International Journal of Humanoid Robotics

Multisensor-based Human Tracking Behaviors with Markov Chain Monte Carlo Methods

TAKAHIRO MIYASHITA

Intelligent Robotics and Communications Laboratories,
Advanced Telecommunications Research Institute International (ATR IRC),
Keihanna Science City, Kyoto, 619-0288, JAPAN
miyasita@atr.jp

MASAHIRO SHIOMI

ATR IRC
and Department of Adaptive Machine Systems, Osaka University,
Yamadaoka, 2-1, Suita, Osaka, 565-0871, JAPAN
m-shiomi@atr.jp

HIROSHI ISHIGURO

ATR IRC
and Department of Adaptive Machine Systems, Osaka University,
Yamadaoka, 2-1, Suita, Osaka, 565-0871, JAPAN
ishiguro@ams.eng.osaka-u.ac.jp

Received (Day Month Year)
Revised (Day Month Year)
Accepted (Day Month Year)

For communication robots, it is important to find a communication partner and attract his or her attention in daily environments. In this paper, we propose a method for communication robots to detect and track a human actively in order to communicate with him or her. We apply Markov chain Monte Carlo methods (MCMC) to human detection and tracking behaviors with a humanoid robot that has four types of sensors. Thus, by utilizing our method, the robot can detect and track humans with irregular motion in complicated daily environments. We verify the validity of our method by performing experiments with a humanoid-type communication robot named Robovie.

Keywords: MCMC; particle filtering; sensor fusion; human tracking; humanoid robot.

1. Introduction

One of the ultimate goals of robotics is to build robots that can move in daily environments and participate in society. In order to realize such robots, one of the most important abilities is to communicate with humans in a natural manner.

In recent years, there has been much research on communication between humans and robots [1-5]. For example, Matsusaka et al. [1,2] tried to realize a natural conversational system with a body and facial expressions and conversational function. Breazeal and Scassellati [3] constructed a robot that exploits natural human social tendencies to convey intentionality through motor actions and facial expressions. Ishiguro et al. [4, 5] estimated the communication between humans and a robot with semantic differential method. However, almost all of these efforts involve passive communication. In these works, it is the human who initially acts and speaks to the robot. Then, the robot replies by utilizing verbal or nonverbal information. In our daily communications, it is important to communicate not only passively but also actively. Active communications include, for example, finding a communication partner and attracting his or her attention.

In this paper, to realize active communications, we propose a human detection and tracking method for humanoid robots with multi sensors. There is much previous research on human detection and tracking tasks [6-8]. Almost all of those works, however, were conducted under the conditions of a static environment or motionless sensors being fixed to the environment. We cannot assume such conditions because we have to deal with mobile humanoid robots in our daily environment to realize active communication. Thus, we cannot utilize these methods for active communication. The key ideas of our method are as follows: First, we combine four types of sensor information taken from a color CCD camera, an omni-directional vision sensor, ultrasonic range sensors, and infrared motion detecting sensors, into a reliability distribution for human existence. Second, we apply Markov chain Monte Carlo methods (MCMC) [8–10] to finding a communication partner from the reliability distribution. We applied our method to a humanoid robot and verified its robustness and effectiveness.

In the next section, we introduce the humanoid robot, named "Robovie," that was used for our experiments. In Sections 3 and 4, we describe our method and experiments, respectively. We verify the validity of the method and discuss future works in Section 5.

2. Everyday Robot – Robovie –

Robovie [4] is a humanoid robot that has been developed by ATR Intelligent Robotics and Communication Laboratories. It can communicate with humans autonomously. We used this robot for the verification of our method. In this section, we introduce Robovie briefly.

As can be seen in Fig. 1, Robovie has many kinds of sensors: two color CCD cameras that can pan, tilt, and zoom individually, an omni-directional camera, a microphone, ultrasonic range sensors, tactile sensors, and pyroelectric infra-red sensors. The shape of its upper body is similar to a human's. Its face consists of the color CCD cameras as the eyes, the microphone as the ears, and a speaker as the mouth of a human. It also has four degrees of freedom (DOFs) for each arm and three DOFs for the neck. Thus, it can generate various gestures for communication, as a human does. Its lower body is a wheeled mobile base that consists of two powered wheels and one free wheel.

The robot has two computers (a. Pentium-III 900 MHz, b. Pentium-4 2.4 GHz, OS: Linux) for processing sensor information and controlling its body. Its software architecture for communication is a kind of behavior network that consists of hundreds of situated modules as nodes. The nodes describe communication behaviors depending on various situations, and transition rules of the modules are represented as arcs [4]. By utilizing this architecture, it can communicate with humans autonomously.

However, it is not enough just to be able to communicate with humans actively without any behavior for tracking and attracting humans.

