

Human Detection and Gesture Recognition Based on Ambient Intelligence

Naoyuki Kubota
Tokyo Metropolitan University
Japan

1. Introduction

Recently, various types of human-friendly robots such as pet robots, amusement robots, and partner robots, have been developed for the next generation society. The human-friendly robots should perform human recognition, voice recognition, and gesture recognition in order to realize natural communication with a human. Furthermore, the robots should coexist in the human environments based on learning and adaptation. However, it is very difficult for the robot to successfully realize these capabilities and functions under real world conditions. Two different approaches have been discussed to improve these capabilities and functions of the robots. One approach is to use conventional intelligent technologies based on various sensors equipped on a robot. As a result, the size of a robot becomes large. The other approach is to use ambient intelligence technologies of environmental systems based on the structured information available to a robot. The robot directly receives the environmental information through a local area network without measurement by the robot itself. In the following, we explain the current status of researches on sensor networks and interactive behavior acquisition from the viewpoint of ambient intelligence.

1.1 Ambient Intelligence

For the development of sensor network and ubiquitous computing, we should discuss the intelligence technologies in the whole system of robots and environmental systems. Here intelligence technologies related with measurement, transmission, modeling, and control of environmental information is called ambient intelligence. The concept of ambient intelligence was discussed by Hagrais et.al. (Doctor et al., 2005). Their main aim is to improve the qualities of life based on computational artifacts, but we focus on the technologies for the co-existence of humans and robots in the same space. From the sensing point of view, a robot is considered as a movable sensing device, and an environmental system is considered as a fixed sensing device. If the environmental information is available from the environmental system, the flexible and dynamic perception can be realized by integrating environmental information.

The research on wireless sensor networks combines three components of sensing, processing, and communicating into a single tiny device (Khemapech et al., 2005). The main roles of sensor networks are (1) environmental data gathering, (2) security monitoring, (3)

and object tracking. In the environmental data gathering, the data measured at each node are periodically transmitted to a database server. While the synchronization of the measurement is very important to improve the accuracy of data in the environmental data gathering, an immediate and reliable emergency alert system is very important in the security monitoring. Furthermore, a security monitoring system does not need to transmit data to an emergency alert system, but the information on features or situations should be transmitted as fast as possible. Therefore, the basic network architecture is different between data gathering and security monitoring. On the other hand, the object tracking is performed through a region monitored by a sensor network. Basically, objects can be tracked by tagging them with a small sensor node. Radio frequency identification (RFID) tags are often used for the tracking system owing to low cost and small size.

Sensor networks and ubiquitous computing have been incorporated into robotics. These researches are called network robotics and ubiquitous robotics, respectively (Kim et al., 2004). The ubiquitous computing integrates computation into the environment (Sato, 2006). The ubiquitous computing is conceptually different from sensor networks, but both aim at the same research direction. If the robot can receive the environmental data through the network without the measurement by sensors, the size of the robot can be easily reduced and the received environmental data are more precise because the sensors equipped in the environment is designed suitable to the environmental conditions. On the other hand, network robots are divided into three types; visible robots, unconscious robots, and virtual robots (Kemotsu, 2005). The role of visible robots is to act on users with their physical body. The role of unconscious robots is mainly to gather environmental data, and this kind of unconscious robot is invisible to users. A virtual robot indicates a software or agent in a cyber world. A visible robot can easily perceive objects by receiving object information from RFID tags, and this technology has been applied for the robot navigation and the localization of the self-position (Kulyukin et al., 2004). Hagaras et al. developed iDorm as a multi-function space (Doctor et al., 2005). Furthermore, Hashimoto et al. proposed Intelligent Space (iSpace) in order to achieve human-centered services, and developed distributed intelligent network devices composed of color CCD camera including processing and networking units (Morioka & Hashimoto, 2004). A robot can be used not only as a human-friendly life-support system (Mori et al., 2005), but also as an interface connecting the physical world with the cyber world.

1.2 Interactive Behavioral Acquisition

In general, the behavioral acquisition used in robotics can be classified into supervised learning and self-learning (Figure 1). The self-learning is defined as unsupervised learning performed by trial-and-error without exact target teaching signals for motion reproduction. Supervised learning in behavior acquisition is divided into social learning and error-based learning. For example, least mean square algorithms are applied for behavioral learning when exact target teaching signals are given to a robot. On the other hand, the observed data, instead of exact target teaching signals, are used in social learning. The social learning is performed between two or more agents. Basically, social learning is divided into imitative learning, instructive learning, and collaborative learning (Morikawa et al., 2001).

