

Rapid Correspondence Finding in Networks of Cortical Columns

Jörg Lücke^{1,2} and Christoph von der Malsburg^{2,3,4}

¹ Gatsby Computational Neuroscience Unit, UCL, London WC1N 3AR, UK

² Institut für Neuroinformatik, Ruhr-Universität, 44780 Bochum, Germany

³ Frankfurt Institute for Advanced Studies, 60438 Frankfurt a. M., Germany

⁴ Computer Science Dept., University of Southern California, LA 90089-2520, USA

Abstract. We describe a neural network able to rapidly establish correspondence between neural fields. The network is based on a cortical columnar model described earlier. It realizes dynamic links with the help of specialized columns that evaluate similarities between the activity distributions of local feature cell populations, are subject to a topology constraint, and gate the transfer of feature information between the neural fields. Correspondence finding requires little time (estimated to 10-40 ms in physiological terms) and is robust to noise in feature signals.

1 Introduction

For various purposes it is necessary for the brain to find point-to-point correspondences between structured neural arrays. Among these is stereo vision and visual motion extraction. There are good reasons to assume that the brain also performs *correspondence-based invariant object recognition* [1]. In the technical domain, this represents state-of-the-art object and face recognition technology [2, 3].

Objects can be recognized in less than 100ms, see e.g. [4], and after each saccade a new system of correspondences needs to be established, making it clear that the mechanism must be very fast. A previous neural model of correspondence-based recognition [5] had problems with the evaluation of feature similarity and with speed. The model [6] is fast but did not attempt to cope with different feature types.

2 Columnar Network Model

The central element of our model is the *cortical column*. As discussed in [7] our column corresponds approximately to what in neuroscience is called a hypercolumn or macrocolumn and in primary visual cortex comprises all neurons that are activated from one point in visual space. A column contains sub-units called *minicolumns* or simply *units* that comprise on the order of one hundred neurons which are connected by mutual excitation. The activity of a unit is described collectively by a variable p , and the units of one column mutually inhibit each

other. The coupling coefficient ν of this inhibition is cyclically driven (ν -cycle), such that when ν is low all units are on, and when ν approaches a critical value some units switch off in sequence, thus reflecting the relative strengths of their afferent input. The dynamics of columns is described in the next section.

A simple model setting for the process of correspondence finding (see Fig. 1B) consists of an input domain \mathcal{I} , left column of large shaded ellipses, and a model domain \mathcal{M} , right column. Both domains consist of neural sheets that represent images by the activity distribution of fields of local feature detectors. (Correspondingly they should be two-dimensional, but for simplicity we limit ourselves here to one-dimensional chains. And the model domain should contain many such sheets to represent objects in memory, but for the time being we focus on just one.)

In each point of the two domains there are two columns, one to represent local features (horizontal ellipses within the shaded regions of Fig. 1B) and one to control links between the two domains (vertical ellipses). This double column (shaded ellipse in Fig. 1B) is called a *node*. *Feature columns* represent, with their activity, the local texture of the image or model, usually represented by units that are excited by different local spatial frequencies and orientations of the image's gray-level distribution. Typical feature distributions as used on our simulations are shown in Fig. 1A, where each row corresponds to one feature column, index i , and each column to one feature type, index α .

The domains communicate through *links*, which connect feature columns by as many fibers as there are feature types. In a *link control column* each unit stands for one link entering the node and does three things. One, it compares the activity distributions of the feature columns at the two ends of the link, two, it tries to be consistent with activities of units controlling parallel links, and three, by its activity it keeps open its link. The situation is shown in more detail in Fig. 2. At the end of a ν -cycle, when only one control unit is left active, all but one of the links into the node are switched off. This unit or link is selected by a combination of two criteria. One is feature similarity, the other is a topology constraint. The latter is to favor those link arrangements that connect neighbors in one domain with neighbors in the other domain, and is implemented by connections between control units in neighboring nodes (symbolized at the extreme right of Fig. 1B).

