Proc. IEEE-INNS-ENNS Int. Joint Conference on Neural Networks (IJCNN 2000), Como (Italy), 2000, pp. V 436-441

Fast and Robust Prediction of Optical Flow Field Sequences for Visuomotor Anticipation

Volker Stephan, Torsten Winkler, Horst-Michael Gross
Department of Neuroinformatics, Ilmenau Technical University
98684 Ilmenau, Germany
volker.stephan@informatik.tu-ilmenau.de

Abstract

In this paper, we present a hybrid neural architecture to predict optical flow fields as consequences of real and hypothetical actions. In this architecture, we introduce a neural field-based method to fuse sensory bottom-up and predicted top-down expectations. All subsystems extensively use confidence estimations to reduce disturbances caused by noise. The facilities of this anticipative preprocessing can be demonstrated by means of an optical flow field based local navigation behavior of the miniature robot KHEPERA. Our anticipative preprocessing enables the robot to bridge gaps of sensory dropouts and, in consequence, to avoid collisions even with very noisy sensory information.

Keywords: sensory prediction, world model, expectation, sensor fusion, sensorimotor anticipation

1 Introduction

Traditional approaches to visual perception are based on the 'information processing paradigm' [9], which can be characterized by a strict separation between sensory perception and generation of behavior (see [11, 10] for a review). In recent years, the appreciation of visual perception as a generative sensorimotor process gained increasing acceptance [3, 1]. The generative aspect of perception has been emphasized especially by [7, 8, 6] who supposed that internal simulation and mental imagery may play an integral role in perception, helping one not only to recognize objects but also to anticipate the consequences of events. 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 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 [5].
In this paper, we present a hybrid network architecture to predict optical flow fields and demonstrate its functionality in a KHEPERA-navigation task. This is done by means of a fusion of sensory bottom-up and expectation-based top-down information. This is a kind of anticipative preprocessing embedded in a cognitive processing cycle of hypotheses generation and verification [8, 6].

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 left). The system's goal is a collision-free local navigation only based on visual information, in this case the optical flow fields. We use the optical flow, because it is largely independent of specific visual details of the objects in the scene and yields implicit information about spatial distances to objects.
In the preprocessing we perform a polar transformation of the original omni-camera-image (see Figure 2 left) to the deskewed form depicted in Figure 2 (right). These transformed images are used directly 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 [2] must not cope with rotated correlation areas, which would be very time consuming.

Fig. 1: Used robot platform KHEPERA equipped with an omni-directional camera (left). Top-view of the environment with the KHEPERA in starting position (right).

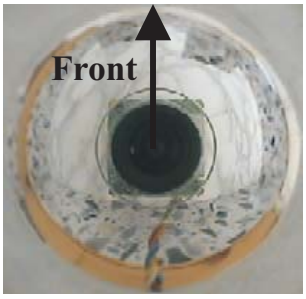

Fig. 2: Left: original image of the omni-camera mounted on top of the KHEPERA obtained in its starting position in the environment (see Figure 1 right). Right: polar transformed image: middle=front, left and right image borders=back.

3 Architecture

As introduced in section 1, we use a hybrid architecture to predict the optical flow fields as a result of the last optical flow field and the real or hypothetical action to be executed. To demonstrate the functionality in a KHEPERA-scenario, we fuse the sensory bottom-up estimation and the top-down expectation in order to reduce the noise and gain robustness against sensory dropouts (see Figure 3).
A central aspect of our anticipative preprocessing in the bottom-up top-down cycle is the usage of flow vector specific confidence estimates organized topographically manner corresponding to the flow field.
These confidence-values of each flow vector are based on the optical flow estimation by evaluating the shape of the correlation function. Sharp and unique minima cause high, whereas flat or ambiguous correlation functions result in low confidence values.

3.1 Expectation generation

The sensorymotor prediction is of central importance for our approach. In previous approaches [5], we used standard neural networks, such as multilayer-perceptrons or a mixture of experts consisting of several action-specific perceptrons. In some cases, these networks had prediction problems, especially if the incoming data were distributed badly. In our present view, the key problem of the used neural networks was the prohibitively high dimensionality of sensory input, the whole optical flow field. A succeeding single optical flow vector, however, depends only on a very small part of the current flow field, but never on the whole field. Hence, a network with completely connected layers first has to find the respective 'source-region' and thereafter to learn to predict the corresponding flow vector for each position of the flow field.

