Cognitive Modeling Based on Perceiving-Acting Cycle in Robotic Avatar System for Disabled Patients

Takenori Obo
Department of Applied Computer Science
Tokyo Polytechnic University
1583 Iiyama, Atsugi, Kanagawa
243-0297, Japan
t.obo@cs.t-kougei.ac.jp

Ryoya Hase
Department of Applied Computer Science
Tokyo Polytechnic University
1583 Iiyama, Atsugi, Kanagawa
243-0297, Japan
c1618068@st.t-kougei.ac.jp

Kohei Kobayashi
Department of Applied Computer Science
Tokyo Polytechnic University
1583 Iiyama, Atsugi, Kanagawa
243-0297, Japan
c1618035@st.t-kougei.ac.jp

Kotaro Sueta
Department of Applied Computer Science
Tokyo Polytechnic University
1583 Iiyama, Atsugi, Kanagawa
243-0297, Japan
c1618045@st.t-kougei.ac.jp

Takeru Nakano
Department of Applied Computer Science
Tokyo Polytechnic University
1583 Iiyama, Atsugi, Kanagawa
243-0297, Japan
c1618062@st.t-kougei.ac.jp

Duk Shin
Department of Electronics and Mechatronics
Tokyo Polytechnic University
1583 Iiyama, Atsugi, Kanagawa
243-0297, Japan
d.shin@em.t-kougei.ac.jp

***Abstract*—In this study, we aim to develop a system of remote-controlled avatar robot for elderly and disabled patients. Most of teleoperation systems have interfaces to visually present the state of the robots including feedback information and receive the control commands manually sent from the operator. However, in elderly and patients with disabilities, they might have difficulty in the manual control of the robot. This paper therefore presents a multimodal interface for remotely controlling a robotic avatar. We furthermore propose a cognitive platform for remotely controlling a robot based on a concept of perceiving-acting cycle. The platform consists of a perceptual system for incremental environment modeling and an action system for extracting patterns of operator behavior. In each system, a self-organized neural network based on unsupervised learning is used. Moreover, we use a spiking neural network for spatial-temporal modeling of interaction between an operator and environment.**

Keywords— avatar robot, perceiving-acting cycle, spiking neural network, cognitive modeling, multimodal interface

I. INTRODUCTION

Various types of social mobile robots have been developed and used in a wide variety of fields, with the rapid advance of artificial intelligence (AI) and wireless communication technology. Government of Japan proposed “Society 5.0” as a future vision of a human-centered society that balances economic advancement with the resolution of social problems by a system that highly integrates cyberspace and physical space [1, 2]. In the society, a large amount of data from wireless sensor devices in real world is stored in the cyberspace. Information derived from the data by AI technologies is provided in suitable style for humans, things and systems. The cyber-physical systems can realize extension of human capabilities, providing seamless connections between virtual space and real space.

Aging and declining population is a serious social issue in Japan. The annual fatalities from brain stroke and cerebrovascular disorders are on downward trend in recent years. However, the number of patients with permanent damage such as higher cerebral dysfunction and motor paralysis remains on the upward trend. These patients are suffering from many disabilities related to loss of independence in speaking, walking, eating, and performing activities of daily living (ADLs). It can lead to their social isolation and loneliness that can induce immobility and depression, producing the vicious cycle. To maintain and improve their Quality of Life (QOL) and Quality of Community (QOC), synthesis of information technology (IT), network technology (NT), and robot technology (RT) can be a prospective way to serve as the catalyst for being reintegrated into society and enhancing individual autonomy. In this study, we aim to develop a system of remote-controlled avatar robot for elderly and disabled patients.

In a teleoperation system of avatar robot, a human operator is required to remotely control a robot. In the related works, some of these systems not only serve a purpose of remote-controlling but also permit the operator to feel that the movement of the robot is the operator’s own [3]. The sense of agency is strengthened by the subjective experience of synchronization between the operator’s actions and the robot movements. Especially, in teleoperation systems, discomfort caused by time delay and movement that differs from operator’s prediction depresses the sense of agency [4, 5]. Most of these systems have interfaces to visually present the state of robots including feedback information and receive the control commands manually sent from operator. However, in elderly and patients with motor disabilities, they might have difficulty in the manual control of robot because of their involuntary movement and reduced range of motion. Gaze tracking system is used as an effective handsfree teleoperation interface [6-8].