3 System Dynamics

The dynamics of the system is described by a set of coupled stochastic differential equations. We first introduce some notation. Let $\mathcal{L} \in \{\mathcal{I}, \mathcal{M}\}$ and $\mathcal{L}' \in \{\mathcal{I}, \mathcal{M}\} \setminus \{\mathcal{L}\}$ be indices for the two domains, i.e., $(\mathcal{L}, \mathcal{L}') = (\mathcal{I}, \mathcal{M})$ or $(\mathcal{M}, \mathcal{I})$. Further, let $p_\alpha^{\mathcal{L}i}$ stand for the activity of the feature unit α in node i of domain \mathcal{L} . We assume α runs from 1 to k and i from 1 to N . Let us designate by $W^{\mathcal{L}i, \mathcal{L}'j}$ the activity of the control unit with index j in node i of domain \mathcal{L} (each control column must contain as many control units as there are nodes in the other domain, in order to control as many links). As introduced and discussed

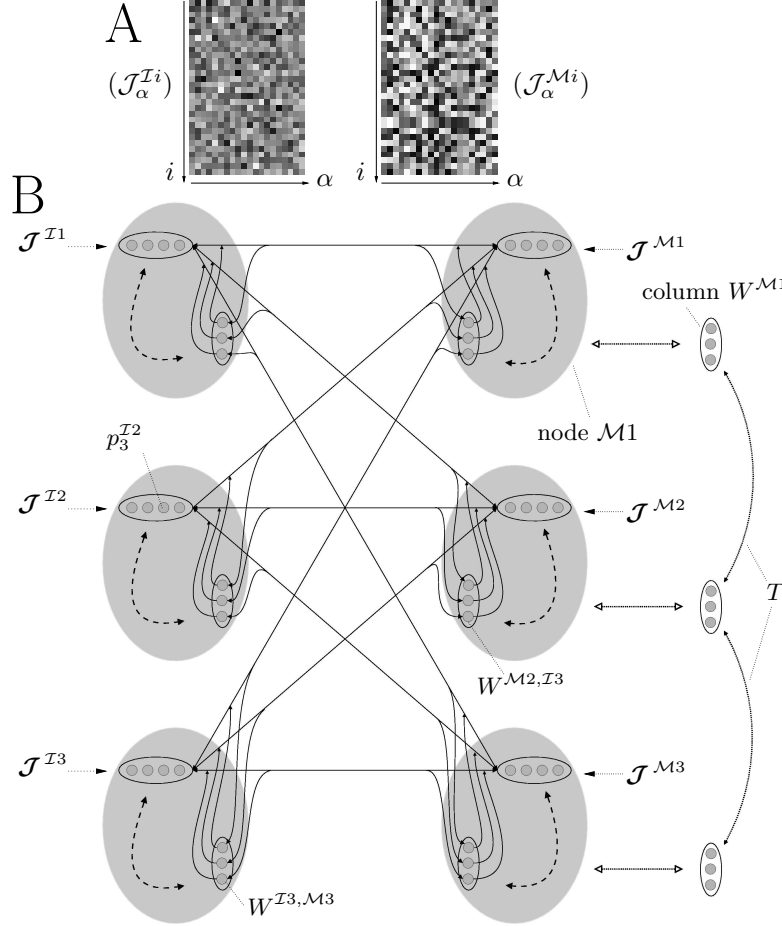


Fig. 1. **A** A collection of feature vectors (rows) with $k = 20$ entries. Model feature vectors on the right-hand-side are iid and uniformly distributed in $[0, 1]$. The input feature vectors on the lhs are noisy copies of the vectors on the rhs. **B** Network of columns for correspondence finding. The network consists of an input domain and a model domain with nodes $\mathcal{I}1$ to $\mathcal{I}3$ and $\mathcal{M}1$ to $\mathcal{M}3$, respectively. Each node consists of a feature column with $k = 4$ minicolumns and of a control column with $N = 3$ minicolumns. Each node in the input layer receives input from each node in the model layer, and *vice versa*. The inputs to a node are modulated by its control column according to the interconnectivity as displayed in Fig. 2. The control columns receive input from the units of feature columns of both layers and from neighboring control columns.

