In Proceedings of IJCNN'2001, Washington, D.C., pp 2165-2170, IEEE/Omnipress, 2001

A Neural Architecture for Expectation Driven Detection of Moving Objects During Egomotion

Stephan, V. and Gross, H.-M.
Ilmenau Technical University, Department of Neuroinformatics, P.O.B. 100565,
D-98684 Ilmenau, Germany
{vstephan,homi}@informatik.tu-ilmenau.de

Abstract

In this paper we present an architecture functionally inspired by biology for drawing selective attention to visually unexpected and therefore conspicuous events or image changes. This ability is investigated by means of a mobile-robot scenario, whereby the observing system is moving around and detects other dynamic objects within the scene based on vision only. After detection of an dynamic object our system is able to track it or to realize a higher behavior like hunting or escaping.

1 Introduction

Traditional approaches to visual perception are based on the 'information processing paradigm' [1], which can be characterized by a strict separation between sensory perception and generation of behavior (see [2, 3] for a review).

In recent years, the appreciation of visual perception as a generative sensorimotor process gained increasing acceptance [4, 5]. The generative 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 at different levels of complexity and for different modalities, 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 present a hybrid network architecture to direct selective visual attention to image subregions with significant mismatches between anticipatory expectation and sensory observation.

The detection of dynamic objects in image sequences is well known from literature as "tracking". But most of these systems crucially require a stationary observing system and therefore interpret any changes between subsequent frames as an indicator to a dynamic object. In the following we try to overcome the restriction of the stationary observer and present an architecture, that draws visual attention to dynamic objects in the scene during egomotion of the observing system. In contrast to other tracking approaches of moving objects during camera motion [11], our architecture does not require any a priori knowledge like camera or movement parameters.

2 Experimental framework

For our experiments, we use the real robot platform KHEPERA, a miniature robot equipped with an omni-directional color-camera (see figure 1) to investigate the proposed attentional mechanisms at the level of behavior. Figure 2 depicts a typical environment with the mobile robot inside. The input of our attention drawing architecture is, as indicated in section 1, exclusively visual. Thus, we use the optic flow as a small fraction

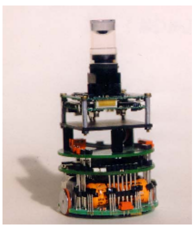

Figure 1: Used robot platform KHEPERA equipped with an omni-directional camera.

Figure 2: View of the environment with the KHEPERA.

of the entire visual input, 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-images we perform a polar transformation (see figure 3 top) to the deskewed form depicted in figure 3 (bottom). These transformed images are used to estimate the optical 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 [12] needs not cope with rotated correlation areas, which would be very time consuming.

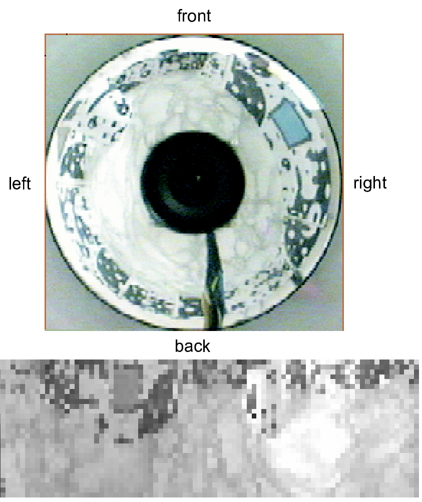

Figure 3: Top: original image taken from the omni-camera mounted on top of the KHEPERA obtained in its position in the environment (see figure 2). Bottom: polar transformed image: middle=front, left and right image borders=back.

The system's goal is first to detect moving objects within the scene. After detection of a dynamic object, the robot has to drive towards or away from it in order to realize a hunting and escaping behavior, respectively. We believe, that the behavior of an autonomous system operating only on this information is a very good indicator of the performance of the system's 'perception' of its environment.

3 Architecture

Figure 4 depicts our architecture to detect dynamic objects. The main principle is a comparison between the really observed and expected optic flow field, which is anticipated based on prior sensorimotor experiences of the mobile system.