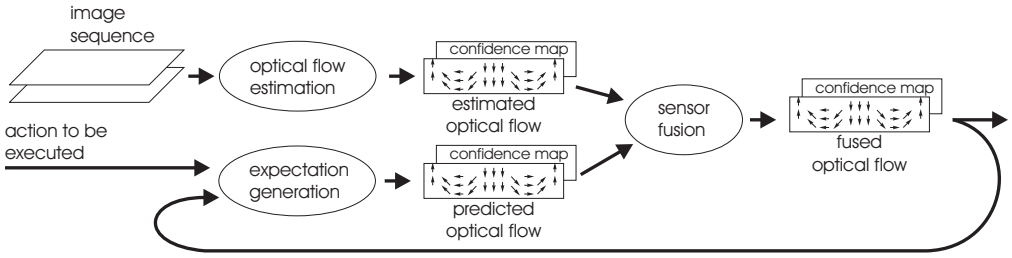

Fig. 3: Hybrid architecture to fuse the sensory bottom-up and the top-down expectation.

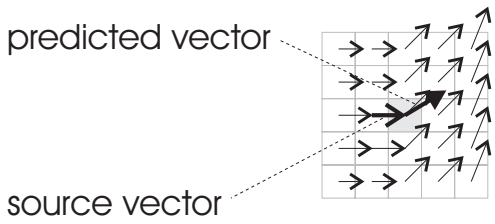

Fig. 4: The flow vector to be predicted depends only on those flow vector(s), that point closest to the position for which the flow vector is to be predicted. That is, because these vectors describe the velocity of the corresponding objects in the scene.

Hence, we developed an alternative approach to overcome this problem. It uses the optical flow inherent property to represent the movements of objects onto the camera-plane. Consequently, the inverse optical flow itself points to that part of the image, where movement must be predicted (see Figure 4).
Equation 1 shows the computation rule for a source-vector $\underline{f}^S_{pq}$ at position (p, q) within the optical flow field. This is a superposition of all vectors $\underline{f}^M_{kl}$ depending on the superposition weight r_{pqkl}. r_{pqkl} reflects, how exactly these optical flow vectors point to the current position $(p, q)^T$ and is normalized by the sum of all superposition weights (see equation 5). The search for this source-vector among all flow vectors $\underline{f}^M_{kl}$ can be reduced to a small region around the current position (k, l), where the size of this search-window corresponds to the size of the search-window of the optical flow estimation determined by n. The actual predicted vector $\underline{f}^P_{pq}$ is thereafter only a scaling in x and y-direction of the source-vector by the weights $w_{xpq}(t)$ and $w_{ypq}(t)$ (equation 2). The confidences c^P_{pq} depend on the confidence c^S_{pq} of the source-vector and on the confidence of the prediction itself w^c_{pq} (equation 4).

$$\underline{f}^S_{pq} = \sum_{k=p-n}^{p+n} \sum_{l=q-n}^{q+n} \underline{f}^M_{kl} \cdot r_{pqkl} \tag{1}$$

$$\underline{f}^P_{pq} = \begin{pmatrix} w_{xpq}(t) & 0 \\ 0 & w_{ypq}(t) \end{pmatrix} \underline{f}^S_{pq} \tag{2}$$

$$c^S_{pq} = \sum_{k=p-n}^{p+n} \sum_{l=q-n}^{q+n} c^M_{kl} \cdot r_{pqkl} \tag{3}$$

$$c^P_{pq} = w^c_{pq} \cdot c^S_{pq} \quad \text{with} \tag{4}$$

$$r_{pqkl} = \frac{\sum_{u=p-n}^{p+n} \sum_{v=q-n}^{q+n} \left\| \begin{pmatrix} u \\ v \end{pmatrix} + \underline{f}^M_{uv} - \begin{pmatrix} p \\ q \end{pmatrix} \right\|}{\left\| \begin{pmatrix} k \\ l \end{pmatrix} + \underline{f}^M_{kl} - \begin{pmatrix} p \\ q \end{pmatrix} \right\|} \tag{5}$$

