Spatial Recognition Model by Extracting Correlated Information between Vision and Motion Information using Neural-Network

Katsunari Shibata
The Research Center for Advanced Science and Technology, University of Tokyo
4-6-1 Komaba, Meguro-ku, Tokyo 153, Japan

Abstract

Spatial recognition ability can be obtained by learning from visual experiments. The author thinks that this ability is to extract correlated information between vision and motion signals. Learning to extract correlated information among multiple inputs can be done without supervisor. The author built a robot with a vision sensor in a computer and it was confirmed that the robot was able to recognize the distance vector of two dimensions from itself to a target object using neural-network without any supervisors.

1. INTRODUCTION

We can realize the spatial recognition ability by learning. The R. Held and A. Hein's experiment[1] shows us that it is important for getting the spatial recognition ability not only to use vision but also to move intentionally. We can think the reason of this fact is that spatial recognition ability is obtained by learning of the correlation between visual sensor information and somato-sensory information or motion information. By developing this ability, we can image how does a sensor information, i.e. the state of environment, change when twe make a motion. And we can get this ability by an efficient unsupervised learning method in which only input information is needed.

The association area in the brain is said to associate motion and sensory information, such as vision, auditory and somato-sensory, with each other. In the view of ontogeny it also grows later than the sensory area or the motion area, and it is affected by visual experiments[2][3]. From this fact, spatial recognition ability is realized in the association area.

We can learn to extract correlated information among multiple inputs without supervisor because the supervisor can be created automatically at the view of extracting correlated information, and the author proposed the method to realize[4]. Then in this paper, the author proposes an idea of spatial recognition model using the method above, and the author report the result of simple simulation in which the robot with vision sensor can learn to recognaize the distance of two dimensions to the target object by itself.

2. Past Psychological and Physiological Experiments

One of the famous experiments about spatial recognition was done by R. Held and A. Hein[1]. In this experiment, two cats are entered in the contrivance, and one can move by its own intention, but the other's motion is passive and same as the previous one. Then the cat moving passively can not form the spatial recognition ability. So it is thought that not only seeing by vision but also moving by itself is necessary to recognize the spatial information.

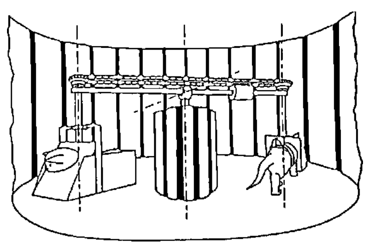

Fig.1 R.Held and A.Hein's experiment[1]

In real living creatures, neurons which are activated by simultaneous inputs of the visual and somato-sensory signals are discovered in the association area. But if the visual information is cut out soon after its birth, the neurons become to activate to only somato-sensory information[5]. This shows us that association of somato-sensory and visual information is done in this area, and tspatial recognition ability can be obtained from experiments after birth.

Furthermore, in Aitken and Bower's experiment a inborn blind baby can become reach his hand to a toy after some tens of presenting by having a transforming equipment from visual information (distance to a target objectand size of it) to the auditory information[6]. This shows us that we can get spatial recognition ability using auditory information, and suggest us that there exists a general sensor-integration learning.

Katayama et al. proposed to obtain the internal representation of the target object by integrating of visual and somato-sensory signals[7]. In this method, as specialized structures is employed and the flow of signal is one way from visual information to somato-sensory information, the abstracting from visual information is realized, but it is not possible to abstract from somato-sensory information. we can't also explain the Aitken and Bower's experiment result that we can form spatial recognition ability using auditory information without visual information from their experiments.

3. METHOD OF EXTRACTING CORRELATED INFORMATION

Correlated information **r** between input information ***x*** and ***y*** is defined as follows.

$$\boldsymbol{r}(t) = \boldsymbol{f}(\boldsymbol{x}(t)) = \boldsymbol{g}(\boldsymbol{y}(t)) \quad (1)$$

$\boldsymbol{f}, \boldsymbol{g}$: vector functions of same dimension as $\boldsymbol{r}$

To extract the correlated such information, the author employs the architecture as shown in Fig.2, and the multiple inputs are entered to each layered network and then exchange the outputs of one network as the supervisors of the other network[3]. Both networks are trained by error-back-propagation method.

