https://doi.org/10.20965/jrm.2010.p0221

Paper:

Multiple-Person Tracking by Multiple Cameras and Laser Range Scanners in Indoor Environments

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.
2-3-6 Otemachi, Chiyoda-ku, Tokyo 100-8141, Japan
E-mail: t.matsu@mri.co.jp
[Received September 30, 2009; accepted January 27, 2010]

In this paper, we propose a method for multiple-person tracking using cameras and laser range scanners. Our method estimates 3D positions of human body and head, and labels them with their identities. Individual particle filters track person correctly by integrating information from laser range scanners and target-specific information from cameras, thus compensating for weak points of each. We also develop a particle filter framework that tracks the human head simultaneously using the estimated body position. Results of experiments demonstrate the effectiveness and robustness of the proposal in tracking multiple persons with multiple scanners and cameras.

Keywords: multiple-person tracking, particle filter, laser range scanner and camera

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 in using robot assistance and recognizing human behavior patterns. 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 should not require wearable devices, especially in long-term measurement.

A variety of tracking approaches have been developed. Kobayashi [1], for example, developed 3-dimensional head tracking using multiple cameras via AdaBoost classification. Matsumoto [2] used multiple-camera head tracking using a particle filter. Kim [3] tracked subjects by segmenting occluded areas on a ground plane with integration of images from multiple cameras. 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 cameras are a rich source of information in subject identification, however, they have problems with occlusion and lighting that cause mistakes in tracking in short time.

Floor sensors [4–7] are also used to monitor persons. 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 involve problems of expense and implementation in the ordinary home environment. The floor sensors also 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.

Laser range scanner provides yet another way of tracking human position sensing because of the scanner's price-reduction in recent years and easiness in deployment. Zhao [8], and Cui [9] developed multiple-person tracking from the scanners at leg height, and Fod [10] also realizes tracking from the scanners at hip height. Glas [11] 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.

Using more than one type of sensing has also been tried, including combinations of camera and wireless de-

vices [12], cameras and floor sensors [13], and floor sensors and embedded RadioFrequency Identification (RFID) [14]. 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 propose using cameras and laser range scanners for multiple-person tracking. Because scanners solve the camera's occlusion problem and cameras compensate for the scanner's identification problem.

Cui [15] developed multiple people tracking using a single camera and scanner based on Bayesian filtering technique. Their approach's target is mainly the combination of single camera and single scanner. Kurazume [16] used multiple cameras and scanners, which was robust for crowds. However, target subjects are walking people and the target area is limited to open space without camera occlusion objects such as tables, sofas, and chairs, and they only tracked the body center. In proposing for multiple-person tracking using multiple cameras and scanners, we estimated the 3D positions of heads and bodies in the environment including camera occlusion objects such as tables and chairs. Measurement of heads and bodies positions contributes to the action recognition such as sitting and standing, which is important in human support. We also labeled persons' identities. The indentificaiton means that the same labels are correctly allocated as the same person in tracking. This identification is essential in support fitting to individual person.

Our proposal assumes a one-room tracking area with cameras at the 4 corners in a home-like environment. 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, scanners are set 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, our proposal does not require a high sampling rate. In home environment, because communication hardware such as wired LAN is stable, communication delay is negligible. Synchronization of sensors are easy. We assume the synchronization is realized between the cameras and the scanners.

2. Multiple-Person Tracking

We utilize particle filtering technique [17] for tracking. The particle filter is a kind of sequential filtering. The filter regards the tracking target as a probability density represented as a finite set of samples including states and likelihoods. The filter recursively updates probability using propagation based on state transition probability and the evaluation of likelihood through observation. The filter has an advantage in sudden subject movement and sensor noise thanks to the probabilistic mechanism. In typical camera-based tracking by a particle filter, a sample includes the 3D position as a state and sample likelihood is evaluated by matching the projected 3D position in samples to 2D camera images. The final subject position is calculated as the expectation of 3D position and likelihood.

