

Interacting Cortical Filters for Object Recognition

N. Petkov and T. Lourens

Department of Mathematics and Computer Science, Rijksuniversiteit Groningen

P.O. Box 800, 9700 AV Groningen, The Netherlands

E-mail: {petkov, tino}@cs.rug.nl

Abstract

It is shown how cortical filters can be used for image analysis and object recognition. Similarly to previous work in this area, we compute functional inner products of a two-dimensional input signal (image) with a set of two-dimensional Gabor functions which fit the receptive fields of simple cells in the primary visual cortex of mammals. We propose a method in which these inner products become the subject of thresholding, orientation competition and lateral inhibition. Each of the resulting cortical images contains only edge lines of a particular orientation and a particular light-to-dark transition direction. In this way, the information which is present in the original image is split in different channels and we show how this splitting can be used for object recognition. The method discriminates between simple geometrical figures, e.g. polygons with different numbers of edges, with reliability of 100% and a recognition rate of 99% has been achieved when the method was applied to a large database of face images.

1 Introduction

The insights in the microstructure of the brain provided by neurophysiological and neurobiological research may open new opportunities for automatic object recognition. Neurophysiological research has delivered a number of interesting results which can inspire new image analysis models. It is, for instance, well known that a large amount of neurons, the so called simple cells, in the primary visual cortex of mammals react strongly to short oriented lines [1]. A more precise study has shown that the receptive field functions¹ of such neurons can be fitted well by Gabor functions [2, 3], differences of offset Gaussians or other similar functions [4]. Using these results, researchers mimic the function of the primary visual cortex by computing the activation of each individual simple cell for a given input 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.

Previous research in this area has focussed on the precise type of functions which best model simple cell receptive fields (see discussion in [4]). 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 information delivered by 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 in such a way that it can be used for image analysis and object recognition.

For this purpose, we propose the following scheme: The pixel values computed by Gabor convolvers are not considered as the actual activations of cortical cells but rather as net inputs to the cortical cells and the actual cell activities

¹The receptive field of a neuron is the mathematical function which describes the response of that neuron to a small spot of light as a function of position.

are computed by *thresholding* of the net inputs. The activities thus computed become the subject of mutual inhibition, a mechanism which according to the results of neurobiological research has almost universal validity in natural neural networks. We propose two mechanisms of inhibition: (i) *orientation competition* between cells whose receptive fields are centered at the same visual field point and have the same size but differ in their orientations and (ii) *lateral inhibition* between cells which have receptive fields of the same size and orientation but are centered on different points of the visual field. The representations obtained in this way exhibit a high degree of information structuring, in that only edge lines of a particular orientation and light-to-dark transition direction are present in each cortical image. The cortical representations of an input image are used to extract a set of descriptors which is used to search for a nearest neighbour in a database. The present work is an extension of our previous work reported elsewhere [5, 6, 7]. By applying the above sketched scheme, we achieve 100% recognition rate of simple geometrical figures and 99% recognition rate on a large database of face images.

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 and for what reason thresholding, orientation competition and lateral inhibition are made a part of the model. Section 4 outlines the transition from cortical images to descriptor sets used for image analysis and object recognition. Section 5 presents our experiments and results on object recognition.

2 Gabor functions and natural vision

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

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

By means of translations parameterized by a pair (ξ, η) , dilations parameterized by an integer number j and rotations parameterized by an angle φ , one gets the following family of two-dimensional Gabor functions (in the following L denotes the size of the visual field):

$$g_{j,\varphi}(x-\xi, y-\eta) = \frac{\alpha^{2j}}{\pi} e^{-\alpha^{2j}(x'^2+y'^2)+i\pi\alpha^j x'} \quad (2)$$

$$j \in \mathbb{Z}, \quad \varphi \in [0, 2\pi), \quad \xi, \eta, x-\xi, y-\eta \in [0, L]$$

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

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

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

$$\lambda_j = \frac{2}{\alpha^j} \quad (3)$$

and a wavevector of orientation φ and magnitude (the magnitude of the wavevector is also referred to as *spatial frequency*)

$$k_j = \pi\alpha^j. \quad (4)$$

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 α^j ($j \in \mathbb{Z}$) corresponds to equidistant sampling of a logarithmic wavelength/spatial-frequency scale that corresponds to the logarithmic dispersion of spatial frequencies found by neurophysiological research [2, 3].

