Flexibility through a Neural Architecture for Visual Orientation in a Natural Environment [1,2]

J. Kopecz

Zentrum für Neuroinformatik GmbH, D–44799 Bochum, Germany,

Email: joerg@neuroinformatik.ruhr-uni-bochum.de

C. Engels H. Janßen

Institut für Neuroinformatik, Ruhr–Universität Bochum ND–04, D–44780 Bochum, Germany,

Email: heja@neuroinformatik.ruhr-uni-bochum.de

Abstract: We propose a framework to model visual information processing, behavioral control on the base of two–dimensional activity distributions with local interactions. These layers are used to represent parameters of visual data and behavioral variables. Interactions within and among such representations defined as systems of linear or nonlinear differential equations support basic operations of neural information processing. The approach is exemplified through three different models in the domains of *visual representation and recognition*, *saccadic control* and *navigation in complex environments.*

Keywords:

Neural Architecture, Neural Fields, Natural Environment, Computer Vision, Active Vi-

[1] Published in Physica D 75, 209-224, 1994, based on a talk given at the Oji Symposium at Numazu, March 1993

[2] Partially supported by BMFT, Grant No. 413-5839-01IN101A6

sion, Robotics, Non-linear Dynamics, Two-dimensional Representations.

1. Introduction

We know from everyday's observations that animals are capable of solving a number of tasks in a complex and varying environment. In evolution their brains have undergone a constant optimization process. Starting from very simple organisms and very simple behavior nature has built an increasingly complex structure for information processing and the generation of behavior. The challenging issue - aside of the complexity and the sheer quantitiy [1] of these structures - is, that at no stage of evolution a complete redesign of the information processing structure was necessary: The solutions to a problem served as a preadaptation for a more complex one. Thus, the hierarchy of information processing is, to a certain extent, preserved in the structure of the brain. If we are able to identify such fundamental structures together with the processing tasks they solve, we can set up a neural instruction set for early visual processing [**?**,**?**].

One crucial problem of visual information processing and behavioral control is caused by the *limitations of sensory information* about the environment. One prominent example in vision is the (ill-posed) problem of the inversion of optics to acquire 3D-information. Therefore it is necessary for every system which wants to act in a natural environment to make strong assumptions: prejudice is more important for the generation of behavior than actual sensory input.

Another problem is the *wealth of sensory information* which even the brain seems to be incapable to process. Instead processing resources seem to be distributed specifically to task dependent information (*attention*).

[1] Following common estimation, the cortex has about 10^{11} neurons, with an average number of connections of 10^4 per neuron.

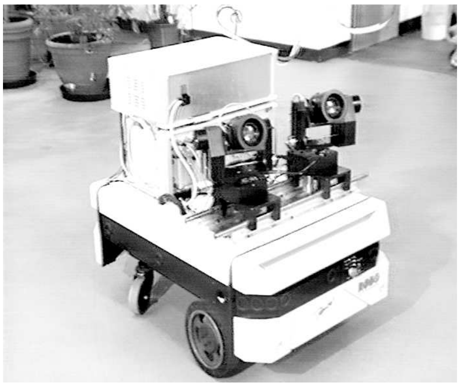

Fig. 1. Mobile Robot platform MARVIN that serves as a testbed for the neural architecture developed at the Institut für Neuroinformatik in Bochum. The camera system has the same degrees of freedom as our eyes (except for the rolling), they can be moved independently of each other with an angular speed of 180 Deg/sec and can perform saccades as well as smooth pursuit.

While we are interacting with our environment, the image on our retina (and thus the information coming to our brain) are changing. Therefore, visual processing cannot be separated from a behavioral task. To put it in a well known phrase: *We don't just see, we look.*

This concept leads us to paradigms like active vision, task related processing structures and to the fact, that perception and behavior cannot be separated from each other.

In the course of this paper we want to show on three examples of models inspired by these principles of biological information processing. We use two–dimensional activity distributions representing parameters of sensory data and behavioral control. Information processing is gained by defining local interactions on such layers in the form of linear or nonlinear differential equations. We think that this type of models can be used to integrate information from different sources and may define a common interface between processing steps or in a behavioral hierarchy.

We motivate our approach by using results from Neurophysiology, and –anatomy, as well as Psychophysics. Due to the large number of neurons in the cortex we feel that an approach resting on neural nets with discrete neurons is not powerful enough. We therefore use continuous formulations whenever possible, which gives us the advantage of simpler analytical description.

The usefulness of our models is checked by applying them to various problems arising in the domain of controlling an autonomous vehicle in a complex environment.

On a low level it should be able to process the necessary visual information, make decisions and plan its actions. This shall be accomplished by *emergent behavior* rather than by explicitly programmed behavioral sequences.

2. Early visual processing: Discrete Parametric Representations

The first step of each information processing system is to acquire the appropriate data. In modelling cortical hypercolumns [**?**], we extract local orientation information by means of two-dimensional Gabor functions that are used as filters. It is reported in the literature, that these functions model very closely the receptive field characteristics of simple cells [**?**,**?**] and display stable behavior for image processing tasks. In the context of this paper a hypercolumn is defined by the area covered by a number of cells having receptive fields sensitive to oriented stimuli ranging from 0 to 180 degrees. Mathematically this can be formulated as a three-dimensional structure consisting of two spatial coordinates for the approximate location of a stimulus on the retina and one coordinate denoting its orientation. Since all these variables vary in discrete steps, such a representation of images is called *discrete parametric representation* (DPR). The detailed description of the construction of such a DPR was given elsewhere [**?**,**?**] so that it is just summarized here:

Each image is divided into a number of overlapping segments at positions i, j on the image. In each segment oriented features are found by convolution with a set of K Gabor functions of different orientations. The process is sketched in Fig. 2. For each segment $S_{i,j}$ the signal energy is determined and taken as a hypercolumnar cell response to a feature of a given orientation.