Since the particle filter represents a tracking target is represented as non-parametric probability density distribution, discriminate probability density distribution is difficult to distinguish for each tracking target from multiple target distributions. The filter works well in estimating position in single-person tracking because number of the samples is large enough to present a 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. Based on the idea, they proposed Mixture Particle Filter (MPF), which tracks multiple targets simultaneously. Although the MPF resolves the problem on miss of target people, it may confuse or lose targets during labeling ID. For example, IDs counterchange or lose when two persons move across each other. The heterogeneous allocation of samples among the targets still causes the ID missing problems.

The filter we propose is similar to MPF. The filter consists of single-target particle filters that track the same person with samples in the filter. The single-target filter also contains target-specific information for labeling the same ID. This configuration provides robust tracking and labeling in multiple-person tracking, as shown 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 multiple-person tracker detects a person, it generates a new ID 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 [19]. Each camera 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. The tracking target is assumed to be an adult human whose body height from 150 cm to 180 cm tall, and

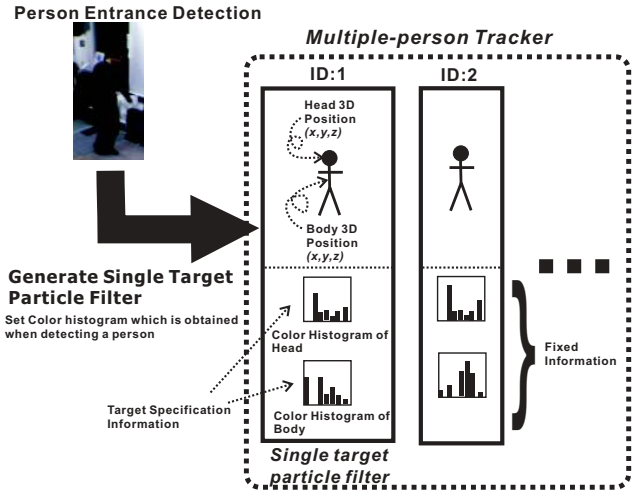


Fig. 1. Multiple-person tracking overview.

Fig. 2. Color histogram initialization.

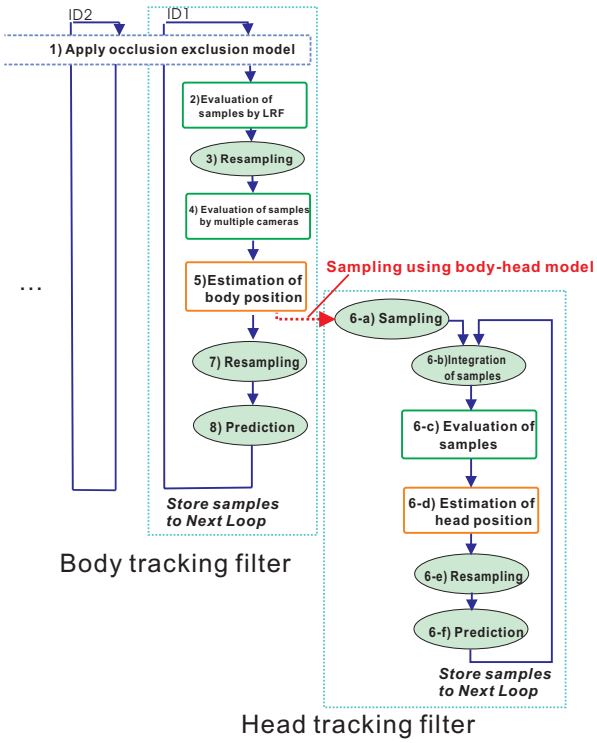


Fig. 3. Head and body tracking flow.

only one person is detected at an entrance. 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 head and body histograms.

