VISUAL SCALE INDEPENDENCE IN A NETWORK OF SPIKING NEURONS

Dipl. Eng. RAUL MUREŞAN

Nivis Research
Gh. Bilascu, Nr. 85, Cluj-Napoca,
Cod 3400, ROMANIA, EUROPE (www.nivis.com)
raulmuresan@personal.ro

ABSTRACT

The scale independence in visual recognition tasks is still one big problem in neurocomputing today. This paper presents a method of obtaining scale independence in a purely feed-forward way, being able to account for ultra-rapid visual categorization. I used a retinotopic architecture of simple spiking neurons with different types of receptive fields, organized in a hierarchical fashion similar to the mammal visual path. Fast shunting inhibition had been implemented using a rank-order coding similar to that described by S. Thorpe [6]. Scale independence had been achieved by using different sized end-stopping bar detectors and combining then in a scalable way to produce scale independent response over a given domain. This solution does not conflict with the saliency based models and offers a great robustness to clutter.

Keywords: Scale independence; Rank order coding; Feed-forward; Receptive field; Neuro-computing.

1. INTRODUCTION

Studies on the mammal visual cortex had shown that cortical simple cells are inside the receptive field of complex and hipercomplex cells. Since Hubel and Wiesel [2] it has been accepted that this hierarchical model is used to obtain position and scale independence. But some argued that such a composition could lead to the loss of relative relationships between different elements of an object, thus "the binding problem" has appeared [4]. The solution to this had been thought to be pulse synchronization, which could retain the relationship information.

Recent studies however, had shown that ultra-rapid visual categorization is possible, in a time magnitude under 150 ms in human visual neocortex. This is exactly the timing necessary for the information to reach the infero-temporal (IT) cortex neurons responsible for object recognition [5].

Under such circumstances, a natural question arises: how can the brain recognize objects (mainly unfamiliar objects, presented only few times to subjects) in a scale independent manner since there is no time for pulse synchronization to occur?

A solution to this problem will be presented in this paper, explaining at the same time, why contour integration is so important and why the bar detectors have such a huge importance in object recognition.

2. METHODS

For testing reasons and modeling, I implemented a neural simulator based on the retinotopic organization of the visual cortex. The simulator, named "RetinotopicNET" can successfully trace networks with millions of neurons and a magnitude of 10^{10} synapses in a matter of seconds. This high performance is due to the event-based type of simulation.

Neurons were simple integrate-and-fire cells with fast shunting inhibition implemented as exponential modulated synapses [6]. No leakage has been included in the model since the amount of current leak in the short period the neuron's state is pooled can be neglected (no rate based coding is present).

Each time an afferent spiked the sensitivity of the synapses had been decreased by a modulation factor as follows:

Sensitivity ← Sensitivity * Modulation

where:
- Sensitivity represents the synaptic sensitivity over all synapses
- Modulation is a number in the range [0..1]

The architecture of the model contains 7 levels of processing, following the retinal, V1, V4 pathway up to the infero-temporal cortex. The first 2 levels are similar to the architecture used by Arnaud Delorme [1].

2.1. Architecture

The seven layers of processing correspond to an ascending feed forward processing with lateral interactions at some of the levels (Fig. 1). The key feature of the model is the use of extensive competition

between different elements of the object to be recognized. The only information used at this time is contour information but blob-type cells could also be included to account for color or intensity patches as well.

Level 1 : Retinal processing

At the first layer of processing the retinal ganglion cells process the incoming image intensities (only intensity 8 bit grayscale images were used). The ON-OFF effect has been achieved by using a classical difference-of-gaussians (DOG), center-ON-surround-OFF and viceversa filter with a ratio of 1 to 3. Then, the image intensity for the two maps has been converted into spike latency and spikes were fed into the "RetinotopicNET" simulator.

Level 2 : V1 Area

The second layer of processing corresponds to the V1 primary cortex area where different orientations are selected by oriented Gabor-like receptive fields. These are the corresponding simple cells, which detect different orientation contrasts.

