Visuomotor Anticipation – a Powerful Approach to Behavior-Driven Perception

Volker Stephan, Horst-Michael Gross

In this paper we present a biologically inspired neural architecture for visual perception based on anticipation. Our approach is able to explain several perceptional abilities of biological systems, for instance, expectation driven perception or the internal simulation of hypothetical action sequences in order to find an optimal action to be performed in reality. We demonstrate the functioning of our anticipatory approach for the visual perception of space of a mobile robot, that has to realize a local navigation behavior in an unknown environment. We claim that perception is not an end in itself, instead it is a sensorimotor process integrating the generation of behavior.

1 Introduction

In the classic understanding, visual perception is a purely data-driven process of sequential image processing stages that are transforming the retinal image into an internal, purely sensory representation of the external world. Afterwards, based on that representation an appropriate behavior is generated. This strict separation between perception and generation of behavior turned out to suffer from a number of conceptual problems. The main problems concern the generation of an internal representation that entails those sensory features that are *relevant* for the problem to be solved. This is critical, because the intentions, the basic motor abilities, and the typical properties of the whole system are without any influence on these representations [7]. Furthermore, the selection of appropriate behaviors from these representations has to be designed externally.

A number of alternative theories of perception have been developed recently that try to overcome these problems by replacing sensory with sensorimotor representations and by considering perception as an active and generative process rather than a pure projection [1], [6].

Based on that background, in section 2 we present selected neuroanatomical properties of the cerebral cortex explaining the mechanisms of perception and behavior generation in biological systems. Afterwards, section 3 describes the mobile robot we used for the experiments described in this article. Section 4 describes a biologically inspired neural architecture, which demonstrates, that perception driven by expectation is more powerful than a purely data-driven perception-approach. Section 5 extends that architecture by the ability for internal simulation in order to find appropriate motor actions. Finally, section 6 gives a short summary.

2 Biological Background

Today it is well known, that visual information about a current situation is carried from the retina to several cortical areas, for instance, to the areas *V1-V5(MT)* (Mediotemporal Cortex). The *posterior parietal cortex (PPC)* effects the integration of different sensory inputs, for instance from visual and primary somatosensory cortex (figure 1). Both the *premotor cortex (PMC)* and the *supplementary motor area (SMA)* receive projections from PPC and are strongly involved in generating hypothetical movements from memory [6], [10]. The structure of the *lateral cerebellum*, the strong connections from the cerebral cortex (MT, PPC, and PMC), and the connections back to premotor and motor cortices make the cerebellum a very suitable brain area to construct sensory predictions [2], [6]. The *basal ganglia (BG)* receive inputs from all parts of the cerebral cortex, including PMC, and have outputs directed strongly towards the premotor and prefrontal cortex via which they might influence movement selection and initiation [4], [11]. A more detailed discussion can be found in [3].

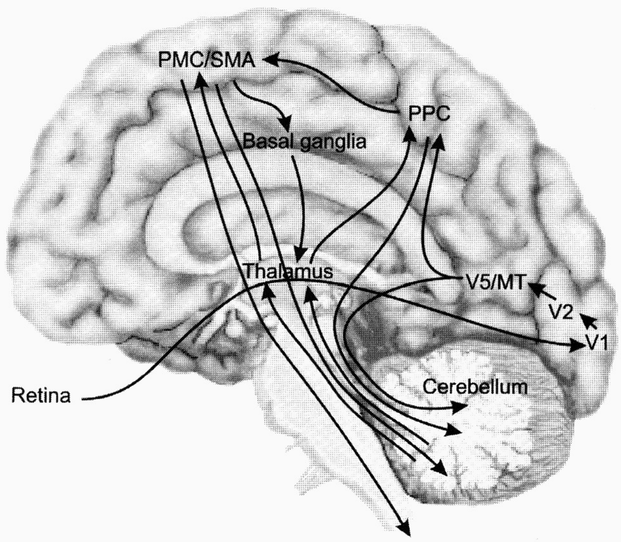

Figure 1: Model-relevant structures of the human brain with their interconnections and interrelations.

