Associative Learning for Cognitive Development of Partner Robot through Interaction with People

Naoyuki KUBOTA
Department of System Design, Tokyo Metropolitan University,
Hino, Tokyo 191-0065, Japan

ABSTRACT

This paper discusses associative learning of a partner robots through interaction with people. Human interaction based on gestures is very important to realize the natural communication. The meaning of gestures can be understood through the actual interaction with a human and the imitation of a human. Therefore, we propose a method for associative learning based on imitation and conversation to realize the natural communication. Steady-state genetic algorithms are applied for detecting human face and objects in image processing. Spiking neural networks are applied for memorizing spatio-temporal patterns of human hand motions, and relationship among perceptual information. Furthermore, we conduct several experiments of the partner robot on the interaction based on imitation and conversation with people. The experimental results show that the proposed method can refine the relationship among the perceptual information, and can reflect the updated relationship to the natural communication with a human.

keywords: Associative Learning, Cognitive Development, Partner Robots, Computational Intelligence

1. INTRODUCTION

Recently, social communication has been discussed from various points of view [1-3]. The capabilities on social communication are required for human-friendly robots such as pet robots, partner robots, and robot-assisted therapy to realize natural communication with people [4-7]. Such a robot requires adaptive perceptual systems to communicate with a human flexibly, and adaptive action systems to learn human behaviors. To realize the learning through interaction with people, we must consider a total architecture of the cognitive development. The cognitive development for robots has been discussed in the fields such as cognitive robotics and embodied cognitive science [8-10]. The study on cognitive development of robots should be performed interdisciplinary from different viewpoints, and the theoretical and technological innovation is expected for the next generation of robotic studies. In the previous research of cognitive robotics, many researchers have proposed the learning methods for the achievement of joint attention, imitative learning, linguistic acquisition from the viewpoints of babies and infants [6,10]. On the other hand, we focus on the refinement of associative memory by using symbolic information used for utterances and patterns based on visual information through interaction with people as cognitive development of robots.

Human interaction based on gestures is very important to realize the natural communication. The meaning of gestures can be understood through the actual interaction with a human and imitation of a human. Therefore, we propose a method for associative learning based on imitation and conversation to realize the natural communication. Basically, imitative learning is composed of model observation and model reproduction. Furthermore, model learning is required to memorize and generalize motion patterns as gestures. In addition, the model clustering is required to distinguish a specific gesture from others, and model selection is also performed for the human interaction. In this way, the imitative learning requires various learning capabilities of model observation, model clustering, model selection, model reproduction, and model learning simultaneously. We proposed a method for imitative learning of partner robots based on visual perception [11,12]. First of all, the robot detects a human based on image processing with a steady-state genetic algorithm (SSGA) [13]. Next, a series of the movements of the human hand are extracted by SSGA used as model observation, and the hand motion pattern is extracted by a spiking neural network (SNN) . Furthermore, SSGA is used for generating a trajectory similar to the human hand motion pattern as model reproduction [14]. In addition to the imitative learning, the robot requires the capability of extracting necessary perceptual information in finite time for the natural communication with a human. Associative memory in the cognitive development is very important for the perception. Therefore we propose a method for the simultaneous associative learning of various types of perceptual information such as colors, shapes, and gestures related with symbolic information used for conversation with a human. Symbolic information used in utterances is very important and helpful for the associative learning, because human language has been improved and refined for long time. The meaning of symbols is neither exact nor precise among people, but the use of linguistic information is very useful and helpful for robots in order to share the meanings of patterns in visual images with people. We apply SNN for associative learning of perceptual information. Furthermore, we conduct several experiments of the partner robot on the interaction with people.

This paper is organized as follows. Section 2 introduces partner robots and computational intelligence technologies. Section 3 explains the method for associative learning of perceptual information. Section 4 shows experimental results of partner robots based on the proposed method.

2. COMPUTATIONAL INTELLIGENCE FOR PARTNER ROBOTS

Partner Robots

We developed two types of partner robots; a mobile PC called MOBiMac [15] and a human-like robot called Hubot [14] in order to realize the social communication with a human (Fig.1). Each robot has two CPUs and many sensors such as CCD camera, microphone, and ultrasonic sensors. Therefore, the robots can conduct image processing, voice recognition, target tracing, collision avoidance, map building, and imitative learning.

