Multiple Persons Tracking with Data Fusion of Multiple Cameras and Floor Sensors Using Particle Filters

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

HAL Id: inria-00326752

https://inria.hal.science/inria-00326752

Submitted on 5 Oct 2008

Multiple Persons Tracking with Data Fusion of Multiple Cameras and Floor Sensors Using Particle Filters

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

Graduate School of Information Science and Technology, The University of Tokyo,
7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan
{tmori, matumoto, simosaka, noguchi}@ics.t.u-tokyo.ac.jp,
tomomasasato@jcom.home.ne.jp

Abstract. Successful multi-target tracking requires locating the targets and labeling their identities. For the multi-target tracking systems, the latter becomes more challenging when the targets frequently interact with each other. In this paper, we propose a method for multiple persons tracking using multiple cameras and floor sensors. Our method estimates 3D positions of human body and head, and labels their identities. The method is composed of multiple particle filters that interact only in the exclusion occlusion model. Each particle filter tracks each person correctly by integrating information from floor sensors 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 new 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 against several complicated movements of multiple persons. The results also demonstrate that this method can maintain correct tracking when the targets are in close proximity.

1 Introduction

Recently, expectations towards robotic systems which enables daily life assistance are rising. Among many, researches on home environment with distributed sensors are active, and believed to offer practical applications. Aware Home[1] and Sensing Room[2] are examples of such systems. When considering the appropriate support by such systems, position and ID of the targets are the most important information. Without tracking position and ID, inappropriate support and accidents might occur. In addition to the human body position, tracking a human part (i.e. a head, a hand) also enables a wide variety of support.

There are a number of researches on human tracking. Various sensors are tried for human tracking in those researches. Cameras[3–6], RFID[7], laser range finder[8], and floor pressure sensors[9–11] are used.

Among a variety of sensors, cameras are especially used in many works. Most of these methods use particle filter framework[12–14]. Kobayashi et al.[3] used cascaded classifier on AdaBoost in the framework. Kim et al.[4] incorporated color histograms of the multiple camera images and segmentation on the ground plane into the framework. These methods succeeded in tracking a single person robustly or multiple persons in each environment. However, they are not sufficient to track against complicated movements of multiple persons.

On the other hand, Fukui et al.[10] proposed a method for human locomotion with probability potential maps by the sensor floor and Murakita et al.[11] proposed a method for human tracking using floor sensors based on the Markov Chain Monte Carlo Method. These methods show the effectiveness of floor pressure sensors against a single person tracking. However, it is very difficult to label their identities by using only floor pressure sensors.

Sensor fusion techniques for human tracking are also proposed. Dore et al.[15] proposed a video-radio fusion technique using particle filter framework. In [7], floor pressure sensors and RFID System are used for multiple persons tracking. However, these methods are not sufficient to estimate identities and positions of multiple persons correctly when the targets are in close proximity.

We then propose a method for multiple persons tracking using multiple cameras and floor sensors, which estimates 3D positions of human body and head, and labels their identities. The method is composed of multiple particle filters. In the particle filter framework, we integrate multiple camera images and floor sensors. Camera images are influenced by illumination, but a great deal of information is provided by cameras. Floor sensors are resistant to environmental changes and cover wide area, but it is difficult to discriminate multiple persons. Integration of these two types of sensors enables complement of each weak point and the correct tracking of the targets. Moreover, to track both human body and head simultaneously, we develop a new particle filter framework that tracks the human head by using the estimated body position.

This paper is organized as follows. We briefly give an overview of the particle filter in section 2. In section 3, we propose a framework for multiple persons tracking. In the framework, section 4 presents a method for human body tracking using multiple cameras and floor sensors. Then section 5 presents extension to head tracking. We mention experimental results of the proposed method in section 6. Finally, conclusion is discussed in section 7.