Imitation (Billard, 2002) is a powerful tool for gestural interaction among children and for teaching how to behave to children by parents. Furthermore, the imitation is often used for communication among children, and the gestures are useful to understand the intentions

and emotional expressions. Basically, imitation is defined as the ability to recognize and reproduce other's actions. The concept of imitative learning has been applied to robotics. In the traditional researches of learning by observation, motion trajectories of a human arm assembling or handling objects are measured, and the obtained data are analyzed and transformed for the motion control of a robotic manipulator. Furthermore, various neural networks have been applied to imitative learning for robots. The discovery of mirror neurons is especially important (Rizzolatti et al., 1996). Each mirror neuron activates not only by performing a task, but also by observing somebody performing the same task. Rao and Meltzoff classified imitative abilities into four stage progression: (i) body babbling, (ii) imitation of body movements, (iii) imitation of actions on objects, and (iv) imitation based on inferring intentions of others (Rao & Meltzoff, 2003). If a robot can perform all stages of imitation, the robot might develop in the same way as humans. While the imitative learning is basically unidirectional from a demonstrator to a learner, the instructive learning is bidirectional between an instructor and a learner. An instructor assesses the learning state of the learner, and then shows additional and suitable demonstrations to the learner. Collaborative learning is slightly different from the imitative learning and instructive learning, because neither exact teaching data nor target demonstration is given to agents beforehand in the collaborative learning. The solution is found or searched through interaction among multiple agents. Therefore, the collaborative learning may be classified as the category of self-learning.

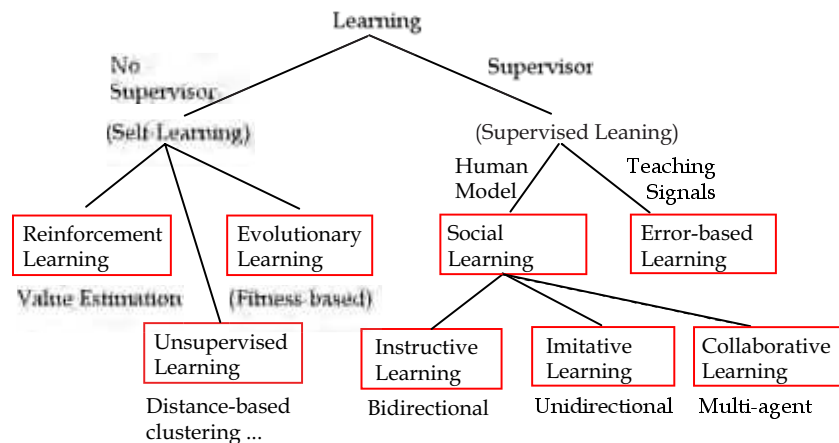


Figure 1. Learning methods in robotics

1.3 Human Detection and Gesture Recognition

Human interaction based on gestures is very important to realize the natural communication. The meaning of gesture can be understood through the actual interaction and imitation. Therefore, we focus on the human detection and gesture recognition for imitative learning of human-friendly network robots. 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 as a result of the human interaction is also performed. In this way, the imitative learning

requires various learning capabilities of model observation, model clustering, model selection, model reproduction, and model learning simultaneously. First of all, the robot detects a human based on image processing with a steady-state genetic algorithm (SSGA) (Syswerda, 1991). Next, a series of the movements of the human hand by image processing as model observation, or the hand motion pattern, is extracted by a spiking neural network (Gerstner, 1999). Furthermore, SSGA is used for generating a trajectory similar to the human hand motion pattern as model reproduction (Kubota, Nojima et al., 2006). In the following, we explain the methods for the human detection and gesture recognition based on ambient intelligence.

2. Partner Robots and Environmental System

We developed two different types of partner robots; a human-like robot called Hubot (Kubota, Nojima et al., 2006) and a mobile PC called MOBiMac (Kubota, Tomioka et al., 2006) in order to realize the social communication with humans. Hubot is composed of a mobile base, a body, two arms with grippers, and a head with pan-tilt structure. The robot has various sensors such as a color CCD camera, two infrared line sensors, a microphone, ultrasonic sensors, and touch sensors (Figure 2(a)). The color CCD camera can capture an image with the range of -30° and 30° in front of the robot. Two CPUs are used for sensing, motion control, and wireless network communication. The robot can take various behaviors like a human. MOBiMac is also composed of two CPUs used for PC and robotic behaviors (Figure 2(b)). The robot has two servo motors, four ultrasonic sensors, four light sensors, a microphone, and a CCD camera. The basic behaviors of these robots are visual tracking, map building, imitative learning (Kubota, 2005), human classification, gesture recognition, and voice recognition. These robots are networked, and share environmental data among each other. Furthermore, the environmental system based on a sensor network provides a robot with its environmental data measured by the equipped sensors.

