Vision-Motor Abstraction toward Robot Cognition

Fady Alnajjar, Abdul Rahman Hafiz, Indra Bin Mohd Zin, and Kazuyuki Murase

Graduate School of Engineering, University of Fukui, Japan
{fady,abdul,indra,murase}@synapse.his.fukui-u.ac.jp

Abstract. Based on indications from neuroscience and psychology, both perception and action can be internally simulated in organisms by activating sensory and/or motor areas in the brain without actual sensory input and/or without any resulting behavior. This phenomenon is usually used by the organisms to cope with missing external inputs. Applying such a phenomenon in a real robot recently has taken the attention of many researchers. Although some work has been reported on this issue, none of it has so far considered the potential of the robot's vision at the sensorimotor abstraction level, where extracting data from the environment to build the internal representation takes place. In this study, a novel vision-motor abstraction is presented into a physically robot through a memory-based learning algorithm. Experimental results indicate that our robot with its vision could develop a simple anticipation mechanism in its memory from the interacting with the environment. This mechanism could guide the robot behavior in the absence of external inputs.

Keywords: Vision-motor abstraction, memory based learning, cognition.

1 Introduction

Real world applications are usually subject to change and very difficult to be predicted. Any sudden changes in the environment can possibly cause temporary lose in communication with the external world. Some organisms, those that have the ability of cognition, can cope with such situations by replacing the external missing sensory data with their own internal representation.

In recent decades, a branch of science called cognitive neuroscience has been established to introduce such a phenomenon to the robot [1]. It is concerned with endowing robots with human-like cognitive capabilities to enable them to accomplish complex tasks in complex environments. In [2], the authors have argued that all living creatures are cognitive to some degree, and therefore, it could be achieved to some extent on the robots [8]. We believe that the level of or how much the robot could be conscious of the surrounding environment depends on how much the robot knows about this environment. One of the possible approaches to measure robot's consciousness is by examining its ability to cope with missing external sensory data during performance of a specific task, i.e., performs blindfolded navigation using only its internal representation.

In recent years therefore, building a complete blindfolded navigation system in a robot has been a challenging task [3,4,6]. For instance, some initial experiments were

C.S. Leung, M. Lee, and J.H. Chan (Eds.): ICONIP 2009, Part II, LNCS 5864, pp. 65–74, 2009.

presented in [3] that aim to contribute toward building a robot that navigates completely blindfolded in a simple environment using a two-level network architecture; (a) low-level abstraction from sensorimotor values to a limited number of simple concepts, following the work done by Linker and Niklasson [4], and (b) higher-level prediction/representation of the agent's interaction with the environment, inspired by the work done by Nolfi and Tani [5]. These efforts have succeeded in allowing the robot to anticipate long chains of future situations. However, they have failed to support a completely blindfolded navigation [6], in which the robot repeatedly uses only its internal representation values for its navigation. The failure partly seems to be due to the short range of the robot's proximity sensors that they used, which limits the data that could be abstracted from the environment. We argue here that improving the robot's sensorimotor abstraction level from the limited short range sensor to a wide vision view could possibly overcome this problem.

To support our argument, we have done psychological experiments, similar to the one introduced by Lee and Thompson [7]. In a series of two experiments, we demonstrated the accuracy with which humans can guide their behavior based only on their internally generated sensory experiences. Two subjects were asked to do a task under different conditions. The first subject (X) was asked to 'look' around in a room and locate a specific target (Fig.1A). He was then blindfolded and asked to locate the target again. The subject performed the task accurately with closed eyes, in the same manner as when he was free to 'look' (Fig.1B). However, he could not predict the exact time needed to turn to the target and this caused a few hits with the obstacle (the empty circles in Fig.1B). The second subject (Y) was not allowed to explore the room with his eyes (no vision input). Instead, he was blindfolded and walked around the room touching things around him until he found the target (Fig.1C). He was then asked to seek the target again blindfolded from the initial position. Though successful in reaching the target, he took more time than that needed by subject (X). In addition, the number of times that he hit the wall or touched it to correct or locate his direction was greater (the empty circles in Fig.1D).

