

## Human Tracking by Importance Sampling Particle Filtering on Omnidirectional Camera Platform

C. J. Song\*, C. M. Huang\*, L. C. Fu\*\*

\*Electrical Engineering Department, National Taiwan University, Taiwan, ROC.

\*\* Electrical Engineering Department and Computer Science and Information Engineering,  
National Taiwan University, Taiwan, ROC.  
(e-mail: [lichen@ntu.edu.tw](mailto:lichen@ntu.edu.tw))

---

**Abstract:** This paper proposes a sequential importance sampling (SIS) particle filtering framework to track the human with overcoming the warping and low resolution. We utilize a foreground-based importance sampling mechanism for efficiently converge to the target distribution. We construct a tracking system with the fusion image likelihood, even the human raises the head. Furthermore, the two-space integration to evaluate the likelihood measurement is proposed to robustly track human for overcome the warping effect. The overall performance has been validated in the experiments.

---

### 1. INTRODUCTION

Visual tracking in a dynamic environment has drawn much attention nowadays. We want to develop a vision system for human tracking under complex environments and covering large space. Utilizing different kinds of cameras will lead to different properties in regard to the above mentioned problem. Compared with other cameras, the omnidirectional cameras have a wider field of view covering large space, which approaches to 360 degrees. However, it suffers from a serious distortion and low resolution problems. Therefore, the approach for the developed omnidirectional camera here should be immune to these drawbacks.

#### 1.1 Related Works

In practice, omnidirectional cameras are often utilized in many applications, like autonomous mobile robot (Yagi *et al.*, 2005), video conference (Wallhoff *et al.*, 2004), military purpose (Boult *et al.*, 2001), virtual reality (Ikeuchi *et al.*, 2004), collaboration with PTZ camera (Scotti *et al.*, 2005) (Scotti *et al.*, 2004), visual tracking. The techniques for detection and tracking using an omnidirectional camera can be grouped into two categories, one in warping space and the other in unwarping space.

For techniques that work in warping space, they usually utilize color information as features and integrate different algorithms for detection and tracking, such as, optical flow, CAMShift, Kalman filters, and particle filters. A simple method for tracking a single person is to extract human features and integrate them with optical flow regions (Yagi *et al.*, 2002). However, optical flow is unsuitable for tracking in the image plane. To overcome this limitation, the work in (Wang *et al.*, 2006) combined optical flow with the CAMShift algorithm. The author used background subtraction to detect the single person, and then the CAMShift algorithm is used for extracting color information of the target in order to facilitate subsequent tracking. The

advantage of integrating optical flow and the CAMShift is that the system can achieve automatic tracking. The work in (Matsumura *et al.*, 2002) modeled skin color with a Gaussian distribution in normalized RGB color space, and used that skin color information under the Kalman filter framework to detect and track people. The state vector of the Kalman filter consists of position and velocity of the target. However, the Kalman filter assumes that humans move linearly and models the probability as a Gaussian distribution. It does not correspond to the actual condition. In (Ortegon-Aguilar *et al.*, 2006), the particle filter was proposed to track foreground regions in an omnidirectional sequence where the foreground regions were detected by background subtraction. The recently proposed approaches only used the color model as the feature without taking any other information, thus making tracking unstable. It is why unwarping space is applied here.

The technique of tracking work in unwarping space is based on a global transforming algorithm. In general, such techniques consist of four steps, i.e., transforming into panoramic image, background subtraction, modeling skin color distribution, and tracking framework. The aim of transform into panoramic image and background subtraction is to obtain an unwarping foreground, whereas that of modeling skin color distribution is to extract features. And the last step, i.e., tracking framework, is to predict the position of the target. For this tracking system in unwarping space, Cielniak *et al.* (Cielniak *et al.*, 2003) presents an appearance-based algorithm for tracking a single person using an artificial neural network to learn the appearance model to be incorporated with the Kalman filter. Features are extracted from the panoramic images, and then an artificial neural network is trained to estimate the distance of the person from the camera. The Kalman filter was used to track the position of the person. Wallhoff *et al.* (Wallhoff *et al.*, 2004) later employed the particle filter to track multiple people. First, an omnidirectional image was transformed to an unwarping image. Next, the skin color was converted from RGB color space into normalized rg-Chroma space and is

modeled with a Gaussian mixture model to be used to segment human regions and to compute face likelihood. Finally, the particle filter based face tracking system is used to track multiple people (Cauchois *et al.*, 2005), which begins with transformation of omnidirectional images to panoramic images, then adopts background subtraction to detect the object, and finally track it by utilizing an CAMShift algorithm. Nevertheless, the property of unwarping space was not utilized to solve the problem of the distortion. It not only loses the meaning of transforming original omnidirectional space into unwarping space but also wastes time.

