# Robust Tracking of Walking Persons by Elite-Type Particle Filters and RGB-D Images

Akari Oshima[1(✉)], Shun'ichi Kaneko[1], and Masaya Itoh[2]

[1] Hokkaido University, Sapporo, Japan
oshima@hce.ist.hokudai.ac.jp, kaneko@ist.hokudai.ac.jp
[2] Hitachi Research Laboratory, Hitachi Ltd., Hitachi, Japan
masaya.itoh.pp@hitachi.com

**Abstract.** In this paper, we propose a robust real-time tracking system using RGB-D image sequence which are obtained through stereo camera. We apply 'Elite-type' particle filter, which is novel structure of particle filter, for tracking multiple persons. In Elite-type particle filter, to be robust to change of appearance and partial occlusion, likelihood is designed based on histogram and each particle possess their own model histogram. The system assign this particle filter to each person, and estimate state of the target person which vary from frame to frame. Furthermore, the system is able to measure the height of person's head, which is effective for analysis human behavior. Real-time tracking performance of multiple persons was confirmed by experiments which simulating a real shop.

**Keywords:** Tracking · Robustness · Particle filter · Color · Stereo sensing · Depth · Likelihood · Human behavior · Walking person · Shopper

## 1 Introduction

Recently, due to the growing awareness of safety and security or crime prevention requirement, surveillance cameras are introduced into many places increasingly. Accordingly, human behavior recognition and analysis technologies based on image sequences acquired from these have been studied [1–4]. Such technologies are used in a variety situations such as marketing design, security and healthcare management. We, for example, have tried to make a tracking system named ISZOT [5] by use of a calibrated single camera to measure rough 2D positions in the shop to analyze shopper behaviors. In the ISZOT system, shopper's zone trajectories could be effectively analyzed, which represent their purchasing and/or wondering behaviors in front of pre-specified zones.

In order to design any effective tracking algorithms, we have had to solve many ill-conditions in the real environment, such as illumination fluctuation, occlusion between walking persons, shadows, and so on. The image data acquired from monochrome or color cameras installed for many security-oriented monitoring, however, are not sufficient for making the tracking algorithms more robust

S. Battiato et al. (Eds.): ICIAP 2017, Part I, LNCS 10484, pp. 108–118, 2017.
https://doi.org/10.1007/978-3-319-68560-1_10

against those ill-conditions due to their limitations of two dimensional (2D) observation. In recent decade, stereo sensing cameras become popular in the real world in performance and price in addition with the availability of rapid network environment to connect them from/to their central controllers. It is getting important to design any effective algorithms to introduce much more 3D real-time sensing functions into the above mentioned systems, and then to utilize the 3D data for robust capturing of the target continuous movements in the scene. In this paper, as our contribution to this field, based on a novel structure of particle filter, a robust real-time tracking system using RGB-D image sequence which are obtained through stereo camera sensing is proposed.

The rest of this paper is organized as follows: Sect. 2 describes algorithms of Elite-type particle filter, Sect. 3 describes how to manage these particle filters in order to track multiple free walking persons, Sect. 4 shows how to adjust the parameters based on target locations relative to the stereo sensors or cameras, Sect. 5 presents the experimental results in the laboratory, and then in Sect. 6 we discuss our conclusive remarks and future works.

## 2 Elite-Type Particle Filter

### 2.1 Overall Structure of Tracking Algorithm

In this research, we develop a tracking algorithm by applying particle filters which has been an approach to estimating the non-linear and transitional statistical distributions of object states by using a large number of particles distributed in the observation space. Many study have been reported in human tracking [6–8]. In general type of it one uses a simple likelihood because of its limited calculation cost for large number of particles. In this research, contrast to these conventional methods, we originally utilizes multiple filters, each of which consists of a smart few particles to follow simultaneously multiple persons. In this independent filter, each particle memorizes which part of the target or person it may be placed on at the previous frame or sampling time based on three likelihoods. The basic structure of the proposed tracking system is shown in Fig. 1. When the person detector finds a set of data probably representing a person in a subtracted depth image calculated from a RGB-D image, the system generates and places a particle filter around it. The updating process includes search and resampling of particles and the state estimation based on likelihoods are repeated in every sampled frame. Each process is described in detail later. In this system, 3D position information which xy plane represent floor is calculated from the RGB-D image, where through a calibrated adjustment by the stereo sensor. We call information of 3D position and color which associated with coordinate value in image space as a data point.