One key feature is the lateral connection within each orientation map. I have used a butterfly-like lateral connection, which has the property of improving contours. This is a form of primitive contour-integration, but due to the lack of iterative loops only a feed forward contour completion is used. Important work on this matter had been conducted by Zhaoping Li [3]. Further improvement on the system may be achieved by implementing a stronger contour integration mechanism.

The Gabor patches were all at the same scale and had a spatial frequency of 0.5 pixels. They covered the range of 0 to 180 degrees with over a total of 8 orientations.

Level 3 : Bar-like detectors

For each orientation, a corresponding set of different scaled-receptive-field maps was used to extract the bar-like feature at the corresponding position.

Each receptive field contained an oriented bar-like, end-stopping type. The central, elongated bar, corresponds to excitation. The surrounding area corresponds to weak inhibition proportional to the level of blackness (Fig. 2). This type of receptive field tunes the neuron to the bar that best matches its excitatory

size. For the same orientation, multiple scaled bar-detectors were used.

Considering one set of oriented scaled maps (multiscaled maps for the same orientation), lateral inhibition has been introduced from large to small sizes, generating a size competition at that orientation. The priority in terms of timing varied from large to small bars. In other words, the maps with a large receptive field had the chance of firing first and, by the means of inter-map lateral inhibition, the smaller bar detectors were inhibited. Such a mechanism ensures that the largest size possible is always detected (instead of composing it from multiple smaller size bars).
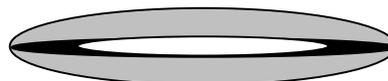

Fig. 2. Bar-like detector for 0^0 orientation

The white bar in the center corresponds to excitatory synapses; the gray and black areas correspond to inhibitory synapses. The black is the strongest inhibition, the gray level being a weaker inhibition.

Level 4 : Multiscale downsample

The fourth layer of processing is responsible for bringing every detail to the same level of spatial importance. In other words, if two long lines are detected, the distance between them has to be brought down to the same distance as the one between the same lines scaled down (in an object scaling operation) to a smaller size. The key mechanism of scale independence is exactly the equivalence of feature distance with feature size. This level is the most important one for scale independence and we shall describe the mechanism in detail.

Let us consider, as an example, a simple object, formed of just two lines, oriented at 0^0, as shown in figure 3.
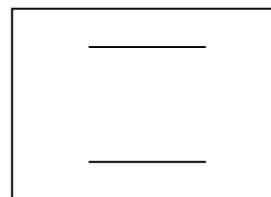

Fig. 3. An object formed of 2 lines, of size 20 pixels each, inside an image.

Image

Level 1 :
Retina

Retina ON units

Retina OFF units

Level 2 :
V1 Area

0^0 V1 units

22.5^0 V1 units

45^0 V1 units

. . .

180^0 V1 units

Level 3 :
Bar-like
multiscale
detectors

0^0 Bar-like Detect Scale S1

22.5^0 Bar-like Detect Scale S1

45^0 Bar-like Detect Scale S1

180^0 Bar-like Detect Scale S1

0^0 Bar-like Detect Scale Sn

22.5^0 Bar-like Detect Scale Sn

45^0 Bar-like Detect Scale Sn

180^0 Bar-like Detect Scale Sn

Extensive
lateral
competition

Level 4:
Multiscale
neural
downsample

0^0 Scale 1

S 2

Sn

. . .

180^0 Scale 1

S 2

Sn

Level 5:
Scale
independent
oriented bars

0^0 Scale combine

. . .