2   Particle Filter

The particle filter is a Bayesian sequential importance sampling technique, which recursively approximates the posterior distribution using a finite set of weighted samples. The posterior distribution at time t is approximated by a set of discrete samples $\{ s_t^{(n)} \}$ $(n = 1...N)$ with importance weights $\{ \pi_t^{(n)} \}$ $(n = 1...N)$. Particle filter simulates this distribution by the following three-step recursion.

1. Selection: Select samples $\{ s_{t-1}^{'(n)} \}$ $(n = 1...N)$ in proportion to weight $\{ \pi_{t-1}^{(n)} \}$ $(n = 1...N)$ corresponding to sample $\{ s_{t-1}^{(n)} \}$ $(n = 1...N)$.

2. Prediction: Propagate samples $\{ s_{t-1}^{'(n)} \}$ $(n = 1...N)$ with state transition probability $p(x_t|x_{t-1} = s'_{t-1})$ and generate new samples $\{ s_t^{(n)} \}$ $(n = 1...N)$ at time t.

3. Update: Update weights $\pi_t^{(n)} = p(y_t|x_t = s_t^{(n)})$ corresponding to sample $s_t^{(n)}$ by evaluating a likelihood through observations. Normalize $\pi_t^{(n)}$ so that the sum of $\{ \pi_t^{(n)} \}$ $(n = 1...N)$ is equal to 1. As a result, estimated state at time t is equal to the expectation of the set of samples $\{ (s_t^{(n)}; \pi_t^{(n)}) \}$ $(n = 1...N)$.

3   Multiple persons tracking

Traditional particle filters perform poorly at consistently maintaining the multi-modality in the target distribution that often results from multiple targets. Vermaak et al.[13] introduced a mixture particle filter (MPF), where each component is modeled with an individual particle filter that forms part of the mixture. The MPF enables tracking multiple targets simultaneously. However, this method is inappropriate for multi-target 3D tracking that requires locating the targets and labeling their identities.

In this paper, our method is composed of multiple single-target particle filters as the MPF. Each of the filters tracks the same single-target and labels same identity through all tracking by holding the target-specific information.

Moreover, for discriminating the distribution of multiple filters, we applied the exclusion occlusion model as the interaction between multiple filters.

3.1   Target-specific information and detection

Each of the filters holds the target-specific information provided by cameras. In our method, we obtain color histogram of the person as the target-specific information and use it in the samples evaluation. The reference color histogram is gathered from the pixels detected by background subtraction of each camera image. When detecting a person, the filter obtains color histogram of the person and labels identity. The filter then starts tracking.

Person is detected by floor sensors. The sum of weight on floor sensors is in proportion to the number of persons, and floor sensors enable correct detection.

3.2   Occlusion exclusion model

When tracking multiple persons, samples of multiple particle filters are often mixed so that multiple particle filters track the same person by mistake. The remedy for this problem is exclusive segmentation of the targets' region. We then proposed the occlusion exclusion model.

When samples of the filters tracking a different person respectively overlap on the ground plane, the confidence of the samples is not high. Then we give samples the square area which side is 100mm, and the weights of the area-overlapped samples are 0. The length of the side is determined by the size of a person. Such method enables exclusive segmentation. Fig.1 shows the occlusion exclusion model.
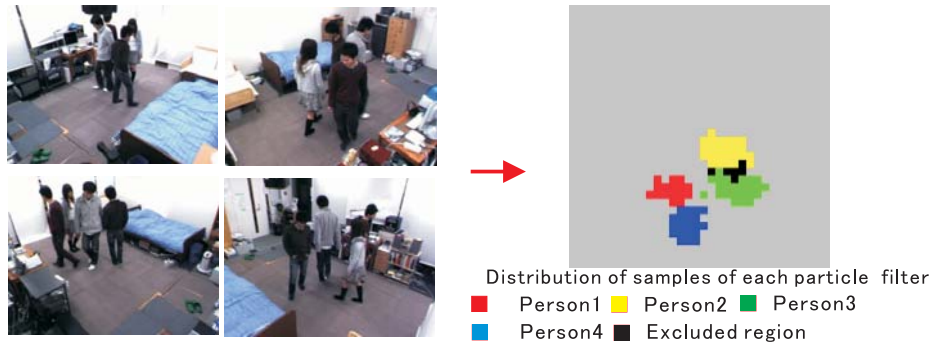
Distribution of samples of each particle filter
- Person1 Person2 Person3
- Person4 Excluded region

