Saliency-based Scene Learning and Recognition based on Competitively Growing Neural Network using Temporal Coding

Masayasu Atsumi
Dept. of Information Systems Sci., Faculty of Eng., Soka University
1-236 Tangi-cho, Hachioji-shi, Tokyo 192-8577, Japan
E-mail: matsumi@t.soka.ac.jp

***Abstract*— This paper discusses saliency-based scene learning and recognition in which objects in attended spots are quickly learned and recognized based on the competitively growing neural network using temporal coding. This neural network represents objects using latency-based temporal coding and grows size and recognizability through learning and self-organization. Through simulation experiments of a robot equipped with a camera, it is shown that quick self-organized learning and glance recognition of objects in scenes are well performed by our model.**

I. Introduction

A human can learn scenes in almost one shot and also recognize them at a glance. In these processes, spatially circumscribed regions of the visual field are selected based on saliency-based attention as well as volition-controlled attention before further processing. The former is rapid, bottom-up and task-independent attention and the latter is slow, top-down and task-dependent attention [3] [4]. These small regions are highly processed through the cortical visual hierarchy, and as a result scene memory is considered to be composed of attended objects, complexes of objects and their spatial relation from the egocentric point of view.

In this paper, we discuss saliency-based scene learning and recognition in which objects in attended spots are quickly learned and recognized based on the competitively growing neural network using temporal coding [5], which is named the COGNET (COmpetitively Growing NEural network using Temporal coding). In learning and recognition, objects in attended spots are sequentially encoded to be invariant with respect to position and size by this network and their positions and sizes are encoded simultaneously [6]. In this network, objects are internally represented using latency-based temporal coding. This network enables fast self-organized learning of objects based on recruiting neurons and similarity-based sorting of neurons and quick glance recognition of objects based on the latency-based temporal coding.

This paper is organized as follows. In section II, a model of saliency-based scene memory is outlined, and then in section III, the COGNET is described in detail. In section IV, learning and recognition performance of objects and scenes, especially quick self-organized learning and glance recognition performance of objects in scenes, are evaluated through simulation experiments of a robot equipped with a camera.

II. A Model of Scene Memory

A model of saliency-based scene memory is shown in Figure 1. In the retina of primates, early visual features such as contrast, its shift and opponent colors are processed before further "what" and "where" visual processing. Also it is reported that the parietal and frontal cortices, the pulvinar nuclei of the thalamus and the superior colliculus are involved in visual attention [7]. In the first phase of our model, contrast and opponent color channels of red, green, blue and yellow are computed at each pixel of a scene image [4]. Then the saliency map is produced to represent saliency at every pixel of the image by combining its contrast and opponent color channels [2]. Next, connected object regions are segmented using a grow-and-merge method on the saliency map and their bounding boxes are extracted as attended spots. Three or less attended spots are extracted.

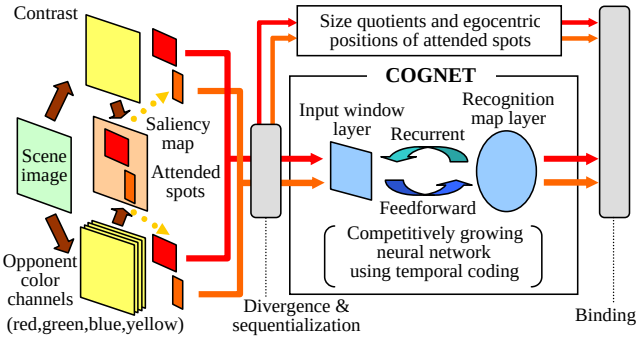

Fig. 1. A model of saliency-based scene memory

In the second phase of our model, attended spots are processed in the order of decreasing saliency at a certain interval. This interval of time corresponds to an interval at which the focus of attention jumps from one attended spot to the next. In this phase, an object in each attended spot is encoded to be invariant with respect to position and size by the COGNET, and the position and the size of the attended spot are encoded at the same time. These correspond to visual processing in the ventral or "what" visual pathway and the dorsal or "where" visual pathway respectively.

