Use of Cortical Filters and Neural Networks in a Self-Organising Image Classification System

Nikolay Petkov

Centre for High Performance Computing and
Institute of Mathematics and Computing Science
University of Groningen
P.O.Box 800, 9700 AV Groningen, The Netherlands
petkov@cs.rug.nl

Abstract. A preprocessor based on a computational model of simple cells in the mammalian primary visual cortex is combined with a self-organising artificial neural network in order to set up an image classification system. After learning with a sequence of input images, the output units of the system turn out to correspond to classes of input images. Notably, a group of output units which are selective for images of human faces emerges. These units mimic the behaviour of face selective cells that have been found in the inferior temporal cortex of primates.

1 Introduction

The system described in this paper was inspired by two facts which are known from neurophysiological research on the visual cortex of primates. The first fact is that the majority of neurons in the primary visual cortex react strongly, in one way or another, to oriented lines, line and curve segments, bars and edges [2, 3]. The second fact is that in certain hierarchically high areas of the visual cortex of monkeys, more precisely in the inferior temporal cortex, cells have been found which react selectively to very complex visual patterns, such as faces [5].

As to the first above mentioned fact, the extensive neurophysiological studies carried out in the past three and a half decades have provided models of primary cortical cells based on so-called receptive field functions. On the basis of such models, one can compute quantities which correspond to the activities of primary visual cortex cells when an arbitrary input image is projected on the retina, i.e. one can compute (an approximation to) the primary cortical representation of that image.

As to the second above mentioned fact, at present there is no computational model of face selective and, more generally, class selective cells. The facts known about the cortical areas between the primary cortex and the inferior temporal cortex are not sufficient to complete the picture and become able to conceive a full computational model which would make possible automatic classification according to and with the efficacy of the mechanisms employed by natural visual systems. In this study, a self-organising artificial neural network is used to bridge this gap in our knowledge of the visual system in order to set up an artificial image classification system.

2 Preprocessor based on cortical filters

The following model is used to compute the response r of a visual neuron characterised by a receptive field function (impulse response) $g(x, y)$ to a composite input visual signal $s(x, y)$, $(x, y) \in \Omega$ (Ω denotes visual field):

An integral

$$\tilde{s} = \iint_{\Omega} s(x, y)g(x, y) \, dx dy \quad (1)$$

is evaluated and the result $\tilde{s}$ is submitted to thresholding

$$r = 0 \quad \text{if} \quad \tilde{s} \leq T, \quad (2)$$

where T is a threshold value, and non-linear local contrast normalisation:

$$r = \frac{r_{\infty} \tilde{s}}{c_{\frac{1}{2}} L + \tilde{s}} \quad \text{if} \quad \tilde{s} > T, \quad (3)$$

where L is the mean illuminance within the receptive field of the concerned neuron and r_{∞} and $c_{\frac{1}{2}}$ are constants which specify the saturation response and the value ($\tilde{s}_{\frac{1}{2}}/L$) of the contrast for which half-saturation is reached, respectively.

In the following, a family of Gabor functions proposed by Daugman [1] to model the responses of so-called simple cells in the mammalian primary visual cortex is used in a slightly modified parametrisation form:

$$g_{\xi, \eta, \sigma, \gamma, \Theta, \lambda, \varphi}(x, y) = e^{-\frac{(x')^2 + \gamma^2 y'^2}{2\sigma^2}} \cos(2\pi \frac{x'}{\lambda} + \varphi) \quad (4)$$

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

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

The arguments x and y specify the position of a light spot in the visual field and $\xi, \eta, \sigma, \gamma, \Theta, \lambda$ and φ are parameters as follows:

The pair (ξ, η) , which has the same domain Ω as the pair (x, y) , specifies the *center of a receptive field* within the visual field. The standard deviation σ the Gaussian factor determines the (linear) *size of the receptive field*. The eccentricity of the Gaussian factor and herewith the eccentricity of the receptive field ellipse is determined by the parameter γ which is referred to as the *spatial aspect ratio*. Its value has been found to vary in a very limited range of $0.23 < \gamma < 0.92$. One constant value $\gamma = 0.5$ is used in this study.

The angle parameter Θ ($\Theta \in [0, \pi)$) specifies the *orientation* of the normal to the parallel excitatory and inhibitory stripe zones (this normal is the axis x' in eq.4). The parameter λ is the *wavelength* of the harmonic factor $\cos(2\pi \frac{x'}{\lambda} + \varphi)$. The ratio σ/λ determines the bandwidth of a cell in the spatial frequency domain and the number of parallel excitatory and inhibitory zones which can be observed in its receptive field. The value $\sigma/\lambda = 0.5$ which has biological relevance [1] is used in this study.