Human detection is one of the most important functions in the ambient intelligence space. The visual angle of the robot is very limited, while the environmental system is designed to observe the wide range of the environment. Therefore, human detection can be easily performed by the monitoring of the environmental system or by the cooperative search of several robots based on the ambient intelligence (Figure 3). In the following sections, we explain how to detect a human based on image processing.

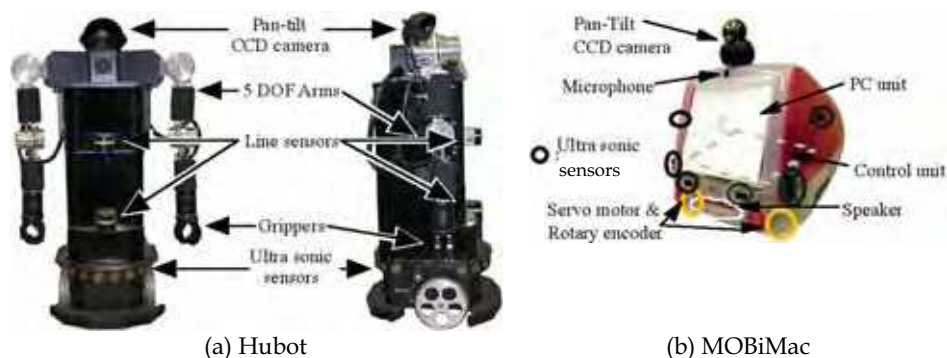


Figure 2. Hardware architecture of partner robots

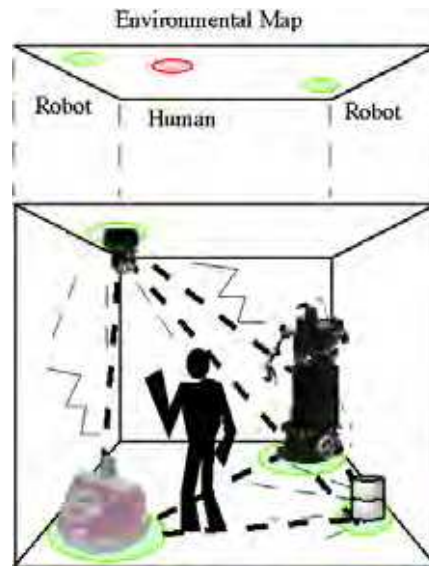


Figure 3. The sensor network among robots and their environmental system

3. Human Detection and Gesture Recognition Based on Ambient Intelligence

3.1 Human Detection

Human detection is performed by both robots and environmental system. Pattern matching has been performed by various methods such as template matching, cellular neural network, neocognitron, and dynamic programming (DP) matching (Fukushima, 2003; Mori et al., 2006). 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 from an image, and the aim of the target recognition is to identify the target from classification candidates. Since the image processing takes much time and computational cost, a full size of image processing to every image is not practical. Therefore, we use a reduced size of image to detect a moving object for fast human detection. First, 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 human detection result; the maximum and minimum of the image sizes are 640×480 and 80×60 , respectively. The differential extraction calculates the difference of the number of pixels between the previous and current images. If the robot does not move, the COG of the difference represents the location of the moving object. Therefore, the main search area for fast human detection can be formed according to the COG for fast human detection.

We use a steady-state genetic algorithm (SSGA) for human detection and object detection as one of search methods, because SSGA can easily obtain feasible solutions through environmental changes at low computational costs. SSGA simulates a continuous model of the generation, which eliminates and generates a few individuals in a generation (iteration) (Syswerda, 1999). Here SSGA for human detection is called SSGA-H, while SSGA for object detection used as human hand detection is called SSGA-O.