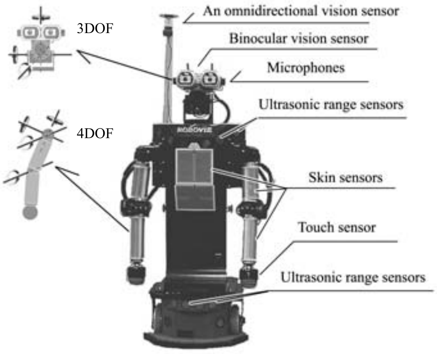


Fig. 1. Everyday robot "Robovie"

3. Multisensor-based Human Detecting and Tracking with MCMC

In this section, we propose a human detection and tracking method for humanoid robots. We begin by giving an outline of our method to detect and track humans with MCMC. Next, we explain several details of the method, especially the reliabilities and properties of Robovie's sensors.

3.1. *Outline of Proposed Method*

Figure 2 shows an outline of the proposed method. In this figure, the grey areas denote the MCMC processes and the slanted line area denotes the non-MCMC processes. The outline of the flow is as follows:

(1) The robot acquires sensor information within a limited range from each sensor. Initially, it uses the entire range of each sensor.
(2) It calculates the reliability distribution of human existence for each sensor from the sensor information (see subsection 3.2).

(3) It combines the distributions of the sensors into the reliability distribution of human existence around itself (see subsection 3.2).
(4) It predicts the next positions of humans from the distribution (see subsection 3.3).
(5) Finally, it limits the range of each sensor based on the predictions (see subsection 3.3).

By iterating these steps, the robot can reduce the sampling area of the sensors and also reduce the processing time. When the robot cannot detect a human from the predictions, it removes the limitation and restarts from the initial process.

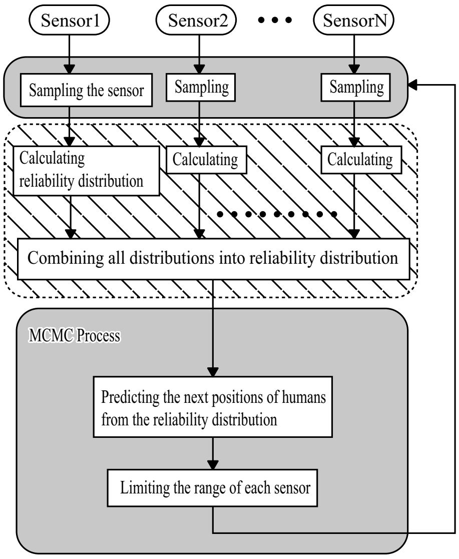

Sensor1
Sensor2
SensorN
Sampling the sensor
Sampling
Sampling
Calculating reliability distribution
Calculating
Calculating
Combining all distributions into reliability distribution
MCMC Process
Predicting the next positions of humans from the reliability distribution
Limiting the range of each sensor

Fig. 2. Outline of proposed method's data processing flow

3.2. *Reliability Distribution of each Sensor*

Human existence is observed by detecting skin color and human face features from visual information, and moving objects from range, infrared and visual information. If we do not detect such features, however, we cannot determine that there is no human presant. Therefore, if we do detect these features, we can increase the reliability of human existence, but, if we do not detect the features, we cannot increase the reliability of human nonexistence. Thus, we define the reliability of human existence around the robot, $R(x)$, as

$$R(x) = 1 - H(x) \tag{1}$$

$$\text{and } H(x) = -P(x)\log_2(P(x)) -(1-P(x))\log_2(1-P(x)) \ , \tag{2}$$

where x, $H(x)$ and $P(x)$ denote a direction around the robot, an information entropy, and a probability of human existence in the effective area of each sensor, respectively (see Fig. 3). In our case, the range of $P(x)$ is from 0.5 to 1.0.

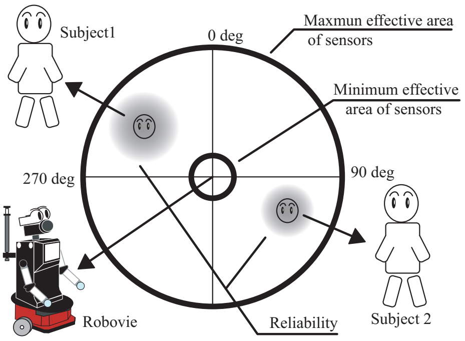

Subject1
0 deg
Maxmun effective area of sensors
Minimum effective area of sensors
270 deg
90 deg
Robovie
Reliability
Subject 2

Fig. 3. Reliability and effective area of sensors

We calculate the reliability from four kinds of sensors: an omni-directional camera, conventional CCD cameras, ring-type ultrasonic range sensors, and pyroelectric infra-red sensors, attached on the robot as shown in Fig. 1.