Mathematically, this DPR is constructed out of a grey–scale image by the following operation:

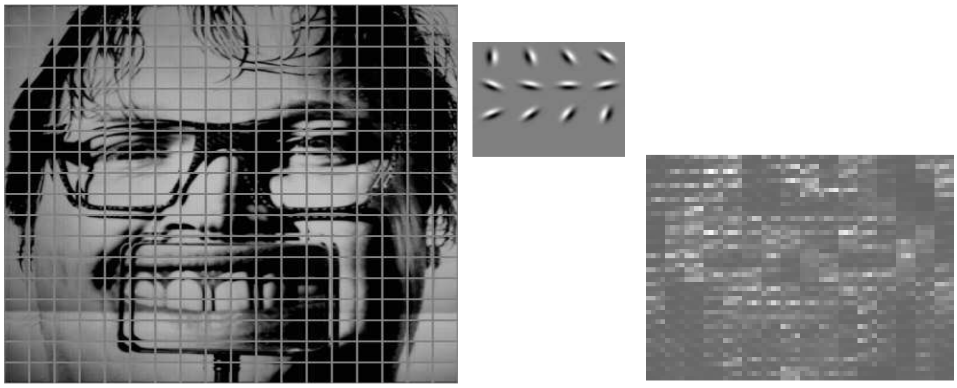

Fig. 2. Construction of a DPR from a camera image. The image is segmented into intervals each of which is analyzed by a set of Gabor filters. The coarse structure of the image is preserved but the exact information about location of features is lost.

$$\eta_{ijk}(x,y) = h_{i,j}(x,y) * g_k(x,y) \tag{1}$$

$$\xi_{u,v} = \mathcal{M} \sum_{(x,y)\in S_{ij}} \eta^2_{ijk}(x,y) \tag{2}$$

where

x, y	Pixel coordinates of grey-scale image
S_{ij}	Segment centered at i, j on the image
$h_{ij}(x,y)$	Image function within the segment S_{ij}
k	Orientation index of a Gabor filter: $\pi k/K = \arctan(\omega_y/\omega_x)$
$g_k(x,y)$	Set of two-dimensional Gabor functions with K different orientations (receptive fields)
$\mathcal{M}$	Mapping operator $\mathbb{R}^3 \mapsto \mathbb{R}^2$
$\xi_{u,v}$	Activity of a hypercolumnar cell at cortical position u, v, coding orientation k within a segment $S_{i,j}$

The operator $\mathcal{M}$ stands for the mapping of the inherently three-dimensional structure onto a two-dimensional cortex.

It is known from Neurophysiology, that the brain frequently "uses" this technique to achieve functional maps such as found in hypercolumns or ocular dominance stripes or binaural columns [**?**,**?**,**?**].

There are a number of algorithms known to develop such a mapping, (see e.g. [?,?]). Here, we use a simplification of a model proposed by Durbin and Mitchison [?] and used earlier by Durbin and Willshaw [?] as elastic net algorithm. It formulates an update rule for the reference vectors $\mathbf{w}$:

$$\Delta \mathbf{w_j} = \alpha f_j(\mathbf{y}, \mathbf{w})(\mathbf{y} - \mathbf{w_j}) + \beta \sum_{k \in N} (\mathbf{w_k} - \mathbf{w_j}) \tag{3}$$

$$\text{and} \qquad f_j(\mathbf{y}, \mathbf{w}) = \frac{F(\mathbf{y}, \mathbf{w_j})}{\sum_l F(\mathbf{y}, \mathbf{w_l})}$$

where $\mathbf{j}$ is the cortical position, $\mathbf{y}$ the stimulus vector in parameter space ($Y \subset \mathbb{R}^3$), $\mathbf{w_j}$ is the coordinate vector of post-synaptic neuron $\mathbf{j}$ and k are the nearest neighbours to $\mathbf{j}$. $F(\mathbf{y}, \mathbf{w_j})$ denotes the gaussian receptive field (RF) of neuron $\mathbf{j}$ and $f_j(\mathbf{y}, \mathbf{w})$ is the normalized RF of neuron $\mathbf{j}$.

Eq. (3) uses two assumptions that are biologically plausible: Firstly, a cortical map will match the stimulus distribution as closely as possible (See [?] for a more detailed explanation), and secondly, the cortical elements will always tend to be similar to their neighbours, as expressed by the second term in (3). During the iteration process (i.e. the learning) the receptive fields $f_j(\mathbf{y}, \mathbf{w_j})$ are constantly normalized, thereby modelling a restriction of post-synaptic resources.

The instabilities arising in dimension reduction processes where first considered by Amari [?]. According to this, an algorithm supposedly suited for the generation of functional maps has to display *transversal instability*, i.e. instability in its fix-points due to the existence of the excess dimension that is to be mapped [2].

In order to analyze the algorithm analytically with this respect, we formulate an averaged learning rate by performing two limits: The discrete positions j on the cortex are transformed into a continuous variable r. Considering the limit in j for the second term of (3) results in the second derivative or a diffusion term. Since we are only interested in fixed points, we can combine α and β to a diffusion constant D_v. It is further assumed,

[2] Kohonen's model [?] also displays this vital feature.

that each element of the signal space Y is mapped with equal probability so that we can average over all signal space by taking the integral. The normalization of the receptive fields is done after each update so that this enters the equation as a nonlinearity.

We then arrive at an equation of motion for the reference vector from the update rule Eq. (3)[?] [3]:

$$\dot{\mathbf{w}}(\mathbf{r}) = \frac{1}{C} \int_{\Lambda} \mathbf{dy} \left[\frac{\mathbf{f}(\|\mathbf{y} - \mathbf{w}(\mathbf{r})\|^{\mathbf{2}})}{\int \mathbf{dxf}(\|\mathbf{y} - \mathbf{w}(\mathbf{r})\|^{\mathbf{2}})} (\mathbf{y} - \mathbf{w}(\mathbf{r})) \right] + \mathbf{D_v} \Delta \mathbf{w}(\mathbf{r}) \tag{4}$$

with C some normalizing constant, D_v a diffusion constant of the information on the cortex surface, and Λ a subspace of $\mathbb{R}^3$. By means of 1. order perturbation theory of a trivial mapping ($\mathbf{w}(\mathbf{r}) = (r_1, r_2, 0)^T$) the eigenvalues can be calculated and analyzed for its stability. In the case of a mapping $\mathbb{R}^3 \mapsto \mathbb{R}^2$ three eigenvalues are obtained:

$$\begin{aligned} \lambda_1 &= -\|\mathbf{k}\|^2 (D_v + \sigma^2 \exp(-\sigma^2 \|\mathbf{k}\|^2)) \\ \lambda_2 &= -\|\mathbf{k}\|^2 D_v \\ \lambda_3 &= -\|\mathbf{k}\|^2 D_v + \frac{a^2}{3\sigma^2}(1 - \exp(-\sigma^2 \|\mathbf{k}\|^2)) \end{aligned} \tag{5}$$

