Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

The Intelligent Control based on Perceiving-Acting Cycle by using 3D-range camera

Hiroyuki Masuta
Tokyo Metropolitan University
Department of System Design
6-6 Asahigaoka Hino Tokyo Japan, 191-0065
masuta-hiroyki@sd.tmu.ac.jp

Naoyuki Kubota
Tokyo Metropolitan University
Department of System Design
6-6 Asahigaoka Hino Tokyo Japan, 191-0065
kubota@tmu.ac.jp

***Abstract*- This paper discusses a integrated perceptual system for intelligent control of a service robot. The robot should be able to perceive the environment flexibly to realize intelligent behaviors. We focus on the perceptual system based on the perceiving-acting cycle discussed in ecological psychology. We propose a perceptual system composed of the retinal model and the spiking-neural network to realize concept of the perceiving-acting cycle. The proposed method is applied to the robot arm equipped with a 3D-range camera. This proposed method features the robot detects the invariant information of a dish, for example the contrast of a distance Information or a luminance information. As experimental results, We show that the integrated perceptual system can adapt effectively in the dynamic environment.**

Keywords- **Arm Robots, Perceiving-Acting Cycle, 3D-Range Camera, Spiking-neural network**

I. INTRODUCTION

RECENTLY, various types of intelligent robots have been developed for the society of the next generation. In particular, intelligent robots should continue to perform tasks not only in given environments like factories but also real environments such as houses, commercial facilities and public facilities. These intelligent robots equip a lot of sensors for measuring environment. The recent sensors are very advanced such as high definition CCD (charge-couple device) cameras, scanning range finders and 3D-range cameras. The intelligent robots can measure huge volume of information at a time. However, these robots are difficult to detect necessary information from the huge volume of information in real time. Despite this, the intelligent robots should perceive the facing situation and take suitable action in real time.

We also are developing intelligent service robot systems for clearing dishes from the table in a restaurant. Fig.1 shows the plan of an intelligent service robot system overview. The interaction tool recognizes an order from human gestures or voice [1]. The robot arm removes dishes from the table to a tray robot. At the same time, the robot arm gets

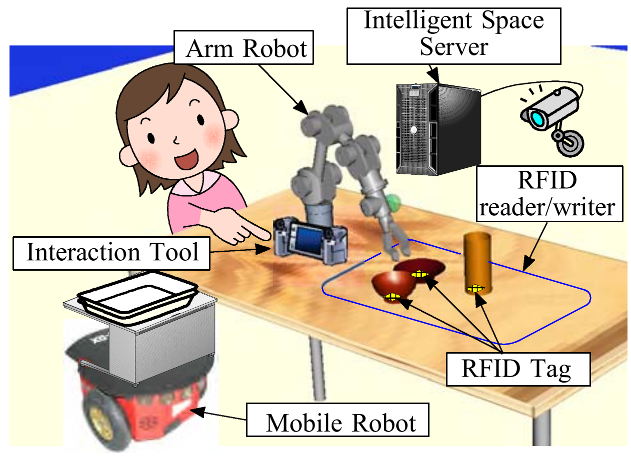


Fig. 1. The intelligent service robot system overview

environmental information such as size, weight and position of dishes from an intelligent space server [2]. An intelligent space server keeps newest environmental information by using CCD cameras and RFID (radio-frequency identification device) tags. In this paper, we develops the control of robot arms for clearing the table by using 3D-range camera. The 3D-range camera can measure depth distance and luminance at the same time. The robot arm with 3D-range camera should consider huge volumes of environmental data.

The researches on visual systems also consider the huge volumes of data. The researches on visual systems are very active because the intelligent robots can get necessary information from CCD cameras by various image processing method. Various types of features such colors and shapes can be extracted from a camera image [3], [4]. These features are used for perceiving the environment. Therefore, we have discussed the methodologies for the active perception by CCD cameras [5]. However, it is difficult to detect necessary information due to the lighting condition and high computational cost.