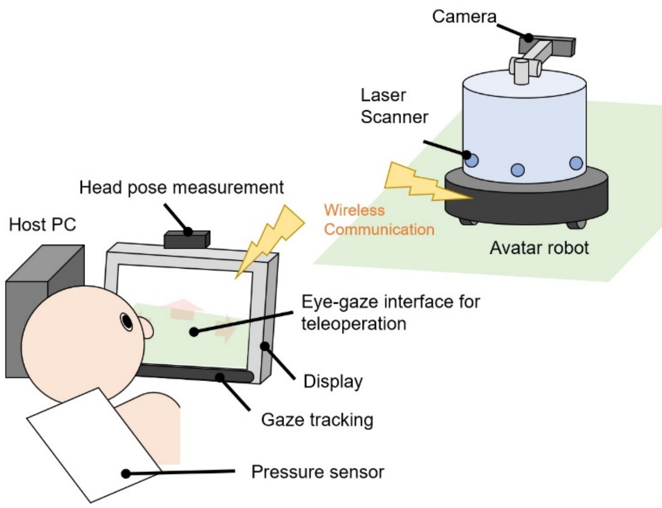

Fig. 1. A concept of remote-controlled robot system for elderly and disabled patients.

Eye-gaze tracking interface has a lot of contribution to move around through the use of a robotic wheelchair and a telepresence robot, despite quadriplegia. Moreover, sensors that can detect body posture and gesture are applied to interfaces for teleoperation systems [9-12]. A virtual reality system to give an experience of an immersive teleoperation integrates different types of sensors for multimodal interaction [13-15]. The multimodal interface is required to present suitable information for an operator corresponding to sensory modalities clarified from the mutuality of their perceiving and acting.

This paper presents a prototype of a robotic avatar system for patients with disabilities. Figure 1 shows a concept of the proposed system divided into two subsystems: the robot system and the multimodal interface. We propose a cognitive platform for remotely controlling a robot based on a concept of perceiving-acting cycle. The platform consists of a perceptual system for incremental environment modeling and an action system for extracting patterns of operator behavior. In each system, a self-organized neural network based on unsupervised learning is used. We furthermore use a spiking neural network for spatial-temporal modeling of interaction between an operator and environment.

This paper is organized as follows: section 2 explains the developed avatar robot system, section 3 describes the details of the proposed cognitive platform based on a model of perceiving-acting cycle, section 4 shows an experimental example and discusses the results, and we finally summarize this paper and mention the future direction of this study in Section 5.

II. Avatar Robot System

A. Robot System

We use a programmable mobile robot iRobot Create 2 developed by iRobot Corporation. The specification is based on that of robotic cleaner Roomba. The robot has two drive wheels and a coaster wheel in front. The vacuum cleaner hardware is removed from the development robot. The robot alternately has a Mini-DIN 7-pin serial port providing serial communication, digital input and output, analog input and output. Moreover, the robot platform provides the protocol of iRobot Roomba Open Interface (ROI) to send and read mode commands, actuator commands, sensor commands, and so on. We equipped the mobile robot with 2D laser scanner, web camera, mini PC, and battery as shown in Fig. 2.

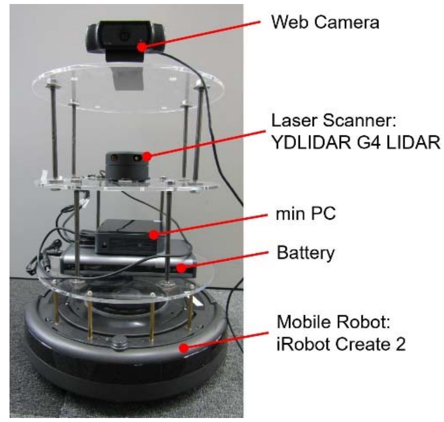

Fig. 2. Avatar robot.

The laser scanner is used for 2D simultaneous localization and mapping (SLAM). The product (YDLIDAR G4 LIDAR) is developed by Shenzhen Yuedeng Technology Co., Ltd. This devise is a 360-degree two-dimensional laser range scanner based on time of flight principal, depending on the principle of triangulation distance measurement to achieve high frequency and high precision measurement. The laser range scanner can measure distance to the surrounding objects, up to approximately 16 [m]. Scan frequency is 5-12 [Hz].