Fig. 1. Occlusion exclusion model

4   Tracking by integration of cameras and floor sensors

Our method uses particle filter framework to track a single human body. We estimate human body position, which is represented as a 3D vector $(x, y, z)^T$ in a 3D state. In the framework, we integrate both information from floor sensors and the target-specific information from multiple cameras. We show a transition model and a observation model which are important when implementing the particle filter.

4.1   Transition model

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

$$s_t = s'_{t-1} + \tau v_{t-1} + \omega, \tag{1}$$

$$v_t = \frac{p_t - p_{t-1}}{\tau}, \tag{2}$$

where τ is the time interval between frames, v_{t-1} is the previous velocity of the target, ω is a system noise added to s'_{t-1}, and p_t is the estimated target's position at time t. We control the diffusion factor ω adaptively by the velocity of the target. Such control of the system noise ω contributes to improving the robustness against sudden abrupt motion and the accuracy of the estimation.

4.2   Observation model for body tracking

In the observation model, cameras and floor sensors are integrated.

In the model, the evaluations of samples are hierarchically executed. The hierarchical evaluation enables to decrease calculation cost. The concrete evaluation procedures are given below.

1. Eliminate a sample when a location of the sample is out of the room.

2. Evaluate a weight $\pi_{f,t}$ by floor sensors as like section 4.3. When the evaluation score is 0, the sample is eliminated.
3. Project a sample into the i-th camera coordinates. When the sample is out of the field of view of the camera, the sample is eliminated.
4. Evaluate a weight $\pi_{bg,i,t}$ by background subtractions of the i-th camera image. The evaluation score is binary. When the evaluation score is 0, the sample is eliminated.
5. Evaluate a weight $\pi_{clr,i,t}$ by color histogram of the i-th camera image as like section 4.4.
6. Repeat the procedures from step 3 to 5 for each camera. Compute a weight $\pi_{b,t}$ for body tracking at time t by multiplying these likelihoods. A weight $\pi_{b,t}$ is $\pi_{f,t} \prod_i \pi_{bg,i,t} \prod_i \pi_{clr,i,t}$.

We incorporate the estimation using floor sensors into the particle filter framework and tracking with data fusion of multiple cameras and floor sensors is realized by multiplying the weights. Floor sensors are resistant to environmental changes and are not influenced by occlusion. Even when the multiple persons are in close proximity and occlusion ambiguities occur in camera images, the evaluation using floor sensors enables estimating true position.

In the evaluation by background subtractions, we improve the background subtraction process to avoid eliminating correct samples We perform background subtraction in HSV color space. After the background subtraction, we use erosions and dilations of the image to decrease noise. Moreover, we remove small regions through the contour process.

The evaluation using color histogram in camera images is indispensable for tracking the correct target. Color histogram gives the target-specific information and enables labeling their identities correctly. The evaluation also contributes to the estimation of human position in high resolution.

Evaluation using floor sensors and evaluation using color histogram are described in the following section.

4.3   Evaluation of samples using floor sensors

The weight $\pi_{f,t}$ is evaluated by weight on floor sensors. The weight $\pi_{f,t}$ is computed by

$$\pi_{f,t} = \frac{f_t(x,y)}{\sum_{m \in L} f_t^m}, \qquad (3)$$

where $f_t(x,y)$ is the weight at a location (x,y) and is included in the cluster given a label L in floor sensors. m is the elements of the cluster L. $\sum_{m \in L} f_t^m$ is the weight of the cluster L and $\pi_{f,t}$ is the proportion of the weight at a location (x,y) to the weight of the cluster L.

