

# Human Motion Tracking for Cognitive Rehabilitation in Informationally Structured Space Based on Sensor Networks

Yuichiro Toda

Tokyo Metropolitan University  
Tokyo, Japan  
toda-yuuichirou@sd.tmu.ac.jp

Yuki Kodai

Tokyo Metropolitan University  
Tokyo, Japan  
yuuki-kodai@sd.tmu.ac.jp

Eriko Hiwada

Tokyo Metropolitan University  
Tokyo, Japan  
hiwada-eriko@sd.tmu.ac.jp

Naoyuki Kubota

Tokyo Metropolitan University  
Tokyo, Japan  
kubota@tmu.ac.jp

**Abstract**—This paper discusses measurement methods of human behaviors based on sensor network and human interaction of rehabilitation using robot partners. First, we explain robot partners and sensor networks for rehabilitation. Next, we apply a steady-state genetic algorithm to extract human motions from 3D distance image. Finally, we discuss the effectiveness of the proposed methods through several experimental results.

**Keywords**—component; Rehabilitation, Human Motion Tracking, Sensor Networks, Robot Partners, Evolutionary Computation

## I. INTRODUCTION

Recently, as the increase of number of elderly people, much more caregivers are required for the support for them. The mental and physical care is very important in order to avoid the progress of dementia of elderly people living alone in home, but such people have little chances to talk with other people and to perform daily exercise. On the other hand, the number of patients of brain damage is also increasing. The cognitive therapeutic exercise is useful in the rehabilitation after brain damage [1]. In the cognitive rehabilitation, before the actual motions, a patient pays attention to the cognitive cycle of the pre-planning of actual motions and the prediction of sensing in the actual motion in order to recover the cognitive ability from perception to action. It takes much time and load for a therapist to observe and navigate the behaviors of a patient in the rehabilitation program. In the above discussion, the motion measurement, natural communication, and human-friendly interaction are very important and necessary for both elderly people and patients. Of course, it is really ideal that these roles should be done by human caregivers and therapists, but the number of caregivers and therapists is not enough in the current situation. The introduction of robot partners instead of people is one of possible solutions. Such a robot partner should understand human behaviors for natural communication with elderly people. However, it is also difficult for a robot partner to observe the human behaviors. Therefore, sensor network and

portable sensing devices can be applied for observing and understanding human behaviors [2,3].

The emerging synthesis of information technology (IT), network technology (NT), and robot technology (RT) is one of the most promising approaches to realize a safe, secure, and comfortable society for the next generation [4]. NT can provide the robot with computational capabilities based on various types of information outside of robots. Actually the robot directly receives the environmental information through a local area network without the measurement by the robot itself [5,6]. Wireless sensor networks [7-9] realize to gather the huge data on environments. However, it is very difficult to store all of huge data in real time. Furthermore, some features should be extracted from the gathered data to obtain the required information. Therefore, intelligent technology is required in wireless sensor networks.

Intelligence technology and information technology have been discussed from various points of view. Information resources and the accessibility within an environment are essential for both people and robots. Therefore, the environment surrounding people and robots should have a structured platform for gathering, storing, transforming, and providing information. Such an environment is called informationally structured space [10,11] (Fig.1). The intelligent technology for the design and usage of the informationally structured space should be discussed from various points of view such as information gathering of real environment and cyber space, structuralization, visualization and display of the gathered information. The structuralization of informationally structured space realizes the quick update and access of valuable and useful information for people. It is very useful for both robots and people to easily access the information on real environments. The information is transformed into the useful form suitable to the features of robot partners and people. Furthermore, if the robot can share the environmental information with people, the communication with people might become very smooth and natural.

In this paper, we discuss the measurement method of human

motions in the informationally structured space including robot partners and sensing devices. First, we apply a steady-state genetic algorithm (SSGA) [12,13] to perform human hand motion tracking by using the time series of 3D distance image based on differential extraction. SSGA is one of evolutionary optimization methods using selection, mutation, and crossover operators [14]. The shape and size of a template is updated by SSGA in this study. Next, we show experimental results, and discuss the effectiveness of the proposed method.