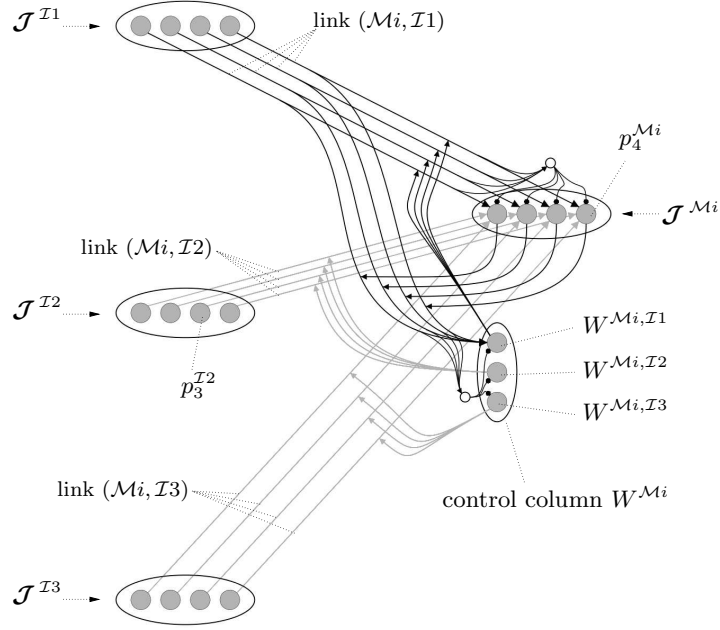


Fig. 2. Detailed connectivity of one node. On the input side three feature columns $\mathcal{I}1$, $\mathcal{I}2$ and $\mathcal{I}3$ are shown, which by their activity distribution represent three input feature vectors. On the model side, one node $\mathcal{M}i$ consisting of feature and control column is shown. For one control unit, $W^{\mathcal{M}i, \mathcal{I}j}$, the connection details are indicated by bold lines. On its input side, the unit evaluates the similarity of feature vectors in terms of their scalar product (multiplicative interactions indicated by arrowheads touching connecting fibers) and with its output gates the incoming link it stands for. Connectivity with neighboring control columns is not shown. Small circles represent neurons which by their inhibition subtract the mean of incoming feature vectors.

in [7], the dynamics of the feature columns is then described by

$$\frac{d}{dt} p_{\alpha}^{\mathcal{L}i} = f(p_{\alpha}^{\mathcal{L}i}, \nu \max_{\beta=1, \dots, k} \{p_{\beta}^{\mathcal{L}i}\}) + \kappa E_{\alpha}^{\mathcal{L}i}, \quad (1)$$

where $E_{\alpha}^{\mathcal{L}i}$ is input to the unit, controlled in its strength by parameter κ , and where

$$f(p, h) = a p (p - h - p^2) + \sigma \eta_t \quad (2)$$

is a control function in the form of a polynomial of third degree, including the Gaussian noise term $\sigma \eta_t$ with variance σ^2 . The inhibition is determined by the most active unit in the column, modulated by the inhibitory coefficient ν , which,

as stated above, is controlled cyclically:

$$\nu(t) = \begin{cases} 0 & \text{if } \tilde{t} < T_{\text{init}} \\ (\nu_{\max} - \nu_{\min}) \frac{\tilde{t} - T_{\text{init}}}{T - T_{\text{init}}} + \nu_{\min} & \text{if } \tilde{t} \geq T_{\text{init}} \end{cases}, \quad (3)$$

where $\tilde{t} = t \bmod T$, which is $t - nT$, with n the greatest integer satisfying $t - nT \geq 0$.

To specify the input $E_{\alpha}^{\mathcal{L}i}$ in (1) we first have to define a few quantities. The feature input to feature unit α in node $\mathcal{L}i$ is designated as $\tilde{\mathcal{J}}_{\alpha}^{\mathcal{L}i} = \mathcal{J}_{\alpha}^{\mathcal{L}i} - \frac{1}{k} \sum_{\beta=1}^k \mathcal{J}_{\beta}^{\mathcal{L}i}$, where the mean is subtracted from the raw feature inputs. The momentary coupling strength from node j in domain \mathcal{L}' to node i in domain \mathcal{L} is set equal to the mean-free activity of the control unit of that link, $\tilde{W}^{\mathcal{L}i, \mathcal{L}'j} = W^{\mathcal{L}i, \mathcal{L}'j} - \frac{1}{N} \sum_{l=1}^N W^{\mathcal{L}i, \mathcal{L}'l}$. We then define the input into feature unit $p_{\alpha}^{\mathcal{L}i}$ in (1) as