The COGNET consists of the input window layer and the

recognition map layer. The input window layer consists of $l_w \times l_w \times 5$ neurons that receive normalized values of contrast and four opponent colors at $l_w \times l_w$ sections in height and width on an attended spot. The recognition map layer is a competitively growing layer where neurons can be arranged in a two-dimensional lattice. An object in an attended spot is encoded by the winner neuron in the recognition map layer. The input window layer and the recognition map layer have the entire reciprocal connection by feedforward and recurrent synapses. In this network, it is possible to recognize an object in an attended spot when it is given to neurons in the input window layer, and it is also possible to recall a mental image of an object when a neuron in the recognition map layer is stimulated. Meanwhile, the position of an object is expressed by the center coordinate of its bounding attended spot in the egocentric coordinate system whose origin is the center of a scene image. The size of an object is expressed by the size quotient $q = \frac{l_s}{l_w}$ where l_s is the larger of height and width of the attended spot. The size quotient represents relative distance of the same object in different views or relative size of different objects in a view.

Finally, consecutive processing results of attended spots in a scene are bound together and a scene memory is given by a set of triplets each of which consists of the winner neuron, the center coordinate and the size quotient obtained for each attended spot in the scene. We call this triplet an attended spot code and a set of triplets a scene code.

III. Competitively Growing Neural Network Using Temporal Coding

A. Latency-based Temporal Coding

In neural network models, information is internally represented in the form of a rate code or a temporal code. In Kohonen's self-organizing map [10] as a representative competitive neural network, rate coding is in general used though self-organizing maps of spiking neurons using temporal coding is also proposed [1]. Several temporal coding schemes have been proposed [9] [11]. In the COGNET, latency from trigger such as pulse transmission or stimulation to a neuron until firing is used as a temporal code to encode information.

Each neuron in the input window layer converts an input value into a latent period so that the larger the input value is, the shorter the latent period is. As a result, a spatial pattern of input strength that codes an object is converted into a spatiotemporal pattern of pulse transmission, and the object is internally represented by this pattern. A neuron in the input window layer is formulated as follows. When an external excitatory input or recurrent pulse transmission is given to a neuron n_i at time t, membrane potential $p_i(t)$ of n_i is computed by

$$p_i(t) = \sum_j (w_{ij}^R \times \delta_{ij}(t)) + ext_i(t) - ip_i(t) \tag{1}$$

where w_{ij}^R is recurrent synaptic efficacy from a neuron n_j in the recognition map layer, $\delta_{ij}(t)$ is a function which takes the value 1 if a pulse is transmitted from a neuron n_j at time t and 0 otherwise, $ext_i(t)$ is an external excitatory input, and $ip_i(t)$ is an inhibitory input. The output function $o_i(p)$ of the neuron n_i returns latency until firing as a function of membrane potential p when $p \geq 0$ and is given by

$$o_i(p) = \lambda \times \max(1 - p, 0) \tag{2}$$

where λ is a constant called the validity term of latency and takes the value 255. According to these formulas, the neuron n_i fires at time $t + o_i(p)$ so long as there is no strong inhibition.

In neurons in the recognition map layer, membrane potential is computed based on a spatiotemporal pattern of pulse arrival, that is, a pattern of latency from pulse arrival until the competition time. For a neuron n_i in the recognition map layer, membrane potential $p_i(t)$ at time t is computed by

$$p_i(t) = \sum_j ep_{ij}(t) + ext_i(t) - ip_i(t) \tag{3}$$

$$ep_{ij}(t) = w_{ij}^F \times \frac{kp_{ij}(t)}{N_i^P(t)} \tag{4}$$

where $ext_i(t)$ is an external excitatory input, $ip_i(t)$ is an inhibitory input, w_{ij}^F is feedforward synaptic efficacy from a neuron n_j in the input window layer, $kp_{ij}(t)$ is a kernel function, and $N_i^P(t)$ is a normalization function. These two functions are given by

$$kp_{ij}(t) = \begin{cases} t - t_{ij}^a & \cdots \quad 0 \leq t - t_{ij}^a \leq \lambda \\ 0 & \cdots \quad otherwise \end{cases} \tag{5}$$

$$N_i^P(t) = \sqrt{\sum_j kp_{ij}(t)^2} \tag{6}$$

where t_{ij}^a is a recent time of pulse arrival from a neuron n_j in the input window layer and λ is the validity term of latency. The $kp_{ij}(t)$ encodes latency from pulse arrival. Firing in the recognition map layer is determined through competition among neurons using $\{p_i(t_c)\}$ at a competition time t_c.