B. Multimodal Interface for Remote Controlling

We developed a mattress pressure sensor with 16 force sensing resistors for measuring user's center of gravity (COG) to control the mobile robot, as shown in Fig. 3. Its direction and velocity of movement are controlled by the position of COG while sitting. For example, the robot moves forward when the operator fall forward slightly on the mattress interface. For the electrically driven motors, the values of output power are given by

$$\begin{aligned} P_R &= \alpha \cdot (-c_x + c_y) \\ P_L &= \alpha \cdot (c_x + c_y) \end{aligned} \tag{1}$$

where (c_x, c_y) indicates an operator's COG position on a coordinate system, α is a coefficient, and P_R and P_L are respectively the values of analog output (PWM) to the right and left motors. Moreover we utilize a belt tactile interface for vibrotactile sensory feedback (Fig. 4). The tactile interface has eight small vibration motors and a microcontroller board to control them. Strength of each motor's vibration varies with speed and direction of the robot's movement.

Furthermore, we use a stationary eye-tracker Tobii Eye Tracker 4C produced by Tobii Technology, Inc. Figure 5 shows an eye-gaze interface for teleoperation. The eye-tracker is a low-cost and easily implementable device that needs to attach to the bottom of a PC display and connects to computer via USB to supply power and to retrieve the gaze and head pose data. The

Fig. 3. Mattress pressure sensor with 16 force sensing resistors for measuring user's center of gravity to control the mobile robot.

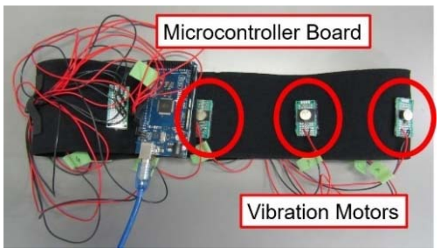

Fig. 4. Belt tactile interface for vibrotactile sensory feedback.

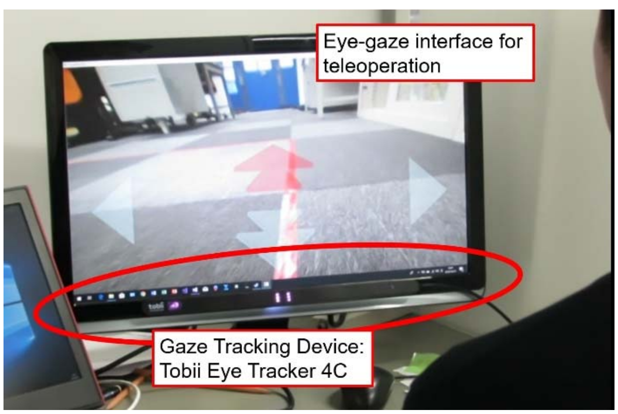

Fig. 5. Eye-gaze interface for teleoperation.

TABLE I. SPECIFICATION OF TOBII EYE TRACKER 4C

Size	17×15×335 [mm]
Operating Distance	50 - 90 [cm]
Track Box Dimensions	40 × 30 [cm] at 75 [cm]
Frame rate	90 [Hz]

specification is presented by Table 2. The gaze tracking system is used for detecting attentional points on the display while teleoperating.

III. COGNITIVE MODELING FOR AVATAR ROBOT

A. Cognitive Modeling Based on Perceiving-Acting Cycle

In this study, we propose a method of human cognitive modeling while teleoperating in the robotic avatar system. Figure 6 presents the proposed approach. The cognitive model is based on perceiving-acting cycle discussed in ecological psychology. The cognitive model is composed of a perceptual