The update of the weights depends on the difference between the predicted vector $\underline{f}^P_{pq}(t)$ and the actually experienced in the next time step $\underline{f}^M_{pq}(t+1)$ (equation 6). The confidence-weights w^c_{pq} are updated according equation 7.

$$\begin{aligned} \Delta w_{xpq}(t) &= \eta \cdot \left(f^M_{xpq}(t+1) - f^P_{xpq}(t) \right) \cdot f^S_{xpq}(t) \\ \Delta w_{ypq}(t) &= \eta \cdot \left(f^M_{ypq}(t+1) - f^P_{ypq}(t) \right) \cdot f^S_{xpq}(t) \end{aligned} \tag{6}$$

$$\Delta w^c_{pq}(t) = \eta \cdot e^{-\frac{\left\| \underline{f}^M_{pq}(t+1) - \underline{f}^P_{pq}(t) \right\|}{2}} - w^c_{pq}(t) \tag{7}$$

3.2 Fusion

With regard to Figure 3, in this section we present the fusion between bottom-up and top-down information (see Figure 5). Each vector of the whole field is represented by a 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 activation of the blobs in the neural field is a measure for the corresponding confidence of this local flow vector. Due to this 2-dimensional representation, it is possible to hold many alternative hypotheses (blobs) for each flow

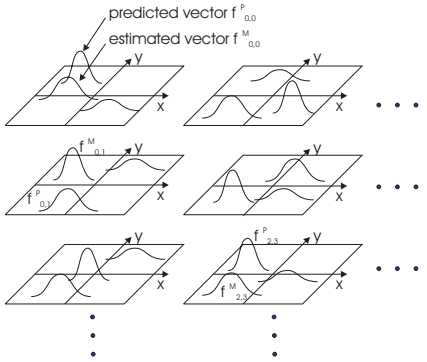

Fig. 5: Each vector of the optical flow field is represented by a 2-dimensional neural field, where the position within the neural field codes possible flow-vectors as blobs, and the activation of the blob is a measure for the corresponding confidence of that respective optical flow vector.

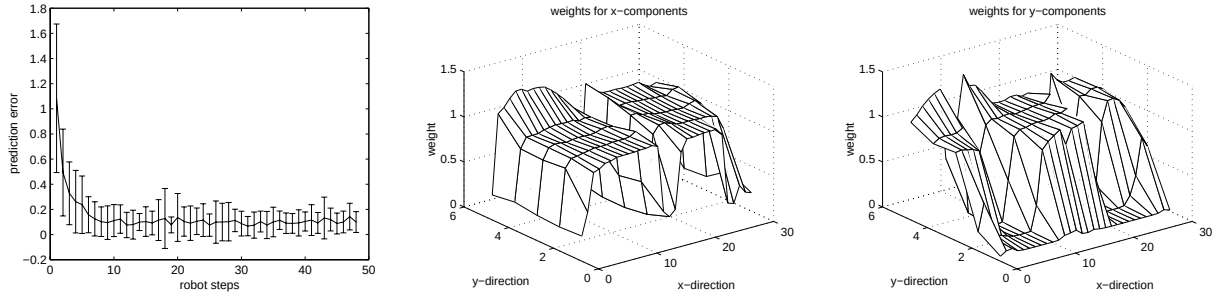

Fig. 6: (Left) development of prediction error during training. Arrays of weights of our sensory predictor for the x (middle) and y-components (right) of the optical flow vector field. For detailed explanations see text.

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. The output results from the hypothesis with the highest confidence (equation 8). 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 9, 10). Equation 9 shows, that the new state $\underline{z}^x_{pq}(t+1)$ is computed by discounting the last state $\underline{z}^x_{pq}(t)$ by $\alpha \in (0 \ldots 1)$ and the superposition of the sensory bottom-up vector $\underline{g}(f^E_{xpq}(t))$ and the top-down expectation $\underline{g}(f^P_{xpq}(t)$ in form of 1 dimensional blobs (equation 11) weighted by their confidences $c(.)$

