Intern. Conference on Artificial Neural Networks (ICANN 2001), Vienna,2001, pp. 885-890, Lecture Notes in Computer Science, vol. 2130, Springer Verlag 2001

Neural Architecture for Mental Imaging of Sequences based on Optical Flow Predictions

V. Stephan and H.-M. Gross

Ilmenau Technical University, Department of Neuroscience, P.O.B. 100565,
D-98684 Ilmenau, Germany
{vstephan,homi}@informatik.tu-ilmenau.de

Abstract. In this paper we present a neural architecture for a mental imaging like generation of image sequences. Mental imaging plays a central role for various perception processes. Thereto, we investigated mechanisms to model this ability of biological systems at a functional level for sequences of images. Because it is impossible to memorize many experienced sequences, we developed an universal, general and very powerful approach based on the ability to predict optic flow fields as consequences of the systems own actions and tested the resulting architecture on a real mobile system.

1 Introduction

Mental imaging is understood as a process, which generates conscious images in mind without any sensory stimulation[1]. This is possible, since experienced sensory (visual) impressions are stored by some neural memory structure and can afterwards be reconstructed.

In literature there is broad consens about the central aspect of these imaging processes. Many authors [1–4] propose, that perception and mental imaging share common mechanisms and support each other. In [5–7] there is described, that mentally generated images are used as an expectation for what the system sees. This so called top-down expectation is then used to enhance the sensory bottom-up data in order to become more robust against noise or sensory dropouts. Other authors postulated, that mental imagery is also used for internal simulation of the consequences of hypothetically executed actions in order to find an optimal action sequence [5, 8, 9].

In general, we and almost any perceiving biological system do not operate on sensory snapshots. Instead, we operate on a continuous flow of data, which requires a continuous flow of corresponding expectations. The generation of single mental images already requires a large memory to store the experienced images. The extension to *sequences* of mental images seems to be unrealizable through memorization, because of the exponential growth of data to be stored. In consequence, a more general approach is required.

By learning universally valid sensorimotor relationships between subsequent images and the executed motor actions of the mobile system such an approach could be developed. The resulting neural architecture is able to anticipate sequences of images, which are expected after the hypothetical execution of a given action trajectory. This ability can be utilized to generate the desired continuous flow of top-down expectations or to realize an internal simulation.

2 Scenario

We investigated our approach within a scenario with the miniature robot Khepera equipped with an omnidirectional camera. Using a real robot in favor to stored image sequences from any video stream is very important, since the executed action *causes* the changes within the image and thus defines the direction of the development of the whole image sequence. This emphasizes the fact, that perception does not end in itself, instead it is a sensorimotor process integrating the generation of behavior [10]. For our experiments we used the scenario depicted in figure 1. The omnidirectional camera images are polar-transformed and then used to estimate optic flow using a correlation based method [11].

Fig. 1. Scenario used for our experiments with the mobile robot Khepera equipped with an omnidirectional camera.

3 Architecture

Our approach to mental imaging of sequences avoids the need of memorization of all experienced image sequences by using an iterative process of subsequent image transformations. Therefore, the required image transformation represents the changes within the image that are caused by egomotion of the observing system. If this transformation can be found, the system can generate sequences of subsequent sensory visual impressions starting from a real or an initially memorized image.

3.1 How to find this Transformation?

The required transformation has to comply the following criteria. It must be invariant to color, shape, texture, and other object or scene specific features. In contrast, motion parameters of the mobile system and the spatial configuration of the objects within the scene are very important. With respect to these requirements, optic flow seems to be well suited for this task, since it exclusively represents spatial relations and movements of objects in the scene caused by egomotion of the system.