To avoid the trivial solution like

$$f(\boldsymbol{x}(t)) = g(\boldsymbol{y}(t)) = const., \quad (2)$$

the Value Range Expanding Operation is held for every cycle, for example every 1000 patterns. First, the degree of independence I which means how independently the output neuron excite from the other output neurons, is defined as follows.

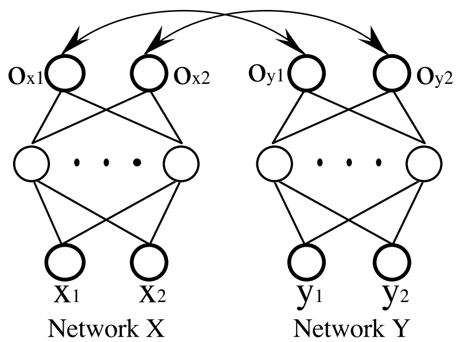

Fig.2 A network structure to get correlated informations among multiple inputs

$$I_j(n) = \frac{\{\sum_i o_{ij}(n) - mid_j\}^2}{\sum_{k \neq j}\{\sum_i o_{ok}(n) - mid_k\}^2 + \alpha} \quad (3)$$

$o_{ik}(n)$: the output of the kth output neuron in the ith network at the nth input pattern

mid_k : the average between the maximum and minimum value of $\sum_i o_{ik}(n)$ in the last cycle

α : small constant value. 0.001 is proper from experiences.

Then the input pattern which gave the maximum degree of independence I when o_{ij} is larger than mid_k in the last cycle re-enter to the network and the network trained by the given supervisor 0.9 and the pattern which the maximum I when o_{ij} is smaller than mid_k also re-enter and trained by the supervisor 0.1. So the output neuron are excited more when the others are less excited, and we can realize to make the correlated information orthogonal to each other. If the number of output neuron is only one, the pattern which gave the maximum or minimum output value in the last cycle is chosen in spite of the maximum of I.

4. SPATIAL RECOGNITION MODEL AND SIMULATION RESULT

4.1 Extracting one-dimensional distance

The author tried to realize the spatial recognition ability similar to the psychological experiments mentioned already using the correlated information extracting network explained before. Concretely we assume a robot with vision as shown in Fig.3 and it looks at the object from the distance which is decided by two moving parts x and ϕ. It has two receptive area and can get sensor output e_1, e_2 which is the ratio of the object in each receptive area. The author prepares 2 network and one of them is for vision (Vision Net) and the other is for motion information (Motion Net), and e_1, e_2 were entered in Vision Net and x and ϕ were entered in Motion Net. Then the author examined if the distance to the object was extracted or not as the correlated information. In these experiments x and ϕ were randomly chosen. Figure 4 shows the output value corresponding to x and ϕ after learning. The curved line like contour in this figure expresses that the distance d is constant. From this figure, we can say that the output does not correspond to x or ϕ, but correspond to distance d. In other words it can be said that the distance d can be extracted.

4.2 Extracting two-dimensional distance

Next, the author examined if the robot can extract two dimensional information efficiently when the freedom of the robot is increased to 2 degrees. In these experiments the robot has three moving parts x, y, ϕ and moves in a two dimensional plane, forward or backward, up or down, by taking x, y, ϕ randomly as shown in Fig.5. It also gets e_1, e_2 as the visual information. And each network has two output neurons and learn for each output to be orthogonal with each other.

Then after learning, Figure 6 could be gotten which shows the two output values corresponding to the two distances to the object, d_1 means the forward or backward distance and d_2 means the up or down distance. Looking corresponding to each distance only, in other words looking at Fig.6 from just side, Figure 7 could be gotten. From this figure it can not be said that d_1 or d_2 can be extracted, because the output value is not decided from the distance. But when the view point is rotated 30 degree, we can get Fig.8. Here d'_1 and d'_2 are the distance rotated 30 degree from d_1 and d_2 respectively. In this figure each output has a one-to-one correspondence to d'_1 or d'_2 respectively. In this case the angle from the direction of d'_1 to that of d'_2 is 90 degree, so it can be said two extracted informations could become orthogonal to each other.