Human skin and hair colors are extracted by SSGA-H based on template matching. Figure 4 shows a candidate solution of a template used for detecting a target. 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 . One iteration is composed of selection, crossover, and mutation. The worst candidate solution is eliminated ("Delete least fitness" selection), and is replaced by the candidate solution generated by the crossover and the mutation. We use elitist crossover and adaptive mutation. The elitist crossover randomly selects one individual and generates an individual by combining genetic information from the randomly selected individual and the best individual. Next, the following adaptive mutation is performed to the generated individual,

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

where f_i^H is the fitness value of the i th individual, f_{\max}^H and f_{\min}^H 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^H and β_j^H are the coefficient ($0 < \alpha_j^H < 1.0$) and offset ($\beta_j^H > 0$), respectively. In the adaptive mutation, the variance of the normal random number is relatively changed according to the fitness values of the population. Fitness value is calculated by the following equation,

$$f_i^H = C_{Skin}^H + C_{Hair}^H + \eta_1^H \cdot C_{Skin}^H \cdot C_{Hair}^H - \eta_2^H \cdot C_{Other}^H \quad (2)$$

where C_{Skin}^H , C_{Hair}^H and C_{Other}^H 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 ($\eta_1^H, \eta_2^H > 0$). Therefore, this problem results in the maximization problem. The iteration of SSGA-H is repeated until the termination condition is satisfied.

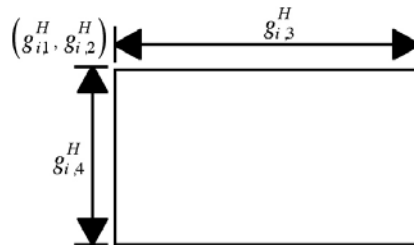


Figure 4. A template used for human detection in SSGA-H

3.2 Human Hand Detection

We proposed a method for human hand detection based on the finger color and edges (Kubota & Abe, 2006), but we assume the human uses objects such as balls and blocks for performing a gesture interacting with a robot. Because the main focus is the gesture recognition based on human hand motion and the exact human hand detection is out of scope in this chapter. Therefore, we focus on color-based object detection with SSGA-O based on template matching. The shape of a candidate template is generated by the SSGA-O. We assume the human uses objects such as balls and blocks for performing a gesture interacting with a robot. Figure 5 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 $(g_{i,1}^O + g_{i,j}^O \cos(g_{i,j+1}^O))$,

$g_{i,2^0} + g_{i,j}^0 \sin(g_{i,j+1}^0)$, $i=1, 2, \dots, n, j=3, \dots, 2 \square m+2$; $O_i (= (g_{i,1}^0, g_{i,2}^0))$ 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 $(g_{i,1}^0, g_{i,2}^0, \dots, g_{i,2m+2}^0)$. We used an octagonal template ($m=8$). The fitness value of the i th candidate template is calculated as follows.

$$f_i^0 = C_{Target}^0 - \eta^0 \cdot C_{Other}^0 \quad (3)$$

where η^0 is the coefficient for penalty ($\eta^0 > 0$); C_{Target}^0 and C_{Other}^0 indicate the numbers of pixels of 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, this problem also results in the maximization problem. 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 $t=1, 2, \dots, T^G$.

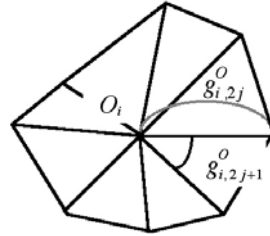


Figure 5. A template used for object detection in SSGA-O

3.3 Human Hand Motion Extraction

Various types of artificial neural networks have been proposed to realize clustering, classification, nonlinear mapping, and control (Jang et al., 1997; Kuniyoshi & Shimozaki, 2003; Rumelhart et al., 1986). Basically, artificial neural networks are classified into pulse-coded neural networks and rate-coded neural networks from the viewpoint of abstraction level (Gerstner, 1999). 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 (Anderson & Rosenfeld, 1988). 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 contexts. Therefore, we apply a SNN for memorizing several motion patterns of a human hand, because the human hand motion has specific dynamics.

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;

$$h_i(t) = \tanh\left(h_i^{syn}(t) + h_i^{ext}(t) + h_i^{ref}(t)\right) \quad (4)$$

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

$$h_i^{syn}(t) = \gamma^{syn} \cdot h_i(t-1) + \sum_{j=1, j \neq i}^N w_{j,i} \cdot h_j^{EPSP}(t) \quad (5)$$