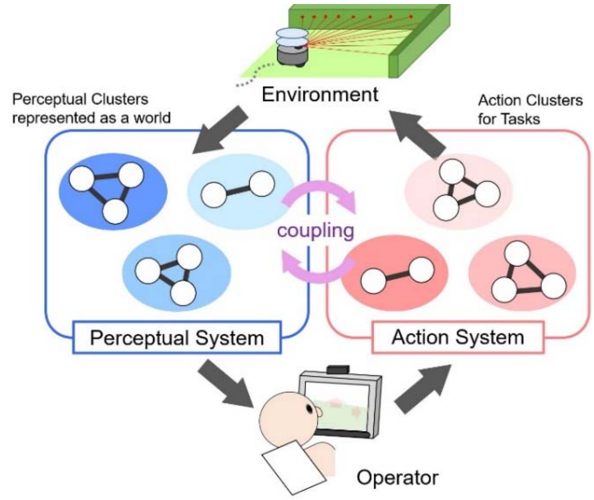

Fig. 6. Cognitive modeling based on a concept of perceiving-acting cycle in the robotic avatar system.

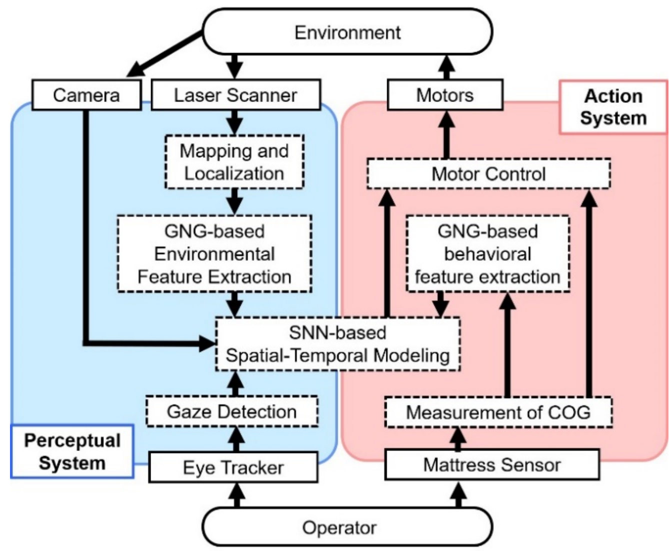

Fig. 7. System architecture and processing for the cognitive modeling.

system for incremental environment modeling and an action system for extracting patterns of operator's behavior. H. Masuta proposed an intelligent control system based on perceiving-acting cycle for an autonomous robot [16], however, we applied the concept to a teleoperation robot system. A coupling between the perceptual system and action system emerged by the interaction with environment can lead to improvement of sense of agency and ownership to measure the embodiment.

Figure 7 presents the system architecture and processing for the cognitive modeling. In perceptual system, the information of environmental map is used for detecting space where the robotic avatar can move around. Moreover, environmental feature is extracted as perceptional information by using a self-organized neural network. In action system, behavioral feature in teleoperating is extracted from the measurement of their COG. We assume that the extracted features have spaciotemporal relationship with each other. Therefore, a spiking neural network is used for the spatial-temporal modeling to organize the cognitive model based on perceiving-acting cycle.

B. Self-Organized Neural Network

In this study, an unsupervised learning method based on a self-organized neural network is used for the cognitive modeling. Self-organized map (SOM) is a typical unsupervised learning method that flexibly generates a topological structure corresponding to data distribution. The number of nodes and the connections in SOM are experimentally predefined; that is, the algorithm does not include a process for incrementing and deleting nodes and edges.

Growing neural gas (GNG) can dynamically change their topological structure based on the adjacent relation (edge) referring to the ignition frequency of the adjacent node according to the error index. The algorithm of GNG is proposed by B. Fritzke [17-22]. In GNG, the nodes and edges are deleted based on a concept of node ageing. The learning algorithm is as follows:

Step 0. Generate two units at random position, $\mathbf{r}_1$, $\mathbf{r}_2$ in $\mathbf{R^n}$. Initialize the connection set.

Step 1. Generate ramdomly an input data v according to the probability density function of the input data.