On the other hand, human can perceive necessary information easily from huge volumes of environmental information. In the field of psychology, Gibson developed ecological approach to visual perception [6]. Ecological approach emphasizes the importance of perception and action

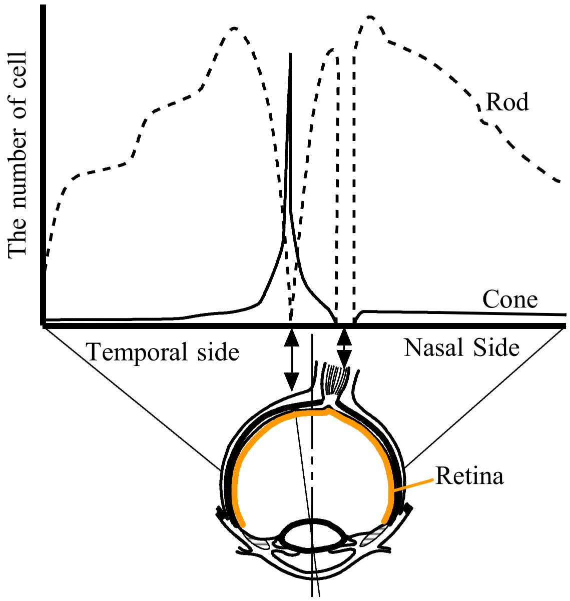


Fig. 2. Cell distribution in the retina

interacting with an environment. There are many important concepts, such as invariant, affordance, resonance and perceiving-acting cycle in ecological approaches. In ecological psychology, the smallest unit of analysis must be the perceiving-acting cycle situated in an intentional context. The coupling of perceptual system and action system is very important. The perceptual system extracts perceptual information to be used for making action outputs. Here the importance is to extract perceptual information for persisting with a series of motions to take an intentional action. In addition, the output of the action system constructs the spatiotemporal context for the specific perception with the dynamics of the environmental change. The perceptual system must search for and select perceptual information required by a specific action. Therefore, we have been proposed the perceptual system based on the ecological psychology [7].

For all of these discussions, we propose the control method based on the perceiving-acting cycle. Especially, we focus on the perceptual system for the active perception and the direct perception of intelligent robots. We experiment with a dish detection task by using a robot arm with 3D-range camera for the effectivity of the proposed method.

This paper is organized as follows; Section II explains human visual perception from viewpoint of neurophysiology and ecological psychology. Section III explains our robot system and explains the proposed method based on the perceiving-acting cycle. Section IV explains experimental results. Finally, Section V concludes this paper.

II. Visuoperceptual Function of Human

This section explains visual perception of a human. The model of visual perception is discussed in terms of neuropsychology and ecological psychology. While neuropsychology can be discussed from the sensing devices, ecological psychology can be discussed from the perceptual systems. Therefore, we should sequentially integrate the theories and concepts of these two fields.

A. The Retinal Structure Based on a Viewpoint of Neurophysiology

The retina is a light sensitive tissue lining the inner surface of the eye. The retinal receptors convert the light patterns into neural signals. The retinal receptors can be divided into two groups of cones and rods. Cone cells are associated with daylight vision and color vision, while rod cells are associated with night vision. Fig.2 shows the cell distribution of the retina [8]. The solid line shows the distribution of cone cells in the center of retina. The cone cells have an advantage of perception of high discriminating sensitivity on colors and shapes, but it is difficult for them to perceive fast-paced action over 5 [Hz]. The density of retinal receptors is not uniform over the retina, but the density is the highest in the fovea and gradually becomes low towards the periphery of the retina. This is called foveal vision. This means the foveal vision can perceive the detailed static information in the narrow range of attentiveness. The dash line shows the distribution of rod cells on the retina. The rod cells have an advantage of perception for quick motions, but they have a disadvantage of low discriminating sensitivity on colors and shapes. This is called peripheral vision. This means the peripheral vision can perceive the dynamic information in a wide range. Thus a human can perceive the environment flexibly by collaboration with different functions.

B. The Invariant for Perceiving-Acting Cycle

This subsection explains the concept of perceiving-acting cycle based on ecological psychology. The concept of perceiving-acting cycle maintains the coupling of perceptual system and action system in ecological psychology [6]. The perceptual system and the action system restrict each other through interaction with the environment. When the perceiving-acting cycle forms a coherent relationship with the environment, the specific perceptual information generates the specific action outputs like reactive motions. We consider the couple of perceptual system and action system situated to the facing environment as one phase of the perceiving-acting cycles. In particular there is an important idea that is invariant on perceptual system in above discussion.