### 1.2 System Overview

In this paper, we propose an approach to perform the human detection and tracking in a more effective manner. Fig. 1 shows the block diagram of the proposed algorithm. The main idea is to utilize the foreground-based importance sampling mechanism to draw particles efficiently. Most, we extract color and contour features and regard these two kinds of features as the likelihood measurement. The multiple cue fusion integrating color and contour features makes track humans more accurate. Evaluating the likelihood measurement on two-space integration can solve distortion drawbacks. Because we simultaneously use the two-space of information to evaluate the likelihood measurement, our system can overcome distortion problems and make the tracking capability more robust. On the other hand, since contour information is used, CAMShift is no longer fit to the tracking rule. Hence, we choose SIS particle filter to overthrow the problem with non-linear movement and lack of Gaussian distribution. SIS particle filter is more effective to draw particles than the particle filter.

This paper is organized as follows. The foreground-based importance sampling mechanism is described in section II. Section III addresses the feature extraction algorithm. The likelihood evaluation based on the two-space integration is described in section IV. In section V, we demonstrate the effectiveness of the developed approach by providing some appealing experimental results. Finally, we conclude the paper in section VI with some relevant discussion.

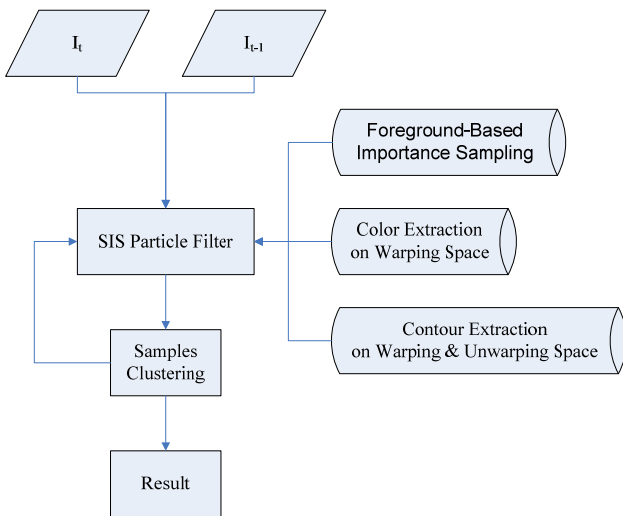


Fig. 1. The block diagram of the proposed algorithm.

## 2. Foreground-Based Importance Sampling

In this paper, we use SIS particle filter to detect and track the human. In order to draw particles efficiently, how to identify the importance sampling function is very crucial. If the importance sampling function is good enough, the particles will be drawn efficiently, and then we can detect the targets fast. Therefore, we integrate the auxiliary knowledge in the form of an importance function, which represents the probability of the possible target locations and is called the foreground-based importance sampling mechanism.

In our tracking system, the targets are heads of humans. Because our omnidirectional camera is installed on the ceiling, the heads of the humans in the images captured by our omnidirectional camera are approximating a circle. So we define the state vector as  $\mathbf{X}_t = (x, y, r)$ , where  $(x, y)$  is the center of the head with the purple point, and  $r$  is the radius of the head as shown in Fig. 2. The state vector  $\mathbf{X}_t = (x, y, r)$  can describe a circle to model a head which we observe from the omnidirectional camera.  $A(\mathbf{X})$  means the every pixels inside the particle  $\mathbf{X}_t = (x, y, r)$ .  $C(\mathbf{X})$  means the every pixels around circumference of the particle  $\mathbf{X}_t = (x, y, r)$ .  $\mathbf{Z}$  is the observer vector containing color and contour information we can observe from the image. So we must estimate the current state  $\mathbf{X}_t$  from observations  $\mathbf{Z}_{0:t} = \mathbf{Z}_0, \mathbf{Z}_1, \dots, \mathbf{Z}_t$  to get the posterior  $p(\mathbf{X}_t | \mathbf{Z}_{0:t})$ .

