ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 345-350

Perception and Action Selection by Anticipation of Sensorimotor Consequences

T. Seiler, V. Stephan, H.-M. Gross

Technical University of Ilmenau
Department of Neuroinformatics
{torsten, vstephan, homi}@informatik.tu-ilmenau.de

Abstract. The anticipatory approach presented here is based on the prediction of sensory consequences of hypothetically performed actions. With the prediction of consequences of actions, it is possible to evaluate sensory sequences. Prerequisite for this is a sufficient prediction of the sensory consequences of executed actions. Based on this anticipatory approach, we propose a neural architecture that is able to evolve an initially reactive behavior into a planning and forecasting behavior.

1. Introduction

The basic idea of the anticipatory approach presented in [7, 8] is to avoid the separation of perception and generation of behavior and to fuse both parts into one neural process. It seems to be more realistic to characterise the visual scenery immediately in categories of behavior. Perception is considered to be the internal simulation of a number of actions and the anticipation of their consequences. On one hand, these hypothetical actions and their anticipated consequences characterise the sensory situation. On the other hand, from this set of descriptive actions those can be selected for execution, which result in a positive effect concerning to the system goal.

For enumerable states and actions there exist some similarities to Sutton's DYNA architecture [11]. In contrast to DYNA our architecture deals with a quasi-continuous action space and uses environment models at the sensory level. Other approaches try to solve the problem of action selection for a given sensory situation by introducing evaluation signals to describe the quality of alternative action sequences with respect to the current sensory situation and the system goal (e.g. Q-Learning [12]). The disadvantages are that learning is very costly and each learned action value represents a complete sequence of experienced state-action transitions. It is not possible to evaluate a composite sequence generated from action sequences that are only partly known, but have never been seen as a whole before.

This work is supported by DFG-Grant(Gr 1378/2-1)

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 345-350

2. Architecture

We understand perception as an active process of generating sensory hypotheses based on detected sensorimotor relations. Hypothesis generation for sequences requires an adequate representation of alternative actions. In order to evaluate action sequences we have to search in a quasi–continuous sensorimotor space. To reduce the search space we select only promising actions using an extended neural field approach and a fixed search depth in our neural architecture.

The sensory information is the optical flow from a monocular video sequence. We use it because of its implicit information about spatial distances to objects in the environment and the causing action. Regardless of typical optical flow problems DUCHON et al. [3] and KRÖSE et al. [5] showed that it is possible to navigate and avoid obstacles using the optical flow only. To estimate the optical flow we use an region–based correlation approach by CAMUS[2].

Figure 1 depicts the principle structure of a *prediction module* consisting of an *action suggestion*, an *optical flow prediction*, an *action selection* and a *hypothesis evaluation*. The *action suggestion* generates a topological motor map that codes a set of alternative actions. A location in this map corresponds to a specific action within a two–dimensional action space (speed and steering angle), while the activation of the corresponding neuron represents the learned evaluation of this action. The *action suggestion* is implemented as a neural function approximator based on an adaptive vector quantization technique in conjunction with a simple reinforcement learning mechanism.

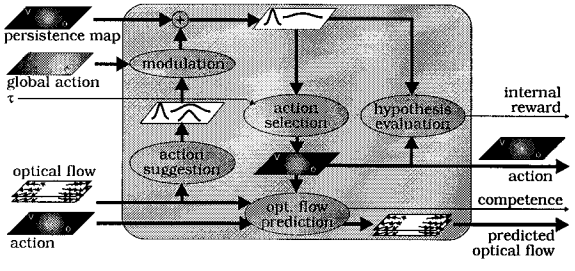

Figure 1: Structure of the Prediction Module. For explanations, see text.

In order to integrate the *prediction module* into a global behavior–based system (fig. 2), we modulate the motor map with a global motor map (fig. 1).

To select a single action within the motor map we use an extended neural field with sequential selection behavior [1, 10]. This mechanism selects the most promising action first and later less promising actions with respect to their evaluation, if the available time for internal simulation allows this. The *optical flow prediction* computes the sensory consequences of the selected actions using the current sensory input. We use a modified "mixtures of expert" approach realized by action–specific perceptrons gated by a neural vector quantization technique [6]. We employ a piecewise linear transformation of the current optical flow field to approximate the succeeding one. The adaptation is performed after the execution of a "real" action by comparing the real and the predicted sensory situation. The quality of prediction is stored in a situation and action