Step 2. Select the unit (winner) s_1 and the second-winner unit s_2 by

$$s_1 = \arg\min_{i \in A} \{d_i\} \tag{2}$$

$$s_2 = \arg\min_{i \in A \backslash \{s_1\}} \{d_i\} \tag{3}$$

$$d_i = \|\mathbf{v} - \mathbf{r}_i\| \tag{4}$$

Step 3. If a connection between s_1 and s_2 does not exist already, create the connection. Set the age of the connection between s_1 and s_2 to zero:

$$a_{s_1,s_2} = 0 \tag{5}$$

Step 4. Add the squared distance between the input data and the winner to a local error variable:

$$E_{s_1} \leftarrow E_{s_1} + d_{s_1} \tag{6}$$

Step 5. Update the reference vectors of the winner and its direct topological neighbors by the learning rate η_1 and η_2, respectively:

$$\mathbf{r}_{s_1} \leftarrow \mathbf{r}_{s_1} + \eta_1 \left(\mathbf{v} - \mathbf{r}_{s_1}\right) \tag{7}$$

$$\mathbf{r}_j \leftarrow \mathbf{r}_j + \eta_2 \left(\mathbf{v} - \mathbf{r}_{s_1}\right) \quad \text{if } C_{s_1,j} = 1 \tag{8}$$

where $C_{S1,j}$ is a connection between s_1 and j-th node (if they are connected with each other, $C = 1$).

Step 6. Increment the age of all edges emanating from s_1:

$$a_{s_1,j} \leftarrow a_{s_1,j} + 1 \quad \text{if } C_{s_1,j} = 1 \tag{9}$$

Step 7. Remove edges with an age larger than $a_{\max}$. If this results in units having no more emanating edges, remove those units as well.

Step 8. If the error E_l is higher than the predefined threshold, insert a new unit as follows.

i. Select the unit m with the maximum accumulated error among the neighbors of l.

ii. Add a new unit n to the network and interpolate its reference vector from l and m.

$$\mathbf{r}_n = 0.5 \cdot \left(\mathbf{r}_l + \mathbf{r}_m\right) \tag{10}$$

iii. Insert edges connecting the new unit n with units l and m, and remove the original edge between l and m.

iv. Decrease the error variables of l and m by a coefficient γ:

$$E_l \leftarrow \gamma E_l \tag{11}$$

$$E_m \leftarrow \gamma E_m \tag{12}$$

v. Interpolate the error variable of n from l and m:

$$E_n = 0.5 \cdot \left(E_l + E_m\right) \tag{13}$$

Step 9. Decrease the error variables of all units:

$$E_i \leftarrow E_i - \beta E_i \quad \left(\forall i \in A\right) \tag{14}$$

Step 10. Continue with step 1 if a stopping criterion (e.g., net size or some performance measure) has not yet been fulfilled.

C. Pulse Neuron Model

The McCulloch–Pitts (MCP) neuron model is one of the most famous models [23, 24]. However, since the neuron model is so simple that the activation function is dependent on only the magnitude of the input signal, the model is very sensitive to the input. Pulse neuron model describes the spatiotemporal dynamics of real neuron in more detail.

In this study, a simplified spike response model is used for computational cost reduction [25, 26]. The neuron model has an internal state given by the following equation:

$$h_i(t) = \gamma \cdot h_i(t-1) + h_i^{syn}(t) + h_i^{ext}(t) + h_i^{ref}(t) \tag{15}$$

where γ is the discount rate ($0 < \gamma < 1.0$), $h^{syn}(t)$ is the synthesis of input from other neurons, $h^{ext}(t)$ is an external input, and $h^{ref}(t)$ is a term to simulate the refractoriness. If the internal state of the neuron exceeds a threshold, the neuron fires a output spike. The spike is given by

$$p(t) = \begin{cases} 1 & \text{if } h(t) \geq q \\ 0 & \text{otherwise} \end{cases} \tag{16}$$,

where q is the threshold.

The output is transmitted to the connected neuron, based on the degree of the presynaptic action potential (PSP). The PSP is represented by the following equation:

$$h_i^{psp}(t) = \begin{cases} 1 & \text{if } h_i(t) \geq q \\ \gamma^{psp} \cdot h_i^{psp}(t-1) & \text{otherwise} \end{cases} \tag{17}$$

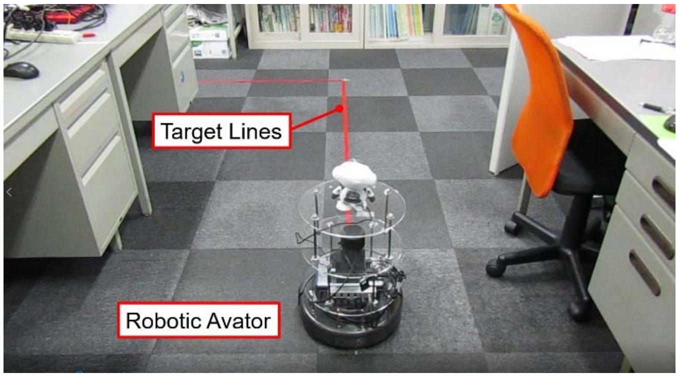

Fig. 8. Experimental environment.