3. Multiple-Person Tracking Using Laser Range Scanners and Cameras

3.1. Head and Body Tracking

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

The body tracking filter calculates the position of body by two-step estimation. The filter first evaluates samples with data of the laser range scanners, which are more accurate than camera data. After evaluation, the filter updates the evaluated samples. The filter then evaluates updated samples with the camera data in terms of position. The filter also evaluates updated samples with containing color histogram, then updates the samples. Sample for body tracking $\boldsymbol{s}_b$ contains 3D position (x, y, z) as a state. Sample for head tracking $\boldsymbol{s}_h$ also includes 3D position. The head tracking filter estimates head position with the calculated body position using the body tracking filter and camera data as follows and as shown in **Fig. 3**.

1. Apply an exclusion model for occlusion [13] to body-tracking samples $\boldsymbol{s}_{b,t}$ in body tracking filters. Weights of samples in overlapping areas become zero.
2. Evaluate weight $\pi_{bl,t}$ of sample $\boldsymbol{s}_{b,t}$ using scanner 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 a Gaussian distribution.
4. Evaluate weight $\pi_{bc,t}$ corresponding to sample $\boldsymbol{s}'_{b,t}$ using camera data.
5. Estimate the center 3D position of body $\boldsymbol{p}_{b,t}$.
6. Measure the head position as follows
   a. Generate samples $\boldsymbol{s}_{bh,t}$ from the center position of the body and model between body and head $\boldsymbol{h}_{t-1}$.
   b. Generate samples $\boldsymbol{s}_{h,t}$ by integrating samples generated from 6-a) and updated samples $\boldsymbol{s}_{rh,t}$ at time $t-1$.
   c. Evaluate weight $\pi_{h,t}$ of samples $\boldsymbol{s}_{h,t}$ using camera data.
   d. Estimate head position $\boldsymbol{p}_{h,t}$ 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 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. Laser Range Scanner and Camera Body Tracking

The 3D body center is tracked by two-step estimation in the filter. Conventionally, this position is estimated during a single step, because sensor measurements are independent, e.g., multiplying by the weighted value evaluated from each sensor for total evaluation. This mechanism makes it is easy to fuse multiple types of sensors, but, it is difficult to design and to adjust parameters – e.g., weights in evaluation at every sensor – in sensor data fusion. The evaluation of one sensor may drastically exceed that of another sensor. This precludes the advantage of sensor fusion. Since no weight in the above example is zero, total evaluation, which should be small, becomes large due to sensor evaluation raised by noise. In our sensor fusion, however, because measurement by laser range scanners is accurate, they completely eliminate incorrect evaluation. Instead of weighted evaluation multiplication, we separate evaluation into two steps for using advantage of the scanner measurement.

(1) The filter evaluates samples by the laser range scanners due to their more accurate distance measurement. Because the accurate distance measurement easily separates people's areas, it limits position candidates into narrow area by selecting samples based on their evaluated weights. (2) The filter evaluates selected samples using camera images. The ID that the filter contains is evaluated using the camera images and color histogram included in the filter, then, samples are selected in proportion to their weights. The 3D body center position is estimated using selected samples.

3.2.1. Body Center Tracking Transition Model

We assume uniform straight movement 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 by

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

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

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-person tracking. To improve sample diffusion in sudden movement and remaining still, we introduce adaptive control of diffusion factor $\boldsymbol{\omega}$, in which diffusion regions change based on tracking target speed. In fast target movement, the diffusion region becomes wide. In implementing control, $\boldsymbol{\omega}$ is represented as white noise within the range $-\boldsymbol{\delta}_t \leq \boldsymbol{\omega} \leq \boldsymbol{\delta}_t$. Three-dimensional vector $\boldsymbol{\delta}_t$ is controlled based on target speed. $\boldsymbol{\delta}_t$ increases in proportion to absolute target velocity.