This paper is organized as follows. Section 2 explains the robot partners, and discusses how to use robot partners for rehabilitation. Section 3 proposes methods of human detection and human hand motion extraction. Section 4 shows experimental results of the proposed method, and Section 5 summarizes the paper, and discusses the future vision of robot partners for rehabilitation.

## II. INFORMATIONALLY STRUCTURED SPACE

### A. Robot Partners

We can use three different types of robot partners from the interactive point of view (Fig.2). One is a physical robot partner. We can interact with the physical robot partner by using multi-modal communication like a human [15]. The next one is a pocket robot partner. The pocket robot partner has no mobile mechanism, but we can easily bring it everywhere and can interact with the robot partner by touch and physical interface. The last one is a virtual robot partner. The virtual robot partner is in the virtual space in the computer, but we can immerse into the virtual space, and interact with it through the virtual person or robot. The interaction style of these three types of robot partners is different, but they share the same personal database and interaction logs, and can interact with the person based on the same interaction rules independent from the style of interfaces.

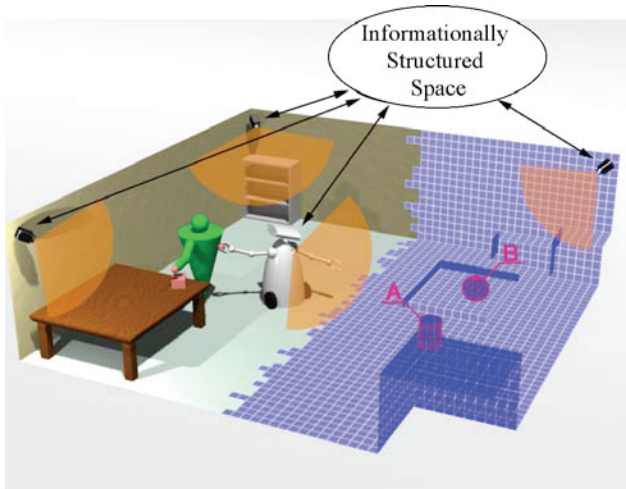


Figure 1. The concept of informationally structured space

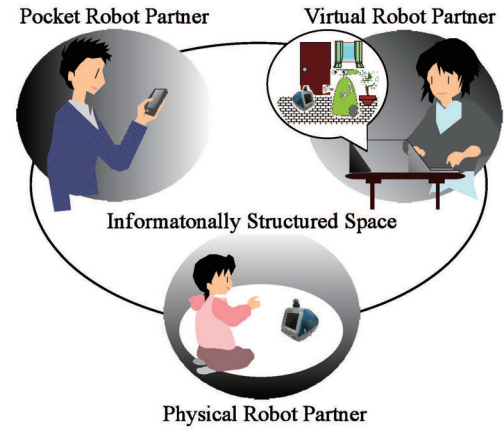


Figure 2. Interaction with robot partners

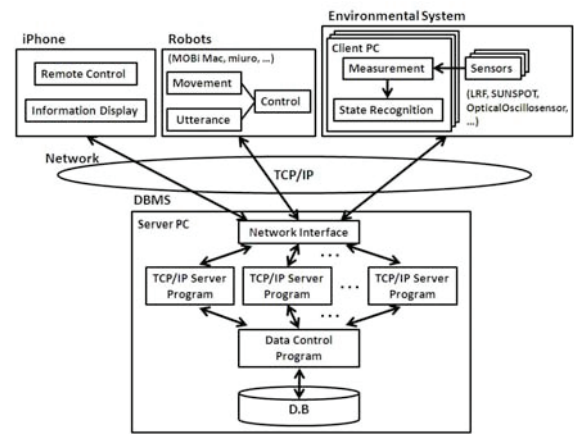


Figure 3. Data flow in informationally structured space

### B. Data Flow in Informationally Structured Space

Figure 3 shows the data flow in the developed system based on informationally structured space. The developed system is divided into four components; (1) database management server, (2) robot systems, (3) environmental systems, and (4) human interface systems. The environmental system is based on a ubiquitous wireless sensor network composed of sensors equipped with wall, floor, ceiling, furniture, and home appliances. These sensors measure the environmental data and human motions. The measured data are transmitted to the database server, and then feature extraction is performed. Each robot can receive the environmental information from the database server, and serve as a partner to the person. Furthermore, the user interface device is used for the person to access the environmental information.

### C. Rehabilitation Support in Informationally Structured Space

Figure 4 shows an example of rehabilitation support by a physical robot partner. A patient can perform the rehabilitation program through interaction with the robot. The robot can repeatedly show the behaviors demonstrated for the rehabilitation programs, and can observe behaviors of the