With σ being the width of the receptive fields, a being the height of the excess dimension and $\mathbf{k}$ being the wave vector of a state. The third eigenvalue was identified as the one responsible for transversal instability and a condition for its becoming *positive* and therefore causing the desired instability was obtained. This is in contrast to the condition for reorganization of a map in the case of the Kohonen model, as obtained by Ritter [?], which is $a > 2.02\sigma$. We find a similar condition that is almost sufficient:

$$a > \sqrt{3}\sigma \tag{6}$$

the exact condition obtained from the model is

$$\frac{3D_v}{a^2} \left(\ln(\frac{3D_v}{a^2}) - 1 \right) + 1 - \frac{3\sigma^2}{a^2} > 0 \quad \text{and} \quad a^2 > 3D_v \tag{7}$$

Regarding the structure of the resulting map the maximally unstable eigenvalue determines the wavelength of the map. It turned out that this gives an ensemble of wave-vectors

[3] We thank K. Kurata for fruitful collaboration on this matter.

satisfying:

$$\|k\| = \frac{1}{\sigma}\sqrt{\ln\frac{a^2}{3D_v}} \tag{8}$$

The question now arises, how to find a parameter regime that produces maps suited for information processing. In order to tackle this problem, a selection criterion was imposed:

If local uniform operators are supposed to perform image processing on such a structure - e.g. cross orientation inhibition for the enhancement of continuous features in images or extracting subjective contours [**?**] - then there has to be a certain distance to travel from one orientation to the next orthogonal one on the cortex. Surprisingly, such a mean *cross orientation distance* does not exist for most of the possible mappings. Cross correlation analysis was carried out on the map. The correlation will be minimal for a pair of cells having orthogonal receptive fields. A given map is a "good map" if a fixed distance can be found for all cells on the modelled cortex. This distance also defines the size of a hypercolumn. By means of this technique and parameter variation in the mapping equation appropriate mappings as shown in Fig. 3 were created [4]. A second criterion is the information contents of the map: A map can be used for image processing only if all elements of the DPR are represented on it. Since Eq. (3) explicitly optimizes this in its first term (contrary to the Kohonen mapping), this condition is also well fulfilled.

It is worth noting that (3) gives a description of *continuous distributions of neurons* which yields a description of *neural fields* rather than discrete samples. Since the algorithm stems from the idea of elastic nets this may be viewed as an elastic field that serves optimal dimension reduction purposes (optimal in the sense mentioned above).

Such a map can be seen as a medium that is excited by some visual stimulus. Information processing on this two-dimensional medium can be modelled by distance dependent forces between the excitations at certain positions. Since little is known about the exact type of interaction on the cortex, we make a phenomenological model to produce an enhancement

[4] We thank W. Fellenz for numerical simulations.

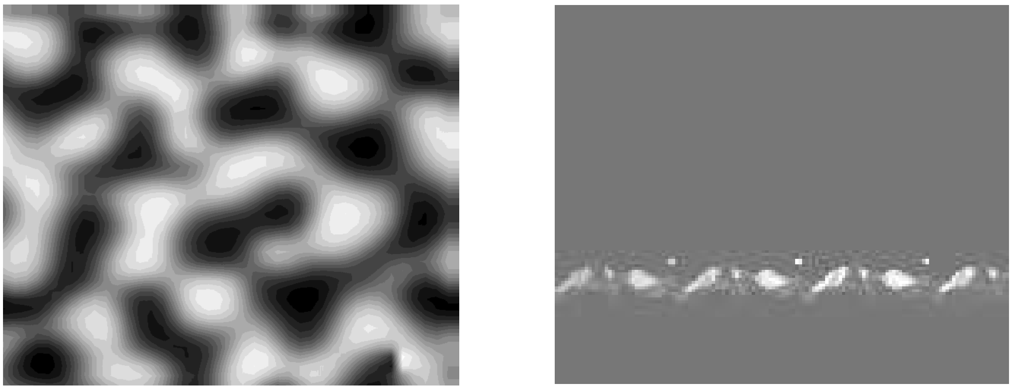

Fig. 3. **Left**: Distribution of orientation on the cortex surface of a hypercolumnar map. Each grey value codes a different orientation ranging from 0 to 180 deg. **Right**: Display of the activity distribution of a straight horizontal line on the modelled cortex. The activity distribution has to have "holes" in it, so that lateral inhibition does not destroy its information.

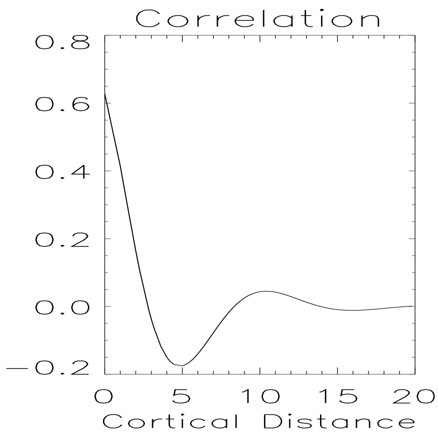

Fig. 4. Result of the cross correlation analysis of the map shown in fig. 3. The minimum of the correlation shows the mean distance of two cortical cells having orthogonally oriented receptive fields. The width of a hypercolumn being $\approx$ 10 cortical units of diameter. This computational value was closely matched by the theoretical prediction from Eq. (8).

of continuous features by a mechanism known as *cross orientation inhibition* [**?**]: This effect is usually ascribed to nonlinear characteristics of receptive fields but may also be seen as a result of interactions which cannot be discriminated from the perception process since they occur on the same time scale [**?**,**?**]. The dynamics chosen was described and motivated elsewhere [**?**,**?**] as coming from the selection dynamics of the hypercycle [**?**] and is just quoted for the case of a two-dimensional excitation field:

$$\dot{\xi}_{uv} = \iota_{uv} - \xi_{uv}\left[1 + A(d)\sum_{nm}\xi_{nm}^2\right] \tag{9}$$

where we have the following notations:

ξ_{uv}	activity on the cortex at position u, v
ι_{uv}	constant external force, either stimulus or memorized pattern
d	COR as measured in Fig. 4
$A(d)$	distance dependent amplitude, e.g, a delta function or DOG
$n, m \in U(uv)$	Surrounding to a position u, v containing the COR

The parameter ι can either be the (clamped) stimulus or a memorized pattern that was triggered by a visual input (see [?] for a discussion of the dynamics) [5]. Depending on this the resulting pattern and hence the final perception of the system is different.

Together with an associative memory that stores these patterns, we built an image recognition system that is running on MARVIN. It is part of a larger system for saccadic exploration of scenes as we will show in the next section.