From the above experiment, we can conclude that subject (X) had collected a sufficient data during his first 'eyes open' navigation. This data could be various dimensions in the room which the subject related to times and distances that helped him to build internally his own image. In contrast, the data that subject (Y) had collected was limited to the objects that his hand touched during his first blindfolded navigation and their relation to his moving steps. This data, however, was not good enough to accurately perform the task.

In the above experiment, subject Y could be a demonstration of the results of the most recently reported works (e.g., [3]), since they used the short-range sensors for building the sensorimotor abstraction level. Our target in this study is to demonstrate the subject X behavior. Here we explore the inner world of a real robot that has a chance to explore the surrounding environment with its camera before it was told to navigate blindfolded in it.

The rest of the paper is structured as follows. Section 2 briefly presents the background. Section 3 describes the robot and the environment. Section 4 explains the architecture of the proposed algorithm and the experiments. The final section presents a brief discussion of the work and gives some future work suggestions.

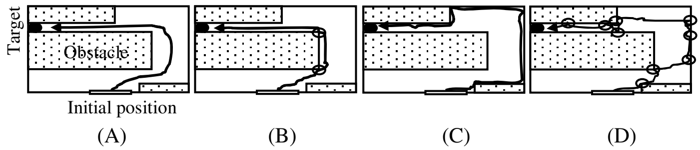


Fig. 1. The track of subject X\Y: (A)\(C) the first move, (B)\(D) the second move

2 Robot Cognition

A number of researchers have tried to investigate the robot's inner world [8]. Hesslow's paper [8], for instance, described a development of three simulation hypotheses in order to explain the robot's inner world. This was also discussed by [3]: The first is covert behavior, which is the ability to generate internally neural motor responses that are not actually externally executed. The second is sensor imagery, which is the ability to internally activate the sensory areas in the brain, so as to produce the simulated experience without actual inputs. The third is anticipation, which is the ability to predict the sensory consequences of the motor response [3,8].

Based on the above hypotheses, the internal sequences of the robot behaviors could be illustrated by Fig.2. In Fig.2A, a situation S1 elicits internal activity s1, which in turn leads to a motor response preparation r1 and thereafter results in the overt behavior R1, which causes a new situation S2. In Fig.2B, because of the robot's past experiences, the response preparation r1 could directly elicit the internal activity s2. In Fig.2C, if the robot trains the network to some degree, then it should be possible to simulate long sequences of motor responses and sensory consequences.

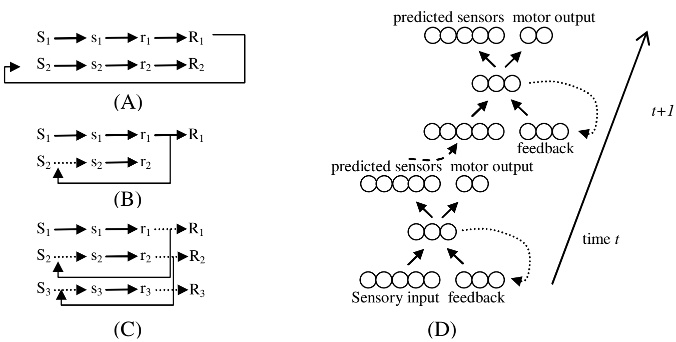


Fig. 2. (A)-(C) The basic principle of Hesslow's hypothesis (Adopted from [8]). (D) The basic approach of simulation of perception in robots used by [3].

Several modelers have used the above approach directly into the robot [2,3]. These modelers have mapped the sensory input to motor output, and also predicted the next time step's sensory input based on the network's experience. Most of these works, however, have considered only the robot's short-range sensors (e.g., IR sensor) in their experiments, which cannot provide enough data about the environment for the robot to build its internal representation.

From the psychological experiments in the above section, we agree with [3] that the weak spot in simulation theories is concerning the matter of the abstraction level at which internal representation is relied on. Therefore, improving the ability of this level could result in better internal representation in the robot, and in consequence, better cognition ability.