### 2.2 Basic Structure of Elite-Type Particle Filter

Figure 2 shows a stereo sensor installed on the ceiling to take images of walking persons on the floor and the basic concept and situation of the proposed particle

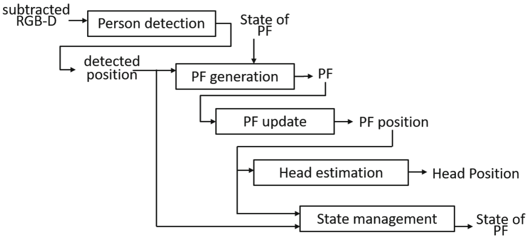

**Fig. 1.** Schematic of proposed algorithm

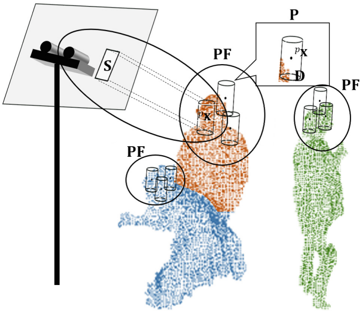

**Fig. 2.** Definition of particle filters

filter. We define a particle filter by coordinating in multiple and dispersive manner each particle of a cylindrical shape, which always lies in the vertical direction in the real space. Because of their frequent attitudes in standing and/or walking, we choose the cylinder of the designed size and shape for this purpose and then it is expected to strong against the changing direction of walking persons due to its invariance in shape in rotation about the vertical axis to the floor. The particle $\mathbf{P}$ is defined by following equation:

$$\mathbf{P} = \{{}^p\mathbf{x}, \mathbf{D}, \mathbf{S}\} \tag{1}$$

where ${}^p\mathbf{x}$, $\mathbf{S}$ and $\mathbf{D}$ define the center position, the image region where it is projected onto, and the set of data points included, respectively. Here, Smart Window Transform (SWT) [9] has been used to project particle to image, which enables to calculate $\mathbf{D}$ efficiently from $\mathbf{S}$. The particle filter (PF) is a set of the neighboring particles that can be coordinated not to belong to any other PF which is defined as $\mathbf{F}$ by the next equation.

$$\mathbf{F}_i = \{\mathbf{P}^i_j \mid j = 1, \dots N_p\} \tag{2}$$

where, $N_p$ is the number of component particles. The position of $\mathbf{F}_i$ is defined as ${}^f\mathbf{x}_i$ which is the average of ${}^p\mathbf{x}^i$. In these Elite-type particles, through somewhat a taking care of each process, we aim to make not so many particles to govern themselves and follow each target autonomously.

A PF is typically generated when a person comes in the scene and the person detector, one of provided libraries, possibly detect him or her in the observation space. Some particles are defined and generated just around the portions of the head through the chest because the upper body of a walking person has less change in shape than the lower body has. In order to realize such arrangement, we use the highest data points in the detected area, where we call them ‘head-top’ of the tracked person. The particle layout is slightly controlled by introducing random factors in 3D space. The condition is represented in the next equation with respect to the number density of data points so that each particle can include enough amount of data points inside it.

$$|\mathbf{D}| > \delta_p \tag{3}$$

During this process, although the person detector find person-like area, if the condition shown in Eq. 3 cannot be satisfied for more than a certain period of time, it is regarded as an error of the detector and the generation process gets to be quitted partially only just in this area. Moreover, in order to prevent multiple placement of PF with severe overlap, we check whether the generated PF can follow the target and then in order to judge any successful PF generation relative arrangement of the existing PFs and the the new one $\mathbf{x}_d$ is used as]break follows:

$$\forall i, |\mathbf{x}_d - {}^{f}\mathbf{x}| > \lambda_s \tag{4}$$

where, the threshold $\lambda_s$ is so important that it can control the relative arrangement of all of the existing PFs by keeping that their mutual distances should be larger than $\lambda_s$.