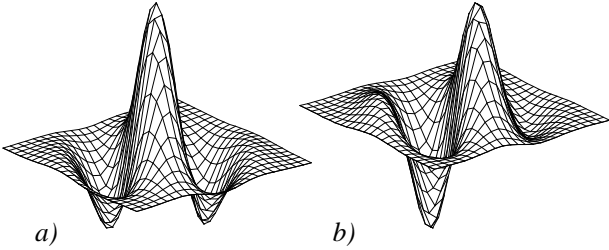


Figure 1: Real (a) and imaginary (b) part of a Gabor function.

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

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

may be considered as the amount of a harmonic wave with wavelength λ_j and wavevector orientation φ in a surrounding area of linear size λ_j centered on a point with coordinates (ξ, η) . In this way, equation (5) represents local spectral analysis which is embedded in global spatial coordinates (ξ, η) . For fixed j and φ and variable ξ and η , $\tilde{s}_{j,\varphi}(\xi, \eta)$ presents a two-dimensional function of the same size as the input image. For ease of reference, in the following we call such a function *Gabor representation*. The coefficient α^{2j}/π 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, $|\tilde{s}_{j,\varphi}(\xi, \eta)| = 1$.

We now assume that the quantities $\tilde{s}_{j,\varphi}(\xi, \eta)$ computed in (5) for the various values of the parameters j, φ, ξ and η correspond to the net inputs to individual cortical cells when the visual system is presented an image $s(x, y)$. Note that the set of Gabor representations $\tilde{s}_{j,\varphi}(\xi, \eta)$ comprises more data than the original image $s(x, y)$. Such data expansion is, however, 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 [8]). We suggest to interpret this data expansion as a means used by the brain to facilitate image analysis and propose a method for further enhancement of information structuring in cortical filters and extraction of representations in a lower-dimension space which can be used for automatic object recognition.

3 Thresholding, orientation competition and lateral inhibition

3.1 Thresholding for distinguishing light-to-dark transition orientation

Note that the quantities $\tilde{s}_{j,\varphi}(\xi, \eta)$ computed in (5) depend linearly on the input image $s(x, y)$ and are complex. For reasons to become clear below, we do not use the real parts of Gabor representations. We consider the imaginary parts of the quantities (5) as net inputs to the cortical cells whereby the output activity $a_{j,\varphi}(\xi, \eta)$ of a cell with receptive field centered

on a point with coordinates ξ and η and characterized by main wavelength λ_j and wavevector orientation φ is determined as the imaginary part of the complex quantity $\tilde{s}_{j,\varphi}(\xi, \eta)$ computed in (5) if this part is positive and is set to zero if this part is negative:

$$a_{j,\varphi}(\xi, \eta) = \Im(\tilde{s}_{j,\varphi}(\xi, \eta)) \text{ if } \Im(\tilde{s}_{j,\varphi}(\xi, \eta)) > 0 \quad (6)$$

$$a_{j,\varphi}(\xi, \eta) = 0 \quad \text{if } \Im(\tilde{s}_{j,\varphi}(\xi, \eta)) \leq 0 \quad (7)$$

The motivation for this transformation is as follows:

- (i) The information provided by the real parts of the Gabor representations is redundant, since the real parts of two Gabor representations for opposite orientations φ and $\varphi + \pi$ are identical. Therefore, the real parts of Gabor representations cannot be used to discriminate between light-to-dark and dark-to-light transitions. At present we do not know how the real part can effectively be used to give information which is additional to that provided by the imaginary part.
- (ii) The information provided by the imaginary parts is also redundant, since the imaginary part of a Gabor representation $\tilde{s}_{j,\varphi}(\xi, \eta)$ for orientation φ is a negative of the imaginary part of the Gabor representation $\tilde{s}_{j,\varphi+\pi}(\xi, \eta)$ for the opposite orientation $\varphi + \pi$. In contrast to the real parts, this redundancy can be alleviated by dropping values which are negative. This choice is in part biologically motivated, since negative, i.e. inhibitory, input to cortical cells cannot cause them to fire. Eqs. (6-7) present a kind of *thresholding* with a threshold zero.

