In Proceedings of 2001 IEEE International Conference On Systems, Man & Cybernetics, pp. 2275-2280.

# NEURAL ANTICIPATIVE ARCHITECTURE FOR EXPECTATION DRIVEN PERCEPTION

**VOLKER STEPHAN and HORST-MICHAEL GROSS**

Ilmenau Technical University, Department of Neuroinformatics
98684 Ilmenau, Germany
volker.stephan@informatik.tu-ilmenau.de

## Abstract

In this paper we present a biologically inspired neural architecture for visual perception based on anticipation. The main goal of this work is to demonstrate, that anticipation is a central key to improve the perception performance of technical systems. The presented approach is able to increase the robustness of the perception process against noise or sensory dropouts. We demonstrate these perceptional improvements through our architecture at the level of local navigation behavior of the miniature robot Khepera. We claim that perception is not an end in itself. Instead it is a sensorimotor process integrating the generation of behavior.

## Keywords

visual perception, anticipation, expectation, local navigation

## 1 Introduction

Visual perception in biological systems is a fast, extremely robust and very powerful process, which is not yet completely understood. Traditional approaches for visual perception try to solve technical problems based on the 'information processing paradigm' [1], which can be characterized by a chain of subsequent signal-processing modules providing a set of features that describes the current situation to the system. Based only on this description another module generates the appropriate behavior. This strict separation between sensory perception and generation of behavior and the resulting problems are reviewed in detail [2, 3].

In recent years, the appreciation of visual perception as a generative sensorimotor process gained increasing acceptance [4, 5]. The generative and active aspect of perception has been emphasized especially by [6, 7, 8] who supposed that anticipation of sensory consequences of actions may play an integral role in perception.

If this holds true, then there must exist structures that are capable of predicting the sensory consequences of actions. Such sensory predictors seem to be multi-functional systems, since they can be used to a) enhance the incoming bottom-up sensory information by a top-down expectation generated previously [9]; b) direct selective attention to those environmental subregions which caused a mismatch of top-down expectation and bottom-up sensory information and c) internally simulate the consequences of action sequences in order to find and execute those actions, that entail positive outcomes for the system [10]. In this paper, we focus on a) and present a corresponding network architecture.

## 2 Experimental framework

For our experiments, we use the robot platform KHEPERA, a miniature robot equipped with an omnidirectional color-camera to investigate the proposed mechanisms at the level of behavior. Figure 1 depicts a typical environment with the mobile robot inside. The

Figure 1: View of the used environment with the robot KHEPERA, equipped with an omni-directional camera.

input of our architecture is, as indicated in section 1, exclusively visual. Thus, we use optic flow as a small fraction of the entire visual input features, because it is largely independent of specific visual details of the objects in the scene, entails implicit information about the 3D-structure of the environment, object motion and the egomotion of the system. In the preprocessing of the original omni-camera-image we perform a polar transformation to the deskewed form depicted in figure 2. These transformed images are used to estimate the optic flow fields, because an action of the robot with a rotational part yields a rotation of the omni-camera-image but only a shift in x-direction of the polar transformed image. This is very advantageous, since the applied correlation based optical flow estimation [11] needs not cope with rotated correlation areas, which would be very time consuming.

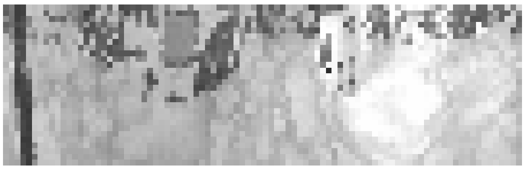

Figure 2: Polar transformed image of the omnidirectional camera of the KHEPERA: middle=front, left and right image borders=back of the robot.

The system's goal is a collision free local navigation based only on visual information. Thus, it has to extract robust optic flow vector fields from the video stream, even if the sensory data stream is temporarily disturbed. We believe, that the behavior of an autonomous system operating only on its perceived information is a very good indicator of the performance of the perceptual system.

## 3 Biological inspiration

The visual perception process in biology is not purely sequential, instead it features many recurrent connections (see figure 3).