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

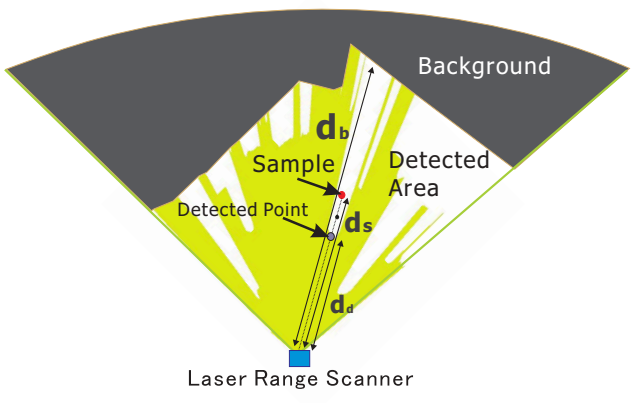


Fig. 4. Laser range scanner observation model parameters.

$\bar{\boldsymbol{v}}_t$ is a 3D vector including absolute value in each axis, $\boldsymbol{\Gamma}$ is a fixed 3D diagonal matrix, and $\boldsymbol{\sigma}$ is a fixed 3D vector. System noise $\boldsymbol{\omega}$ control improves position estimation and reduces tracking error in sudden movement.

3.2.2. Laser Range Scanner Observation Model

Samples are first evaluated by scanners with 3D sample positions projected onto a horizontal plane. The laser range scanner captures a single scan without person as background information in advance, then measures distance d_d in the i-th direction in a single scan. Sample weight $\pi_{bl,t}$ is evaluated as follows 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 as shown in **Fig. 4**:

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

This evaluation excludes conversion from polar to 2D coordinates, so the weight is calculated quickly. Calculation cost depends on the number of scanners and samples.

3.2.3. Camera Observation Model

Samples are then evaluated using spatial room information, matching background-subtracted images of individual cameras and color histograms. The filter evaluates samples hierarchically from camera images to minimize calculation time as follows:

1. Weights of samples outside of the room size become zero.
2. The 3D sample positions are projected onto individual 2D camera planes. If a sample's projected position is outside of all camera images, the sample is eliminated, i.e., its weight becomes zero.
3. If a sample's projected position is outside of background-subtracted images of all cameras, the sample is eliminated.

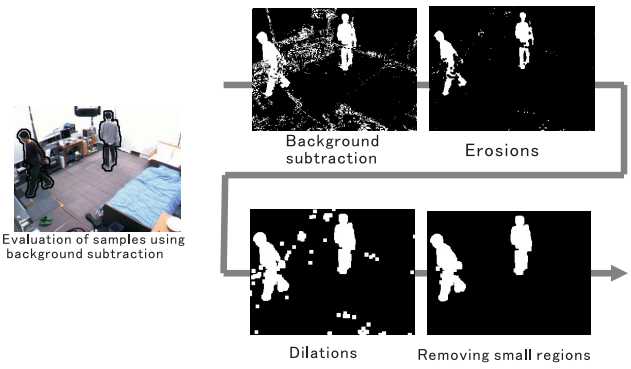


Fig. 5. Background subtraction.

4. Remaining samples are evaluated using color histograms.

The filter first evaluates samples with spatial room information. Room size is defined in advance by actual measurement. Samples whose positions are outside of the measured room size are eliminated, i.e., their weight becomes zero.

The sample's 3D position is projected onto the 2D camera image plane. If projected samples are outside of all camera images, samples are eliminated.

A camera image is processed in each frame. For background subtraction, we use HSV color space for image processing because of its greater robustness against changes in light conditions, compared to RGB color space. Large region including person is detected to absorb camera calibration error. Processing for a single camera image consists of subtraction from background capture in advance, erosion, dilation, contour detection, a contour inside labeling and elimination of small regions. The background is subtracted in H, S and V color space. This subtracted value is binarized at each threshold. The threshold for V color space is large, because the value in V color space is sensitive to environmental lighting condition. The binarized result in each color space is integrated using logical addition. Image processing procedure is shown in **Fig. 5**. If the projected sample position is on a foreground pixel, the sample weight is one and, if not, zero.