Fig. 2 shows a synthetic input image which is used in the following to illustrate the proposed method. Fig. 3 shows the absolute values of the imaginary part of the Gabor representations (absolute values have to be taken for graylevel visualization, although taking two colours to distinguish positive from negative values would be much more informative) for fixed magnitude of the wavevector ($\lambda \approx L/40$) and different wavevector orientations φ ($\varphi_i = 2\pi i/16, i = 0 \dots 15$). Since absolute values are visualized, the imaginary parts of Gabor representations computed for φ and $\varphi + \pi$ appear as identical (compare the pairs of images in the first and third and second and fourth row for each column of Fig. 3). If shown in two different colours used to represent positive and negative values, they would, however, look different since they are negatives of each other. This asymmetry becomes explicit by dropping the negative values in the thresholded representations.

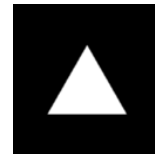


Figure 2: A simple input image.

3.2 Orientation competition for increasing orientation sensitivity

Although the above mentioned redundancy has been removed by the thresholding, there is still a considerable redundancy left. In particular, each edge line is enhanced in several thresholded representations. These results seem to be in contrast with the high orientation sensitivity of the visual system of mammals as confirmed by psychophysical and neurophysiological experiments.

Elsewhere [6, 7] we proposed a method to improve the orientation sensitivity of the cortical representations by a winner-takes-all competition between all quantities $a_{j,\varphi}(\xi, \eta)$ with the same values of ξ, η and j but with different values of φ . One should point out that there is a certain biologically motivated

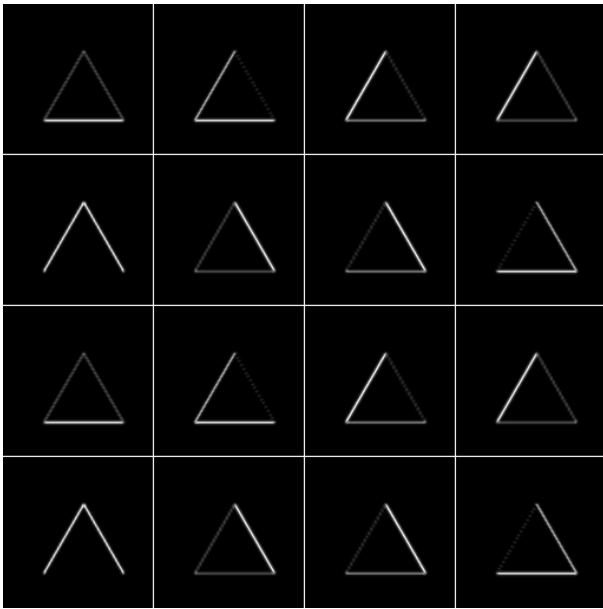


Figure 3: Images based on the absolute values of the imaginary part of the Gabor representations. The first row of images correspond, left to right, to orientations $\varphi_i = 2\pi i/16, i = 0 \dots 3$. The second, third and fourth row correspond to orientations $\varphi_i = 2\pi i/16, i = 4 \dots 7, i = 8 \dots 11$ and $i = 12 \dots 15$ respectively.