Such evaluation contributes to a dense set of samples around the true position, which consequently enables the amelioration of the estimation accuracy and the tracking rate of multiple persons. Fig.2 shows floor sensors images and camera images. Here, we convert weight of floor sensors into gray scale image.

Fig. 2. Weight on floor sensors

4.4   Evaluation of samples using color histogram

Our method uses the data similarity between the reference color histogram p obtained in the initialization and a candidate color histogram q around a sample for samples evaluation. A candidate color histogram q is gathered within a 30×90 pixel-sized region set around a sample.

Following Perez's method[6], our method adopt a sample evaluation based on HSV color histograms. HSV is insensitive to illumination effects. Color information is however only reliable when both the saturation and the value are not too small. Hence, we populate an HS histogram with $N_H N_S$ bins using only the pixels with saturation and value larger than two thresholds set to 0.1 and 0.07 respectively in our experiments. The remaining "color-free" pixels can however retain crucial information when tracked regions are mainly black and white. We thus populate N_V additional value-only bins with them. The resulting histogram is composed of $N = N_H N_S + N_V$ bins, where N_H, N_S, N_V are set to 10.

We apply the Bhattacharyya similarity coefficient. We thus compute the weights $\pi_{clr,i,t}$ by color histogram in the i-th camera image as below.

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

$$\pi_{clr,i,t} \propto e^{-\lambda D^2(p,q)}, \qquad (5)$$

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

5   Extension to head tracking

Our method estimates not only human body position but also human head position. In the framework of single person tracking, to track both body and head simultaneously, we develop a new particle filter framework.

5.1   Head tracking using estimated body position

In addition to a sample set represents body state, another sample set represents head state $s_{h,t} = (x, y, z, \theta)^T$ is added, where θ denotes the pan of the head. In

our method, the filtering step for head tracking is also repeated individually. In sampling and evaluation step, we link head tracking to body tracking by using the estimated body position. By using the estimated body position, tracking performance when the head moves up and down is significantly improved. Moreover, the link with body tracking enables labeling their identities of the head.

First, we show the single-target tracking procedures below.

1. Apply the exclusion occlusion model to the sample set for body tracking $\{ s_{b,t}^{(n)} \}$ $(n = 1...N)$ generated at time $t - 1$.
2. Update weights $\{ \pi_{b,t}^{(n)} \}$ $(n = 1...N)$ as like section 4.2.
3. Estimate the body position $p_{b,t}$ at time t.
   (a) Generate the new sample set $\{ s_{h,b,t}^{(n)} \}$ $(n = 1...\alpha M)$ for head tracking around the body position $p_{b,t}$ and propagate the samples with body-head model h_{t-1}.
   (b) Mix samples $\{ s_{h,b,t}^{(n)} \}$ $(n = 1...\alpha M)$ with samples $\{ s_{h,r,t}^{(n)} \}$ $(n = 1...(1 - \alpha)M)$ generated at time $t - 1$ by resampling, and generate the sample set $\{ s_{h,t}^{(n)} \}$ $(n = 1...M)$.
   (c) Update weights $\{ \pi_{h,t}^{(n)} \}$ $(n = 1...M)$ as like section 5.3.
   (d) Estimate the head position $p_{h,t}$ at time t.
   (e) Select samples $\{ s_{h,r,t}^{'(n)} \}$ $(n = 1...\alpha M)$ in proportion to weight $\{ \pi_{h,t}^{(n)} \}$ $(n = 1...M)$.
   (f) Propagate samples $\{ s_{h,r,t}^{'(n)} \}$ $(n = 1...\alpha M)$ with state transition probability $p(x_{t+1}|x_t)$ and generate new samples $\{ s_{h,r,t+1}^{(n)} \}$ $(n = 1...\alpha M)$.
4. Select samples $\{ s_{b,t}^{'(n)} \}$ $(n = 1...N)$ for body tracking in proportion to weight $\{ \pi_{b,t}^{(n)} \}$ $(n = 1...N)$.
5. Propagate samples $\{ s_{b,t}^{'(n)} \}$ $(n = 1...N)$ with state transition probability $p(x_{t+1}|x_t)$ and generate new samples $\{ s_{b,t+1}^{(n)} \}$ $(n = 1...N)$.