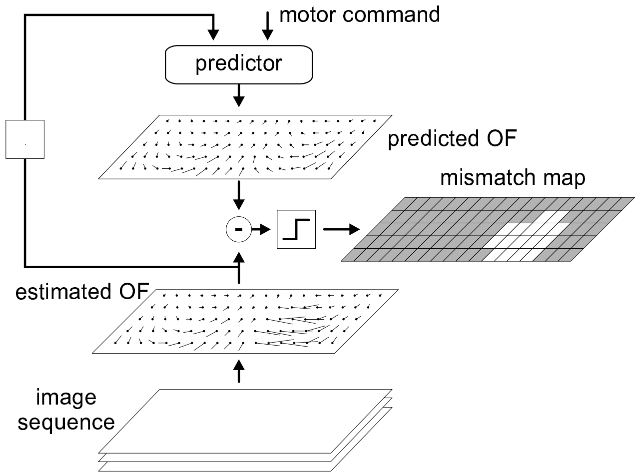

Figure 4: Anticipatory architecture to detect mismatches between the really observed and expected optic flow.

First, based on the current polar-transformed image sequence, the sensory optical flow is estimated in the bottom-up pathway. The sensory predictor anticipates the next expected flow field and operates on the last optic flow field and the currently executed motor command of the robot. That generated top-down expectation is compared to the real observation, whereby the resulting mismatch-map describes large differences between the corresponding flow vectors.

3.1 Sensorimotor prediction

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 a recurrent artificial neural network on really experienced optical flow data of the mobile robot within a static environment. Thus, the applied Jordan-network [13] learns universally valid sensorimotor relationships between subsequent optic flow fields under influence of the systems own motor commands. Consequently, the neural network 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.

3.2 Mismatch detection

To detect large mismatches between the top-down expectation $\vec{f}^E_{pq}(t)$ and the sensory bottom-up observation $\vec{f}^O_{pq}(t)$, the difference Δ_{pq} for all flow vectors is computed by equation 1. A threshold operation with $\theta = 0.67$ defines the value of the corresponding element b_{pq} in the mismatch-map (equation 2).

$$\Delta_{pq} = (f^E_{xpq} - f^O_{xpq})^2 + (f^E_{ypq} - f^O_{ypq})^2 \quad (1)$$

$$b_{pq} = \begin{cases} 1 & : \ \Delta_{pq} > \theta \\ 0 & : \ \text{else} \end{cases} \quad (2)$$

Afterwards, a simple morphological processing eliminates outliers. Finally, an image-region with large discrepancies between the expected and really experienced optic flow fields is obtained.

4 Results

Since the sensory predictor was trained within an static environment, the behavior of dynamic objects is not predictable in this context. Therefore, large discrepancies between the internally expected and actually experienced visual inputs refer to one or more moving dynamic objects in the scene.

4.1 Stationary observer

Despite the recently defined system goal of detection of dynamic objects during egomotion of the observer, we first show the ability of our architecture to detect dynamic objects even without observer-egomotion. Thereto, we use a second KHEPERA, which moves as indicated by the dashed line in figure 5.

The detection results of the middle column were obtained from mismatches between sensory observation and anticipatory expectation. In the right column indicated long optical flow vectors a moving object. This approach is similar to classical tracking systems based on differences between subsequent images. As can be seen, both approaches are able to detect the moving robot very precisely.

4.2 Dynamic observer

Another, more interesting experiment is presented in the following section, where the observing system is moving too. In that case, the classical approach purely

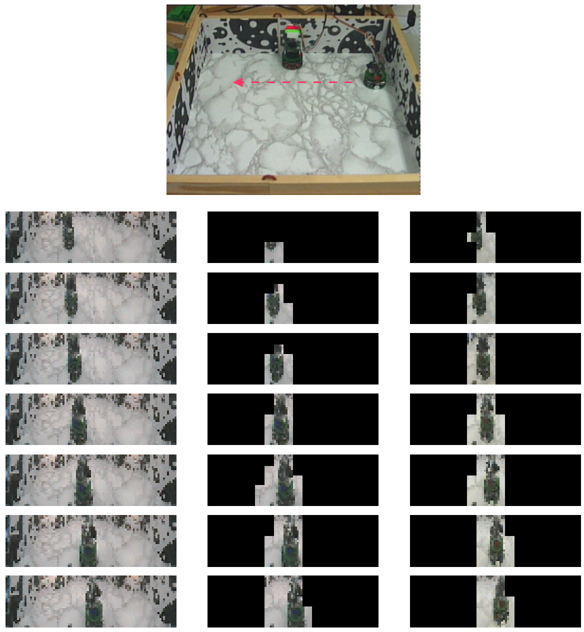

Figure 5: Stationary observer. **Top:** the observing robot remained in place, while the passively moving, second robot is driving from right to left along the dashed line. **Left Column**, from top downwards: sequence of really experienced polar-transformed images. **Middle column:** black areas mark regions without noticeable differences between the anticipatory expectation and observation, the non-black areas mark regions with large mismatches. **Right column:** as in the middle, except that large optic flow vectors triggered selective visual attention.

