Paper: Rb*--**-******

Multiple Persons Tracking by Multiple Cameras and Laser Range Scanners in Indoor Environment

Hiroshi Noguchi*, **Taketoshi Mori***, **Takashi Matsumoto****, **Masamichi Shimosaka*** and **Tomomasa Sato***

*The University of Tokyo, 7-3-1 Hongo,Bunkyo-ku, Tokyo 113-8656 Japan
E-mail:{noguchi,tmori,simosaka,tomo}@ics.t.u-tokyo.ac.jp
**Mitsubishi Research Institute, Inc., 3-6, Otemachi 2-chome Chiyoda-ku, Tokyo, 100-8141, Japan
E-mail:t.matsu@mri.co.jp

[Received 00/00/00; accepted 00/00/00]

In this paper, we propose a method for multiple persons tracking using multiple cameras and laser range scanners. Our method estimates 3D positions of human body and head, and labels their identities. The method is composed of multiple particle filters, and each particle filter tracks each person correctly by integrating information from laser range scanners and the target-specific information from multiple cameras. Integration of these two types of sensors enables complement of each weak point and the correct tracking of the target. Moreover, we develop a particle filter framework that tracks the human head by using the estimated human body position simultaneously. Our experimental results demonstrate the effectiveness and robustness of the method in multiple persons tracking with multiple scanners and cameras.

Keywords: Laser Range Scanner, Particle Filter, Multiple Person Tracking

1. Introduction

Human monitoring in home environment is one of important topics for intelligent human support. Especially, measurement of people's postures and positions in a room is essential to support with home robots and to recognize pattern of human daily behavior. In order to monitor human, wearable sensors such as RFID tags, ultra sonic devices, and infra-red devices realize easy and robust measurement of occupant's position. However, wearing special sensor devices constricts human daily behaviors and is troublesome for the habitants. Monitoring without wearable devices is desirable for long-term measurement. We focus on human behavior monitoring with room-equipped sensors.

Measurement of human posture and position with multiple cameras is one of the most popular topics among monitoring with sensors in environments. Kobayashi[9] developed 3d-tracking technique of heads from multiple cameras via classification based on Ada Boost. Matsumoto[12] also developed head tracking from multiple cameras by using particle filter. Kim[8] realized people tracking by segmentation of occluded people's areas with integration of images from multiple cameras on ground plane. The camera-based tracking is weak at occlusion occurred from multiple people and such objects in a room as furniture and appliances. Light condition of the room frequently changes due to usage of lighting apparatus and light condition of room outside. Generally speaking, the modification of light condition decreases performance of camera-based tracking. While camera provides rich information for human identification, occlusion and lighting problems usually cause miss of tracking in short time.

Floor sensors[11, 14, 16, 17] are also utilized for human position monitoring. The floor sensors detect human position from output of distributed pressure sensors or switch sensors. Since the occupants always touch the floor plane with their legs, the floor sensors do not miss the people's position. However, the floor sensors contain several practical problems. The floor sensors are expensive. Introduction of the sensors into the home environment is difficult. The measurement area is limited. The floor sensors have a disadvantage on identification and occlusion, which is overlapping of detected area. Thus, once tracking system with a floor sensor misses an occupant, it is difficult for the system to decide whether detected area is the occupant's area or not.

Currently, laser range scanner is often utilized for human position sensing because of the scanner's price-reduction and easy introduction into environment. Zhao[20] and Cui[2] develop multiple people tracking technique from the scanners at leg height. Fod[5] also realizes tracking from the scanners at hip height. Glas[6] proposed the method for measurement of human position and direction from the scanners at leg height. The scanner can measure accurate distance. The scanner is suitable for accurate position sensing. The occlusion problem occurs at the scanners. While the cameras are difficult to separate area overlapping of people, the outer shape of human from the scanner data is more easily separable than the camera. However, since it is difficult to recognize the person from his/her outer shape by the scanners, the recovery of tracking from missing of the people is as difficult as the floor sensors.

People tracking from only single kind of sensor has merits and demerits. Some researchers combine two kinds of sensors in order to measure human position. The ways how they merge the sensors are combination of camera and wireless devices[4], mixture of cameras and floor sensors[13] and blend of floor sensors and embedded RFID[15]. These approaches are effective for robust people tracking, because the sensor covers the other sensor's drawback. We also adopt the same approach. As we mentioned above, the wearable sensors are unsuitable because of their constriction for users. The floor sensors have high cost and are troublesome for equipment. RFID has the same problem. The RFID can identify the person easily but it is useless for measurement of position. The floor sensor and RFID are also inappropriate. We utilize cameras and laser range scanners for people tracking. While the scanners cover the occlusion problem of the camera, the cameras compensate for the identification problem of the scanner.

Cui[1] develops multiple people tracking from single camera and scanner based on Bayesian filtering technique. Their approach's target is mainly the combination of single camera and single scanner. Kurazume[10] also develops multiple people tracking with multiple cameras and scanners. Their methods are robust for crowd of people. However, the target area is open space without occlusion objects such as tables, sofas, and chairs. Their method only tracks the body center of position. We propose the method for multiple person tracking using multiple cameras and scanners. Our method can estimate not only the 3D positions of bodies but also 3D positions of heads in the environment including low occlusion objects such as tables and chairs. Measurement of heads and bodies positions contributes to the action estimation such as sitting and standing, which is important in human support. The method can also label persons' identities. In the paper, the identification is not global specification among people. The identification that means the same labels is correctly allocated to the same person in tracking. This identification is essential for human fitting support.

