Large Scale Natural Vision Simulations

T. Lourens, N. Petkov and P. Kruizinga

Department of Mathematics and Computer Science, Rijksuniversiteit Groningen
P.O. Box 800, 9700 AV Groningen, The Netherlands
E-mail: {tino, petkov, peterkr}@cs.rug.nl

Abstract

A computationally intensive approach to pattern recognition in images is developed and applied to face recognition. Similarly to previous work, 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. The proposed model includes nonlinearities, such as thresholding, orientation competition and lateral inhibition. The output of the model is a set of cortical images each of which contains only edge lines of a particular orientation in a particular light-to-dark transition direction. In this way the information of the original image is split into different channels. The cortical images are used to compute a lower-dimension space representation for object recognition. The method was implemented on the Connection Machine CM-5[1] and achieved a recognition rate of 97% when applied to a large database of face images.

1 Introduction

The progress in parallel computing has pushed high-performance computing forward in the dozens of Gflops/s domain and the first Tflops/s parallel supercomputers have already been announced. The awareness of the new possibilities offered by high-performance computers has led to ambitious new projects based on the computational approach to solving problems in natural and engineering sciences such as physics, chemistry, astronomy, fluid dynamics, electrical engineering, etc. At the same time, the newly available computing power is hardly used outside the mentioned traditional areas of supercomputing. For instance, while neurobiological research and progress in artificial neural networks have meanwhile led to the insight that mimicing the abilities of the human brain would require tremendous computing power (10^{17}-10^{18} flops/s), little has been done to develop computationally intensive models that would make use of even a fraction of such power and be realizable with contemporary or future supercomputers. Large scale computer simulations are still far from becoming an instrument of neuroscience, a task that they successfully fulfill for a number of years in the other branches of science mentioned above.

The insights in the microstructure of the brain provided by neurophysiological and neurobiological research and progress in parallel computing may open new opportunities for pattern recognition in images. Neurophysiological research has delivered interesting results which can inspire new image analysis models. It is 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 functions 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 quantities which correspond to the activations of simple cells when an input image is projected on the retina. This approach is sometimes popularly referred to as 'applying cortical filters'.

The research carried out until now has given rise to a number of open questions. Among these we consider as most interesting the questions of how the information delivered by cortical filters can be used to analyse images and 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.

We propose the following model: The values computed by Gabor convolvers are not considered as the actual activations of cortical cells but rather as net inputs to the cortical cells. The actual cell activities are computed by *thresholding* of the net inputs. These activities become the subject of two types of mutual inhibition. We use *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 *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

[1]Most of the computation 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).

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 computed cortical images are used to extract a lower-dimension space representation which is used to search for a nearest neighbour in a database. By applying the above sketched model to the problem of face recognition, we achieve a recognition rate of 97% on a large database of face images.

The paper is organized as follows: In Section 2 we introduce the reader to two-dimensional Gabor functions and their relation to natural vision. Section 3 describes the different steps (thresholding, orientation competition and lateral inhibition) used in the model. Section 4 outlines the transition from cortical images to a lower dimension-space representation which is used for image recognition. Section 5 presents our face recognition experiments and results.

2 Gabor filters

The receptive field function 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. In general, some background stimulus, such as constant illumination or random noise, is used to get a certain excitation level and the response to the bright spot stimulus is measured relative to this level. This gives the possibility to measure inhibitory effects. Note that, due to the use of a background stimulus, the receptive field function *cannot* be considered as the impulse response.

The basic two-dimensional Gabor function we use is defined as follows:

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

By means of translations parameterized by a pair (ξ, η), which has the same domain as the coordinate pair (x, y), 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):

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

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

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

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

where α^j is a dilation factor (see below). 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 of orientation φ and magnitude (spatial frequency)

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

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 dilation 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]. We take the basic dilation factor to be $\alpha = \sqrt{2}$.

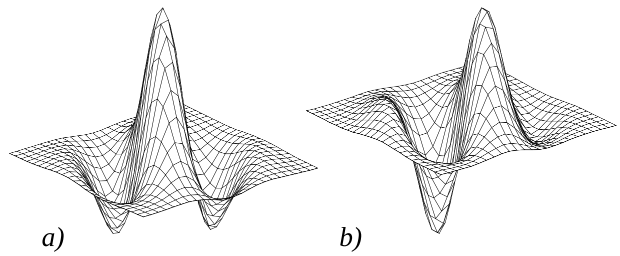


Figure 1: Real (a) and imaginary (b) part of a Gabor function.