To generate a hypothetically image sequence for a given action sequence, an anticipation of the action consequences is crucially required. Instead of realizing a scene depending mapping from the current image $I(t)$ to the following one $I(t+1)$, our system learns the scene-independent mapping from past optic flow fields and actions to the following one by a time-delayed neural network (TDNN):

$$OF(t-n)\dots OF(t) \quad \times \quad a(t-n)\dots a(t) \quad \mapsto \quad OF(t+1)$$

Thus, the resulting image transformation depicted in figure 2 is based on universally valid sensorimotor relationships, represented by optic flow.

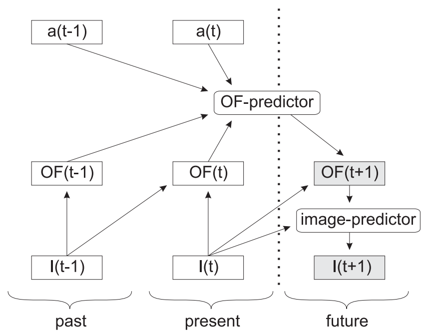

Fig. 2. Architecture for mental imaging. Based on universally valid sensorimotor relationships the next optic flow $OF(t+1)$ is predicted by a neural network (TDNN). The image-prediction uses the predicted optic flow $OF(t+1)$ to transform the current image $I(t)$ to the mentally generated image $I(t+1)$.

3.2 Image Generation

Based on the predicted optic flow field $OF(t+1)$ and the current image $I(t)$ it is possible to generate the predicted image $I(t+1)$. This is realized by shifting the pixels along their corresponding predicted flow field vectors. First, for each pixel of the new image $I(t+1)$ a source region in $I(t)$ is computed. For this purpose those flow vectors $\underline{f}^{pq}$ are searched, which point most closely to the new pixel position (i,k) (figure 3).

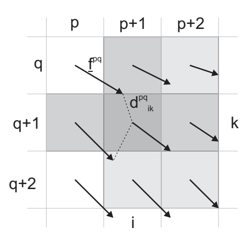

Fig. 3. Computation of distance d_{ik}^{pq} (dashed line) between the end-point of the neighbored optic flow vector $\underline{f}^{pq}$ originating from the pixel at position (p,q) and the pixel-coordinates at position (i,k).

To compute the source region for the new pixel at position (i,k) all distances d_{ik}^{pq} of the flow vectors $\underline{f}^{pq}$ out of the neighborhood N to the new position (i,k) are computed (equation 1). Then, the color of the new pixel $\underline{c}_{i,k} = (r,g,b)^T$ results from a superposition of all pixels out of neighborhood N weighted by a distance-dependent factor (equation 2).

$$d_{ik}^{pq} = \sqrt{(i - f_x^{pq} + p)^2 + (k - f_y^{pq} + q)^2} \qquad (1)$$

$$\underline{c}_{ik} = \sum_{p,q \in N} \frac{1}{(d_{ik}^{pq})^8} \cdot \underline{c}_{pq} \tag{2}$$

Thus, an interpolated transformation of the image $I(t)$ to the predicted image $I(t+1)$ is defined, which uses the predicted optic flow $OF(t+1)$.

4 Results

We tested the above described image transformation approach for several image sequences. First, we put the robot into the situation illustrated in figure 4 top. At each time step the process of mental imaging generates a sequence of im-