In this paper, we focus on the cognitive development of partner robots through interaction with people. As a basic policy of this study, we use flexible and adaptive methods for search and learning. Various types method for the search and learning have been proposed, but we use steady-state genetic algorithms (SSGA) for the search, and spiking neural networks (SNN) for memorizing spatio-temporal information.

Steady-State Genetic Algorithm

A steady-state genetic algorithm (SSGA) is used as one of stochastic search methods, because SSGA can easily obtain feasible solutions through environmental changes with low computational cost . SSGA simulates a continuous model of the generation, which eliminates and generates a few individuals in a generation (iteration) [16-17]. The genotype is represented by $g_{i,j}$ (i=1,2,..., G, j=1,2,..., M) and fitness value is represented by f_i. One iteration is composed of selection, crossover, and mutation. The worst candidate solution is eliminated ("Delete least fitness" selection strategy), and is replaced with the candidate solution generated by the crossover and the mutation.

We use the elitist crossover and adaptive mutation [15]. The elitist crossover randomly selects one individual and generates an individual by combining genetic information from the selected individual and the best individual with the crossover probability. If the crossover probability is satisfied, the elitist crossover is performed. Otherwise, a simple crossover is performed between two randomly selected individuals. Next, the following adaptive mutation is performed to the generated individual,

$$g_{i,j} \leftarrow g_{i,j} + \left(\alpha_j \cdot \frac{f_{\max} - f_i}{f_{\max} - f_{\min}} + \beta_j \right) \cdot N\left(0,1\right) \tag{1}$$

where f_i is the fitness value of the ith individual, f_{max} and f_{min} are the maximum and minimum of fitness values in the population; $N(0,1)$ indicates a normal random variable with a mean of zero and a variance of one; α_j and β_j are the coefficients ($0<\alpha_j<1.0$) and offset ($\beta_j>0$), respectively. In the adaptive mutation, the variance of the normal random number is relatively changed according to the fitness values of the population in case of maximization problems.

Spiking Neural Networks

Various types of artificial neural networks have been proposed to realize clustering, classification, nonlinear mapping, and control [18-20]. Basically, artificial neural networks are classified into pulse-coded neural networks and rate-coded neural networks from the viewpoint of abstraction level [18]. A pulse-coded neural network approximates the dynamics with the ignition phenomenon of a neuron, and the propagation mechanism of the pulse between neurons. Hodgkin-Huxley model is one of the classic neuronal spiking models with four differential equations. An integrate-and-fire model with a first-order linear differential equation is known as a neuron model of a higher abstraction level. A spike response model is slightly more general than the integrate-and-fire model, because the spike response model can choose kernels arbitrarily. On the other hand, rate-coded neural networks neglect the pulse structure, and therefore are considered as neuronal models of the higher level of abstraction. McCulloch-Pitts and Perceptron are well known as famous rate coding models [19,20]. One important feature of pulse-coded neural networks is the capability of temporal coding. In fact, various types of spiking neural networks (SNNs) have been applied for memorizing spatial and temporal context.

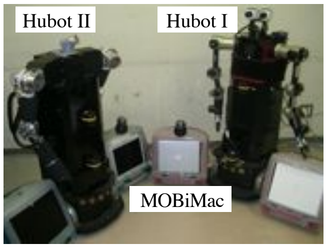


Fig.1. Partner robots; MOBiMac and Hubot

We use a simple spike response model to reduce the computational cost. First of all, the internal state $h_i(t)$ is calculated as follows;

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

Here hyperbolic tangent is used to avoid the bursting of neuronal fires, $h_i^{ext}(t)$ is the input to the ith neuron from the external environment, and $h_i^{syn}(t)$ including the output pulses from other neurons is calculated by,