Derived Model: The retina converts the incoming optical data stream into a neural representation which is processed along the visual pathways via lateral geniculate nucleus to areas V1, V2, V3, V4 and into V5/MT. Area V5/MT contains an optic flow map [5], which is distributed to several cortical and subcortical areas, for instance, to the basal ganglia, to the lateral cere-

Künstliche Intelligenz, Heft 2/03, Seiten 12-17,
ISSN 0933-1875, arendtap Verlag, Bremen

bellum, and to the PPC. The massive afferent sensory information in PPC from area V5/MT, from somatosensory cortex and from thalamus allow this structure to generate a description of the system's current status. The output of PPC is transferred to the motor areas of the cerebral cortex, especially to PMC and SMA [11], which uses the current system status to generate several action strategies. Thus, in SMA/PMC, afferent sensory information is transformed into a set of possible motor commands describing the current sensory situation.

Cortico – basal ganglionic loop: The motor alternatives generated in SMA/PMC are projected to the basal ganglia, which associate evaluation signals, learned from dopamine neurons in the midbrain, to these actions [4], [11]. Based on this information, the local connectivity of BG selects a subset of SMA/PMC suggested actions, that had positive outcomes in the past. This set of appropriate motor commands is fed back via thalamus to SMA/PMC closing the loop between cortex and basal ganglia [8], [4]. Thus, in our model, the cortico - basal ganglionic loop evaluates and selects actions.

Cortico – cerebellar loop: From SMA/PMC, the information about appropriate movements is projected to the lateral cerebellum [4], which uses the sensory context coming from area V5/MT to predict the sensory consequences of the proposed motor commands. This sensory prediction is projected back to the motor cortices and to PPC via thalamus closing the cortico-cerebellar loop. Thus, the cortico - cerebellar loop seems to anticipate the sensory outcomes of hypothetical movements suggested by SMA/PMC and evaluated by basal ganglia.

Integration of both loops: Now, the cycle of information processing described above can be reentered time and again in order to simulate and evaluate entire sequences of hypothetical sensorimotor states, while the best evaluated motor sequence may be buffered in SMA/PMC. The hypothetical sensory situation, predicted by the cerebellum, can be used as a starting point for the next anticipation cycle. Then, SMA/PMC can suggest a new set of possible actions for the hypothetical sensory situation, the basal ganglia evaluate and select appropriate movements, the cerebellum predicts their sensory consequences, and so on. Finally, the initial motor command of the sequence best evaluated out of all simulated ones is fed to primary motor cortex, which is responsible to execute the selected motor commands in reality. The real consequences of this action are observed by the sensory system (visual, tactile) and can be used to adapt the sensory prediction by learning. Subsequently, the whole process of internal simulation starts again.

In our approach, perception of space and shape is regarded to be an active process which anticipates the sensory consequences of alternative hypothetical interactions with the environment, that could be performed by the sensorimotor system. This point of view emphasizes the generative and anticipative character of perception considering both sensory and motor aspects of the action-perception-cycle. Our model does not attempt to account for a detailed description of specific cortical or subcortical structures, but we try to capture some general properties of architecture and processing that we found in neuroanatomical literature. To the best of our knowledge, other concrete computational neural models allowing to handle the problem of anticipatory search and internal simulation in order to explain perception and generation of behavior at the level of sensorimotor intelligence are not yet known.

3 Experimental Framework

Because perception and behavior generation are just two aspects of one consistent neural process of internal simulation, it appears useful to investigate the developed architecture by means of an autonomous sensorimotor system. For this, we used a mobile robot Khepera, which features an omnidirectional camera for visual data acquisition and two motors for the execution of actions. The robots task is a collision-free local navigation behavior based only on visual data within an environment depicted in figure 2.

