Available at
www.ComputerScienceWeb.com
POWERED BY SCIENCE DIRECT®

Robotics and Autonomous Systems 45 (2003) 211–221

Robotics and Autonomous Systems

www.elsevier.com/locate/robot

Towards selective attention: generating image features by learning a visuo-motor map

Takashi Minato*, Minoru Asada

Department of Adaptive Machine Systems, Graduate School of Engineering, Osaka University, Yamadaoka 2-1, Suita, Osaka 565-0871, Japan

Received 28 October 2002; received in revised form 11 August 2003; accepted 22 September 2003

Abstract

Robots require a form of visual attention to perform a wide range of tasks effectively. Existing approaches specify in advance the image features and attention control scheme required for a given robot to perform a specific task. However, to cope with different tasks in a dynamic environment, a robot should be able to construct its own attentional mechanisms. This paper presents a method that a robot can use to generating image features by learning a visuo-motor map. The robot constructs the visuo-motor map from training data, and the map constrains both the generation of image features and the estimation of state vectors. The resulting image features and state vectors are highly task-oriented. The learned mechanism is attentional in the sense that it determines what information to select from the image to perform a task. We examine robot experiments using the proposed method for indoor navigation and scoring soccer goals.

Keywords: Image feature; Visual attention; Task-oriented; Visual cortex

1. Introduction

Through billions of years of evolution, biological systems have acquired their organs and strategies to survive in hostile environments. Visual attention can be regarded as a combination of such organs and strategies: vision captures a huge amount of data about the external world, and attentional mechanisms extract information necessary for the system to achieve the mission at hand. This capability is desirable for artificial systems, and it has remained one of the most formidable issues in robotics and AI for many years.

Human beings can readily exploit attentional mechanisms in various kinds of situations, and much research focuses on the early visual processing of human beings [8,16,19,20]. Some research applies Shannon's information theory to the observed image to select the focus of attention in the view [14]. The main emphasis of this work is the analysis of human visual processing and the explanation of our own attentional mechanisms.

Some computer vision researchers focused on the viewpoint selection (i.e., where to look) problem [1,13] in order to disambiguate the descriptions for the observed image that is obtained by matching the image with a model database. The selection criterion is based on the statistics of image data. The actions (gaze control) are intended to obtain a better observation for object recognition, but are not directly related

* Corresponding author.
E-mail addresses: minato@ams.eng.osaka-u.ac.jp (T. Minato), asada@ams.eng.osaka-u.ac.jp (M. Asada).

0921-8890/$ – see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.robot.2003.09.001

to the physical actions needed to accomplish a given task beyond those required for recognition.

Some robot researchers focused on the attention problem of robot vision. Thrun [15] and Vlassis et al. [17] extracted image features correlated with the mobile robot's self-localization information from the observed images based on a probabilistic method. Kröse and Bunschoten [7] decided the robot direction, that is, the camera direction, by minimizing the conditional entropy of the robot position given the observations. These methods are considered to be task-relevant visual attention but are not related to any physical actions.

As mentioned above, there are many methods to construct attentional mechanisms. However, existing methods do not generally take the robot's (or human's) physical actions into consideration.

Meanwhile, in biological systems, feature extracting cells in the visual cortex develop depending on visual and motor experiences. The functions of these cells are not innate, but are adaptively acquired depending on visual experiences in early development after birth [2,5]. Furthermore, self-produced movement with its concurrent visual feedback is necessary for the development of visuo-motor coordination [4].

In light of these neurophysiological studies, the visual organs of a robot should develop depending on its visual and motor experiences. Linsker [9,10] showed that an orientation-selective cell emerged in an artificial multilayered network using modified Hebbian learning. This result shows that the artificial system can learn a visual function similar to the one found in the brain. However, this is a closed system with respect to visual experiences.

In this paper, we focus on extracting image features as an attentional mechanism, and propose a method for image feature (e.g., edges or color regions) generation by visuo-motor map learning depending on the experience gathered by the robot while performing a task. The training data constructs the visuo-motor mapping that constrains image feature generation and state vector estimation for the selection of actions. That is, the state space is constructed so that the correlation between a state and a given instruction can be maximized. The resultant image feature and state vector are task-oriented. The method is applied to indoor navigation and scoring soccer tasks.

There are some existing methods to construct the visual state spaces through task execution (e.g. [6,12]). These methods can construct the task-oriented state vector, but they have not focused on image features. The proposed method constructs the task-oriented visual state space and image features that are useful for selective attention.

The remainder of the paper is organized as follows. First, we describe the basic idea of image feature generation along with the learning formulation. We use the projection matrix from the extracted image feature to the state vector to determine the optimal action. Next, we give experimental results to show the validity of the proposed method. Finally, we conclude with a discussion on the attentional mechanism suggested by the current results.