Fig. 1. An input image (a) and cortical images (b–h) computed from it using receptive field functions of different orientations and symmetries: b) $\Theta = 0$, $\varphi = -\frac{1}{2}\pi$; c) $\Theta = 0$, $\varphi = \frac{1}{2}\pi$; d) $\Theta = 0$, $\varphi = \pi$; e) $\Theta = \frac{1}{2}\pi$, $\varphi = -\frac{1}{2}\pi$; f) $\Theta = \frac{1}{2}\pi$, $\varphi = \frac{1}{2}\pi$; g) $\Theta = \frac{1}{4}\pi$, $\varphi = -\frac{1}{2}\pi$; h) $\Theta = \frac{3}{4}\pi$, $\varphi = -\frac{1}{2}\pi$.

Finally, for the phase parameter φ we use the following values: 0 for symmetric receptive fields with an excitatory central lobe, π for symmetric receptive fields with an inhibitory central lobe and $-\frac{1}{2}\pi$ and $\frac{1}{2}\pi$ for antisymmetric receptive fields with opposite polarity.

Substituting a receptive field function $g_{\xi,\eta,\sigma,\gamma,\Theta,\lambda,\varphi}(x,y)$ in eqs.1-3, one can compute the response $r_{\xi,\eta,\sigma,\gamma,\Theta,\lambda,\varphi}$ of a simple visual cortical cell modelled by this function to an input image $s(x,y)$. Fig.1a shows an input image for which a number of such quantities are computed and shown in Fig.1b-h, grouped together in so-called *cortical images*. The quantities grouped in one such image are computed with receptive field functions of the same values for all parameters but (ξ,η) ; the later specify the coordinates of a pixel to which the value of such a quantity is assigned. The filters which generate such images are referred to as *cortical filters* or *channels*.

Roughly speaking, the effect of such a filter is to enhance luminance transitions of a given orientation and at a give scale. The filters with antisymmetric receptive field functions are more selective for edges and those with symmetric receptive field functions are more selective for bars of a given width. These selectivity properties of the cortical filters can be further enhanced by additional non-linear mechanisms in which the four filters responsible for the same orientation but having different phase (symmetry) interact with each other to produce cortical images in which either edges or bars (but not both at the same time) are enhanced in a cortical image [6].

3 Classifier based on a self-organising neural network

Fig.2 shows a two-layer artificial neural network constructed of a set of one-dimensional Kohonen networks [4]. The first layer consists of m independent networks, each of n nodes. Each first-layer network accepts a cortical image as input and the number m of such networks is equal to the number of cortical channels used. In this pilot study, a limited number of $m = 32$ cortical channels were used.

Fig. 2. A separate one-dimensional Kohonen network is associated with each cortical filter channel, accepting the corresponding cortical image as an input. The two-dimensional activity pattern produced by the set of networks in the first layer is used as an input pattern to a second-layer (output) Kohonen network.

As to the number of nodes n , it is chosen depending on the target application taking into account the following considerations: (i) n has to be greater than the number of classes which have to be discriminated and (ii) n should not be very large to prevent a very sparse assignment of units to classes and a large ensuing number of learning cycles needed to achieve convergence. In this study, the number of units used for each first-layer network was taken to be $n = 20$. This number was chosen according to the above mentioned considerations and a particular test application in which five classes of objects (images of faces, bottles, armchairs, tea cups and triangles) had to be discriminated.

For each first-layer network, the learning sequence consisted of 30 cortical images computed from 30 corresponding input images (six images per class for each of the above mentioned classes). Convergence was achieved amazingly very quickly — typically in less than ten epochs (presentations of the learning sequence). The fast convergence can be explained by the fact that single cortical images are simple patterns containing a small number of features (oriented elongated connected activity regions). In most of the cases these patterns are disjoint from class to class so that, after such a pattern is assigned to a unit, it cannot be modified by patterns which arise from images of a different class. As to self-organisation, as expected the units which are activated by images of the same

class tend to build clusters, but interleaved units of different classes or units which are never activated as well as mixed-class units, i.e. units which react to image patterns that belong to different classes, can also be observed.

The existence of mixed-class units means that, if one would use just one single-orientation cortical channel with one associated Kohonen network, one can get misclassifications. This is not amazing, since for complex visual input patterns, individual cortical images are not necessarily characteristic of the classes of the corresponding input images. However, *combinations of such images* are characteristic of the class to which an input image belongs.