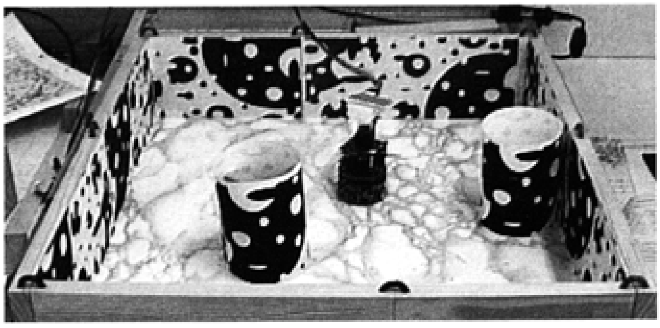

Figure 2: Used environment with the visuomotor system, a robot Khepera, equipped with an omni-directional camera.

Thereto, we perform a polar transformation of the omnidirectional camera-image and estimate the optic flow, which may serve as a representative for the whole spectrum of visual information.

4 Expectation Driven Perception

In this section, we describe an interesting subprocess of the biological model developed in section 2. The projection from the lateral cerebellum back to PPC, carrying the predicted sensory consequences of a simulated action, can be used to match the experienced sensory observation with the predicted outcomes of the currently executed motor command. Thus, the sensory prediction may serve as an internal expectation about what should be observed (see figure 3).

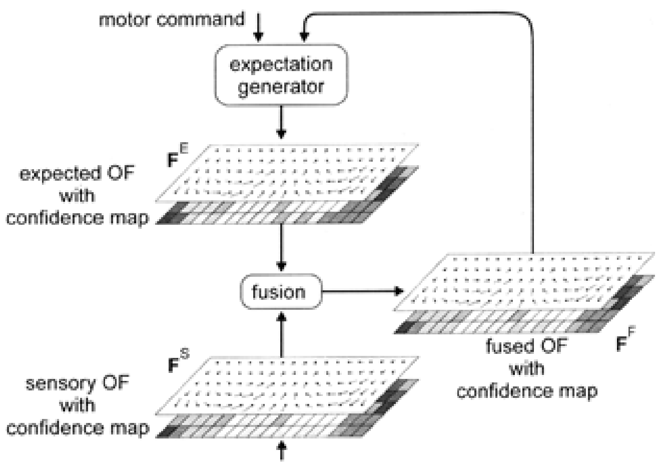

Figure 3: Hybrid architecture to fuse the sensory bottom-up data and the top-down expectation.

Normally, these two hypotheses should fit well, but in case of sensory disturbances, the resulting erroneous measurements can be replaced by the corresponding expectation.

The expectation generation requires the ability to learn a mapping from the current flow field *OF(t)* and the executed action *a(t)* to the resulting situation *OF(t+1)*. Therefore, we investigated a large set of different artificial neural networks. A detailed discussion of all tested networks with their performances is given in [9]. The best performing predictor turned out to be the recurrent Jordan-network.

To reliably fuse the top-down expectation with the sensory bottom-up observation, each vector of the flow field is represented by an activity distribution within a 2-dimensional neural field (blob), where the position of this blob within this 2D-space codes the x- and y-components of the flow-vector, and the height of the blobs is a measure for the corresponding confidence of this flow vector. The sensory bottom-up confidence is provided by the optic flow estimator and the confidence of the top-down expectation is computed from the approximation-quality of the neural-network learning the sensory prediction. Due to the 2-dimensional representation, it is possible to hold many alternative hypotheses (blobs) for each flow vector. Consequently, both the sensory bottom-up and the top-down expectation can add their hypotheses into the corresponding representation, whereby similar hypotheses result in a superposition of the blobs at the same position. The output results from the hypothesis with the highest confidence. Hence, this algorithm selects those of all hypotheses, which support each other. This is reasonable, since similar information in both streams implies, that this information is reliable and trustworthy.