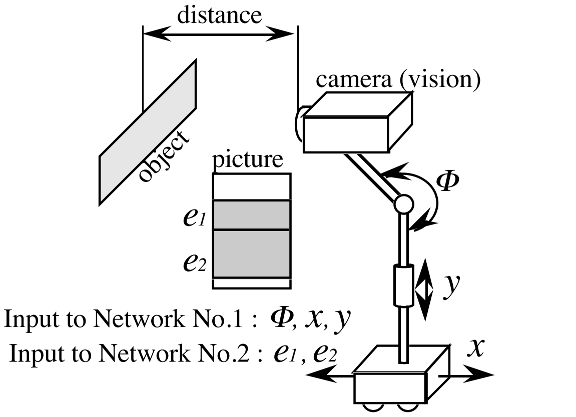

Fig.5 A robot with vision assumed in the simulation which moves in two dimensional area

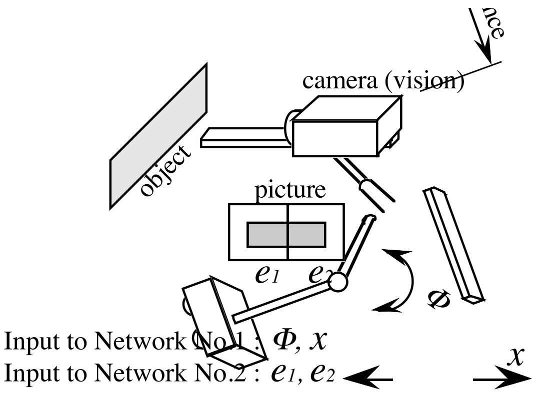

Fig.3 A robot with vision assumed in the simulation

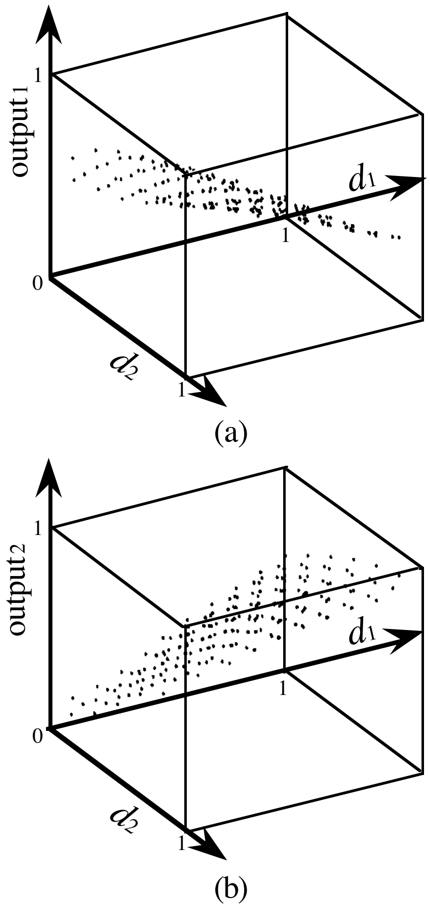

Fig.6 Output of network corresponding to two direction of distance d_1, d_2

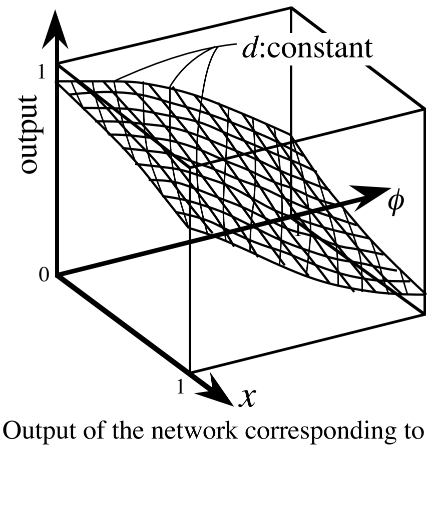

Fig.4 Output of the network corresponding to x and ϕ

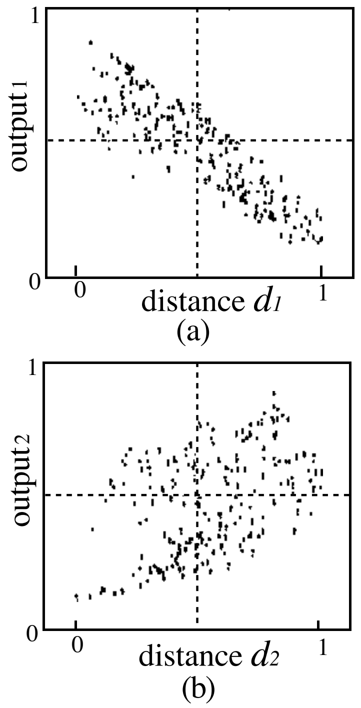

Fig.7 Output of the network corresponding to each distance d_1, d_2

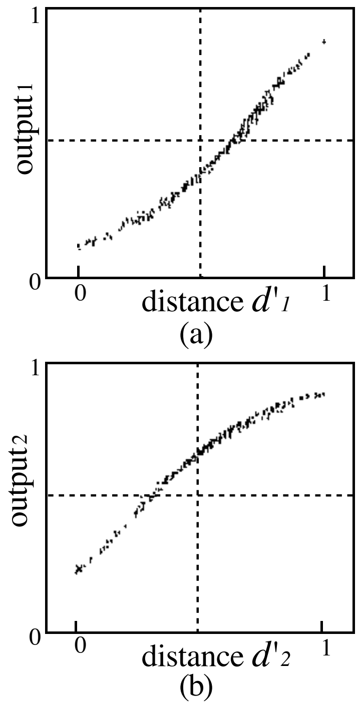

Fig.8 Output of the network corresponding to each distance d'_1, d'_2 (rotate 30 degree from d_1, d)

5. CONCLUSION

The author has proposed simple spatial recognition model using Correlated Information Extracting Network and confirmed to be able to learn the distance vector of two dimensions to an object without any supervisor in a simulation experiments. In the spatial recognition model, the directions of distance to be get by learning are different corresponding to the machine. The author thinks it is a very interesting subject to become the same direction through communications among the machines.

Acknowledgments