In this paper, we suppose tracking area is one-room-size area. The camera and scanners are deployed at 4 corners. The cameras are equipped with the ceiling of the room for avoidance of occlusion. Usually speaking, the scanners are arranged at ankle height because of no dependence on body height and avoidance of occlusion among people. However, in home environment, there are many objects at ankle height. In order to resolve the problem, we deploy the scanners at hip height. Thus, our tracking targets are the persons whose hips are observed with the scanners. In multi-sensor fusion research, synchronization of sensor data collection is an important topic. In our research, since speed of human walking is slow in home environment, the high sampling rate of sensors is not needed. In home environment, communication hardware such as wired LAN is stable, we suppose the synchronization is realized between the cameras and the scanners.

2. Multiple Persons Tracking

We utilize particle filtering technique[7] for people tracking. The particle filter is a kind of sequential filtering. The filter regards tracking target as probability density. The probability density is represented as a finite set of samples that include states and likelihoods. The filter recursively updates the probability using propagation based on state transition probability and evaluation of likelihoods through observations. The filter has an advantage on sudden movement of tracking persons and noise of sensors because of probabilistic mechanism. In typical camera-based tracking by particle filter, a sample includes 3D position as a state. The samples' likelihoods is evaluated by matching of projected 3D position in the samples with 2D images of cameras. The final position of body is calculated as expectation of samples' 3D positions and likelihoods.

Since tracking target is represented as non-parametric probability density distribution in the particle filter, it is difficult to discriminate probability density distribution about each tracking target from the distributions of multiple targets. The filter has high performance of position estimation at single person tracking because number of the samples is large enough to represent the distribution. On the contrary, in multiple people tracking, sometimes one tracking target gathers many samples and the other targets contain a few samples. The targets that include a few samples tend to be missed because small number of samples cannot represent the distribution well. Against this problem, Vermaark[18] proved that the probability density distribution including all tracking target is represented as a sum of the weighted distributions about every targets. Base on the idea, they propose mixture particle filter (MPF), which tracks multiple targets simultaneously. While the MPF resolves the problem on miss of target people, the filter is insufficient for multiple persons tracking during labeling ID. IDs counterchange or lose when two persons move across each other. The heterogeneous arrangement of samples among the targets still causes the ID missing problems.

We propose the similar filter to MPF. The filter consists of the single-target particle filters about target people. The single-target filter tracks the same person with samples in the filter. The single-target filter also contains target-specific information for labeling the same ID. This configuration of the filter realizes robust tracking and labeling in multiple persons tracking. Our method for multiple persons tracking is illustrated in Fig. 1.

The target-specific filter contains the information for person identification. In our method, each filter updates contained samples with target-specific information. This mechanism helps robust labeling of IDs. When our method detects a person, the method generates new ID for the detected person and obtains information about the tracking target. As the target-specific information, we adopt color histogram of each person, which represents characteristic of wearing cloth. This idea is popular approach. This approach is utilized in [3]. Each camera

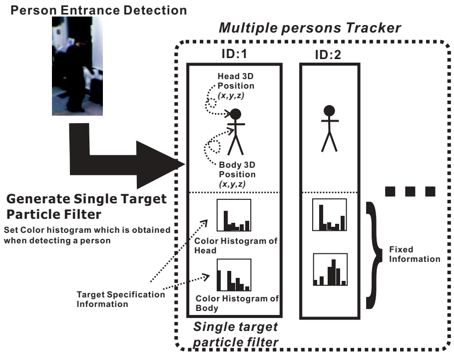

Fig. 1. Outline of multiple persons tracking

captures color histogram in background subtracted image of rectangle area defined in advance. The capturing scene is shown in Fig. 2. In the figure, the red region is a rectangle area defined in advance. The upper area is region for head histogram and the lower area is region for body histogram. We suppose a person for tracking is adult human whose body height is from 150 [cm] to 180 [cm] and only single person pass the detection area at entrance detection. Since the posture in entrance is usually standing posture and capturing area is large enough to capture whole body, this simple approach is sufficient for capturing body and head histograms.

Fig. 2. Initialization of color histogram

3. Multiple Persons Tracking using Laser Range Scanners and Cameras

3.1. Procedures of Head and Body Tracking

In tracking, the target-specific filter contains a particle filter for body tracking and a filter for head tracking. Each filter includes color histograms about body and head for person identification.

The body tracking filter calculates the position of body by two-step estimation. Firstly, the filter evaluates samples with data of the laser range scanners because the scanner data is more accurate than the camera data. After evaluation, the filter updates the evaluated samples. Secondly, the filter evaluates updated samples with data of the cameras in terms of position. The filter also evaluates updated samples with containing color histogram. After evaluation, the filter updates the particles. The head tracking filter estimates head position with calculated body position using the body tracking filter and the camera data. The tracking procedure is shown as following. The procedure is illustrated in Fig. 3.

A sample for body tracking $\boldsymbol{s}_b$ contains the 3D position (x, y, z) as a state. A sample for head tracking $\boldsymbol{s}_h$ also includes the 3D position.

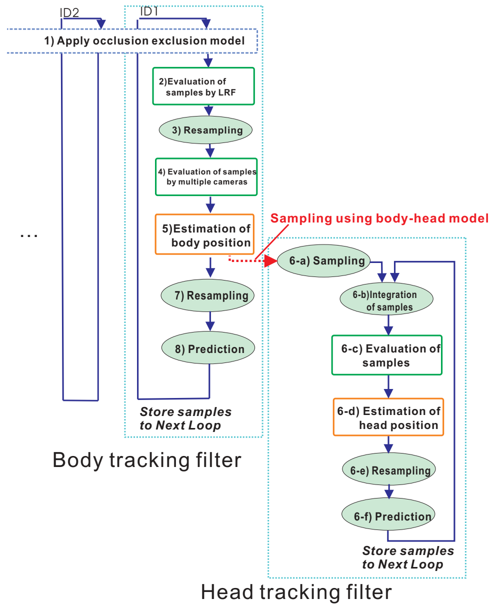

Fig. 3. Flow of body and head tracking

1 Apply exclusion model for occlusion[13] to body-tracking samples $\boldsymbol{s}_{b,t}$ in the body tracking filters. The weights of particles in overlap area become zero.