The invariant is not obtained by inferences from input data but is obtained by direct perception. For example, Lee verified that the τ-coupling is used as a direct perception when a human perceives the approach of objects [9]. Turvey verified that the dynamic touch is also considered as one of direct perceptions when a human perceives the length of objects without using eyesight [10]. The common theory of the above examples is that a human directly perceives the important information required for taking suitable actions like τ-coupling. The invariants extracted as the perceptual information can be obtained based on the coupling structure of the physical body with the environment. Therefore, we

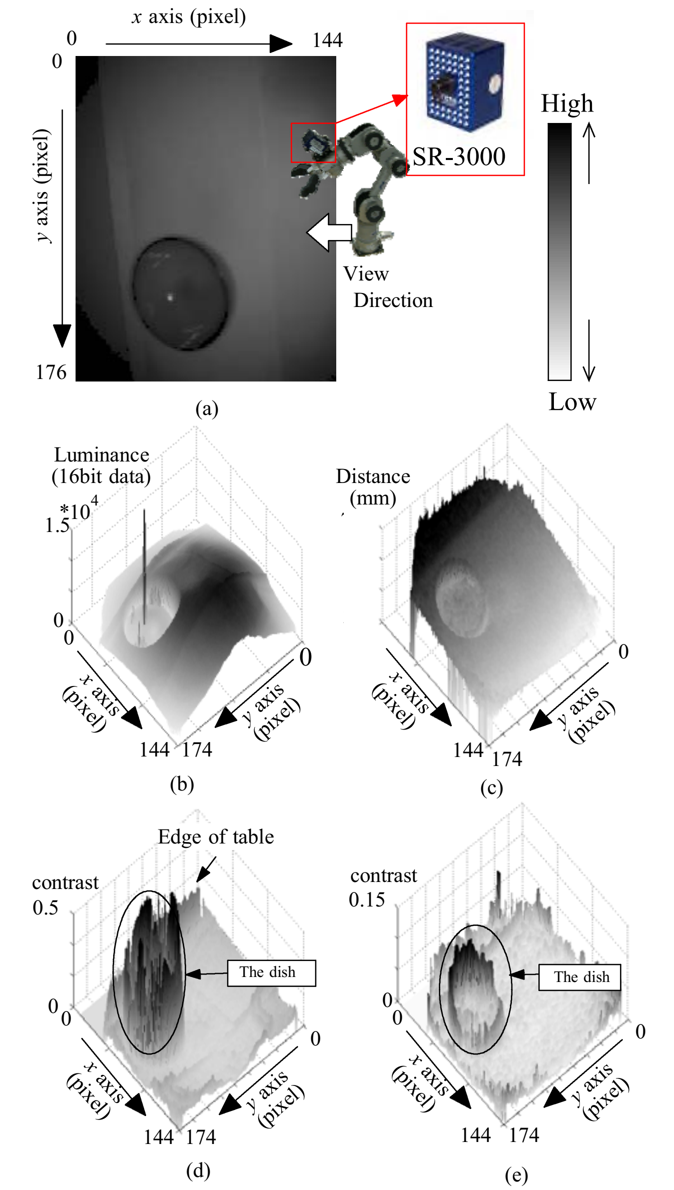


Fig. 3. Preliminary experimental results of contrast extraction filter (a)The image data view by gray scale image. (b)Luminance data is measured by SR-3000. (c)Distance data is measured by SR-3000. (d)The result of contrast extraction filter output by luminance data. (e)The result of contrast extraction filter output by distance data.

should take the concept of direct perception into account to realize the flexible perception, and we apply the concept of perceiving-acting cycle for the control of the perceptual system. The essence of the perceiving-acting cycle is to extract and use invariants.

III. Perceptual System Based on Perceiving-Acting Cycle

This section explains the proposed method based on the perceiving-acting cycle. First, we explains the robot arm and the 3D-range camera for experiments. Next, we propose the perceiving-acting cycle model based on the Ecological Psychology and the Neurophysiology.