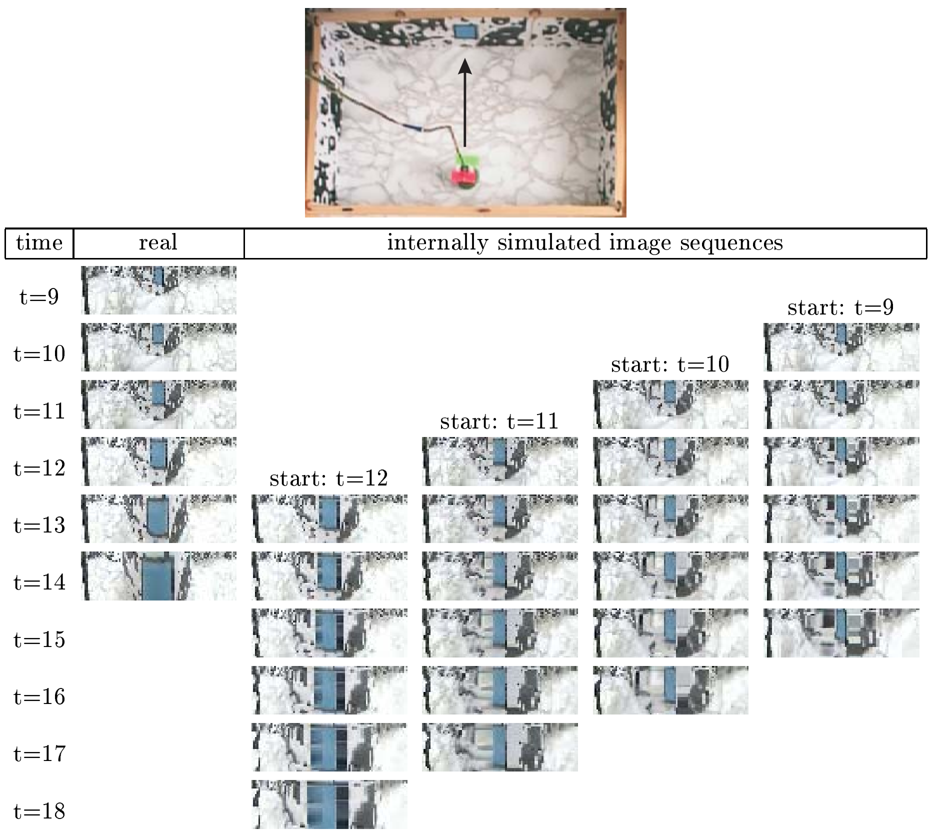

Fig. 4. Internal simulation of images realized by our anticipatory system during straight ahead movement of the mobile robot towards the blue rectangular mark (top). The really experienced image sequence (leftmost column) plotted since time $t = 9$ and four image sequences with different starting points, generated by means of our mental imaging approach, are plotted on the right.

ages (vertical columns), that reflect the expected image changes caused by the locomotion of the mobile system.

The leftmost column of really experienced images shows the growing rectangular blue mark in the image caused by the oncoming opposite marked wall. The rightmost mentally generated sequence based on the real image at time step $t = 9$ shows in contrast to the real sequence only a vertical growth of the central blue mark. The later the starting point for mental imagination, the better the similarity to the last really experienced image at time $t = 14$ can be seen. Likewise, the image quality decreases over simulation time, since both optic flow prediction errors and image interpolation errors are accumulated over simulation time.