2. Image feature generation

2.1. The basic idea

In the visual cortex, there are many kinds of cells that extract basic features such as edges from retinal signals. Higher level processes, such as recognition, are performed according to the cell's responses (bottom–up signals) and memory or appetite (top–down signals). That is, various kinds of features are extracted from an input image in a bottom–up process, and the necessary features are selected according to the task in a top–down process. In our model shown in Fig. 1, we decompose the state estimating process into image feature extraction and state estimation. The former is similar to the bottom–up process, while the latter is similar to the top–down process. The robot learns the functions to extract the image features and to estimate the state vector.

The orientation to which the orientation-selective cells in a kitten's visual cortex respond adaptively changes according to its visual experiences during early development [2,5]. Orientation-selectivity is, however, regarded as innate. Inspired by this biological suggestion, we draw the following analogy: in our method, using an image filter to extract features is innate, but the parameters of the image feature are learned from experience.

The robot generates the filtered image I_f from the observed image I_o using the filter F, then extracts its

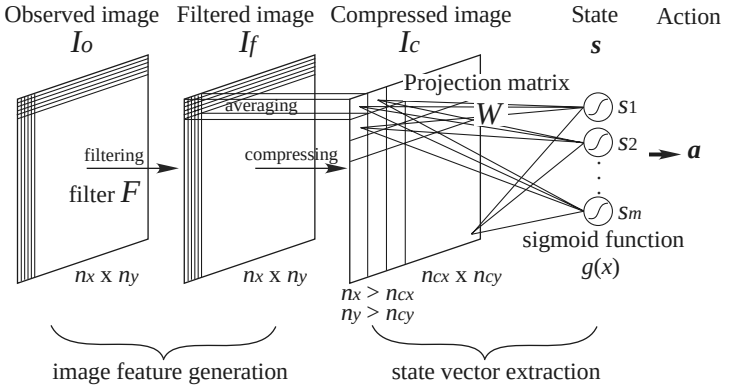


Fig. 1. Image feature generation and action selection model.

state $\boldsymbol{s}$ from I_f and decides the action appropriate to the current state $\boldsymbol{s}$. To avoid the *curse of dimensionality*, we reduce the size of the filtered image I_f to the compressed image I_c from which the state vector is estimated by a projection matrix W. Therefore, the problem is how to learn F and W.

In this two-stage model, we expect the projection matrix to be more task-oriented than the feature filter. During feature extraction, the interactions between raw data are limited to local areas, while the connections between the filtered image and the state spread over the entire space to represent more global interactions. A similar structure can be found in the synaptic connections of our brain, where the receptive fields in the retina and V1 of the visual cortex are relatively narrow and signal transmissions are limited to local areas while later connections are less localized (e.g. [3]). We can also regard W as a kind of attentional mechanism because it connects the compressed image I_c to the state space, that is, it tells which part in the view is more important to estimate at each state and, finally, to decide the optimal action.

As mentioned in the last section, the robot should learn F and W depending on experience gathered while performing tasks. To learn these task constraints, we presented to the system a training set composed of positive instances including the robot's actions and observed images. F and W are learned so that the robot can decide which action should correspond to the observed image. This learning constraint is implemented by minimizing the conditional entropy of the action given the state.

To learn the coefficients of the filter, we prepare a 3×3 spatial filter F_s and a color filter F_c as follows:

- a 3×3 spatial filter $F_\mathrm{s} = (f_{\mathrm{s}ij}) \in \Re^{3\times 3}$:

$$\begin{aligned}\bar{I}_{xy} = {} & f_{\mathrm{s}11} I_{x-1y-1} + f_{\mathrm{s}12} I_{xy-1} + f_{\mathrm{s}13} I_{x+1y-1} \\ & + f_{\mathrm{s}21} I_{x-1y} + f_{\mathrm{s}22} I_{xy} + f_{\mathrm{s}23} I_{x+1y} \\ & + f_{\mathrm{s}31} I_{x-1y+1} + f_{\mathrm{s}32} I_{xy+1} + f_{\mathrm{s}33} I_{x+1y+1},\end{aligned} \tag{1}$$

$$I_{\mathrm{f_s}xy} = g(\bar{I}_{xy}). \tag{2}$$

- a color filter $F_\mathrm{c} = (f_{\mathrm{c}i}) \in \Re^3$:

$$\bar{I}_{xy} = f_{\mathrm{c}1} I_{\mathrm{r}xy} + f_{\mathrm{c}2} I_{\mathrm{g}xy} + f_{\mathrm{c}3} I_{\mathrm{b}xy}, \tag{3}$$

$$I_{\mathrm{f_c}xy} = g(\bar{I}_{xy}), \tag{4}$$

where x and y denote the position of the pixel, I, I_r, I_g and I_b the gray, red, green and blue components of the observed image, respectively, and $g(\cdot)$ a sigmoid function. For example, the following F_s and F_c represent a vertical edge filter and a brightness filter, respectively:

$$F_\mathrm{s} = \begin{pmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{pmatrix}, \tag{5}$$

$$F_\mathrm{c} = \begin{pmatrix} 0.2990 & 0.5870 & 0.1140 \end{pmatrix}^\mathrm{T}. \tag{6}$$

As mentioned above, the robot is given the type of filter F_s or F_c and learns the parameters of the filter coefficients $f_{\mathrm{s}ij}$ or $f_{\mathrm{c}i}$.