Furthermore, $h_i^{ref}(t)$ indicates the refractoriness factor of the neuron; $w_{j,i}$ is a weight coefficient from the j th to i th neuron; $h_j^{EPSP}(t)$ is the excitatory postsynaptic potential (EPSP) of the j th 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;

$$h_i^{EPSP}(t) = \sum_{n=0}^T \kappa^n p_i(t-n) \quad (6)$$

where κ is the discount rate ($0 < \kappa < 1.0$); $p_i(t)$ is the output of the i th 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,

$$h_i^{ref}(t) = \begin{cases} \gamma^{ref} \cdot h_i^{ref}(t-1) - R & \text{if } p_i(t-1) = 1 \\ \gamma^{ref} \cdot h_i^{ref}(t-1) & \text{otherwise} \end{cases} \quad (7)$$

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

$$p_i(t) = \begin{cases} 1 & \text{if } h_i^{ref}(t) \geq q_i \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

where q_i is a threshold for firing. Here spiking neurons are arranged on a planar grid (Figure 6) 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;

$$h_i^{ext}(t) = \exp\left(-\frac{\|\mathbf{c}_i - \mathbf{G}(t)\|^2}{2\sigma^2}\right) \quad (9)$$

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 $\mathbf{G}(t)$. The weight parameters are trained based on the temporal Hebbian learning rule as follows,

$$w_{j,i} \leftarrow \tanh\left(\gamma^{wgt} \cdot w_{j,i} + \xi^{wgt} \cdot h_j^{EPSP}(t-1) \cdot h_i^{EPSP}(t)\right) \quad (10)$$

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

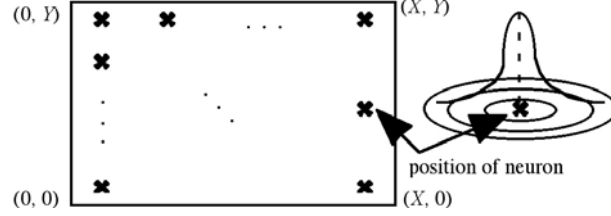


Figure 6. Spiking neurons for gesture recognition

3.4 Gesture Recognition and Learning

This subsection explains a method for clustering human hand motions. Cluster analysis is used for grouping or segmenting observations into subsets or clusters based on similarity. Self-organizing map (SOM), K -means algorithm, and Gaussian mixture model are often applied as clustering algorithms (Hastie et al., 2001; Kohonen, 2001). SOM can be used as incremental learning, while K -means algorithm and Gaussian mixture model use all observed data in the learning phase (batch learning). In this paper, we apply SOM for clustering spatio-temporal patterns of pulse outputs from the SNN, because the robot observes a human hand motion at a time. 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,

$$\mathbf{v} = (v_1, v_2, \dots, v_N) \quad (11)$$

where v_i 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 \mathbf{r}_i , the Euclidian distance between an input vector and the i th reference vector is defined as

$$d_i = \|\mathbf{v} - \mathbf{r}_i\| \quad (12)$$

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

$$k = \arg \min_i \{\|\mathbf{v} - \mathbf{r}_i\|\} \quad (13)$$

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

$$\mathbf{r}_i \leftarrow \mathbf{r}_i + \zeta^{SOM} \cdot \zeta_{k,i}^{SOM} \cdot (\mathbf{v} - \mathbf{r}_i) \quad (14)$$

where ζ^{SOM} is a learning rate ($0 < \zeta^{SOM} < 1.0$); $\zeta_{k,i}^{SOM}$ is a neighborhood function ($0 < \zeta_{k,i}^{SOM} < 1.0$). Accordingly, the selected output unit is the nearest pattern among the previously learned human hand motion patterns.

4. Experiments

This section shows several experimental results of human detection and gesture recognition. We conducted several experiments of human detection by the environmental system and the robot (Kubota & Nishida, 2006; Sasaki & Kubota, 2006). Figures 7 and 8 show human detection results by SSGA-H from the ceiling view and from the robot view, respectively. The environmental system detects two people in the complicated background in Figure 7. In Figure 8 (a), first the robot used the high resolution of images to detect a walking human. Afterward, as the human gradually came toward the front of the robot, the robot used the lower resolution of images to reduce computational cost and detected the human (Figure 8 (b)).



