A Novel Algorithm for Image Segmentation Using Time Dependent Interaction Probabilities

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 and the background. Classical algorithms to segment a visual scene have an implicit representation of the image in the connection structure. We propose a new model that uses an image representation in the time domain, operating on stimulus dependent latencies. Such stimulus dependent temporal differences are observed in biological sensory systems. In our system they will be used to define the interaction probability between the different image parts. The gradually changing pattern of active image parts will thereby lead to the assignment of the different labels to different regions which leads to the segmentation of the scene.

1 Introduction

The segmentation of a visual scene is a fundamental process of early vision, where elementary features are grouped together into discrete objects and objects are segregated from each other and the background. In the brain of the higher vertebrates it has been suggested that this could be achieved by synchronization between cells [3], [5], [7]. It has also been supposed that temporal differences of neuronal signals could play an important role in the perception of higher vertebrates [2], [6], [8].

In this paper we present an labeling algorithm that utilizes stimulus dependent temporal differences (latencies) to segment visual scenes. This temporal structure is the only representation of the image in our system.

We will first give an overview of the used labeling model. Then we will show an example how the system segments a visual scene containing a shaded surface, a square and a disk. Finally we will discuss the results.

2 The Model

For the task of image segmentation a labeling algorithm is used which is based on the interaction of labels in order to minimize the energy of the system as given in Eq. 1. The energy function is not related to observables of a physical system, but only a quantity to describe the interaction of labels. The dynamics of the system will tend to assign the same label to units representing one object and a different label to units representing another object.

The units are arranged on a two dimensional lattice of size $N = N_x x N_y$. Each unit $i = (i_x, i_y)$ can take k different labels $\sigma_i \in \{1..k\}$.

The energy of the system is depending on the label configuration and the connection strength between units.

$$E_{tot}(t) = \sum_{i=1}^{N} \sum_{\substack{j \\ \|j-i\| < d_{max}}} -w_{i,j}\delta_{\sigma_i\sigma_j} P^B_{i,j}(t) \quad (1)$$

N	: number of units at the two dimensional lattice ($N = N_x x N_y$),
σ_i,σ_j	: labels of unit i and j,
$w_{i,j}$	: connection strength between unit at location i and j,
$P^B_{i,j}(t)$	: time dependent binary coupling term $P^B_{i,j} \in \{0,1\}$ which indicates, if two units are interacting at anyone point in time or not,
$\delta_{\sigma_i\sigma_j}$	: kronecker function,
$\|j-i\|$	: distance of unit i and j,
d_{max}	: maximal allowed range of interaction.

In classical label algorithms for image segmentation [4] the connection strength $w_{ij} = w_{ji}$ between unit i and j defines the similarity between two locations i and j. In our approach we will show that it is sufficient to have a constant connection strength for all units, while the probability for an interaction P^B_{ij} is given in the time domain.

Thus, the coupling term $P^B_{ij}(t)$ adds an additional time dependence to the dynamics of the system. At first each unit is assigned a time dependent probability described by

$$P_i(t) = \alpha \exp(-\frac{(t-t_i)^2}{s_i^2}). \quad (2)$$

t_i is the characteristic latency of unit i given as a function of the contrast C of pixel i ($t_i \to \infty,\ if\ C \to 0$). The probability for an interaction P^I_{ij} of unit i and j is given by the product of the corresponding probabilities P_i (Eq. 2) of unit i and P_j of unit j. For a given point in time we have $P^I_{ij}(t) = P_i(t)P_j(t)$. Finally we restrict the approach to binary interactions $P^B_{ij} \in \{0,1\}$. The probability that a binary interaction at time t is actually taking place ($P^B_{ij} = 1$) is given by $P^I_{i,j}(t)$. In Fig. 1 the interaction probability is shown as a function of time for three units having different characteristic latencies.

Two units (i, j) have rather similar latencies while the third unit (k) has a much longer latency. The probability of interaction $P^I(t)$ at time t_0 is much larger for unit i and j as compared to unit i and k (Fig. 1). The effect of the time dependent interaction term is that units representing bright objects are responding earlier than units which represent darker objects. This yields to a temporal separation of the different parts of the input image. In Fig. 2 it is shown that the temporal structure by itself is sufficient to segment a visual scene (the connection strengths w_{ij} are constant for all units).

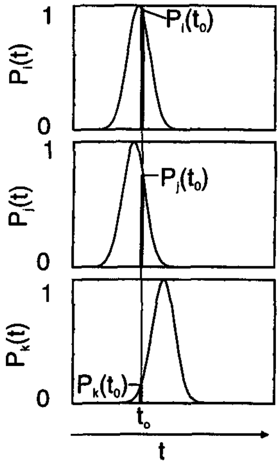

Fig. 1. The probability P for the interaction of three units i,j and k, as a function of time.

