Multiple-Person Tracking for a Mobile Robot using Stereo

Junji Satake Jun Miura
Toyohashi University of Technology
1-1 Hibarigaoka, Tempaku-cho, Toyohashi, Aichi 441-8580, Japan
{satake, jun}@ics.tut.ac.jp

Abstract

This paper describes a method of detecting and tracking two or more persons from images captured with a moving camera. We propose a method of person tracking using distance information calculated with a stereo camera. Each person's position is estimated by using Kalman filter in consideration of movement of a mobile robot and persons. Additionally, we report an experimental result on control of mobile robot following a specific person.

1 Introduction

Following a specific person is an important task for service robots. Visual person following in public spaces entails tracking of multiple persons by a moving camera. Though there are a lot of researches on person detection and tracking from images [1, 2, 3], many of them use a fixed camera. In the case of using moving camera, it is one of important problems to separate moving objects from the background. By using distance information obtained from a stereo camera, Beymer and Konolige [4] detect persons quickly, and Kang et al. [5] track moving objects. There is a system that detects pedestrians from a moving vehicle [6], and the segmentation is relatively easy because the camera moves usually regularly along a road.

In this paper, we propose a method of detecting and tracking multiple persons for a mobile robot. Laser range finders are widely used to detect obstacles for mobile robots, and several researches use them for person detection and tracking [7, 8, 9]. Image information, such as color and texture, is, however, necessary for segmenting and/or identifying each person. Several person tracking methods for mobile robots are proposed which use an omnidirectional camera [10, 11, 12] or a stereo camera [13, 14, 15]. However, it is difficult to analyze a complex scene using an omnidirectional camera due to its resolution. Therefore, these works detect person regions by using distance information obtained from stereo camera, and track persons by using appearance models. On the other hand, we propose a method to track persons by using silhouette models. We also present an EKF-based algorithm that considers movement of persons and a mobile robot, and can therefore track persons correctly even when they are occluded by other persons for a short period of time.

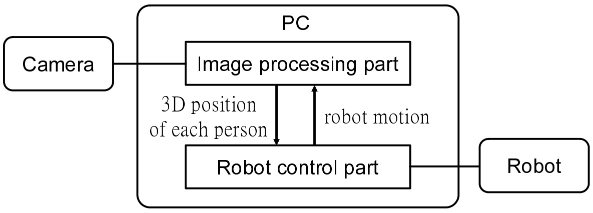


Figure 1: Configuration of our system.

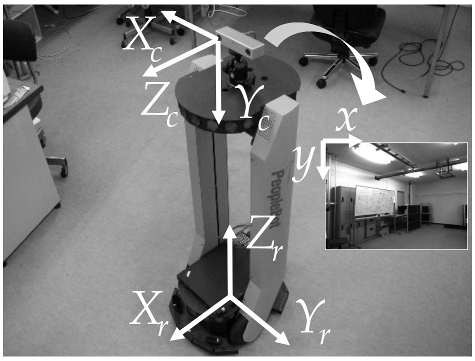


Figure 2: Definition of coordinate systems.

2 Person Tracking using Stereo Camera

2.1 Configuration of our system

Our system consists of the following two parts as shown in Fig. 1. The image processing part estimates the 3D position of each person using stereo. Using this estimated position information, the robot control part controls the robot motion. The robot motion information is sent back to the image processing part to be considered in person tracking. The detail of robot control to follow a specific person is explained in Section **3**.

Figure 2 illustrates the coordinate systems attached to our mobile robot and stereo system. The relation between the robot and the camera coodinate system is given by

$$Z_c \left[\, x \;\; y \;\; 1 \,\right]^T = \boldsymbol{A} \left[\, \boldsymbol{R} \mid \boldsymbol{T} \,\right] \left[\, X_r \;\; Y_r \;\; Z_r \;\; 1 \,\right]^T , \quad (1)$$

where $\boldsymbol{A}$, $\boldsymbol{R}$ and $\boldsymbol{T}$ show the intrinsic parameters matrix, the rotation matrix, and the translation vector, respectively.