2 Evaluate weight $\pi_{bl,t}$ of sample $\boldsymbol{s}_{b,t}$ by the scanners data.

3 Select new samples $\boldsymbol{s}'_{b,t}$ in proportion to weight $\pi_{bl,t}$ corresponding to samples $\boldsymbol{s}_{b,t}$ and add Gaussian distribution.

4 Evaluate weight $\pi_{bc,t}$ corresponding to sample $\boldsymbol{s}'_{b,t}$ by the cameras data.

5 Estimate center 3D-position of body $\boldsymbol{p}_{b,t}$.

6 Measure head position.

  a Generate samples $\boldsymbol{s}_{bh,t}$ from center position of body and model between body and head $\boldsymbol{h}_{t-1}$.

  b Generate samples $\boldsymbol{s}_{h,t}$ by integration of generated samples from 6-a) and updated samples $\boldsymbol{s}_{rh,t}$ at time $t-1$.

  c Evaluate weight $\pi_{h,t}$ of samples $\boldsymbol{s}_{h,t}$ by the cameras data.

  d Estimate head position $p_{h,t}$ from samples $\boldsymbol{s}_{h,t}$.

  e Select samples $\boldsymbol{s}'_{rh,t+1}$ in proportion to their weights $\pi_{h,t}$.

f Propagate samples $\boldsymbol{s}'_{rh,t+1}$ with state transition probability and generate new samples $\boldsymbol{s}_{rh,t+1}$.

7 Select samples $\boldsymbol{s}'_{b,t+1}$ in proportion to their weights $\pi_{b,t}$ from the samples $\boldsymbol{s}_{b,t}$.

8 Propagate samples $\boldsymbol{s}'_{b,t+1}$ with state transition probability and generate new samples $\boldsymbol{s}_{b,t+1}$

3.2. Body Tracking by Laser Range Scanners and Cameras

The 3D-position of body center is tracked by two-step estimation in the filter. In traditional approach, the position of body center is estimated during single step, because the sensor measurements are independent. For example, multiplying by weighted value evaluated from each sensor is total evaluation. The estimation in single step has the merit that mechanism is easy enough to fuse the multiple kinds of sensors. However, it is difficult to design and to adjust parameters (in the example case, weights in evaluation at every sensor) in sensor data fusion. Sometimes evaluation of a sensor drastically exceeds evaluation of another sensor. This exceeding of single-sensor evaluation decreases benefit of the sensor fusion. In the above example, since no weight is zero, the total evaluation value that should be small indeed becomes large, when the certain sensor evaluation is large by the noise. In our problem of sensor fusion, because measurement by laser range scanners is accurate, the scanners can absolutely eliminate wrong evaluation. We do not adopt the approach that multiplication of weighted evaluation. We separate the evaluation process into two steps for benefit of the scanner measurement. In first step, the filter evaluates the samples by the laser range scanners, because the laser range scanners can measure more accurate distance than the cameras and can easily separate people's areas. This procedure means that the filter limits position candidates into narrow area. The samples are selected based on evaluated their weights. In second step, the filter evaluates the selected samples by the camera images. In this step, ID that the filter contains is evaluated with the camera images and included color histogram. Finally, the samples are selected in proportion to their weights. The 3D position of body center is estimated with selected samples.

We explain about our state transition model and observation model of the sensors.

3.2.1. Transition Model for Body Center Tracking

We assume uniform straight motion of a target 3D position between two successive image frames. Transition model $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1} = \boldsymbol{s}'_{t-1})$ is denoted as below.

$$\boldsymbol{s}_t = \boldsymbol{s}'_{t-1} + \tau \boldsymbol{v}_{t-1} + \boldsymbol{\omega} \qquad (1)$$

Where τ is the time interval between frames, $\boldsymbol{v}_t$ is the previous velocity of the target, $\boldsymbol{\omega}$ is a system noise added to $\boldsymbol{s}'_{t-1}$, and $\boldsymbol{s}_t$ is the estimated target's position at time t. System noise $\boldsymbol{\omega}$ contributes to improvement of the robustness against sudden motion and the accuracy of the estimation.

In the general particle filter, large diffusion of samples decreases approximation accuracy of the probability density distribution with the samples. Incorrect approximation of the distribution causes wrong estimation of the position and mistake of tracking in multiple people tracking. In order to improve diffusion of samples in sudden movement and remaining still, we introduce adaptive control of diffusion factor $\boldsymbol{\omega}$.

In the adaptive control, the diffusion regions change in accordance with the speed of the tracking target. When the target moves fast, the diffusion region becomes wide. As for the implementation of the control, $\boldsymbol{\omega}$ is represented as a white noise limited in the range $-\boldsymbol{\delta}_t \leq \boldsymbol{\omega} \leq \boldsymbol{\delta}_t$. The 3-dimensional vector $\boldsymbol{\delta}_t$ is controlled corresponding to the speed of the target. $\boldsymbol{\delta}_t$ is denoted below. $\boldsymbol{\delta}_t$ increases in proportion to absolute value of the velocity of the target.

$$\boldsymbol{\delta}_t = \boldsymbol{\Gamma} \bar{\boldsymbol{v}}_t + \boldsymbol{\sigma} \qquad (2)$$

Where $\bar{\boldsymbol{v}}_t$ is 3-dimensional vector including absolute value in each axis, $\boldsymbol{\Gamma}$ is fixed 3-dimensional diagonal matrix and $\boldsymbol{\sigma}$ is fixed 3-dimensional vector. Control of system noise $\boldsymbol{\omega}$ improves position estimation and reduces mistake of tracking in sudden movement.