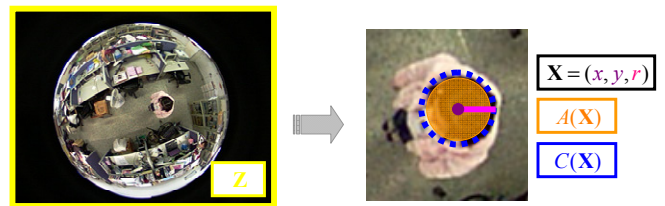


Fig. 2. Mathematical definition of our system.

In the foreground-based importance sampling mechanism, there are two major steps. The first one is the foreground segmentation, and the second one is sampling distribution. The targets can be detected by foreground segmentation the first step (Adams *et al.*, 1994) (Huang, 2007), and then sampling distribution is the second step can be processed, where the particles can be drawn inside the foreground.

### 2.1 Sampling Distribution

In our human tracking system, we define  $\{q_m(\mathbf{X}), m = 1, \dots, M\}$  as the importance functions representing the probability distribution of the  $M$  targets. Each importance function takes charge of a tracking task for a single specific target. Now, let  $q_1(x, y)$  denote the probability distribution projected on the  $x$  and  $y$  axis and  $q_2(r)$  be just a uniform probability distribution. Therefore,  $q(\mathbf{X})$  satisfies the following equation:

$$q(\mathbf{X}) = q(x, y, r) = q_1(x, y)q_2(r), \quad (1)$$

where

$$q_1(x, y) = P(X, Y) = P(Y | X)P(X) \quad (2)$$

with

$$P(X = x) = \sum_j B(x, j) / \sum_{(i,j)} B(i, j) \quad (3)$$

and

$$\begin{cases} B(i, j) = 0 & \text{if } (i, j) = 0 \\ B(i, j) = 1 & \text{otherwise} \end{cases} \quad (4)$$

In Fig 3, we first project every pixel on the x axis to make  $P(X)$  distribution. Then we randomly choice a position on the  $P(X)$  distribution where the pixel number larger than one. So we just project every pixel of a range regarding the position as the center on the y axis to make  $P(Y | X)$  distribution. Also we randomly choice a position on the  $P(Y | X)$  distribution where the pixel number larger than one. Finally, we can randomly get a position on the 2D image. According to the obtained probability distribution, we can draw significant samples. Fig. 4 shows the overall foreground-based importance sampling mechanism.

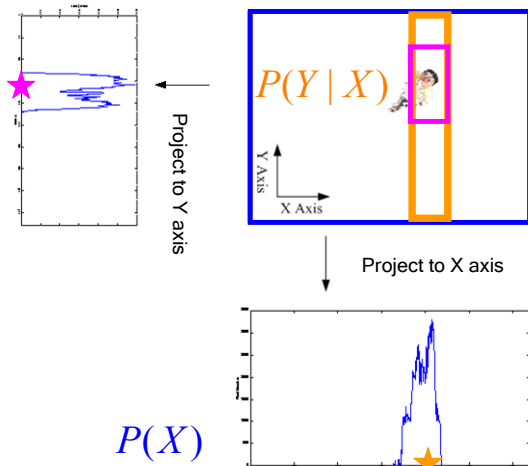


Fig. 3. Sampling distribution from foreground detection.