B. Outline of Neural Dynamics with Growth

Neural dynamics in learning and recognition of objects is controlled by a discrete-time clock as illustrated in Figure 2 and outlined as follow:

1) firings of neurons in the input window layer caused by an external input of $l_w \times l_w \times 5$ values sampled from an attended spot,
2) transmission of pulses to the recognition map layer and control of the competition time by the pre-competition inhibition imposed on neurons in the recognition map layer,
3) competitive firing of neurons in the recognition map layer at the competition time, which involves recruitment of a new neuron, modulation of feedforward synaptic efficiency and self-organization of neurons at learning,
4) transmission of pulses to the input window layer and control of firing by the post-competition inhibition imposed on neurons in the input window layer,

5) a repetitive external input, which involves modulation of recurrent synaptic efficiency at learning,
6) and repetition of these processes from 1) to 4) with no learning, in which if winner neurons in two repetitions are the same, it is judged to be a neuron that encodes the external input.

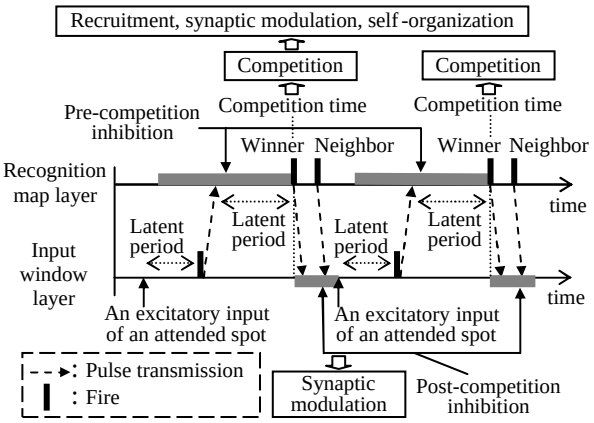

Fig. 2. An outline of neural dynamics in which an external input is inputted to the input window layer two times at a specified interval. Learning is performed only in the first repetition but not in the second repetition.

C. Competition with Recruitment

For the first external input, neurons in the input window layer fire with latent periods in $[0, \lambda]$ which are computed by the formulas (1) (2) whose inhibitory input is set to 0. This firing is transmitted to the recognition map layer. In the recognition map layer, the pre-competition inhibition continues in default for a period of λ and controls firing timing of neurons, that is, the competition time. Membrane potentials of all neurons at the competition time are computed using a spatiotemporal pattern of pulse arrival according to the formulas (3) (4) (5) (6). If the maximum membrane potential of them is above a threshold, that is called the threshold of discrimination, a neuron that takes the maximum membrane potential is selected as the winner neuron. Otherwise, a new neuron is recruited as the winner neuron, whose synaptic efficacy is initialized so that it takes the maximum membrane potential to the spatiotemporal pattern of pulse arrival.

Suppose t_{ij}^a to be a pulse arrival time from a neuron n_j in the input window layer to a recruit neuron n_i and t_c be a competition time. Then, feedforward synaptic efficiency w_{ij}^F from n_j to n_i and recurrent synaptic efficiency w_{ji}^R from n_i to n_j are chosen by

$$w_{ij}^F = \frac{kp_{ij}(t_c)}{N_i^P(t_c)} \tag{7}$$

$$w_{ji}^R = \frac{kp_{ij}(t_c)}{\lambda} \tag{8}$$

where

$$kp_{ij}(t) = \begin{cases} t - t_{ij}^a & \cdots \quad 0 \leq t - t_{ij}^a \leq \lambda \\ 0 & \cdots \quad otherwise \end{cases} \tag{9}$$

$$N_i^P(t) = \sqrt{\sum_j kp_{ij}(t)^2} \tag{10}$$

and λ is the validity term of latency. According to these formulas and (3) (4) (5) (6), membrane potential of the recruit neuron is 1. Since the maximum membrane potential is 1 when there is no external input, it is clear that the recruit neuron becomes the winner neuron.