patience. The observed data can be useful and important information for the therapist to understand the state of rehabilitation, and to update the rehabilitation program. Furthermore, such a physical robot partner can avoid the progress of dementia of elderly people and urge daily routine exercises of the elderly people to make them maintain the physical and mental health.



Figure 4. A conceptual figure of rehabilitation by a robot partner



(a) Wireless accelerometer (b) Sensing direction

Figure 5. Wireless accelerometer developed by ZMP, Inc.



Figure 6. Wireless accelerometer built in a building block

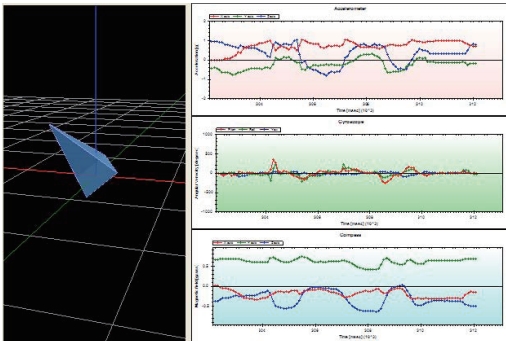


Figure 7. A measurement result of human hand motion

Next, we explain the sensor network devices. We use 3-axis accelerometer to measure the human hand posture of grasping blocks. Figure 5 shows the appearance of 3-axis accelerometer named IMU-Z developed by ZMP, Inc. [16]. The size of this device is 42[mm] x 52.5[mm] x 20.5[mm], but this device includes accelerometer, Jyro, electrical compass. The sampling interval is 10 [msec]. Furthermore, this device can be connected with PC through Bluetooth wireless communication. This device is built in a building block used for rehabilitation (Fig.6). Figure 7 shows a measurement result of posture of human grasping behavior of the block including the wireless sensor device. The experiment result shows the time-series of angle of X, Y, and Z axes. This shows the change of the posture according to the grasping behavior.



Figure 8. 3D distance image sensor of M Kinect sensor

TABLE I. SPECIFICATION OF KINECT SENSOR

|                          |                  |
|--------------------------|------------------|
| Horizontal Field of View | 57 degrees       |
| Vertical Field of View   | 43 degrees       |
| Physical Tilt Range      | $\pm 27$ degrees |
| Depth Sensor Range       | 1.2M – 3.5M      |
| Number of Pixels         | $640 \times 480$ |

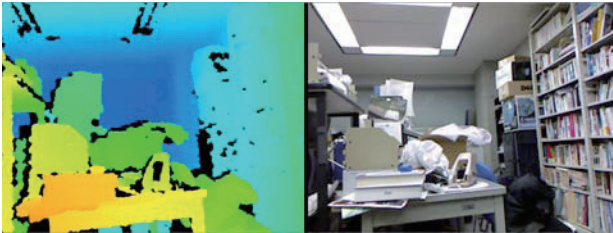


Figure 9. 3D distance image and color image

3D distance image sensor can measure the distance from the device like a camera in realtime. We use a Kinect sensor developed by Microsoft [17]. This device includes 3D distance image sensor, microphones, RGB camera, accelerometer, and tilting-up mechanism. This device can be connected with PC through USB. Table 1 shows the main specification of 3D distance image sensor. Figure 9 shows an example of the distance and color image measured by Kinect sensor.