The filter evaluates a sample with similarity between color histogram included in the filter and color histogram that is calculated from pixels near projected sample positions. We define near pixels in camera image as the region between u_t pixels on the u-axis and 20 pixels on the v-axis. u_t is variable and is calculated from 1300 mm height in the real world, estimated subject position and calibration parameters in each frame. The histogram is captured in HSV color space for insensitivity to lighting conditions. In SV color space, a value lower than a threshold is regarded as the value of no color, i.e., white or black, and is inserted into the V bin in the histogram. When the value exceeds the threshold, the value is inserted into the HS bin. We define the number of bins for H color space N_{H} is 10, the number of bins for S color space N_{S} is 10, and the number of bins for V color space N_{V} is 10, The total number of bins N is 110 ($N = N_{\mathrm{H}}N_{\mathrm{S}} + N_{\mathrm{V}} = 110$). Histogram calculation is shown in **Fig. 6**. We use Bhattacharyya's similarity coefficient as histogram similarity, calculated by

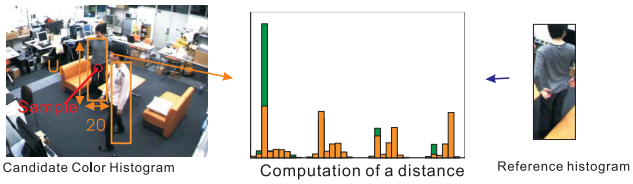


Fig. 6. Color histogram evaluation in body tracking.

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

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

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

3.3. Head Tracking Based on Body Tracking Estimation

The head tracking filter generates new samples $\boldsymbol{s}_{bh,t}$ from an estimated body position and relationship model $\boldsymbol{h}_t$ between head and body, and merges the new samples into samples generated in the previous frame, and evaluates the samples by estimated body position and camera images.

3.3.1. Head Tracking Transition Model

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

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

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

$\boldsymbol{\omega}$ is 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. The transition model is the same as that for body tracking. We define $a = 0.33$ empirically.

3.3.2. Head Tracking Observation Model

Samples $\boldsymbol{s}_{h,t}$ are evaluated as follows:

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

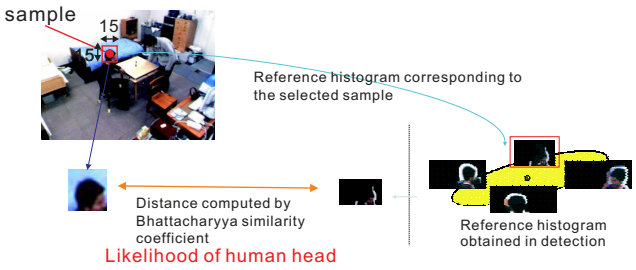


Fig. 7. Color histogram evaluation for head tracking.

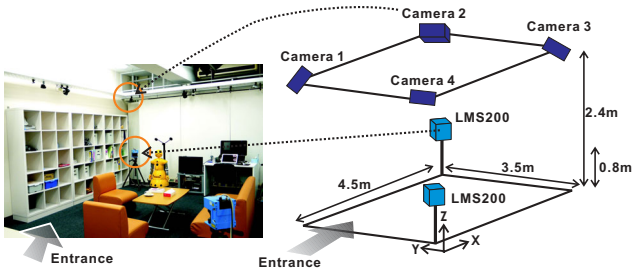


Fig. 8. Camera and laser range scanner configuration.

2. If the distance between the estimated center 3D position of the body and the 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 sample weights using color histogram at each camera and multiply the evaluated likelihoods.

The subtracted images and the information about room size are the same as that for the body tracking.

We introduce another evaluation step here to reflect the relationship between the head and body in the filter. The evaluation is based on the distance between sample positions and estimated body center position, eliminating samples whose positions are incorrect. This contributes to improvement of head position estimation.

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

4. Experiment in Multiple-Person Tracking

4.1. Experiment Conditions

As stated, we used four Dragonfly2 cameras at the room's four ceiling corners. Individual images were captured at a resolution of 320×240. We calculate intrinsic and extrinsic camera parameters using Zhang method [20] in advance.