The winner neuron fires at the competition time t_c. Let n_s be the winner neuron. Then every neuron n_k whose membrane potential is above a threshold η is selected as a neighbor neuron of n_s and fires with a lag proportional to difference of membrane potential between n_k and n_s. That is, n_k fires at $t_c + lag(n_s, n_k, t_c)$ which is given by

$$lag(n_s, n_k, t_c) = \gamma \times (p_s(t_c) - p_k(t_c)) \tag{11}$$

where η is called the threshold of neighborhood and γ is called the constant of firing lag at competition.

D. Self-organized Learning

The winner neuron and its neighbor neurons modulate their synaptic efficacy just after firing so that they may become easy to fire for the same pattern of pulse arrival. That is, synaptic modulation is performed for the winner neuron so as to memorize an object that is encoded by the pulse pattern, and for neighbor neurons so as to bring their memory close to the one of the winner neuron. Synaptic efficacy w_{sj}^F of the winner neuron n_s is modulated according to the following rule at the competition time t_c:

$$w_{sj}^F \leftarrow \frac{w_{sj}^F + \Delta w_{sj}^F}{N_s^W} \tag{12}$$

where

$$\Delta w_{sj}^F = \alpha \times (\frac{kp_{sj}(t_c)}{N_s^P(t_c)} - w_{sj}^F) \tag{13}$$

$$N_s^W = \sqrt{\sum_j (w_{sj}^F + \Delta w_{sj}^F)^2}. \tag{14}$$

In these formulas, α is a modulation rate, $kp_{sj}(t_c)$ and $N_s^P(t_c)$ are values given by (5) and (6) respectively, and N_s^W is a normalization factor. On the other hand, synaptic efficacy w_{kj}^F of each neighbor neuron n_k whose firing lag $u_k = lag(n_s, n_k, t_c)$ is modulated according to the following rule:

$$w_{kj}^F \leftarrow \frac{w_{kj}^F + \Delta w_{kj}^F(u_k)}{N_k^W} \tag{15}$$

where

$$\Delta w_{kj}^F(u) = \alpha \times (\frac{kp_{kj}(t_c)}{N_k^P(t_c)} - w_{kj}^F) \times G(u) \tag{16}$$

$$N_k^W = \sqrt{\sum_j (w_{kj}^F + \Delta w_{kj}^F(u_k))^2}. \tag{17}$$

In these formulas, α, $kp_{kj}(t_c)$, $N_k^P(t_c)$ and N_k^W are the same as above, and $G(u)$ is the Gaussian function which gives a decrease rate of modulation due to the firing lag. Let σ be a

specified standard deviation of firing lags. Then $G(u)$ is given by

$$G(u) = \exp(-\frac{u^2}{2\sigma^2}). \quad (18)$$

In Kohonen's self-organizing map [10], learning is performed for the winner neuron and its topological neighbor neurons on the arrangement lattice. Though this achieves topology-preserving mapping, it takes a lot of time to form the mapping because they are only near in topological distance. In the COGNET, neighborhood learning is performed for neurons near to the winner neuron in their patterns of synaptic efficacy. Since this achieves fast learning but does not achieve the topology preservation, the following sorting of neurons in the recognition map layer is performed in descending order of membrane potential of neighbor neurons to achieve the topology preservation (Figure 3). Let n_s be the winner neuron, n_k be its neighbor neuron, and $(n_s.y, n_s.x)$ and $(n_k.y, n_k.x)$ be their positions on a two-dimensional lattice. Firstly, a quadrant in which the position $(n_k.y, n_k.x)$ belongs is obtained in the two dimensional lattice space whose origin is the position $(n_s.y, n_s.x)$. Then, for the Manhattan distance d_k from $(n_s.y, n_s.x)$ to $(n_k.y, n_k.x)$, a set L of lattice points in the quadrant whose Manhattan distance from $(n_s.y, n_s.x)$ is shorter than d_k is obtained as a set of candidate target points to which the neighbor neuron n_k may moves. Secondly, a target point is selected in L according to the following conditions that the point has the minimum Manhattan distance from $(n_s.y, n_s.x)$, and in addition it has the minimum Manhattan distance from $(n_k.y, n_k.x)$, and moreover it does not occupied as a target point of a neighbor neuron whose membrane potential is larger than the one of n_k. If such a target point is found and it is not occupied by any neuron, the neighbor neuron n_k is moved to the point. Otherwise, after the neighbor neuron n_k is moved to the selected target point, each neuron on the way from the target point toward $(n_k.y, n_k.x)$ is permuted with each next neuron one by one until a non-occupied point appears.

