Image Point Correspondences from a Wavelet Representation and a Hierarchical Dynamic Link Network

Rolf P. Würtz and Christoph von der Malsburg

Ruhr-Universität Bochum
Institut für Neuroinformatik
D-44780 Bochum, Germany

Phone: +49 234 700-7996; Fax: +49 234 7094-210
email: rolf@neuroinformatik.ruhr-uni-bochum.de

1 Introduction

Conventional neural networks try to solve the problem of object recognition in one step — the network learns the known examples and builds up a classifier. The result is usually coded as the activity of a single or a small set of cells. We take a different approach and assume that recognition is an active process involving neuronal dynamics and fast synaptic plasticity driven by activity correlation. This scheme has been proposed by von der Malsburg [5, 6, 4] under the name "dynamic link architecture". Dynamic links are synaptic connections that, in contrast to neuronal learning schemes, have time constants of the same order of magnitude as the neuronal dynamics themselves These dynamics rapidly converge to an ordered link constellation, which in itself codes for the percept or the intermediate results to be read off by other subsystems in the brain.

The basic problem for visual object recognition under realistic circumstances is a solution of the *Correspondence Problem*:

> *Given two images of the same object, decide which pairs of points correspond to the same point on the physical object. The fact that some points do not have corresponding partners in the other image must also be established.*

This is problematic because local features do not have unique counterparts in the other image. These ambiguities can only be resolved by their relative geometrical arrangement. In the dynamic link architecture this is achieved by neuronal layers that are wired such that a localized region of activity (blob) moves across the layers. The growth of dynamical links is governed by a combination of correlated layer activity and feature similarity.

In this system the sorting out of the feature ambiguities is partly sequential and requires an unacceptably long processing time for layers of realistic size. We therefore present a pyramidal image representation and a dynamics well adapted to this representation. It consists of a pair of layers for every level of the pyramid and solves the correspondence problem by proceding from coarse to fine scales.

A further novel property of the system described here is the possibility to find correct correspondences in the presence of structured background. Due to the very limited space many details of the system have to be omitted. They can be found in [8].

2 Representation of Images and Models

The processing of a retinal grey-level image I in the primary visual cortex can be modeled by a wavelet transform based on complex-valued Gabor functions with an extra term that removes their DC-component:

$$\begin{aligned}(\mathcal{W}I)\left(\vec{k},\vec{x_0}\right) &:= \left(\psi_{\vec{k}} * I\right)(\vec{x_0}) \;, \\ \psi_{\vec{k}}\left(\vec{x}\right) &= \frac{\vec{k}^2}{\sigma^2}\exp\left(-\frac{\vec{k}^2\vec{x}^2}{2\sigma^2}\right)\left[\exp\left(i\vec{k}\vec{x}\right)-\exp\left(-\sigma^2/2\right)\right] .\end{aligned}$$

The single wavelet is parameterized by its spatial frequency $\vec{k}$, a two-dimensional vector described by length and orientation. The responses of all spatial frequencies of some fixed length form a frequency level, which assigns a small feature vector to all image points