Since as the one of our applications of this algorithm we aim shopper behavior analysis for effective marketing, we have designed an estimator of head-top positions of persons in each sampled frame because the gaze orientation is one of the most important demands in such application, extending the possibility of our proposed algorithm. In order to do this, H-Mask is designed to cover a head of the average size of Japanese. Figure 3 show the procedure of estimating the head position ${}^{h}\mathbf{X}$ independently of any tracking process. We first calculate the head-top position of the target as in the same way as the PF generation. Secondly, the upper center of the H-Mask is defined to coincide it with the head-top. Finally, the position corresponding to the center part of the H-Mask is taken as ${}^{h}\mathbf{X}$.

### 2.3 Likelihoods

The particle memorizes its position in the previous sampled frame and then it searches its own possible location in the current frame for fixing itself in some range around the previous position. This position determination process is performed based on the likelihoods which evaluate three types of similarities with respect to color, height, and trajectory. The likelihood of color $L_c$ uses color features in their 2D histogram $\mathbf{H}_c^{(t)} = \{h_c^{(t)}(i,j)\}$ is made from the data point set $\mathbf{D}^{(t)}$ at the frame $t$. We use color phase $ab$ in the Lab color coordination for making bins of the histogram. Besides, $L_c$ is calculated by the following equation which is the intersection evaluation [10] of $\mathbf{H}_c^{(t)}$ and $\mathbf{H}_c^{(t-1)}$.

$$L_c = \sum_{i=1}^{n_a}\sum_{j=1}^{n_b} \min(h_c^{(t)}(i,j), h_c^{(t-1)}(i.,j)) \tag{5}$$

where $n_a$ and $n_b$ indicate the number of bins in the histogram, respectively. The likelihood of height $L_h$ addresses the similarity based on the height histograms.

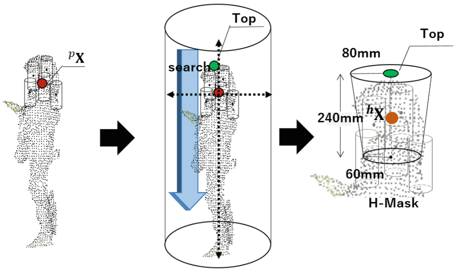

**Fig. 3.** Head-mask

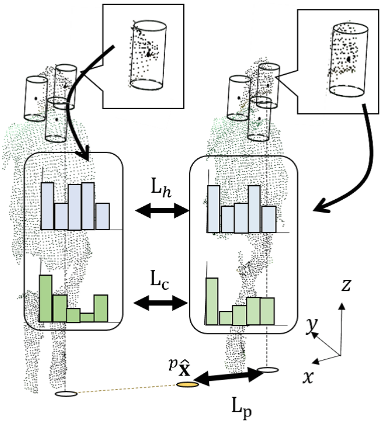

**Fig. 4.** Three likelihood

Similarly to $L_c$, a height histogram $\mathbf{H}_h^{(t)} = \{h_h^{(t)}(i)\}$ is made from $\mathbf{D}^{(t)}$. It is calculated as

$$L_h = \sum_{i=1}^{n_h} \min(h_h^{(t)}(i), h_h^{(t-1)}(i)) \tag{6}$$

where $n_h$ indicates the number of bin in the height histogram. The likelihood of trajectory $L_p$ is calculated from the difference between the current position of the particle: ${}^p\mathbf{x}$ and the estimated position ${}^p\hat{\mathbf{x}}$ by use of the velocity and the position at the previous frame as follows:

$$L_p = \frac{1}{K_r|{}^p\hat{\mathbf{x}} - {}^p\mathbf{x}| + 1} \tag{7}$$

where $K_r$ is a weight parameter. Here, we have $0 \leq L_p \leq 1$ and hence the likelihood decreases as the particle moves away from the estimated location. The total likelihood $L$ is finally calculated by the following equation as the combination of the above three likelihoods.

$$L = \alpha_c L_c + \alpha_h L_h + \alpha_p L_p \tag{8}$$

where $\alpha_c, \alpha_h, \alpha_p, (\alpha_c + \alpha_h + \alpha_p = 1)$ are the weights for each likelihood, and in this paper we give those values empirically. Figure 4 shows the concept of three elemental likelihoods. Moreover, when multiple persons approach and then sometimes occlude each other, it is afraid that any PF tracking them may fall into ill-condition and then possibly lose their correct trajectories or follow another person vice versa as misrecognition. In order to deal with these cases, we utilize a prohibited area $\mathbf{C}_i$ for PF replacement as shown in the next expression.