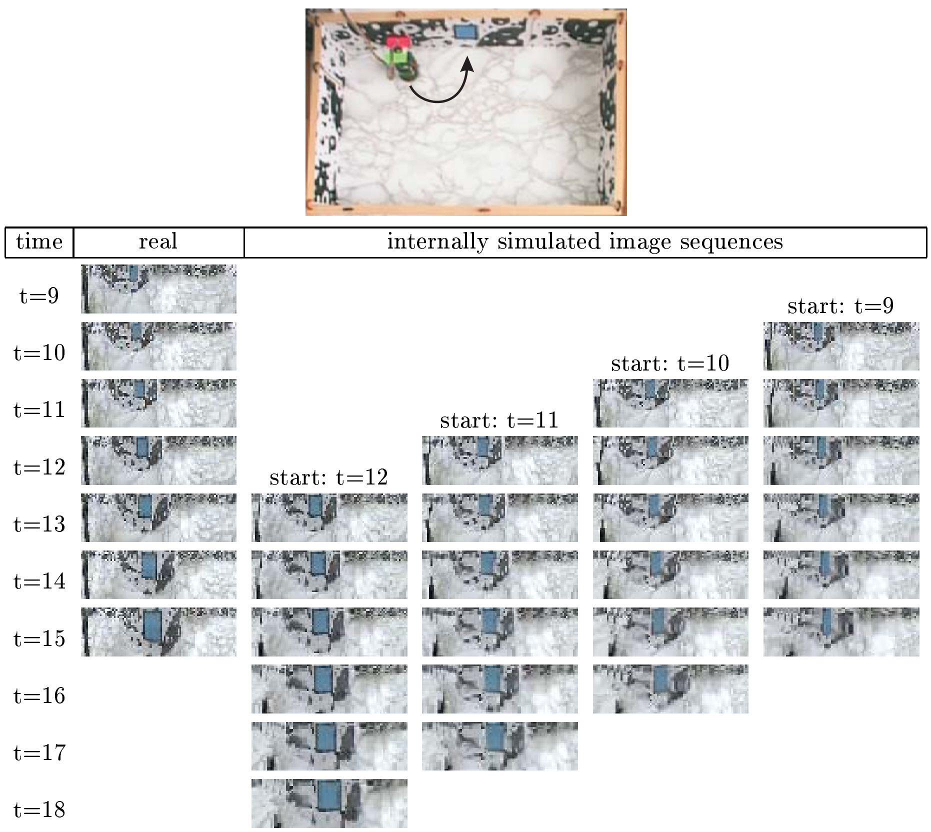

Fig. 5. Internal simulation of images realized by our anticipatory system during a left-turn movement of the mobile robot towards the blue rectangular mark (top). Image arrangement as in figure 4.

Nevertheless, the principal ability of our approach to internally generate image sequences representing the course of the changing reality caused by the systems actions can be demonstrated.

Similar results are depicted in figure 5, where the robot drove a left turn towards the blue mark. In contrast to the previous experiment, the driven curve causes a horizontal shift of the polar-transformed image over time, which also is predicted by our mental imaging architecture.

5 Outlook

The presented neural architecture for optic flow-based mental sequence-imaging internally generates image sequences representing the consequences of a mobile systems own hypothetically executed actions. These image sequences can be used to generate the desired continuous flow of top-down expectations or to realize an internal simulation process for action planning on image sequences.

In future work we intend to investigate these challenging anticipative mechanisms for mobile systems at the behavioral level in more detail.

References

1. M.J. Farah. The neurobiological basis of mental imagery: A componential analysis. *Frontiers in Cognitive Neuroscience*, pages 559–573, 1992.
2. S.M. Kosslyn. *Image and Brain: The Resolution of the Imagery Debate.* MIT Press, 1996.
3. J. Demiris and G. Hayes. Active imitation. In *AISB'99 Symposium on Imitation in Animals and Artifacts*, 1999.
4. L. Deecke. Planning, preparation, execution and imagery of volitional action. *Cognitive Brain Research*, 3(2):59–64, 1996.
5. 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.
6. Y. Miyashita, M. Ohbayashi, M. Kameyama, Y. Naya, and M. Yoshida. Origin of visual imagery: role of temporal cortex and top-down signal. In M. Ito and Y. Miyashita, editors, *Integrative and Molecular Approach to Brain Function*, pages 133–145. Elsevier, 1996.
7. V. Stephan and H.-M. Gross. Improvement of optical flow estimates by visuomotor anticipation. In *Proceedings of Workshop Dynamische Perzeption, Ulm, Germany, 2000.*, pages 191–194. akademische Velagsgesellschaft, infix, 2000.
8. A.R. Damasio and H. Damasio. Making images and creating subjectivity. In R. Llinas and P. Churchland, editors, *The Mind-Brain Continuum*, pages 19–27. MIT Press, 1996.
9. 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.
10. R. Pfeifer and C. Scheier. From perception to action: The right direction? In *Proc. PerAc'94*, pages 1–11. IEEE Computer Society Press, 1994.
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.