A. The Intelligent Service Robot System Overview

We use the robot arm for clearing the table robustly in a restaurant. We use the robot arm "Katana" that has been developed by Neuronics AG. The Katana has 5-DOF (degree-of-freedom) and has 6 motors including a gripper. The Katana is bolted to the table. The control orders of Katana are each motor speed and position. In the control of Katana, we apply a concept of a robot technology middleware (RT-Middleware) [11]. The Katana equips the lower part of the hand position with a 3D-range camera "SR-3000" for measuring environment (see Fig.3 (a)). The SR-3000 has been developed by MESA Corporation [12]. The SR-3000 is range camera that can measure 3-dimensional distance up to 7.5 [m]. The SR-3000 can output the measuring data as luminance data and distance data on the Cartesian axis. The resolution of SR-3000 data is a spatial QCIF (176×144 pixels) (see Fig.3 (a)). Therefore, the intelligent robot with SR-3000 should consider huge volumes of data, in this case the distance data and luminance data are 25344-direction respectively at a time.

B. The Invariant for the Detection of a Dish

This subsection explains the proposed method based on the invariant. The basic function of a retina detects the high contrast areas in the visual information on neuropsychologic study. Therefore, we apply a filter of contrast extraction based on neighborhood processing. The value of the point, $I[n_0,n_1]$, in an original image is updated into the value of the point, $I'[n_0,n_1]$, by the filter based on the mask of 5×5 pixels in the following equation (1).

$$I'[n_0,n_1] = \frac{v_{max} - v_{min}}{v_{max} + v_{min}} \quad (1)$$

where v_{max} is a maximum value in the mask, v_{min} is a minimum value in the mask.

Fig.3 shows the preliminary experimental results of a contrast extraction filter applied to the luminance and the depth distance. Fig.3 (a) is a gray scale image based on the luminance data measured by SR-3000. There is a bowl in the bottom of the image. Fig.3 (b) and (c) are original data view of the luminance and the depth distance respectively. Fig.3 (d) and (e) are contrast the results obtained by the extraction filter of the luminance and the depth distance respectively. In these results, the contrast of a bowl is strengthened on both the luminance and the depth distance. However, the edge of table is strengthened in the result of luminance, because the difference of shading is strengthened (Fig.3 (b), (d)). On the other hand, the difference of concavo-convex shape is strengthened in the result of depth distance (Fig.3 (c), (e)). Therefore, the filter used to the distance data is less affected to the edge of the table. The luminance and depth distance should be considered based on the environmental situation.

C. The Retinal model based on foveal vision

This subsection shows the proposed method based on the retina. The example of removing dishes, humans perceive a size, a color and a shape of a dish, while they do not need