2.2. Learning method

First, the robot collects supervised successful instances of the given task, and then learns F and W based on them. In the teaching stage, the robot collects the ith pair

$$T_i = \langle I_{oi}, \boldsymbol{a}_i \rangle, \tag{7}$$

where I_o is the observed image, $\boldsymbol{a} \in \Re^l$ is the supervised robot action executed after the robot observes I_o and i denotes the data number.

The state of the robot $\boldsymbol{s} \in \Re^m$ is estimated by $W \in \Re^{m \times n_{cx} n_{cy}}$. Let $\boldsymbol{i}_c \in \Re^{n_{cx} n_{cy}}$ be the one-dimensional representation of I_c, then

$$s_j = g((W\boldsymbol{i}_c)_j), \tag{8}$$

where $(W\boldsymbol{i}_c)_j$ is a jth element of $W\boldsymbol{i}_c$.

To evaluate F and W, we use the conditional entropy of an action $\boldsymbol{a} \in A$ given a state $\boldsymbol{s} \in S$:

$$H(A|S) = -\int p(\boldsymbol{s}) \int p(\boldsymbol{a}|\boldsymbol{s}) \log p(\boldsymbol{a}|\boldsymbol{s}) \, \mathrm{d}\boldsymbol{a} \, \mathrm{d}\boldsymbol{s}, \tag{9}$$

where $p(\cdot)$ is the probabilistic density, A the action set and S the state set, respectively. To approximate $H(A|S)$, we use the following function R [17]:

$$R = -\frac{1}{N} \sum_i^N \log p(\boldsymbol{a}_i|\boldsymbol{s}_i) = -\frac{1}{N} \sum_i^N \log \frac{p(\boldsymbol{a}_i, \boldsymbol{s}_i)}{p(\boldsymbol{s}_i)}, \tag{10}$$

where N is the size of the training set. To model $p(\boldsymbol{a}, \boldsymbol{s})$ and $p(\boldsymbol{s})$, we use kernel smoothing [18]:

$$p(\boldsymbol{s}) = \frac{1}{N} \sum_q^N K_s(\boldsymbol{s}, \boldsymbol{s}_q), \tag{11}$$

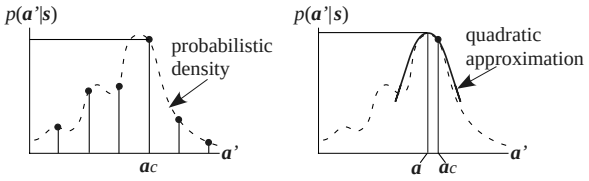
$$p(\boldsymbol{a}, \boldsymbol{s}) = \frac{1}{N} \sum_q^N K_a(\boldsymbol{a}, \boldsymbol{a}_q) K_s(\boldsymbol{s}, \boldsymbol{s}_q), \tag{12}$$

where

$$K_s(\boldsymbol{s}, \boldsymbol{s}_q) = \frac{1}{(2\pi)^{m/2} h_s^m} \exp\left(-\frac{\|\boldsymbol{s} - \boldsymbol{s}_q\|^2}{2h_s^2}\right), \tag{13}$$

$$K_a(\boldsymbol{a}, \boldsymbol{a}_q) = \frac{1}{(2\pi)^{l/2} h_a^l} \exp\left(-\frac{\|\boldsymbol{a} - \boldsymbol{a}_q\|^2}{2h_a^2}\right), \tag{14}$$

h_s and h_a are the width of the kernels. R can be regarded as the Kullback–Leibler distance between $p(\boldsymbol{a}|\boldsymbol{s}_i)$ and a unimodal density sharply peaked at $\boldsymbol{a} = \boldsymbol{a}_i$. By minimizing R, we can bring $p(\boldsymbol{a}|\boldsymbol{s})$ close to the unimodal density, that is, the robot can uniquely decide the action $\boldsymbol{a}$ from the state $\boldsymbol{s}$.



(a) Coarse search (b) Quadratic interpolation method

Fig. 2. A coarse-to-fine strategy.