$$E_{\alpha}^{\mathcal{L}i} = C_E \tilde{\mathcal{J}}_{\alpha}^{\mathcal{L}i} + (1 - C_E) \sum_{j=1}^N \sum_{\beta=1}^k \tilde{W}^{\mathcal{L}i, \mathcal{L}'j} R_{\alpha\beta}^{\mathcal{L}i, \mathcal{L}'j} p_{\beta}^{\mathcal{L}'j}, \quad (4)$$

where the parameter $C_E \in [0, 1]$ controls the relative strength of the two sources of input. The matrix $R_{\alpha\beta}^{\mathcal{L}i, \mathcal{L}'j}$ defines feature-preserving interconnections between feature columns in the two domains: $R_{\alpha\beta}^{\mathcal{L}i, \mathcal{L}'j} = \delta_{\alpha\beta} - \frac{1}{k}$. Here finally are the dynamic equations of the control units:

$$\frac{d}{dt} W^{\mathcal{L}i, \mathcal{L}'j} = f(W^{\mathcal{L}i, \mathcal{L}'j}, \nu \max_{l=1, \dots, N} \{W^{\mathcal{L}i, \mathcal{L}'l}\}) + \kappa I^{\mathcal{L}i, \mathcal{L}'j}, \quad (5)$$

$$I^{\mathcal{L}i, \mathcal{L}'j} = C_I \sum_{\alpha, \beta=1}^k p_{\alpha}^{\mathcal{L}i} R_{\alpha\beta}^{\mathcal{L}i, \mathcal{L}'j} p_{\beta}^{\mathcal{L}'j} + \overbrace{(1 - C_I) \sum_{a, b=1}^N T_{ab}^{\mathcal{L}i, \mathcal{L}'j} \tilde{W}^{\mathcal{L}a, \mathcal{L}'b}}^{\text{topology term}}, \quad (6)$$

where $C_I \in [0, 1]$ controls the relative influence of the two terms in (6). The first term evaluates feature similarity. It resembles a scalar product with Euclidean metric between the activity vectors $\mathbf{p}^{\mathcal{L}i}$ and $\mathbf{p}^{\mathcal{L}'j}$ (other choices of $R^{\mathcal{L}i, \mathcal{L}'j}$ would correspond to other metrics). However, the situation is somewhat more complicated, as the activities of feature units do reflect feature values more in terms of the timing of their switching off in the course of the ν -cycle (later for stronger values) than by their firing strength at any moment.

The topology term in (6) implements link-to-link interactions. With vanishing topology term, $C_I = 1$, dynamics (1) to (6) would converge to a one-to-one connectivity that connected the most similar feature vectors in the model and input domains. Unfortunately it turns out that if there are non-trivial differences between model and image of the same object many nodes find their most similar feature vector in non-corresponding points of the other domain [8]. To remedy this problem, system dynamics should favor link arrangements that preserve