The LMS-200 laser range scanner has scan range of 100°, an angle resolution of 0.25° and a scan speed of 18.75 Hz. The scanners are located in the two corners of the room. The experimental room layout is shown in **Fig. 8**.

We conducted experiments offline with a sensor dataset captured at 10 Hz. Three persons – first 2 together, followed by the third – from 165 cm to 175 cm tall walked into the room, moved around the table, sat down on the sofas and bent down. Total number of data frames is 1178.

In the experiments, the number of the samples in head tracking filter and body tracking filter are 150 and 250, respectively. In observation model parameters for the scanner, we define $\alpha = 0.1, \beta = 1.0, \gamma = 10.0$. We use an identity matrix 3 by 3 as $\mathbf{\Gamma}$ 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. Experimental Results

No identification error or no miss of tracking occurred in total frames in experiments. The 3D position estimation results are shown in **Fig. 9** with positions projected onto the x-y plane at left and heights at right. The upper part is body tracking result and the lower part is head tracking result. The graphs at the left shows the movement trajectory avoiding the table and the sofas. The dense positions on the sofas are seating positions. The graphs at the right shows relationship between frames and subjects' head heights. The heights become low when the subjects sit on the sofas or bend. The body heights are more unstable than the head heights because evaluating body samples from cameras are sensitive to occlusion by the sofas.

The typical tracking scene is shown in **Fig. 10**. A 3D image is generated with estimation results projected onto the camera plane. The rectangle includes all samples. The area means detected person area. The rectangle near the head means detected head area. **Fig. 11** shows a typical estimation sequence when three people are in the room. The result images show the proposed filter tracks and labels heads and bodies.

As for the body tracking, in frames 880 and 920, the scan data is occluded. Half of the outer shape of person ID-1 disappears and is distorted due to sitting posture in frame 880. Since 4 cameras capture 3 regions of people in the camera images, the identification and tracking of people are succeeded. Frame 920 also shows how multiple cameras support identification and tracking. The camera data is occluded from frame 890 to frame 910. In camera 1 and 3 images at frame 900, regions of people overlap, and for camera 3, person ID-1 is not visible. In the scan image of the frame, shapes of 3 persons are separated accurately. The 3D position estimation and identification is achieved successfully. These examples show the complementation mechanism of the cameras and the laser range scanners at tracking and identification works well.

The head tracking was successful in all frames, even though this is difficult in sitting due to fast head's movement. Frames 880 and 890 show 3D head position at person ID-1 is estimated accurately. In frames 910 and 920,

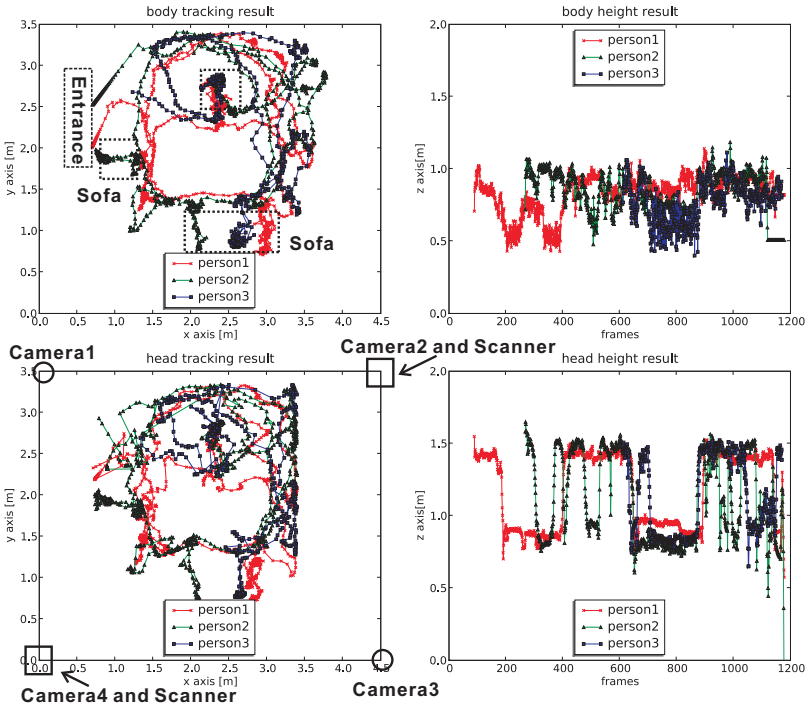


Fig. 9. Head and body position in experiments.