We define the reliability distributions of human existence for each sensor by conducting preliminary experiments. In the experiments, we record time series data of each sensor output and subject's position with a 3-D motion capture system when he or she walks around the robot. We assume that the distributions are able to approximate normal density distribution, and calculate parameters of them from the recorded data. Details of them are described below.

Omni-directional camera

The omni-directional camera can obtain omni-directional visual information as shown in Fig. 4. We calculate the reliability distribution for this information by the features of a moving object taken from the camera. The features of the moving object are its region size and position calculated by the interframe difference of RGB values of the omni-directional visual information. We calculate the reliability distribution for the omni-directional camera, $f_o(x)$, as

$$f_o(x) \sim N(\mu_o, {\sigma_o}^2) \times \alpha_o \ , \tag{3}$$

$$\alpha_o = 0.3 \times M_{size} \ , \tag{4}$$

$$\mu_o = M_\theta \ , \tag{5}$$

$$\text{and} \ \ {\sigma_o}^2 = M_d \ , \tag{6}$$

where $N(\mu_o, {\sigma_o}^2)$, M_{size}, M_d and M_θ denote the normal distribution with a mean μ_o and a variance ${\sigma_o}^2$, the region size of the moving object, the distance between the robot and the object, and the direction to it with respect to coordinates fixed to the robot, respectively. The variables and the distribution are shown in Fig. 5.

We do not use this sensor while the robot is moving because the features of moving objects consist of both the object motion and the robot motion, which are difficult to distinguish between. Thus, in (3), while the robot is moving, α_o is 0.

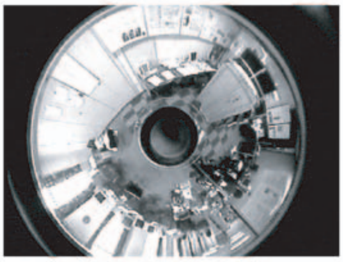


Fig. 4. Omni-directional visual information

Conventional CCD camera

We calculate the reliability distribution for the conventional CCD camera by the sizes and positions of a human face, skin and clothes taken from it. The size and position of the face can be calculated by Gabor filter banks [11]. We also calculate the region size and the position of the skin by using the HSV color information. The size and position of the clothes are calculated by tracking a color region beneath the face.

We define the reliability distributions calculated by the face, the skin and the clothes information as $f_f(x)$, $f_s(x)$ and $f_c(x)$, respectively. These distributions conform to a normal distribution such as (3). The means and variances for them are

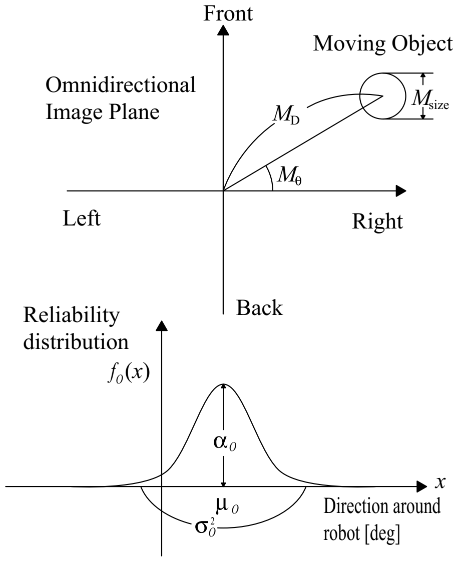


Fig. 5. Reliability based on features of moving object taken from omni-directional camera

as follows.

$$\alpha_f = 0.5 \times F_{size}, \tag{7}$$

$$\alpha_s = 0.3 \times S_{size}, \tag{8}$$

$$\alpha_c = 0.2 \times C_{size}, \tag{9}$$

$$\mu_f = F_\theta,\ \mu_s = S_\theta,\ \mu_c = C_\theta, \tag{10}$$

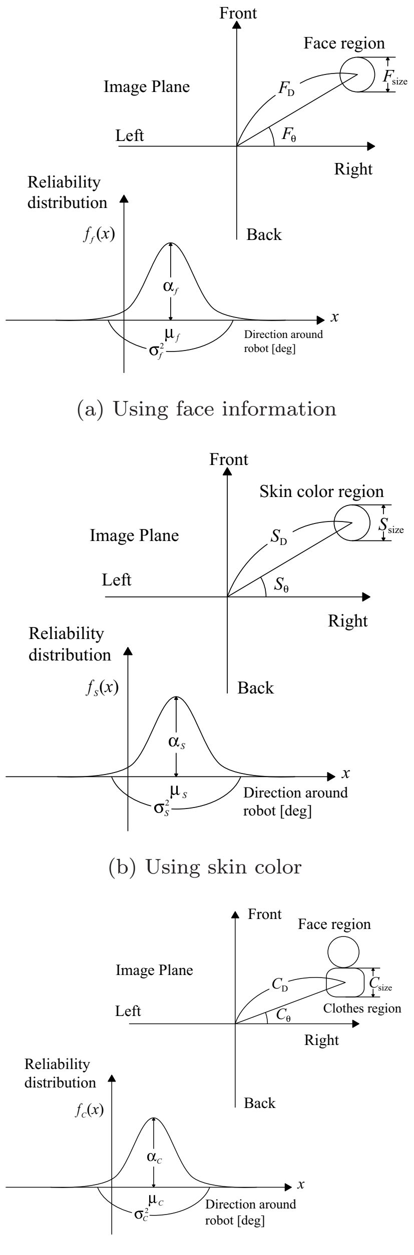
$$\sigma_f{}^2 = F_d, \sigma_s{}^2 = S_d, \text{and } \sigma_c{}^2 = C_d, \tag{11}$$