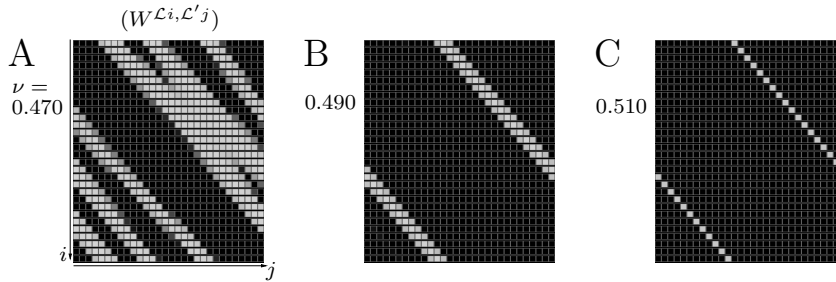


Fig. 3. Time course of network activities ($W^{\mathcal{L}_i, \mathcal{L}'_j}$) during a ν -cycle for $C_I = 0$. In a ν -cycle, ν increases from 0.4 to 0.52 in about 25ms. In the beginning all minicolumns ($W^{\mathcal{L}_i, \mathcal{L}'_j}$) of all columns are equally active (light grey). **A** For $0.45 < \nu \leq 0.47$ minicolumns start to be deactivated. **B** Because of the special choice of $(T_{ab}^{\mathcal{L}_i, \mathcal{L}'_j})$, diagonally arranged minicolumns are exciting each other and survive the increasing inhibition longer. Note that diagonals in $(W^{\mathcal{L}_i, \mathcal{L}'_j})$ correspond to neighborhood-preserving connectivity patterns between input and model domain. **C** Finally just one minicolumn per control column survives and the connectivity matrix ($W^{\mathcal{L}_i, \mathcal{L}'_j}$) is a shifted diagonal.

C_I . We simulate dynamics (1) to (7) with $C_E = 0.6$, so that feature columns are slightly more sensitive to their own feature vector than to input from the other layer, and with $C_I = 0.5$, giving equal weight to feature similarities and neighborhood relationships in the control of links. As input and model we use feature vectors ($\mathcal{J}^{\mathcal{I}^i}$) and ($\mathcal{J}^{\mathcal{M}^i}$) as given in Fig. 1A. The input feature vectors are noisy versions of the model feature vectors⁵. In Fig. 4 the result of a simulation with these feature vectors and parameters is shown for one ν -cycle. As can be observed, the dynamics converges to a symmetric one-to-one connectivity pattern between input and model layer. For visualization purposes we have chosen input feature vectors that were not translated w.r.t. the model feature vectors. For translated input feature vectors ($\mathcal{J}^{\mathcal{I}^i} = (\mathcal{J}^{\mathcal{I}^{i+\text{const}}})$) (respecting the cyclic boundary conditions) the system converges the corresponding shifted diagonal. For input generated as above the system reliably finds the right correspondences for noise levels up to about $\sigma = 0.6$ which shows a remarkably high noise tolerance. For $k > 20$ results improve and for $k < 20$ the error rate

⁵ We use $N = 30$ model and $N = 30$ feature vectors, each with $k = 20$ entries. Correspondingly, the numbers of nodes per layer is N and the number of minicolumns per feature column is k . The model feature vectors consist of randomly ordered copies of 10 different feature vectors whose entries contain equally, identically, and independently distributed random values between zero and one. An input feature vector ($\mathcal{J}^{\mathcal{I}^i}$) is generated from the model vector by adding Gaussian white noise with $\sigma = 0.6$ to the values ($\mathcal{J}^{\mathcal{M}^i}$). Subsequently, the set of all values ($\mathcal{J}^{\mathcal{I}^i}$) is rescaled such that all feature vector entries lie in the interval $[0, 1]$ again. The resulting image ($\mathcal{J}^{\mathcal{I}^i}$) has on average smaller component deviations from the mean, due to the rescaling after adding noise.