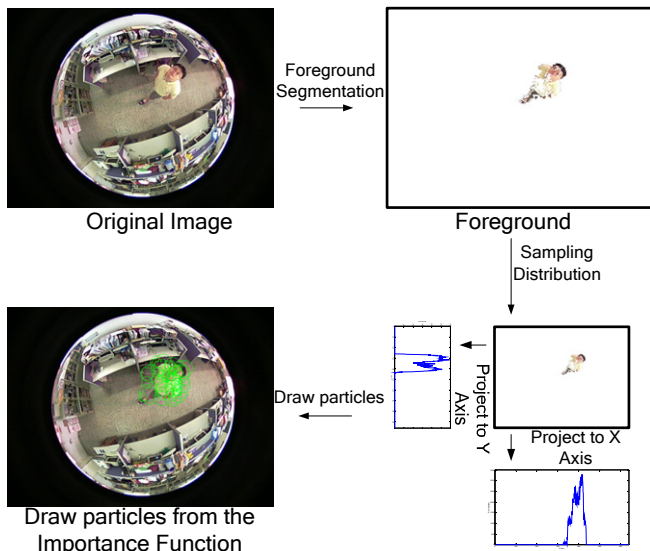


Fig. 4. Foreground-based importance sampling mechanism.

### 3. Multiple Cue Fusion

In order to get enough features to track targets robustly, we use the color and contour information to model and update the likelihood, which can be factored as

$$p(\mathbf{Z}_t | \mathbf{X}_t) \propto p_{contour}(\mathbf{Z}_t | \mathbf{X}_t) \cdot p_{color}(\mathbf{Z}_t | \mathbf{X}_t). \quad (5)$$

#### 3.1 Contour Likelihood

We obtain the contour images from the original warping images with the Sobel operator. Let the gradient magnitude as  $G(x, y)$ , and the contour distance  $D_{contour}(\mathbf{Z} | \mathbf{X})$  of a specific particle  $\mathbf{X}$  can be defined as

$$D_{contour}(\mathbf{Z} | \mathbf{X}) = 1 - \frac{1}{|C(\mathbf{X})|} \sum_{(x,y) \in C(\mathbf{X})} G(x, y). \quad (6)$$

Then, the normalized contour distance  $\bar{D}_{contour}(\mathbf{Z} | \mathbf{X})$  of a specific particle  $\mathbf{X}$  can be defined as

$$\bar{D}_{contour}(\mathbf{Z} | \mathbf{X}) = \frac{D_{contour}(\mathbf{Z} | \mathbf{X}) - D_{contour}^{min}}{D_{contour}^{max} - D_{contour}^{min}}. \quad (7)$$

As a result, the contour likelihood measurement function can be expressed as

$$p_{contour}(\mathbf{Z} | \mathbf{X}) \propto \exp[-\lambda_{contour} \bar{D}_{contour}(\mathbf{Z} | \mathbf{X})], \quad (8)$$

where  $\lambda_{contour}$  is the coefficient of the contour likelihood measurement function.

#### 3.2 Color Likelihood

The color distribution can be exploited as a useful feature to enhance the tracking performance. As shown in Fig. 5, we define an  $N$ -bin color histogram reference model as

$$\mathbf{q}^* = \{q_0(n)\}_{n=1}^N, \quad (9)$$

which is the color distribution of region  $A(\mathbf{X})$ . Here, we use the normalized histograms, i.e.  $\sum_{n=1}^N q_0(n) = 1$ . Now, we consider the color histogram in the associated image region as

$$\mathbf{q}(\mathbf{X}) = \{q(n; \mathbf{X})\}_{n=1}^N, \quad (10)$$

which is the color distribution of the template. Then, comparing these two histograms, we define the distribution likelihood as

$$p_{color}(\mathbf{Z} | \mathbf{X}) \propto \exp[-\lambda_c D_{color}^2(\mathbf{q}^*, \mathbf{q}(\mathbf{X}))], \quad (11)$$

where  $\lambda_c$  is the coefficient of the color likelihood measurement function, and  $D(\mathbf{q}^*, \mathbf{q}(\mathbf{X}))$  is the Bhattacharyya distance, which is defined as

$$D(\mathbf{q}^*, \mathbf{q}(\mathbf{X})) = \left[ 1 - \sum_{n=1}^N \sqrt{q_0(n) \cdot q(n; \mathbf{X})} \right]^2. \quad (12)$$

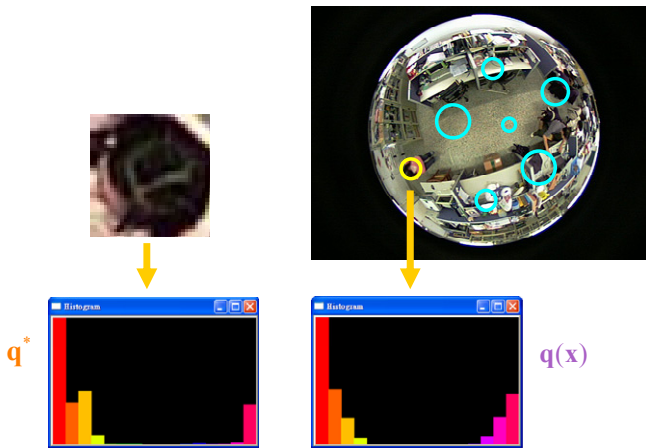


Fig. 5. An explanation of histogram likelihood.