Learning with recruitment and topology-preserving sort of neurons enables fast learning of objects, fast self-organization of memory structure of objects, and growth of the recognition map layer caused by increase of objects to be stored.

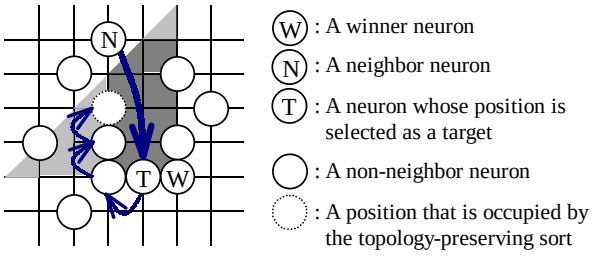

Fig. 3. An illustration of the topology-preserving sort. A halftone area contains candidate target points to one of which a neighbor neuron moves. A dark halftone area takes higher priority as target points than a light halftone area. A thick arrow shows movement of a neighbor neuron. Thin arrows show permutation of neurons for the topology preservation.

E. Reciprocal Learning

Firing of the winner neuron and its neighbor neurons is transmitted to the input window layer. As shown in Figure 2, excitation of neurons in the input window layer caused by this transmission is counterbalanced by the post-competition inhibition according to the formula (1). For the second external input to neurons in the input window layer after the post-competition inhibition, the neurons fire with the same latent period as for the first input according to the formulas (1)(2). Just after firing, synaptic efficacy of neurons in the input window layer is modulated dependent on firing latency and the distribution of recurrent pulse arrival time. Let t_0 be the second external input time. For a neuron n_j in the input window layer, let t_j be the firing time, t^a_{js} be the pulse arrival time from the winner neuron n_s, and t^a_{jk} be the pulse arrival time from a neighbor neuron n_k in the recognition map layer. Then synaptic efficacy w^R_{ji} from a neuron n_i, which is n_s or n_k, to n_j is modulated as follows:

$$w^R_{ji} \leftarrow w^R_{ji} + \Delta w^R_{ji} \quad (19)$$

where

$$\Delta w^R_{ji} = \alpha \times (\frac{\lambda - (t_j - t_0)}{\lambda} - w^R_{ji}) \times G(t^a_{ji} - t^a_{js}). \quad (20)$$

In this rule, α is a modulation rate, λ is the validity term of latency, and $G(u)$ is the Gaussian function which gives a decrease rate of modulation due to the lag of pulse arrival from n_i against pulse arrival from n_s. Let σ be a specified standard deviation of firing lags. Then $G(u)$ is given by

$$G(u) = \begin{cases} \exp(-\frac{u^2}{2\sigma^2}) & \cdots \quad u \geq 0 \\ 0 & \cdots \quad u < 0 \end{cases}. \quad (21)$$

According to this rule, each recurrent synaptic efficacy w^R_{ji} is modulated to be similar with corresponding feedforward synaptic efficacy w^F_{ij}, that is, w^R_{ji} is modulated to be a constant times w^F_{ij}. As a result, symmetrical synaptic efficacy is acquired in reciprocal connection, which enables recall of an object's mental image when an external excitatory input is given to a neuron in the recognition map layer [6].