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 345-350

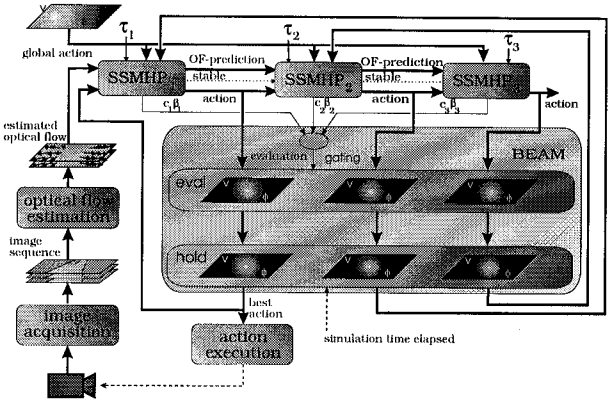

Figure 2: Chaining of prediction modules for simulation of sequences of actions.

specific competence. The *hypothesis evaluation* computes an evaluation signal based on the selected action.

The adaptation of the two pathways, *action suggestion* and *optical flow prediction*, is independent of each other. Hence, a non–optimal selected action is no problem for the training of the sensory prediction. On the other hand, an erroneous prediction has no influence on the *real reward* signal. Only the selection process is affected. A single *prediction module* is able to sequentially generate several individual actions and to estimate resulting sensory consequences together with corresponding evaluations. By chaining of multiple prediction modules working on staggered time scales it is possible to generate sequences of actions and their sensory predictions (Figure 2). The first prediction module works on the real sensory input, the second on the prediction of the first and so on. Each successive module is only given that limited time to generate alternative hypotheses during which the output of its predecessor remains stable. The choice of the parameters for the internal dynamics controls the breadth of search. The maximum depth of search depends on the number of replicated prediction modules organized in this temporal hierarchy. A stable action sequence is fed into a short–term memory called *best evaluated action memory* if its evaluation is higher than that of the stored one. After the simulation time elapsed, the best simulated action sequence is copied to the corresponding prediction modules (fig. 2) in order to prefer the best simulated action sequence in the next simulation. The first action of this sequence is executed. Our mechanism guarantees that after a setup time the process can be interrupted at any time. It puts out that action sequence which has been judged best up to this point. Therefore it can be considered as a neural implementation of an anytime algorithm.

3. Experimental Results

First we investigated the reactive behavior. Therefore we trained the *action suggest* module to satisfy the reinforcement function shown in figure 4. The desired behavior is a straight–ahead motion that is as fast as possible. Three

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 345-350

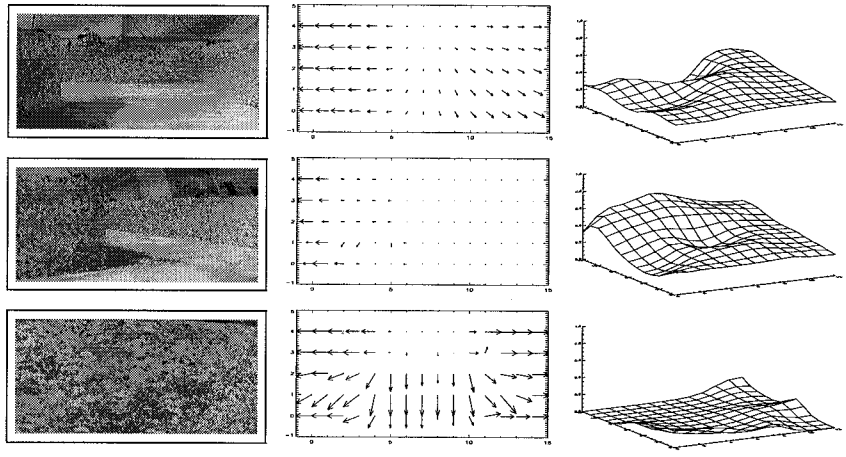

Figure 3: Three different sensory situations (left), the corresponding optical flow fields (middle) and the suggested action maps (right) as the results of the module *action suggestion*. Highly evaluated actions generate higher, lower evaluated generate lower activities in the motor map. Velocity is coded in y–direction and steering angle in x–direction. In the upper situation the robot avoids left turns, in the middle the robot can move straight ahead. The last situation represents an obstacle in front of the robot, therefore no "move forward" hypotheses are generated, merely slow turn left or turn right movements.

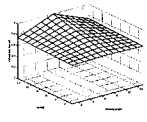

Figure 4: The already defined heuristic reinforcement function favours fast and straight ahead movements. Additionally, any collision is evaluated with a zero reward.

typical sensory situations are shown in figure 3, the corresponding optical flow fields and the suggested actions learned in preceding action perception cycles. For a reactive behavior we only use one prediction module. In this way, the action selection depends only on the current situation and the immediately expected reward.