#### 4. Two-Space Integration

In general, the circumference of the circle on warping space is the width of the image on unwarping space and the radius of the circle on warping space is the height of the image on unwarping space as shown in Fig. 7.

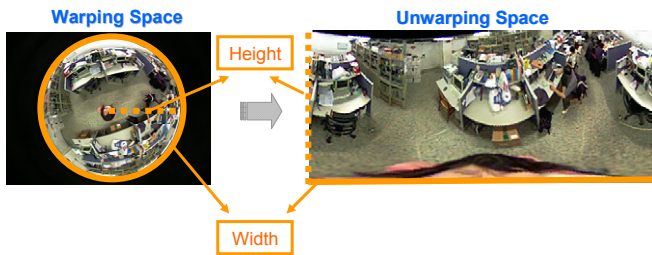


Fig. 6. Global unwarping transformation.

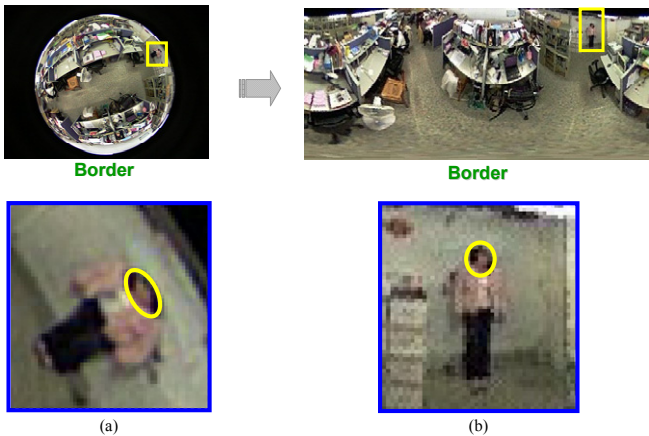


Fig. 7. The person on the border of the image on two-space.

#### 4.1 Color and Contour Properties on Two-Space

In this section, we introduce the different influences based on warping and unwarping space for color and contour features. Despite the person in different positions, it is obvious for color information no matter what the person locate. So we evaluate the color likelihood measurement  $p_{color}(Z_t | X_t)$  on

warping space. But there is heavy distortion for contour information when the person is on the center of the image on unwarping space as shown in Fig. 6. So we evaluate the contour likelihood measurement  $p_{contour}(Z_t | X_t)$  on warping space when the person is near the center of the image. We observe that there are better circular contours on unwarping space as shown in Fig. 7. So, we evaluate the contour likelihood measurement  $p_{contour}(Z_t | X_t)$  on unwarping space when the person is near the border of the image.

#### 4.2 Local Unwarping Transformation

Because we just transform the border of the image on warping space to unwarping space, the transformation method is local unwarping transformation as shown in Fig. 9. It is not like the Global unwarping transformation as shown in Fig. 6. We utilize the sampling method to do the local unwarping transformation. The detail transformation is shown in Fig. 10. We use inner circumference to be the width on unwarping space. If we use outer circumference to be the width on unwarping space, it will be several pixels corresponding to only one pixel. It can not solve the distortion. In our method, it samples the pixels to transform warping space to unwarping space. In the implementation, a circle map as shown in Fig. 8 is established. The image is separated into three parts. The black part is useless. The particles  $X$  in the gray part are transformed into unwarping space to evaluate contour likelihood measurement. The contour likelihood measurement is evaluated on warping space when the particles  $X$  are in the white part.