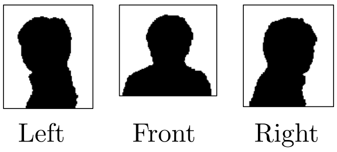
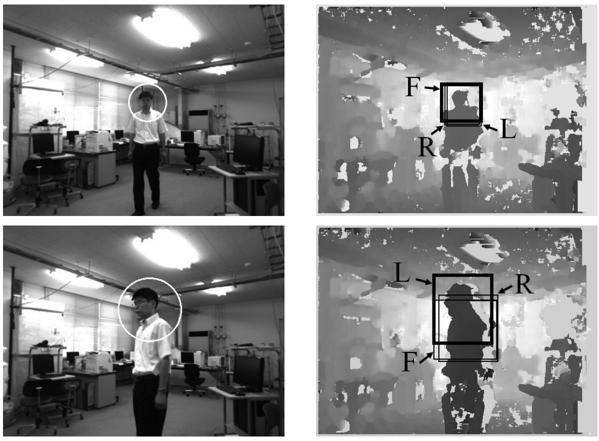


Figure 3: Distance templates.



(a) Input images (b) Depth images

Figure 4: Examples of tracking result using distance templates.

2.2 Person tracking based on distance information

To track persons stably with a moving camera, we use *distance templates*, which are the templates for human upper bodies in depth images (see Fig. 3). We made the templates from the depth images where the target person was at 2m away from the camera. When the templates are used for person detection, their sizes and the depth values are adjusted accoding to the distance from the camera to the target person. By matching the template to the depth image, the 2D position of person (x, y), the distance from camera D, and the evaluation value are obtained. In addition, we use three templates with different direction of body, and take the one with the highest evaluation value as a tracking result in each frame. The tracking algorithm is described in Section **2.4**.

Figure 4 shows examples of tracking using the distance templates. Three rectangles in each depth image are tracking results with the three templates, and the one with the highest evaluation value is shown in bold line. Even when the direction of the body changed, it is possible to track person stably by using multiple templates.

Objects with similar silhouette to person, shown in Fig. 5, are sometimes detected using only depth information. We reject such false detections using color and texture information.

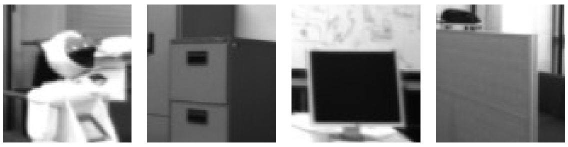


Figure 5: Examples of false detection.

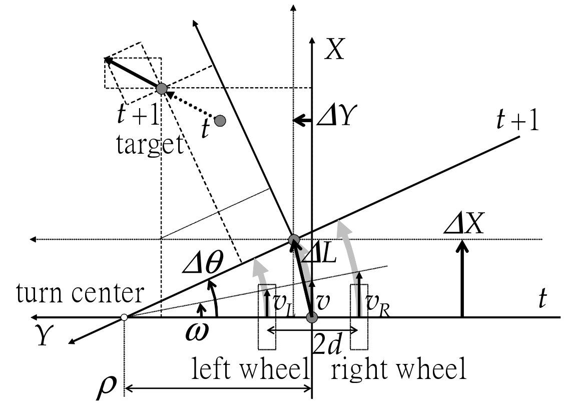


Figure 6: Control of wheeled mobile robot.

2.3 Estimation of 3D position using EKF

2.3.1 State equation

In the robot coordinate system, the person's position at time t is defined as (X_t, Y_t, Z_t). The state variable $\boldsymbol{x}_t$ is defined as

$$\boldsymbol{x}_t = \begin{bmatrix} X_t & Y_t & Z_t & \dot{X}_t & \dot{Y}_t \end{bmatrix}^T,$$

where $\dot{X}_t$ and $\dot{Y}_t$ denote velocities in the horizontal plane.