where F, S and C denote the face, skin and clothes, respectively. The subscripts $_{size}$, $_d$ and $_\theta$ denote the size of a region, the distance between the robot and the center of the region, and the region's direction. The variables and the distributions are shown in Fig. 6.

The robot utilizes only this sensor while it is moving. It, however, can calculate three reliability distributions, f_f, f_s, and f_c, based on six kinds of information that are the region sizes and the positions of the face, the skin color, and the clothing color. Thus it is still able to track a human robustly while it is moving.

Ultrasonic range sensors

The robot can obtain omni-directional range information from the ring-type ultrasonic range sensors. We calculate its reliability distribution from changes in the range information, which describe the existence of moving objects.



(a) Using face information

(b) Using skin color

(c) Using clothes color

Fig. 6. Reliability based on visual information taken from conventional CCD camera

The distribution also conforms to a normal distribution. Its mean and the variance are

$$\alpha_u = \frac{0.2 \times D_{min}}{D}, \tag{12}$$

$$\mu_u = U_\theta, \tag{13}$$

$$\text{and } \ {\sigma_u}^2 = D, \tag{14}$$

where D, D_{min}, and U_θ denote the distance between the object and the robot, the minimum distance that can be measured by the sensor, and its direction. The variables and the distribution are shown in Fig. 7. While the robot is moving, α_u is 0.

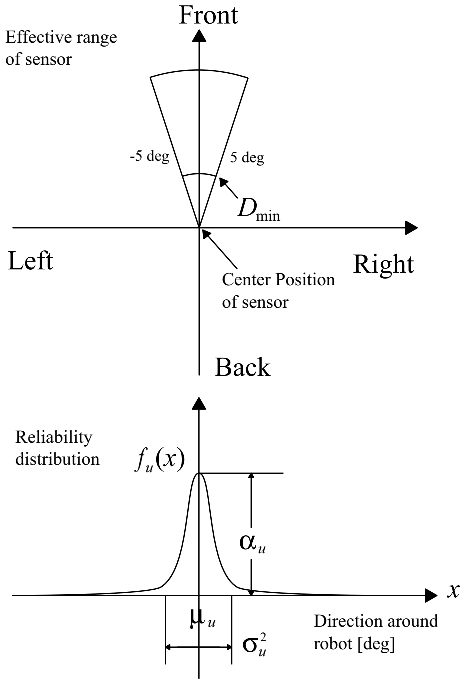


Fig. 7. Reliability based on change in range information taken from an ultrasonic range sensor

Pyroelectric infrared sensors

The pyroelectric infra-red sensors are used to detect the moving objects by measuring changes in infra-red readings. The sensors' output is either 1 (detected) or 0 (not detected).

We define these sensors' reliability distribution, $f_p(x)$, as

$$f_p(x) \sim \begin{cases} 0.3 \times \alpha_p \ (-30 \leq x_p \leq 30), \\ 0 \qquad\qquad (\text{otherwise}), \end{cases} \tag{15}$$

where α_p is the weight value 0 while the robot is moving and 1 while it is not moving. The distribution is shown in Fig. 8.

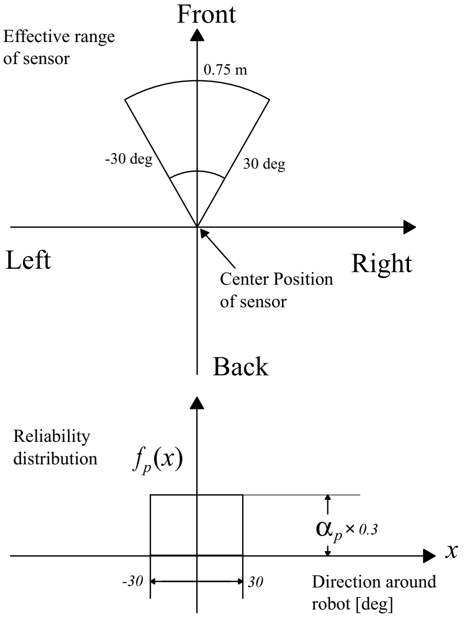


Fig. 8. Reliability based on change in IR information taken from pyroelectric IR sensor