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

Furthermore, $h_i^{ref}(t)$ indicates the refractoriness factor of the neuron; $w_{j,i}$ is a weight coefficient from the jth to ith neuron; $h_j^{EPSP}(t)$ is the excitatory postsynaptic potential (EPSP) approximately transmitted from the jth neuron at the discrete time t; N is the number of neurons; γ^{syn} is a temporal discount rate. The presynaptic spike output is transmitted to the connected neuron according to EPSP. The EPSP is calculated as follows;

$$h_i^{EPSP}(t) = \sum_{n=0}^{T} \kappa^n p_i(t-n) \tag{4}$$

where κ is the discount rate ($0<\kappa<1.0$); $p_i(t)$ is the output of the ith neuron at the discrete time t; T is the time sequence to be considered. If the neuron is fired, R is subtracted from the refractoriness value in the following,

$$h_i^{ref}(t) = \begin{cases} \gamma^{ref} \cdot h_i^{ref}(t-1) - R & if \quad p_i(t-1) = 1 \\ \gamma^{ref} \cdot h_i^{ref}(t-1) & otherwise \end{cases} \tag{5}$$

where γ^{ref} is a discount rate. When the internal potential of the ith neuron is larger than the predefined threshold, a pulse is outputted as follows;

$$p_i(t) = \begin{cases} 1 & if \quad h_i^{ref}(t) \ge q_i \\ 0 & otherwise \end{cases} \qquad (6)$$

where q_i is a threshold for firing. The weight parameters are trained based on the temporal Hebbian learning rule as follows,

$$w_{j,i} \leftarrow \tanh\left(\gamma^{wgt} \cdot w_{j,i} + \xi^{wgt} \cdot h_j^{EPSP}(t-1) \cdot h_i^{EPSP}(t)\right) \qquad (7)$$

where γ^{wht} is a discount rate and ξ^{wgt} is a learning rate.

Clustering Methods

Cluster analysis is used for grouping or segmenting observations into subsets or clusters based on similarity. Self-organizing map (SOM), *K*-means algorithm, growing neural gases, and Gaussian mixture model are often applied as clustering algorithms [21]. SOM can be used as incremental learning, while *K*-means algorithm and Gaussian mixture model use observed all data in the learning phase (batch learning). In this paper, we apply SOM for clustering spatio-temporal patterns of pulse outputs from the SNN. Furthermore, the neighboring structure of units can be used in the further discussion for the similarity of clusters.

SOM is often applied for extracting a relationship among observed data, since SOM can learn the hidden topological structure from the data. The inputs to SOM is given as the weighted sum of pulse outputs from neurons,

$$\mathbf{v} = \left(v_1, v_2, ..., v_N\right) \qquad (8)$$

where *vi* is the state of the *i*th neuron. In order to consider the temporal pattern, we use $h_i^{EPSP}(t)$ as v_i, although the EPSP is used when the presynaptic spike output is transmitted. When the *i*th reference vector of SOM is represented by **r***i*, the Euclidian distance between an input vector and the *i*th reference vector is defined as

$$d_i = \left\|\mathbf{v} - \mathbf{r}_i\right\| \qquad (9)$$

Where $\mathbf{r}_i=(r_{1,i},\ r_{2,i},\ \ldots\ ,\ r_{N,i})$ and the number of reference vectors (output units) is *M*. Next, the *k*th output unit minimizing the distance d_i is selected by

$$k = \arg\min_i \left\{\left\|\mathbf{v} - \mathbf{r}_i\right\|\right\} \qquad (10)$$

Furthermore, the reference vector of the *i*th output unit is trained by

$$\mathbf{r}_i \leftarrow \mathbf{r}_i + \xi^{SOM} \cdot \zeta_{k,i}^{SOM} \cdot \left(\mathbf{v} - \mathbf{r}_i\right) \qquad (11)$$

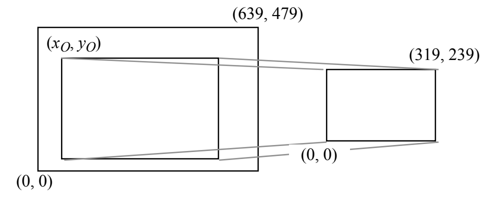
where ξ^{SOM} is a learning rate ($0<\xi^{SOM}<1.0$); $\zeta_{k,i}^{SOM}$ is a neighborhood function ($0<\zeta_{k,i}^{SOM}<1.0$).