We first consider the case where the robot does not move. The system equation is given by

$$\boldsymbol{x}_{t+1} = \boldsymbol{F}_t \boldsymbol{x}_t + \boldsymbol{G}_t \boldsymbol{w}_t \tag{2}$$

where

$$\boldsymbol{F}_t = \begin{bmatrix} 1 & 0 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & 0 & \Delta t \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{G}_t = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ \Delta t & 0 \\ 0 & \Delta t \end{bmatrix},$$

$$\boldsymbol{Q}_t = cov(\boldsymbol{w}_t) = E\left[\boldsymbol{w}_t \boldsymbol{w}_t^T\right] = \sigma_w^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

We then consider the case where the robot moves. Figure 6 shows how a wheeled mobile robot moves. The distance of two wheels is denoted as $2d$. When each wheel rotates with speed v_L and v_R, the velocity v, the angular velocity ω, and the turning radius ρ of the robot have the following relation. First, we have

$$v = \rho\omega.$$

For the both wheels, we also have

$$v_L = (\rho - d)\omega, \quad v_R = (\rho + d)\omega.$$

From these equations, we have

$$v = (v_R + v_L)/2, \quad \omega = (v_R - v_L)/2d,$$
$$\rho = d(v_R + v_L)/(v_R - v_L).$$

The rotation angle $\Delta\theta$ and the moved distance ΔL during time Δt are obtained respectively as

$$\Delta\theta = \omega\Delta t, \quad \Delta L = 2\rho\sin(\Delta\theta/2).$$

In addition, the robot movement ΔX and ΔY seen from the robot position at time t are obtained respectively as

$$\Delta X = \Delta L\cos(\Delta\theta/2), \quad \Delta Y = \Delta L\sin(\Delta\theta/2).$$

Therefore, the person position $\boldsymbol{x}_{t+1}^{(t+1)}$ at time $t+1$ seen from the robot position at time $t+1$ is shown as follows by the use of the person position $\boldsymbol{x}_{t+1}^{(t)}$ at time $t+1$ seen from the robot position at time t.

$$\begin{aligned}
X^{(t+1)} &= (X^{(t)} - \Delta X)\cos\Delta\theta + (Y^{(t)} - \Delta Y)\sin\Delta\theta,\\
Y^{(t+1)} &= -(X^{(t)} - \Delta X)\sin\Delta\theta + (Y^{(t)} - \Delta Y)\cos\Delta\theta,\\
\dot{X}^{(t+1)} &= \dot{X}^{(t)}\cos\Delta\theta + \dot{Y}^{(t)}\sin\Delta\theta - v,\\
\dot{Y}^{(t+1)} &= -\dot{X}^{(t)}\sin\Delta\theta + \dot{Y}^{(t)}\cos\Delta\theta.
\end{aligned}$$

By the combination of these equations and equation (2), the state equation that considers the robot movement $\boldsymbol{u}_t = [v_L \ v_R]^T$ is expressed as

$$\boldsymbol{x}_{t+1} = \boldsymbol{f}_t(\boldsymbol{x}_t, \boldsymbol{u}_t) + \boldsymbol{G}_t\boldsymbol{w}_t, \qquad (3)$$

where

$$\boldsymbol{f}_t(\boldsymbol{x}_t, \boldsymbol{u}_t) = \left[\begin{array}{c}
(X_t + \Delta t\dot{X}_t - \Delta X)\cos\Delta\theta + (Y_t + \Delta t\dot{Y}_t - \Delta Y)\sin\Delta\theta\\
-(X_t + \Delta t\dot{X}_t - \Delta X)\sin\Delta\theta + (Y_t + \Delta t\dot{Y}_t - \Delta Y)\cos\Delta\theta\\
Z_t\\
\dot{X}_t\cos\Delta\theta + \dot{Y}_t\sin\Delta\theta - v\\
-\dot{X}_t\sin\Delta\theta + \dot{Y}_t\cos\Delta\theta
\end{array}\right].$$