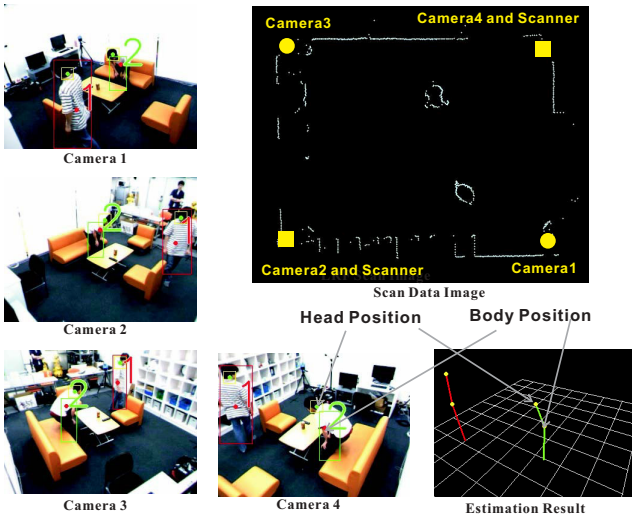


Fig. 10. Camera images and scan image in experiments.

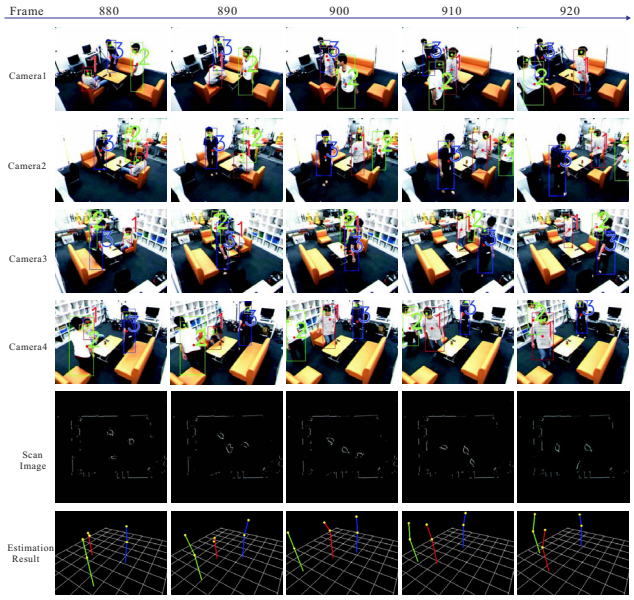


Fig. 11. Experimental results in three-person tracking.

the head position of person ID-3 shifts slightly. The head histogram is similar to the other head, although body histograms differ enough to distinguish the persons. This similarity causes incorrect observation in persons' crossing in the filter.

The calculation time per frame is within 100 ms stably. This time makes our proposal applicable in practical use.

5. Conclusions

We have proposed multiple-person tracking using cameras and laser range scanners in a particle filter framework involving head and body tracking filters. The filter estimates 3D positions of head and body with identification. Scanner and camera data are merged in a two-step particle filter framework. In step 1, samples are evaluated using scanner data because the scanner measures accurate distance. In step 2, multiple camera images evaluate samples for 3D position estimation and identification. The filter compensates for weak points of cameras and scanners. Head and body positions are estimated in 3D are labeled with their identifies. The head position is also estimated by the particle filter with calculated 3D position of the body and camera images. The experiment demonstrated robustness of multiple-person tracking for behav-

iors such as sitting, bending and walking in real home environment. In the future, we will challenge the posture estimation from cameras and scanners.