3. ASSOCIATIVE LEARNING OF PERCEPTUAL INFORMATION

Visual perception of the robots is realized by image processing based on images taken from the CCD camera. In this paper, the robots use perceptual modules for image processing such as differential extraction, human detection, object detection, and human hand motion recognition. Furthermore, we propose the associative learning of perceptual information in this section.

Human Detection and Tracking

Various types of pattern matching methods such as template matching, cellular neural network, neocognitron, and dynamic programming (DP) matching, have been applied for the human detection in image processing. In general, pattern matching is composed of two steps of target detection and target recognition. The aim of target detection is to extract a target candidate from an image, and the aim of the target recognition is to identify the target from classification candidates.



(a) Original Image (b) Attention range

Fig. 2 Human face detection for joint attention

Since the image processing takes much computational time and cost, the full size of image processing to every image is not practical. Therefore, we use the reduced size of image to detect a moving object for the fast human candidate detection. First, an image of RGB color space is taken by a CCD camera equipped with the partner robot. Next, the robot calculates the center of gravity (COG) of the pixels different from the previous image as the differential extraction. The size of image used in the differential extraction is updated according to the previous result of human detection. Here the area generated by the differential extraction is called an attention range. If the robot does not move, the COG of the difference represents the location of the moving object. Therefore, the main search area for the human detection can be formed according to the COG in the attention range for the fast human detection. In this paper, the original size of an image is 640×480, and the size of this image is reduced into 320×240 as an attention range according to the reduction level ($1.0\le RL \le 2.0$) and the origin (x_O, y_O) of the attention range (Fig.2). If the reduction level is 1, the same resolution of the image is cut off from the original image. Otherwise, each pixel on the attention range is interpolated according to the four surrounding pixels based on the reduction level.

The robot must recognize a human face from complex background speedily. Therefore, we use SSGA for human detection as one of search methods. The human face candidate positions based on human skin and hair colors are extracted by SSGA with template matching. Figure 3 shows a candidate solution of a template used for detecting a human face. A template is composed of numerical parameters of $g_{i,1}^H$, $g_{i,2}^H$, $g_{i,3}^H$, and $g_{i,4}^H$. The number of individuals is G^H. The fitness value of the *i*th individual is calculated by the following equation,

$$f_i^H = C_{Skin}^H + C_{Hair}^H + \eta_1^H \cdot C_{Skin}^H \cdot C_{Hair}^H - \eta_2^H \cdot C_{Other}^H \qquad (12)$$

where C^H_{Skin}, C^H_{Hair} and C^H_{Other} indicate the numbers of pixels of the colors corresponding to human skin, human hair, and other colors, respectively; η_1^H and η_2^H are the coefficients (η_1^H, $\eta_2^H>0$). Therefore, this problem results in the maximization problem. The iteration of SSGA is repeated until the termination condition is satisfied. Here we used the fixed number of generations for the termination condition. SSGA for the human detection is called SSGA-

H. Since SSGA extracts the area of skin colors and hair colors in the human detection, various objects except humans might be detected. Therefore, the human tracking is performed according to the time series position of the *i*th human candidate ($g_{i,1}^H$, $g_{i,2}^H$) obtained by SSGA-H. The position of the *j*th human candidate in the human tracking ($X_{k,1}$, $X_{k,2}$) is updated by the nearest human candidate position within the tracking range. In addition, the width and height of the human candidate for the human tracking ($X_{k,3}$, $X_{k,4}$) are updated by the size of the detected human ($g_{i,3}^H$, $g_{i,4}^H$). The update is performed as follows (j=1,2,3,4);

$$X_{k,j} = (1-\lambda)X_{k,j} + \lambda \cdot x_{i,j} \tag{13}$$

Furthermore, the time counter for the reliability of human tracking is used. If the human candidate position in the human tracking is performed, the time counter is incremented. Otherwise, the time counter is decremented. If the time counter is larger than the threshold (*HT*), the human count is started. Sometimes, several human candidates are close each other, because several human candidates in a single human can be generated by the human detection. Therefore, the removal processing is performed when human candidates are coexisting within the tracking range.

The facial direction can be approximately extracted by using the relative positions of human hair and human face. We apply spiking neurons to extract the direction of the detected human face. We use the relative position of COG of areas corresponding to human hair and human face. The relative positions of the COG against the central position of the detected face region are used as inputs to the spiking neurons for extracting the human facial direction.