180^0 Scale combine

Level 6:
Orientation
independence
(optional)

V4 final stage

Level 7:
Infero-Temporal
Cortex
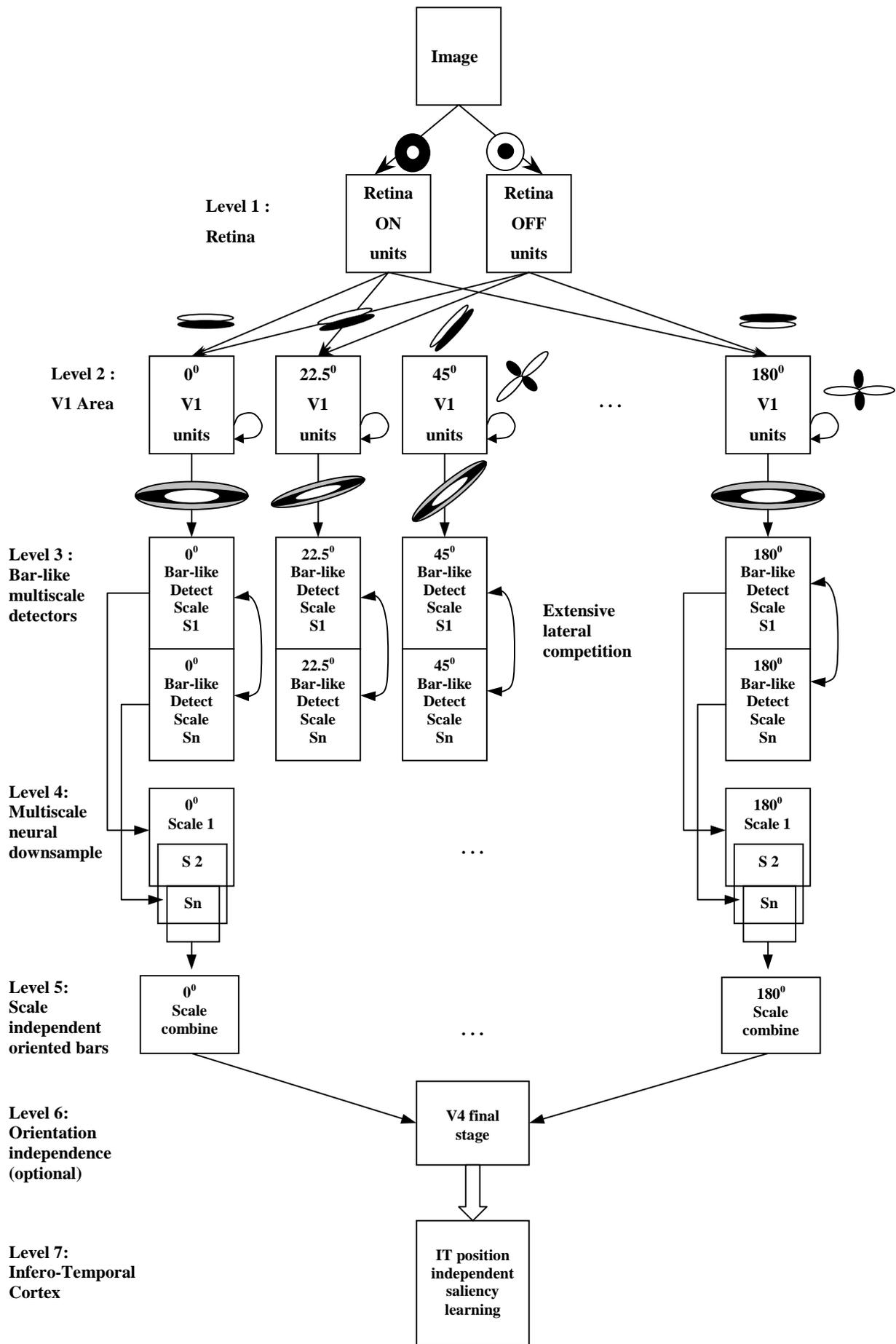
IT position independent saliency learning

Fig. 1. Model architecture

For explanation purpose, let us consider that our scaled bar detectors range over 10 to 30 neurons (pixels) and that the lines in the original image (Fig. 3) have a size of 20 pixels. We have 3 maps of 0^0 orientation with bar detectors at 10, 20 and 30 neurons (pixels).