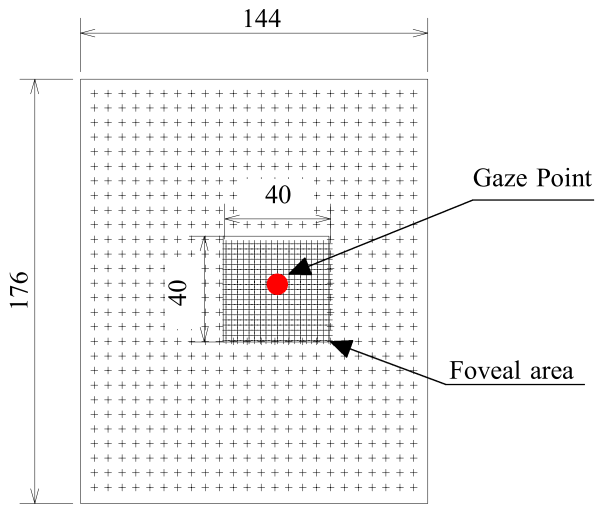


Fig. 4. A retinal model for foveal vision

perceive a quick action. In this case humans perceive a dish by the ability of the foveal vision basically. Therefore, we focus on the mechanism of the foveal vision, and propose a method for updating the density corresponding to the visual attention used for the information extraction. In this paper, we propose the retinal model to perceive a dish based on the foveal vision. Fig.4 shows a model for realizing the mechanism of the foveal vision. The center of the visual field is defined a 40×40 pixel area at the surround of a gaze point as the foveal area, because the angle of 40 pixels in SR-3000 is corresponding to the angle of 10 degrees in foveal vision (see Fig.2). In this area measured at every 2 [pixel], and the periphery of the visual field is located at every 6 [pixel]. Also the gaze point can be controlled for detecting the necessary information. By using this mechanism, the robot can pay attention to only necessary information and ignore other unnecessary information. In addition, the number of inputs is reduced into 1060 from 25344. Therefore, the proposed method can be expects to reduce high computational cost.

D. Spiking-Neural Network for Perceptual System

In the perceiving-acting cycle model, we should consider the interaction with a perceptual system and an action system, and also should consider a spatiotemporal context generated by their interaction. The artificial neurons are classified into pulse-coded and rate-coded neuron models from the viewpoint of abstraction level. A pulse-coded neural network is often called a spiking neural network (SNN). We apply the SNN to realize the concept of perceiving-acting cycle for extracting a dish. There are two main reasons for applying the SNN. The first one is that the SNN can be considered as a model dealing with the spatiotemporal context easily than the other NN, because the SNN has internal state based on the capability of temporal coding and the capability of noise reduction by integration of input data. The second one is that the SNN is highly harmonic with the discrete-time system like embedded device, because the SNN outputs time series of pulses.

We use a simple spike response model to reduce computational cost [13]. The internal state of the ith neuron, $h_i(t)$, is calculated as follows;

$$h_i(t) = \tanh\left(h_i^{syn}(t) + h_i^{ref}(t) + q_i(t)\right) \tag{2}$$

$$h_i^{syn}(t) = \gamma^{syn} \cdot h_i(t-1) + \sum_{j=1, j\neq i}^{N} w_{j,i} \cdot p_j^{EPSP}(t-1) \tag{3}$$

where γ^{syn} is the discount rate (0< γ^{syn} <1.0); $w_{j,i}(t)$ is weight from jth neuron to ith neuron; $p_j^{EPSP}(t)$ is an excitatory postsynaptic potential (*EPSP*) of the jth neuron at the discrete time t; $h_i^{ref}(t)$ is the refractoriness of the neuron; $q_i(t)$ is an external input; N is the number of neurons. When the ith neuron is fired, R is subtracted from $h_i^{ref}(t)$ in the following,

$$h_i^{ref}(t) = \begin{cases} \gamma^{ref} \cdot h_i(t-1) - R & if \quad p_i(t-1) = 1 \\ \gamma^{ref} \cdot h_i(t-1) & otherwise \end{cases} \tag{4}$$

where γ^{ref} is the discount rate (0< γ^{ref} <1.0). When the $h_i(t)$ is larger than the predefined threshold θ_i for firing, a presynaptic spike output is done as follows;

$$p_i(t) = \begin{cases} 1 & if \quad h_i(t) \geq \theta_i \\ 0 & otherwise \end{cases} \tag{5}$$

The presynaptic spike output is transmitted to the next layer according to *EPSP*. The *EPSP* of the ith neuron, $p_i^{EPSP}(t)$, is calculated as follows;

$$p_i^{EPSP}(t) = \frac{1}{T}\sum_{n=0}^{T} \kappa^n p_i(t-n) \tag{6}$$

where κ is the discount rate (0<κ<1.0); T is the time sequence to be considered. The learning of weight parameters is performed based on the temporal Hebbian rule in the following;

$$w_{j,i} \leftarrow \tanh(\gamma^{wgt} w_{j,i} + \xi^{wgt} \cdot p_i^{EPSP}(t) \cdot p_j^{EPSP}(t-1)) \tag{7}$$

where γ^{wgt} is the discount rate for weight updating, ξ^{wgt} is the learning rate.