Figure 10 shows an example of robot partner with Kinect sensor. The robot has three behavior modes of normal mode, interaction mode, and rehabilitation mode. In the normal mode, the behavior mode is transited to the interaction mode. Figure 11 shows an example of human interaction. In the normal mode, the kinect sensor is in the normal position ( $\theta = 0^\circ$ ) shown in Fig.11 (a). When a person approaches to a chair, the Kinect sensor is tilted up toward the human face ( $\theta = 27^\circ$ ) shown in Fig.11 (b). If the person tries to touch iPad built with the robot,

the robot supports the edges of iPad (Fig.11 (c)). Finally, the behavior mode is transited to the normal mode after the person leaves the chair. In this way, the robot changes the behavior mode according to the human interaction.



Figure 10. A rehabilitation support system using a robot partner

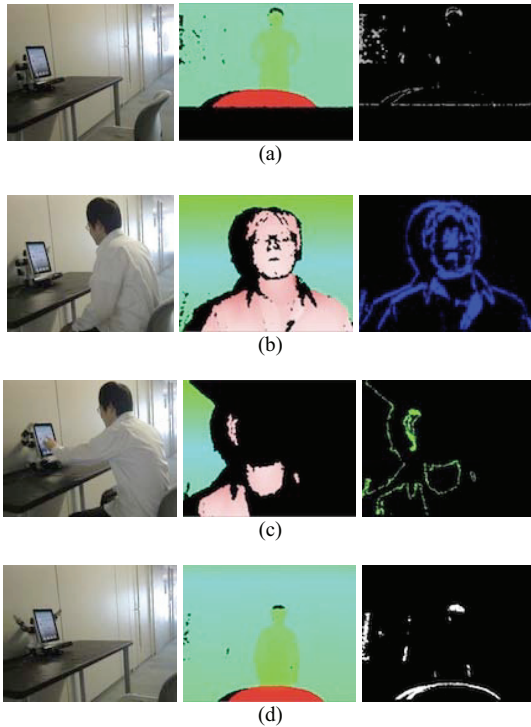


Figure 11. Interaction mode of the robot partner

### III. HUMAN HAND MOTION TRACKING

#### A. Measurement by 3D Distance Image Sensor

We have proposed the concept of evolutionary robot vision based on the active vision and evolutionary search [15]. The time series of evolutionary search in dynamic environment is the most important in evolutionary robot vision. The structure of environmental changes is applied to the next visual search. We use a reliability space for human localization. The reliability space is built according to the temporal difference based on distance image. Figures 12 and 13 show the procedure of human hand motion extraction and an example of image processing, respectively. In the initialization, the long-term memory images of illumination and distance;  $a_{i,j}^L$  is generated using the current images of distance,  $a_{i,j}^C$  where  $i=X_{\min}, \dots, X_{\max}$  and  $j=Y_{\min}, \dots, Y_{\max}$ . Furthermore, each value of the reliability

space  $b_{i,j}$  ( $0 \leq b_{i,j} \leq 1$ ) is initialized at 1. Next, the new images of illumination and distance are taken from the 3D distance image sensor. Figure 13 (a), (b), and (c) show the original images of distance, the result of edge detection, and the long-term memory images of distance, respectively. When a person enters into the view angle of the camera, we obtain the temporal difference images  $d_{i,j}$  by the differential extraction between the current images and the long-term memory images in the following;

$$d_{i,j} = |a_{i,j}^C - a_{i,j}^L| \quad (1)$$

Next, the reliability space is updated in the following;

$$b_{i,j} \leftarrow \begin{cases} b_{i,j} - \alpha_1^B & \text{if } d_{i,j} > \theta^B \\ b_{i,j} + \alpha_2^B & \text{otherwise} \end{cases} \quad (2)$$

where  $0 < \alpha_2^B < \alpha_1^B < 1$  and  $\theta^B > 0$ . The reliability of a pixel ( $i,j$ ) decreases, as the temporal difference is large.