$$\mathbf{C}_i = \{\mathbf{x}_s || \mathbf{x}_s - {}^f\hat{\mathbf{x}}_j| \leq \lambda_c, j = 1, \ldots, N_f, j \neq i\} \tag{9}$$

where $\lambda_c$ indicates the threshold and ${}^f\hat{\mathbf{x}}_j$ is estimated position of $\mathbf{F}_j$ which should be distinguished from the estimated position of particle in Eq. 7, for example. Since the probability of another target person's existence is high in the estimated position of their PF, it is regarded as prohibition area.

### 2.4 Resampling

In order to improve the tracking performance, the resampling process is performed after the searching as in the normal approaches. One can reproduce the particles by use of this process so that those placed at any 'wrong' location or at non-target ones could be moved to a possible location of the same target. The resampling process is very important to maintain Elite-type PF in better activity and in this paper we need the following three procedures: (1) Compatibility evaluation, (2) Grouping, and (3) Reliability check. First, we judge a compatibility of each particle in tracking by using the number density of data points included in the particle through the Eq. 3 which is same as PF generation. The condition seems as a simple one however we need introduce a novel scheme as shown in Sect. 4 to adjust the threshold values including the above one with respect to their distances from the stereo sensors. If all of the particles have disappeared at a frame, the PF is judged as it lost its target and then transited to 'standby' state, where the detail of this state transition mechanism will be described in the next Sect. 3.

Secondly, the particles belonging to the same PF are grouped according to their distances. Let $\mathbf{F}'$ be a set of all of the $\mathbf{P}$ that survive through the previous compatibility evaluation, $\mathbf{G}_i (i = 1, 2, \ldots, N_g)$ be the direct sum decomposition of $\mathbf{F}'$, and $\mathbf{I}_i$ be their subscript set. Here $\mathbf{G}$ means the group, each of which should satisfies the following condition.

$$\begin{aligned} &\mathbf{G}_i = \{\mathbf{P}_n \mid n \in \mathbf{I}_i\} \\ &n \in \mathbf{I}_i, m \in \mathbf{I}_j, i \neq j \Rightarrow |{}^p\mathbf{x}_n - {}^p\mathbf{x}_m| > \Gamma_p \end{aligned} \tag{10}$$

The above condition means that any member particle to an arbitrary group and other non-member particles is separated to have larger distance than the threshold $\lambda_g$. By appropriately setting of this $\lambda_g$, it is possible to satisfactorily separate the particles placed in the target and non-target subjects.

Finally, we calculate a reliability $\gamma_i$ for each group $\mathbf{G}_i$. This represents how firmly it is placed with fitting to just the target person as follows:

$$\gamma_i = \alpha_\gamma \gamma_i^p + (1 - \alpha_\gamma)\gamma_i^d \tag{11}$$

where $\alpha_\gamma$, $\gamma_i^p$, and $\gamma_i^d$ are a weighting coefficient, the position-based reliability, and the data-point-based reliability, respectively, and furthermore $\gamma_i^p$ is calculated as follows:

$$\gamma_i^p = \frac{1}{K_r |{}^f\hat{\mathbf{x}} - {}^g\mathbf{x}_i| + 1} \tag{12}$$

where ${}^{f}\hat{\mathbf{x}}$ and ${}^{g}\mathbf{x}_i$ are the estimated position of PF and the average position of particles belonging to $\mathbf{G}_i$, respectively, and using the same $K_r$ as used for PF generation. In addition, $\gamma_i^d$ is calculated using the following equation.

$$\gamma_i^d = \frac{|{}^{g}\mathbf{D}_i|}{\sum_{j=1}^{n_g} |{}^{g}\mathbf{D}_j|} \tag{13}$$

where $|{}^{g}\mathbf{D}_i|$ is the total number of data points of particle belonging to the group. Only the particles belonging to $\mathbf{G}$ which have sufficiently high reliability to continue better tracking, and the other particles may be deleted. Finally, to supplement new particles is provided randomly around ${}^{g}\mathbf{x}$ of $\mathbf{G}$ so that the total number of particles in any PF is kept as $N_p$.