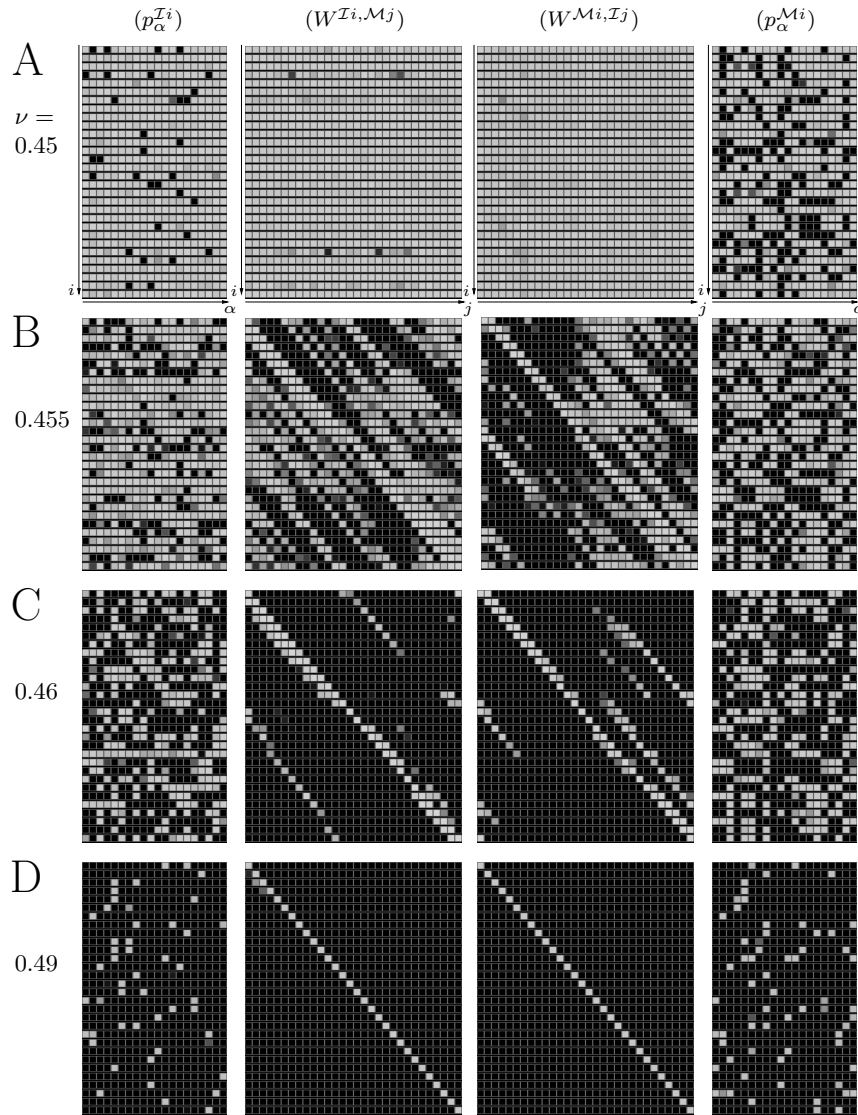


Fig. 4. Time course of the dynamical variables of the system displayed in Fig. 1B if feature vectors Fig. 1A are used. During a ν -cycle, ν increases from 0.4 to 0.52 in about 25ms. In the beginning of the ν -cycle all minicolumns are active (high gray values) and start deactivating at about $\nu = 0.45$ (see **A**). Note that the feature columns start first to deactivate their minicolumns because the input they get from their feature vectors is patterned. Feature vectors on the model side have inputs of higher variance and deactivate their minicolumns earlier. **B-C** Minicolumns of the control columns are deactivated according to similarities in feature columns and activities of other control columns. **D** Finally, control minicolumns remain active that correspond to diagonals in $(W^{Ii, Mj})$ and $(W^{Mi, Ij})$. The system has found the right correspondences as a neighborhood preserving mapping between similar features.

increases. Note in this context that in our technical applications feature vectors typically have $k \geq 40$ entries [2, 3] and that cortical columns in primary sensory areas are estimated to contain about $k = 80$ minicolumns [10].

On the basis of feature similarities alone, $C_I = 1$, a system with otherwise the same parameters and noise level $\sigma = 0.6$ converges to one-to-one connectivities which are not neighborhood preserving and in which 80 – 90% of the surviving links connect non-corresponding points.

5 Conclusion

Finding homomorphic, that is, structure-preserving, mappings between neural fields —the correspondence problem— is a capability of fundamental importance for the brain, not only for the visual system (stereo matching, motion field extraction) or perceptual systems in general (invariant pattern recognition), but more fundamentally for the application of abstract schemas to concrete situations and analogical thinking, and thus for intelligence on all levels. By its very nature, correspondence requires for its establishment and expression neural implementation media for the formulation of structural relationships and for the expression of dynamic links.

Both roles are played in our system by control columns, whose implementation turned out to be possible with fairly standard neurons. Our model describes minicolumn activity by abstract continuous variables, but as shown in previous work [9] this is capturing the essential properties of a more direct modeling of a system of spiking neurons [11]. Our model makes essential use of sigma-pi neurons, requiring sums of products of signals, *cf.* the second term in (4) and the first term in (6). Both cases involve control neurons, on the input side in one case, the output side in the other.