In order to test this architecture, we put the robot with a pretrained sensory predictor in the situation depicted in figure 4. The robot moved along the arrow straight ahead with fixed velocity towards the upper wall. At the positions marked by the white arcs, the camera was occluded partially in order to emulate a sensory distortion. In consequence, the sensory data entail almost no information about the oncoming obstacle (upper wall in figure 4) on the robots left.

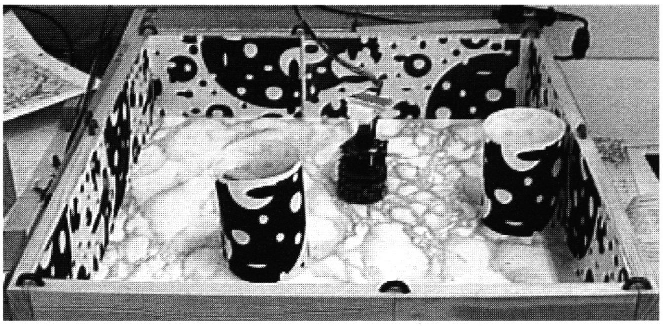

Figure 4: Test-scenario for expectation driven perception.

Figure 5 depicts the observed camera images and flow fields during that locomotion. As can be seen in the non-disturbed sequence, the oncoming obstacle on the left causes growing flow vectors especially in the left-middle (sensory non-disturbed OF). In contrast, the applied disturbance prevents a correct estimation of optical flow vectors in the corresponding part of the camera image. Nevertheless, through active generation of an expectation about the external world and the fusion with the noisy sensory information, our anticipative system is able to maintain a valid representation of the oncoming obstacle (anticipatory fused OF).

To demonstrate the facilities of the presented anticipatory preprocessing at the behavioral level, we placed the robot in an unknown environment to navigate through a narrow passage without collision. For this benchmark, we used the balancing approach, which tries to equalize the optical flow in both hemispheres of the robot, which results in a collision-free locomotion in the middle of hallways. Figure 6 (left column) shows a top view of this scenario with the collision-free traces of our robot.

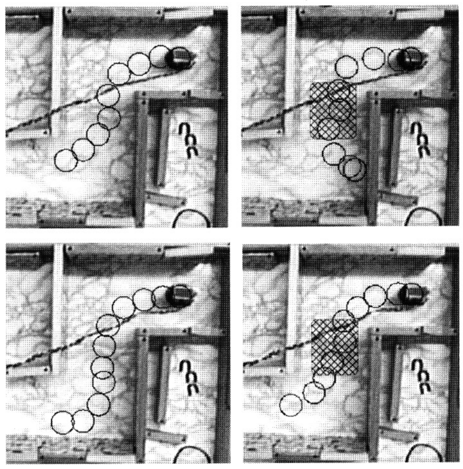

Figure 6: Navigation based on the estimated optical flow starting at the upper right corner and moving to the opposite one. As can be seen, both the navigation on the pure estimated optical flow (top left) and on the expectation driven preprocessed optical flow (bottom left) allow a collision-free locomotion of the robot KHEPERA through the environment. In contrast, a significant disturbance of the optical flow estimated by fluctuating ambient light in the areas marked by the hatched areas causes a collision at the end of the plotted trace, if no anticipative preprocessing is applied (top right). The anticipative preprocessing overcomes the problems and allows a collision-free locomotion (bottom right).

If a perturbation is applied in this experimental situation, the navigation based only on estimated optical flow fields must fail, because the very noisy sensory input entails almost no information about close obstacles (top right). In contrast, our anticipatory preprocessing allows the system to bridge the time gap of sensory dropouts with the generated expectation and is therefore able to extract relevant information in order to avoid the oncoming obstacles (bottom right).