3 Robot and Environment

All the experiments in here were conducted in a physical mobile robot “Hemisson”. (www.k-team.com) (Fig.3A). The robot is able to avoid obstacles, detect ambient light intensity and follow a line on the floor. Hemisson is equipped with IR sensors and a wireless camera to transmit video images to a receiver that is connected to a PC for simple image processing.

In this study, the robot’s camera view has been divided into 4 parts, as illustrated in Fig.3C. Each part covers a number of pixels that represent the distance to the obstacles. The combination of these parts was used to indentify the current concept of the robot’s view (CC). The average of F_L and F_R were used to calculate the real distance (D) cm to the frontal obstacles. We have applied the idea of the flood fill algorithm [9] to filter the robot’s view and to easily clarify the boundaries between the floor and the obstacles. We arranged an ideal environment for the robot to avoid a large amount of image processing. A simple neural network was trained by Back-Propagation (BP) to convert the number of pixels in each part into a real distance.

The environment we used was similar to the one used by [3], consisting of two different-sized rooms connected by a short corridor (Fig.3B).

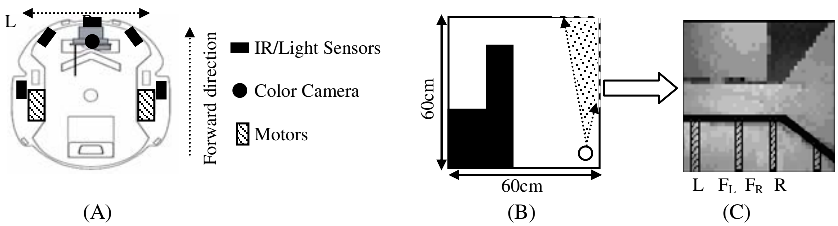


Fig. 3. (A) Schematic drawing of Hemisson. (B) Robot environment. The empty circle illustrates the robot. The dots area illustrates the range of robot’s view. (C) Snapshot of the Robot’s view taken from the position shown in B. Black thick line illustrates the lower edge of the obstacle. L, F_L, F_R and R, illustrate the left, left-front, right-front and right pixels range reading, respectively.

4 Experiments

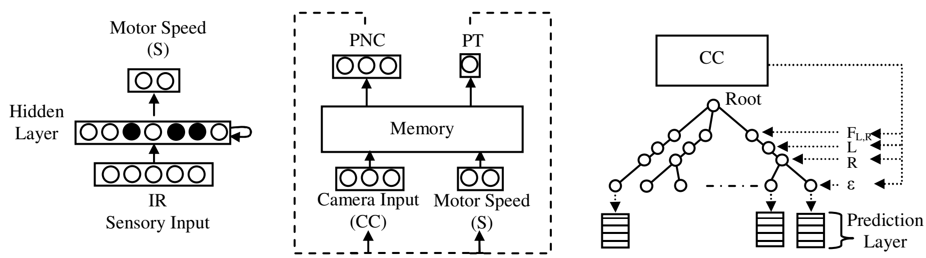
4.1 The Proposed Architecture

The general network architecture in this study was inspired by the architecture in [3]. In their work, they used two-level neural network architecture. The lower level consisted of an unsupervised vector quantizer that categorized the current IR-sensor and motor values into a more abstract level; they called “concepts”, such as “corner” or “corridor”. The higher level consisted of a recurrent neural network that trained by either BP or by GA to try to predict the sequence of the lower-level concepts and their respective duration [3].

Our architecture, in contrast, differs in two main aspects. First, instead of using only the IR sensors as an input for the abstract level, we added the robot’s visual sensors to the system. Second, the robot has two separated networks. The first, to control the robot’s navigation (Fig.4A), while the second, to represent the robot’s memory (Fig.4B). The second network is used to abstract data from both the first network (motors’ speed) and the environment. It also learns the relationships between these data to build the robot’s internal representation, and to predict both the upcoming concept (PNC) and the time needed to go through each concept (PT).

We tried to train the second network with both GA and BP, as was suggested by [3,6]. However, the error ratio in both algorithms was very high, even when we trained the robot for a very long time. The reasons for these failures could be that the number of concepts generated by the robot’s camera has a sequence which is too complex for such algorithms to learn; especially we are dealing with a physical robot that makes learning through these types of evolutionary algorithms quite impossible.