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

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

may, roughly speaking, 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 interpretation equation (5) represents local spectral analysis which is embedded in global spatial coordinates (ξ, η).

In the following we 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 each such quantity is complex, so more precisely it represents the net inputs to a pair of cells, with the one cell having a symmetric (real part of the Gabor function) and the other antisymmetric (imaginary part of the Gabor function) receptive field profile. Neurophysiological research has shown that such pairs of cells actually exists [5].

3 The cortical image model

3.1 Antisymmetric functions

At present we are primarily interested in form information as contained in the edges of the objects. A simple cortical cell which is characterized either by the real or the imaginary part of a 2-D Gabor function will react strongly to an edge in its receptive field. While both symmetric and antisymmetric receptive field functions can be used to detect edges, only the antisymmetric functions can give information about the direction of the light to dark transition. Therefore, we only use antisymmetric receptive field functions (imaginary parts of Complex 2-D Gabor functions) in the following. (For more details on this topic see [9].)

3.2 Thresholding

The imaginary parts of the quantities $\tilde{s}_{j,\varphi}(\xi,\eta)$ contain redundant information, since it holds:

$$\Im(\tilde{s}_{j,\varphi}(\xi,\eta)) = -\Im(\tilde{s}_{j,-\varphi}(\xi,\eta)) \tag{6}$$

This redundancy can be rectified by simply skipping all negative quantities. No information is lost, since, if a negative quantity $\Im(\tilde{s}_{j,\varphi}(\xi,\eta))$ is skipped (i.e. set to zero) this quantity can be restored from the quantity $\Im(\tilde{s}_{j,-\varphi}(\xi,\eta))$ which is positive and therefore not skipped.

We use the imaginary parts of the quantities (5) as net inputs to the cortical cells with antisymmetric receptive field functions and the output activity $a_{j,\varphi}(\xi,\eta)$ of a cell with receptive field centered on a point with coordinates (ξ,η) and characterized by main wavelength λ_j and wavevector orientation φ is determined as the imaginary part of the complex quantity $\tilde{s}_{j,\varphi}(\xi,\eta)$ if this part is positive and is set to zero if this part is negative:

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

$$a_{j,\varphi}(\xi,\eta) = 0 \quad \text{if} \quad \Im(\tilde{s}_{j,\varphi}(\xi,\eta)) \leq 0 \tag{8}$$

In this way, we apply *thresholding* to the quantities computed in (5). As mentioned at the beginning, skipping the negative values is used to reduce redundant information. On the other hand, there is a certain biological motivation for thresholding, since it is known that negative (i.e. inhibitive) input to neurons will not cause them to fire but rather prevents them from firing.

3.3 Orientation competition

Redundancy cannot be removed completely by thresholding: identical edge lines, for instance, are enhanced in several thresholded representations and this can be considered as an expression of redundancy [9]. Next we try to reduce this redundancy 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. Elsewhere [7, 8] we used this method to improve the orientational selectivity of cortical filters. This winner-takes-all *orientation competition* is realized as follows:

$$b_{j,\varphi}(\xi,\eta) = a_{j,\varphi}(\xi,\eta) \tag{9}$$
$$\text{if} \quad a_{j,\varphi}(\xi,\eta) = max\{a_{j,\phi}(\xi,\eta) \mid \forall\, \phi\}$$

$$b_{j,\varphi}(\xi,\eta) = 0 \tag{10}$$
$$\text{if} \quad a_{j,\varphi}(\xi,\eta) < max\{a_{j,\phi}(\xi,\eta) \mid \forall\, \phi\}$$

whereby the quantities $b_{j,\varphi}(\xi,\eta)$ should be considered as the new cortical representations. (As to the question whether these quantities should be interpreted as activities of simple or higher cortical cells, see the discussion in [9].) 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.