3. Visual Scene Analysis by means of saccadic camera movements

Saccades, i.e. fast "jumps" of our eyes, are the most common kind of eye movement. The ability for fast shifts of our eyes corresponds to the space dependent receptor density of our retina with a small high resolution fovea (2–3 deg.) and a low resolution periphery (cone density relation approximately 1 : 10). Therefore the computationally expensive processing of visual information can be selectively distributed to arbitrary locations, *(saccadic scanpath)*, using the environment as a kind of external buffer for visual information [?].

Several influences to the selection of fixation targets have been reported. In general a

[5] It is worth noting that the role of the stimulus reflects the time scales on which the organization process of Eq. (4) and the interaction process (9) are working: In the first case, the stimulus is a fast varying variable over which we can average since the time-scale of the organization is very slow. In the latter the stimulus may be considered constant since cortical interaction is very fast (some 10 msec or shorter).

basic distinction [?] can be made:

– *Preattentive* saccades are made to moving patterns in the scene (a very strong effect) or to simple outstanding image features like color blobs or texture discontinuities. In those cases latency time does not depend on the complexity of the image - suggesting parallel processing in the brain.

– For *attentive* saccades some kind of sequential attentional process selects the targets *(overt attention)*, and eventually cognitive tasks determine the scanpath [?].

While quantitative properties of preattentive gaze selection are known [?,?], for the higher level "cognitive" phenomena so far only phenomenological models exist (e.g. [?], [?] for critical discussion).

Another crucial question is the interaction of different levels of saccadic control: how can preattentive and attentive/cognitive fixation demands be integrated ?

Our approach is to generate camera movements for a robot vehicle which are similar to the human saccades to obtain visual exploration and scene recognition. This way we hope to also understand more about the specifics and advantages of neural information processing which by now is still far more capable than any technical solutions in most areas.

The models discussed in this section define a "modular" system generating saccadic camera movements which resemble the movements of human eyes in the cases of "preattentive" exploration of a scene and recognition of already presented scenes [6]. Two activity distributions with inherent dynamics - called *interest map* here - are used to achieve behavioral control in the sense of competition and cooperation of points in the scene to gaze at next.

The whole system (Fig. 6) consists of the following parts:

[6] Part of this work was done in collaboration with G.–J. Giefing. Prelimanary studies for the problem of scene recognition were made together with B. Olbrecht.

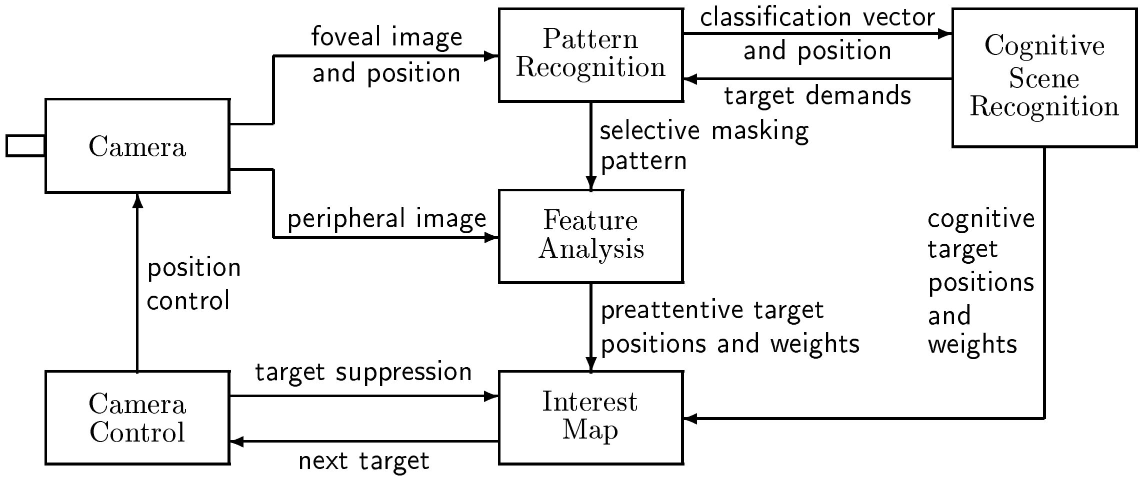

Fig. 5. System overview

– A preattentive peripheral feature detection extracts "salient" points in the periphery of the actually seen image - based on simple image measures like feature discontinuities, which can be weighted to allow for "selective attention".

– A multi resolution foveal image identification classifies the actual input image based on a hypercolumnar representation - this is a simple but efficient type of image recognition system [?].

– A scene recognition classifies by means of temporal integration of foveal pre-classifications. The underlying scene definition is a set of foveal views plus their spatial relationship (saccadic shifts). This has the advantage that the representation of the scene is expressed in terms of behavior resp. relevant information processing. The output is a "saliency-map" defining expected locations for other relevant parts of the scene; thus the scene recognition tries to verify its scene hypothesis by stimulating saccades.

– An egocentered interest map integrates the two types of fixation demands and selects the next fixation position.

Only the last part is of primary concern in this context, a discussion of the whole system and its incorporation in the behavioral context of our robot vehicle can be found in [**?**] resp. [**?**].

Input to the interest map are the three 2-dimensional excitation distributions supplied by the "preattentive" (movement detector and static feature detector) and "cognitive" (scene recognition) target demands. The goal is to give the simplest possible model with the following properties:

– A cooperation/competition mechanism between the different possible gaze targets must be installed selecting the location with maximal accumulated evidence.

– "Inhibition of return" must be incorporated, i.e. it must be assured, that the scanpath for a given scene does not simply oscillate between the two most salient positions.

– While it is questionable if the human saccadic system with its angle of view of nearly 180 degree uses any buffering, most current technical systems are limited to about 1/4 of that. With our hardware being limited to less than 30 degrees for the image diagonal, we definitely need some kind of "saliency memory" to be able to scan the environment.

– If one takes the accumulating error of saccadic shifts relative to a fixed (head centered) coordinate system into account, some kind of positional forgetting will inhibit false fixations.

The interest map is defined as a system of two linear differential equations:

– A *target map* $\psi(\mathbf{x}, t)$ [7] sums the inputs and defines a "leaky memory" function by a relaxation term. Diffusion compensates for the accumulating position error over time and models "positional forgetting".

– A *inhibition map* $\rho(\mathbf{x}, t)$ memorizes the locations already gazed at. Again relaxation and diffusion allow for effects of memory limitations.