After calculating all reliability distributions for all sensors, we combine the distributions into a reliability distribution of human existence around the robot by the following equation.

$$F(x) = \max\left(f_o(x), f_f(x), f_s(x), f_c(x), f_u(x), f_p(x)\right). \tag{16}$$

3.3. *MCMC process*

We predict the next positions of humans and limit the ranges of each sensor by utilizing particle filtering method [9, 10] that is based on MCMC. A process of the prediction and the limitation is as follows:

(1) Acquire a weighted sample-set $\{(s_{t-1}^{(n)}, \pi_{t-1}^{(n)}), n = 1, \ldots, N\}$ at time-step $t-1$ from a previous iteration. $s_{t-1}^{(n)}$ and $\pi_{t-1}^{(n)}$ denotes the n-th sampling point (a

direction with respect to the robot coordinate) and its weight at $t-1$, respectively. The sample-set $\{(s_{t-1}^{(n)}, \pi_{t-1}^{(n)})\}$ represents approximately a conditional state-density $R(X_{t-1}|Z_{t-1})$ where X_{t-1} and Z_{t-1} denotes a state and an observation feature, respectively. $R(X_{t-1}|Z_{t-1})$ denotes a normalized reliability distribution of human existence at time-step $t-1$. We decide a number of sampling points N as 1000 in the experiments.

(2) Calculate cummulative weights $c_{t-1}^{(n)}$ as

$$c_{t-1}^{(0)} = 0, \quad \text{and} \quad c_{t-1}^{(n)} = c_{t-1}^{(n-1)} + \pi_{t-1}^{(n)} \ . \tag{17}$$

Select a sample set $\{s_t'^{(n)}\}$ by iterating following steps [9].

(a) generate a random number $r \in [0,1]$, uniformly distributed.
(b) find, by binary subdivision, the smallest j for which $c_{t-1}^{(j)} \geq r$
(c) set $s_t'^{(n)} = s_{t-1}^{(j)}$

(3) Assume that humans walk randomly with the speed less than 5 [km/h], the distance between the robot and human is approximately 1 [m], and the sampling rate of the sensors is 100 [ms]. Then we define the dynamical model which denotes predicted position of human at the next time-step as

$$R(X_t|X_{t-1}) \sim N(X_{t-1}, (\Delta/6)^2) \ , \tag{18}$$

where $N(X_{t-1}, \left(\frac{\Delta}{6}\right)^2)$ denotes the normal distribution with a mean X_{t-1} and a variance $\left(\frac{\Delta}{6}\right)^2$. X_{t-1} denotes a state variable at time-step $t-1$. We decide Δ as 15 [deg] in the preliminary experiment. This distribution indicates that the predicted position of human is between -15 and 15[deg] with a probability of 0.9974.

(4) Predict a sample-set $\{s_t^{(n)}, n = 1, \ldots N\}$ by sampling from $R(X_t|X_{t-1} = s_t'^{(n)})$.

(5) Acquire an observation model $R(Z|X)$ by normalizing the reliability distribution $F(x)$ calculated by (16) as

$$R(Z|X) = \frac{F(x)}{\sum_{i=0}^{360} F(i)} \ , \tag{19}$$

where Z denotes observation features.

(6) Measure the observation features Z_t at the new sample position $s_t^{(n)}$. Then, calculate the weight $\pi_t^{(n)}$ as

$$\pi_t^{(n)} = \frac{R(Z_t|X_t = s_t^{(n)})}{\sum_{n=0}^{N} R(Z_t|X_t = s_t^{(n)})} \tag{20}$$

By iterating the above process, we can reduce the sampling points and the range of each sensor by utilizing the reliability from the previous time-step. An outline of the process is shown in Fig. 9