3.2.2. Observation Model of Laser Range Scanners

In the first step of the filter, the samples are evaluated by the scanners. The 3D positions of samples are projected into horizontal plane. The laser range scanner captures the single scan in absence of the people as background information in advance. The scanner measures distance d_d in i-th direction in single scan. The weight $\pi_{bl,t}$ of the sample is evaluated by the following equation with the parameter measured distance d_d, background distance d_b, and distance d_s between projected 2D position of sample and the scanner position (Fig. 4).

$$\pi_{bl,t} = \prod_i \begin{cases} \alpha & \text{if } d_s \leq d_d \\ \beta & \text{if } d_s \geq d_d \cap d_s \geq d_b \ (\alpha < \beta < \gamma) \\ \gamma & \text{if } d_s \geq d_d \cap d_s \leq d_b \end{cases} \qquad (3)$$

Since this evaluation equation excludes conversion from polar-coordinates to 2D-coordinates, the weight is calculated in short time. The calculation cost depends on number of scanners and number of samples.

3.2.3. Observation Model of Camera

In second step, the samples are evaluated by room spatial information, matching with background subtracted image of each camera and matching with color histogram. The filter evaluates samples hierarchically from camera images in order to reduce calculation time. The evaluation step is shown as following.

1 Weights of samples outside of the room size become zero.

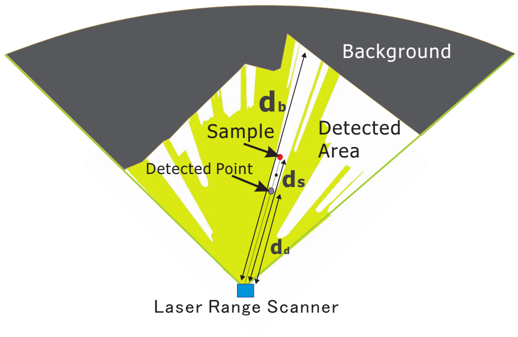

Fig. 4. Parameters for observation model of laser range scanner

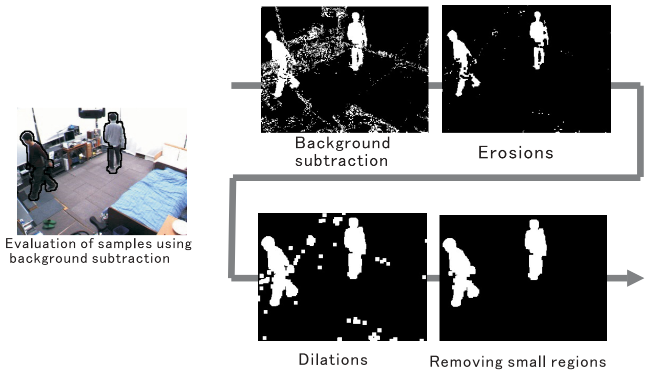

Fig. 5. Procedure in background subtraction

2 The 3D positions of samples are projected into each 2D camera plane. If projected position of a sample is outside of all camera images, the sample is eliminated (i.e. weight of the sample become zero).

3 If the projected position of a sample is outside of background subtracted images of all cameras, the sample is eliminated.

4 The remaining samples are evaluated using color histogram.

Firstly, the filter evaluates samples with room spatial information. The room size is defined in advance with actual measurement. The samples whose positions are outside of the measured room size are eliminated (i.e. weight of the sample become zero).

Secondly, 3D position of the sample is projected into 2D camera image plane. If the projected samples are outside of the all camera images, the samples are eliminated.

Camera image is processed in each frame. For background subtraction, we adopt HSV color space for image processing because HSV color space is more insensitive to change of light condition than RGB color space. We have an approach that large region including person is detected in order to absorb error of camera calibration. The processing for single camera image is subtraction from the background capture in advance, erosion, dilation, contour detection, labeling of contour inside and elimination of small region. The background is subtracted in H, S and V color space. The subtracted value in H, S and V color space is binarized in each threshold. The threshold for V color space is large, because the value in V color space is sensitive to environment lighting condition. The binarized result in each color space is integrated by logical addition. The image processing procedure is shown in Fig. 5. If the projected position of the sample is on the foreground pixel, the weight of the sample is one. If not, the weight is zero.

Finally, the filter evaluates a sample with similarity between color histogram included in the filter and color histogram that is calculated from pixels near the projected positions of the sample. We define the near pixels in the camera image as the region between u_t pixels in u axis and 20 pixels in v axis. u_t is variable and is calculated from 1300 [mm] height in real world, estimated human position and calibration parameters in every frame. The histogram is captured in HSV color space for insensitive to lighting condition. In SV color space, when value is lower than the threshold, the value is regarded as the value of no color (i.e. white or black) and is inserted into V bin in the histogram. When the value is higher than the threshold, the value is inserted into HS bin. We define number of bins about H color space N_H is 10, number of bins about S color space N_S is 10 and number of bins about V color space N_V is 10. Thus, total number of bins N is 110 ($N = N_H N_S + N_V = 110$). Histogram calculation is illustrated in Fig. 6. We apply Bhattacharyya similarity coefficient as histogram similarity. The similarity is calculated as below.

$$D(p,q) = \sqrt{1 - \sum_{k=0}^{N} \sqrt{p_k q_k}} \quad \ldots\ldots \quad (4)$$

$$\pi_{bc,t} \propto e^{-\lambda D^2(p,q)}, \quad \ldots\ldots \quad (5)$$

where both p and q are normalized values in bins and λ is a constant which is experimentally determined.