Figure 7. Human detection results from the ceiling camera of the environmental system by SSGA-H

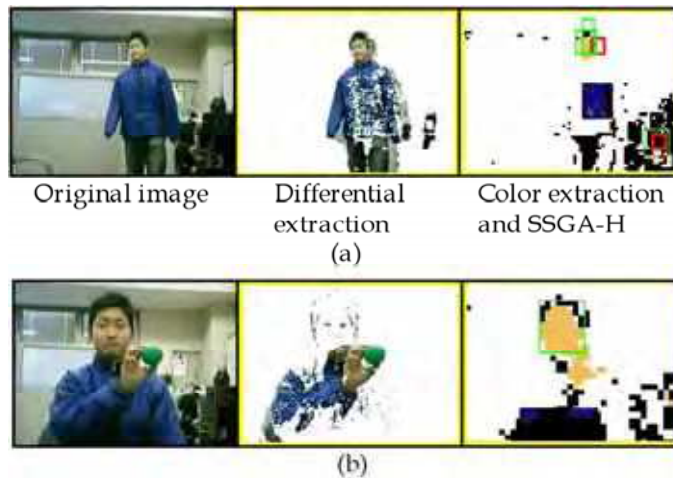


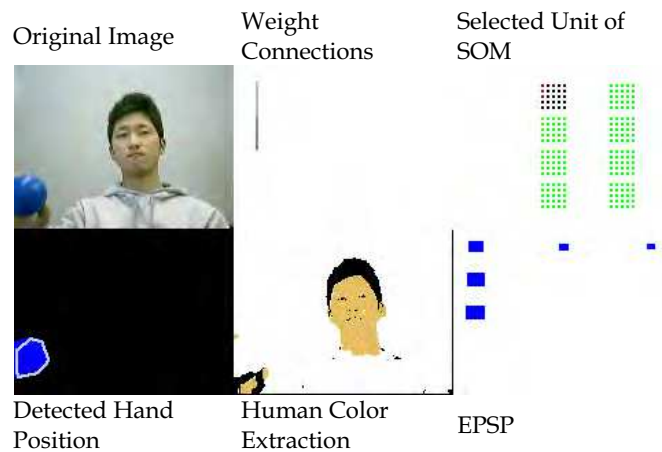
Figure 8. Human detection results from the robot by SSGA-H

We conducted several experiments of gesture recognition. The number of units used in SOM is 8. Figures 9 and 10 show examples of human hand motion and the learning of SNN, respectively. The human moves his hand from the upper left to the lower right through upper right on the human position. The EPSP based on spike outputs does not cover the human hand motion at the first trial (Figure 10 (a)), but after learning, the EPSP successfully covers the human hand motion based on the trained weight connections where the depth of color indicates the strength of weight connections between two neurons (Figure 10 (b)).

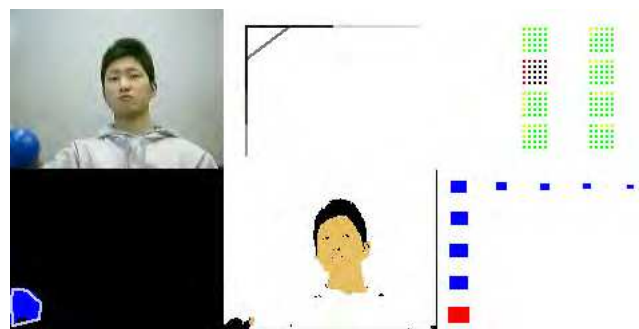
Figure 11 shows the learning results of various human hand motion patterns. The weight connections are trained according to the frequency of local movements between two neurons according to the human hand motion patterns. A different unit of SOM is selected when the different human hand motion is shown to the robot. The selected unit is depicted as dark boxes where each small box indicates the magnitude of the value in the reference vector. The 8th and 5th units are selected as gestures according to human hand motions, respectively. Furthermore, the comparison between SNN and RNN and the detailed analysis are discussed in (Kubota, Tomioka et al., 2006).