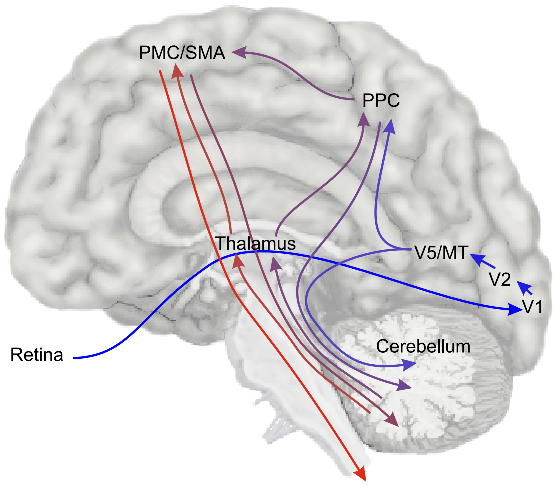


Figure 3: Model-relevant structures of the human brain with its connections.

Starting from the retina, a sequential processing of simple image features by the early visual areas V1-V2 is followed by the estimation of optic flow in area MT/V5, which projects also to the posterior parietal cortex (PPC). In addition to that sensory data stream also projections from the lateral cerebellum arrive, which contribute a sensory prediction of the consequences of real or hypothetical actions generated by the premotor areas SMA/PMC [10]. Thus, area PPC may be able to fuse the sensory bottom up and the expected top-down information in order to replace faulty information from any source and to extract more robust visual features.

## 4 Architecture

Inspired by that aspect of biological information processing, we developed a functionally similar architecture, which predicts optical flow fields as visual consequences of real or hypothetically executed actions of our mobile robot (see figure 4).

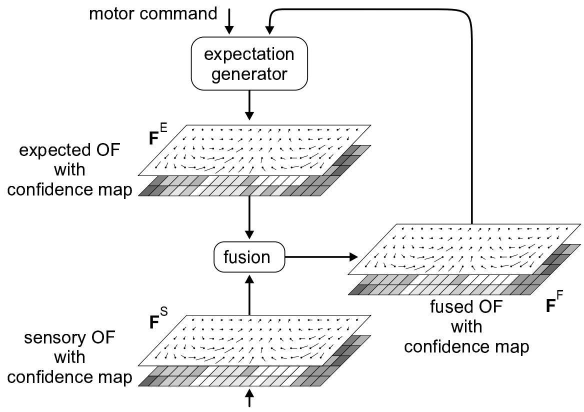


Figure 4: Hybrid architecture to fuse the sensory bottom-up data and the top-down expectation.

A central aspect of our anticipative processing in the bottom-up/top-down cycle is the usage of flow vector specific confidence estimates organized topographically corresponding to the flow field. These confidence-values of each flow vector are based on correlation-based optical flow estimation [11] by evaluating the shape of the correlation function. Sharp and unique minima cause high confidence values, whereas flat or ambiguous correlation functions result in low ones.

### 4.1 Expectation generation

As indicated in figure 4, the sensory prediction of the consequences of the mobile systems own actions is an essential part of our architecture. Therefore, we trained an artificial neural network on really experienced optical flow data of the mobile robot within a static environment. Thus, the applied neural network learns universally valid sensorimotor relationships between subsequent optic flow fields under influence of the systems own motor commands. Consequently, the predictor realizes the anticipative function approximation $OF(t) \times a(t) \mapsto \hat{OF}(t+1)$, where $OF(t)$ is the current optic flow, $a(t)$ is the currently executed motor action of the robot and $\hat{OF}(t+1)$ is the predicted optic flow field. More details about the applied networks can be found in [12, 13].

### 4.2 Fusion

With regard to figure 4, in this section we present the fusion between bottom-up and top-down information (see figure 5).

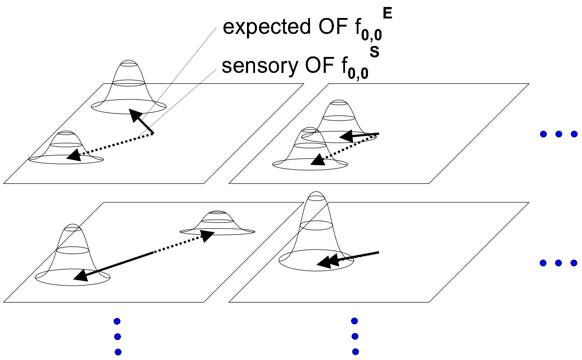


Figure 5: Each vector of the optical flow field is represented by a 2-dimensional neural field, where the position of activation blobs within the neural field codes possible flow-vectors.