$$a_{i,j}^L \leftarrow \begin{cases} (1 - \alpha_1^L) a_{i,j}^L + \alpha_1^L \cdot a_{i,j}^C & \text{if } b_{i,j} > \theta^L \\ (1 - \alpha_2^L) a_{i,j}^L + \alpha_2^L \cdot a_{i,j}^C & \text{otherwise} \end{cases} \quad (3)$$

where  $0 < \alpha_2^L < \alpha_1^L < 1$  and  $0 < \theta^L < 1$ . If the amount of the update of the long-term memory is large as the reliability is high. Next, we construct the search space,  $c_{i,j}$  ( $0 \leq c_{i,j} \leq 1$ ) by combining the temporal images of illumination and distance in the following weighted sum;

$$c_{i,j} = \alpha_1^S \cdot d_{1,i,j} + \alpha_2^S \cdot d_{2,i,j} \quad (4)$$

where  $0 < \alpha_2^S, \alpha_1^S < 1$ .

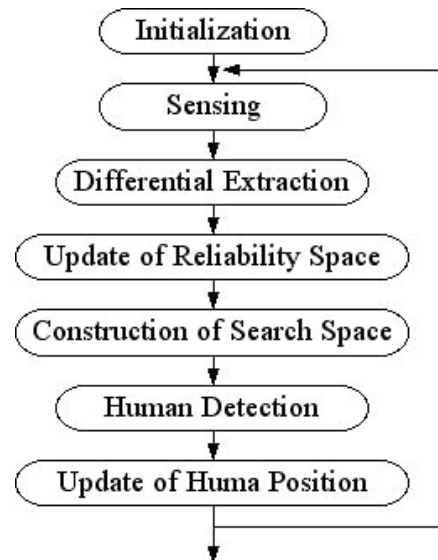


Figure 12. Procedure of human localization

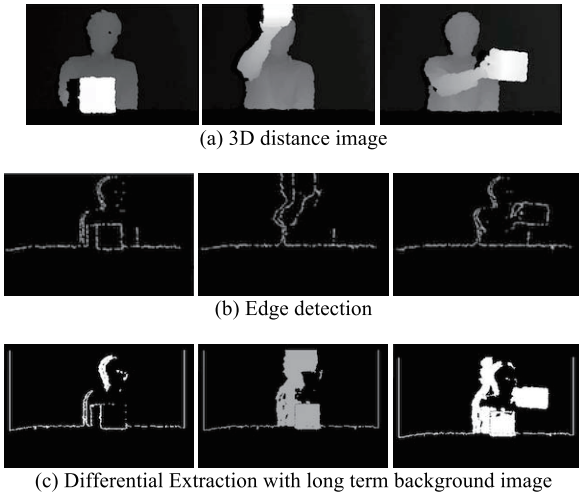


Figure 13. Image processing based on distance data

### B. Human Detection and Tracking

We use a steady-state genetic algorithm (SSGA) [12] based on template matching for extracting a human shape from the background image [12,13,18,19]. We assume that a person is moving in the image. Therefore, we use the search space constructed by the temporal difference. The shape of a candidate template is generated by the SSGA. We used an octagonal template. Figure 14 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  $(g_{i,1}^O + g_{i,j}^O \cos(g_{i,j+m}^O), g_{i,2}^O + g_{i,j}^O \sin(g_{i,j+m}^O))$ ,  $i=1, 2, \dots, G^O$ ,  $j=3, 4, \dots, 2 \times m+2$ ;  $O_i = (g_{i,1}^O, g_{i,2}^O)$  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 ( $m=8$ ), respectively. Therefore, a candidate template is composed of numerical parameters of  $(g_{i,1}^O, g_{i,2}^O, \dots, g_{i,2m+2}^O)$ .