5 Internal Simulation

The central idea of the biological model described in section 2 are the two reentrant loops, the cortico-basal ganglionic-loop and the cortico-cerebellar-loop, which are passed through repeatedly in order to simulate internally alternative sensorimotor sequences. Since the internal simulation process in our biological model is executed in parallel by reentrant information flow in the two loops, not only a single, but a whole set of alternative movements, evaluations and predictions has to be represented in the same neural architecture at the same time. In this case, not only an extremely large sensorimotor manifold would have to be represented, but also a correct causal and temporal binding of the related sensory and motor components would be required in order to evaluate and compare whole sequences and to select the best one.

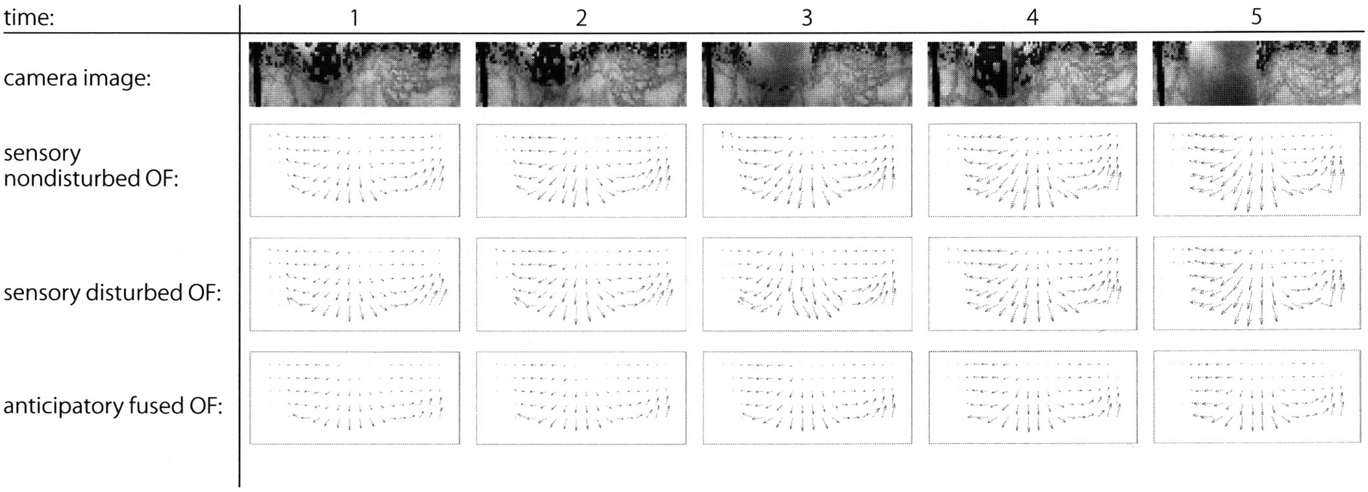

Figure 5: Expectation-driven perception during locomotion of the mobile robot depicted in figure 4. row 1 camera images with disturbance observed by the robot. row 2: Optic flow (OF) sequence without disturbance. row 3: OF sequence with disturbance. row 4: anticipatory fused OF.

In consideration of these, at least at present, unsolvable problems arising from the massively parallel computing and coding in the biological model, we have investigated conceptionally equivalent but better realizable anticipation strategies for our computational model. It is based on a cascaded neural architecture which, instead of searching for appropriate actions in parallel, uses an efficient sequential search.

This sequential search is realized by a concatenation of replicative, so called *Prediction Modules* (PM) (see Figure 7) operating at staggered time-scales of a time hierarchy. The first *Prediction Module* (PM1) operates on the real sensory input, generates a sequence of alternative, evaluated action hypotheses, and predicts their sensory consequences. The second module (PM2) just operates on the predicted sensory consequences of the simulated PM1-actions, generates itself a set of alternative actions, and anticipates their sensory consequences. This internal simulation and prediction process on hypothetical data can be continued in subsequent modules. In this way, a whole search tree of possible sensorimotor sequences is developed, which has to be evaluated to find the best sequence.