The details of the sampling and evaluation using body position are described in the following section.

5.2   Sampling using body-head model

In the sampling step of samples for head tracking, a new sample set is composed of the mixture of samples $s_{h,r,t}$ generated by normal resampling step and samples $s_{h,b,t}$ generated around the position based on the estimated body position.

Samples $s'_{h,b,t}$ are generated around the estimated body position and propagated with body-head model h_{t-1}. Body-head model h_t denotes the distance between head and body of the target at time t. Samples $s_{h,b,t}$ are given by

$$s_{h,b,t} = s'_{h,b,t} + h_{t-1} + \omega, \tag{6}$$

$$h_t = p_{h,t} - p_{b,t}, \tag{7}$$

where ω is a 4D Gaussian noise which is adaptively controlled by the velocity of the target as like section 4.1.

Samples $s_{h,r,t}$ are selected in proportion to weight $\pi_{h,t}$ and propagated with state transition probability $p(x_{t+1}|x_t)$. Transition model $p(x_t|x_{t-1})$ is the same as that of body tracking in Eq.1.

5.3   Observation model for head tracking

In the observation model, the evaluation of samples $s_{h,t}$ by using the estimated body position is incorporated. The evaluation procedures are given below.

1. Eliminate a sample when a location of the sample is out of the room.
2. Evaluate a likelihood by distance from body position. When the evaluation score is 0, the sample is eliminated.
3. Project a sample into camera coordinates. When the sample is out of the field of view of the camera, the sample is eliminated.
4. Evaluate a likelihood by background subtractions of the camera image. The evaluation score is binary. When the evaluation score is 0, the sample is eliminated.
5. Evaluate a likelihood by color histogram.
6. Repeat the procedures from step 3 to 5 for each camera. Compute a weight $\pi_{h,t}$ for head tracking at time t by multiplying these likelihoods.

In the second step of the evaluation, the distance from body position is used. Such evaluation enables head tracking linked to body tracking.

The evaluation of color histogram is based on that of body tracking. Our method uses the data similarity between the reference color histogram obtained in the initialization and a candidate color histogram around a sample as like section 4.4. The reference color histogram is changed into histogram around human head. A candidate color histogram is gathered within a 20×20 pixel-sized region set around a sample. In the evaluation of the sample $s_{h,t}$, the reference color histogram corresponding to the pan θ of the sample is selected from multiple reference color histograms obtained by multiple camera images.

6   Experimental Results

6.1   The tracking system

Floor pressure sensors are distributed on the floor at 200mm intervals. There are 252 pressure sensors. This floor detects pressure data with 200mm spatial resolution. Force Sensing Resisters (FSRs) are utilized as the pressure sensor. Pressure can be converted into the voltage value, because a FSR has a character that its electric resistance decreases by the increase in the power added. Pressure sensitivity range is $10kg/cm^2$. It produces an 8-bit output and can detect human existing areas by finding high-pressure areas. Sampling rate of the pressure sensors is 10Hz.

We used four cameras on the four corners of the ceiling of the room. The camera is IEEE1394camera DragonFly2. Each image was captured at a resolution of 320×240. 4 server PCs (Dell Precision 690, Intel Xeon Processor 5060 3.20 GHz × 2, 4GB Memory) were prepared, and a camera was connected to each server PC. Cameras were completely synchronized during capture, and intrinsic / extrinsic camera parameters were calculated beforehand.