Each vector of the whole field is represented by a small 2-dimensional neural field, where the position within the neural field codes the x- and y-components of the flow-vector as a blob, and the height of the blobs in the neural field is a measure for the corresponding confidence of this flow vector. Due to this distributed 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 about the real optical flow vector into the corresponding neural field, whereby similar hypotheses result in a superposition of the blobs at the same position.

For reasons of simulation resources, we split the 2-dimensional neural field into 2 one-dimensional neural vectors representing the x- and y-direction of the flow vector separately (equations 1, 2). Thus, the time to fuse two flow fields containing $18 \times 5$ vectors each could be reduced to $275ms$ (Pentium 200MHz). Equation 1 shows, that the new state in the fusion-map $\underline{z}^x_{pq}(t)$ is computed by the superposition of the discounted previous state $\underline{z}^x_{pq}(t-1)$ with $\alpha \in (0 \ldots 1)$, the sensory bottom-up vector $\underline{\vartheta}(f^S_{xpq}(t))$ and the top-down expectation $\underline{\vartheta}(f^E_{xpq}(t))$ in form of one-dimensional gaussian blobs (equation 3) weighted by their confidences $c(.)$. These one-dimensional blobs $\underline{\vartheta}$ realize a topological coding of the sharp x and y coordinates of the corresponding flow vectors and allow to represent multimodal hypotheses. The fused output $\underline{f}^F_{pq}(t)$ results from the hypothesis with the highest confidence (equation 4).

$$\begin{aligned}
\underline{z}^x_{pq}(t) &= \alpha \underline{z}^x_{pq}(t-1) + \underline{\vartheta}(f^S_{xpq}(t)) \cdot c^S_{pq}(t) \\
&\quad + \underline{\vartheta}(f^E_{xpq}(t)) \cdot c^E_{pq}(t) && (1) \\
\underline{z}^y_{pq}(t) &= \alpha \underline{z}^y_{pq}(t-1) + \underline{\vartheta}(f^S_{ypq}(t)) \cdot c^S_{pq}(t) \\
&\quad + \underline{\vartheta}(f^E_{ypq}(t)) \cdot c^E_{pq}(t) && (2) \\
\vartheta_k(u) &= e^{-\frac{(u-k)^2}{2\sigma^2}} && (3) \\
\underline{f}^F_{pq}(t) &\Leftarrow \begin{pmatrix} argmax(\underline{z}^x_{pq}(t)) \\ argmax(\underline{z}^y_{pq}(t)) \end{pmatrix} && (4) \\
c^F_{pq}(t) &= \frac{max(\underline{z}^x_{pq}(t)) + max(\underline{z}^y_{pq}(t))}{2} && (5)
\end{aligned}$$

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.

## 5 Results

To demonstrate the facilities of the presented anticipatory preprocessing, we placed the robot in the unknown environment depicted in figure 6, where it drove along the arrow straight forward towards the wall. In the last section of that pathway, a temporal, ar-

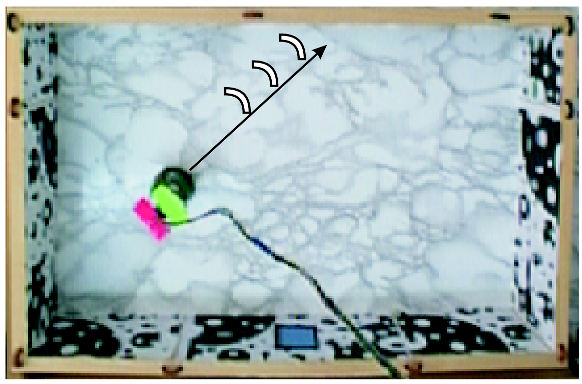

Figure 6: Path of the mobile robot for investigation of expectation-driven perception.

tificial disturbance was introduced by covering an area of about $90°$ of the omnidirectional camera image by a white paper (see white circular arcs in figure 6). In consequence, the sensory data entail almost no information about the oncoming obstacle on the left of the robot.