Using the steepest gradient method, we obtain a pair of F and W that minimize R:

$$F \leftarrow F - \alpha_f \frac{\partial R}{\partial F}, \qquad W \leftarrow W - \alpha_w \frac{\partial R}{\partial W}, \tag{15}$$

where α_f and α_w are the step size parameters.

After learning the robot executes the action $\boldsymbol{a}$ derived from its state $\boldsymbol{s}$ computed from the observed image as follows:

$$\boldsymbol{a} = \arg\max_{\boldsymbol{a}'} p(\boldsymbol{a}'|\boldsymbol{s}). \tag{16}$$

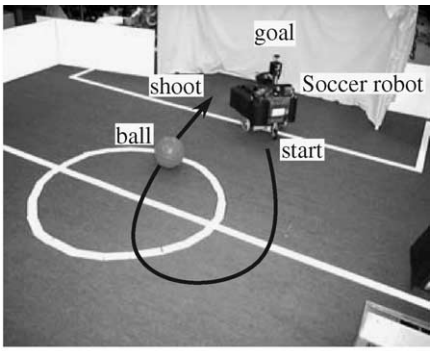
To find the maximum value, we adopt a coarse-to-fine search strategy. At first, we coarsely search approximate maximum density $p(\boldsymbol{a}_c|\boldsymbol{s})$ on the whole action space (Fig. 2(a)), and then compute maximum density $p(\boldsymbol{a}|\boldsymbol{s})$ on the proximity space of $\boldsymbol{a}_c$ by quadratic interpolation method (Fig. 2(b)).

3. Experiments

3.1. Tasks and assumptions

We applied the proposed method to an indoor navigation task with the Nomad mobile robot, Fig. 3(a), and a ball shooting task with a soccer robot, Fig. 3(b). Although the mobile robot shown in Fig. 3(a) is equipped with stereo cameras, we use only the left camera image. The soccer robot shown in Fig. 3(b) is equipped with a single camera directed ahead. Each robot must move along the given path to the destination using the camera image. The size of the observed image I_o is 64×54 pixels and the pixel values of I, I_r, I_g and I_b are normalized to $[0, 1]$. The size of

(a) Task: 1 simple invigation



(a) Task: 2 shooting a ball

Fig. 3. Task.

the compressed image I_c is 8×6 pixels, and each pixel value is the average value of the corresponding region in I_f. The robots can execute translational speed v and steering speed ω independently, so the action vector is represented as

$$\boldsymbol{a} = (v, \omega)^{\mathrm{T}}, \tag{17}$$

where v and ω are normalized to $[-1, 1]$, respectively. We defined the dimension of state as $m = 2$. The sigmoid function g is

$$g(x) = \frac{1}{1 + \exp(-(x - \theta)/c)}, \tag{18}$$

where $\theta = 0.0$ and $c = 0.2$.

3.2. Learning results

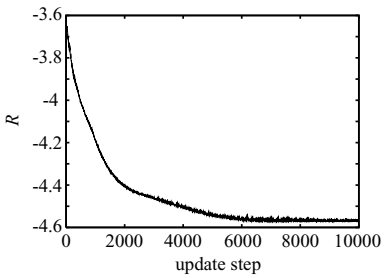
In the teaching stage, we provided positive instances by operating the robot to move along the given path. At each time step, the robot collects an instance composed of a given action and an observed image. In each task, the robot learns a spatial filter F_s and a color filter F_c, separately. We initialized the components of W to small numbers and

$$F_s = \begin{pmatrix} 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 \end{pmatrix} \quad \text{(smoothing filter)}, \tag{19}$$

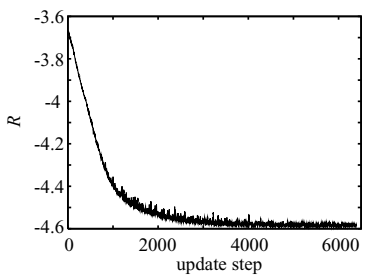
$$F_c = \begin{pmatrix} 0.2990 & 0.5870 & 0.1140 \end{pmatrix}^{\mathrm{T}} \quad \text{(filter to extract brightness)}. \tag{20}$$

3.2.1. Task 1: simple navigation

In the teaching stage, we provided 158 instances. Fig. 4 shows the changes in R for the F_s and F_c models. R decreased almost monotonically. Fig. 5 shows the distributions of the state on the training set for the model with F_s. To visualize the relationship between

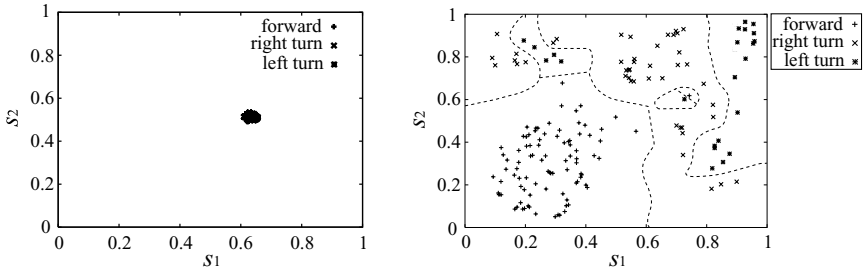


(a) Model with F_s



(b) Model with F_c

Fig. 4. Learning curves of R.