## 3 Transitional Management of PF States

When one imagines some indoor scenes having freely walking people, there may be some happenings in observation, such as appearing and disappearing in/from the scene, crossover between any two persons, and sudden stopping and standing. It is not so easy to deal with all of these cases, however, we try to attack some of the problems by recognizing transitional states of all of the PF under control of a management algorithm proposed in this paper. The states of PF can basically be divided into the following two: an active state $S_a$ and a standby state $S_r$ which are simply shown in Fig. 5. The former one has been described so far, however, the latter one needs to explain here. That is the state where any PF may lose its target person temporally.

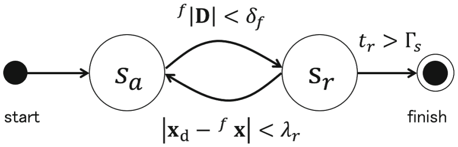
${}^{f}|\mathbf{D}| < \delta_f$
$t_r > \Gamma_s$
$S_a$
$S_r$
start
finish
$|\mathbf{x}_\mathrm{d} - {}^{f}\mathbf{x}| < \lambda_r$

**Fig. 5.** State transition management of PF

As shown in Fig. 5, any PF must be in the active state in the beginning, and then during its 'active' life, it is expected to perform normal tracking. Since sometimes it gets to lose its target and to have only few data points, it must be transited to the standby state by checking the total number of data points belonging to it (${}^{f}|\mathbf{D}|$) falls below the threshold $\delta_f$. During the 'standby' life, the PF must be in exemption from any process for PF except for keeping in the position just before the transition to wait re-activation. Any standby PF can have two possibilities as follows: the transition again to the active state if its target person may be detected just in the neighboring range of the distance $\lambda_r$ from its position. Or the disappearance from the scene if the elapsed time of its standby state $t_r$ exceeds $\Gamma_s$, where we may find absence of the target person from the observation space. By managing the state of PF, it is possible to obtain a better adaptation as a real facility.

## 4 Parameter Adjustment

In any stereo sensing, the spatial resolution of measured coordinates basically decreases according to the increasing ego-centric distance from the camera to the objects, in addition, the measurement errors may have an opposite tendency to increase together with the distance. For not so short period to keep observation of moving persons by the stereo camera, we should have designed some scheme to adjust important parameters in our proposed mechanism to the change. From fundamental experiments, there must be some sensitive but important parameters as follows: the threshold value for compatibility check $\delta_p$ in Eq. 3, the specified distance between any two PF $\lambda_s$ in Eq. 4, and then the threshold value to determine the standby state of PF in Fig. 5 $\delta_f$. In the case of $\delta_p$, the number of data points may increase as looming persons to the camera as their projected sizes on the camera plane increase. Thus, it is necessary to adjust their values smoothly within a predetermined range according to the distance from the camera. In order to realize this requirement we adopted the sigmoid function which has two representative values. For example, $\delta_p$ is calculated as follows:

$$\delta_p = \frac{\delta_{p2} - \delta_{p1}}{1 + e^{\alpha_s(|^p\mathbf{x}^c| - \lambda_d)}} + \delta_{p1} \tag{14}$$

where $\delta_{p_1}$ and $\delta_{p2}$ are the lower and the upper limits, which can be used in the large and the small distance from the camera, respectively, according to the ego-centric distance $|^p\mathbf{x}^c|$ of $\mathbf{P}$ from the camera. In addition, $\lambda_d$ gives the distance at which the controlled parameter has the middle value and $\alpha_s$ realizes an arbitrary rate of smooth change. Figure 6 shows how $\delta_p$ varies according to the distance from the camera. Even though two particles contain the same number of data points, the particle near to the camera is judged incompatible, while the other one far from the camera is judged compatible. The remaining two parameters