A client PC was also prepared in the same specification. The client PC integrated data of two sensors and executed the tracking process. In this configuration, we processed 150 samples represent head state and 250 samples represent body state per one particle filter within 50ms when three persons tracking. The tracking runs online at a frame rate of 20 Hz. Fig.3 shows their configuration.
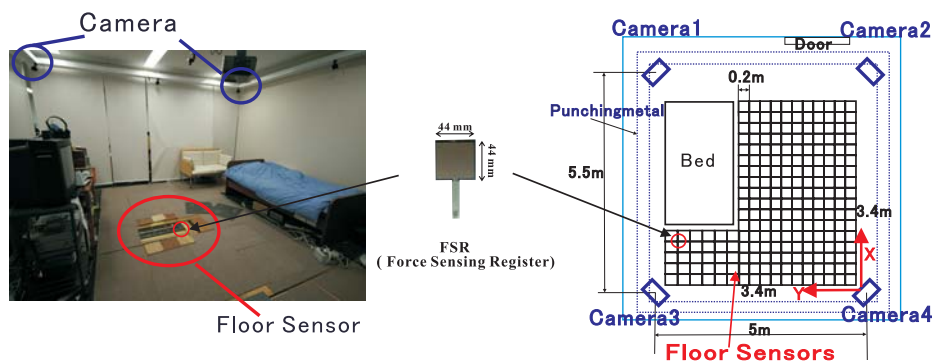

Fig. 3. Configuration of cameras and floor sensors

6.2   Tracking multiple persons

To demonstrate the effectiveness and robustness of the method against the complicated movements of multiple persons, we performed several experiments. In the experiments, we assumed a variety of movements of multiple persons: crossing, turning after coming close, going around, and going back after coming close.

Fig.4 shows the tracking results against two movements: crossing, going back after coming close. Tracking against going back after coming close is very difficult because the movement is contrary to the transition model. Even against the complicated movements, our method was able to estimate positions of multiple persons and label their identities correctly. This is attributed to data fusion of multiple cameras and floor sensors. When the multiple persons are in close proximity and occlusion ambiguities occur in camera images, floor sensors give true positions of the persons and cameras gives the target-specific information.

Next, the tracking experiment in a home environment was performed. Fig.5 shows an example of the images obtained in the experiment. We can see from Fig.5 that the human head position is well estimated even when the head moves up and down.
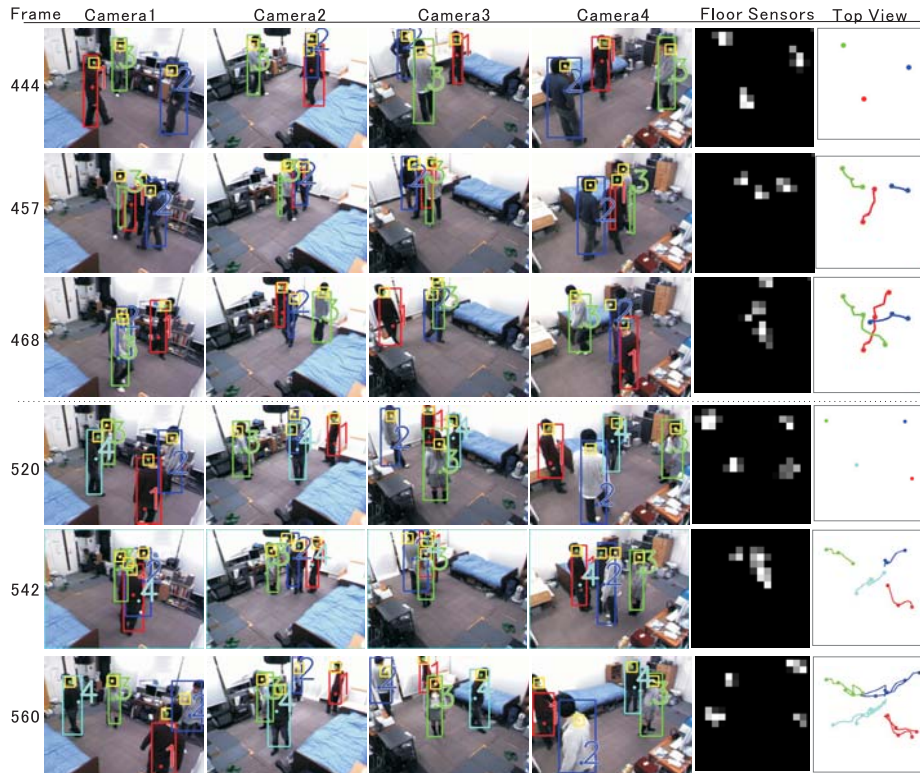
Fig. 4. Tracking of multiple persons