(a) Initial state space (b) Learned state space

Fig. 5. State distributions (task 1, F_s).

the states and actions, we labeled the action indices as follows:

- $v \geq 0.6$ for any ω: forward,
- $v \leq -0.6$ for any ω: backward,
- $-0.6 < v < 0.6$ and $\omega < 0.0$: right turn, and
- $-0.6 < v < 0.6$ and $\omega > 0.0$: left turn.

The dotted lines in Fig. 5(b) are boundaries of each class of actions derived from Eq. (16). As we can see from these figures, the state space can be roughly classified in terms of actions. That is, the state space is constructed so that the correlation between classes of action and classes of state can be maximized. Therefore, the robot can almost uniquely decide the action $\boldsymbol{a}$ from the state $\boldsymbol{s}$.

The generated F_s and F_c are shown below

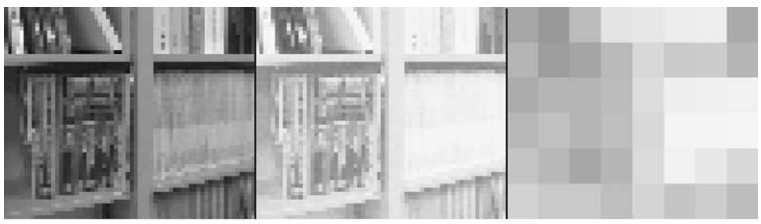
$$F_s = \begin{pmatrix} -0.8915 & -0.5995 & -0.06528 \\ -0.9696 & -0.4790 & 1.357 \\ -0.2482 & 0.1021 & 2.756 \end{pmatrix}, \quad (21)$$

$$F_c = \begin{pmatrix} -0.4233 & 1.464 & -0.1718 \end{pmatrix}^{\mathrm{T}}. \quad (22)$$

Figs. 6 and 7 show examples of the filtered images. As we can see from Fig. 6, F_s can extract vertical and horizontal edges. However, F_c does not show any

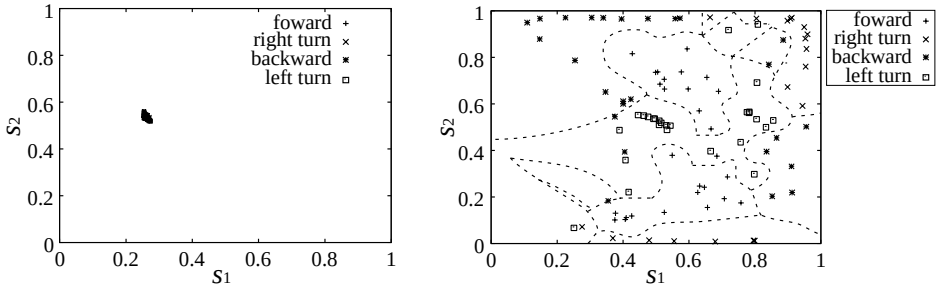
(a) Io (b) If (c) Ic

Fig. 6. An example of the filtered image (F_s).



(a) Io (b) If (c) Ic

Fig. 7. An example of the filtered image (F_c).



(a) Initial state space (b) Learned state space

Fig. 8. State distributions (task 2, F_c).

marked characteristics because there are no salient colored objects in the environment (our laboratory). The edge image extracted by the filter F_s is suitable for recognizing the scene in the clutter environment, because the robot could discriminate between areas that was more textured and areas that was less textured (the floor of the environment).

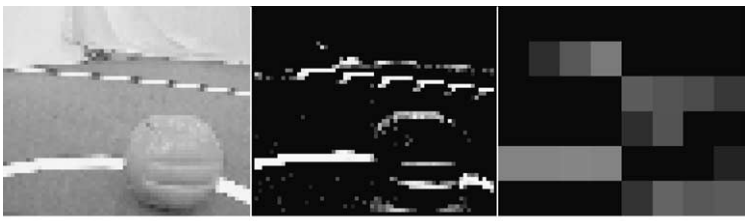
3.2.2. Task 2: shooting a ball

We used the shooting behavior of a soccer robot for RoboCup as a training set. The number of instances was 100. The generated F_s and F_c are shown below:

$$F_s = \begin{pmatrix} -3.384 & -1.953 & -1.686 \\ 0.3491 & -1.350 & 0.5363 \\ 1.656 & -1.208 & 5.223 \end{pmatrix}, \tag{23}$$