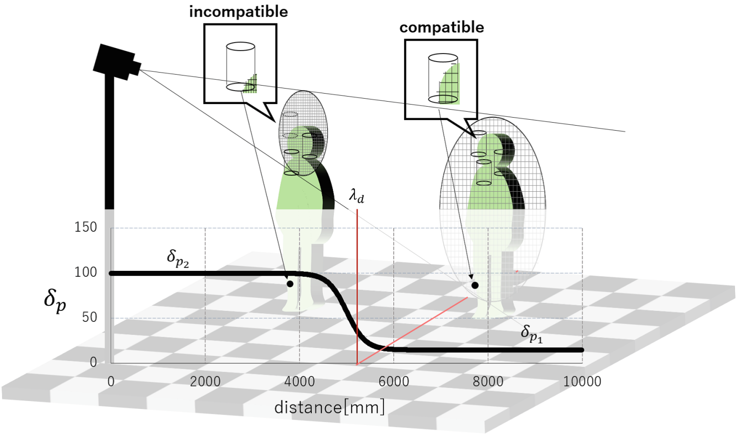

**Fig. 6.** Parameter adjustment

$\lambda_s$ and $\delta_f$ could be adjusted by use of the similar mechanism above mentioned successfully in our experiments.

# 5 Experiment

## 5.1 Specifications

Since the developed prototype system is a total and somewhat complex one for the behavior analysis of shoppers, we could not find any other one developed for the same purpose so far. Therefore, for this reason, we could not include any simple comparison with the other methods in this paper. In the lab room, we have installed a stereo sensor near the ceiling for simulation of tracking multiple shoppers in the small shops. Eight zones were prepared by desks, pillars, and walls in the observation spaces together with different product-like items, such as stuffed toys and stationeries etc. Five walking persons have performed some types of shoppers, each of whom simply walks, searches around for their products, walks and stops often to be interested in their products, walks with accompanying persons, repeatedly stands and crouches, and then frequently picks up the products. In addition, some persons have entered twice into the observation space. Table 1 shows the specifications of the cameras installed in the stereo sensor. We have acquired the RGB-D image sequence of the scene in which the maximum five persons could walk at the same time as a typical complicated situation such as persons passing and occlusion is frequently occurred between them. We have tried analyzing the shopper's behaviors with using parameters shown in Table 2.

**Table 1.** Camera functions

| Param | Value |
|---|---|
| Angle of dip | 30° |
| Baseline length | 150 mm |
| Size of image | 640 × 480 |
| Frame rate | 10 fps |

**Table 2.** Experimental specifications

| Param | Value | Param | Value | Param | Value |
|---|---|---|---|---|---|
| $\alpha_c$ | 0.48 | $\delta_{p1}$ | 10 | $\lambda_{s1}$ | 700 mm |
| $\alpha_h$ | 0.42 | $\delta_{p2}$ | 80 | $\lambda_{s2}$ | 1800 mm |
| $\alpha_p$ | 0.10 | $\delta_{f1}$ | 5 | $\lambda_g$ | 180 mm |
| $\alpha_\gamma$ | 0.38 | $\delta_{f2}$ | 20 | $K_r$ | 0.0028 |

## 5.2 Results and Discussion

In the experiments, 6 PFs were generated for the same image data sequence, where the number of person who appeared in the measurement space was 6 (one appeared twice). As the result, we have confirmed that all of the generated PF could track their corresponding target persons without any losing. Figure 7 shows the sampled shots of the results. The colored bold quadrilaterals and the finer ones show the H-masks and the particle, respectively, both of which are projected onto the camera plane through the SWT. We could see that continuous tracking can be realized without losing, even if people pass each other. Table 3 shows

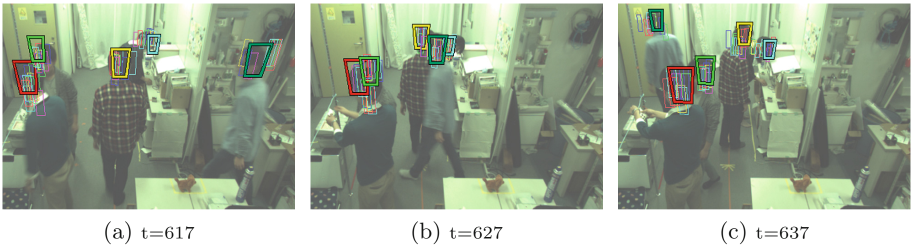

**Fig. 7.** Experimental results (Color figure online)

**Table 3.** Distance error of the system

| | xy | z |
|---|---|---|
| ID1 | 166 | 118 |
| ID2 | 88 | 135 |
| ID3 | 101 | 123 |
| ID4 | 59 | 123 |
| ID5 | 100 | 130 |
| ID6 | 98 | 139 |
| Average | 102 | 128 |