If the particles  $X$  are in the gray part, they are transformed into unwarping space. We transform  $C(X)$  formed by the particle  $X$  to unwarping space  $C'(X)$  using the transform function  $T(P(X), P'(X))$ . The transform function  $T(P(X), P'(X))$  is shown in the following:

$$\begin{aligned}
 x &= x_0 + (R_2 + y') \cos \theta, \\
 y &= y_0 + (R_2 + y') \sin \theta, \\
 r &= r', \\
 \theta &= (2\pi x') / \text{Width}_{\text{warp}},
 \end{aligned}
 \tag{13}$$

where  $X = (x, y, r) \in C(X)$  on warping space and  $X' = (x', y', r') \in C'(X)$  on unwarping space.  $(x_0, y_0)$  is the center of the image on warping space.  $R_2$  and  $\text{Width}_{\text{warp}}$  are shown in Fig. 10.  $r'$  is a average distance between  $C'(X)$  and  $(x', y')$ .



Fig. 8. Circle map.



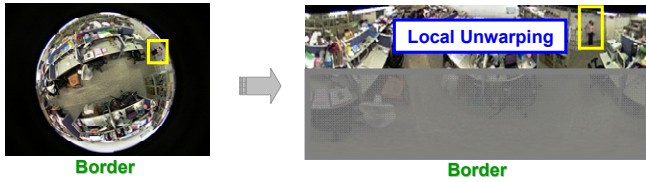


Fig. 9. Local unwarping transformation.

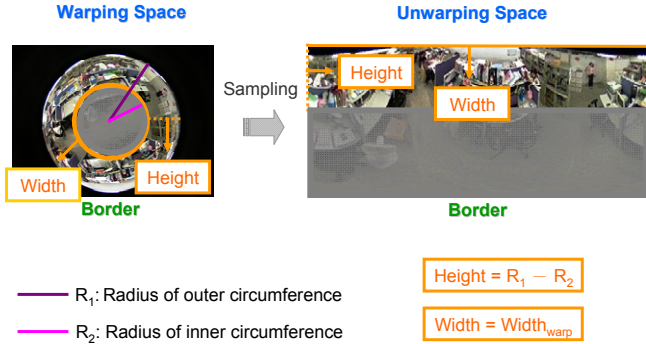


Fig. 10. Local unwarping transformation method.

### 4.3 Contour Likelihood Evaluation Redefine

We redefine the likelihood measurement in (5). The new likelihood measurement can be factored as

$$p(\mathbf{Z}_t | \mathbf{X}_t) \propto p'_{contour}(\mathbf{Z}_t | \mathbf{X}_t) \cdot p_{color}(\mathbf{Z}_t | \mathbf{X}_t), \quad (14)$$

where

$$p'_{contour}(\mathbf{Z} | \mathbf{X}) \propto \exp[-\lambda_{contour} \bar{D}'_{contour}(\mathbf{Z} | \mathbf{X})], \quad (15)$$

with

$$\bar{D}'_{contour}(\mathbf{Z} | \mathbf{X}) = \begin{cases} r_{p_c} \leq R_2 & \bar{D}_{contour}(\mathbf{Z} | \mathbf{X}) \\ r_{p_c} > R_2 & \bar{D}_{unwarp}^{contour}(\mathbf{Z} | \mathbf{X}) \end{cases}, \quad (16)$$

and  $r_{p_c}$  means the distance between the particle  $\mathbf{X}$  and the center of the image on warping space.

## 5. Experiment Results

A well-known open source library, OpenCV, is used to facilitate the development of the system. The fisheye camera is combined the fisheye lens and the camera platform of AXIS 216 Network Camera. The angle of the view is 185 degrees. In our experiment, we use the 640×480 resolution which can be captured 30 frames per second. Furthermore, we process the image on-line with a computer which is Intel Core2 1.83GHz and the RAM is 1GB.

Fig. 11 and Fig. 12 are the tracking result evaluating likelihood by color and contour information and integrating two-space. In Fig. 11, regardless of the person near the center of the image in frame#68 and frame#85 or on the boundary of other images, the results show that the target can be detected and tracked robustly even when the person changes the pose, like raising his head. In Fig. 12, it shows the robust and correct tracking when two persons walk randomly and change pose.