$$\underline{f}^M_{pq}(t) = \begin{pmatrix} argmax_x(\underline{z}^x_{pq}(t)) \\ argmax_y(\underline{z}^y_{pq}(t)) \end{pmatrix} \quad (8)$$

$$\underline{z}^x_{pq}(t+1) = \alpha \underline{z}^x_{pq}(t) + \underline{g}(f^E_{xpq}(t)) \cdot c(f^E_{xpq}(t)) + \underline{g}(f^P_{xpq}(t)) \cdot c(f^P_{xpq}(t)) \quad (9)$$

$$\underline{z}^y_{pq}(t+1) = \alpha \underline{z}^y_{pq}(t) + \underline{g}(f^E_{ypq}(t)) \cdot c(f^E_{ypq}(t)) + \underline{g}(f^P_{ypq}(t)) \cdot c(f^P_{ypq}(t)) \quad (10)$$

$$g_k(u) = e^{-\frac{(u-k)^2}{2\sigma^2}} \quad (11)$$

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.

4 Exemplary results

To train the flow field predictor, we put the robot KHEPERA into its starting position depicted in Figure 1 (right) several times and drove with fixed speed straight forward up to the opposite wall. During this training period, the KHEPERA experienced several optical flow field configurations with obstacles on the left, on the right, and, finally, also in front of the robot. Figure 6 shows the decreasing prediction error over the training period (left) and the learned weight matrices (middle and right). As can be seen, the predictor

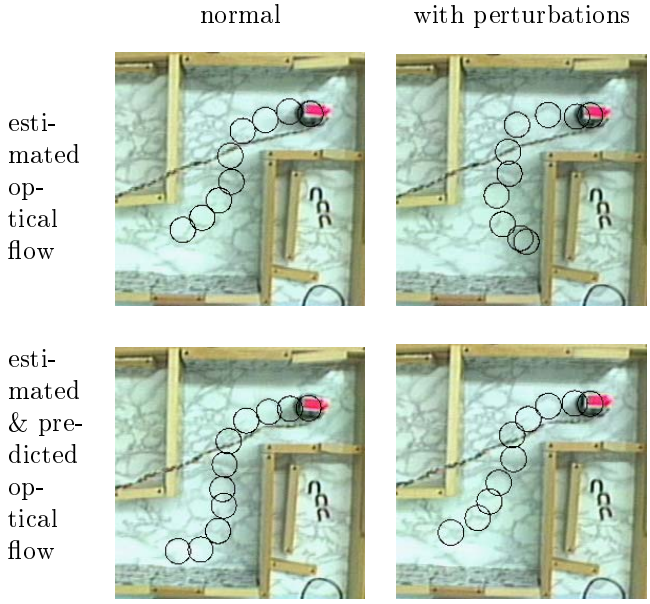

Fig. 7: Navigation based on the estimated optical flow applying the well known balancing approach [4]. As can be seen, both the navigation on the pure estimated optical flow (top left) and on the expectation driven preprocessed optical flow (bottom left) allow a collision-free locomotion of the robot KHEPERA through the environment. In contrast, a significant disturbance of the optical flow estimation by means of fluctuating ambient light causes a collision at the end of the plotted trace, where no anticipative preprocessing is applied (top right) (bottom right). The anticipative preprocessing overcomes the problems and allows a collision-free locomotion (bottom right)

amplifies the y-components (Figure 6, right) in front (around an x-direction of 15) and in the back (left and right border in x-direction) of the robot. The x-components (Figure 6, middle) are amplified in the lateral parts of the polar transformed image (x-directions $5 \ldots 10$ and $20 \ldots 25$).
To demonstrate the facilities of the presented anticipatory preprocessing, we also placed the robot in unknown environments to navigate without collisions through the narrow passage. For this benchmark , we used the balancing approach [4], 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 7 (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 usage of pure estimated optical flow fields fails, because the very noisy sensory input entails no information about near obstacles. 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.
At this point, we have to ask the question: would a pure feedback without any sensory prediction (see Figure 3) result in the same behavior? In this case, the fusion of the noisy and very unconfident estimated optical flow and the relative confident expectation would return the last fusion output. Thus, such an architecture is nothing but a low-pass filter over time. The advantage of our anticipative preprocessing in contrast to this approach is depicted in Figure 8. In this case, the robot stood in front of an obstacle at a distance of about 10cm and drove with slight right turns avoiding a collision. As can be seen, the system with anticipative preprocessing is able to recursively predict the shift of the central large flow vectors representing the close obstacle to the left, whereby the system without sensory prediction once again can only store the last sensory situation, which becomes more and more obsolete over time.