6.3   Estimation errors of human position

To investigate the estimation accuracy of the human body by the proposed method, we conducted comparative experiments. We confirm the effectiveness of the evaluation using floor sensors and the evaluation using color histogram. However, it is difficult to measure human body positions accurately. Therefore, we assumed that a test subject walks along the straight test course precisely. Fig.6 shows the trajectory of the estimated body position by three comparative methods. We can see from Fig.6 that the estimation accuracy is sufficient to estimate the body position. The effectiveness of the evaluation using color histogram can be also confirmed.

Next, in order to investigate quantitatively the estimation accuracy of the human head by the proposed method, the human head in the image was manually specified. The 3D coordinates obtained by inverse projection from 4 camera images was assumed to be the true position, and was compared to the result of estimation. The means of the error on the X, Y and Z axis are 4.9, 5.4 and 4.5 cm respectively. These errors are below the radius of human head and the tracking can be considered sufficiently accurate.
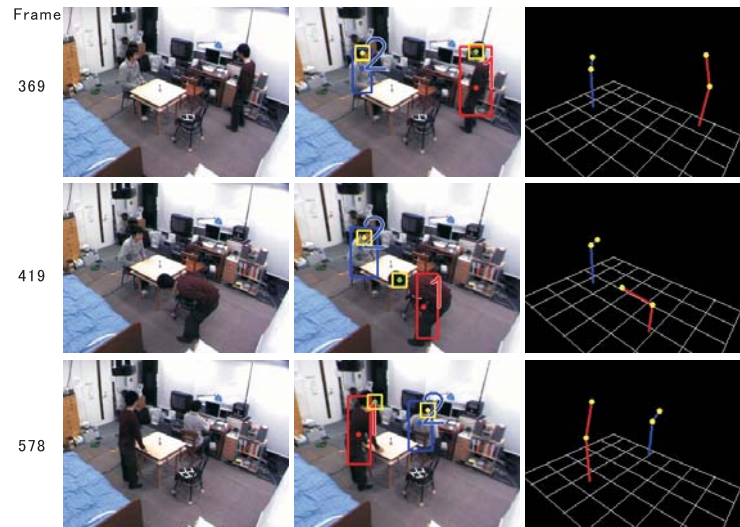
Fig. 5. Tracking of multiple human heads in indoor environment

7   Conclusion

In this work, we have proposed a method for tracking multiple persons by integrating multiple cameras and floor sensors. Our method estimates positions of human body and head, and labels their identities. Our method is composed of multiple particle filters, and each filter tracks each person correctly by integrating information from floor sensors and target-specific information from cameras.

Successful tracking was demonstrated in a tracking experiment for multiple persons. Even when the targets are in close proximity and move complicatedly, our method was able to estimate positions of the targets and label their identities correctly. This is due to the integration of multiple cameras and floor sensors. Because floor sensors are resistant to environmental changes and is not influenced by occlusion, the evaluation using floor sensors enabled estimating true position. Camera provides a great deal of information such as color histogram and background subtraction, and the evaluation using color histogram enabled labeling their identities correctly. It was shown that robust tracking of multiple persons was realized by the integration of these two types of sensors.