Figure 7 depicts the observed camera images and flow fields during that locomotion in detail. As can

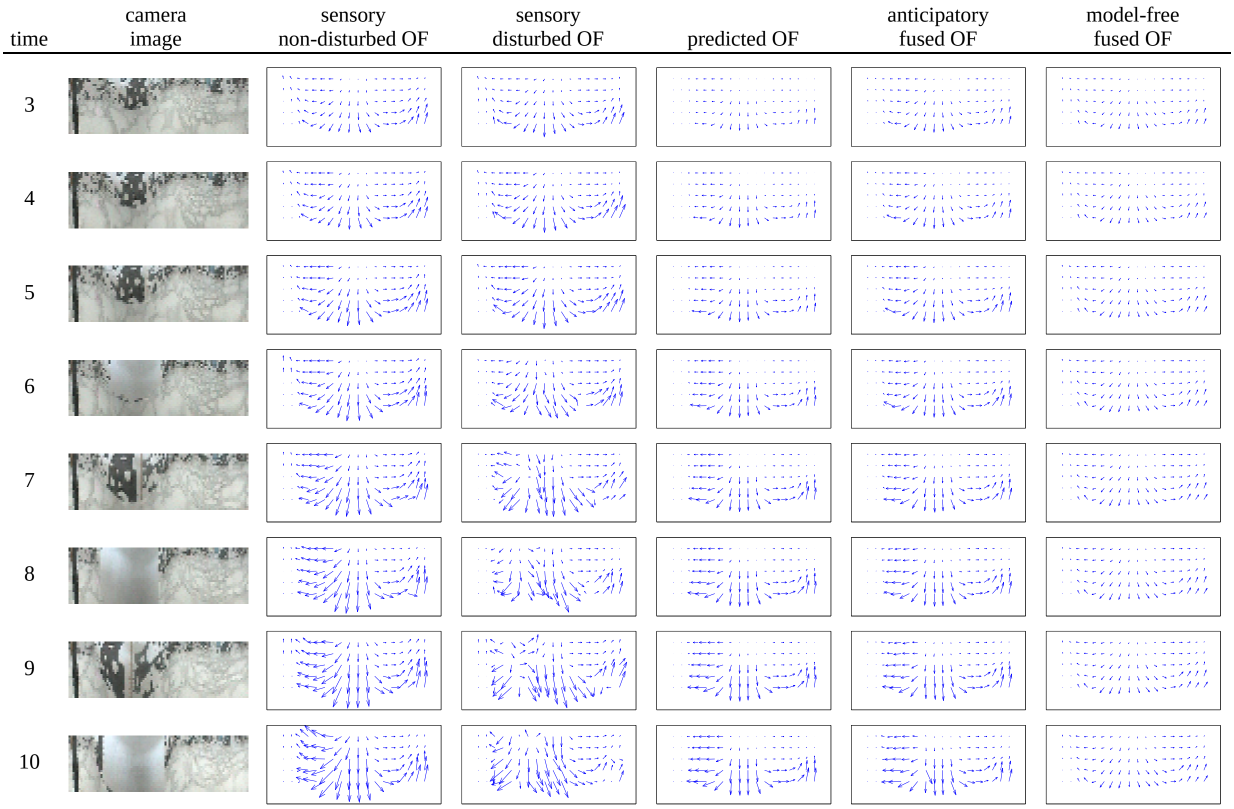


Figure 7: Results of expectation-driven perception during locomotion of the mobile robot depicted in figure 6. *left:* camera images with disturbance observed by the robot. *2. column:* OF sequence without disturbance. *3. column:* OF sequence with disturbance. *4. column:* predicted/expected OF. *5. column:* anticipatory fused OF. *6. column:* model-free fused OF.

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 was able to maintain a valid representation of the oncoming obstacle (anticipatory fused OF).

At this point, we have to ask the question: would a pure feedback without any sensory prediction (see figure 4) result in the same behavior? In this case, the fusion of the noisy and very unreliable estimated optical flow with the more reliable previous fusion result would probably just return the last fusion output. Thus, such an architecture is nothing but a low-pass filter over time. As can be seen in the rightmost column in figure 7, the system without anticipative preprocessing is not able to predict the enlarging flow vectors representing the close obstacle. The repeatedly generated last reliable sensory situation deprecates more and more over time and does therefore not represent the external reality.