5 Conclusions and Outlook

In this paper, we presented a hybrid neural architecture to predict optical flow fields as consequences of actions. Further, we introduced a neural field-based method to fuse sensory bottom-up estimations and top-down expectations. All proposed subsystems extensively use confidence measurements in order to prevent disturbance by noise. The facilities of this anticipative preprocessing could be demonstrated by means of a local navigation behavior of the real robot platform KHEPERA. The presented sensory prediction can be very useful for various 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 many action sequences in order to find an optimal action sequence according to the current system state [5].
Future work will address the improvement of the sensory prediction. The current network causes problems

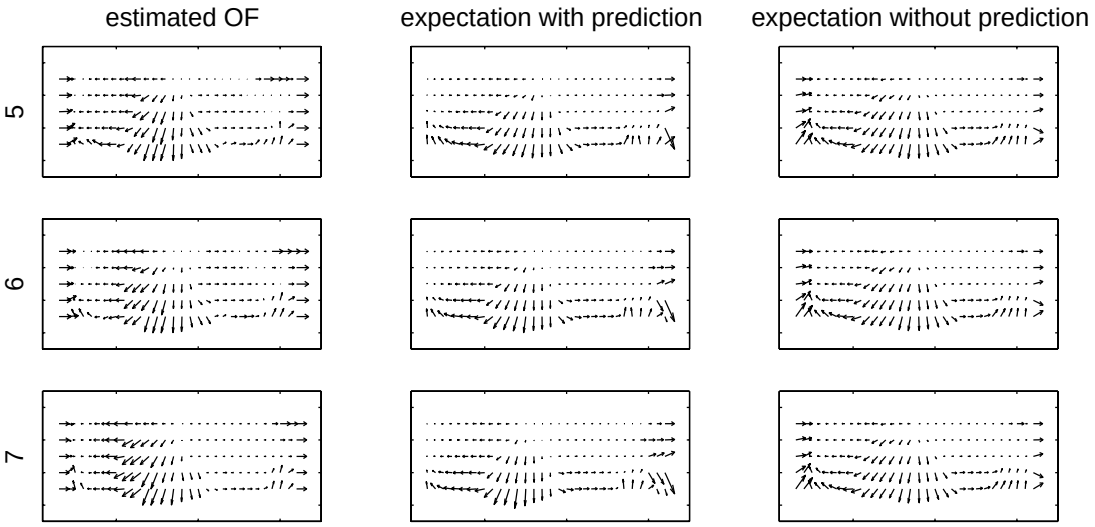

Fig. 8: Sequences of optical flow fields for a turn to the right: (left) sequence of real flow fields estimated in 3 subsequent steps of movement; (middle) internally simulated sequence of flow fields starting at the real flow field in time step 5. Each predicted flow field is the result of a confidence controlled top-down superposition of the last prediction and the succeeding one. (Right) same as in the middle, except that in this case the top-down expectation is the last fused optical flow field instead of predicted. As can be seen, this architecture is unable to generate expectations reflecting the changes of the environment caused by the executed actions. Hence, the sensory predictor is an essential part of our anticipative preprocessing.

with large steering angles and cannot cope with different speeds of the robot. Moreover, problems emerging from object occlusions have to be solved, to allow the robust prediction of longer sequences in order to apply the predictor network to our model for anticipation based on sensory imagery.

References

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

[2] 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.

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

[4] 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.

[5] 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.

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

[7] 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.

[8] 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.

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

[10] R. Moeller and H.-M. Gross. Perception through Anticipation. In *Proc. PerAc'94 - From Perception to Action*, pages 408–411. Los Alamitos: IEEE Computer Society Press, 1994.

[11] R. Pfeifer and C. Scheier. From Perception to Action: The Right Direction? In *Proc. PerAc'94*, pages 1–11. IEEE Computer Society Press, 1994.