Object Recognition

We explain a method for object recognition. We focus on color-based object and shape recognition with SSGA based on template matching. Here the SSGA for object recognition is called SSGA-O. The shape of a candidate template is generated by the SSGA-O. We used an octagonal template of the angle fixed at 45º. Figure 4 shows a candidate template used for detecting a target where the *j*th point $g_{i,j}^O$ of the *i*th template is represented by ($g_{i,,1}^O$,+$g_{i,,j}^O$cos($g_{i,j+m}^O$), $g_{i,,2}^O$+$g_{i,,j}^O$sin($g_{i,,j+m}^O$)), i=1, 2, ... , G^O, j=3, 4, ... , 2×m+2; O_i (=($g_{i,1}^O$, $g_{i,2}^O$)) is the center of a candidate template on the image; *n* and *m* are the number of candidate templates and the searching points used in a template, respectively. Therefore, a candidate template is composed of numerical parameters of ($g_{i,1}^O$, $g_{i,2}^O$,... , $g_{i,2m+2}^O$). Fitness value is calculated as follows.

$$f_i^O = C^{\mathrm{O}}_{\mathrm{Target}} - \eta^O \cdot C^{\mathrm{O}}_{\mathrm{Other}} \tag{14}$$

where η^O is a coefficient for penalty (η^O>0); C_{Target}^O and C_{Other}^O indicate the numbers of pixels of the colors corresponding to a target and other colors included in the template, respectively. The target color is selected according to the pixel color occupied mostly in the template candidate. Therefore, the largest area of a single color is extracted on the reduced color space of the image.

Furthermore, we apply k-means algorithm for the clustering of candidate templates in order to find several objects simultaneously. The inputs to *K*-means algorithm are the central position of templates candidates; v_j (=($g_{i,1}^O$, $g_{i,2}^O$), j=1,2, ..., K). After selecting the nearest reference vector to each input, the *i*th reference vector is updated by the average of the inputs belonging to the *i*th cluster. If the update is not performed at the clustering process, this updating process is finished. The crossover and selection are performed with the template candidates of each cluster. Therefore, SSGA-O tries to find different objects within each cluster according to the spatial distribution of objects in the image. Spiking neurons are applied for shape recognition of the detected color objects. We use four sensor neurons for extracting circle, triangle, rectangle, and complicated shape. The number of acute angles in a template candidate is used to sensor neurons. If the sensor neuron corresponding to each shape is fired, the spike output is transmitted to utterance system.

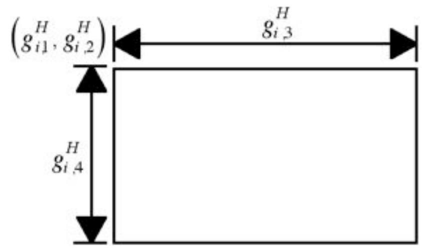


Fig. 3. A template used for human detection in SSGA-H

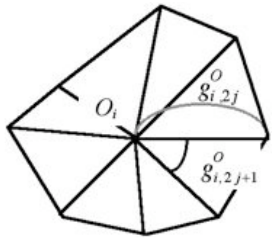


Fig. 4. A template used for object detection in SSGA-O

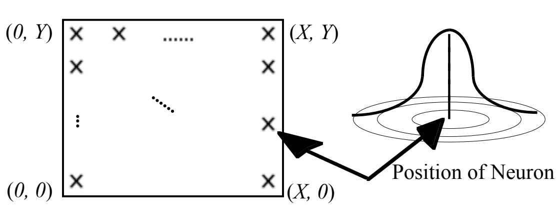


Fig. 5. Spiking neurons for gesture recognition

Human Hand Motion Extraction and Learning

The robot extracts human hand motion from the series of images by using SSGA-O where the maximal number of images is T_G. The sequence of the hand positions is represented by $\mathbf{G}(t)$=($G_x(t)$, $G_y(t)$) where t=1, 2, ... , T_G. Here spiking neurons are arranged on a planar grid (Fig.5) and N=25. By using the value of a human hand position, the input to the *i*th neuron is calculated by the Gaussian membership function as follows;