We, therefore, shifted the learning process in the second network so to be based on the contents of the robot’s memory. That is while the robot is navigating in the environment using the first network; the second network tries to memorize the environment and predicts them from its experience the next view or action, and corrects later its data by the actual fact when it faces it. This algorithm turned out to be reasonably successful.



(A) (B) (C)

Fig. 4. (A) Architecture of the SNN used in the first network. White\black circles represent excitatory\inhibitory neurons which have positive\negative connection. The neurons in the hidden layer are fully connected to each other (Adopted from [10]). (B) Architecture of the second network. *PNC* and *PT* are the output of the robot’s memory at each time step. The dashed line illustrates the feedback connection done in experiment 3. (C) Tree-memory structure used in this study (Adopted from [11]). Each concept has its own prediction layer. Prediction layer contains *PNC, PT* and the end action of each concept, e.g., turn right or turn left.

A tree-type memory structure has been introduced to the second network, similar to the one introduced in [11] (Fig.4C). This memory has a dynamic structure and simple storing and retrieving mechanism. It was also supported by forgetting and clustering mechanisms to control its general size and to provide a maximum memorizing ability. The memory was divided into five levels. The first three levels were used to store the robot camera inputs to identify each concept. The fourth level, (ε), was used to count the number of concepts in each environment to identify the environment. The last level represents the prediction layer.

During the navigation, the robot built its memory and used it to predict its future action. The flowchart in Fig.5 shows the working mechanism of the memory. According to the chart, when the robot returns to its straight state after performing the turning action, two phases are operated sequentially: the learning phase (the thick lines in the chart), where the robot learns and updates its memory with the currently available facts, and the predicting phase (the thin lines in the chart), where the robot explores the environment and gradually builds its experiences.

4.2 Experimental Results

4.2.1 Wall-Following Behavior

In order to control the robot's initial behavior, the robot's first network was programmed to perform wall following behavior using a pre-trained self-organizing spiking neural network [10] Fig.4A.

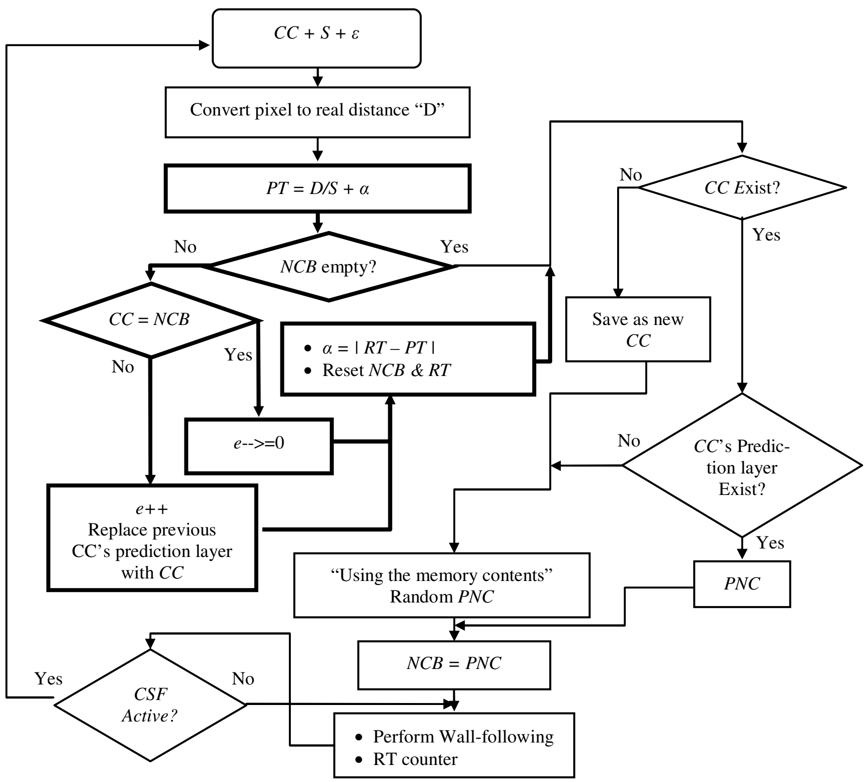


Fig. 5. The working mechanism of the robot's memory. Where NCB: next concept buffer. RT: real time. *e:* error prediction rate. CSF: changing-state flag.