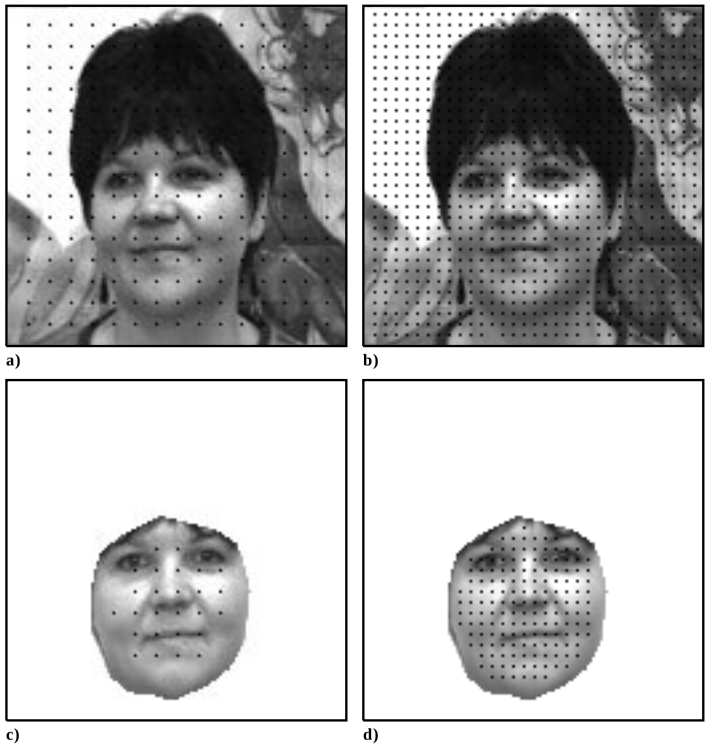

Figure 1: The location of the layer neurons. The black spots mark the location of the layer neurons in image and model, respectively. **a)** and **b)** show the image representation on level 0 and 1, **c)** and **d)** the model representation. The layers are rectangular, those neurons without a location in the model representation are part of the layer dynamics but do not make or receive dynamic links.

on an appropriate sampling grid (see figure 1). The components of the feature vectors correspond to the various orientations of the spatial frequency.

This pyramidal arrangement has the advantage that all responses which are influenced by the background can be discarded (see figure 1 c) and d)). The stored model (or prototype) is segmented and its representation contains only the responses of Gabor functions whose receptive fields fall completely inside the segmented area. The image representation consists of a full pyramid.

3 Hierarchical Dynamic Link Matching

For the matching dynamics each frequency level of image and model is assigned a pair of neuronal layers. Each pair of layers is interconnected reciprocally by dynamic links. The layer dynamics have the general form:

$$\begin{aligned} \tau_a \frac{d}{dt} a(\vec{x}) &= -a(\vec{x}) + c_\kappa \left(\kappa(\vec{x}) - c_g\right) * \vartheta(a(\vec{x})) + c_c - c_h h(\vec{x}) + c_s s(\vec{x}) + c_\xi \xi \\ \frac{d}{dt} h(\vec{x}) &= \begin{cases} \tau_{s+}^{-1} \left(a(\vec{x}) - h(\vec{x})\right) & : \quad a(\vec{x}) > 0 \\ \tau_{s-}^{-1} \left(a(\vec{x}) - h(\vec{x})\right) & : \quad a(\vec{x}) \leq 0 \end{cases} \end{aligned}$$

On the lowest frequency level the decision must be made which part of the image matches the stored model best. Both layers are wired with short-range excitation and global inhibition. In the presence of noise this supports a stable state with only one

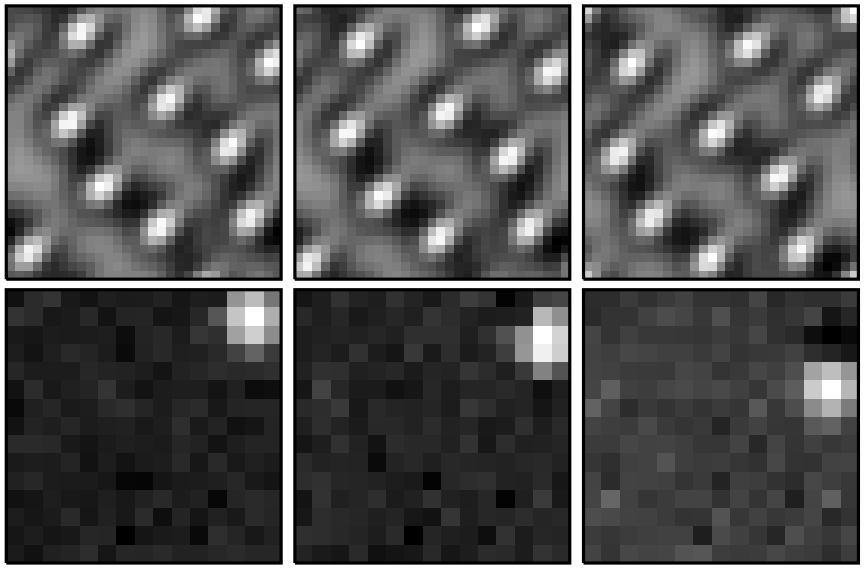

Figure 2: Layer dynamics on level 1 and 0. As a visualization of the dynamic activity on layer this figure shows three snapshots of a moving blob spaced by 25 simulation time steps (layer 0, below) and three snapshots of the multiple blobs spaced by 4 time steps (layer 1, above).

connected activity region (blob) [1]. A system of delayed self-inhibition ($h(\vec{x})$) is used to make this blob move across the layer.