## 6. Conclusion

In this paper, we perform the human tracking using the SIS particle filter with the single omnidirectional camera. In order to draw particles efficiently, we extend the particle filter to the SIS particle filter. A foreground-based importance sampling mechanism is proposed to evaluate the importance function. It draws the particles from the importance function rather than the prior distribution. Moreover, we fuse the contour features and color features as the likelihood measurement function for evaluating the particle sets of the human. Fusion of multiple cues for likelihood evaluation tracks the human in a more accurate way. Furthermore, combining the color and contour features of warping space and the contour features of unwarping space makes the features more significant. Likelihood evaluation by integrating two-space enhances the robustness of the tracking system. The experimental results demonstrate that our proposed method can successfully track the human even under a complex environment with a large space coverage.

## REFERENCES

- Yagi, Y., K. Imai, K. Tsuji, and M. Yachida (2005). Iconic memory-based omnidirectional route panorama navigation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 78-87.
- Wallhoff, F., M. Zobl, G. Rigoll, and I. Potucek (2004). Face tracking in meeting room scenarios using omnidirectional views. *Proceedings of the 17th International Conference on Pattern Recognition*, pp. 933-936.
- Boult, T. E., R. J. Micheals, G. Xiang, and M. Eckmann (2001). Into the woods: Visual surveillance of noncooperative and camouflaged targets in complex outdoor settings. *Proceedings of the IEEE*, vol. 89, pp. 1382-1402.
- Ikeuchi, K., M. Sakauchi, H. Kawasaki, and I. Sato (2004). Constructing Virtual Cities by Using Panoramic Images. *International Journal of Computer Vision*, vol. 58, pp. 237-247.
- Scotti, G., L. Marcenaro, C. Coelho, F. Selvaggi, and C. S. Regazzoni (2005). Dual camera intelligent sensor for high definition 360 degrees surveillance. *IEE Proceedings Vision, Image and Signal Processing*, vol. 152, pp. 250-257.
- Scotti, G., L. Marcenaro, C. Coelho, F. Selvaggi, and C. S. Regazzoni (2004). A novel dual camera intelligent sensor for high definition 360 degrees surveillance. *IEEE Intelligent Distributed Surveillance Systems*, pp. 26-30.
- Yagi, Y., and M. Yachida (2002). Omnidirectional sensing for human interaction. *Proceedings of Third Workshop on Omnidirectional Vision*, pp. 121-127.
- Wang, M. L., C. C. Huang, and H. Y. Lin (2006). An Intelligent Surveillance System Based on an Omnidirectional Vision Sensor. *IEEE Conference on Cybernetics and Intelligent Systems*, pp. 1-6.
- Matsumura, A., Y. Iwai, and M. Yachida (2002). Tracking people by using color information from omnidirectional

images. *Proceedings of the 41st SICE Annual Conference*, pp. 1772-1777.

Ortegon-Aguilar, J. and E. Bayro-Corrochano (2006). Omnidirectional Vision Tracking with Particle Filter. *18th International Conference on Pattern Recognition*, pp. 1115-1118.

Cielniak, G., M. Miladinovic, D. Hammarin, L. Goranson, A. Lilienthal, and T. Duckett (2003). Appearance-based tracking of persons with an omnidirectional vision sensor. *Omnivis Workshop with ECCV: Omnidirectional Vision and Camera Networks*.

Cauchois, C., F. de Chaumont, B. Marhic, L. Delahoche, and M. Delafosse (2005). Robotic assistance: an automatic

wheelchair tracking and following functionality by omnidirectional vision. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2560-2565.

Adams, R., and L. Bischof (1994). Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, pp. 641-647.

Huang, S.S. (2007). Segmentation and Pose Estimation. *Ph.D. dissertation*, Dept. Computer Science and Information Engineering, National Taiwan Univ., Taiwan, ROC.