Of course, a complete search in this tree would be very time expensive. Moreover, numerous branches include sequences with already experienced, low evaluations that do not have to be simulated time and again. To reduce the complexity of search in the sensorimotor space, we use an active dynamic approach for a combined search in width and depth, which only simulates the most promising sensorimotor transitions at each prediction level of the cascaded architecture. That way, the search space for the motor commands is restricted and the effort in time is reduced.

Figure 8 depicts the internal architecture of a single *Prediction Module* (PM) as the basic building block of the cascaded architecture. Each sensory input is processed in two pathways that correspond to the cortico-basal ganglionic and the cortico-cerebellar loop of the biological model:

- *Sensorimotor Mapping* with the subsystems *Action Suggestion*, *Action Selection*, and *Hypotheses Evaluation* (cortico-basal ganglionic loop) and
- *Motorsensory Prediction* with the *Sensory Prediction* subsystem (cortico-cerebellar loop). In these pathways, hypothetical actions and their expected sensory consequences can be simulated.

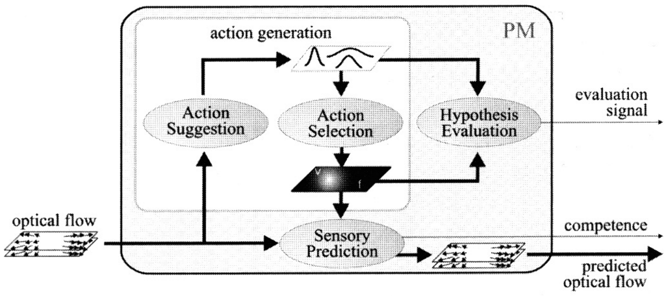

Figure 8: Architecture of a single Prediction Module.

Like the basal ganglia, the evaluation of actions (*Action Suggestion*) utilizes the learned associations between really executed actions, sensory context and experienced rewards (see [3] for more details on the applied neural network realizing a simple reinforcement-learning). The local competitive connectivity of striatal neurons realizes the selection only of appropriate action hypotheses (*Action Selection*). The second, predictive pathway (*Sensory Prediction*) is supposed to have a biological counterpart in the cortical loop through lateral cerebellum. The process of storage and evaluation of whole action sequences (*Hypothesis Evaluation*) is assumed to be performed in the SMA/PMC. It should be stressed, that the *Hypothesis Evaluation* has no direct biological counterpart, since it is a direct result of the sequential search strategy. We assume, that the final selection of the best action sequence is done by the motor cortices.

The application of this technical realization of the biological model within the experimental framework introduced in section 3 documents its principal functioning. Thereto, both the neural networks predicting sensory consequences and the network learning evaluations were trained based on experienced interactions of the robot with its environment. Figure 9 depicts the internal simulation process realized by three subsequent prediction modules.

The really observed optic flow field (left) features large vectors in the middle caused by an oncoming frontal obstacle.