The proposed two-layer network structure is based on this assumption and its function can intuitively be explained as follows: Each of the first-layer networks associated with the corresponding cortical channels makes its own classification of the cortical image it receives as input. One can think of the first-layer networks associated with different cortical channels as voting for different classes. These votes are counted by another network arranged in the second layer and the class which collects the largest number of votes wins.

The second-layer network has a structure which is similar to the structure of the individual networks in the first layer. It is also one-dimensional and has the same number of units n and the same neighbourhood relations. The only difference is that, while the inputs to the first-layer networks are cortical images of size $k \times k$ ($k = 32$ in this particular case), the inputs to the second-layer network are $m \times n$ binary activity patterns produced by the m first-layer networks, each of n units ($n = 20$, $m = 32$ in this case).

The learning process for the second layer is started after the learning in the first layer is completed. The learning sequence for the second-layer network consisted of 30 activity patterns induced in the first-layer networks for the corresponding 30 input images. Similar to the convergence behaviour of the first-layer networks, the learning process for the second layer took not more than ten epochs. The classification of the learned input patterns was always correct, i.e., images from different classes always activated different units whereby the units corresponding to a class tend to form a cluster. The units of one of these clusters are activated only by input images of faces and in that they mimic the function of face selective cells in the inferior temporal cortex. As to test patterns, the system was tested only for the classification of images taken under similar conditions as the learning images. The system succeeded to classify all test images correctly.

4 Implementation and conclusions

Computing one cortical image on a powerful contemporary workstation takes approximately 5 seconds. In this study only 32 cortical images were computed for each input image. In an ongoing extension, a set of 320 cortical images is computed for each input image. This computation takes more than half an hour on a powerful workstation. In order to accelerate computations, the set of cortical filters has been implemented on a Connection Machine CM-5 scale 3

parallel supercomputer (16 nodes, 64 vector units, 2 Gflop/s, 512 Mbyte) and effective acceleration by a factor of nearly forty was achieved.

The computational effort connected with the neural network part of the system depends on the number of units used. In our first experiments, this number was relatively small, so that computing time did not present a major problem both for learning and classification. The real time learning and classification for a very large number of cortical channels and thousands of different image classes exceed the power of the currently available parallel supercomputers.

The presented system does not incorporate any a priori model of the visual patterns which have to be learned and classified. No knowledge of the visual world, such as the fact that a face has two eyes, a nose, a mouth, etc., or that a bottle has a body and a neck, is used. The system computes its own internal representations of visual patterns, memorises them and uses them to classify new patterns. The output units of the classifier subsystem correspond to classes of visual patterns and not just to individual patterns (generalisation property). This in a way corresponds to the neurophysiological observation that face selective cells are broadly tuned, in that such a cell would react to different faces instead of just one individual face.

As demonstrated by the above example, studying and simulating the principles and mechanisms employed by natural vision systems can lead to new image analysis and object recognition techniques which may have the potential to outperform traditional machine vision approaches. For further details and discussion the reader is referred to [6, 7]. The current paper is a short version of [7].

References

1. Daugman J.G.: "Uncertainty relations for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters", *Journal of the Optical Society of America A*, Vol. 2 (1985) No. 7, pp.1160-1169.
2. Hubel D.H. and Wiesel T.: "Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex", *J. Physiol. (London)*, Vol. 160 (1962), pp.106-154.
3. Hubel D.H.: "Explorations of the primary visual cortex, 1955- 1978" (1981 Nobel Prize lecture), *Nature*, Vol. 299 (1982) pp.515-524.
4. Kohonen T.: *Self Organization and Associative Memory* (Berlin: Springer-Verlag, 1990, third edition)
5. Perrett D.I., Hietanen J.K., Oram M.W., and Benson P.J.: "Organization and functions of cells responsive to faces in the temporal cortex", *Phil. Trans. Roy. Soc. Lond. B: Biolog. Sci.*, Vol. 335 (1992) pp.23-30.
6. Petkov N.: "Biologically motivated computationally intensive approaches to image pattern recognition", *Future Generation Computer Systems*, special issue *HPCN'94*, ed. W. Gentzsch, 1995, in print.
7. Petkov N.: "Image classification system based on cortical representations and unsupervised neural network learning", Proc. Int. Conf. on *Computer Architectures for Machine Perception, CAMP'95*, Villa Olmo, Como, Italy, September 18-20, 1995 (IEEE Press, 1995, in print).