[7] Note that $\mathbf{x}$ denotes a coordinate on a spherical surface. To model the human saccadic system, Listing coordinates would be most appropriate.

$$
\begin{aligned}
\frac{d\psi(\mathbf{x},t)}{dt} &= -\tau_\psi \psi(\mathbf{x},t) + D_\psi \Delta\psi(\mathbf{x},t) + I_d(\mathbf{x},t) + I_p(\mathbf{x},t) + I_r(\mathbf{x},t) \\
\frac{d\rho(\mathbf{x},t)}{dt} &= -\tau_\rho \rho(\mathbf{x},t) + D_\rho \Delta\rho(\mathbf{x},t) + I_f(\mathbf{x},t)
\end{aligned}
$$

$\psi(\mathbf{x},t)$,$\rho(\mathbf{x},t)$ target map, inhibition map

$I_d(\mathbf{x},t)$,$I_p(\mathbf{x},t)$ input from movement- and preattentive feature detection

$I_r(\mathbf{x},t)$ input from scene recognition

$I_f(\mathbf{x},t)$ input from camera control (actual fixation)

Note that $I_d(\mathbf{x},t)$ and $I_p(\mathbf{x},t)$ are unequal zero only in the relatively small area of the image, while the maps and eventually $I_r(\mathbf{x},t)$ are defined for the whole scene i.e. the area the sensors can be shifted to.

The next fixation position is determined by the position of maximum activity in the difference of the maps: $\mathbf{x}_{t_n} = \max_{\mathbf{x}} \Big(\psi(\mathbf{x},t_n) \ - \rho(\mathbf{x},t_n)\Big)$.

Given a completely static scene and therefore constant input to the maps, the system will generate a cyclic scanpath dependent on the excitation distribution of the input: Relatively high preattentive input over the whole scene will generate longer periodic sequences. The relaxation parameters τ_ψ and τ_ρ determine the "buffer function" for targets resp./ the tendency to avoid already fixated positions. The ratios of D_ψ and D_ρ relative to the fixation time adjust the interaction radius of different inputs resp. the minimal target distance.

In a pre-run the system was started "dumb" in the hall of our institute and explored the environment with ca. 50 saccades. This scene information (foveal patterns and saccadic shifts) were learned into memory. A typical intermediate state of the interest map during this exploration is shown in fig. 6.

In the main run, we put MARVIN with this information into the hall again but changed one part of the scene rather drastically (objects were shifted etc.). The starting behavior

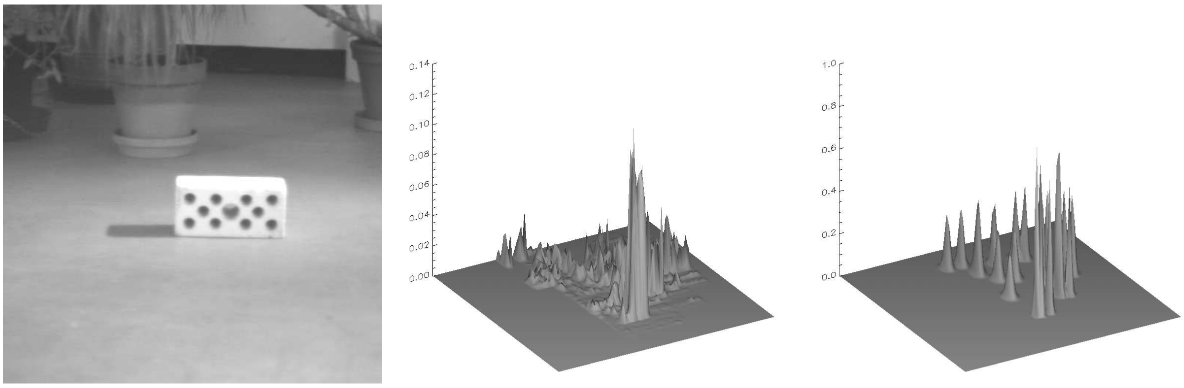

Fig. 6. Interest map after the 20th fixation: shown are the input image (left) and neural maps for target selection (mid) and inhibition of return (right). The time course of fixations can be easily seen from the inhibition map (starting in the middle, the going from top to bottom in a curve). In the target map the high contrast brick generates high priority saccade targets.

was of course exploration again, but after finding some already known foveal views the hypothesis was formed, that the scene was the one learned before. However, after continuing with this hypothesis, saccades to the changed parts of the scene correctly falsified the hypothesis again. The time course of recognition in the main run shows all principal shifts of behavior: sole exploration, scene hypothesis verification and falsification.

4. Local path planning in closed loop

Finally we want to provide an architecture for generation of complex behavior in the domain of autonomous mobile robots by use of *neural fields*[8], see [?,?]. A *neural field* represents parameters by localized neural activity, receives and sends weak external stimuli from respectively to other fields and performs nonlinear lateral interactions, see [?].
The behavioral task is to plan a trajectory of a vehicle in a simulated 2–D environment

[8] We are greatly indebted to Gregor Schöner for the guidance of our work and many valuable ideas. Furthermore we want to thank Michael Dose, Wolfram Erlhagen and Klaus Kopecz for fruitful discussions.

consisting out of goals and obstacles based on locally sensed information. Recent neurobiological findings in the motor cortex of the monkey give evidence that a population of neurons represents task related parameters. The sum of activities u_i of the individual neurons i weighted by the vector of their preferred parameters $\mathbf{c}_i$ is regarded as the population vector $\mathbf{p}$ which is representing actual parameter values in the space of $\mathbf{c}$.

$$\mathbf{p}(t) = \sum_i u_i(t)\mathbf{c_i} \tag{10}$$

Before a certain motor task is actually performed the population vector changes from the initial value $\mathbf{p}(t_0)$ to its final value $\mathbf{p}(t_1)$ which is solving the task. It is important to note that the population vector rotates continuously due to excitatory and inhibitory interactions among neurons [**?**]. Neurophysiological findings in the superior colliculus of the cat [**?**] imply that this rotation results from a continuous motion of the excitation of a localized population of neurons within the motor map. Further biological principles which are relevant to our approach are two–dimensional parametric maps organizing cortical areas and mappings between different cortical areas, see section 2. The biological results inspired the construction of our system in the following way:

(i) Points within two–dimensional maps represent a vector of the preferred parameter values $\mathbf{c}(x)$ within a topological order where $x \in \mathbb{R}$. Due to equation 10 the activity distribution $u(\mathbf{x})$ of these maps is coding a vector of actual parameter values $\mathbf{p}$. The activity distribution consists of either actual sensory $u_{\mathrm{s}}(\mathbf{x})$, memory $u_{\mathrm{m}}(\mathbf{x})$ or planning $u_{\mathrm{p}}(\mathbf{x})$ activity.