Figure 9. A human hand motion

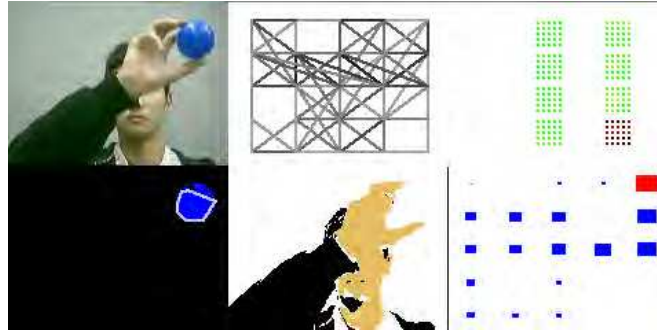


(a) The initial state of EPSP at the first trial

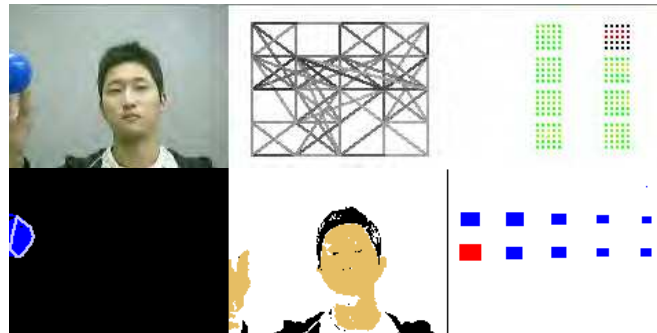


(b) The state of EPSP after learning

Figure 10. Learning of SNN with the human hand motion of Figure 9



(a) A recognition result of human hand motion of the 8 figure; the 8th unit is selected.



A recognition result of the human hand motion from the left to right; the 5th unit is selected.

Figure 11. Gesture recognition by SNN and SOM after learning

5. Summary

This chapter introduced the methods for human detection and gesture recognition based on ambient intelligence. The experimental results show the effectiveness of the methods for human detection and gesture recognition, but we should use various types of sensory information in addition to visual images. We proposed the concept of evolutionary robot vision. The main sensory information is vision, but we can integrate and synthesize various types of sensory information for image processing. Possible sensory information to improve the performance is distance. We developed a 3D modeling method based on CCD cameras and a laser range finder with a small and compact pan-tilt mechanism. We intend to develop a gesture recognition method based on various types of sensory information. Furthermore, network robots based on ambient intelligence are an attractive approach to realize sophisticated services in the next generation society. One of attractive approaches is the SICE City project. The aim of SICE city project is to build up the methodology and concept in the city design to realize sophisticated services for residents based on measurement technology, network technology, control technology, and systems theory. We will discuss the applicability of human detection, human recognition, gesture recognition, and motion recognition based on ambient intelligence to human-friendly functions in the city design.

6. Acknowledgment

The author would like to thank K.Nishida, T. Shimizu, and Y. Tomioka for all their help. This research is partially supported by grants-in-aid for scientific research on priority areas (B) 17700204 from Japan Society for the Promotion of Science.

7. References

- Anderson, J. A. & Rosenfeld, E. (1988). *Neurocomputing*, The MIT Press, Cambridge, Massachusetts, US
- Billard, A. (2002). Imitation, *Handbook of Brain Theory and Neural Networks*, Arbib, M. A., (Ed.), 566-569, MIT Press, Cambridge, Massachusetts, US
- Doctor, F.; Hagaras, H. & Callaghan, V. (2005). A type-2 fuzzy embedded agent to realise ambient intelligence in ubiquitous computing environments, *Information Sciences*, Vol.171, Issue 4, pp.309-334
- Fukushima, K. (2003). Neural network model restoring partly occluded patterns, *Knowledge-Based Intelligent Information and Engineering Systems, (7th International Conference, KES 2003)*, Part II, eds: V. Palade, R. J. Howlett, L. Jain, Berlin - Heidelberg: Springer-Verlag, pp. 131-138
- Gerstner, W. (1999). Spiking neurons, *Pulsed Neural Networks*, Maass, W. & Bishop, C.M., (Ed.), 3-53, MIT Press, Cambridge, Massachusetts, US
- Hastie, T.; Tibshirani, R. & Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer-Verlag, New York
- Jang, J.-S.R.; Sun, C.-T. & Mizutani, E. (1997). *Neuro-Fuzzy and Soft Computing*, Prentice-Hall Inc.
- Kemmotsu, K.; Tomonaka, T.; Shiotani, S.; Koketsu, Y. & Iehara, M. (2005). Recognizing human behaviors with vision sensors in network robot systems, *Proceedings of The 1st Japan-Korea Joint Symposium on Network Robot Systems (JK-NRS2005)*
- Khemapech, I.; Duncan, I. & Miller, A. (2005). A survey of wireless sensor networks technology, *Proceedings of The 6th Annual PostGraduate Symposium on The Convergence of Telecommunications, Networking and Broadcasting*
- Kim, J-H.; Kim, Y-D. & Lee, K-H. (2004). The third generation of robotics: ubiquitous robot, *Proceedings of 2nd International Conference on Autonomous Robots and Agents(ICARA 2004)*, pp.1-12, Palmerston North, New Zealand
- Kohonen, T. (2001). *Self-Organizing Maps*, 3rd Edition, Springer-Verlag, Berlin, Heidelberg
- Kubota, N. (2005). Computational Intelligence for Structured Learning of A Partner Robot Based on Imitation, *Information Science*, No. 171, pp.403-429
- Kubota, N. & Abe, M. (2006). Human hand detection for gestures recognition of a partner robot, *Proceedings of World Automation Congress (WAC) 2006*, ifmip_214, Budapest, Hungary
- Kubota, N. & Nishida, K. (2006). Cooperative perceptual systems for partner robots based on sensor network, *International Journal of Computer Science and Network Security (IJCSNS)*, Vol. 6, No. 11, pp. 19-28, ISSN 1738-7906
- Kubota, N.; Nojima, Y.; Kojima, F. & Fukuda, T. (2006). Multiple Fuzzy State-Value Functions for Human Evaluation through Interactive Trajectory Planning of a Partner Robot, *Soft Computing*, Vol.10, No.10 pp.891-901