$$h_i^{ext}(t) = \exp\left(-\frac{\left\| \mathbf{c}_i - \mathbf{G}(t) \right\|^2}{2\sigma^2} \right) \tag{15}$$

where $\mathbf{c}_i$=($c_{x,i}$, $c_{y,i}$) is the position of the *i*th spiking neuron on the image; σ is a standard deviation. The sequence of pulse outputs $p_i(t)$ is obtained by using the human hand positions $\mathbf{G}(t)$. Because the adjacent neurons along the trajectory of the human hand position are easily fired as a result of the temporal Hebbian learning, the SNN can memorize the temporally firing patterns of various gestures.

Next, we explain a method for clustering human hand motions. The inputs to SOM is given as the weighted sum of pulse outputs from neurons,

$$\mathbf{v} = \left(v_1, v_2, \ldots, v_N\right) \tag{16}$$

where v_i is the state of the ith neuron. In order to consider the temporal pattern, we use $h_i^{EPSP}(t)$ as v_i, although the EPSP is used when the presynaptic spike output is transmitted. Accordingly, the selected output unit is the nearest pattern among the previously learned human hand motion patterns.

Utterance System and Associative Learning

This subsection explains a method for associative learning in the perceptual system for cognitive development. Symbolic information is very useful and helpful to learn the relationship among patterns. In this paper, we focus on the refinement of the association of other information from the perceptual information. We use spiking neural networks.

Various types of utterance systems and language processing systems have been proposed [3,4]. In this paper, we propose an utterance system composed of two stages; (1) Utterance group selection and (2) word or sentence selection. Basically, an utterance group is composed of different presentational words or sentences with the same meaning, e.g., "hello"={hi, hello, ya}. The utterance group is selected according to the context of the conversation and perceptual information, and one word or sentence is stochastically selected from the group according to the state of feelings. The total selection strength of the ith utterance group s_i^U is calculated as follow,

$$s_i^U = s_i^S + v_j^R \cdot s_{j,i}^R + v_h^H \cdot s_{h,i}^H + \sum_{k=1}^{K} (w_{k,i} \cdot p_k) \tag{17}$$

where s_i^S is the suppression factor; $s^R_{j,i}$ is the selection strength of the ith utterance after the robot speaks the jth utterance; v_i^R and v_i^H are the validity parameters; $s^H_{h,i}$ is the selection strength of the ith utterance after the human speaks the hth utterance, that is recognized by the robot; $w_{k,i}$ is the connection strength between the kth perceptual information and the ith utterance; p_k is the input value to the utterance system based on the perceptual result; K is the number of perceptual information. The selection probability (s_i^P) of the ith utterance group is calculated by using a Boltzmann selection scheme as follows,

$$s_i^P = \frac{\exp(s_i^U / \tau^U)}{\sum_{j=1}^{J} \exp(s_j^U / \tau^U)} \tag{18}$$

where τ^U is a positive parameter called the temperature. When the temperature is high, the robot randomly selects an utterance group. As the temperature decreases, the robot deterministically selects the utterance group with the high selection strength. Furthermore, if the utterance group is selected, s_i^S is set at a negative large value in order to avoid the continuous selection of the same utterance group. Otherwise, $s_i^S = \tau^S \cdot s_i^S$ where τ^S is the discount rate.

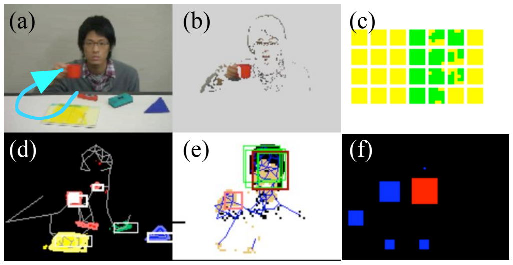


Fig. 6 Experimental results of human tracking, object recognition, and gesture recognition

The learning of the symbolic information is performed by the associative learning between the neurons corresponding to the utterance word and the neuron corresponding to patterns of gestures, colors, and shapes.

