ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 187-192

Improving object recognition by using a visual latency mechanism

R. Opara and F. Wörgötter
Dept. of Neurophysiology, Ruhr-Universität, 44780 Bochum, Germany

Abstract

For a consistent analysis of a visual scene the different features of an individual object have to be recognized as belonging together and separated from other objects. In the brain of higher vertebrates it has been suggested that this is achieved by synchronization of the activity of disjunct nerve cell assemblies. However cross-talk between spatially adjacent image parts occurs preventing efficient synchronization. As a consequence image segmentation can be heavily impaired in network models. In these cases, temporal differences, naturally introduced by stimulus latencies in every biological sensory system, can strongly improve the performance of the network.

1 Introduction

The segmentation of a visual scene is a fundamential process of early vision, where elementry features are grouped together into discrete objects and objects are segregated from each other and the background. Image segmentation by synchronized cell activity requires that a minimal phase difference exists between cells belonging to the same assembly, while the phase difference to cells of another assembly should be much larger. The temporal separation of two objects achieved by such a mechanism, however, cannot exceed more than one half-cycle of the underlying oscillation period, which is about 10-30 ms in the visual cortex [2]. Retinal latencies are contrast dependent and can span much larger ranges of 20-120 ms [4],[5],[6], measured after a significant proportion of the response has built up (e.g., 4th spike, [5]). The visual latencies in the cortex [7],[8] can, therefore, be used for a contrast dependent grouping of the neuronal activity, leading to a larger temporal separation of the cell assemblies, which strongly improves undisturbed internal synchronization.

The system we present combines a conventional synchronization algorithms with a conceptionally new algorithm that utilizes latency differences to enhance the speed of synchronization for a cell assembly which represent one feature.

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 187-192

2 The Model

The network consists of nine layers (Fig. 1). Neurons of each layer are arranged in a three dimensional grid (for simplicity only the two-dimensional case is shown in figure 1). The central column of Fig. 1 shows the connectivity pattern of the multi-layer neural network. The architecture is related to the basic anatomy of the primary visual pathway, however, largely omitting the spatial structure of real receptive fields.

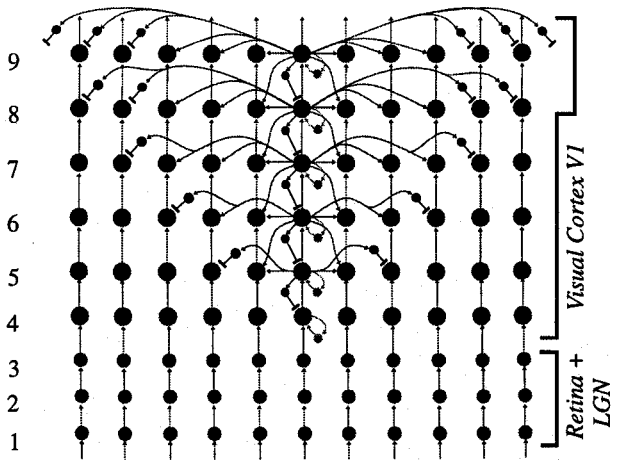

Figure 1: Schematic diagram of the network. Lateral connections are only drawn for the central column.

The membrane potential U for all layers l is described by the following differential equation (δ= Kronecker function):

$$\frac{dU(x,y,l)}{dt} = \frac{\delta_{1,l} I(x,y)+1}{\tau}[1 - U(x,y,l)] + \sum_{x',y',l'} V_\epsilon(x-x', y-y', l-l', l); \quad if\ U < \Theta$$

$$U = 0; \qquad else.$$

This is a low-pass differential equation with a time constant that in the first layer depends on the image intensity I. The first part of the equation describes the oscillator dynamics, while the second part is a spatial coupling term which is incorporating excitatory ($V_\epsilon > 0$) and inhibitory ($V_\epsilon < 0$) connections. If a neuron at position $p'=(x', y', l')$ fires, the membran potential U of a neuron at $p = (x, y, l)$ is shifted by V_ϵ. Since U describes a concave function, synchronization occurs necessarily as soon as V_ϵ is unequal zero [12]. The coupling V_ϵ leads to an excitatory interaction within layers over short distances while inhibitory coupling is predominate at long distances. The coupling strength decays exponentially with increasing distance. Θ is the firing threshold. If the membrane potential U

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 187-192

reaches the threshold Θ a spike is generated and U is reset to zero.

In the network two functionally different areas can be distinguished. The first area is mainly responsible for the latency differences. In Fig. 1 it is labeled as *Retina and LGN*. The second part labeled *V1* causes fast synchronization and feature binding, which is achieved by standard circuitry [9].

2.1 Latency Layer

In the visual system latencies arise due to complex electrochemical processes, with which we are not concerned. Instead, at the input stage we use frequency coding, that depends on feature contrast. To excite neurons in the next layer long integration times (2-3 cycles) are used. As a result neurons with a high frequency (high features contrast) at the input stage need less time (short latency) to excite cells in layer two and three than neurons with a low frequency (low feature contrast). This effectively mimics the different latencies of a real retinal signal.