As previously stated, the robot used this network exclusively for performing the navigation task; no data abstraction from the environment was processed in this level. The activation of CSF at every new state excited the second network to do its task. To simplify the second network's task, which partly depended on the motor output from the first network, we limited the robot's forward speed to a value equal to the average of the robot's forward speed in 10 success rounds in the environment (S=1.25cm/sec).

4.2.2 Data Abstraction and Prediction

The main objective of this stage is to examine the validity of the second network to build an internal representation, so that, the robot can keep tracking its own relative position in the environment and to anticipate the upcoming event.

Similar work to [3] has been carried out in this stage with some variations. The sensorimotor abstract level in the second network was supported by the robot's vision and a tree-memory structure. In the experiment, we left the robot, using the first network, to perform the wall-following task for 5 rounds, simultaneously with the existence of the second network whenever CSF was activated. At the beginning of each concept, the network trained both PNC and PT.

Figure 6A&B show the number of concepts that the robot could identify from the environment using its camera. From the figure, we can see that our method abstracted 7 different concepts from the environment, while in [3] only 5 concepts were found by using the short-range IR sensors. Notice that the number of concepts indicates to what degree the robot is sensitive about the environment, and as a consequence, it would results in better anticipation.

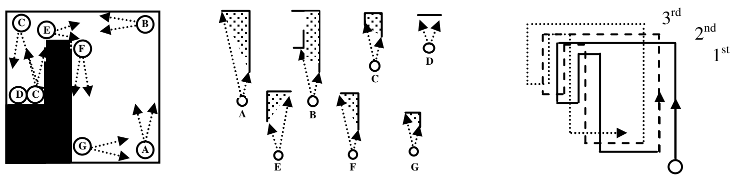
Table 1 shows the evolvement of the learning and predicting phases in the robot's memory during the 3 complete rounds in the environment. From the table, at the first round, the robot is not able to predict neither the PNC nor PT correctly. The robot set these values randomly since it does not have experience about them. In the learning phase, however, it updates both of these values in each concept by learning online from the value of RNC and RT, respectively. It is worthwhile to mention that the robot built a suitable knowledge about the surrounding environment within the first two rounds. Within the 3^{rd} round the robot was able to predict all the PNC correctly, i.e., PNC = RNC and e decreases to 0. Although, the robot was unable to predict PT 100% correctly, i.e., PT≠RT, however, its value comes very close to RT and α turned out to be very close to 0. We have also introduced different environment to the robot; the robot could easily adapt to the new environment and in short time could update its memory to predict the sequence of the new environment (experiment results were not shown due to the limitation).

4.2.3 Blindfolded Navigation

The objective of this stage is to examine the ability of the robot to replace all of its external sensory input with its own internal representation, i.e., repeatedly using the sequences of its own prediction for a certain number of times without external sensory input, see the dashed lines in Fig.4B. In other words, have the robot navigates itself blindfolded in the environment.

In this stage, after the robot trained in the original environment for 5 rounds, we removed the surrounding environment completely, eliminated the external sensory inputs, and let the robot move in a wide space using only the last found sequence of the concepts in its memory.

Figure 6C shows the robot's best behavior. Interestingly, the robot built experiences about the environment in its memory sensitive enough so that it could navigate in the environment without any interaction with the external world. All the concepts were memorized correctly, and the robot moved according to the environment's layout. Unfortunately, the robot has slightly shifted its movement in each round, and this was probably due to the error in predicting the time of each concept.



(A) (B) (C)

Fig. 6. (A) Number of concepts that robot's camera could identify in the original environment. (B) The layout of the robot's view in each concept. (C) Robot's behavior during 3 rounds simulating only its internal representation.