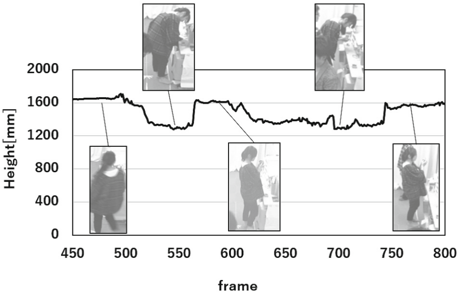

**Fig. 8.** Measurement trajectory (along z)

the distance errors in the xy plane as 102 mm and the z direction as 128 mm of each ID respectively, which were not so large in order to use in supermarkets and so on.

Due to the simple way of locating the H-mask, the error of z direction is larger than the ones in the xy plane. However, in Fig. 8, we have shown some measured profiles of variation of height of the head in each frame. We could see that the estimated values of the head could represent the person postures, such as standing, crouching, or being seated.

From the above results, one could find that by use of the proposed method described in this paper it may be possible to realize the simultaneous and robust tracking of multi-persons in the real environments.

## 6 Conclusions

A robust tracking approach of multiple walkers was proposed by using RGB-D image sequence obtained from a stereo sensor. An 'Elite-type' particle filter can be adopted in the proposed method, where three likelihoods based on color, height, and trajectory are effectively utilized and one can estimate positions of heads of the target persons by using some specialized mask operation. We designed our own state transition model for state management of PF for their application to the real facility where any target to track often changes its situation very frequently. In addition, we proposed a unique mechanism to adjust parameters according to the camera distance, which is one of the important process in using any stereo sensors. Experimental results simulating a real store showed the effectiveness of the proposed method.

## References

1. Hampapur, A., Brown, L., Connell, J., Ekin, A., Haas, N., Lu, M., Merkl, H., Pankanti, S.: Smart video surveillance: exploring the concept of multiscale spatiotemporal tracking. IEEE Sig. Process. Mag. **22**, 38–51 (2005)
2. Jayawardena, C., Kuo, I.H., Unger, U., Igic, A., Wong, R., Watson, C.I., Stafford, R.Q., Broadbent, E., Tiwari, P., Warren, J., Sohn, J., MacDonald, B.A.: Deployment of a service robot to help older people. In: 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 5990–5995. IEEE Press, Taipei (2010)
3. Fieguth, P., Terzopoulos, D.: Color-based tracking of heads and other mobile objects at video frame rates. In: 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 21–27. IEEE Press (1997)
4. Nguyen, H.T., Worring, M., Van Den Boomgaard, R.: Occlusion robust adaptive template tracking. In: Eighth IEEE International Conference on Computer Vision, vol. 1, pp. 678–683. IEEE Press (2001)
5. Etchuya, T., Nara, H., Kaneko, S., Li, Y., Miyoshi, M., Fujiyoshi, H., Shishido, K.: Integration of image and ID-POS in ISZOT for behavior analysis of shoppers. In: Tutsch, R., Cho, Y.-J., Wang, W.-C., Cho, H. (eds.) Progress in Optomechatronic Technologies. LNEE, vol. 306, pp. 3–14. Springer, Cham (2014). doi:10.1007/978-3-319-05711-8_1
6. Nummiaro, K., Koller-Meier, E., Van Gool, L.: An adaptive color-based particle filter. Image Vis. Comput. **21**, 99–110 (2003)
7. Wang, J., Chen, X., Gao, W.: Online selecting discriminative tracking features using particle filter. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 1037–1042. IEEE Press (2005)
8. Zhao, X., Satoh, Y., Takauji, H., Kaneko, S.: Hybrid feature and adaptive particle filter for robust object tracking. World Acad. Sci. Eng. Technol. **59**, 2486–2491 (2011)
9. Li, Y., Ito, M., Miyoshi, M., Fujiyoshi, H., Kaneko, S.: Human detection using smart window transform and edge-based classifier. In: Proceedings of JSPE Semestrial Meeting, vol. 2011A, pp. 920–921 (2011). (in Japanese)
10. Swain, M.J., Ballard, D.H.: Indexing via color histograms. In: Third International Conference on Computer Vision, pp. 390–393. IEEE Press (1990)