3 Results

As an example Fig. 2b shows the label distribution of the units if a stimulus (Fig. 2a) is given to the system. The stimulus consists of two objects with uniform distributed gray values g a square ($\overline{g}_{square}$=240) and a disk $\overline{g}_{disk}$ = 40), a shaded surface (g_{min} = 60, g_{max} = 220) and a background ($\overline{g}_{background}$ = 128).
The simulation starts with a random label configuration thus each label is represented by nearly the same number of units (Fig. 2c).
During the first iterations only the most salient objects (low latency) are processed (square and bright parts of the shaded surface) (Fig. 2b iteration 10 and 350 and Fig. 2d). According to the dynamics of the system the units of the square will receive the same label. This is indicated in Fig. 2c. The number of assigned units to the label *square* is increased after a few iterations. The plateau for the square (Fig. 2c) in the beginning, however, also contains units from other regions which by chance have the *square* label. These will eventually be removed and the plateau reaches its final level. Due to the latency the parts with a lower luminance are not processed at this time (Fig. 2d).
At later times the case is quite different. Now the bright objects are processed with a small probability, while the darker image parts are processed with a high probability (Fig. 2b, 2d).
One remarkable effect occures after $\sim$ 350 iterations (Fig. 2), where the probability distributions of the shaded surface and the background are overlapping. In spite of the overlapping probability distributions, the background and the shaded surface are segmented with different labels, except for a small part at the border of background and shaded surface. The assignment of different labels to the background and the shaded surface occures due to the history of the

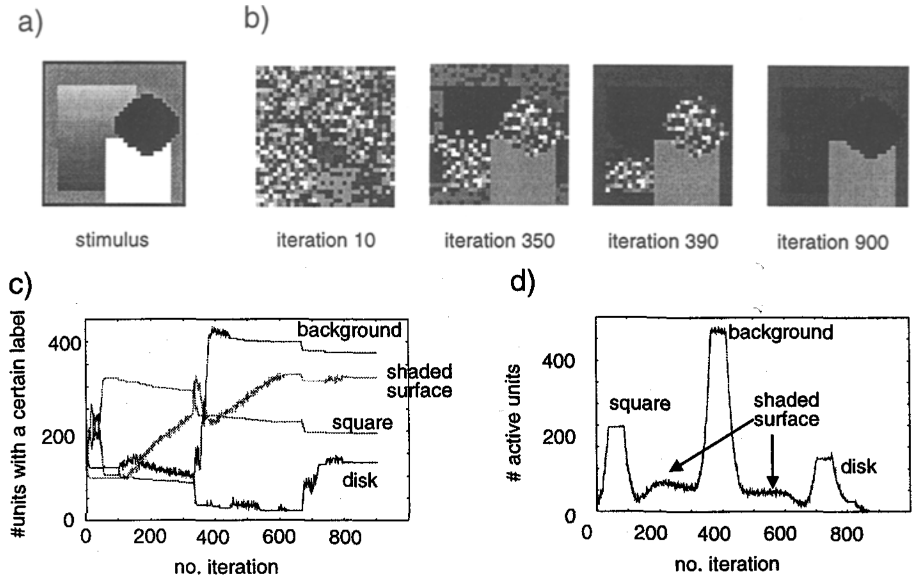

Fig. 2. a: The stimulus consists of two objects with a uniform distributed gray value, a shaded surface and the background b: Label distribution of the system at different times. The labels are coded as gray values. Units with the same label (same gray value) belong to the one object, while different gray levels indicates the assignment to different objects. c: The number of units assigned to a certain label. d: The number of units which interact at a given point in time.

processing of the objects and the small overlapping area.
In standard algorithms for image segmentation [4] such smooth transitions between two objects are also hard to segment, because at this area the similarity between the objects is nearly the same as within an object.

4 Discussion

In our conceptual framework we show that image segmentation can effectively be realised in the time domain. Therefore no implicit representation of the image in the connection structure is needed. Features with a high contrast are thereby favored and will be processed earlier than objects that do not "jump to the eye". This mechanism thereby mimics human perception [1] but more importantly it efficiently limits the information flow which needs to be evaluated at any point in time.
To this end we were only concerned with the separation of image parts by contrast. With only a few restrictions other features (color, spatial frequency etc.) could also be included in the system. The integration of other features would lead to a more robust segmentation.

Acknowledgements

The authors acknowledges the support of the Deutsche Forschungsgemeinschaft (grant WO388 4-2, 5-2).

References

1. Burgi, P. Y. and Pun, T. Figure-ground separation: evidence for asynchronous processing in visual perception? *Perception* **20**, 69, (1991).
2. McClurkin, J. W., Optican, L. M., Richmond, B. J. and Gawne, T. Concurrent processing and complexity of temporally encoded neural messages in visual perception. *Science* **253**, 675-677, (1991).
3. 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).
4. Geman, D., Geman, S., Graffigne, C. & Dong, P. Boundary detection by constrained optimization. *IEEE Trans. Pattern Analysis Machine Intelligence*, **12**(7), 609-628, (1990).
5. Gray, C.M., König, P., Engel, A.K. & Singer, W. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, **338**, 334-337 (1989).
6. Hopfield, J. J. Pattern recognition computation using action potential timing for stimulus representation. *Nature*, **376**, 33-36, (1995).
7. v.d. Malsburg, C. The correlation theory of brain function. Int. report 81-2, Dept. of Neurobiol. Max-Planck-Institute for Biophysical Chemistry, Göttingen, (1981).
8. Wörgötter, F., Opara, R., Funke, K. & Eysel, U. Utilizing latency for object recognition in real and artificial neural networks. *NeuroReport*, **7**, 741-744 (1996).