2.3.2 Observation equation

The observed person's position in the robot coordinate system is denoted as $\boldsymbol{y}_t$. The observation equation is expressed as

$$\boldsymbol{y}_t = \boldsymbol{H}_t\boldsymbol{x}_t + \boldsymbol{v}_t, \qquad (4)$$

where

$$\boldsymbol{y}_t = \begin{bmatrix} X_r \\ Y_r \\ Z_r \end{bmatrix}, \quad \boldsymbol{H}_t = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix},$$

$$\boldsymbol{R}_t = cov(\boldsymbol{v}_t) = E\left[\boldsymbol{v}_t\boldsymbol{v}_t^T\right] = \sigma_v^2 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

2.3.3 Kalman filter

Based on the state equation (3) and the observation equation (4), the person's position can be estimated by using the Extended Kalman Filter (EKF) shown by the following equations.

$$\begin{aligned}
\hat{\boldsymbol{x}}_{t|t-1} &= \boldsymbol{f}_t(\hat{\boldsymbol{x}}_{t-1|t-1}, \boldsymbol{u}_{t-1}), && (5)\\
\boldsymbol{P}_{t|t-1} &= \boldsymbol{F}_t\boldsymbol{P}_{t-1|t-1}\boldsymbol{F}_t^T + \boldsymbol{G}_t\boldsymbol{Q}_{t-1}\boldsymbol{G}_t^T, && (6)\\
\boldsymbol{K}_t &= \boldsymbol{P}_{t|t-1}\boldsymbol{H}_t^T\left[\boldsymbol{H}_t\boldsymbol{P}_{t|t-1}\boldsymbol{H}_t^T + \boldsymbol{R}_t\right]^{-1}, && (7)\\
\hat{\boldsymbol{x}}_{t|t} &= \hat{\boldsymbol{x}}_{t|t-1} + \boldsymbol{K}_t\left[\boldsymbol{y}_t - \boldsymbol{H}_t\hat{\boldsymbol{x}}_{t|t-1}\right], && (8)\\
\boldsymbol{P}_{t|t} &= \boldsymbol{P}_{t|t-1} - \boldsymbol{K}_t\boldsymbol{H}_t\boldsymbol{P}_{t|t-1}, && (9)
\end{aligned}$$

where

$$\boldsymbol{F}_t = \left.\frac{\partial\boldsymbol{f}_t}{\partial\boldsymbol{x}_t}\right|_{\hat{\boldsymbol{x}}_{t|t-1}, \boldsymbol{u}_t}$$
$$= \left.\begin{bmatrix}
\cos\Delta\theta & \sin\Delta\theta & 0 & \Delta t\cos\Delta\theta & \Delta t\sin\Delta\theta\\
-\sin\Delta\theta & \cos\Delta\theta & 0 & -\Delta t\sin\Delta\theta & \Delta t\cos\Delta\theta\\
0 & 0 & 1 & 0 & 0\\
0 & 0 & 0 & \cos\Delta\theta & \sin\Delta\theta\\
0 & 0 & 0 & -\sin\Delta\theta & \cos\Delta\theta
\end{bmatrix}\right|_{\boldsymbol{u}_t}.$$

2.4 Tracking algorithm

The image processing part (see Fig. 1) works as follows:

1) Stereo processing The depth image is made with a stereo camera.

2) Person tracking Each person is tracked by using the EKF described in Section **2.3**.

2.1) Prediction The 3D position at the current time t is predicted from the state variable at the previous time $t-1$ by equation (5). And then, it is projected to 2D image by equation (1).

2.2) Observation The person is searched for around the predicted position by the method described in Section **2.2**. The templates used to search are made based on the distance to the person. After the search, the person's 3D position $\boldsymbol{y}_t$ is calculated by equation (1) based on image coordinates (x, y) and distance from camera $Z_c = D$.

2.3) Update The state variable is updated by equation (8).

3) Detection The persons who appear newly in image are detected with distance templates.

4) Communication The estimated position is sent to the robot control part, and the rotational speeds of the left and right wheels are received.