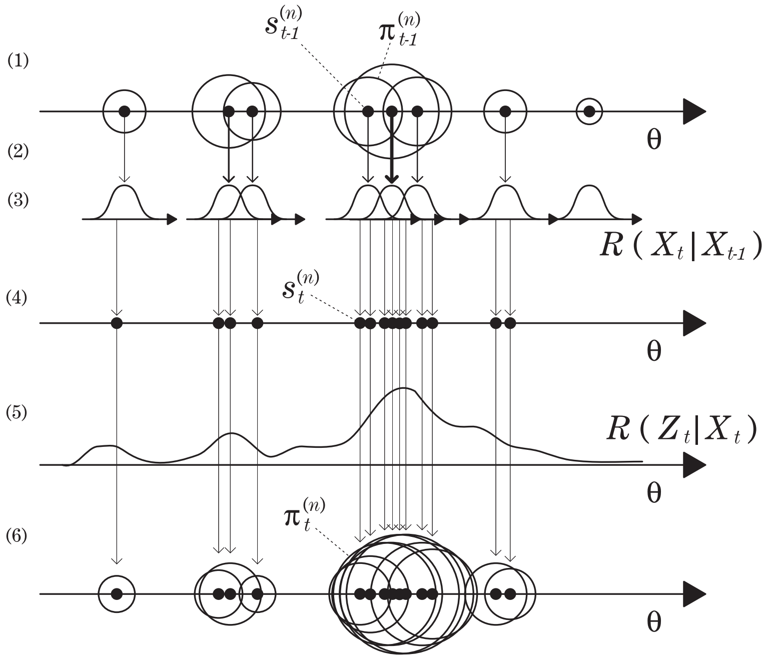


Fig. 9. Outline of the MCMC process

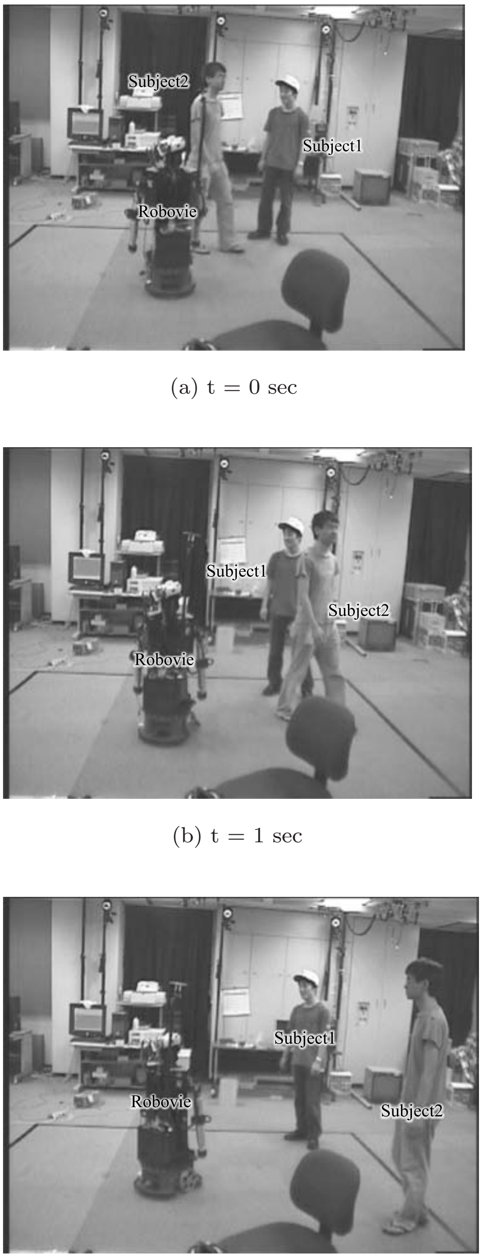
4. Experiments

We applied our method to the humanoid robot Robovie and conducted experiments with the robot to verify the method's validity. The robot and two subjects were situated in the experimental room as shown in Fig. 10. The task of the robot was to detect and track a human. We compared the proposed method with a method that always utilizes the entire range of each sensor. The trial period of the experiments is 2 minutes. The number of trials is 5 for each method.

We use a 3-D motion capture system, VICON, to measure positions and postures of the subjects and the robot at a 120 Hz sampling rate.

4.1. *Results of Experiments*

Figure 11 shows an example of the loci of each subject and the robot taken from the motion capture system under the conditions of the proposed method. In the figure, the circles indicate the positions of them measured at intervals of 20 seconds. Figure 12 shows the directions of each subject with respect to the coordinate fixed to the robot under the conditions of the proposed method. As shown in these figures, the robot detects and tracks Subject1. Figures 13 and 14 also show the same information while utilizing the entire range of each sensor. Under this method, the robot has difficulty tracking the subject continuously because the processing time



(a) t = 0 sec

(b) t = 1 sec

(c) t = 2 sec

Fig. 10. Subjects and robot in experimental room

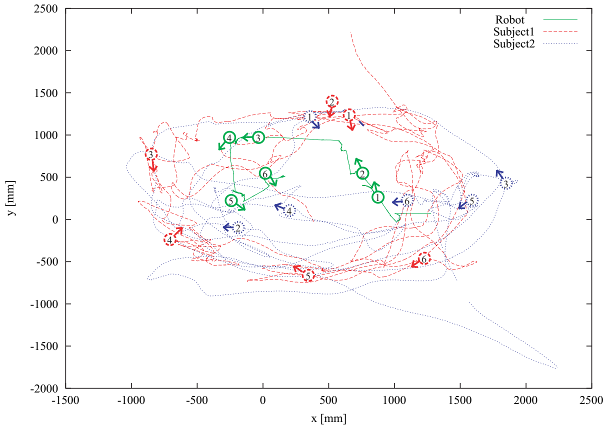


Fig. 11. Results of human detecting and tracking task with proposed method (MCMC)