Fig.5 shows the layout of the spiking neurons composed of three types of sensor neurons. First one is sensor neurons for detecting the input of contrast in SR-3000 data. Each neuron is assigned uniformly in the same interval, and sums up the output after the retinal model. Second one is action input neurons for the target velocity of lengthwise arm movement. The last one is action input neurons for the target velocity of widthwise arm movement. These sensor neurons are called D-Neuron, L-Neuron, and W-Neuron respectively. Fig.6 shows the relationship between D-neurons, L-neurons and W-neurons. D-Neurons are connected each other, and also L-Neurons and W-Neurons are connected to each D-neurons. However, unidirectional connection exists from the L-Neurons to W-Neurons, and from W-Neurons to L-Neurons. L-Neurons and W-Neurons play the role in the prediction of the future position of the detection of a dish occurring from the moving behavior of the robot arm.

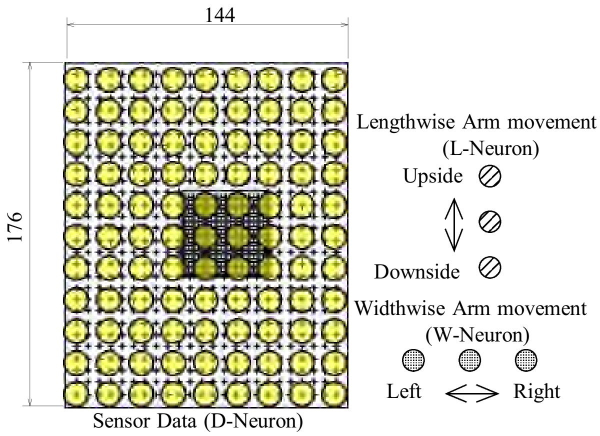


Fig. 5. Layout of the SNN neurons

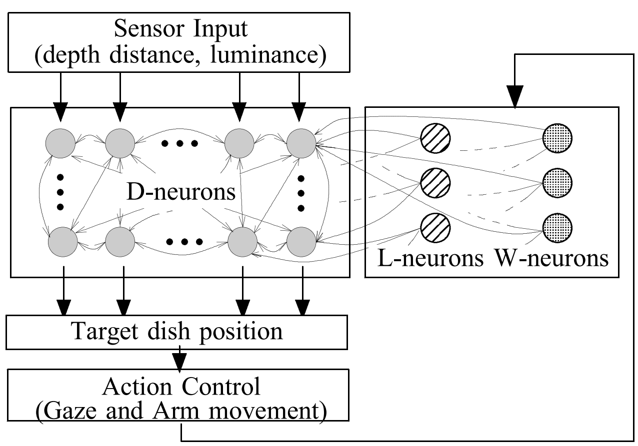


Fig. 6. The relationship between D-neurons, L-neurons and W-neurons

The most frequently fired neuron, that means the neuron with the highest *EPSP*, is selected for calculating the relative position. The position of the selected neuron's is related to that a dish when the highest *EPSP* value is larger than the threshold including hysteresis. Furthermore, the central position of the retinal model is updated by using the relative position. In this way, the neurons detect the relative position of the dish, and the measured distance corresponding to the sensing direction used in the neuron, is available as the input for the retinal model.

Fig.7 shows the flow chart of all algorithms. We consider that the perceptual system is composed of the integration with different specific perceptual module. In previous researches, we have shown the effectivity of the perceiving-acting cycle model based on uni-modal sensory input. However, the proposed method based on uni-modal sensory input could not consider the suitable sensory input in the facing environment [7]. Therefore, we propose the perceptual system that integrates with the luminance perception and the depth distance perception. The luminance perception module and the depth distance module are calculated in parallel at the same time. And the integrated outputs $PI_j^{EPSP}(t)$ are calculated as follows.