F. Glance Recognition

It is possible for a human to recognize visual scenes and objects at a glance. This glance recognition is considered to be achieved based on quickly capturing partial information that represents distinctive features of objects. In the COGNET, this ability can be realized by shortening the length of the pre-competition inhibition period at recognition, which is in default set to λ. Then the winner neuron is selected at the competition time brought forward by using only early pulse arrivals. In this case, late pulse transmissions after the competition time are suppressed by the post-competition inhibition in the input window layer. In many cases correct object recognition can be achieved based on only early pulse arrivals because they encode large contrast or opponent color values that capture distinctive features of objects. Moreover, even in case only partial information of an object is carried

by early pulse arrivals under severe time constraint, the whole information of the object can be restored from partial information through reciprocal connection because the object is encoded in both feedforward and recurrent synapses of a winner neuron.

IV. Experimental results

A. A Testbed

To evaluate the self-organized learnability and the glance recognizability of objects in scenes, simulation experiments of a robot equipped with a color camera were conducted using the simulation tool Webots [8]. Figure 4(a) shows an experimental T-maze world. A robot slowly moves right-handed in the maze not to bump against a wall based on the Braitenberg algorithm using six infrared sensors, where two in the front and two each on both sides. A robot captures scene images of 64×64 pixels in height and width and processes them about once a second. Figure 4 shows an example of a saliency map in (b) and attended spots in (c) computed on a scene image.

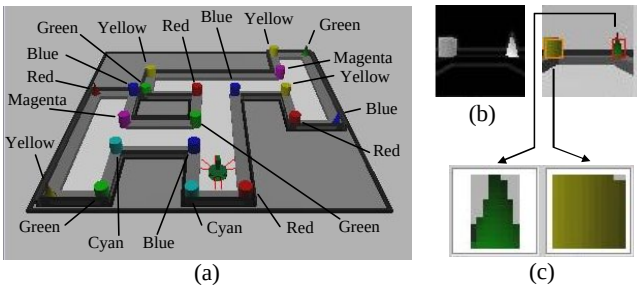

Fig. 4. (a)An experimental world. A total of 20 objects, where three each of red, blue, green and yellow cylinders, two each of magenta and cyan cylinders, and one each of red, blue, green and yellow cones, are arranged as landmarks. (b)A saliency map. (c)Attended spots.

Main parameter values used in experiments are as follows. The l_w of the input window layer is 12. In the recognition map layer, no neuron is arranged initially but neurons are recruited at learning. The threshold of discrimination is 0.75 and the threshold of neighborhood is 0.5. The constant of firing lag at competition is 16. The modulation rate of synaptic efficacy is 0.1. The standard deviation σ of the Gaussian function is 4.

B. Self-organized Learnability

As for self-organized learning, performance of one-shot object learning and fast self-organization in the recognition map layer were evaluated, and also invariant and discriminative object learning performance and scene learning performance were evaluated. In experiments, a robot moves one lap around the maze while learning scenes, then moves another lap while recognizing scenes and recording scene codes without learning. After that, he/she repeatedly searches for a target scene that is picked out every 25 codes from the recorded scene code sequence.

Fig. 5(a) shows a series of concordance between attended object images and winner neurons for them and a series of the number of neurons as the number of attended spots

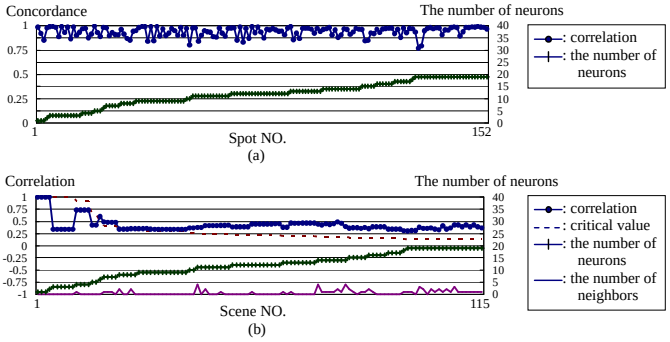

Fig. 5. (a)One-shot learning performance. (b)Fast self-organization performance.

increases at learning. The concordance is given by $\sum_j (w^F_{sj} \times ext_j) / \sqrt{\sum_j (ext_j)^2}$, where w^F_{sj} is a feedforward synaptic efficacy between the winner neuron s and a neuron j in the input window layer and ext_j is an input image element to the neuron j. Since the concordance always takes high values that are larger than the threshold of discrimination, which is due to recruiting new neurons, we can conclude that attended objects were learned quickly.