In SSGA, only few existing solutions are replaced with the candidate generated by genetic operators in each generation. In this paper, the worst candidate solution is eliminated ("Delete least fitness" selection), and it is replaced with the candidate solution generated by the crossover and mutation. The fitness value of  $k$ th is calculated by the following equation;

$$fit_i = \alpha_1 f_1 + \alpha_2 f_2 + \alpha_3 f_3 + \alpha_4 f_4 \quad (5)$$

where  $f_1$  indicates the number of pixels;  $f_2$  indicates the evaluation value based on differential extraction;  $f_3$  indicates the evaluation value on the degree of surface of the target building block;  $f_4$  indicates the evaluation value of the movement of human hand;  $\alpha_1, \alpha_2, \alpha_3$  and  $\alpha_4$  are coefficients.

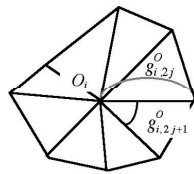


Figure 14. A template used for human detection in SSGA

We use an elitist crossover and an 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. 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,

$$g_{i,j} \leftarrow g_{i,j} + \left( \beta_{1,j} \cdot \frac{f_{\max} - f_i}{f_{\max} - f_{\min}} + \beta_{2,j} \right) \cdot N(0,1) \quad (6)$$

where  $f_i$  is the fitness value of the  $i$ th 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;  $\beta_{1,j}$  ( $\beta_{1,j} > 0$ ) and  $\beta_{2,j}$  ( $\beta_{2,j} > 0$ ) are the coefficients and offset, 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. Furthermore, we use k-means clustering algorithm to detect people, not a single person. The k-means is used for the clustering according to the distribution of the candidate solutions [18]. The genetic operators of selection and crossover are restricted within the subpopulation corresponding to the cluster as the result of the clustering. The candidate solutions are inherited and used in the next image. According to the evaluation based on the fitness function, some of candidate solutions are replaced with the randomly generated candidate solutions. In this way, the evolutionary search is performed to time series of 3D distance images.

The human hand motion tracking is performed according to the position of the  $m$ th candidate solution obtained by SSGA. The position of the  $h$ th human candidate in the human tracking,  $(X_{h,1}^H, X_{h,2}^H)$ , is updated by the nearest human hand candidate position within the tracking range. The update is performed as follows ( $k=1, 2$ );

$$X_{h,k}^H \leftarrow (1 - \alpha^H) X_{h,k}^H + \alpha^H \cdot \hat{x}_{m,k} \quad (7)$$

where  $\alpha^H$  is the update rate. The distance information  $a_{i,j}^C$  of each pixel  $(i,j)$  is transformed into the position  $(x,y,z) = (x_{1,i,j}, x_{2,i,j}, x_{3,i,j})$  in the Cartesian space. Therefore, the update amount  $\hat{x}_{m,k}$  is obtained by the average of the position information of  $(x_{1,i,j}, x_{2,i,j})$  of pixel  $(i,j)$  in the template of the  $m$ th candidate solution.

## IV. EXPERIMENTAL RESULTS

This section shows experimental results of the proposed method. The population size of SSGA is 100, and the number of iterations in the search is 200 per 3D distance image. The sampling interval of Kinect sensor is approximately 30 ms.

First of all, we show experimental results of human hand extraction and tracking of the triangular gesture of. Figure 15 shows the examples of the search result of the object

corresponding to human hand by SSGA. Figure 15 (a) shows the 3D distance image where the degree of gray depth indicates the distance from the 3D distance image sensor; (b) shows the result of differential extraction; and (c) hand detection and tracking results ( $t=10, 20, 30$ ). The experimental results show that the proposed method can extract human hand position according to the size of the template and the distance from the distance image sensor.