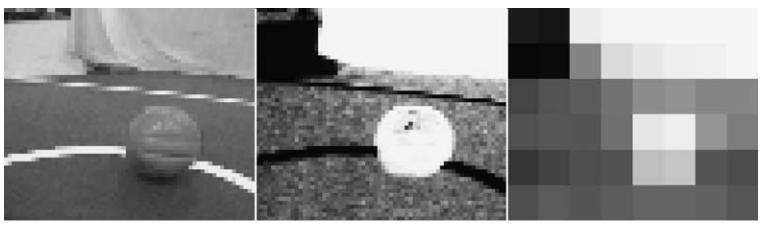
$$F_c = \begin{pmatrix} 1.836 & 1.616 & -4.569 \end{pmatrix}^{\mathrm{T}}. \tag{24}$$

The state distributions in the case of the model with F_c are shown in Fig. 8, and the examples of the filtered image are shown in Figs. 9 and 10.



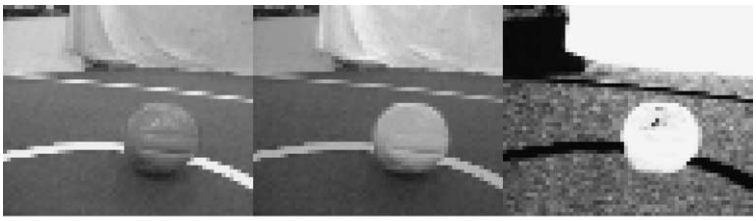
(a) Io (b) If (c) Ic

Fig. 9. An example of the filtered image (F_s).



(a) Io (b) If (c) Ic

Fig. 10. An example of the filtered image (F_c).



(a) input image (b) 1st principal component image of color space (c) our result *If* (*Fc*)

Fig. 11. A first principal component image.

F_s exhibits the characteristic of extracting horizontal edges (see Fig. 9). F_c emphasizes the red ball and yellow goal but inhibits the white line and wall. This is equivalent to a reversed U component in a YUV image. The red ball and white wall are discriminated from other objects in the image extracted by the filter F_c. Therefore, the generated F_c is suitable for a soccer task in the colored soccer field. The feature extracted by F_c is not just a statistical characteristic of the environment. To calculate the statistics of the color component of the input images, we applied PCA. Fig. 11(b) shows the first principal component of color. This is almost entirely the yellow component of the image, because red (ball), yellow (goal), and green (field) are observed with high frequency in this environment. The image extracted by F_c is obviously suitable for deciding an appropriate action; however, the first principal component image is not. This result shows that our method reflects not only the statistical characteristics of the observed images, but the actions needed to accomplish the given task.

3.3. Learned behavior

To verify the validity of the learned model, we applied the model with F_s (task 1) to a navigation task of the Nomad mobile robot (see Fig. 3(a)). Fig. 12 shows a sequence of the acquired behavior. The observed images in this experiment do not exactly coincide with the images in the training set, but the robot accomplished the task. Hence, the acquired image feature and state vector are effective for the task and environment.

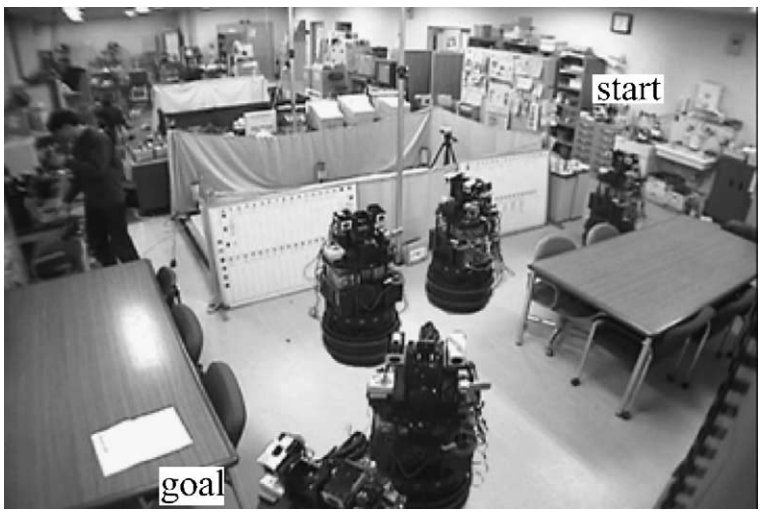
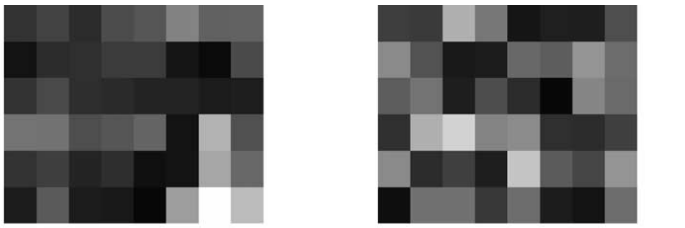


Fig. 12. An acquired behavior.