For the anticipatory experiments we use three concatenated prediction modules. The function of our architecture is shown in figure 5. In contrast to the reactive system, the anticipating one takes not only the immediate best reward into consideration, but also future rewards. The bottom sequence was selected because the overall reward is higher than that of the top sequence, although the expected reward of the first action in the top line is highest. In order to get comparable performance measures we placed the robot in front of a large obstacle approximately 10 cm away. Its task is to maximize its overall reward considering the heuristically defined reinforcement function (fig. 4). The anticipatory approach avoids an obstacle earlier than the reactive (figure 6). Mean rewards for this experiment are shown in table 7.

In a given sensorimotor situation the system can select those local actions from the action map which are suited best for the global action. In figure 8 there

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 345-350

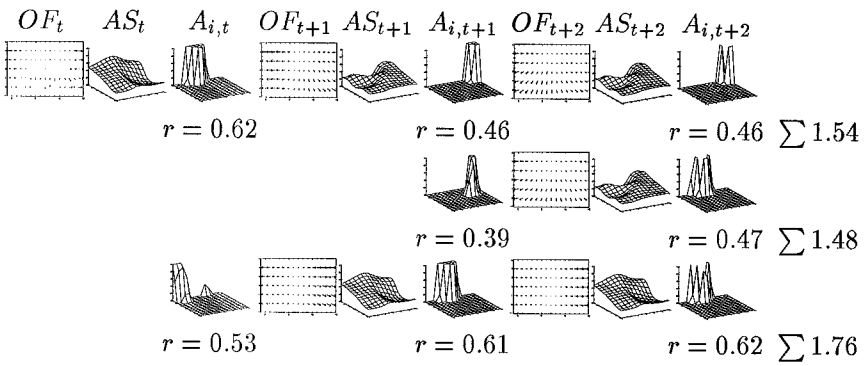

Figure 5: Tree of simulated action sequences. The first proposed action is the best action considering no simulation. With internal simulation, the last line has an higher expected overall reward than the top line.

is a constant global action "move left ahead" modulated into the prediction module. As can seen, the system avoids movements to the left as long as an obstacle is in the robot's left view. After the obstacle has vanished the robot turns left and follows the global intention.

4. Conclusions and Outlook

The usage of information about the future is part of strong reinforcement approaches. TD-(λ) or Q-Learning, for example, propagate the evaluation of a situation back to the one preceeding it. The evaluation of a action sequence depends on its complete execution. Our approach can combine parts of sequences into a new sequence and evaluate it at once. A drawback is that we can build sequences only of a size equal to the number of prediction modules. This is not critical because the changes in global actions are more dominant for larger time scales. In future work we will investigate the use of prediction differences in the sensory information to detect dynamical obstacles in order to give some hints for active vision systems for interesting areas in the visual input. Further investigations have to be carried out in comparisons with strong reinforcement approaches.

References

[1] Shun–Ichi Amari. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27:77–87, 1977.

[2] Ted Camus. *Real-Time Optical Flow.* PhD, Brown University, Department of Computer Science, Providence, RI 02912, USA, 1994.

[3] Andrew P. Duchon, William H. Warren, and Leslie Pack Kaelbling. Ecological robotics: Controlling behaviour with optical flow, 1994.

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 345-350

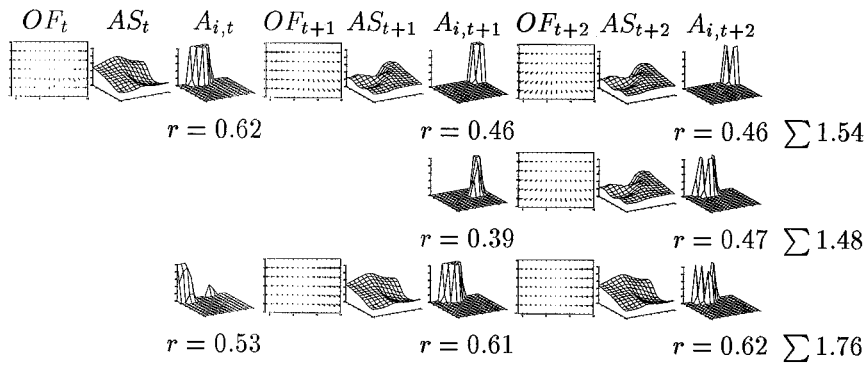

Figure 5: Tree of simulated action sequences. The first proposed action is the best action considering no simulation. With internal simulation, the last line has an higher expected overall reward than the top line.

is a constant global action "move left ahead" modulated into the prediction module. As can seen, the system avoids movements to the left as long as an obstacle is in the robot's left view. After the obstacle has vanished the robot turns left and follows the global intention.