The dynamic links between the layers are initialized to the feature similarities. They grow with a rate proportional to the feature similarity plus the product of the output values of the pair of neurons they connect. (This will be referred to as correlation and denoted as $\mathrm{Corr}(\vec{x}, \vec{y})$.) The growth is constrained by thresholds for the total strength of outgoing and of incoming links, respectively. The link dynamics take the form:

$$\begin{aligned} \tau_W \frac{d}{dt} W(\vec{x}, \vec{y}) &= W(\vec{x}, \vec{y}) \mathrm{Corr}(\vec{x}, \vec{y}) \,, \\ & \int W(\vec{x}, \vec{y}) d^2x \leq 1 \,, \\ & \int W(\vec{x}, \vec{y}) d^2y \leq 1 \,. \end{aligned}$$

In the beginning both blobs move freely and independently on their corresponding layers (see figure 2, lower part). Correlations make some links between the layers grow, others decay. After some time, the links have become strong enough that the image blob can only exist inside the region which corresponds to the model. From then on, the blob decays outside this region after a while and spontaneously reforms inside the region. When the links have grown even stronger, the image blob does not leave the region any more, and the correct links grow until a one-one mapping has been reached.

The layer dynamics on the higher levels have mexican-hat-type interaction with a kernel whose maximum is slightly off center. With appropriate parameters these dynamics support a structure of many small blobs moving coherently across the layer. The delayed self-inhibition is not active on the higher levels ($c_h = 0$) See the upper part of figure 2 for a visualization of these dynamics.

The link dynamics on a higher level are triggered once some link on the previous level has reached a threshold. Their growth rates have the same form as on the lowest level, with an extra term that supports only the connections that have strong links on the previous layer.

On the basis of images of human faces these dynamics can establish rough correspondences on the lowest level, which are then refined on the higher ones. This has been demonstrated by simulation of the first two levels. See figure 3 for the development of the links on the lowest and the next level.

The multitude of blobs on the higher levels allows a partly parallel refinement of the correspondences estimated on lower ones. As the processing time increases with the layer size this system is potentially faster than others that use only one pair of layers [3, 7, 2], although direct comparisons have not been carried out yet.

4 Recognition

In order to demonstrate the capability for object recognition the time consuming simulation of the dynamics has been replaced by a hierarchical template matching scheme. This was successful in the recognition of human faces independently of hairstyle (background) from a database containing 83 different persons. The background independence makes the system proposed here superior to an earlier system described in [4]. This is not part of this poster but underpins the usefulfulness of the hierarchical dynamics described here (see [8] for details).