Figure 8 depicts the development of a vector at the middle-left position (row 5, column 7) of the flow field over time. The optic flow vector generated by fusion of the predicted and the noisy sensory estimate is constantly growing over time and thus reflects the oncoming obstacle. In contrast, both the pure sensory estimation and the model-free fused vectors are significantly affected by the disturbance. A navigation based on these noisy data would probably be impossible.

Figure 9 depicts the resulting steering angles from the optic flow vector fields depicted in figure 7 following the balancing approach [14].

Based on the non-disturbed flow fields a steering command to the right would be generated in order to avoid the collision. In contrast, utilizing the disturbed signal, very small and later on, steering angles towards the wall would be executed. Likewise, the model-free approach is also unable to prevent the oncoming collision, because of its obsolete representation. Obvi-

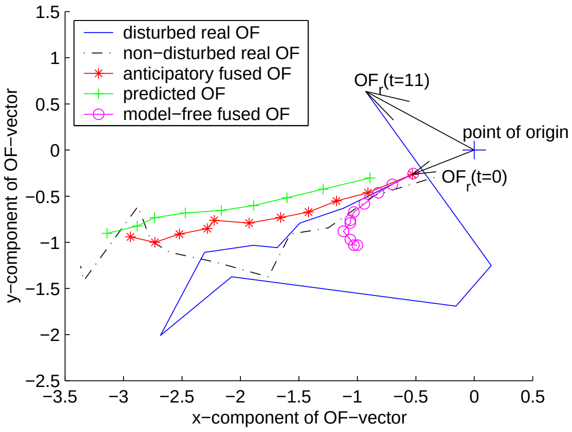


Figure 8: Development in time and space of vector at position $(5, 7)$ within the flow field. The markers on the lines correspond to the end-points of the flow-vectors over time, where all vectors originate at the point of origin. The sensory flow vectors at time-steps $t = 0$ and $t = 11$ of the partially disturbed sequence are additionally plotted.

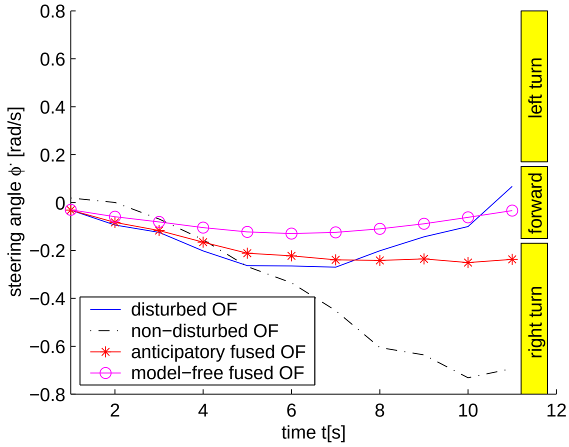


Figure 9: Development of hypothetical steering angles over time based on different optic flow fields depicted in figure 7.

ously, only the architecture utilizing the expectation-driven fusion can maintain a valid internal representation of the oncoming wall, and generates steering angles to the right.

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 [14], which tries to equalize the optical flow in both hemispheres of the robot, which results in a collision-free locomotion in the middle of such an hallway. Figure 10 (left column) shows a top view of this scenario with collision-free traces of our robot. If a perturbation is applied in this experimental situation, the navigation based only on estimated optical flow fields fails, 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 arising obstacles (bottom right).

## 6 Conclusion & Outlook

In this paper, we presented a hybrid neural architecture to model the biologically inspired principle of expectation driven visual perception at a functional level. We used neural networks to learn universally valid relationships between the motor actions of our mobile system and the corresponding visual consequences. That