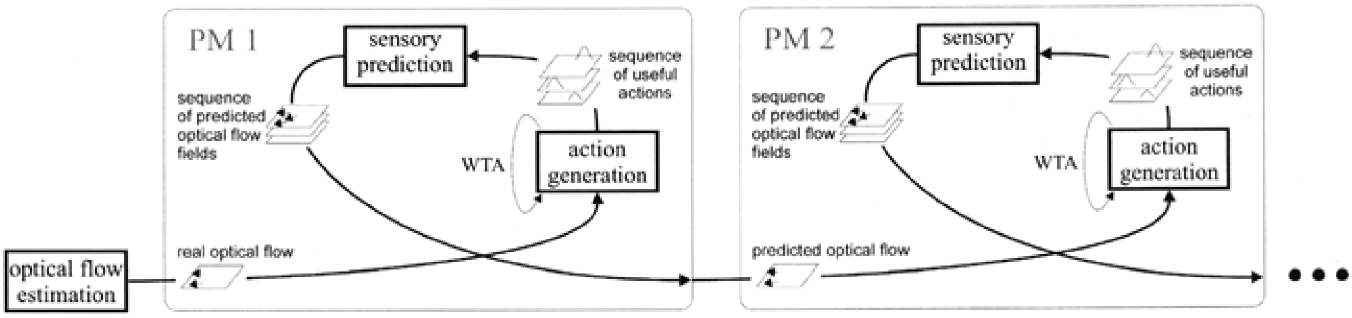

Figure 7: Chaining of Prediction Modules (PMs) for internal simulation of sequences of hypothetical actions and expected sensory consequences. This cascaded architecture realizes a sequential search strategy instead of a parallel one to find the best evaluated sensorimotor sequence out of all simulated sequences.

Through internal simulation of two subsequent straight-ahead-movements the anticipatory system is able to predict a difficult situation (*OF*(t+2) top), where the obstacle is very close and even movements to the left or right cannot prevent a collision. For that reason, this trajectory is evaluated pretty poor ($R = 2.746$). In contrast, the simulation of a turning action to the left followed by two drive-forward-commands (row 2) passes by the collision, is evaluated much better ($R = 2.918$) and would be preferred. This tree-step internal simulation procedure including at each stage three different actions to be investigated takes on a standard state of the art PC much less than a second!

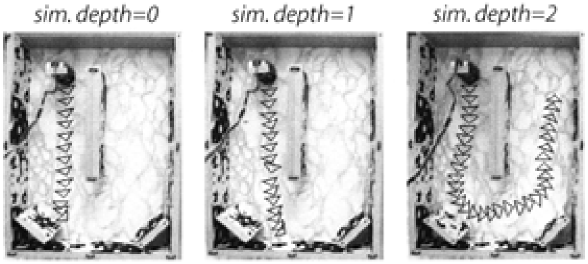

Figure 10: Local navigation performance of the mobile robot for different depths of internal simulation (sim. depth). Each simulation stage is realized by one prediction module. The robot is always starting in the upper left corner and should move without collisions through the U-shaped environment. The black triangles mark the trace with the corresponding orientation of the robot. The trace ends in case the robot reaches the upper right corner or on any collision.

Figure 10 shows the functioning of our architecture at a behavioral level. As can be seen, the more steps can be simulated internally, the better the local navigation performance becomes.

6 Summary and Conclusions

The basic idea of our anticipatory approach to perception is to avoid the common separation between perception and generation of behavior. Our approach is strongly inspired by neuroanatomical knowledge and tries to explain the phenomenon of perception at the level of sensorimotor intelligence from a behavior-oriented point of view.

In section 4, we used the anticipated sensory consequences to improve the really experienced optic flow field by an expectation driven fusion process. Through merging these two datastreams, it is possible to eliminate noisy or invalid observations in order to make the collision-free navigation behavior more robust. We can demonstrate, that even in the presence of massive artificial disturbances the robot is able to maintain a valid sensory representation and, thus, is able to navigate properly.

In section 5, we used the internal simulation process of our architecture to find appropriate action sequences that fulfill the system goal: a collision-free local navigation. It could be shown, that through our sensorimotor perception-approach, the robot became able to perceive its environment and to realize the desired navigation behavior.

This biologically inspired anticipatory approach for sensorimotor perception, overcomes the conceptual problems of the classical information processing paradigm by integrating both perception and behavior generation into one process. Its functioning could be demonstrated for a real sensorimotor system, a mobile robot, which learned to perceive and to behave properly within its environment.

References

[1] M.A. Arbib. *The Metaphorical Brain: An Introduction to Cybernetics as Artificial Intelligence and Brain Theory.* Wiley Interscience, 1972.

[2] C. Ghez. The cerebellum. In *Principles of Neural Science,* pages 626–646. Appleton & Lange, 1991.

[3] H.-M. Gross, A. Heinze, T. Seiler, and V. Stephan. Generative character of perception: A neural architecture for sensorimotor anticipation. *Neural Networks,* 12:1101–1129, 1999.