Even when persons with different distances from the camera occlude other person, it is possible to track each person because the size and the pixel value of the template are set based on each person's 3D position.

3 Control to Follow a Specific Person

The robot with two-wheel drive can follow a circular trajectory from the current to the target position (path A in Fig. 7). In this case, the speeds for the wheels to

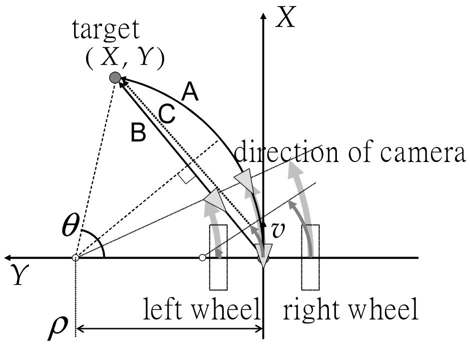


Figure 7: Path to target position.

move the robot at velocity v is calculated as follows. From the equation

$$\rho^2 = \left\{ \left(\frac{X}{2}\right)^2 + \left(\frac{Y}{2}\right)^2 \right\} + \left\{ \left(\frac{X}{2}\right)^2 + \left(\rho - \frac{Y}{2}\right)^2 \right\},$$

we have

$$\rho \ = \ \frac{X^2+Y^2}{2Y}.$$

Then we can calculate the velocities as:

$$\begin{aligned} v_L &= v\left(1-\tfrac{d}{\rho}\right) &&= v\left(1-\tfrac{2dY}{X^2+Y^2}\right), \\ v_R &= v\left(1+\tfrac{d}{\rho}\right) &&= v\left(1+\tfrac{2dY}{X^2+Y^2}\right). \end{aligned}$$

When the robot follows this circular path, since the turning rate of robot orientation is relatively slow, the target person tends to go out of the field of view. On the other hand, the robot first turns and then moves straight toward the target like path B, the robot movement is not smooth. We thus use the one like path C, on which the robot turns to the target while moving ahead. In this case, the velocity of each wheel is adjusted as follows:

$$v_L = v\left(1-k\tfrac{2dY}{X^2+Y^2}\right), \quad v_R = v\left(1+k\tfrac{2dY}{X^2+Y^2}\right).$$

This means the turning radius ρ is reduced to ρ/k.

4 Experimental Result

In the experiment, we used a wheel movement robot and a stereo camera shown in Fig. 2. The robot is PeopleBot (max speed 900mm/sec) made by the Mobile Robots company, and it is controlled by using serial communication from a note PC (ThinkPad, Intel Core2Duo 2.6GHz) on the robot. The stereo camera is Bumblebee2 made by the Point Grey Research company (XGA, 20fps, 100-deg HFOV), and it was set up in the upper part of the robot.

We implemented the software modules for person detection and tracking, motion planning, and robot control as *RT components* in the *RT-middleware* environment [16] for easier development and maintenance.

Figure 8 shows a result of tracking. The left row images are the results of person detection. Each circle in the image shows the result of observation with distance templates, and each small point shows the 3D head position estimated using EKF. The right row images show the positions of the robot and the persons taken by a ceiling camea. In addition, the curves in the final frame (#156) shows the traces of the robot and the persons. The processed image size is 512 × 384, and the processing speed is about 70msec/frame. The robot moved toward person A who was detected first. Even when person B and C passed between the robot and person A, the target person was correctly tracked.

5 Conclusions

In this paper, we described a method of detecting and tracking multiple persons for a mobile robot by using distance information calculated with the stereo camera. We presented an EKF-based algorithm that considers movement of persons and the robot and can therefore track persons correctly even when they are occluded by other persons for a short period of time. Finally, we reported an experimental result on control of mobile robot following a specific person. In the future work, we will plan the path of the robot in consideration of moving obstacles.

Acknowledgment

The authors would like to thank Yuki Ishikawa for his help in implementing the system. This work is supported by NEDO (New Energy and Industrial Technology Development Organization, Japan) Intelligent RT Software Project.