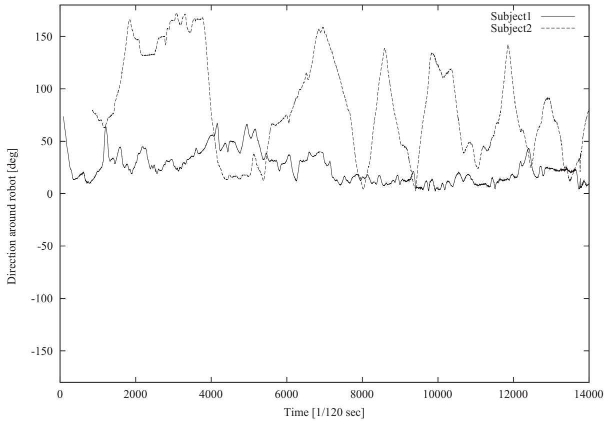


Fig. 12. Results of human detecting and tracking task with proposed method (MCMC)

becomes three times longer than for the proposed method shown in Table. 1.

Table 1. Average processing times

	Proposed method	Using all samplings
Average [msec]	30.8	94.7

Figure 15 shows the change in the range limitation of each sensor under the proposed method. In this graph, the X-axis denotes the direction [deg] around the

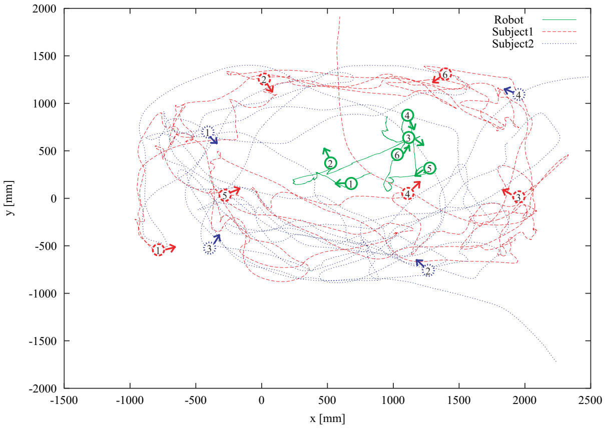


Fig. 13. Results of human detecting and tracking task with proposed method (MCMC)

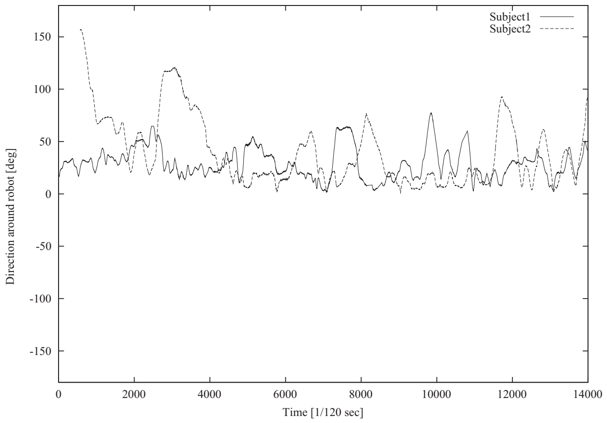


Fig. 14. Results of human detecting and tracking task with all-sampling-method

robot and the Y-axis denotes the number of frames [× 33 msec]. The black area denotes the limited range of each sensor. As shown in the figure, the robot can limit the range of each sensor. When it loses a human within the limited range, it uses the entire range of each sensor.

5. Conclusions and Future work

In this paper, we proposed a method for a robot to detect and track humans by calculating the reliability of human existence around the robot with MCMC. In the experiments, we indicated the validity of the proposed method by comparing it with a method that always utilizes the entire range of each sensor. The robot was able

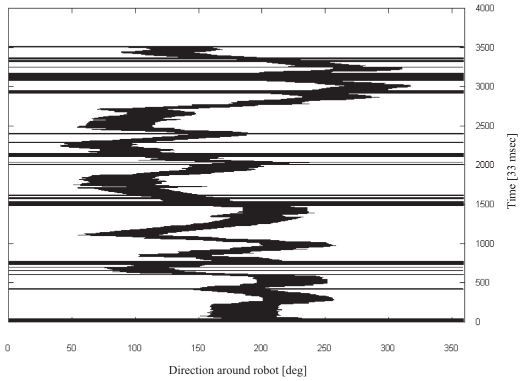


Fig. 15. Change in sampling area with MCMC

to continuously track the first subject detected by using the proposed method.

In the method, the robot combines the reliability distributions of all sensors into one distribution. However, if the robot has the observation model of each sensor, this process is not necessary and it can predict human existence directly from the reliability distribution of each sensor with MCMC. In future work, we will try to apply the observation models, eliminate the combining process, and verify the validity of the method.