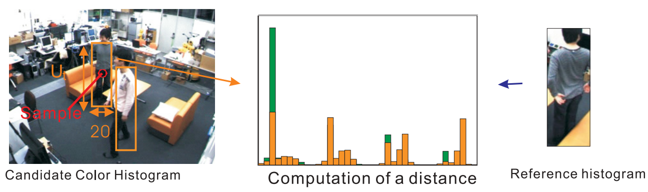

Fig. 6. Evaluation by color histogram for body tracking

3.3. Head tracking based on Body Tracking Estimation

The head tracking filter generates new samples $s_{bh,t}$ from estimated body position and relationship model h_t between body and head, and merges the new samples into generated samples in previous frame, and evaluates the samples by estimated body position and camera images.

3.3.1. Transition Model for Head Tracking

The sample for head tracking $\boldsymbol{s}_{h,t}$ is generated by the mixture distribution based on both state transition $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$ based on head movement model and the distribution $p(\boldsymbol{x}_t|\boldsymbol{y}_t)$ based on estimated body. Total N samples are separated into aN samples $\boldsymbol{s}_{bh,t}$ generated by movement model and $(1-a)N$ samples $\boldsymbol{s}_{h,t}$ selected based on previous estimation. The $s_{bh,t}$ is calculated by the propagation of samples $\boldsymbol{s}'_{bh,t}$ based on body-head model $\boldsymbol{h}_{t-1}$. samples $\boldsymbol{s}_{rh,t}$ is generated around the estimated center position of the body. Body-head model $\boldsymbol{h}_t$ represents relative position between body and head. Body-head model $\boldsymbol{h}_t$ denotes the distance between head and body of the target at time t. Samples $\boldsymbol{s}_{bh,t}$ are given by

$$\boldsymbol{s}_{bh,t} = \boldsymbol{s}'_{bh,t} + \boldsymbol{h}_{t-1} + \boldsymbol{\omega} \quad \cdots\cdots\cdots \quad (6)$$

$$\boldsymbol{h}_t = \boldsymbol{p}_{h,t} - \boldsymbol{p}_{b,t} \quad \cdots\cdots\cdots \quad (7)$$

Where $\boldsymbol{\omega}$ is a fixed 4D Gaussian noise. The samples $\boldsymbol{s}_{rh,t}$ are selected in proportion to weight $\pi_{h,t-1}$ and propagated with state transition probability. Transition model is the same as that of body tracking. We define $a = 0.33$ empirically.

3.3.2. Observation Model for Head Tracking

The samples $\boldsymbol{s}_{h,t}$ are evaluated by the following.

1 Remove the samples whose 3D positions are outside of the room.

2 If the distance between the estimated center 3D position of the body and 3D position of a sample is larger than the threshold, remove the sample.

3 If the projected position of a sample into 2D camera plane is out of image size in all cameras, eliminate the sample.

4 If the projected position of a sample does not occupy background subtracted images in all cameras, eliminate the sample.

5 Evaluate weights of samples by color histogram at each camera and multiply the evaluated likelihoods.

The subtracted images and information about room size is the same as that at the body tracking.

In addition to the same evaluation as the body tracking, we introduce another evaluation step in order to reflect the relationship between body and head into the filter. The evaluation is based on distance between position of samples and estimated position of body center. This evaluation eliminates samples whose positions are wrong. This contributes to improvement of head position estimation.

Evaluation based on color histogram is the same as the body tracking. We define the near pixels in the camera image as the region between 15 pixels in u axis and 15 pixels in v axis. This evaluation is shown in Fig. 7.

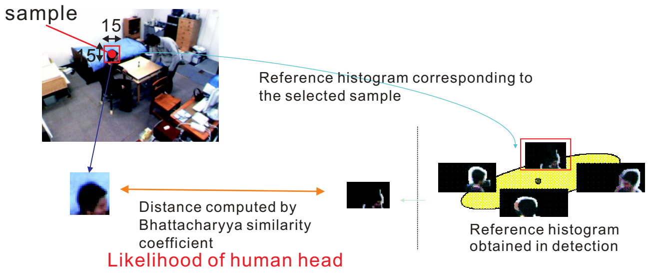

Fig. 7. Evaluation by color histogram for head tracking

4. Experiment on Multiple People Tracking

4.1. Experiment Condition

We utilized four cameras at the four corners on the ceiling of the room. The camera is Dragon Fly2. Each image was captured as a resolution of 320 x 240. We calculate intrinsic parameters and extrinsic parameters of the cameras by Zhang method[19] in advance.

We utilized LMS-200 as a laser range scanner. The LMS-200 is configured at scan range 100 [deg], angle resolution 0.25 [deg] and scan speed 18.75Hz. The two scanners are equipped on the two corners of the room. The experimental room layout is shown in Fig. 8.