based on changes between subsequent images must fail, because the egomotion causes permanent image changes. The functioning of our anticipatory architecture in this more realistic case is depicted in figure 6. In this experiment both mobile robots are driving, the passively moving one is locomoting along the dashed, the observing one moves along the solid arrow in the opposite direction. As can be seen, the anticipatory approach is able to track the other moving robot during egomotion of the observer. In contrast, the classical approach exclusively based on large image changes marks in addition to the moving object also regions in the lower image, where the egomotion caused large flow vectors of the ground.

Another scenario is depicted in figure 7. Both robots move with the same speed and direction along the illustrated arrows. In consequence, the passively mov-

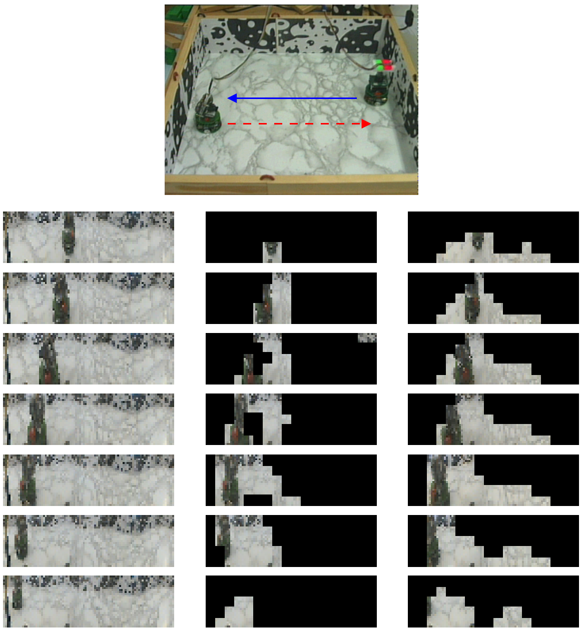

Figure 6: Dynamic observer I. **Top:** the observing robot moves along the solid and the passively locomoting robot along the dashed arrow. **Left Column**, from top downwards: really experienced images. **Middle column:** non black areas mark regions with large mismatches between observation and anticipatory expectation. **Right column:** non-black areas mark regions with large optic flow vectors.

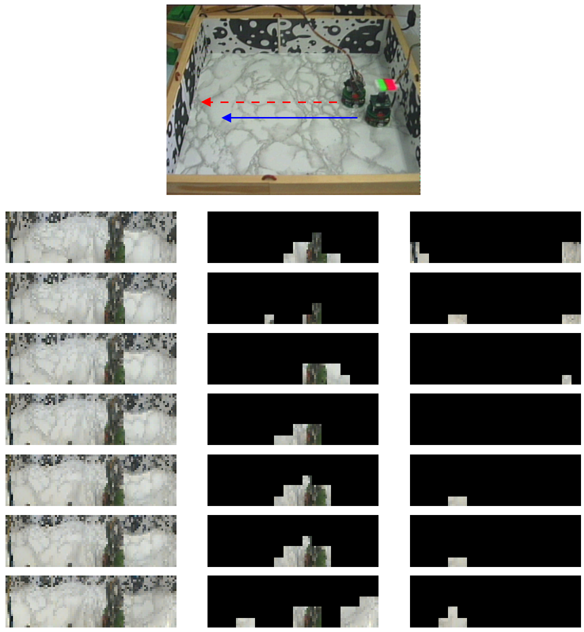

Figure 7: Dynamic observer II. **Top:** both robots move with the same speed and direction along their arrows. **Left Column**, from top downwards: really experienced images. **Middle column:** non black areas mark regions with large mismatches between observation and anticipatory expectation. **Right column:** non-black areas mark regions with large optic flow vectors.

ing robot causes almost no changes in the corresponding region of the observed image. In contrast to previous experiments, in this case the dynamic object is not characterized by large optic flow vectors. Instead, abnormal short vectors in the frontal area mark the other dynamic object in the scene. In consequence, the classical algorithm based on detection of large image changes is as expected absolutely unable to detect the other moving robot in front of the observer. The approach based on the mismatch between anticipatory expectation and sensory observation provides pretty good detection-results, as can be seen in the middle column of figure 7.