rationale behind that decision. It is known from neurophysiological research that the simple cells in the primary visual cortex of mammals are organized in columns and hypercolumns in which they are strongly interconnected. The orientation competition we propose corresponds to inhibitive interconnections between cortical cells whose receptive fields have the same size and are centered on the same point of the visual field, but are oriented in different directions. More precisely, we model this winner-takes-all *orientation competition* in the following way

$$b_{j,\varphi}(\xi, \eta) = a_{j,\varphi}(\xi, \eta) \quad (8)$$

$$\text{if } a_{j,\varphi}(\xi, \eta) = \max\{a_{j,\phi}(\xi, \eta) \mid \forall \phi\}$$

$$b_{j,\varphi}(\xi, \eta) = 0 \quad (9)$$

$$\text{if } a_{j,\varphi}(\xi, \eta) < \max\{a_{j,\phi}(\xi, \eta) \mid \forall \phi\}$$

whereby the quantities $b_{j,\varphi}(\xi, \eta)$ should be considered as the new representation (i.e. the new cortical cell activities after the competition is completed). This scheme better discriminates among different orientations. The effect of orientation competition is that if an edge line is enhanced in a representation corresponding to a given orientation φ , the same line is suppressed in the representations which correspond to neighbouring orientations.

3.3 Lateral inhibition for removing ‘shadow’ lines

An interesting effect is that if an edge line is enhanced in a representations which corresponds to a given orientation φ , the same line appears in the representation corresponding to orientation $\varphi + \pi$ in the form of a pair of parallel lines whose intensity is weaker and which enclose the actual edge line. We refer to such lines as *shadow lines*. To remove this evident redundancy, we next introduce a *lateral inhibition* mechanism in such a way that the actual edge line can suppress the shadow lines. For this purpose, we let a strongly activated cell with receptive field parameters j and φ suppress all less activated

cells having the same receptive field parameters j (size) and φ (orientation) but centered on neighbouring positions within a distance λ_j along a line with orientation φ . More precisely, we compute new representations $c_{j,\varphi}(\xi, \eta)$ as follows:

$$c_{j,\varphi}(\xi, \eta) = b_{j,\varphi}(\xi, \eta) \quad (10)$$

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

$$c_{j,\varphi}(\xi, \eta) = 0 \quad (11)$$

$$\text{if } b_{j,\varphi}(\xi, \eta) < \max\{b_{-j,\varphi}(\xi + \nu\lambda_j\cos\varphi, \eta + \nu\lambda_j\sin\varphi) \mid \forall \nu \in (-1, 1)\}$$

and refer to them as *cortical images*. Fig. 4 shows the cortical images computed in this way for the input image of Fig. 2. Note that each edge line in the input image is enhanced in a distinct cortical image and this property of the method can be interpreted as decomposition of a geometric object into edge lines. In this way, the cortical images computed with the involvement of thresholding, orientation competition and lateral inhibition deliver more structured information than a traditional edge detector such as a Laplacian operator and the previously considered representations.

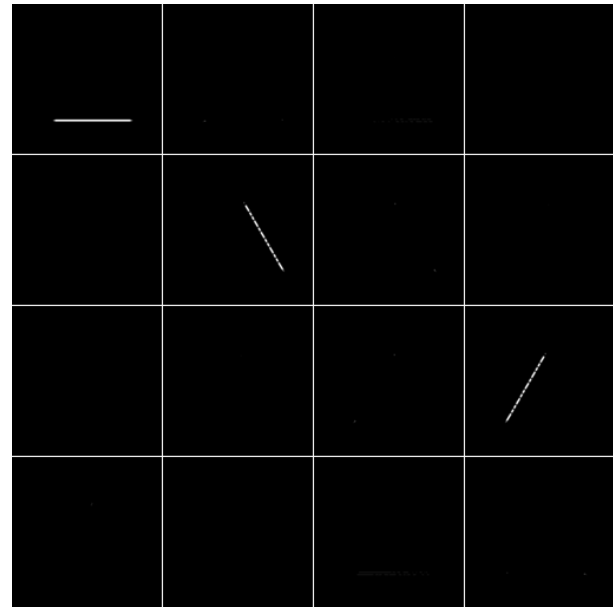


Figure 4: Cortical representations computed by applying thresholding, orientation competition and lateral inhibition.