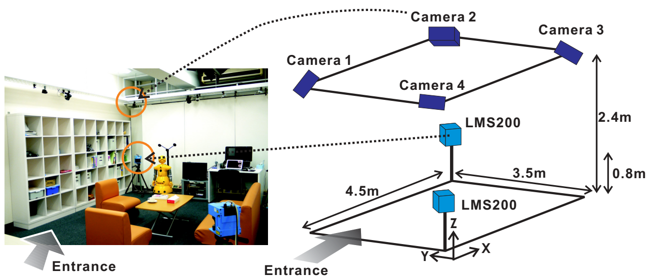

Fig. 8. Configuration of cameras and laser range finders

We conduct the experiment offline with sensor data set captured in 10Hz. In the experiments, the three persons whose body heights are from aprox. 165[cm] to aprox. 175[cm] walked, sited on the sofas in the room and bended down. At first, two persons entered the room. After entrance, the third person entered the room. These three persons moved around the table and sited on the sofas. Total number of frames is 1178.

As for the parameters of the filters, number of the samples in head tracking filter and body tracking filter are 150 and 250, respectively. In parameters of observation model for the scanner, we define $\alpha = 0.1, \beta = 1.0, \gamma = 10.0$. We utilize identity matrix whose size is 3 by 3 as $\boldsymbol{\Gamma}$ of the adaptive control parameter on the transition model. The thresholds for bin determination of color histogram are $S = 0.1$ and $V = 0.07$.

4.2. Experiment Result

In the experiment, there are no mistake of identification and no miss of tracking in total frames. The 3D position estimation result is shown in Fig. 9. The projected positions into x-y plane are shown in the left part of the figure. The graph represents the moving trajectory avoiding the tables and sofas. The dense group of positions on the sofa means sitting positions. The right part of the figure means relationship between the frames and head heights of people. The heights become low as the person sits on the sofa or bends. The heights of the bodies are more unstable than the heights of heads. This is because evaluating samples of bodies from cameras are sensitive to the occlusion of sofas.

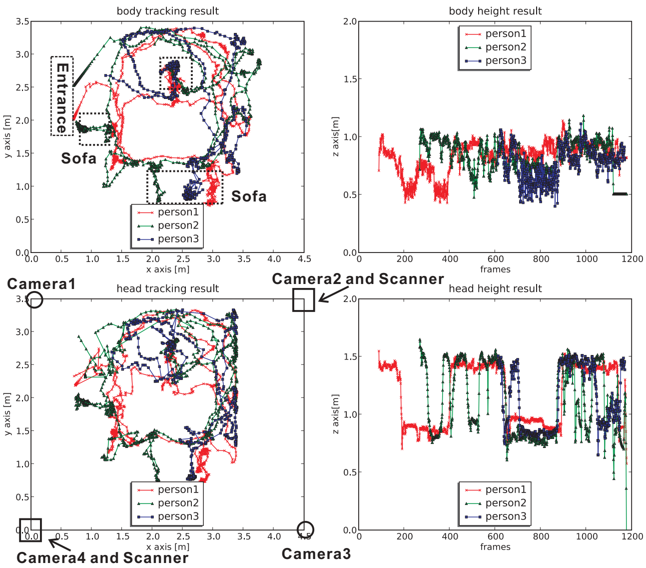

Fig. 9. 3D-Positions of bodies and heads in the experiments

The typical tracking scene is shown in Fig. 10. 3D image is generated with projection of estimation result to camera plane. The rectangle in the image represents area including all samples. The area means detected person area. The rectangle near the head means detected head area. Fig. 11 demonstrates the typical estimation sequence when three people exist in the room. The result image shows the proposed filter tracks and labels bodies and heads of the people.

As for the body tracking, at frame no. 880 and no. 920, the scan data is occluded. Half of the outer shape of ID-1 person disappears and the shape is distorted because of sitting posture at frame no. 880. Since 4 cameras capture 3 regions of people in the camera images, the identification and tracking of people are succeeded. Frame no. 920 is also the same situation that multiple cameras support the identification and tracking. From frame no. 890 to frame no. 910, the camera data is occluded. For example, in images of camera 1 and camera 3 at frame no. 900, regions of people overlap each other. At camera 3, ID-1 person is invisible. However, in the scan image of the frame, shapes of 3 peoples are separated accurately. The 3D position estimation and identification is achieved. These examples show the complementation mechanism of the cameras and the laser range scanners at tracking and identification works well.

The head tracking is also succeeded in all frames. Although head tracking in sitting behavior is difficult because of head's fast movement, frame no. 880 and 890 show 3D position of head at ID-1 person is estimated accurately. At frame no. 910 and 920, the head position of ID-3 person shifts slightly. The histogram of head is the similar to that of the other head, while the histograms of the bodies are different enough to distinguish the people. This similarity causes wrong observation in people' crossing movement in the filter.

The calculation time in one frame is within 100ms stably. This means the method is practicable in calculation time.

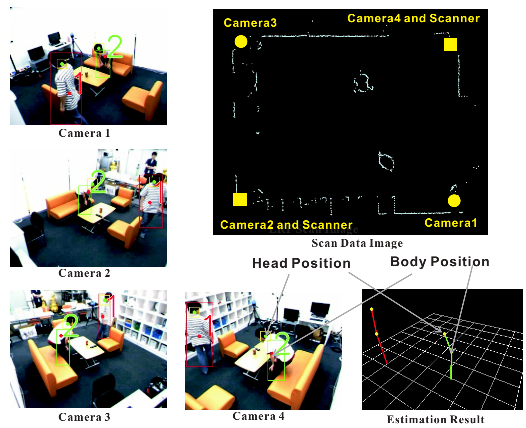

Fig. 10. Camera images and scan image in experiments

5. Conclusion

We propose the method for tracking multiple persons from cameras and laser range scanners in particle filter framework. The method consists of body tracking filters and head tracking filters. The method estimates 3D positions of body and head with identification. The scanner data and camera data are fused by two steps in particle filter framework. At first step, samples are evaluated by accurate scanner data. At the second step, multiple camera images evaluate samples for 3D position estimation and identification. Due to the features, the method covers each weak point of the cameras and the scanner. The method can estimate 3D positions of bodies and heads of the people with labeling IDs. The head position is also estimated by the particle filter with calculated 3D position of the body and camera images. The experiment demonstrates robustness of the tracking people which walk and sit in real home environment. In the future, we will challenge the posture estimation from cameras and scanners.