where γ^{psp} is the discount rate ($0 < \gamma^{psp} < 1.0$). Spikes are integrated as an input to other neuron, based on the weight parameters represented as the strength of the synaptic connections. The input is given by

$$h_i^{syn}(t) = \sum_{j=1,\, j\notin i}^{N} w_{j,i} \cdot h_j^{psp}(t-1) \tag{18},$$

where $w_{j,i}$ is the weight parameter from the j-th to i-th neuron, and N is the number of the neurons. If the weight parameter is positive, the postsynaptic action potential is excitatory.

In physiology, the refractory period is a phase during which the neuron is incapable of firing. The refractoriness is represented by the following equation:

$$h_i^{ref}(t) = \begin{cases} -1 & \text{if } h_i(t) \geq q \\ \gamma^{ref} \cdot h_i^{ref}(t-1) & \text{otherwise} \end{cases} \tag{19}$$

where γ^{ref} is the discount rate ($0 < \gamma^{ref} < 1.0$).

IV. Experimental Example

This section shows an experimental result using the proposed system. We aim to discuss the cognitive modeling during a teleoperation according to the results of unsupervised learning in perceptional and action systems. In the experiment, the subject is required to control the robotic avatar remotely and trace a red color line marked on the floor. Figure 8 and 9 show the experimental environment and movement of the robot.

Figure 10 shows a result of GNG-based environmental feature extraction in perceptional system. In the figures, the robot and extracted feature points are respectively represented by blue and green circles. The topological map generated by GNG is indicated in the green nodes and edges depicted in the figure. The topological map can be flexibly extended corresponding to the progress of map building. The feature points in the topological space can be regarded as the abstracted world in perceptional system.

On the other hand, Fig. 11 presents a result of GNG-based behavioral feature extraction in action system. The topological map is generated in the operator's COG coordinate system. According to the result, the distance between the nodes is dependent on the operational purpose in each situation. The feature points in action system can be spatiotemporally associated with those in perceptional system. In the proposed

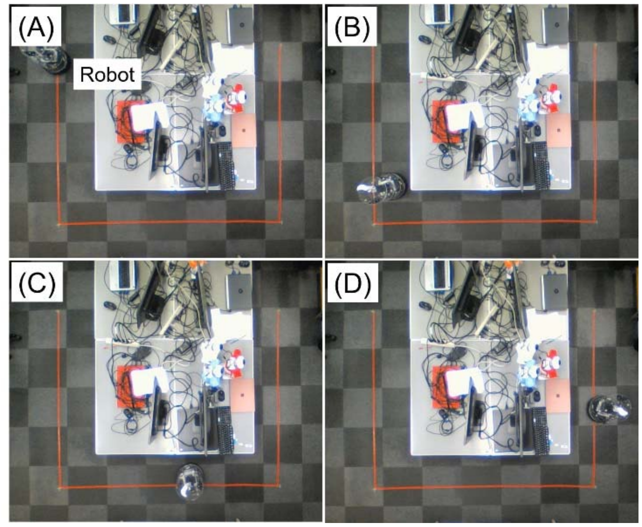

Fig. 9. An example of robot movement in the experimental environment.