The activity of control columns and of feature columns is described here by the same type of stochastic differential equation (equations 1 and 5), but feature neurons and control neurons are probably of a different nature, the two types of columns playing very different roles. Control columns evaluate the similarity of local structure expressed by feature columns (this similarity being defined by the R -matrices in equations 4 and 6) and define, by interactions with each other, the homomorphy aspect of the correspondence (topology term). Feature columns, on the other hand, express local structure and are able to transmit it over distance. The handling of feature columns as integrated entity in the evaluation of similarities makes it possible to represent whole feature spaces, instead of single sample points in such spaces, as are represented by the combination-coding neurons that are conventionally used to represent higher features.

Our system solves several problems with previous models. One of them is the evaluation of feature similarities, which was a problem for [1], [6] and [5]. Another is excessive time requirement in [5]. As we demonstrate here, neural correspondence finding is possible in time-scales well below 100 ms because of the use of population rates. In fact, our simulations show that convergence to the right correspondences is possible within a critical period of a single ν -cycle of

a few tens of ms (25 ms for our simulations) which would correspond to gamma range oscillations. During the critical phase (see Fig. 4) neurons typically spike only few times (<10) as discussed in [11]. In the limit of short period lengths with still reliable convergences (≈ 10 ms) neurons have time to spike only 1 to 2 times in this period.

There are some challenges ahead of us. A full visual object recognition system will need a two-dimensional version and a model domain with many dozens of thousands of models. This threatens to require excessive numbers of control units. However, by using the maplet idea [1] and intermediate layers between the image domain and the model domain [6] this dragon could likely be tamed. Another, more formidable challenge concerns the ontogenetic development of the highly specific network structures involved in our model.

Acknowledgments. We thank J. D. Bouecke in helping us acquiring simulation results and gratefully acknowledge funding by the Gatsby Charitable Foundation, the Hertie Foundation, and by the EU project FP6-2005-015803.

References

1. J. Zhu and C. von der Malsburg. Maplets for correspondence-based object recognition. *Neural Networks, Special Issue 'New Developments in Self-Organizing Systems'*, 17/8-9:1311 – 1326, 2004.
2. P. Jonathan Philips, Hyeonjoon Moon, Syed A. Rizvi, and Patrick J. Rauss. The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000.
3. Kieron Messer et al. Face authentication test on the BANCA database. In *Proceedings of ICPR 2004, Cambridge*, volume 4, pages 523 – 532, 2004.
4. S. Thorpe, D. Fize, and C. Marlot. Speed of processing in the human visual system. *Nature*, 381:520–522, June 1996.
5. Laurenz Wiskott and Christoph von der Malsburg. Face recognition by dynamic link matching. In J. Sirosh, R. Miikkulainen, and Y. Choe, editors, *Lateral Interactions in the Cortex: Structure and Function*, chapter 4. WorldWideWeb, www.cs.utexas.edu/users/nn/book/bwdraft.html, 1995. ISBN 0-9647060-0-8.
6. B. A. Olshausen, C. H. Anderson, and D. C. Van Essen. A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *The Journal of Neuroscience*, 13(11):4700–4719, 1993.
7. J. Lücke. Dynamics of cortical columns – sensitive decision making. In *Proc. ICANN*, LNCS 3696, pages 25 – 30. Springer, 2005.
8. L. Wiskott. The role of topographical constraints in face recognition. *Pattern Recognition Letters*, 20(1):89–96, 1999.
9. J. Lücke and J. D. Bouecke. Dynamics of cortical columns – self-organization of receptive fields. In *Proc. ICANN*, LNCS 3696, pages 31 – 37. Springer, 2005.
10. V. B. Mountcastle. The columnar organization of the neocortex. *Brain*, 120:701 – 722, 1997.
11. J. Lücke and C. von der Malsburg. Rapid processing and unsupervised learning in a model of the cortical macrocolumn. *Neural Computation*, 16:501 – 533, 2004.