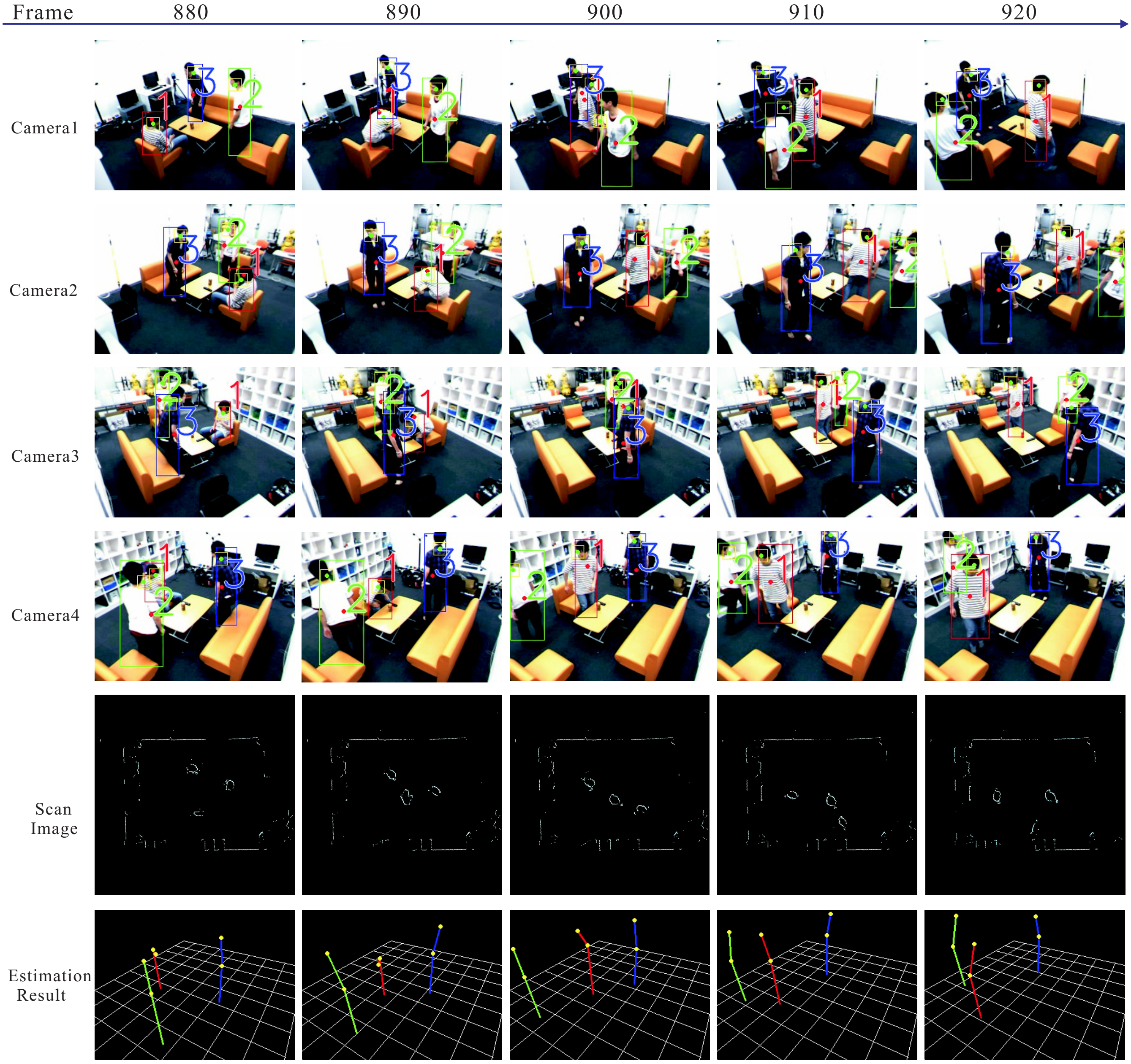

Fig. 11. Experimental results on three people tracking

References:

[1] J. Cui, H. Zha, H. Zhao, and R. Shibasaki, "Tracking Multiple People using Laser and Vision", In International Conference on Intelligent Robots and Systems, pp. 1301–1306, 2005.

[2] J. Cui, H. Zha, H. Zhao, and R. Shibasaki, "Laser-based Interacting People Tracking Using Multi-level Observations", In IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 1799–1804, 2006.

[3] J. Czyza, B. Risticb, and B. Macqa, "A Particle Filter for Joint Detection and Tracking of Color Objects", Image and Vision Computing, 25(8):1271–1281, 2007.

[4] A. Dore, A. F. Cattoni, and C. S. Regazzoni, "A Particle Filter Based Fusion Framework for Video-Radio Tracking in Smart Spaces", In International Conference on Advanced Video and Signal based Surveillance(AVSS 2007), pp. 99–104, 9 2007.

[5] A. Fod, A. Howard, and M. J. Mataric, "A Laser-Based People Tracker", In IEEE International Conference on Robotics and Automation, pp. 3024–3029, 2002.

[6] D. F. Glas, T. Miyashita, H. Ishiguro, and N. Hagita, "Laser Tracking of Human Body Motion Using Adaptive Shape Modeling", In IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 602–608, 2007.