Fig. 5(b) shows series of the degree of self-organization in the recognition map layer, the number of neighbors and the number of neurons as the number of scenes increases at learning. The degree of self-organization is obtained as the Kendall's rank correlation coefficient between the similarity distance and the Manhattan distance for all neuron pairs on a two-dimensional lattice. The similarity distance of each neuron pair is calculated as the cosine of the angle between synaptic efficacy vectors, which is defined by $\sum_k (w^F_{ik} \times w^F_{jk})$ where w^F_{ik} and w^F_{jk} are feedforward synaptic efficacy for a pair of neurons i and j. A dotted line in the figure shows the critical value of a positive correlation at the significance level (right-sided probability) of 0.5%. We can observe that the degree of self-organization is quickly recovered though it sometimes falls when a neuron is recruited especially in the first stage. It is also observed that significant and steady self-organization is achieved as the number of neurons increases in response to increase of the object variety.

Fig. 6 shows an example of object encoding in the recognition map layer and discriminative and invariant rates of object recognition. The invariant rate indicates whether the same object at different positions and sizes in scenes is encoded to a neuron invariantly and the discriminative rate indicates whether different objects are encoded to different neurons. In addition to simple cylinders and cones, some complexes of cylinders that have certain patterns of arrangement were also extracted as attended objects. Almost every simple object and complex of objects was encoded to a unique different neuron in the recognition map layer. As for target scene search, all of five searches succeeded in this example. By repeating experiments, it was confirmed that invariant object recognition with respect to position and size was achieved with a high probability as a

result of learning. Since target scene search succeeded almost perfectly, it was also confirmed that positions and sizes of objects were encoded suitably enough for scene recognition.

ID	Class of objects encoded	Discrimi-native rate	Invariant rate	ID	Class of objects encoded	Discrimi-native rate	Invariant rate
R0	Green cyl.	100%	100%	R1	Blue cyl.	93%	100%
R2	Red cyl.	100%	100%	R3	Magenta cyl.	100%	100%
R4	Yellow cyl.	95%	100%	R6	Blue cone	100%	100%
R8	Green cone	100%	100%	R13	Cyan cyl.	100%	100%
R14	Red cone	100%	100%	R16	Yellow cone	100%	100%
R7	Magenta cyl. and yellow cyl. behind	100%	100%				
R9	A row of blue, red and green cyl.	100%	100%				
R10	Red cyl. and green cyl. behind	100%	100%				
R11	Green cyl. and blue cyl. behind	63%	100%				
R12	Blue cyl. and green cyl. in front	100%	100%				
R17	Blue cyl. and Magenta cyl. in front	100%	100%				
R18	Blue cyl., O1* and Magenta cyl. in front	100%	100%				

O1*: Yellow cyl. or a row of yellow and green cyl.

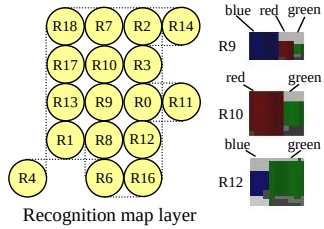

Fig. 6. An example of discriminative and invariant object learning performance. For each neuron, a class of objects is the one that occupies the maximum number of objects encoded to it. The rate of discriminative recognition means the ratio of the number of objects in the class to the number of all objects encoded to the neuron. The rate of invariant recognition means the ratio of the number of objects in the class encoded to the neuron to the number of objects in the class encoded to all neurons. The lower right figures show neuron arrangement in the recognition map layer and complex objects encoded by R9, R10 and R12 respectively.

C. Glance Recognizability

In order to evaluate the glance recognizability using only early pulse arrivals, it is tested whether recognition succeeds for 147 object images that the robot paid attention on the maze when the length of the pre-competition inhibition period is shortened.