References

1. Kidd, C.D., Orr, R., Abowd, G.D., Atkeson, C.G., Essa, I.A., MacIntyre, B., Mynatt, E., Starner, T.E., Newstetter, W.: The Aware Home: A Living Laboratory for Ubiquitous Computing Research. In: CoBuild. (1999) 191–198
2. Mori, T., Noguchi, H., Takada, A., Sato, T.: Sensing Room: Distributed Sensor Environment for Measurement of Human Daily Behavior. In: INSS. (2004) 40–43
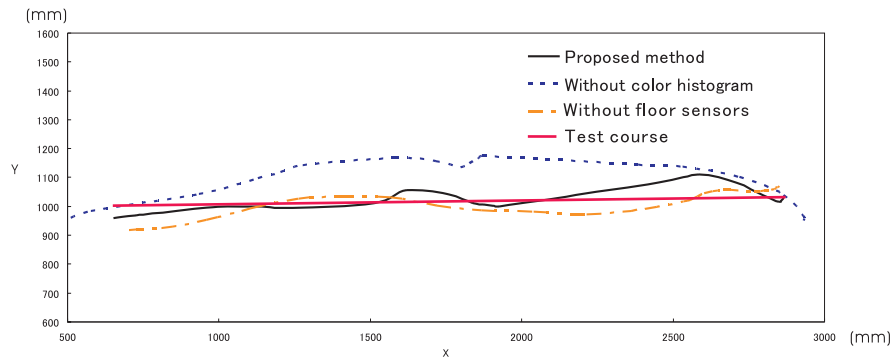
Fig. 6. Walking trajectory of a person

3. Kobayashi, Y., Sugimura, D., Sato, Y., Hirasawa, K., Suzuki, N., Kage, H., Sugimoto, A.: 3d head tracking using the particle filter with cascaded classifiers. In: The British Machine Vision Conference. (2006) 37–46

4. Kim, K., Davis, L.S.: Multi-camera tracking and segmentation of occluded people on ground plane using search-guided particle filtering. In: ECCV (3). (2006) 98–109

5. Yang, C., Duraiswami, R., Davis, L.: Fast multiple object tracking via a hierarchical particle filter. In: International Conference on Computer Vision and Pattern Recognition. Volume 1. (2005) 212–219

6. Perez, P., Hue, C., Vermaak, J., Gangnet, M.: Color-based probabilistic tracking. In: European Conference on Computer Vision(ECCV2002). (2002) 661–675

7. Mori, T., Suemasu, Y., Noguchi, H., Sato, T.: Multiple people tracking by integrating distributed floor pressure sensors and rfid system. In: IEEE International Conference on System Man and Cybernetics. (2004) 5271–5278

8. Fod, A., Howard, A., Mataric, M.J.: A laser-based people tracker. In: IEEE International Conference on Robotics and Automation. (2002) 3024–3029

9. Addlesee, M.D., Jones, A.H., Livesey, F., Samaria, F.S.: The orl active floor. IEEE Personal Communications **4** (1997) 35–41

10. Fukui, R., Morishita, H., Sato, T.: Expression method of human locomotion records for path planning and control of human-symbiotic robot system based on spacial existence probability model of humans. In: International Conference on on Robots and Automation(ICRA2003). (2003) 4178–4184

11. Murakita, T., Ikeda, T., Ishiguro, H.: Human tracking using floor sensors based on the markov chain monte carlo method. In: International Conference on Pattern Recognition(ICPR2004). Volume 4. (2004) 917– 920

12. Isard, M., Blake, A.: Condensation-conditional density propagation for visual tracking. Int.J.Com-put **29** (1998) 5–28

13. Vermaak, J., and Patrick Perez, A.D.: Maintaining multi-modality through mixture tracking. In: International Conference on Computer Vision. (2003) 1110–1116

14. Du, W., Piater, J.: Tracking by cluster analysis of feature points using a mixture particle filter. In: AVSBS05. (2005) 165–170

15. Dore, A., Cattoni, A.F., Regazzoni, C.S.: 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). (2007) 99–104