References

[1] S. Munder, C. Schnorr, and D. M. Gavrila: "Pedestrian Detection and Tracking Using a Mixture of View-Based Shape-Texture Models," *IEEE Trans. Intelligent Transportation Systems*, vol. 9, no. 2, pp. 333-343, 2008.

[2] B. Han, S. W. Joo, and L. S. Davis: "Probabilistic Fusion Tracking Using Mixture Kernel-Based Bayesian Filtering," In *Proceedings of the 11th Int. Conf. on Computer Vision*, 2007.

[3] D. M. Gavrila, "A Bayesian, Exemplar-Based Approach to Hierarchical Shape Matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 8, pp. 1408-1421, 2007.

[4] D. Beymer and K. Konolige: "Real-Time Tracking of Multi-ple People using Continuous Detection," In *Proceedings of the 7th Int. Conf. on Computer Vision*, 1999.

[5] J. Kang, I. Cohen, G. Medioni, and C. Yuan: "Detection and Tracking of Moving Objects from a Moving Platform in Presence of Strong Parallax," In *Proceedings of the 10th Int. Conf. on Computer Vision*, pp. 10-17, 2005.

[6] A. Howard, L. H. Matthies, A. Huertas, M. Bajracharya, and A. Rankin: "Detecting pedestrians with stereo vision: safe operation of autonomous ground vehicles in