Figure 7(a) shows the mean, standard deviation and distribution of the lower limit for success of the glance recognition when the length of the pre-competition inhibition period is gradually shortened. Figure 7(b) shows the mean, standard deviation and distribution of the lower limit for success of the glance recognition against the number of effective pulse transmissions, that is pulse arrivals before the competition time, which is decreased by shortening the length of the pre-competition inhibition period. The mean of the lower limit length of the pre-competition inhibition period for success is 136.1, which is about half of the default length 255 at learning. Also the mean of the lower limit number of effective pulse transmissions for success is 69.0, which is 9.6% of 720 that is the number of pulse transmissions in case the length of the pre-competition inhibition period is 255. Figure 8

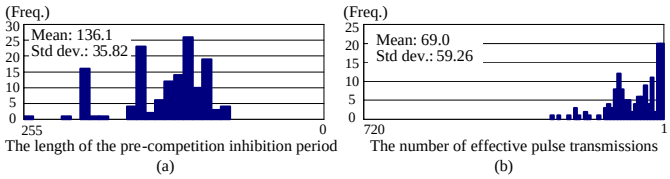

Fig. 7. The lower limit for success of the glance recognition. Data are summed up every 8 interval.

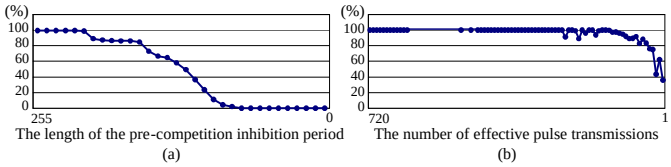

Fig. 8. The success rate of the glance recognition. Data are summed up every 8 interval.

shows the success rate of the glance recognition for the length of the pre-competition inhibition period and the number of effective pulse transmissions. We can observe that a high recognition success rate is achieved using only a small number of early pulse arrivals by shortening the length of the pre-competition inhibition period. These results conclude that the glance recognition is achieved using only early pulse arrivals by bringing the competition time forward.

V. Conclusions

We have discussed saliency-based scene learning and recognition, especially the self-organized learnability and the glance recognizability of objects in scenes by the COGNET. As for learnability, it was confirmed that quick one-shot object learning and fast self-organization of object memory structure were performed, invariant object recognition with respect to position and size was achieved and also positions and sizes of objects were encoded suitably enough for scene recognition. As for recognizability, it was confirmed that glance object recognition was achieved using only early pulse arrivals which encoded distinctive feature of objects.

Acknowledgment

This work was supported in part by Grant-in-Aid for Scientific Research No.13680466 and No.15500075 from Japan Society for the Promotion of Science.

References

[1] B. Ruf and M. Schmitt, “Self-Organizing Maps of Spiking Neurons Using Temporal Coding”, *Computational Neuroscience Trends in Research 1998*, pp.509-514, Plenum Press, 1998.

[2] C. Breazeal and B. Scassellati, “A Context-dependent Attention System for a Social Robot”, Proc. of 16th Int. Joint Conf. on Artificial Intelligence, pp.1146-1151, 1999.

[3] E. Niebur and C. Koch, “Computational Architecture for Attention”, *Attentive Brain*, pp.163-186, The MIT Press, 2000.

[4] L. Itti, C. Koch and E. Niebur, “A Model of Saliency-based Visual Attention for Rapid Scene Analysis”, IEEE Trans. on Pattern Analysis and Machine Intelligence, 20(11), pp.1254-1259, 1998.

[5] M. Atsumi, “Growing Competitive Spiking Neural Network for Saliency-based Scene Recognition”, Proceedings of Workshop on Self-Organizing Maps (WSOM’2003), pp.299-304, 2003.

[6] M. Atsumi, “Saliency-based Scene recognition based on Growing Competitive Neural Network”, SMC 2003 Conference Proceedings, 2003 IEEE Int. Conf. on Systems, Man & Cybernetics, pp.2863-2870, 2003.

[7] M. J. Webster and L. G. Ungerleider, “Neuroanatomy of Visual Attention”, *Attentive Brain*, pp.19-34, The MIT Press, 2000.

[8] O. Michel, *Webots*, Cyberbotics Ltd.

[9] S. Thorpe, A. Delorme and R. V. Rullen, “Spike-based Strategies for Rapid Processing”, Neural Networks, 14, pp.715-725, 2001.

[10] T. Kohonen, *Self-Organizing Maps*, Springer-Verlag, 1995.

[11] W. Gerstner and W. Kistler, *Spiking Neuron Models*, Cambridge University Press, 2002.