4. EXPERIMENTAL RESULTS

This section shows experimental results of human detection, object recognition, gesture recognition, and associative learning of a partner robot. The number of utterance words 50. The population size of SSGA-H and SSGA-O is 100. The number of the spiking neurons in the gesture recognition is 25. The number of gestures in SOM is 50. The gesture recognition for object handling starts if the position of hand and object is near and the velocity is also similar. In this experiments, a person showed the cup, the book and triangle block to the robot repeatedly for 10 minutes.

Figure 6 shows snapshot of image processings; (a) original image and the trajectory of the human hand motion, (b) differential extraction, (c) the reference vectors of SOM corresponding to gestures, (d) object recognition results by SSGA-O (e) human detection results by SSGA-H (f) EPSP of the spiking neurons. The person was reaching the red cup, and held it in this example. The method for human detection and tracking extracted the his face and his hand. In Fig.6 (e), a green box indicates the candidates of human facial position by SSGA-H; a red box indicates the human face position by the human tracking; a pink box indicates the human hand position. SSGA-O detected a red cup, yellow book, read marker, a green eraser, and triangle block in Fig.6 (d). Figure 6 (f) shows the degree of EPSP from a spiking neuron, and this indicates the spatio-temporal pattern captures his hand motion. Figure 6 (c) shows the reference vectors of SOM learned through the interaction with him.

Figure 7 shows the experimental results of the weight connection in the associative learning. Figure 7 (1) shows the learning state after one minute and (2) shows the learning state after ten minutes. In the beginning, the gesture is not learning, and only the weight connection among utterance words, color information, and shape information is updated. In this stage, the robot makes an utterance according to the scenario designed beforehand. Afterward, the gestures are gradually learned and the weight connection among all information is updated. As a result, the robot tried to make utterances according to the learned association. In this way, the robot can perform the natural communication adapted to the specific human interaction.

5. CONCLUSIONS

This paper discussed the capability of associative learning for partner robots through interaction with a human. We proposed the methods of computational intelligence for human detection and tracking, object recognition, and associative learning. The experimental results show the effectiveness of the methods for human detection and gesture recognition, and the robot can learn the relationship among the symbolic information used for utterances and the visual information. As a result, the associative capability of the robot is refined through the actual interaction with a human.

As future works, we will conduct experiments on the actions of partner robot, Hubot through interaction with people, and imitative learning on the object handling.

6. REFERENCES

[1] M.W.Eysenck, Psychology: **An Integrated Approach, Harlow**, Essex: Longman, 1998.

[2] D.Sperber and D.Wilson, **Relevance - Communication and Cognition**, Blackwell Publishing Ltd., 1995.

[3] N.Chomsky, **Reflections on language**, New York: Pantheon, 1975.

[4] J.Cassell, "Embodied Conversational Agents: Representation and Intelligence in User Interface", **AI Magazine**, Vol.22, No.3, pp.67-83, 2001.

[5] Y. Nakauchi, R. Simmons, "A Social Robot that Stands in Line", **Journal of Autonomous Robots**, Vol.12, No.3, pp.313-324, 2002.

[6] M. Imai, T. Ono, H. Ishiguro, "Physical Relation and Expression: Joint Attention for Human-Robot Interaction", **IEEE Transactions on Industrial Electronics**, Vol.50, No.4, pp.636-643, 2003.

[7] H.Ishiguro, M.Shiomi, T.Kanda, D.Eaton, N. Hagita, "Field Experiment in a Science Museum with Communication Robots and a Ubiquitous Sensor Network", **Proc. of Workshop on Network Robot System at ICRA2005**, 2005.

[8] R.Pfeifer, C.Scheier, **Understanding Intelligence**, The MIT Press, 1999.

[9] K.Morikawa, S.Agarwal, C.Elkan, and G.Cottrell, "A taxonomy of computational and social learning", **Proc. Workshop on Developmental Embodied Cognition**, 2001.

[10] R.P.N. Rao, A. N. Meltzoff, "Imitation leaning in infants and robots: towards probabilistic computational models", **Proc. of Artificial Intelligence and Simulation of Behaviors**, pp. 4-14, 2003.

[11] N.Kubota, K.Tomoda, "Behavior Coordination of A Partner Robot based on Imitation", **Proc. of 2nd International Conference on Autonomous Robots and Agents**, pp.164-169, 2004.