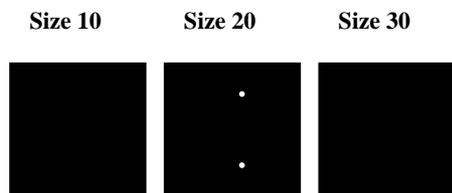The response of the 3 maps to the presented image is presented in Fig. 4.

Size 10   **Size 20**   **Size 30**

Fig. 4. The response of the 3 maps to the original object with lines of size 20.

Because of the strong lateral inhibition and competition, neurons in the map of size 10 can't fire. Neurons in the map of size 30 have not enough stimulation in the excitatory area to be driven by the 2 lines. Thus, only in the map of size 20 the activity will exist.

Now let us scale down by a factor of two the original image (Fig. 5).
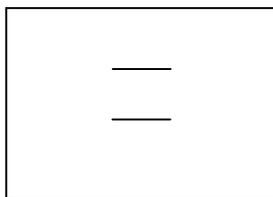

Fig. 5. The original object scaled down by a factor of 2.

The result of the down scaling is that activity will move down to lower sized detectors by a distance proportional to the ratio between the size of different detector maps (Fig. 6).
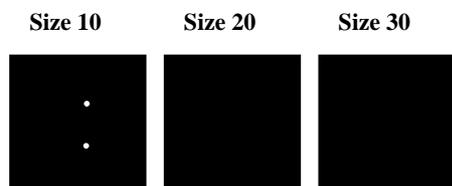

Size 10   **Size 20**   **Size 30**

Fig. 6. The response of the 3 maps to the scaled object with lines of size 10.

We take a look at the distance between the two lines: by scaling the object down, the distance between its parts is also scaled down yielding a cortical response with scaled distances between bar detector neurons. All the system has to do, in this simple case, is to scale down the second map by a factor of 2 and

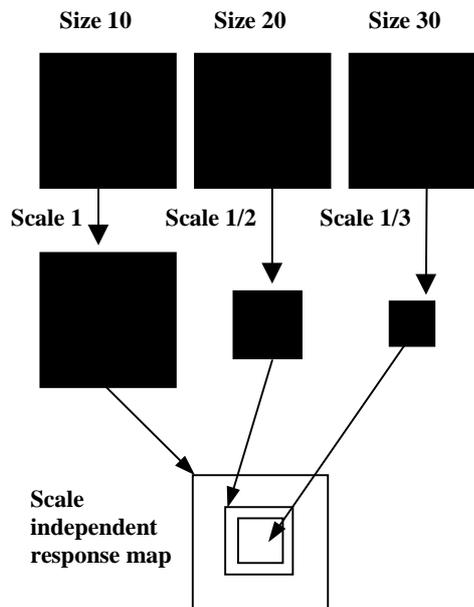the third by a factor of 3 and feed all of these resulting maps into a scale invariant map (Fig. 7).

Fig. 7. Neural downsampling and combination to achieve scale independence

Some may argue that for an object that has lines which will be detected by different sized detectors at the equivalent locations (as dictated by the scale of detection), the neural responses will overlap each other in the final scale invariant map. But we have to take into account that for that specific location, the final neuron will be have a multiplied feeding and thus a multiplied fire rate (the multiplication depends on the number of responses that overlap). The fire rate captures the equivalence information and the location captures the relative position of features! The binding problem is not applicable in our case because relative position of features is encoded into the final map in a salient manner.

The neural downsampling is achieved by using window-like receptive fields which could be associated with center-ON surround-OFF receptive fields in area V4, taking into account the fact that the surround-OFF is a very silent small inhibition which could be used for stability and normalization purposes.

Level 5 : Scale independence maps

At the next layer, at each orientation, the downsampled oriented maps are combined into a scale independent map corresponding for that specific orientation (Fig. 7). The mechanism of combining them should take into account the scaling center of the object. Further improvement, as position independence of features, could be implemented at this level.