(ii) Intrinsic lateral dynamics forms localized excitation defined by positive activity within a bounded region of the map.

$$\tau\dot{u} = \mathcal{L}(u) \tag{11}$$

(iii) External input in form of mappings from n different maps

$$\tau\dot{u} = \mathcal{L}(u) + \sum_{i=1}^{n} \mathcal{M}_i(u_i) \tag{12}$$

defines attractors and repellors for a localized excitation in u. For stability reasons the dynamics receiving input has to act on a fast time-scale compared to the input dynamics

$\dot{u}_i \ll \dot{u}$, known as slaving principle.

For analytical tractability we choose the Amari dynamics for the formation of a localized excitation peak [**?**].

$$\tau du(\mathbf{x},t)/dt = \underbrace{-u(\mathbf{x},t) + \int w(\mathbf{x}-\mathbf{x}')f(u(\mathbf{x}',t))d\mathbf{x}' + h}_{\mathcal{L}(u)} + \underbrace{\xi(\mathbf{x},t)}_{\sum_{i=1}^{n}\mathcal{M}_{\mathrm{i}}u_{\mathrm{i}}} \tag{13}$$

where $w(\mathbf{x})$ denotes a long–range inhibition, short range–excitation connectivity distribution, $f(u)$ a step–function with zero–threshold, h a global inhibition parameter and $\xi(\mathbf{x})$ an external input. Amari showed that the condition for a localized excitation peak of width a in the one–dimensional case as a stationary stable solution is given by

$$W(a) - h = 0 \qquad \wedge \qquad dW(a)/da < 0 \tag{14}$$

where $W(a) = \int_0^a w(x)dx$. External input displaces the localized excitation peak solution with maximum at position z due to the equation of motion

$$\dot{z} = \frac{\epsilon}{\tau c}[\xi(z + a/2) - \xi(z - a/2)] \tag{15}$$

where c designates the slope at the boundaries of the localized excitation and ϵ indicates that the amount of external input is small compared to the localized excitation. Equilibria $\dot{z} = 0$ are stationary stable if $d\dot{z}/dx < 0$ is valid. These properties are applied to come to behavioral decisions by bifurcation and to stabilize control by hysteresis. Suppose we have an external input distribution like in figure 7b and we are modifying the distance d between the maxima then we obtain a pitchfork–bifurcation, see figure 7a. The stationary stable solution becomes bimodal which implies a decision among two possible positions of the localized excitation. Next if we are varying the amplitude A of the left maximum we obtain a hysteresis effect, see figure 7c, which stabilizes the persistence in one solution.

After laying out the general concepts we want to demonstrate the application in the path planning domain. The specific architecture consists out of four maps: The activity distributions u_{g} and u_{r} of the sensory maps are representing the directions ω and distances r to the targets respectively the obstacles in ego–centered coordinates and exhibit no

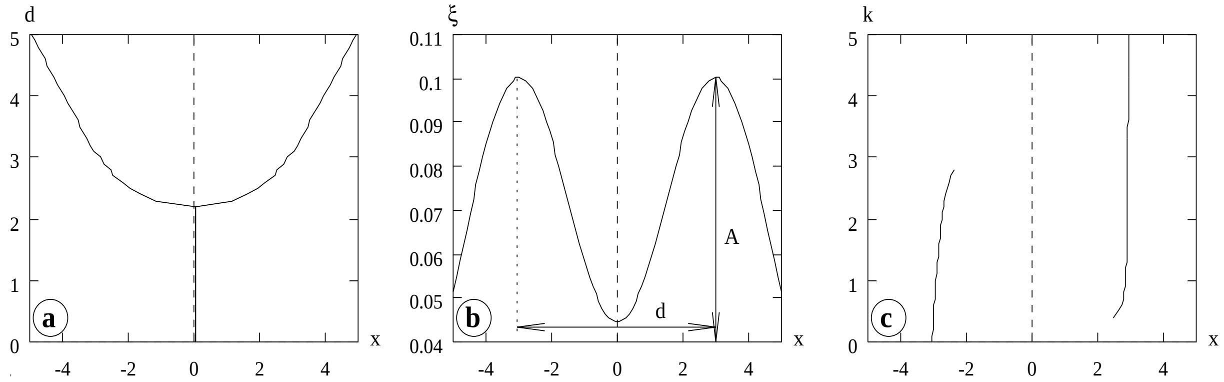

Fig. 7. (**a,c**) Stationary stable solutions depending on the distribution of an external input $\xi(x)$ (**b**) in the one–dimensional case. The abscissa indicates the spatial position x in the map. In diagram (**a**) the distance d between the two maxima of the input distribution is modified while figure (**c**) shows a variation of the amplitude A $= 0.1 *$ k of the right maximum of the input distribution. The width of localized excitation comes to $a = 4$.

lateral dynamical interaction. The distribution $u_{\rm m}$ of the global memory map is coding the representatives of obstacles in cartesian world–coordinates x, y and the distribution of the planning map $u_{\rm p}$ is contains the heading direction ω of the vehicle, see figure 8. The resulting dynamical system is described by

$$\tau_{\rm m}\dot{u}_{\rm m} = \mathcal{L}(u_{\rm m}) + \mathcal{M}_{\rm rm}(u_{\rm r}) \tag{16}$$

$$\tau_{\rm p}\dot{u}_{\rm p} = \mathcal{L}(u_{\rm p}) + \mathcal{M}_{\rm mp}(u_{\rm m}) + \mathcal{M}_{\rm gp}(u_{\rm g}) \tag{17}$$

where $\mathcal{M}_{ij}$ denotes a mapping from map u_i to u_j and the time constants are constrained by $\tau_{\rm m} \ll \tau_{\rm p}$. The coordinate transformations within $\mathcal{M}_{\rm rm}$ and $\mathcal{M}_{\rm mp}$ between the ego– and allo–centered coordinate frames require the position of the vehicle in world coordinates.

The closed loop simulation shows a planned path in a complex environment, see figure 9a. Note that there are several dead ends out of which the vehicle escapes unlike in most reactive potential methods [?]. In the part of the trajectory where the vehicle leaves the central box of obstacles the movement is made into the opposite direction as the target direction while it still finds its way to the target location.