Table 1. Second Network Evolvements for 3 Rounds

(Each concept assigned by CC's value that automatically generated in sequential manner). *RNC* = the *CC* of the next step.

	Predicting phase								Learning phase				
Round	D(*cm*)	***CC***	*CC Value*	*ε*	*S cm/sec*	*D/S*	*PT*	***PNC***	*RT*	*α*	*RNC*	*e*	***e+α***
1	40.04	**A**	1	1	1.25	32.032	32.032	**A**	39.5	7.5	B	1	**8.5**
1	40.04	**B**	2	3	1.25	32.032	32.032	**A**	39.1	7.1	C	2	**9.1**
1	24.88	**C**	3	6	1.25	19.904	19.904	**B**	21.0	1.1	D	3	**4.1**
1	5.01	**D**	4	10	1.25	4.008	4.008	**D**	5.1	1.1	C	4	**5.1**
1	24.55	**C**	3	7	1.25	19.64	19.64	**D**	18.5	-1.1	E	5	**3.9**
1	32.65	**E**	5	12	1.25	0	0	**B**	4.5	4.5	F	6	**10.5**
1	36.8	**F**	6	18	1.25	29.44	29.44	**C**	33.8	4.4	G	7	**11.4**
1	27.7	**G**	7	25	1.25	22.16	22.16	**B**	21.3	-0.9	A	8	**7.1**
2	39.4	**A**	1	24	1.25	31.52	39.0	**B**	39.1	0.1	B	7	**7.1**
2	40.04	**B**	2	22	1.25	32.032	39.1	**C**	39.4	0.3	C	6	**6.3**
2	23.84	**C**	3	25	1.25	19.072	20.2	**E**	20.5	0.3	D	7	**7.3**
2	5.01	**D**	4	21	1.25	4.008	5.1	**C**	5.1	0.0	C	6	**6.0**
2	24.88	**C**	3	18	1.25	19.904	18.8	**D**	19.5	0.7	E	7	**7.7**
2	32.17	**E**	5	13	1.25	0	4.5	**F**	5.0	0.5	F	6	**6.5**
2	36.06	**F**	6	7	1.25	28.848	33.2	**G**	33.0	-0.2	G	5	**4.8**
2	27	**G**	7	0	1.25	21.6	20.7	**A**	20.4	-0.4	A	4	**4.4**
3	39.4	**A**	1	1	1.25	31.52	39.1	**B**	40.2	1.1	B	3	**4.1**
3	39.09	**B**	2	3	1.25	31.272	38.6	**C**	38.9	0.3	C	2	**2.3**
3	24.36	**C**	3	6	1.25	19.488	20.9	**D**	20.5	-0.4	D	1	**1.4**
3	5.01	**D**	4	10	1.25	4.008	5.1	**C**	5.0	-0.1	C	0	**0.1**
3	24.88	**C**	3	7	1.25	19.904	19.5	**E**	18.8	-0.7	E	0	**0.7**
3	36.06	**E**	5	12	1.25	0	5.0	**F**	5.0	0.0	F	0	**0.0**
3	36.8	**F**	6	18	1.25	29.44	33.6	**G**	33.5	-0.1	G	0	**0.1**
3	27	**G**	7	25	1.25	21.6	20.4	**A**	21.0	0.6	A	0	**0.6**

5 Conclusion

We have presented some initial experiments with the aim to contribute toward the development of robot models in sensorimotor abstraction, simulation and anticipation. In particular, and unlike the most previous related work, we have here presented (a) a robot equipped with a video camera to extract data from the environment during its navigation, and (b) a tree-type memory structure to store this data in a simple manner to use it to anticipate upcoming events and to guide the robot's behavior in the absence of the external inputs.

Our experiments show that the proposed algorithm successfully built internal representations of the environment. These representations were capable of predicting upcoming concepts and of navigating the robot blindfolded in the environment, replacing missing sensory input.