(a) The first component of s (b) The second component of s

Fig. 13. A projection matrix W (task 1, F_s).

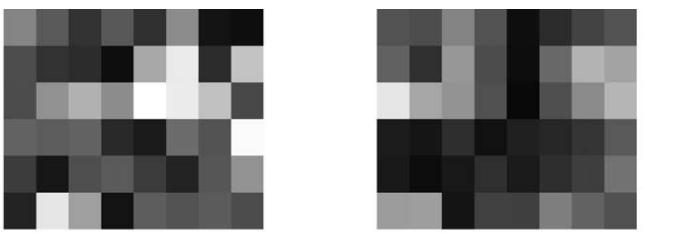
4. Discussion and future work

We proposed a method to generate an image feature and to learn a projection matrix from the filtered image to the state that suggests which part of the view is important, that is, a gaze selection by visuo-motor mapping. The generated image features are appropriate for the task and environment. Also the acquired projection matrices give appropriate gaze selection for the task and environment. To show this, we illustrate the absolute values of W acquired in the model with F_s of task 1 and F_c of task 2 in Figs. 13 and 14. Parts (a) and (b) of each figure show the values of components of W to calculate s_1 and s_2, respectively. Each component of W is a weight of the pixel value of I_c. In these figures, brighter pixels are more closely related to the state vector, that is, the robot gazes at these parts of its view. Therefore, we can regard that a projection matrix provides gaze selection.

The model learned by our method tolerates some environmental changes depending on the learned filter and projection matrix. For example, the image extracted by F_c in Fig. 10 is not affected by changes in environmental illumination. The changes in the partially observed image corresponding to the components of W whose values are nearly zero do not affect the calculation of the state. In general, however, environmental changes require new training data.

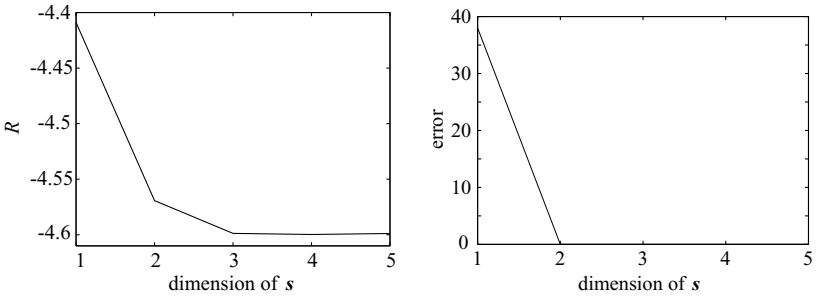
In our method, we use kernel smoothing to calculate the probabilities. The number of kernels used in this method is the same as the number of instances. Therefore, the cost of calculating R is proportional to the square of the number of instances. To avoid the increase in the learning time accompanied by a huge training set, we can randomly sample the instances to calculate the probabilities.

Initially, in this work, we heuristically defined the dimensionality of the state vector to be 2. As a result of experimentation, this dimensionality seems to be appropriate. Fig. 15(a) shows the relationship between the dimension of the state and the value of R for learning the model with F_s in task 1, and (b) shows the relationship between the dimension and the sum of action error. To estimate the action error, we presented the image in the training set to the system and calculated the error norm between the output action and the supervised action corresponding to the image. If we evaluate the action error, we can see that



(a) The first component of s (b) The second component of s

Fig. 14. A projection matrix W (task 2, F_c).



(a) R (b) Estimation error

Fig. 15. An effect of the dimension of the state vector.

a two-dimensional state is enough for deciding an appropriate robot action.

In the sequence in Fig. 12, there are some cases where the robot decides the actions with relatively low probability $p(\boldsymbol{a}|\boldsymbol{s})$, that is, the robot is not so sure about its action decision. Therefore, it seems necessary for the robot to select multiple image features from the image feature set to accomplish more complicated tasks. Now, we are investigating how to integrate the proposed method and the image feature selection method based on the information theoretic criterion [11].

5. Conclusion

In this paper, we have proposed a method in which a robot learns the image feature and state vector that are effective in the given tasks through its experiences. It is considered that the brains of mammals including a human being develop through not only their perception but also the interaction between their bodily movements and the surrounding environment. Our results suggest that we can draw an analogy between generating the image features and developing feature cells in the visual cortex. We hope that our result can be a clue to studying the development of the brain. To this end, our next step is to study the learning of image features in adapting to environmental changes.

References

[1] T. Arbel, F.P. Ferrie, Viewpoint selection by navigation through entropy maps, in: Proceedings of the Seventh International Conference on Computer Vision, 1999, pp. 248–254.