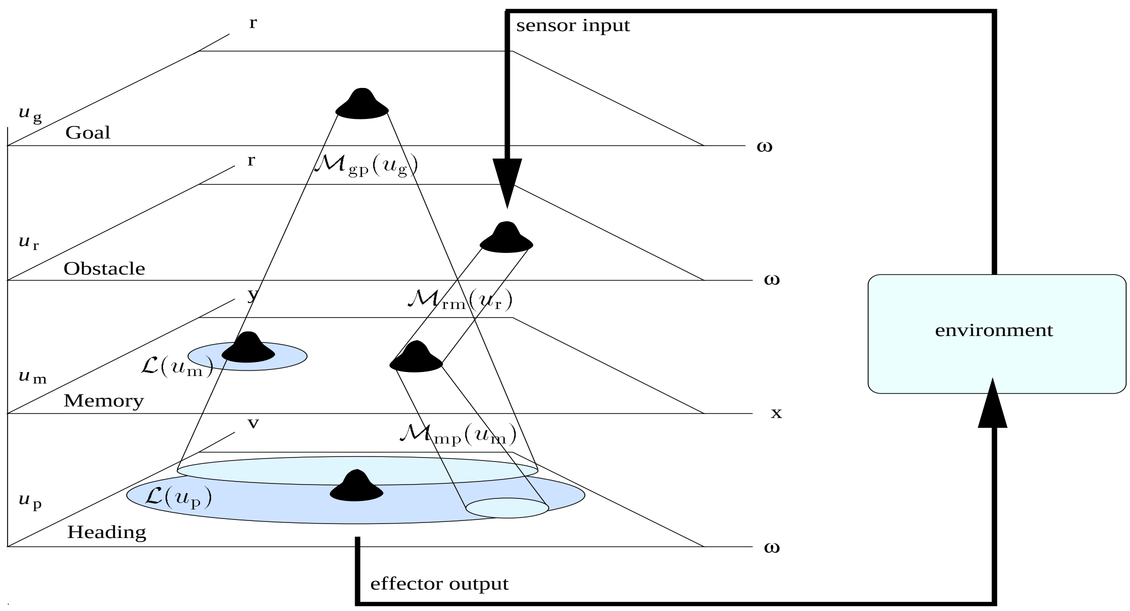

Fig. 8. Architecture of the path–planning system in closed loop with its environment. Mapping range is depicted as light grey area while lateral interaction range is shown as dark grey area. In our simulations the velocity in the heading map is set to $v =$ const.

Additionally the system builds a global map of its obstacle environment, see figure 9b. The obstacle representatives keep a certain distance to each other in order to avoid undesired local minima in the distribution $\mathcal{M}_{mp}(u_{\mathrm{m}})$ in case of cluttered environments. This world map could be used to perform higher cognitive tasks like global path planning.

5. Conclusion

In this report we have presented three different models which share the characteristics of continuous activity distributions within a map, lateral interactions and mappings between different layers. Each model performs a different behavioral respectively perceptual task: First, self–organization of *discrete parametrized repesentations* in the domain of early

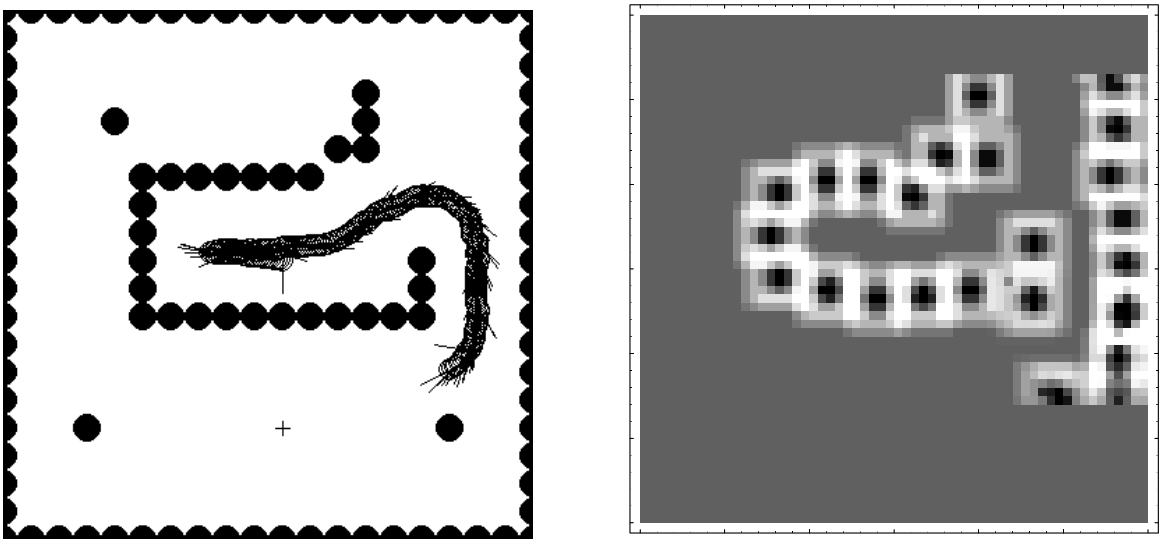

Fig. 9. (**a**) Trajectory of vehicle motion in a simulated environment where obstacles are shown as filled circles, the target location as cross and the vehicle as open circle with radiant as heading direction. A security distance to obstacles which the robot has to keep is equal to the length of the radiant and rotational symmetric. (**b**) Memory map is showing obstacle representatives as localized excitations (black grey–level values) acquired from sensory information during vehicle motion.

vision and *cross orientation inhibition* within this representation, second, integration of *preattentive* and *cognitive* saccade targets by means of a diffusion system, and third, local planning of a trajectory of an autonomous mobile robot and aquisition of a global obstacle map by interactions of *neural fields*. We think, the applied mechanisms can be regarded as a powerful toolkit in the development of an architecture which is able to generate complex behavior.

Acknowledgement

We wish to express our special thanks to Dr. G. Sch"oner for valuable comments and critical remarks. The neural field approach applied to the path planning problem we

described in section 4 mainly stems from his ideas and will be subject of a special paper currently under preparation.

References

[Ama77a] S. Amari. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27:77–87, 1977.

[Ama77b] S.-I. Amari. Neural theory of association and concept–formation. *Biol. Cybern.*, 26:175 – 185, 1977.

[Ark90] R. C. Arkin. Integrating behavioral, perceptual and world knowledge in reactive navigation. *Robotics and autonomous systems*, 6:105–122, 1990.

