database. The search was considered to be successful if the nearest neighbour turned out to be an image of the same person, as illustrated by Fig.5, and not successful if it was an image of a different person, as illustrated by Fig.6.

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.

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

A recognition rate of 99% has been achieved. These results are better than our earlier results on face recognition based on features extracted directly from Gabor representations. The quantitative results on the recognition rate seem to score pretty well in comparison to recent results achieved by other researchers but we should note that a fair comparison is yet hardly possible, since different methods have been applied to different face image databases.

We are rather confident that interaction of cortical filters, as exemplified above by orientation competition and lateral inhibition, is needed to facilitate the process of image analysis and that this might be one of the actual mechanisms used by the brain in the early stages of the visual system. In spite of the excellent results achieved in our experiments on recognition of simple geometric objects and human faces, we have to note that a lot of work has still to be done. In particular, better ways for the extraction of a lower-dimension (preferably syntactic) representation, which preserves the local information delivered by the cortical filters, has to be found. Further work is in progress will be reported elsewhere.

## REFERENCES


1. 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.
2. 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.
3. A.J. Goldstein, L.D. Harmon, and A.B. Lesk: "Identification of human faces", In *Proc. IEEE*, Vol. 59 (1971) pp. 748.
4. A.L. Yuille: "Deformable templates for face recognition", *Journal of Cognitive Neuroscience*, Vol.3 (1991) No.1, pp. 59-70.
5. 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
6. 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.
7. Zi-Quan Hong: "Algebraic feature extraction of image for recognition" *Pattern Recognition* Vol. 24 (1991) No.3, pp. 211-219.
8. O. Nakamura, S. Mathur, and T. Minami: "Identification of human faces based on isodensity maps", *Pattern Recognition*, Vol. (1991) No.3, pp.263-272.
9. 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*
10. N. Petkov, P. Kruizinga, T. Lourens: "Biologically motivated approach to face recognition", *Proc. International Workshop on Artificial Neural Networks*, June 9-11, 1993, Sitges, Spain (Berlin: Springer Verlag, in print).
11. N. Petkov, T. Lourens, P. Kruizinga: "Orientation competition in cortical filters", Report CS9302, Dept. of Computer Science, University of Groningen, April 1993.
12. N. Petkov and T. Lourens: "Interacting cortical filters for object recognition", Report CS9303, Department of Computer Science, University of Groningen, May 1993.
13. N. Petkov: *Systolic Parallel Processing* (Amsterdam: North Holland, 1993)