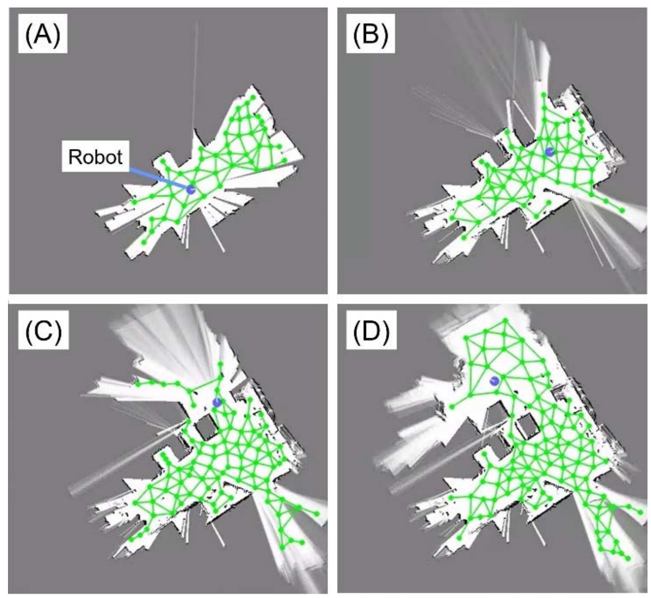

Fig. 10. GNG-based environmental feature extraction in perceptional system.

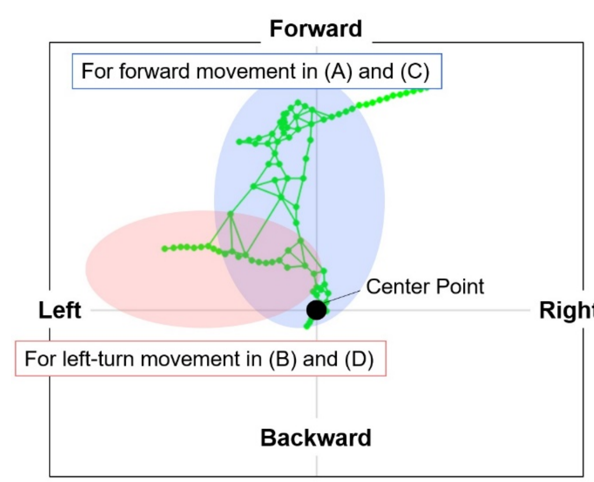

Fig. 11. GNG-based behavioral feature extraction in acting system.

method, the spatiotemporal pattern can be extracted by the SNN as depicted in Fig. 7. For example, in (A) and (C) described in the figures, the robot was controlled by the operator in order to move forward along the path. The feature points generated in the forward direction of the action system were associated with the nodes of the perceptional system by Hebbian learning. The weight parameters in the associative memory can be used for sensitivity control of the interface. In future works, we intend to discuss the applicability of the proposed approach to evaluate patients' cognitive and physical abilities in terms of their rehabilitation.

V. Summary

This paper presented a robotic avatar system for supporting patients with disabilities. We proposed a method of cognitive modeling while teleoperating in the robotic avatar system based on perceiving-acting cycle. A self-organized neural network based on unsupervised learning is used for the modeling of sensory and action information described through interaction with environment. Furthermore, a spiking neural network was organized for the spatial-temporal modeling of interaction between an operator and environment. According to the results of an experimental example, the teleoperation system can change the sensitivity of the interface corresponding to their operation. As future works, we intend to conduct some additional experiments in a clinical setting and evaluate the effectiveness of the proposed approach to support patients with disabilities.

References

[1] Y. Shiroishi, K. Uchiyama, N. Suzuki, "Society 5.0: For Human Security and Well-Being," Computer, Vol.51, Issue.77, 2018.

[2] M. Kitsuregawa, "Transformational Role of Big Data in Society 5.0," In Proceedings of IEEE International Conference on Big Data 2018 (Big Data 2018), 2018.

[3] G. Hirzinger, B. Brunner, J. Dietrich, J. Heindl, "ROTEX-the first remotely controlled robot in space" In Proceedings of IEEE International Conference on Robotics and Automation 1994 (ICRA 1994), 1994.

[4] T. Kataoka, K. Funahashi, K. Tanida, K. Yashiro, "A study of sense of self-agency focused on cross-correlation and delay between action and effect in continuous operation," In Proceedings of International Conference on Artificial Reality and Telexistence & Eurographics Symposium on Virtual Environments 2016 (ICAT-EGVE 2016), 2016.

[5] T. Kataoka, K. Funahashi, K. Tanida, K. Yashiro, "A study of sense of self-agency focused on noise and delay in multiple input operation," In Proceedings of IEEE 6th Global Conference on Consumer Electronics 2017 (GCCE 2017), 2017.