4 Application to object recognition

Next we use the cortical images for extracting a set of image descriptors to be used for object recognition. Since we have no hints from neurophysiological research about how cortical images could be used in the process of object recognition, we have to make hypotheses about the further representation and processing of visual information. Let us consider the following quantities to be called in the following the *descriptors*:

$$C_{j,\varphi} = \int c_{j,\varphi}(\xi, \eta) d\xi d\eta, \quad j \in \mathbf{Z}, \varphi \in [0, 2\pi). \quad (12)$$

Each of them can be interpreted as the cumulative activity of all cells with the same wavevector orientation φ and main spatial frequency $\pi\alpha^j$, independently of their positions (ξ, η) in the visual field. This naive interpretation is based on

the premise 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 (12) 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 $C_{j,\varphi}$ according to (12) might however make sense for one reason: they are not sensitive to the particular position of an object in the visual field,

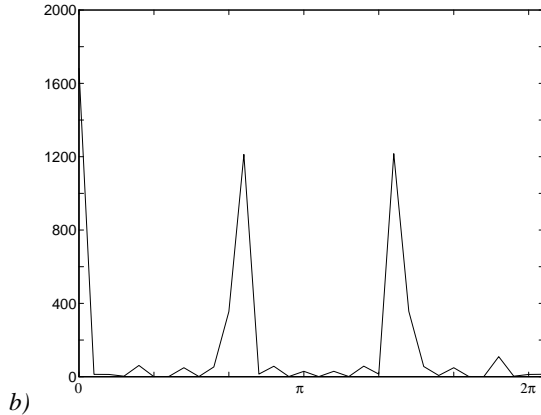


Figure 5: The ultimate cortical representations $c_{j,\varphi}(\xi, \eta)$ computed using thresholding, orientation competition and lateral inhibition.

Fig. 5 shows a plot of the descriptors $C_{j,\varphi}$ for one fixed value of j ($\lambda_j \approx L/40$) and different values of φ , $\varphi \in [0, 2\pi)$. Each of the plotted values of $C_{j,\varphi}$, $\varphi_i = 2\pi i/16$, $i = 0 \dots 15$, is simply the energy of the respective image in Fig. 4. The plot exhibits three very clear peaks which can be interpreted as three dominant edge lines in the original input image.

The developed method is very robust for translations, rotations and scaling. If, for instance, the triangle in the input image is shifted, it would produce virtually the same plot as the one shown in Fig. 5. A rotation of the triangle would lead to a circular shift of the plot. If a triangle of different form is taken (unequal edge lengths), there will be a change in the strength and relative positions of the maxima which can be compensated by dynamic programming.

5 Experimental results

We use the above developed method for the recognition of simple geometric objects. Input images, each containing one polygon, were generated and a system implementing the above described method was requested to classify the objects according to the number of edges they have. Position, size, form and orientation of each polygon were generated at random. In a very large number of trials, the method discriminated with 100% reliability between polygons with different numbers of edges. With the same reliability the method allows to reject objects which do not fall within this class.

An interesting question is whether the method can be applied to more complex objects. For this purpose, we applied the method to the problem of face recognition. A database of 300 different face images of 40 persons has been constructed. Technical details on the database can be found in [5]. For each of the face images, a lower-dimension representation has been computed according to (12) and based on this representation a nearest-neighbour was searched for each of the face images. The search was considered to be successful if the nearest neighbour turned out to be an image of the same person and not successful if it was an image of a different person. A recognition rate of 99% has been achieved. These

results are better than our earlier results on face recognition based on more simple models [5, 6, 7].

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 lots of work has still to be done. In particular, better ways for the extraction of lower-dimension (preferably syntactic) representations have to be found. Further work in progress will be reported elsewhere.

6 Acknowledgement

Most of the computations within this research were carried out on the Connection Machine CM-5 of the University of Groningen. The investments in this supercomputer were partly supported by the Netherlands Computer Science Research Foundation (SION) with financial support from 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] 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.
- [3] 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.
- [4] 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.
- [5] 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
- [6] 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.
- [7] 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)
- [8] M. Connolly 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.