[4] J.C. Houk and S.P. Wise. Distributed modular architectures linking basal ganglia, cerebellum, and cerebral cortex: Their role in planning and controlling action. *Cerebral Cortex,* 5(2): 95–110, 1995.

[5] E.R. Kandel. Perception of motion, depth and form. In *Principles of Neural Science,* pages 440–466. Appleton & Lange, 1991.

[6] S.M. Kosslyn and A.L. Sussman. Roles of imagery in perception: Or, there is no such thing as immaculate perception. In *The Cognitive Neuroscience,* pages 1035-1042. MIT Press, 1995.

[7] R. Möller. *Wahrnehmung durch Vorhersage. Eine Konzeption der handlungsorientierten Wahrnehmung.* PhD thesis, Technische Universität Ilmenau, 1996.

[8] F.A. Middleton and P.L. Strick. Anatomical evidence for cerebellar and basal ganglia involvement in higher cognitive function. *Science,* 266:458–461, 1994.

[9] V. Stephan. *Visuomotorische Antizipation: eine handlungsorientierte Sicht auf die visuelle Wahrnehmung.* PhD thesis, Technische Universität Ilmenau, 2002.

[10] W.T. Thach. On the specific role of the cerebellum in motor learning and cognition: Clues from pet activation and lesion studies in man. *Behavioral and Brain Sciences,* 19(3):411–431, 1996.

[11] J. Wickens. Basal ganglia: structure and computations. *Network: Computation in Neural Systems,* 8:77–109, 1997.

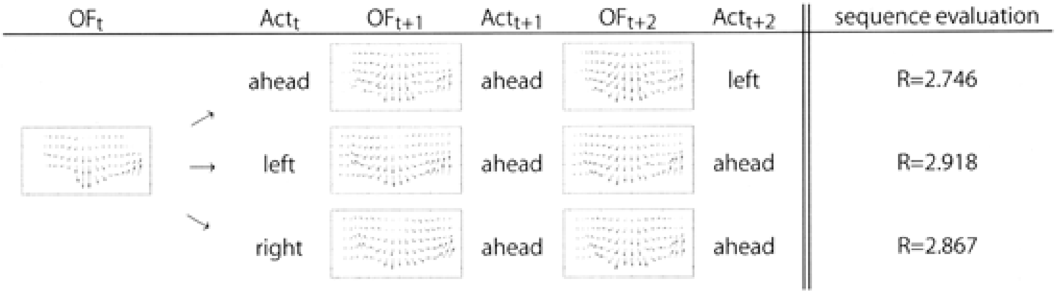

Figure 9: Process of internal simulation of the anticipatory system. Starting from the really observed one (left), for each optic flow field OF_t an action evaluation map is generated (not shown here) and different actions (Act) are selected sequentially, evaluated, and their sensory consequences OF_{t+x} predicted. The planning horizon runs from left to right, the simulation progress from top to bottom.

Contact
Dr. Volker Stephan
TU Ilmenau
Department of Neuroinformatics
Postfach 100565, 98684 Ilmenau
volker.stephan@informatik.tu-ilmenau.de

Prof. Dr. Horst-Michael Gross
TU Ilmenau
Department of Neuroinformatics
Postfach 100565, 98684 Ilmenau
Horst-Michael.Gross@tu-ilmenau.de

Volker Stephan is a member of the department of neuroinformatics at the Technical University Ilmenau since 1996. He is working in the field of reinforcement-learning and visually guided robot navigation using neural networks.

Horst-Michael Gross is professor of Neuroinformatics at the Ilmenau Technical University and heads the Department of Neuroinformatics since 1993. He received his Diploma degree in Electrical Engineering and his Doctorate degree in 1985 and 1989, respectively. Among his main research interests are neural computing, autonomous robots, reinforcement learning, and vision-based navigation.