References:

[1] 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 Conf., pp. 37-46, 2006.

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

[3] 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.

[4] T. Mori, H. Noguchi, A. Takada, and T. Sato, "Sensing Room Environment: Distributed Sensor Space for Measurement of Human Dialy Behavior," Trans. of the Society of Instrument and Control Engineers, E-S-1 No.1, pp. 97-103, 2006.

[5] W.-H. Liau, C.-L. Wu, and L.-C. Fu, "Inhabitants Tracking System in a Cluttered Home Environment Via Floor Load Sensors," IEEE Trans. on Automation Science and Engineering, Vol.5, No.1, pp. 10-20, 2008.

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

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

[8] H. Zhao and R. Shibasaki, "A Novel System for Tracking Pedestrians Using Multiple Single-row Laser-Range Scanners," IEEE Trans. on Systems, Man and Cybernetics Part A: Systems and Humans, Vol.35, No.2, pp. 283-291, 2005.

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

[10] A. Fod, A. Howard, and M. J. Mataric, "A Laser-Based People Tracker," In IEEE Int. Conf. on Robotics and Automation, pp. 3024-3029, 2002.

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

[12] A. Dore, A. F. Cattoni, and C. S. Regazzoni, "A Particle Filter Based Fusion Framework for Video-Radio Tracking in Smart Spaces," In Int. Conf. on Advanced Video and Signal based Surveillance (AVSS 2007), pp. 99-104, Sep. 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, Y. Suemasu, H. Noguchi, and T. Sato, "Multiple People Tracking by Integrating Distributed Floor Pressure Sensors and RFID System," In IEEE Int. Conf. on System Man and Cybernetics, pp. 5271-5278, 2004.

[15] J. Cui, H. Zha, H. Zhao, and R. Shibasaki, "Tracking Multiple People using Laser and Vision," In Int. Conf. on Intelligent Robots and Systems, pp. 1301-1306, 2005.

[16] 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 Int. Conf. on Intelligent Robots and Systems (IROS), pp. 3838-3844, 2008.

[17] M. Isard and A. Blake, "Condensation-Conditional density propagation for visual tracking," Int. J. on Computer Vision, Vol.29, No.1, pp. 5-28, 1998.

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

[19] J. Czyza, B. Risticb, and B. Macqa, "A Particle Filter for Joint Detection and Tracking of Color Objects," Image and Vision Computing, Vol.25, No.8, pp. 1271-1281, 2007.

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

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 Academic Societies:
- The Robotics Society of Japan (RSJ)

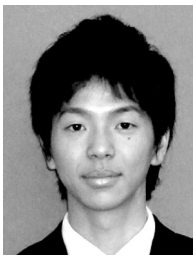
Name:
Taketoshi Mori

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 Academic Societies:
- The Robotics Society of Japan (RSJ)
- The IEEE Robotics and Automation Society (IEEE-RAS)
- The IEEE Computer Society (IEEE-CS)
- The Japan Society of Mechanical Engineers (JSME)
- The Japanese Society for Artificial Intelligence (JSAI)
- The Institute of Electronics, Information and Communication Engineers (IEICE)
- The Society of Instrument and Control Engineers (SICE)



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 Academic Societies:
- The Robotics Society of Japan (RSJ)
- The IEEE Robotics and Automation Society (IEEE-RAS)
- The IEEE Computer Society (IEEE-CS)
- The Japan Society for Artificial Intelligence (JSAI)
- The Institute of Electronics, Information and Communication Engineers (IEICE)

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-Informatics of the University of Tokyo

Membership in Academic Societies:
- The Robotics Society of Japan (RSJ)
- The IEEE Robotics and Automation Society (IEEE-RAS)
- The Japan Society for Artificial Intelligence (JSAI)
- The Japan Society of Mechanical Engineers (JSME)
- The Society of Instrument and Control Engineers (SICE)