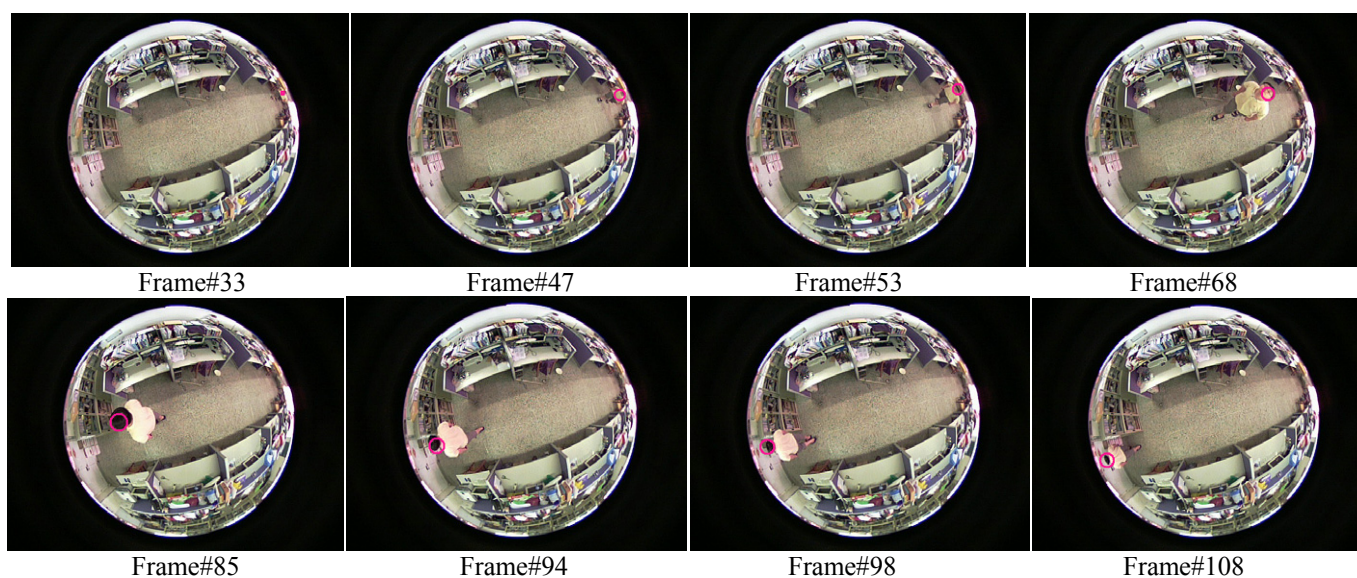


Fig. 11. Single human tracking.

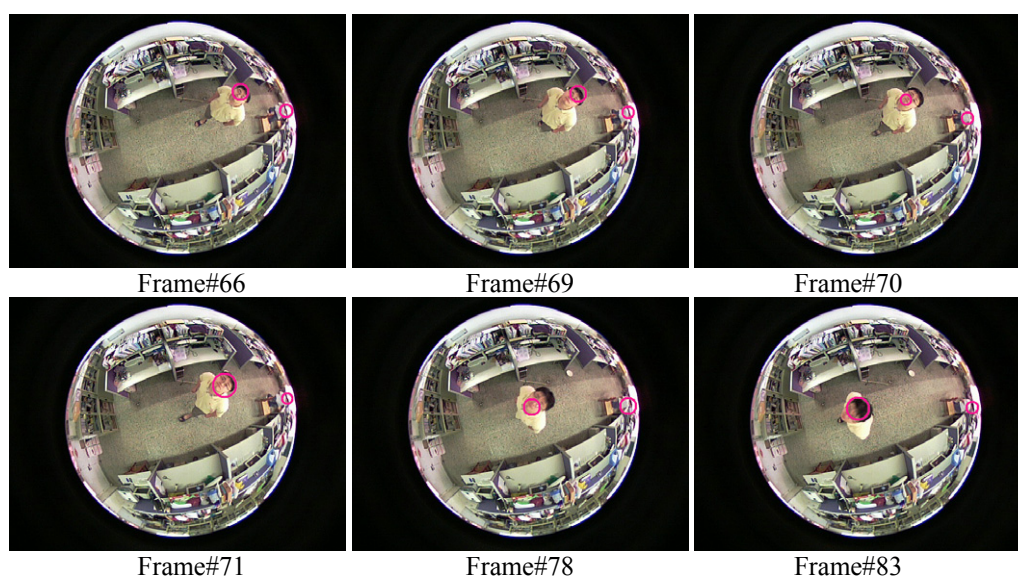


Fig. 12. Multiple human tracking.