- Kubota, N.; Tomioka, Y. & Abe, M. (2006). Temporal coding in spiking neural network for gestures recognition of a partner robot, *Proceedings of Joint 3rd International Conference on Soft Computing and Intelligent Systems and 7th International Symposium on advanced Intelligent Systems (SCIS & ISIS 2006)*, pp. 737-742, Tokyo, Japan
- Kulyukin, et al., (2004). RFID in robot-assisted indoor navigation for the visually impaired, *Proceedings of 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2004)*, pp. 1979-1984
- Kuniyoshi, Y. & Shimozaki, M. (2003). A self-organizing neural model for context-based action recognition, *Proceedings of IEEE EMBS Conference on Neural Engineering*, pp. 442-445, Capri Island, Italy.
- Mori, A.; Uchida, S.; Kurazume, R.; Taniguchi, R.; Hasegawa, T. & Sakoe, H. (2006). Early recognition and prediction of gestures, *Proceedings of International Conference on Pattern Recognition*, pp. 560-563, Hong Kong, China
- Morikawa, K.; Agarwal, S.; Elkan, C. & Cottrell, G. (2001). A taxonomy of computational and social learning, *Proceedings of Workshop on Developmental Embodied Cognition*
- Morioka, K. & Hashimoto, H. (2004). Appearance based object identification for distributed vision sensors in intelligent space, *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2004)*, pp.199-204
- Nolfi, S. & Floreano, D. (2000). *Evolutionary Robotics: The Biology, Intelligence, and Technology of Self-Organizing Machines*, The MIT Press, Cambridge, Massachusetts, US
- Rao, R. P. N. & Meltzoff, A. N. (2003). Imitation learning in infants and robots: towards probabilistic computational models, *Proceedings of Artificial Intelligence and Simulation of Behaviors*, pp. 4-14
- Rumelhart, D. E.; McClelland, J. L. & the PDP Research Group. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volumes 1*, The MIT Press, Cambridge, Massachusetts, US
- Sasaki, H. & Kubota, N. (2006). Map building and monitoring for environmental information gathering based on intelligent networked robots, *Proceedings of Environmental Modelling and Simulation 2006*, 556-070, ISBN 0-88986-619-8, St. Thomas, USVI, USA.
- Satoh, I. (2006). Location-based services in ubiquitous computing environments, *International Journal of Digital Libraries*, ISSN 1432-1300
- Syswerda, G. (1991) A study of reproduction in generational and steady-state genetic algorithms, *Genetic Algorithms*, Morgan Kaufmann Publishers Inc., San Mateo



Face Recognition

Edited by Kresimir Delac and Mislav Grgic

ISBN 978-3-902613-03-5

Hard cover, 558 pages

Publisher I-Tech Education and Publishing

Published online 01, July, 2007

Published in print edition July, 2007

This book will serve as a handbook for students, researchers and practitioners in the area of automatic (computer) face recognition and inspire some future research ideas by identifying potential research directions. The book consists of 28 chapters, each focusing on a certain aspect of the problem. Within every chapter the reader will be given an overview of background information on the subject at hand and in many cases a description of the authors' original proposed solution. The chapters in this book are sorted alphabetically, according to the first author's surname. They should give the reader a general idea where the current research efforts are heading, both within the face recognition area itself and in interdisciplinary approaches.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Naoyuki Kubota (2007). Human Detection and Gesture Recognition Based on Ambient Intelligence, Face Recognition, Kresimir Delac and Mislav Grgic (Ed.), ISBN: 978-3-902613-03-5, InTech, Available from: http://www.intechopen.com/books/face_recognition/human_detection_and_gesture_recognition_based_on_ambient_intelligence

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2007 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](#), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.