[2] C. Blakemore, G.F. Cooper, Development of the brain depends on the visual environment, Nature 228 (1970) 477–478.

[3] D.J. Fellman, D.C.V. Essen, Distributed hierarchical processing in the primate cerebral cortex, Cerebral Cortex 1 (1991) 1–47.

[4] R. Held, A. Hein, Movement-produced simulation in the development of visually guided behavior, Comparative and Physiological Psychology 56 (1963) 872–876.

[5] H.V.B. Hirsch, D.N. Spinelli, Visual experience modifies distribution of horizontally and vertically oriented receptive fields in cats, Science 168 (1970) 869–871.

[6] H. Ishiguro, M. Kamiharako, T. Ishida, State space construction by attention control, in: Proceedings of the 16th International Joint Conference on Artificial Intelligence, 1999, pp. 1131–1137.

[7] B.J.A. Kröse, R. Bunschoten, Probabilistic localization by appearance models and active vision, in: Proceedings of the 1999 IEEE International Conference on Robotics and Automation, 1999, pp. 2255–2260.

[8] P. Laar, S. Gielen, Task-dependent learning of attention, Neural Networks 10 (6) (1997) 981–992.

[9] R. Linsker, From basic network principles to neural architecture: emergence of spatial-opponent cells, Proceedings of the National Academy of Science of the United States of America 83 (19, 21, 22) (1986) 7508–7512, 8390–8394, 8779–8783.

[10] R. Linsker, Self-organization in a perceptual network, Computer 21 (3) (1988) 105–117.

[11] T. Minato, M. Asada, Selective attention mechanism for mobile robot based on information theory, in: Proceedings of the 18th Annual Conference of the Robot Society of Japan, 2000, pp. 811–812 (in Japanese).

[12] T. Nakamura, Development of self-learning vision-based mobile robots for acquiring soccer robots behaviors, in: Proceedings of the 1998 IEEE International Conference on Robotics and Automation, 1998, pp. 2592–2598.

[13] S.K. Nayar, H. Murase, S.A. Nene, Parametric Appearance Representation in Early Visual Learning, Oxford University Press, Oxford, 1996, Chapter 6.

[14] Y. Takeuchi, N. Ohnishi, N. Sugie, Active vision system based on information theory, Systems and Computers in Japan 29 (11) (1998) 31–39.
[15] S. Thrun, Bayesian landmark learning for mobile robot localization, Machine Learning 31 (1) (1998).
[16] A. Treisman, A. Gelade, A feature integration theory of attention, Cognitive Psychology 12 (1980) 97–136.
[17] N. Vlassis, R. Bunschoten, B. Kröse, Learning task-relevant features from robot data, in: Proceedings of the 2001 IEEE International Conference on Robotics and Automation, 2001, pp. 499–504.
[18] M.P. Wand, M.C. Jones, Kernel Smoothing, Chapman and Hall, London, 1995.
[19] J.M. Wolfe, K.R. Cave, S.L. Franzel, Guided search: an alternative to the feature integration model, Journal of Experimental Psychology: Human Perception and Performance 15 (3) (1989) 419–433.
[20] K. Yokosawa, Multiresolutional model of attention for analysis of visual search performance, in: Proceedings of the Fourth International Conference on Visual Search, 1994.

Takashi Minato received his B.E. and M.E. degrees in mechanical engineering from Osaka University in 1996 and 1998, respectively. He was a researcher of CREST, JST since December 2001. He has been a Research Associate of the Department of Adaptive Machine Systems, Osaka University since September 2002. He is a member of The Robotics Society of Japan.

Minoru Asada received his B.E., M.E., and Ph.D. degrees in control engineering from Osaka University, Osaka, Japan, in 1977, 1979, and 1982, respectively. From 1982 to 1988, he was a Research Associate of Control Engineering, Osaka University, Toyonaka, Osaka, Japan. Since April 1989, he became an Associate Professor of Mechanical Engineering for Computer-controlled Machinery, Osaka University, Suita, Osaka, Japan. Since April 1995, he became a Professor of the same department. Since April 1997, he has been a Professor of the Department of Adaptive Machine Systems at the same university. From August 1986 to October 1987, he was a visiting researcher of Center for Automation Research, University of Maryland, College Park, MD.

He has received the 1992 best paper award of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS92), and the 1996 best paper award of RSJ (Robotics Society of Japan). He was a general chair of IEEE/RSJ 1996 International Conference on Intelligent Robots and Systems (IROS96). His team was the first champion team with the USC team in the middle-size league of the first RoboCup held in conjunction with IJCAI-97. In 2001, he received a Commendation by the Minister of Education. Since 2002, he has been the President of the International RoboCup Federation.