[Bal87] D. H. Ballard. Eye movements and spatial cognition. Technical Report TR218, Computer Science Department, University of Rochester, 1987.

[cDgJSC90] Y. c. Diao, W. g. Jia, N. V. Swindale, and M.S. Cynader. Functional organization of the cortical 17/18 border region of the cat. *Exp. Brain. Res.*, 79:271 – 282, 1990.

[Dau85] J. G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two– dimensional visual cortical filters. *Opt.–Soc. Am. A*, 2:1160 – 1169, July 1985.

[DBK90] H. R. O. Dinse, J. Best, and K. Krüger. A new dynamic model of receptive field organization. In G. Hauske R. Eckmiller, G. Hartmann, editor, *Parallel Processing in Neural Systems and Computers*, pages 61 – 64, Amsterdam, 1990. North–Holland.

[DKMB91] H. Dinse, K. Krüger, H. Mallot, and J. Best. Temporal Structure of Cortical Information Processing: Cortical Architecture, Oscillations, and Non-Separability of Spatio-Temporal Receptive Field Organization. In K. Krüger, editor, *Springer Series in Synergetics: Neuronal Cooperativity.* Springer Verlag, 1991.

[DM90] R. Durbin and G. Mitchison. A dimension reduction framework for understanding cortical maps. *Nature*, 343:644 – 647, 1990.

[DW87] R. Durbin and D. Willshaw. An analogue approach to the travelling salesman problem using an elastic net method. *Nature*, 326:689 – 691, 1987.

[Eng93] C. Engels. Neural fields as abstract datatypes. Technical Report 93-04, Institut für Neuroinformatik, 1993.

[ES79] M. Eigen and P. Schuster. *The Hypercycle, A Principle of Natural Self–Organisation*. Springer Verlag, 1979.

[Ess85] D. C. van Essen. Functional organization of primate visual cortex. In Alan Peters and Edward G. Jones, editors, *Cerebral Cortex, Volume 3: Visual Cortex*, pages 259 – 329. Plenum Press, New York and London, 1985.

[GJM92] G. J. Giefing, H. Janßen, and H. Mallot. Saccadic object recognition with an active vision system. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, pages 664 – 667, 1992.

[GTL93] A. P. Georgopoulos, M. Taira, and A. Lukashin. Cognitive neurophysiology of the motor cortex. *Science*, 260:47–51, 1993.

[HTS88] Y. Hata, T. Tsumoto, and H. Sato. Inhibition contributes to orientation selectivity in visual cortex of cat. *Nature*, 335:815 – 817, 1988.

[HW62] D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology (London)*, 160:106 – 154, 1962.

[HW77] D. Hubel and T. Wiesel. Functional architecture of macaque visual cortex. *Proc. R. Society London B*, 198:1 – 59, 1977.

[JB83] B. Julesz and J. R. Bergen. Textons, the fundamental elements in preattentive vision and the perception of textures. *Bell System Technical Journal*, 62:1619 – 1644, 1983.

[JK91] H. Janßen and J. Kopecz. Image Recognition in Hypercolumnar Scale Space by Sparsely Coded Associative Memory. In T. Kohonen et al., editor, *Artificial Neural Networks*, pages 1203 – 1206, 1991.

[JKM90] H. Janßen, J. Kopecz, and H. A. Mallot. Image Representation in Hypercolumnar Structure by Means of Associative Memory. In A. Gruen and E. P. Baltsavias, editors, *Close–Range Photogrammetry Meets Machine Vision*, pages 1050 – 1057. SPIE 1395, 1990.

[Koh82] T. Kohonen. Self–organized formation of topologically correct feature maps. *Biol. Cybern.*, 43:59 – 69, 1982.

[Kop92] J. Kopecz. *Modellierung lokaler Wechselwirkungen in diskreten cortikalen Strukturen mittels nichtlinearer Dynamik.* PhD thesis, FB Physik, Ruhr–Universität Bochum, 1992.

[Kop93] J. Kopecz. A cortical structure for real world information processing. In *Proceedings of the ICNN*, San Francisco, 1993. IEEE.

[MKvS92] H. A. Mallot, J. Kopecz, and W. v. Seelen. Neuroinformatik als empirische Wissenschaft. *Kognitionswissenschaft*, 4, 1992.

[MPG90] D. P. Munoz, D. Pelisson, and D. Guitton. Movement of neural activity on the superior colliculus motor map during gaze shifts. *Nature*, 251:1358–1360, 1990.

[MTT78] J.A. Movshon, I.D. Thompson, and D.J. Tolhurst. Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *J. Physiol.*, 283:53 – 77, 1978.

[NS71] David Noton and L. Stark. Scanpaths in saccadic eyemovements while viewing and recognizing patterns. *Vision Research*, 11:929 – 942, 1971.

[RS86] H. Ritter and K. Schulten. On the Stationary State of Kohonen's Self–Organizing Sensory Mapping. *Biol. Cybern.*, 54:99 – 106, 1986.

[SD92] G. Schöner and M. Dose. A dynamical systems approach to task-level system integration used to plan and control autonomous vehicle motion. *Robotics and*

Autonomous Systems, 10:253–267, 1992.

[TG80] A. Treisman and G. Gelade. A feature–integration theory of attention. *Cognitive Psychology*, 12:97 – 136, 1980.

[Tre85] A. Treisman. Preattentive processing in vision. *Computer Graphics and Image Processing*, 31:156 – 177, 1985.

[vdM73] C. von der Malsburg. Self–organization of orientation sensitive cells in the striate cortex. *Kybernetik (= Biol. Cybern.)*, 14:85 – 100, 1973.

[vdMW77] Ch. von der Malsburg and D.J. Willshaw. How to label nerve cells so that they can interconnect in an ordered fashion. *Proc. Natl. Acad. Sci. USA*, 74(11):5176 – 5178, November 1977.

[Viv90] P. Vivani. Eye movements in visual search: cognitive, perceptual and motor control aspects. In E. Kowler, editor, *Eye Movements and Their Role in Visual and Cognitive Processes.* Elsevier Science Publishers B.V., 1990.

[vSBKT93] W. v. Seelen, S. Bohrer, J. Kopecz, and W. Theimer. Visual Information Processing in Neural Architecture, 1993. Int. Journal of Computer Vision, submitted.

[vSJ92] W. von Seelen and H. Janßen. Structural principles in visually giuded autonomous vehicles. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, pages 302 – 311, 1992.

[Yar67] A. Yarbus. *Eye Movements and Vision.* Plenum Press, New York, 1967.