The results in the 2nd experiment indicated that our algorithm had memorized the sequences of concepts found in the environment and the robot's behavior in each one. The robot's internal representation had captured the topology of the original environment. With such memory structure, the robot's previous knowledge could be recalled easily when it needed. In the final experiment, the robot indeed was able to navigate, to some degree, blindfolded using only its own internally built representation. The robot used its 'mind' to navigate from one concept to another by operating through a series of actions and situations that it learned. The robot's memory was not very good at predicting the real time needed for each concept, but neither can humans (Fig.1). This predicting error caused a little delay in the robot's movement.

Although some studies have reported on the issue of robot imaginations and anticipations in different ways (e.g., SLAM [12]), where the robot can use its sensorimotor representation in the brain to simulate its movement internally before the actual movement and to reason about its ability to perform the task in a short time and a safe manner, the robot, however, has been told the layout of the environment and/or the position of the targets in advance. We showed in this study, that the robot could build an environment's map and an appropriate sequence of events in it through its own experiences. The robot can then use this data to recover any missing or corrupted data and even plan its future movement within its internal representation before any actual move.

We believe that the work presented here illustrates some promising directions for further experimental investigations of vision-motor abstraction and for further developments of the synthetic phenomenology approach in general.

As a possible future set of experiments, it would be interesting to try to improve the learning algorithm of the second network by building a higher-level to control the prediction-layer operations in the prediction phase. This may decrease the learning time for newly created prediction-layers.

Acknowledgments. This work was supported by grants to KM from Japanese Society for Promotion of Sciences and from the University of Fukui.

References

1. Gazzaniga, M.S.: The Cognitive Neurosciences III. MIT Press, Cambridge (2004)
2. Varela, F.J., Thompson, E., Rosch, E.: The embodied mind: cognitive science and human experience. MIT Press, Cambridge (1991)

3. Stening, J., Jacobsson, H., Ziemke, T.: Imagination and abstraction of sensorimotor flow: Towards a robot model. In: Chrisley, R., Clowes, R., Torrance, S. (eds.) Proceedings of the Symposium on Next Generation Approaches to Machine Consciousness, Hatfield, UK, pp. 50–58 (2005)
4. Linåker, F., Niklasson, L.: Extraction and inversion of abstract sensory flow representations. In: Proceedings of the Sixth international Conference on Simulation of Adaptive Behavior, From Animals to Animates, vol. 6, pp. 199–208. MIT Press, Cambridge (2000)
5. Nolfi, D.S., Tani, J.: Extracting regularities in space and time through a cascade of prediction networks: The case of a mobile robot navigating in a structured environment. Connection Science 11(2), 125–148 (1999)
6. Stening, J.: Exploring Internal Simulations of Perception in a Mobile Robot using Abstractions. Masters Thesis, School of Humanities and Informatics, University of Skövde, Sweden (2004)
7. Lee, D.N., Thompson, J.A.I.: Vision in action: the control of locomotion. In: Ingle, D., Goodale, M.A., Mansfield, R.J.W. (eds.) Analysis of Visual Behavior, pp. 411–433. MIT Press, Cambridge (1982)
8. Hesslow, G.: Conscious thought as simulation of behavior and perception. Trends in Cognitive Science 6(6), 242–247 (2002)
9. Taylor, T., Geva, S., Boles, W.W.: Monocular vision as a range sensor. In: Mohammadian, M. (ed.) Proceedings of International Conference on Computational Intelligence for Modeling, Control and Automation, pp. 566–575 (2004)
10. Alnajjar, F., Murase, K.: Self organization of spiking neural network that generates autonomous behavior in a real mobile robot. International Journal of Neural Systems 16(4), 229–239 (2006)
11. Alnajjar, F., Mohd Zin, I., Murase, K.: A Hierarchical Autonomous Robot Controller for Learning and Memory: Adaptation in Dynamic Environment. Adaptive Behavior 17(3), 179–196 (2009)
12. Vaughan, R., Zuluaga, M.: Use your illusion sensorimotor self-simulation allows complex agents to plan with incomplete self-knowledge. In: Nolfi, S., Baldassarre, G., Calabretta, R., Hallam, J.C.T., Marocco, D., Meyer, J.-A., Miglino, O., Parisi, D. (eds.) SAB 2006. LNCS (LNAI), vol. 4095, pp. 298–309. Springer, Heidelberg (2006)