This research was held mainly in the Inoue-Inaba Laboratory Dept. of Mechanical Engineering of the Univ. of Tokyo between 1987-1989. The author wants to say special thanks to Prof. Inoue for giving me a chance to do research, and thanks to him, Associate Prof. Inaba and other laboratory members to discuss with me. The author also wants to say thanks to Prof. Okabe for stimulating me to contribute in this conference, and thanks to him and other laboratory members to discuss. Finally the author want to say thanks to Pablo, E. A. for correcting my poor English.

Reference

[1] Held, R. and Hein, A. (1963). Movement-produced stimulation in the development of visually guided components. *J.Comp. Physiol. Psychol.*, **56** : 872-876

[2] Flechsig, P. (1898). Neue Untersuchungen über die Markbildung in den menschlichen Grosshirnlappen. *Neurol. Centrblt.*, **17**: 977-996

[3] Tsumoto, C. (1986). Nou to Hattatsu (Brain and Development), Asakura-Shoten (in Japanese)

[4] Shibata, K. (1993). A Neural-network to get Correlated Informations among Multiple Inputs, submitting IJCNN'93 Nagoya

[5] Hyvärinen, J., Hyvärinen, L. and Linnankoski, I. (1981). Modification of parietal association cortex and functional blindness after binocular deprivation in young monkeys. *Exp. Brain Res.*, **42**, 1-8

[6] Aitken, S. and Bower T. G. R. (1982). Intersensory substitution in the blind. *Jornal of Experimental Child Psychology*, **33** : 309-323

[7] Katayama, M. and Kawato, M. (1990). Neural Network Model Integrating Visual and Somatic Information. *Journal of the Robotics Society of Japan*, **8** : 757-765 (in Japanese)