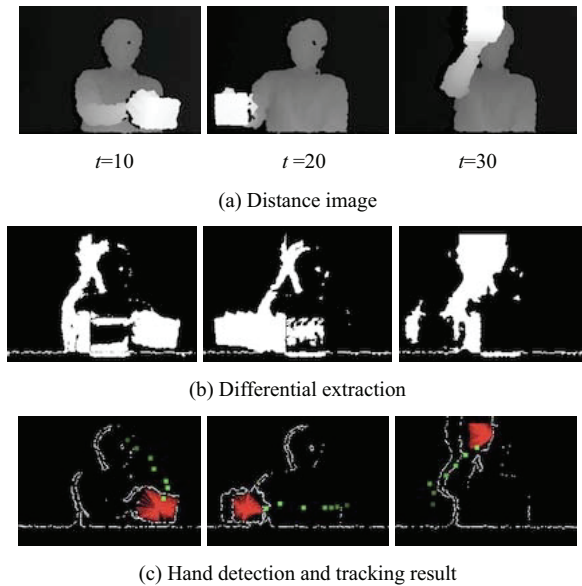


Figure 15. A trajectory of human hand motion by human hand tracking (triangular gesture)

Figures 16 show the tracking results of human hand motions of a horizontal circular gesture. Figure 16 (a) shows the result of differential extraction; (c) hand detection and tracking results; and (c) the 3D distance image where the degree of gray depth indicates the distance from the 3D distance image sensor but the viewpoint is changed to the bird view from the person ( $t=10, 20, 30$ ). Basically, it is very difficult to extract this gesture only by image processing, but the proposed method based on the distance image sensor can extract this complicated gesture. Furthermore, the transformed image of viewpoint from the person is very useful for a therapist to understand the hand motion of a patient.

## V. SUMMARY

This paper developed a measurement system of human hand motion based on sensor network and robot partners for rehabilitation. First, we proposed the interaction system for a robot partner. Next, we applied a steady-state genetic algorithm to extract human hand motions from the 3D distance image. Experimental results show the effectiveness of the proposed methods. The obtained results are available for understanding and analysis of human hand motions.

As a future work, we intend to combine the voice recognition and visual perception to realize natural interaction of robot partners for rehabilitation. Furthermore, we use iPhone and iPod touch for cognitive rehabilitation therapy.

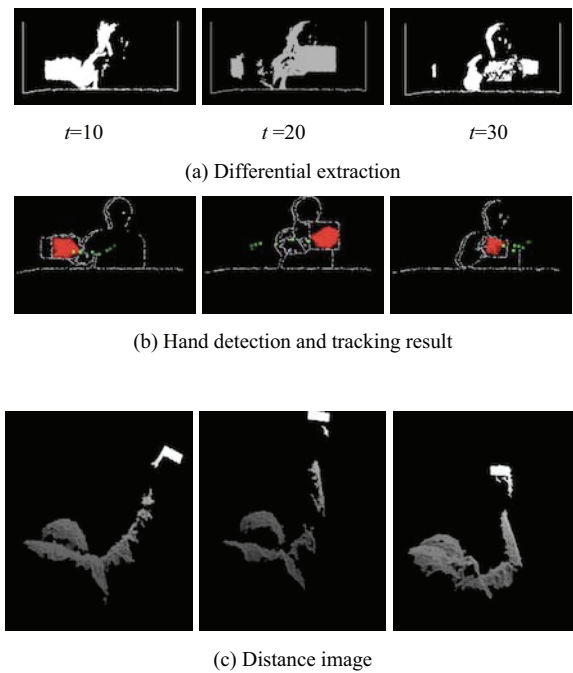


Figure 16. A trajectory of human hand motion extracted by SSGA (a horizontal circular gesture)

## ACKNOWLEDGMENT

This research was partially supported by Grant-in-Aid for Scientific Research (B) No.22300082 from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