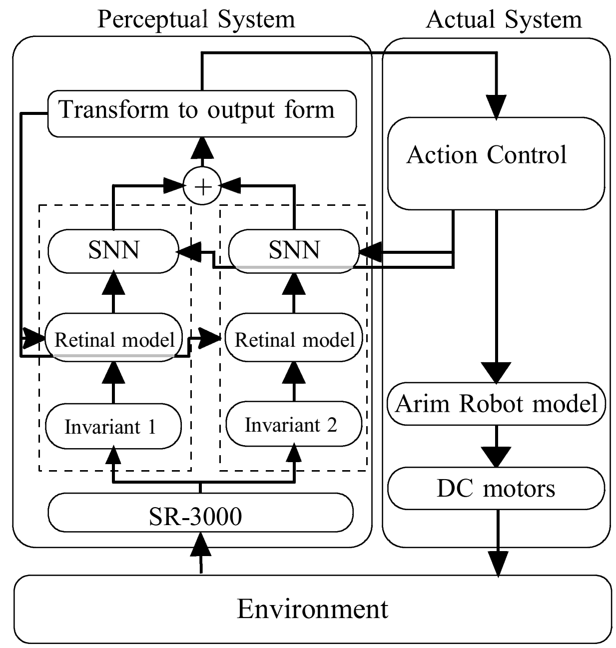


Fig. 7. Flowchart based on perceiving-acting cycle in multiple perceptual system

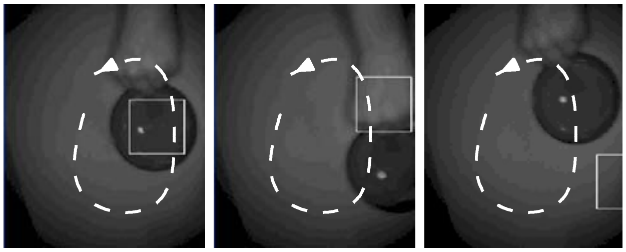
$$PI_i^{EPSP}(t) = PL_i^{EPSP}(t) + PD_i^{EPSP}(t) \tag{8}$$

$PL_i^{EPSP}(t)$ and $PD_i^{EPSP}(t)$ are the output of (6) that the input is the luminance and the depth distance respectively. The highest *EPSP* neuron is selected and the next gaze position. This proposed method is expected to detect a dish well, because the dish on the table has higher luminance contrast and the higher depth distance contrast.

IV. EXPERIMENTAL RESULTS

This section shows experimental results. The experimental condition is shown as follows. The position of the robot arm is fixed and the change on the table is observed by SR-3000. A bowl is moved around on the table by human hand. The task is to detect and trace a bowl by the proposed perceptual system in such a dynamic environment. We experiment on the dish detection by using 3 types of the sensory input. The first is the proposed method with both inputs of luminance and depth distance (Case 1). The second is the proposed method with only inputs of depth distance (Case 2). The third is the proposed method with only inputs of luminance (Case 3).

Fig.8 and Fig.9 shows experimental results of the integrated perceptual system in the dynamic environment. Fig.8 (a), (b) and (c) show snapshots of Case 1, Case 2 and Case 3 respectively. Case 1 can detect a dish correctly (see Fig.8 (a)). Case 2 detects a human hand, and Case 3 detects the edge of image. Fig.9 shows experimental results of gaze movement. Fig.9 (a) shows gaze movement of lengthwise direction on SR-3000 image. Fig.9 (b) shows gaze movement of widthwise direction on SR-3000 image. The heavy green line indicates the trajectory of a dish movement. In this



(a) (b) (c)

Fig. 8. Snap shots of each experimental results (a)The sensory inputs are depth distance and luminance (b)The sensory input is only depth distance (c)The sensory input is only luminance.

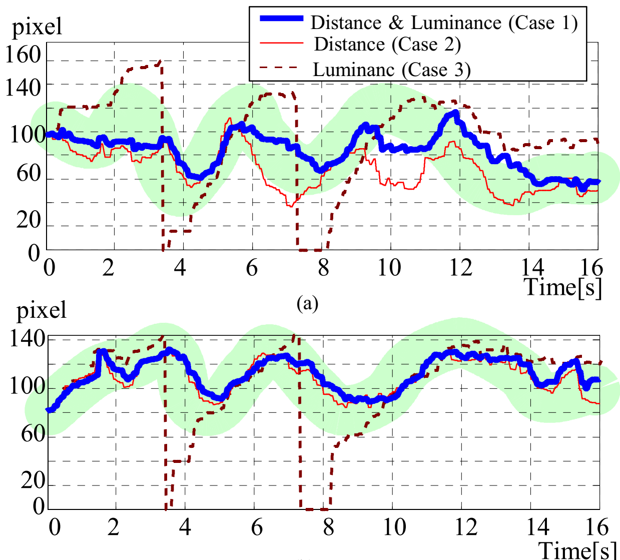


Fig. 9. Experimental results of gaze movement (a)Gaze movement of widthwise direction (b)Gaze movement of lengthwise direction