[7] M. Isard and A. Blake, "Condensation-Conditional density propagation for visual tracking", Internanional Journal on Computer Vision, 29(1):5–28, 1998.

[8] K. Kim and L. S. Davis, "Multi-camera Tracking and Segmentation of Occluded People on Ground Plane Using Search-Guided Particle Filtering.", In ECCV, pp. 98–109, 2006.

[9] Y. Kobayashi, D. Sugimura, Y. Sato, K. Hirasawa, N. Suzuki, H. Kage, and A. Sugimoto, "3D Head Tracking using the Particle Filter with Cascaded Classifiers", In The British Machine Vision Conference, pp. 37–46, 2006.

[10] R. Kurazume, H. Yamada, K. Murakami, Y. Iwashita, and T. Hasegawa, "Target Tracking Using SIR and MCMC Particle Filters by Multiple Cameras and Laser Range Finders", In IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 3838–3844, 2008.

[11] W.-H. Liau, C.-L. Wu, and L.-C. Fu, "Inhabitants Tracking System in a Cluttered Home Environment Via Floor Load Sensors", IEEE Transactions on Automation Science and Engineering, 5(1):10–20, 2008.

[12] Y. Matsumoto, T. Kato, and T. Wada, "An Occlusion Robust Likelihood Integration Method for Multi-Camera People Head Tracking", In International Conference on Networked Sensing Systems, pp. 235–242, 2007.

[13] T. Mori, T. Matsumoto, M. Shimosaka, H. Noguchi, and T. Sato, "Multiple Persons Tracking with Data Fusion of Multiple Cameras and Floor Sensors Using Particle Filters", In ECCV Workshop on Multi-camera ad Multi-modal Sensor Fusion Algorithms and Applications(M2SFA2), 2008.

[14] T. Mori, H. Noguchi, A. Takada, and T. Sato, "Sensing Room Environment: Distributed Sensor Space for Measurement of Human Dialy Behavior", Transaction of the Society of Instrument and Control Engineers, E-S-1(1):97–103, 2006.

[15] T. Mori, Y. Suemasu, H. Noguchi, and T. Sato, "Multiple People Tracking by Integrating Distributed Floor Pressure Sensors and RFID System", In IEEE International Conference on System Man and Cybernetics, pp. 5271–5278, 2004.

[16] H. Morishita, R. Fukui, and T. Sato, "High Resolution Pressure Sensor Distributed Floor for Future Human-Robot Symbiosis Environment", In International Conference on Intelligent Robots and Systems(IROS), pp. 1246–1251, 2002.

[17] T. Murakita, T. Ikeda, and H. Ishiguro, "Human Tracking using Floor Sensors based on the Markov Chain Monte Carlo Method", In International Conference on Pattern Recognition(ICPR), pp. 917–920, 2004.

[18] J. Vermaak and A. D. a. Perez, "Maintaining Multi-Modality through Mixture Tracking", In International Conference on Computer Vision, pp. 1110–1116, 2003.

[19] Z. Zhang, "A Flexible New Technique for Camera Calibration", Microsoft Research Technical Report,MSR-TR-98-71, 1998.

[20] H. Zhao and R. Shibasaki, "A Novel System for Tracking Pedestrians Using Multiple Single-row Laser-Range Scanners", IEEE Transacations on Systems, Man and Cybernetics Part A:Systems and Humans, 35(2):283–291, 2005.

Name:
Hiroshi Noguchi

Affiliation:
Research Associate,The University of Tokyo

Address:
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656 JAPAN

Brief Biographical History:
2006- Research Associate, The University of Tokyo (Researcher, CREST, Japan Science and Technology)

Membership in Learned Societies:
- The Robotics Society of Japan

Name:
Taketoshi Mori

Affiliation:
Associate Professor,The University of Tokyo

Address:
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656 JAPAN

Brief Biographical History:
1995- Research Associate, The University of Tokyo
1998- Assistant Professor, The University of Tokyo
2002- Associate Professor, The University of Tokyo

Membership in Learned Societies:
- The Robotics Society of Japan
- The IEEE Robotics and Automation Society
- The IEEE Computer Society
- The Japan Society of Mechanical Engineers
- The Japanese Society for Artificial Intelligence
- The Institute of Electronics, Information and Communication Engineers
- The Society of Instrument and Control Engineers

Name:
Takashi Matsumoto

Affiliation:
Advanced Information Technology Group, Research Center For Information Technology, Mitsubishi Research Institute, Inc.

Address:
3-6, Otemachi 2-chome Chiyoda-ku, Tokyo, 100-8141, Japan

Brief Biographical History:
2009- Mitsubishi Research Institute, Inc.

Name:
Masamichi Shimosaka

Affiliation:
Research Associate,The University of Tokyo

Address:
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656 JAPAN

Brief Biographical History:
2006- Research Associate, The University of Tokyo

Membership in Learned Societies:
- The Robotics Society of Japan
- The IEEE Robotics and Automation Society
- The IEEE Computer Society
- The Japan Society for Artificial Intelligence
- The Institute of Electronics, Information and Communication Engineers

Name:
Tomomasa Sato

Affiliation:
Professor,The University of Tokyo

Address:
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656 JAPAN

Brief Biographical History:
1976- Electrotechnical Laboratory (ETL) of the Ministry of Industrial Science and Technology
1991- Professor, Research Center for Advanced Science and Technology (RCAST) of the University of Tokyo
1998- Professor, Department of Mechano-Infomatics of the University of Tokyo

Membership in Learned Societies:
- The Robotics Society of Japan
- The IEEE Robotics and Automation Society
- The Japan Society for Artificial Intelligence
- The Japan Society of Mechanical Engineers
- The Society of Instrument and Control Engineers