4. Conclusions and Outlook

The usage of information about the future is part of strong reinforcement approaches. TD-(λ) or Q-Learning, for example, propagate the evaluation of a situation back to the one preceeding it. The evaluation of a action sequence depends on its complete execution. Our approach can combine parts of sequences into a new sequence and evaluate it at once. A drawback is that we can build sequences only of a size equal to the number of prediction modules. This is not critical because the changes in global actions are more dominant for larger time scales. In future work we will investigate the use of prediction differences in the sensory information to detect dynamical obstacles in order to give some hints for active vision systems for interesting areas in the visual input. Further investigations have to be carried out in comparisons with strong reinforcement approaches.

References

[1] Shun–Ichi Amari. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27:77–87, 1977.

[2] Ted Camus. *Real-Time Optical Flow.* PhD, Brown University, Department of Computer Science, Providence, RI 02912, USA, 1994.

[3] Andrew P. Duchon, William H. Warren, and Leslie Pack Kaelbling. Ecological robotics: Controlling behaviour with optical flow, 1994.

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 345-350

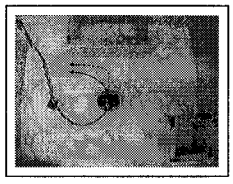

Figure 6: Result of anticipation of sensory consequences. The anticipatory agent moves earlier to the left (solid line). Therefore its mean reinforcement is higher than that of the reactive one (dotted line).

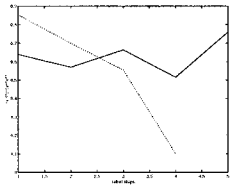

Figure 7: Comparison of the reinforcement values of the reactive (dotted line) and the anticipative behavior (solid line). The reactive robot gets first a higher reinforcement but can often not avoid a collision. The anticipative can look ahead and select the action with higher reinforcements in future.

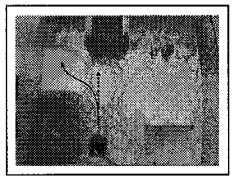

Figure 8: The left trajectory shows an global interaction to prefer a left turn. This preference takes effect after the obstacle on the left side has vanished from the view. The right trajectory is an example for no global interaction. The robot moves straight ahead until an obstacle is blocking its way.

[4] H.-M. Gross et al. A neural network hierarchy for data and knowledge controlled selective visual attention. In *Proc. of ICANN'92, Brighton*, pp 825–828. Elsevier Science Publishers, 1992.

[5] B. Kröse, A. Dev, X. Benavent, F. Groen. Vehicle navigation on optic flow, 1997.

[6] Th. Martinetz and K. Schulten. A "neural gas" network learns topologies. In T. Kohonen, K. Mäkisara, O. Simula and J. Kangas, editors, *Artificial Neural Networks*, pp 397–402. Elsevier Amsterdam, 1991.

[7] R. Möller and H.-M. Gross:. Perception through anticipation. In P. Gaussier and J.-D. Nicoud, editors, *PerAc'94, Los Alamitos.*, pages 408–411. IEEE Computer Society Press, 1994.

[8] Ralf Möller. *Wahrnehmung durch Vorhersage – Eine Konzeption der handlungsorientierten Wahrnehmung.* PhD thesis, TU-Ilmenau, Juni 1996.

[9] V. Stephan. Vision-basierte Ansätze für ein selbstorganisierendes Explorationsverhalten eines mobilen Miniaturroboters in einer Labyrinthwelt. Diplomarbeit, TU Ilmenau, FG Neuroinformatik, 1996.

[10] V. Stephan and H.-M. Gross. Formerhaltende sequentielle visuelle Aufmerksamkeit in columnar organisierten neuronalen Feldern. Submitted to 19. DAGM-Symposium, September 1997.

[11] R.S. Sutton. Integrated Modeling and Control Based on Reinforcement Learning and Dynamic Programming. In R. P. Lippmann, J. E. Moody, and D. S. Touretzky, editors, *Advances in Neural Information Processing 3*, pp 471–478, San Mateo, California, 1991. Morgan Kaufmann Publishers.

[12] C. Watkins and P. Dayan. Q-learning. *Machine-Learning*, 8:279–292, 1992.