[12] N.Kubota, "Computational Intelligence for Structured Learning of A Partner Robot Based on Imitation", **Information Science**, No.171, pp.403-429 , 2005.

[13] N.Kubota, "Computational Intelligence for Human Detection of A Partner Robot", **Proc. (CD-ROM) of World Automation Congress 2004**, 2004.

[14] N.Kubota, Y.Tomioka, M.Abe, "Temporal Coding in Spiking Neural Network for Gestures Recognition of a Partner Robot", **Proc. Joint 3rd Int. Conf. on Soft Computing and Intelligent Systems and 7th International Symposium on advanced Intelligent Systems**, pp. 737-742, 2006.

[15] N. Kubota, K. Nishida, "Development of Internal Models for Communication of A Partner Robot Based on Computational Intelligence," **Proc. of 6th Int. Symp. on Advanced Intelligent Systems**, pp. 577-582, 2005.

[16] Syswerda, G. "A study of reproduction in generational and steady-state genetic algorithms", **Foundations of Genetic Algorithms**, Morgan Kaufmann Publishers Inc., San Mateo, 1991.

[17] D.B.Fogel, **Evolutionary Computation**, IEEE Press, 1995.

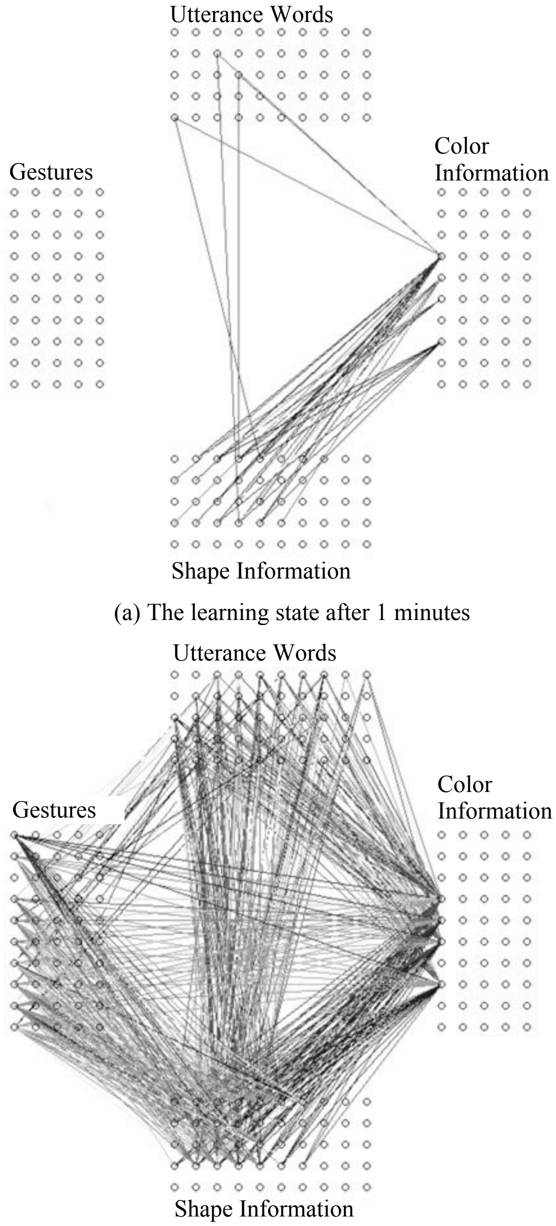
[18] W.Gerstner, **Pulsed Neural Networks**, W.Maass, and C.M.Bishop, (Ed.), pp. 3-53, MIT Press, Cambridge, Massachusetts, US, 1999.

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

[20] J. A.Anderson, E.Rosenfeld, **Neurocomputing**, Cambridge, Massachusetts: The MIT Press, 1988.

[21] T.Kohonen, **Self-Organizing Maps**, 3rd Edition, Springer-Verlag, Berlin, Heidelberg, 2001.

[22] T. Hastie, R.Tibshirani, J.Friedman, **The Elements of Statistical Learning: Data Mining, Inference, and Prediction**, Springer-Verlag, New York, 2001.



(a) The learning state after 1 minutes

(b) The learning state after 10 minutes

Fig. 7. Associative learning through interaction with a human