Only sparse connectivity is used at this stage, to ensure a fast information flow to *V1* with less interference between different cell assemblies.

2.2 Synchronization Layer

The next stage synchronizes the neurons, and therefore binds the features. In this area no additional latencies are introduced. The propagation time through the layers is constant and neurons of a cell assembly, representing a certain feature, can group their activity fast. The grouping is undisturbed, because there is no interference between objects with different contrast due to their different latencies.

Layers in *V1* have four different types of connections.

- Direct feedback inhibition between layers, to uncouple them from the input activity.
- Lateral excitatory feedback connections to speed up the propagation of lagging activity, which is caused by noise in feature intensity or latency times. The radius of lateral excitation on the same layer increases in the upper layers. This makes it possible that increasingly larger regions are involved in the synchronization process.
- Lateral excitatory connections between neuron at neighbouring location.
- Lateral inhibitory connections over a large distance.

3 Results

As an example Fig 2 shows the activity and phase differences of the network response to a stimulus (inset) which consists of two square objects (α,β) with equal contrast and a third (γ) low-contrast object, all subject to 45% noise. The contrast difference between (α,β) and (γ) is 0.22 (enhanced in the graphic).

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 187-192

The phase difference $\Delta\Phi(x,y) = \Phi(x,y) - \overline{\Phi}$ of the current phase $\Phi(x,y)$ to the averaged phase $\overline{\Phi} := \sum_{x,y} \Phi(x,y)$ of each neuron is coded as grey level. As a consequence, for a *single* simultaniously active cell assembly, the phase difference $\Delta\Phi(x,y)$ is equal to zero (in Fig. 2 coded with middle grey). On the other hand, if two cell assemblies are firing with maximal phase difference, one cell assembly is coded as black while the other assembly is coded as light grey.

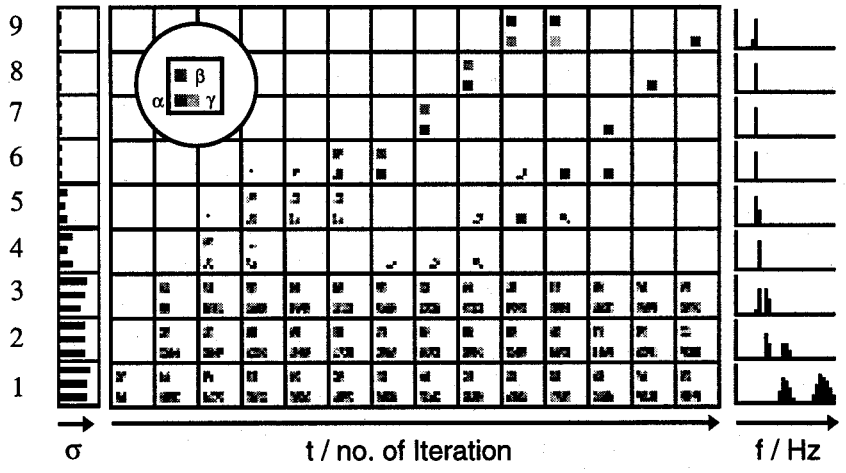

Figure 2: Left: Horizontal bars indicate the standard deviation of the firing phase σ of the neurons that belong to one object (top=α, middle=β, bottom=γ) averaged for the complete run. Standard deviation drops to zero in the higher layers. Middle: Time course of the activity distribution. Pixel grey values represent the firing phase. Above layer three two objects (α,β) are separated in *time* from the third (γ) which has a longer latency. Neurons, which represent one object, acquire synchrony (i.e., constant *phase* which is indicated by constant grey values) in the upper layers and phases are different for different objects. Right: Average firing frequency for all cells in arbitrary units. The frequency is higher in lower layers with two peaks representing the two objects and constant at a single lower value in the model cortex.

In the first three layers the lack of lateral interactions leads to random phase difference $\Delta\Phi$ between all neurons.

Due to the latency mechanism, features with a high feature contrast against the background (α and β in Fig. 2) need less time to propagate through the first layers (retina and LGN) than features with a low contrast (γ in Fig. 2). Features with a high contrast difference are separated across the model layers after a few iterations (column 2 in Fig. 2). This spatial separation of the activity permits an independent processing of features with different contrast, leading to less interference between the features after the first processing stages. As a consequence the speed of the synchronization between neurons coding the same feature is strongly enhanced and in the higher layers individual features are bound together to form a consistent representation of the objects in the visual scene.

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 187-192

The response strength remains unaffected by the latency mechanism. Thus, even contrast gain control mechanisms could be applied at higher levels without interfering with the already existing temporal separation. In addition, more complex spatio-temporal filter operations (e.g., oriented receptive fields) could be introduced after the initial latency dependent separation, which would allow for a more realistic feature detection (e.g., orientation discrimination, [10]).

4 Performance

The performance of the network is determed in a series of simulations, in which the principal parameters, contrast and latency, are varied. The performance test shows the quality of object segmentation in layer 9 measured by the firing coherence rate between two adjacent objects. A high coherence rate of 1 indicates a strong binding between the cell assemblies, which implies, that they belong to one object. On the other hand two objects are indicated by a low coherence rate. Only in a range of 30% up to 70% coherence the interference between the cells is ambiguous. In such a case the system cannot clearly decide if the cell assemblies are belonging to one or two objects. The goal for the system is to minimize this ambiguity range.