Level 6 : Orientation independence

This layer is optional, and in used only for reducing the number of synapses with the infero-temporal map. It corresponds to the final stage in the V4 area. There is no reason for which one might consider different orientations as being equivalent but this type of combination can be used to simulate the hypercomplex cells. Orientation equivalence can be used to improve generalization capability of the recognition in the IT cortex.

Level 7 : Infero-Temporal Cortex

The infero-temporal cortex is responsible for object recognition. In the architecture presented, learning is performed by increasing the synapse strength with the current sensitivity value as resulted from successive modulator effects generated by the firings in the level 6 map. This mechanism is similar to the one used by Arnaud Delorme [1]. Every neuron in the final IT map has a retinotopic type of receptive field, covering most of the level 6 map. The synaptic strengths are shared among all neurons, yielding a good position indepen-dence.

3. RESULTS

Using the "RetinotopicNET" simulator I calibrated the system for face detection (and recognition). The test database had been generated using a QuickCam web camera and consisted of the faces of three different persons with different face expression.

Image sizes were fixed at 92 x 112 grayscale bitmaps and the faces were scaled in a range 1 to 0.58 the original scale (Fig. 8).
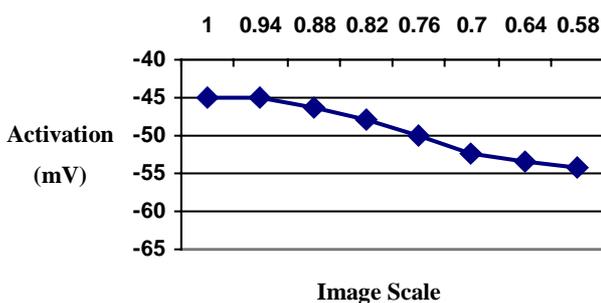

Fig. 8. Activation of infero-temporal cortex for different image scales. The IT layer has been trained to reach the exact threshold activation of -45 mV for the face at the original size.

The number of scales used at level 4 was 7, bar-like detectors ranging from a length of 7 to 13 (7,8,...13). At the infero-temporal level, a strong shunting inhibition had been used to provide enhanced selectivity on learning (for the face recognition case).

The selectivity map of the trained infero-temporal neuron is shown in figure 9 and different sized details can be observed at different positions.
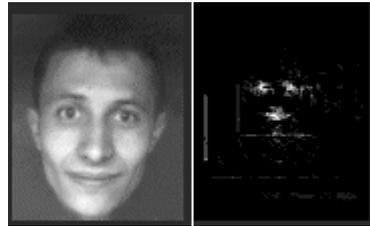

Fig. 9. Training image and the selectivity of the infero-temporal cortex

The given results are surprisingly good, taking into account the limited scale levels used and the size of the details in the image (the details are by far larger than the detectors used). Increasing the size of the bar-detectors and their number (to cover a wider field of scales) can increase accuracy of recognition. At the same time, I expect that the usage of more orientations can increase accuracy because of the better localized bar detection at the level 3 of the architecture.

4. REFERENCES

[1] A. Delorme, S. Thorpe, "Face recognition using one spike per neuron: resistance to image degradation", *Neural Network in press,* 2001.

[2] D. Hubel, T. Wiesel, "Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat", *J. Neurophysiol.* 28, 229–289, 1965.

[3] Z. Li, "A neural model of contour integration in the primary visual cortex", *Neural Comput,* *10*(4), 903-40, 1998.

[4] C. von der Malsburg "The What and Why of Binding: The Modeler's Perspective", *Neuron,* vol. 24, 95-104, Sept., 1999.

[5] S. Thorpe, D. Fize, and C. Marlot, "Speed of processing in the human visual system", *Nature,* *381*(6582), 520-522, 1996.

[6] S.J. Thorpe, J. Gautrais, "Rank order coding", In J. Bower, *Computational neuroscience: Trends in research, 1998* (pp. 113-118). New-York: Plenum Press.