3.4 Lateral inhibition

An interesting effect is that if an edge line is enhanced in a representation $b_{j,\varphi}(\xi,\eta)$ which corresponds to a given orientation φ, the same line is enhanced in a representation $b_{j,\varphi+\pi}(\xi+\Delta\xi,\eta+\Delta\eta)$ corresponding to orientation $\varphi+\pi$ and displaced from (ξ,η) at a certain distance $(\Delta\xi,\Delta\eta)$ which is within a wavelength λ_j; $|\Delta\xi|,|\Delta\eta| < \lambda_j$. To remove this redundancy, we next introduce a *lateral inhibition* mechanism in which a strongly activated cell with receptive field parameters j and φ suppresses all less activated cells having the same receptive field size j and opposite orientation $\varphi+\pi$ 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:

$$c_{j,\varphi}(\xi,\eta) = b_{j,\varphi}(\xi,\eta) \tag{11}$$
$$\text{if} \quad b_{j,\varphi}(\xi,\eta) = max\{b_{j,\varphi+\pi}(\xi+\nu\lambda_j\cos\varphi, \eta+\nu\lambda_j\sin\varphi) \mid \forall\, \nu \in (-1,1)\}$$

$$c_{j,\varphi}(\xi,\eta) = 0 \tag{12}$$
$$\text{if} \quad b_{j,\varphi}(\xi,\eta) < max\{b_{j,\varphi+\pi}(\xi+\nu\lambda_j\cos\varphi, \eta+\nu\lambda_j\sin\varphi) \mid \forall\, \nu \in (-1,1)\}$$

3.5 Cortical Images

For fixed j, φ and variable (ξ,η), $c_{j,\varphi}(\xi,\eta)$ are two-dimensional functions to which we refer as *cortical images*. Fig.2 shows three input images which are used to illustrate the proposed method.

Fig.3, 4 and 5 show the cortical images computed for the input images of Fig.2. The first row of images correspond, left to right, to orientations $\varphi_i = 2\pi i/16$, $i = 0\ldots3$. The second, third and fourth row correspond to orientations $\varphi_i = 2\pi i/16$, $i = 4\ldots7$, $i = 8\ldots11$ and $i = 12\ldots15$, respectively.

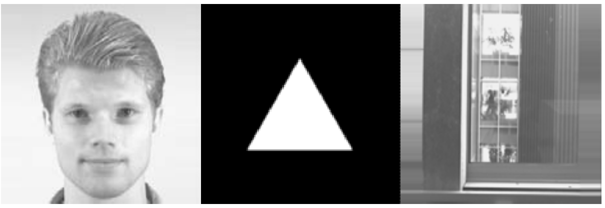


Figure 2: Three different input images.

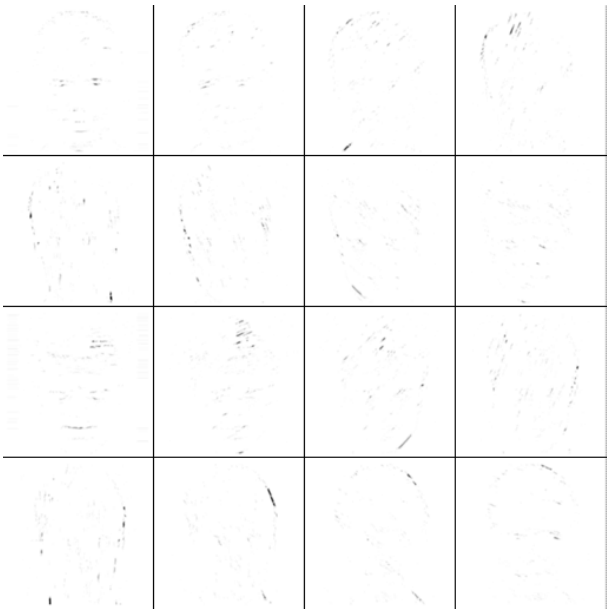


Figure 3: Cortical representations of the face image.