4.3 Dynamic observer with static obstacles

After demonstrating the abilities of our anticipatory architecture to detect dynamic objects during egomotion of the observer, the following experiment illustrates, that our approach is able to distinguish between static and dynamic obstacles. This is possible, since the pretrained sensory predictor is able to anticipate the optic flow vectors belonging to static obstacles. Thereto, we used the scenario depicted in figure 8, where in addition to figure 6 a cylindric static obstacle is included in the upper part.

Similar to the experiment without static obstacle (figure 6), the anticipatory approach is able to detect the other moving robot during egomotion of the observer. Furthermore, the cylindric static obstacle in the right part of the polar transformed image causes correctly no large differences between sensory observation and anticipatory expectation and is therefore not marked as dynamic obstacle.

4.4 Further constellations

In addition to the previously presented experiments, we investigated numerous other constellations with and without static obstacles in order to obtain more reliable information about the performance of our approach. In figure 9) there are some scenarios sketched. The approach for selective attention to dynamic objects during egomotion of the observing system produced similar results to the experiments discussed in more detail in the previous sections.

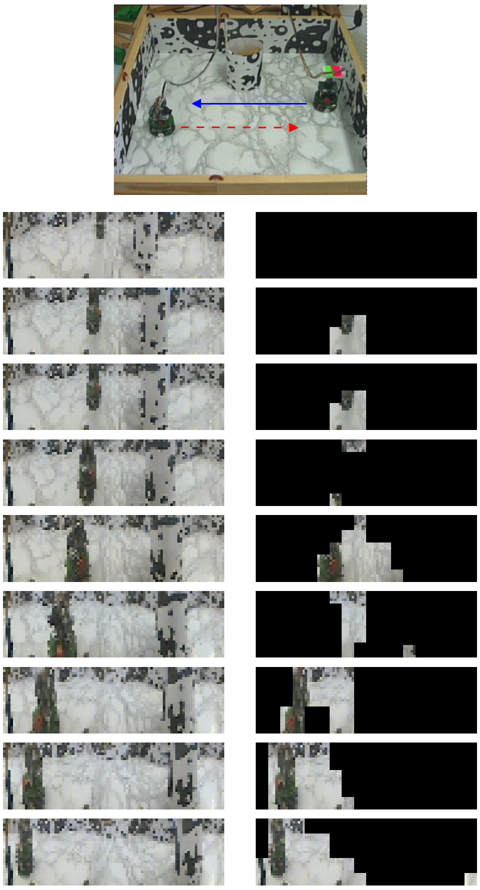

Figure 8: Dynamic observer with static obstacle. **Top:** the observing robot moves along the solid and the passively locomoting robot along the dashed arrow. **Left Column**, from top downwards: really experienced images. **Right column:** non black areas mark regions with large mismatches between observation and anticipatory expectation.

4.5 Realization of higher behaviors

Based on these very encouraging detection-results, we extended our architecture by a very simple behavior generation. Thereto, we computed the center of gravity of the mismatch-map in x-direction c_x and deduced an steering angle ϕ of the observing robot by equation 3, where c_x is the x-coordinate of the center of the mismatch-region in the range between 0.0 (leftmost) and 1.0 (rightmost) within the image.

$$\phi \quad = \quad \alpha \left[c_x - \frac{1}{2} \right] \tag{3}$$

α is a gain constant, that defines, whether the observing robot drives towards or away from the detected dynamic object. If for instance, a dynamic object is

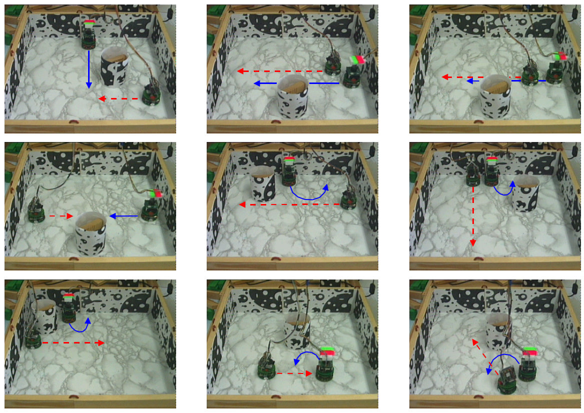

Figure 9: Overview of other investigated scenarios for selective visual attention to dynamic objects during egomotion of the observing system. All scenarios were tested with and without static objects.

detected on the left, $\alpha = 1$ causes a steering angle to the left and vice versa.