[6] F. Kobayashi and H. Takahashi, "Target selection based on gaze likelihood in gaze input interface," In Proceedings of International Conference on Knowledge and Smart Technology 2016 (KST 2016), 2016.

[7] R. Y. Cristanti, R. Sigit, T. Harsono, D. C. Adelina, A. Nabilah, N. P. Anggraeni, "Eye gaze tracking to operate android-based communication helper application," In Proceedings of International Electronics Symposium on Knowledge Creation and Intelligent Computing 2017 (IES-KCIC 2017), 2017.

[8] S. Chandra, G. Sharma, S. Malhotra, D. Jha, A. P. Mittal, "Eye tracking based human computer interaction: Applications and their uses," In Proceedings of International Conference on Man and Machine Interfacing 2015 (MAMI 2015), 2015.

[9] H. Wang, X. Dong, Z. Chen, B. E. Shi, "Hybrid gaze/EEG brain computer interface for robot arm control on a pick and place task," In Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2015 (EMBC 2015), 2015.

[10] M. A. Eid, N. Giakoumidis, A. E. Saddik, "In Proceedings of A Novel Eye-Gaze-Controlled Wheelchair System for Navigating Unknown Environments: Case Study With a Person With ALS," IEEE Access, Vol.4, 2016.

[11] Y. K. Meena, H. Cecotti, K. Wong-Lin, G. Prasad, "A multimodal interface to resolve the Midas-Touch problem in gaze controlled wheelchair," In Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2017 (EMBC 2017), 2017.

[12] D. Gêgo, C. Carreto, L. Figueiredo, "Teleoperation of a mobile robot based on eye-gaze tracking," In Proceedings of Iberian Conference on Information Systems and Technologies 2017 (CISTI 2017), 2017

[13] S. Vickers, H. Istance, A. Hyrskykari, "Performing Locomotion Tasks in Immersive Computer Games with an Adapted Eye-Tracking Interface," ACM Transactions on Accessible Computing, Vol.5, No.1, 2013.

[14] J. Chun, B. Bae, S. Jo, "BCI based hybrid interface for 3D object control in virtual reality," In Proceedings of International Winter Conference on Brain-Computer Interface 2016 (BCI 206), 2016.

[15] E. L. M. Naves, T. F. Bastos, G. Bourhis, Y. M. L. R. Silva, V. J. Silva, V. F. Lucena, "Virtual and augmented reality environment for remote training of wheelchairs users: Social, mobile, and wearable technologies applied to rehabilitation," In Proceedings of International Conference on e-Health Networking, Applications and Services 2016 (Healthcom 2016), 2016.

[16] H. Masuta, and N. Kubota, "The Intelligent Control based on Perceiving-Acting Cycle by using 3D-range camera," In Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics (SMC2009), 2009.

[17] T. Kohonen, "Self-Organization and Associative Memory," Springer-Verlag, 1984.

[18] T. M. Martinetz and K. J. Schulten, "A "neural-gas" network learns topologies," Artificial Neural Networks, vol. 1, pp. 397-402, 1991.

[19] B. Fritzke, "A growing neural gas network learns topologies," Advances in Neural Information Processing Systems, vol. 7, pp. 625-632, 1995.

[20] B. Fritzke, "Growing self-organizing networks – why?," European Symposium on Artificial Neural Networks, pp. 61-72, 1996.

[21] B. Fritzke, "Unsupervised clustering with growing cell structures," Neural Networks, vol. 2, pp. 531-536, 1991

[22] B. Fritzke, "Growing cell structures – a self-organizing network in k dimensions," Artificial Neural Networks, vol. 2, no.2, pp. 1051-1056, 1994.

[23] J.A. Anderson and E. Rosenfeld, "Neurocomputing," The MIT Press, Cambridge, Massachusetts, US, 1988.

[24] J.S.R. Jang, C.T. Sun, and E. Mizutani, "Neuro-Fuzzy and Soft Computing," Prentice-Hall Inc., 1997.

[25] W. Gerstner and W.M. Kistler, "Spiking Neuron Models. Single Neurons, Populations, Plasticity," Cambridge University Press, 2002.

[26] W. Gerstner, "Pulsed Neural Networks," W. Maass and C.M. Bishop (Ed.), MIT Press, Cambridge, Massachusetts, pp.3-53, 1999.