Acknowledgement

This research was supported in part by the National Institute of Information and Communications Technology of Japan.

References

1. Yosuke MATSUSAKA, Shinya FUJIE and Tetsunori KOBAYASHI: “Modelling of Conversational Strategy for the Robot Participating in the Group Conversation,” in Proc. of ISCA-EUROSPEECH, 2001.
2. Tsuyoshi Tojo, Yosuke Matsusaka, Tomotada Ishii, and Tetsunori Kobayashi: “A Conversational Robot Utilizing Facial and Body Expressions,” in Proc. of IEEE International Conference on System, Man and Cybernetics (SMC2000), pp. 858–863, 2000.
3. Cynthia Breazeal and Brian Scassellati: “How to build robots that make friends and influence people,” in Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS’99), 1999.
4. Takayuki Kanda, Hiroshi Ishiguro, Michita Imai, Tetsuo Ono, and Kenji Mase,: “A constructive approach for developing interactive humanoid robots,” in Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2002), pp. 1265–1270, 2002.
5. Hiroshi Ishiguro, Tetsuo Ono, Michita Imai, Takeshi Maeda, Takayuki Kanda, and Ryohei Nakatu: “Robovie: an interactive humanoid robot,” in International Journal of Industrial Robot Vol. 28, No. 6, pp. 498–503, 2001.

6. S. Ioffe and D. Forsyth: “Human Tracking with Mixtures of Trees,” in Proc. of IEEE International Conference on Computer Vision (ICCV 2001), 2001.
7. T. Wilhelm, H. Boehme, and H. Gross,: “Sensor Fusion for Vision and Sonar Based People Tracking on a Mobile Service Robot,” in Proc. of International Workshop on Dynamic Perception 2002, pp. 315–320, 2002.
8. Kiam Choo and David J. Fleet: “People Tracking Using Hybrid Monte Carlo Filtering,” in Proc. of IEEE International Conference on Computer Vision (ICCV 2001), Vol. II, pp. 321–328, 2001.
9. Michael Isard and Andrew Blake: “A smoothing filter for Condensation,” in Proc. of 5th European Conference on Computer Vision, Vol. 1, pp. 767–781, 1998.
10. Arnaud Doucet, Nand de Freitas and Neil Gordon (eds.): “Sequential Monte Carlo Methods in Practice,” (Springer-Verlag, New York, 2001), pp. 339–357.
11. I. Fasel, Bartlett and J. Movellan: “A comparison of methods for automatic detection of facial landmarks,” in Proc. of IEEE International Conference on Automatic Face and Gesture Recognition, 2002.

Takahiro Miyashita received his B.S., M.S., and Ph.D. degree in engineering for computer-controlled machinery from Osaka University, Japan in 1993, 1995, and 2002, respectively. From 1998 to 2000, he was Research Fellow of the Japan Society for the Promotion of Science. From April to September 2000, he was Researcher of Symbiotic Intelligent Group at ERATO Kitano Symbiotic Systems Project of Japan Science and Technology Agency. From October 2000 to July 2002, he was Assistant Professor at Wakayama University. Now, he is a senior research scientist at Intelligent Robotics and Communication Laboratories (IRC), Advanced Telecommunications Research Institute International (ATR). He is also a member of the Robotics Society of Japan and the Japanese Society for Artificial Intelligence.

Takahiro Miyashita is the author of over 40 technical publications, proceedings, editorials and books. His research interests include computer vision, vision based robots, control for multi-degrees of freedom robots, and humanoid robots in daily environment. He is a member o the Robotics Society of Japan and the

Japanese Society for Artificial Intelligence.

Masahiro Shiomi received his M.S. Graduate School of Engineering from Osaka University, Japan in 2004, respectively. Now, he is a Intern researcher at Intelligent Robotics and Communication Laboratories (IRC), Advanced Telecommunications Research Institute International (ATR) and Ph.D student of Osaka University.

Hiroshi Ishiguro received D. Eng. degree from Osaka University, Japan, in 1991. In 1991, he started working as a research assistant of Department of Electrical Engineering and Computer Science, Yamanashi University, Japan. Then, he moved to Department of Systems Engineering, Osaka University, Japan, as a research assistant in 1992. In 1994, he was an associate professor of Department of Information Science, Kyoto University, Japan, and started research of distributed vision using omnidirectional cameras. From 1998 to 1999, he worked in Department of Electrical and Computer Engineering, University of California, San Diego, USA, as a visiting scholar. From 1999, he is a visiting researcher in ATR Media Information Science Laboratories and he has developed interactive humanoid robots, Robovie. In 2000, he moved to Department of Computer and Communication Sciences, Wakayama University, Japan, as an associate professor and then he became a professor in 2001. Now he is a professor of Department of adaptive machine systems, Osaka University, Japan. He is also a group leader of IRC, ATR.