Figure 10 shows two different scenarios, wherein the generation of behavior based on the detection of dynamic objects was investigated. As can be seen, the

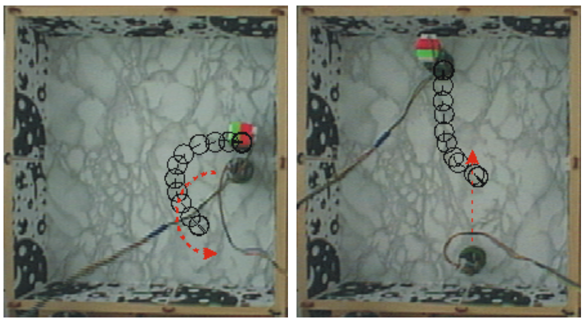

Figure 10: Top view of two scenarios for investigation of behavior generation. The passively moving robot drives along the dashed arrow and the observing system moves along the trace indicated by the black circles.

observing and actively moving robot drives towards the passively moving robot. This kind of hunting behavior can be inverted by using a negative gain constant $\alpha = -1$ to an escape behavior. In that case, the observing robot drives away from any detected dynamic object in its environment. Figure 11 illustrates the resulting escape behavior of the observing robot. As soon as any dynamic object can be detected, a corresponding

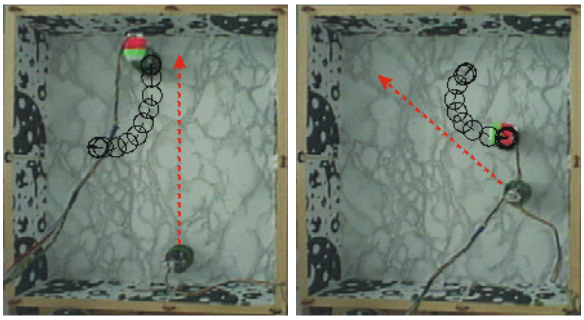

Figure 11: Top view of two scenarios for investigation of behavior generation. Again, the passively moving robot drives along the dashed arrow and the observing system moves along the trace of black circles.

steering command is executed, increasing the distance between the two robots. Due to the omnidirectional visual field of the camera, also objects in the back of the system can be detected (figure 11, right).

5 Conclusion

In this paper we presented an architecture for guiding selective visual attention to dynamic objects during egomotion of the observing system. The presented architecture is inspired by biology at a purely functional level and bases on the ability to learn and predict universally valid sensorimotor relationships.

Through comparison of the predicted sensory outcomes of the systems own actions with the really observed sensory situation, the system can draw its focus of attention to those regions in order to analyse these areas in more detail. Because our system was trained in a static environment, measured significant discrepancies between these two data streams can be interpreted as moving objects. This is possible, since the sensory predictor has learned to anticipate the consequences of its own motor actions for the static objects of the environment. In consequence, absolutely not or at least very poor predictable sensorimotor observations point to dynamic objects. Thus, our anticipative architecture draws, like biological systems, selective attention to visually unexpected and therefore conspicuous events or image changes.

This ability is investigated by means of a mobile-robot scenario, whereby the observing system is moving around and detects other dynamic objects within the scene based on vision only. After detection of an dynamic object our system is able to track it or to realize a higher behavior like hunting or escaping.

This work demonstrates, that learning technical systems inspired by biological vision-mechanisms are able to solve real world problems, like tracking of moving objects during egomotion, without any a priori information about the system itself.

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] S. Feyrer and A. Zell. Detection, tracking and pursuit of humans with an autonomous mobile robot. In *International Conference on Intelligent Robots and Systems 1999*, pages 864–869, 1999.

[12] 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.

[13] M.I. Jordan. Atractor dynamics and parallelism in a connectionist sequential machine. In *Eighth Annual Conference of the Cognitive Science Society*, pages 531–546. Erlbaum, Hillsdale NJ, 1986.