## REFERENCES

- [1] S.F.Szekeress, M.Ylvisaker, S.B.Cohen, A Framework for Cognitive Rehabilitation Therapy, Community Re-entry for Head Injured Adults (edited by M.Ylvisaker, E.M.R.Gobble), Butterworth-Heinemann, pp.87-136, 1987.
- [2] S.Atygalle, M.Duff, T.Rikakis, Jiping.H, Low-cost, at-home assessment system with Wii Remote based motion capture, In Proceedings of the Conference on Virtual Rehabilitation, pp.168-174, 2008.
- [3] M.Duff, S.Atygalle, Jiping.H, T.Rikakis, A Portable, Low-cost Assessment Device for Reaching Times, Engineering in Medicine and Biology Conference (EMBC2008), pp.4150-4153, 2008.
- [4] C. Corderio, D. Agrawal, Ad Hoc & Sensor Networks, World Scientific, 2006.
- [5] I. Khemapech, I. Duncan and A. Miller, "A Survey of Wireless Sensor Networks Technology," Proc. of The 6th Annual PostGraduate Symposium on The Convergence of Telecommunications, Networking and Broadcasting, 2005.
- [6] J-H. Kim, Y-D. Kim, and K-H. Lee, "The Third Generation of Robotics: Ubiquitous Robot," Proc. of 2nd International Conference on Autonomous Robots and Agents, 2004, pp.1-12.
- [7] N.Kubota, Y.Shimomura, Human-Friendly Networked Partner Robots toward Sophisticated Services for A Community, Proc. of SICE-ICCAS 2006, pp. 4861-4866, 2006.
- [8] N.Kubota, K.Nishida, Perceptual Control Based on Prediction for Natural Communication of A Partner Robot, IEEE Transactions on Industrial Electronics, Vol.54, No.2, pp.866-877, April, 2007.
- [9] N.Kubota, T.Obo, T.Fukuda, An Intelligent Monitoring System based on Emotional Model in Sensor Networks, The 18th IEEE International

- Symposium on Robot and Human Interactive Communication, pp.346-351, 2009.
- [10] N.Kubota, H.Sotobayashi, T.Obo, Human Interaction and Behavior Understanding Based on Sensor Network with iPhone for Rehabilitation, Proc. of International Workshop on Advanced Computational Intelligence and Intelligent Informatics, 2009.
  - [11] N.Kubota, A.Yorita, Topological Environment Reconstruction in Informationally Structured Space for Pocket Robot Partners, 2009 IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA 2009), pp.165-170 Daejeon, Korea, December 15-18, 2009.
  - [12] G.Syswerda, A Study of Reproduction in Generational and Steady-State Genetic Algorithms, In Foundations of Genetic Algorithms, Morgan Kaufmann Publishers, Inc., 1991.
  - [13] N.Kubota, K.Nishida, Cooperative Perceptual Systems for Partner Robots Based on Sensor Network, International Journal of Computer Science and Network Security, Vol.6, No.11, pp.19-28, 2006.
  - [14] D.B.Fogel, "*Evolutionary Computation*", IEEE Press , 1995.
  - [15] N.Kubota, T.Shimizu, M.Abe, Joint Attention Between a Human Being and a Partner Robot Based on Computational Intelligence, Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol.11, No.10, pp. 1275-1281, 2007.
  - [16] Z.Lin, M.Zecca, S.Sessa, H.Ishii, A.Takanishi, Ultra-Miniaturized Inertial Measurement Unit WB-3 for Jaw Movement Analysis during Free Chewing, Proc. of 2010 International Symposium on Robotics and Intelligent Sensors, 2010.
  - [17] <http://www.xbox.com>
  - [18] N.Kubota, Computational Intelligence for Structured Learning of A Partner Robot Based on Imitation, Information Sciences, No.171, pp.403-429, 2005.5.
  - [19] 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.