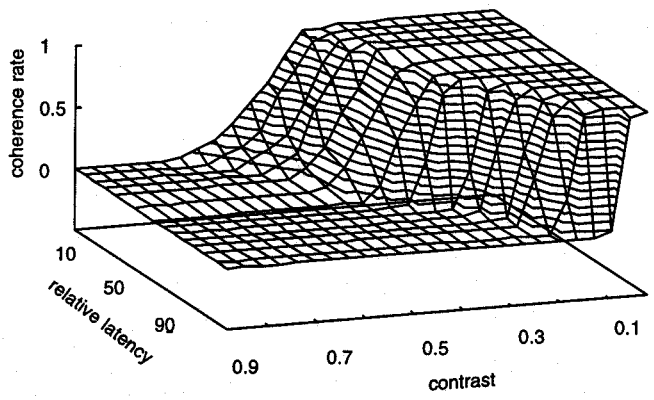

Figure 3: Coherence rate between two cell assemblies representing two adjacent objects as a function of contrast and latency.

In Fig. 3 the coherence rate is shown as a function of contrast and latency. At long latencies the range of ambiguity is small because transition from area 1 to area 2 is steep. Down to shorter latencies the quality of separation decreases. This is indicated by a larger transition range.

5 Discussion

In our conceptual framework the combined application of several physiologically inspired aspects of real neural networks (i.e., oscillation, synchronization) to-

gether with the newly introduced latency mechanism simultaneously achieves efficient object binding and image segmentation. To this end we were only concerned with the separation of image parts by contrast. The enhancement of assembly formation achieved as the result of the latency mechanism can be used as a first but highly efficient step in image analysis problems. It is very likely that these originally formed cell assemblies provide not more than the basis for an immediately following regrouping by other mechanisms to analyze more complicated visual problems (e.g., objects with a graded shading). While we did not attempt this, it is clear that the model provides no restrictions for a possible rearrangement of the activity, which could for instance be achieved by feedback from higher visual areas [11].

References

[1] Gray, C.M., König, P., Engel, A.K. & Singer, W. Oscillatory responses in cat visual cortex exhibit inter-columnar sychronization which reflects global stimulus properties. Nature, 338, 334-337 (1989).

[2] Eckhorn, R., Frien, A., Bauer, R., Woelbern, T. & Kehr, H. High frequency (60-90 Hz) oscillations in primary visual cortex of awake monkey. NeuroReport, 4, 243-246 (1993).

[3] Sompolinsky, H., Golomb, D. and Kleinfeld, D. Global processing of visual stimuli in a neural network of coupled oscillators. Proc. Natl. Acad. Sci. USA, 87, 7200-7204 (1990).

[4] Levick, W.R. Variation in the response latency of cat retinal ganglion cells. Vision Res., 13, 837-853 (1973).

[5] Bolz, J., Rosner, G. and Wässle, Response latency of brisk-sustained (X) and brisk-transient (Y) cells in the cat retina. H. J. Physiol, 328, 171-190, (1982).

[6] Sestokas, A.K., Lehmkuhle, S. and Kratz, K.E. Visual latency of ganglion X- and Y-cells: a comparison with geniculate X- and Y-cells. Vision Res., 27, 1399-1408 (1987).

[7] Raiguel, S.E., Lagae, L. Gulyàs, B. and Orban, G. Response latencies of visual cells in macaque areas V1, V2 and V5. Brain Res. 493, 155-159 (1989).

[8] Ikeda, H and Wright, M.J. Retinotopic distribution, visual latency and orientation tuning of sustained and transient cortical neurones in area 17 of the cat. Exp. Brain Res., 22, 385-398 (1975).

[9] v.d.Malsburg, C. and Buhmann, J. Sensory segmentation with coupled neural oscillators. Biol. Cybern., 67, 233-242 (1992).

[10] Wörgötter, F. and Koch, C. J. A detailed model of the primary visual pathway in the cat: comparison of afferent excitatory and intracortical inhibitory connection schemes for orientation selectivity. Neurophysiol. 11, 1959-1979 (1991).

[11] Tononi, G., Sporns, O. and Edelman, G. Reentry and the problem of integrating multiple cortical areas: simulation of dynamic integration in the visual system. Cerebral Cortex, 2, 310-335 (1992).

[12] Ernst, U., Pawelzik, K. and Geisel, T. Multiple phase clustering of globally pulse coupled neurons with delay. In M. Marinaro & P.G.Morasso (eds.), *ICANN '94: Proceedings of the International Conference on Artificial Neural Networks, Volume I*, (Springer, London, 1994), p. 1063-1066.

Acknowledgements

The authors acknowledges the support of the Deutsche Forschungsgemeinschaft.

ESANN'1995 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 19-20-21 April 1995, D-Facto public., ISBN 2-9600049-3-0, pp. 187-192