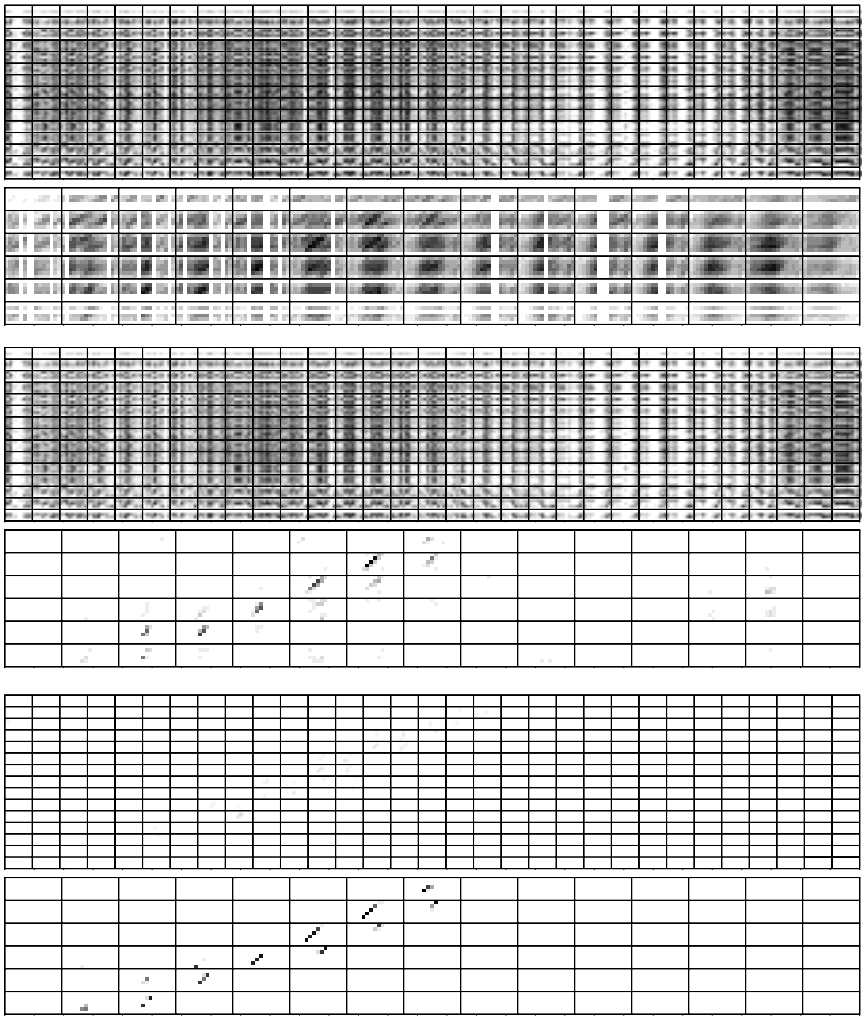

Figure 3: **The development of the dynamic links.** Each little rectangle contains the link strenghts between one horizontal scanline in model and image, respectively. Ideal correspondences (e.g. for identical images) would show no links besides black diagonals in the rectangles belonging to corresponding lines. In the beginning (top two figures) the links reflect only the feature similarities, which are highly ambiguous. After 390 time steps the lower level has sorted out the correct correspondences, and the first links have grown above the threshold where they are allowed to influence the higher frequency level. In the bottom figures (a snapshot after 1000 time steps) the links on both frequency levels are restrained to the correct correspondences, the remaining ambiguities are due to the coarse sampling.

5 Discussion

Concerning the main questions of the workshop this poster supports the view that for working cognitive systems both representation and dynamics must be appropriate, and these questions are not completely separable. The dynamic link architecture offers a framework within which there is sufficient variability to build systems which can model cognitive capabilities.

In this poster a dynamical system has been presented which solves the correspondence problem in a hierarchical fashion and thus cuts down the processing times. The system works well in the presence of slight distortions *and* structured background.

There are many open questions concerning the properties of such networks. The only parts that are understood fairly well analytically are the growth of a single blob and the multiple blobs on the higher layers. The link dynamics with stationary neuronal activity can be described in terms of evolution equations. Any ideas that the participants of the workshop may have about a possible analytical treatment of the combined blob/link dynamics will be highly welcome.

References

[1] S. Amari. Dynamical stability of formation of cortical maps. In M.A. Arbib and S. Amari, editors, *Dynamic Interactions in Neural Networks: Models and Data*. Springer, 1989.

[2] W. Konen and J.C. Vorbrüggen. Applying dynamic link matching to object recognition in real world images. In S. Gielen, editor, *Proceedings of the International Conference on Artificial Neural Networks*. North-Holland, Amsterdam, 1993.

[3] Wolfgang Konen, Thomas Maurer, and Christoph von der Malsburg. A fast dynamic link matching algorithm for invariant pattern recognition. *Neural Networks*, 7(6/7):1019–1030, 1994.

[4] Martin Lades, Jan C. Vorbrüggen, Joachim Buhmann, Jörg Lange, Christoph von der Malsburg, Rolf P. Würtz, and Wolfgang Konen. Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*, 42(3):300–311, 1993.

[5] Christoph von der Malsburg. The correlation theory of brain function. Technical report, Max-Planck-Institute for Biophysical Chemistry, Postfach 2841, Göttingen, FRG, 1981. Reprinted 1994 in: Schulten, K., van Hemmen, H.J. (eds.), *Models of Neural Networks*, Vol. 2, Springer.

[6] Christoph von der Malsburg. Pattern recognition by labeled graph matching. *Neural Networks*, 1:141–148, 1988.

[7] Laurenz Wiskott and Christoph von der Malsburg. Dynamic link matching with running blobs. In preparation, 1994.

[8] Rolf P. Würtz. *Multilayer Dynamic Link Networks for Establishing Image Point Correspondences and Visual Object Recognition*. PhD thesis, Institut für Neuroinformatik, Ruhr-Universität Bochum, D-44780 Bochum, Germany, 1994. Submitted to the physics department in October 1994.