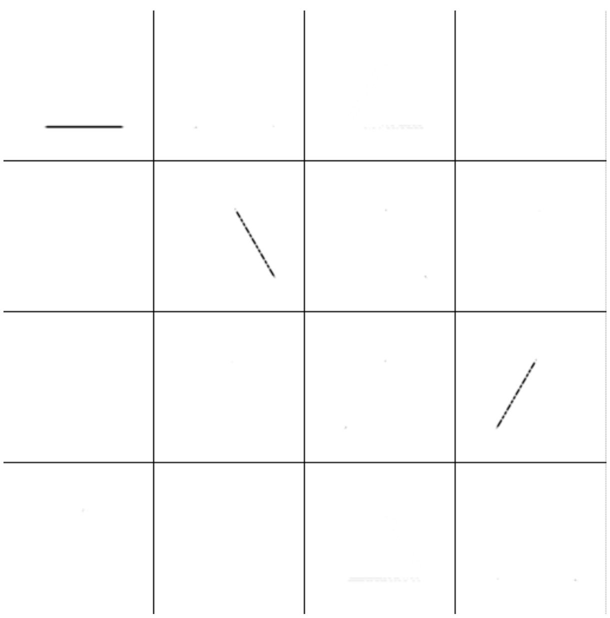


Figure 4: Cortical representations of the triangle.

The same basic wavelength λ is used for all computed images. Note that 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, in that each cortical image contains only transitions of particular size and orientation.

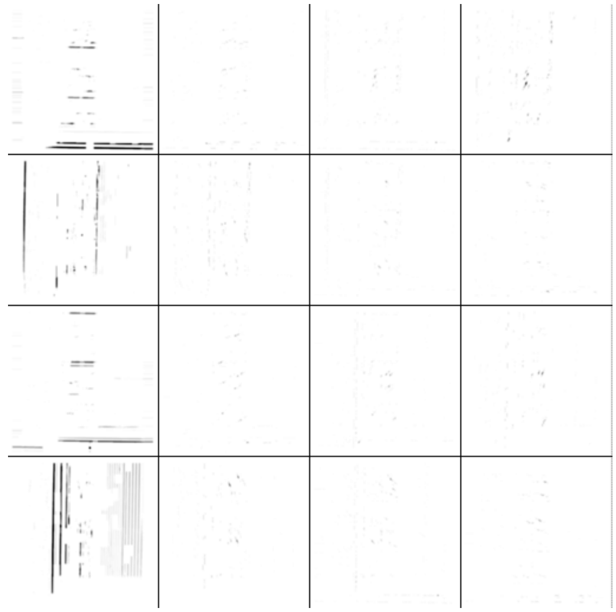


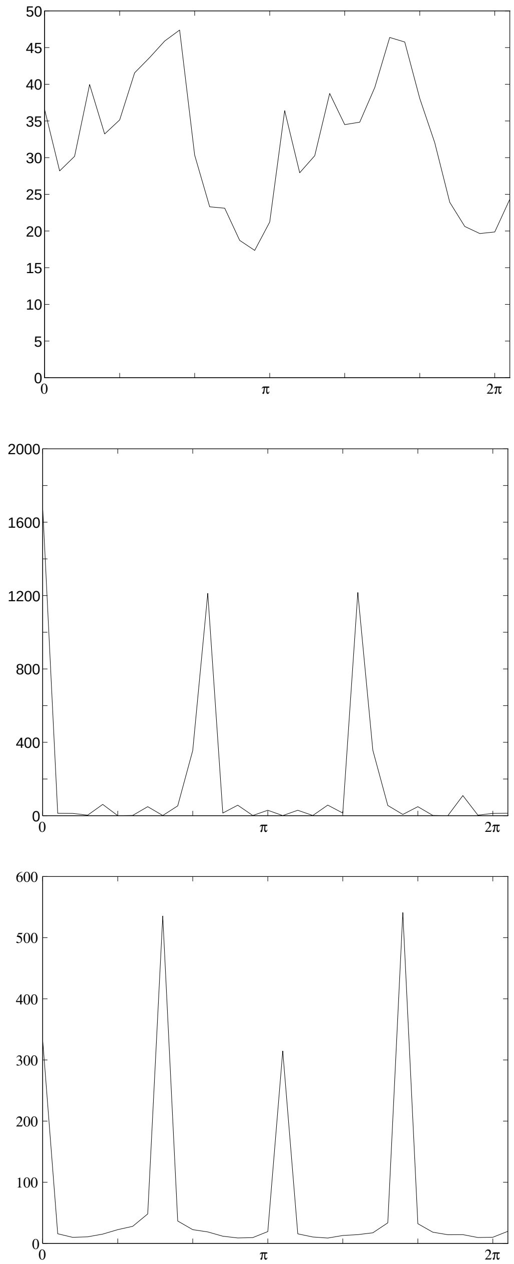
Figure 5: Cortical representations of the building image.