result, Case 1 and Case 2 can trace a dish, however Case 2 fails detection of a dish like Fig.8 (b) at 10 second. This reason is that it is difficult to discriminate between the height of a dish and the height of a human hand from the table. Case 3 fails detection of a dish like Fig.8 (c) at 3 or 8 second, because the contrast of luminance is high at edge of image. That is the reason why the contrast of luminance is high near the peripheral area of the image, because SR-3000 is an active sensor and its irradiating light is not uniform. Therefore luminance has highly difference between the central area of the image and the peripheral area of the image. In these experimental results, the integrated perceptual system can perceive environment suitably, because the proposed method consider both features of depth distance and luminance. Therefore, the integrated perceptual system can adapt effectively in the dynamic environment.

V. Conclusion

This paper proposed the perceptual system for an intelligent service robot based on the perceiving-acting cycle. The model of perceiving-acting cycle constructed from the retinal model and the SNN. We show that the integrated perceptual system can adapt effectively in the dynamic environment. Furthermore, the integrated perceptual system design is easy because each specific perception is designed separately. On the other hand, we do not give a specific target because the main purpose verifies the model based on the perceiving-acting cycle. To realize target dish detection, the robot should update the input parameters based on the specific target. As future works, we need consider a planning phase that important to concept of the situation-intention cycle.

Acknowledgment

This research is being as a part of the "Development Project for an Intelligent Technology of Next-Generation Robots" consigned by New Energy and Industrial Technology Development Organization (NEDO).

References

[1]E.Sato, T.Yamaguchi, and F.Harashima, "Natural Interface Using Pointing Behavior for Human-Robot Gestural Interaction", IEEE transactions on Industrial Electronics, Vol.54, No.2, 2007, pp.1105-1112.

[2]N. Y. Chong, H. Hongu, K. Ohba, S. Hirai, K. Tanie, "Knowledge Distributed Robot Control Framework," Proc. Int. Conf. on Control, Automation, and Systems, Gyeongju, Korea, Oct. 22-25, 2003.

[3] K.Deguchi, "Active Acquisition of 3D Map in Robot Brain by Combining Motion and Perceived Images", First International Symposium on Brain, Vision, and Artificial Intelligence, 2005, pp.386-395.

[4] D.Marr, "Vision: A Computational Investigation into the Human Representation and Processing of Visual Information", W H Freeman & Co, 1982.

[5]N.Kubota, K.Nishida, "Perceptual Control Based on Prediction for Natural Communication of A Partner Robot", IEEE Transactions on Industrial Electronics, Vol.54, No.2, 2007, pp.866-877.

[6] J.J.Gibson, "The ecological approach to visual perception", Hillsdale, NJ: Lawrence Erlbaum Associates, 1979.

[7]H.Masuta, N.Kubota, "Perception in A Partner Robot Based on Human Visual System", International Journal of factory automation, robotics and soft computing, Issue 2, 2008, pp.48-55.

[8]C.A.Curcio, K.R.Sloan, R.E.Kalina, "Human photoreceptor topography", Journal of Comparative Neurology, 292, 1960, pp.497-523.

[9]D.N.Lee, "Guiding movement by coupling taus", Ecological psychology, 10(3-4), 1998, pp.221-250.

[10]M.T.Turvey, R.E.Shaw, "Ecological foundations of cognition", I. Journal of Consciousness Studies, No. 6, 1999, pp.95-110.

[11]N.Ando, et al., RT-Middleware: Distributed Component Middleware for RT (Robot Technology), 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2005), pp.3555-3560, 2005.

[12]T. Oggier, et al., "An all-solid-state optical range camera for 3D-real-time imaging with sub-centimeter depth-resolution (SwissRanger)", Proceedings of SPIE Vol. 5249, 2003, pp.634-545.

[13]W.Maass, and C.M.Bishop, "Pulsed Neural Networks", The MIT Press, 1999.