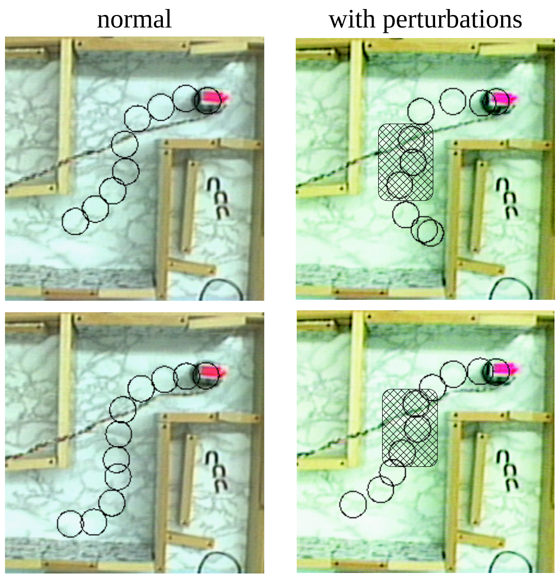


Figure 10: Navigation based on the estimated optical flow applying the well known balancing approach [14] starting at the upper right corner and moving to the opposite one. As can be seen, both the navigation on the pure estimated optical flow (top left) and on the expectation driven preprocessed optical flow (bottom left) allow a collision-free locomotion of the robot KHEPERA through the environment. In contrast, a significant disturbance of the optical flow estimate 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).

sensory predictor generates afterwards the required top-down expectation, which is fused with the sensory bottom-up data stream by a neural field dynamics.

The facilities and usefulness of the expectation-driven preprocessing could be demonstrated by means of a local navigation behavior of the real robot platform KHEPERA. The presented anticipatory neural architecture was able to stabilize the perception process against noise or temporary sensory dropouts in order to maintain a valid representation of the changing and only partially observable environment.

The presented sensory prediction can be very useful for various further tasks, such as the dynamic control of visual attention to regions, where a mismatch of expectation and sensation occurred, or the internal simulation and evaluation of longer action sequences in order to find an optimal action sequence according to the current system state [10].

Future work will address the improvement of the fusion process within the neural field in combination with an extension of the currently unimodal sensory prediction to a multimodal version. That means, that for all hypotheses represented by a non-zero activity within the neural field, a corresponding prediction has to be generated. A system with that ability could evolve many hypotheses in parallel and therefore would be more flexible and powerful.

## References


[1] D. Marr. *Vision*. San Francisco: Freeman, 1982.

[2] R. Pfeifer and C. Scheier. From perception to action: The right direction? In *Proc. PerAc'94*, pages 1–11. IEEE Computer Society Press, 1994.

[3] R. Möller. *Wahrnehmung durch Vorhersage. Eine Konzeption der handlungsorientierten Wahrnehmung*. PhD thesis, Technische Universitt Ilmenau, Fakultt fr Informatik und Automatisierung, 98684 Ilmenau, PF 100565, 1996.

[4] D.T. Cliff. *Computational Neurothology: A Provisional Manifesto*. University of Sussex, School of Cognitive and Computing Sciences, 1990.

[5] M.A. Arbib, P. Erdi, and J. Szentagothai. *Neural Organization: Structure, Function and Dynamics*. MIT Press, 1998.

[6] S.M. Kosslyn, N.M. Alpert, and W.L Thompson. Visual mental imagery activates topographically organized visual cortex: Pet investigations. *Journal of Cognitive Neuroscience*, 5(3):263–87, 1993.

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

[8] S.M. Kosslyn. *Image and Brain: The Resolution of the Imagery Debate*. MIT Press, 1996.

[9] T. Stephan, V. Winkler and H.-M. Gross. Fast and robust prediction of optical flow field sequences for visuomotor anticipation. In *Proceedings of IJCNN'2000, Como, Italy, July 24-27, 2000. Vol. V*, pages 436–441, 2000.

[10] 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.

[11] J.L. Barron, D.J. Fleet, and S.S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision, 12:1*, pages 43–77, 1994.

[12] V. Stephan and T. Winkler. Improvement of optical flow estimates by visuomotor anticipation. In Boehme Gross, Debes, editor, *Proceedings of SOAVE'2000, Ilmenau, Germany, October 5-7, 2000.*, pages 14–21. VDI-Verlag, 2000.

[13] V. Stephan and H.-M. Gross. A neural architecture for expectation driven detection of moving objects during egomotion. In *Proceedings of IJCNN 2001*, 2001.

[14] A.P. Duchon, W.H. Warren, and L.P. Kaelbing. Ecological robotics: Controlling behavior with optical flow. *Proceedings of the 17th Annual Cognitive Science Conference. J.D. Moore and J.F. Lehman (eds.) Lawrence Erlbaum Associates.*, pages 164–169, 1995.