Fig.4 illustrates best what the effect of the used model is. The input image, a triangle in this case, is split in edges which appear in different channels. The edges of the "building image", are also clearly detected and appear in different cortical images (Fig.5). A face has smooth transitions in every direction, therefore there is activity in all cortical channels and the splitting of visual information is rather uniform across all channels.

4 Lower-dimension space representation

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

$$C_{j,\varphi} = \int c_{j,\varphi}(\xi,\eta) d\xi d\eta, \qquad j \in \mathbb{Z},\ \varphi \in [0, 2\pi). \quad (13)$$



$a)$ $b)$ $c)$

Figure 6: Lower-dimension space representations $C_{j,\varphi}$ plotted for a fixed j and different φ: representations of the face (a), triangle (b) and building (c).

Fig.6 shows plots of the quantities $C_{j,\varphi}$ for one fixed value of j and different values of φ computed for the three example input images. Each of the plotted values of C_{j,φ_i}, $\varphi_i = 2\pi i/16,\ i = 0 \ldots 15$, is simply the energy of the respective images in Fig.3, 4 and 5.

Note that a considerable activity in a given cortical image is directly translated in a peak in the respective lower-dimension space representation. This is best illustrated by Fig.4 and Fig.6b where the edges in three of the cortical images of Fig.4 are translated into peaks of the plot shown in Fig.6b. Similarly the increased activity in four of the cortical images shown in Fig.5 leads to four peaks in the corresponding plot of the lower-dimension space representation shown in Fig.6c.

One has to note that after the summation (13) all local information is lost so that even a partial reconstruction of the original image is not possible. Therefore, a certain interpretation of the lower-dimension space representation is possible only if there is some a priory knowledge: if for instance it is known that an input image is allowed to contain only one convex polygon, one may interpret three peaks in the plot of Fig.6b as a triangle. In the same context, the plot in Fig.6c can be misinterpreted as due to a square in the input image.

In spite of these evident limitations, the proposed lower-dimension space representations are quite different for different input images as illustrated by Fig.6 and one can try to use them for object recognition. Similar input images give rise to similar lower-dimension space representations as illustrated by the plots of Fig.6a and 7 which correspond to two different face images. On the other hand, it is important that the representations computed contain sufficient information to enable discrimination of images of the same class. Our face recognition experiments reported below show that this is the case.

Of course, a representation which is to be used for object recognition should be robust for translations, rotations and scaling. This is the case with the proposed representation. If, for instance, the face image is shifted, it would produce virtually the same plot as the one shown in Fig.6a. A rotation of the image would lead to a circular shift of the plot and can easily be compensated. Scaling of an image would lead to the same plot but obtained for a different basic wavelength (value of j), this can also be easily compensated.

5 Experimental results

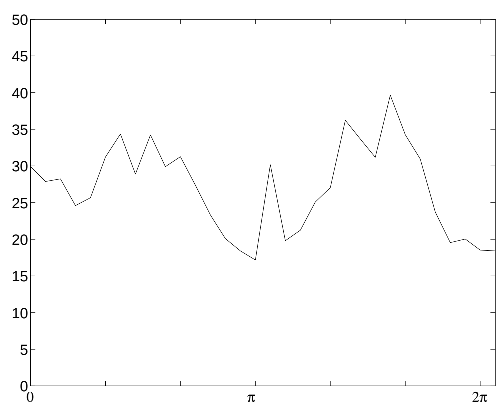


Figure 7: The lower-dimension space representation of another face.

We applied the developed 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 [6]. For each of the face images in the database, a lower-dimension representation has been computed according to (13). 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 another image of the same person (Fig.8) and not successful if it was an image of a different person (Fig.9). With the above described method we achieved a recognition rate of 97%.

Figure 8: Examples of successful searches.

Figure 9: Examples of failures of the model.

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. In spite of the excellent results achieved in our experiments 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.

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

[6] 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.

[7] 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.

[8] 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) pp. 821-826.

[9] N. Petkov, P. Kruizinga and T. Lourens: "Orientation Competition in Cortical Filters - An Application to Face Recognition", in H.A. Wijshoff (ed.) Proc. of *Computing Science in the Netherlands*, Nov. 9-10, 1993, Utrecht, The Netherlands, pp. 285-296.

[10] 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.