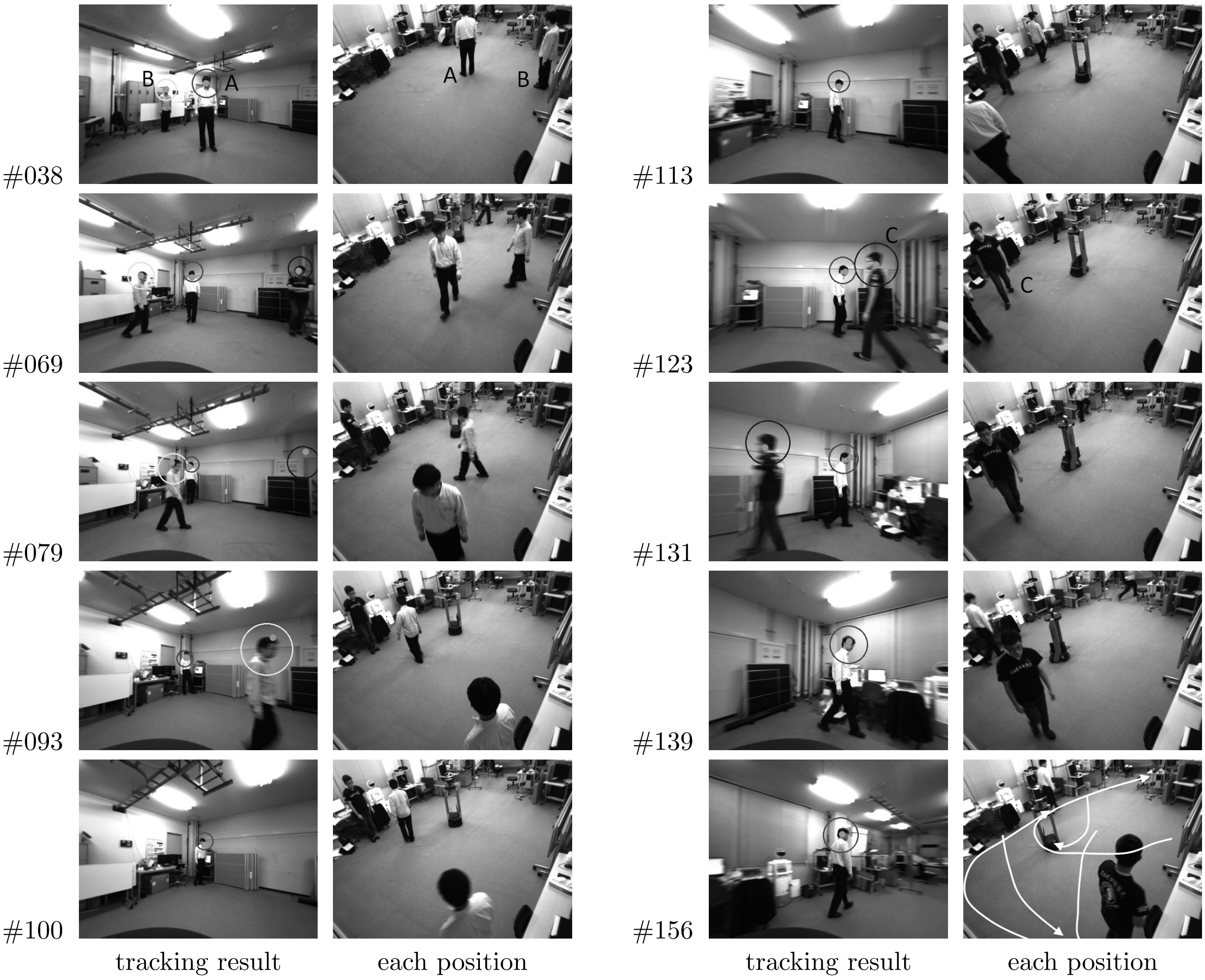


Figure 8: Experimental result.

dynamic environments," In *Proceedings of the 13th Int. Symp. of Robotics Research*, 2007.

[7] D. Schulz, W. Burgard, D. Fox, and A. B. Cremers: "People Tracking with a Mobile Robot Using Sample-based Joint Probabilistic Data Association Filters," *Int. J. of Robotics Research*, vol. 22, no. 2, pp. 99-116, 2003.

[8] N. Bellotto and H. Hu: "Multisensor Data Fusion for Joint People Tracking and Identification with a Service Robot," In *Proceedings of the 2007 IEEE Int. Conf. Robotics and Biomimetics*, pp. 1494-1499, 2007.

[9] C. Y. Lee, H. G. Banos, and J. C. Latombe: "Real-Time Tracking of an Unpredictable Target Amidst Unknown Obstacles," In *Proceedings of the 7th Int. Conf. on Control, Automation, Robotics and Vision*, pp. 597-601, 2002.

[10] L. Spinello, R. Triebel, and R. Siegwart: "Multimodal Detection and Tracking of Pedestrians in Urban Environments with Explicit Ground Plane Extraction," In *Proceedings of the 2008 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 1823-1829, 2008.

[11] M. Kobilarov, G. Sukhatme, J. Hyams, and P. Batavia: "People tracking and following with mobile robot using omnidirectional camera and a laser," In *Proceedings of the 2006 IEEE Int. Conf. on Robotics and Automation*, pp. 557-562, 2006.

[12] H. Koyasu, J. Miura , and Y. Shirai: "Recognizing Moving Obstacles for Robot Navigation using Real-time Omnidirectional Stereo Vision," *J. of Robotics and Mechatronics*, vol. 14, no. 2, pp. 147-156, 2002.

[13] D. Calisi, L. Iocchi, and R. Leone: "Person Following through Appearance Models and Stereo Vision using a Mobile Robot," In *Proceedings of VISAPP 2007 Workshop on Robot Vision*, pp. 46-56, 2007.

[14] A. Ess, B. Leibe, and L. V. Cool: "Depth and Appearance for Mobile Scene Analysis," In *Proceedings of the 11th Int. Conf. on Computer Vision*, 2007.

[15] A. Ess, B. Leibe, K. Schindler, and L. V. Cool: "A Mobile Vision System for Robust Multi-Person Tracking," In *Proceedings of the 2008 IEEE Conf. on Computer Vision and Pattern Recognition*, 2008.

[16] N. Ando, T. Suehiro, K. Kitagaki, T. Kotoku, and W.-K. Yoon: "RTMiddleware: Distributed Component Middleware for RT (Robot Technology)," In *Proceedings of 2005 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 3555-3560, 2005.