LIDAR-based Body Orientation Estimation by Integrating Shape and Motion Information

Masanobu Shimizu[†], Kenji Koide[†], Igi Ardiyanto[‡], Jun Miura[†], and Shuji Oishi[†]
[†]Department of Computer Science and Engineering
Toyohashi University of Technology
[‡]Department of Electrical Engineering and Information Technology
Faculty of Engineering, Universitas Gadjah Mada

***Abstract*— Body orientation gives useful information on assessing a human's state and/or predicting his/her future actions. This paper presents a method of reliably estimating human body orientation using a LIDAR on a mobile robot by integrating shape and motion information. A shape database is constructed by collecting body section shape data from various viewpoints. The result of matching between an input shape with the database is combined with a UKF-based tracker which utilizes a relationship between the body orientation and the motion orientation. The experimental results shows the effectiveness of the proposed method.**

I. Introduction

There is an increasing demand for personal service robots that attend people and provide use-specific services. Such a robot has to have various functions such as person detection and identification. There are many works on human detection and tracking using LIDAR (Laser Imaging Detection and Ranging) [1], [2], images [3], [4], or depth cameras [5], [6]. These works are mainly concerned with human position and motion. In addition to these kinds of information, human body orientation could provide additional information such as a possible direction to move. Therefore this paper deals with body orientation estimation using LIDAR for mobile robots.

Glas et al. [7] developed a method of estimating body orientation by using multiple LIDARs observing a human from various directions. They use a person model with arms and estimate the pose by combining model-based shape prediction with laser scan data. Matsumoto et al. [8] proposed to use sectional-contour for determining human posture such as standing, sitting, and bowing, obtained from multiple LIDARs. Since the LIDARs in these works are supposed to be fixed at different positions, they cannot be directly applicable to mobile robots. Weinrich et al. [9] developed a method of classifying human upper body orientation using HOG [3] and an SVM decision tree. The resolution of image-based orientation estimation is usually not as high as LIDAR-based methods.

When a person walks, his/her moving direction and body orientation are correlated to some extent, and this correlation generally becomes larger at a higher speed. Chen et al. [10] used this observation in a combined person tracking and body orientation estimation using particle filter in visual surveillance. Ardiyanto and Miura [11] dealt with a similar problem by appearance-motion integration using UKF (unscented Kalman filter). Svenstrup et al. [12] used an observation that a person walking at a higher speed changes its orientation less probably for making the orientation estimate more accurate, although they did not use appearance nor shape information.

In this paper, we propose a method of estimating human body orientation by integrating shape information obtained by a LIDAR and motion information obtained by a UKF (unscented Kalman filter) tracker. The shape-motion integration part is based on our previous method [11].

The rest of the paper is organized as follows. Sec. II describes a shape-based orientation estimation using a 2D LIDAR. Sec. III describes our UKF formulation in shape-motion integration for more reliable body orientation estimation. Sec. IV shows experiments for accuracy validation and those in human tracking situation. Sec. V concludes the paper and discusses future work.

II. Shape-Based Body Orientation Estimation

A. Body shape models

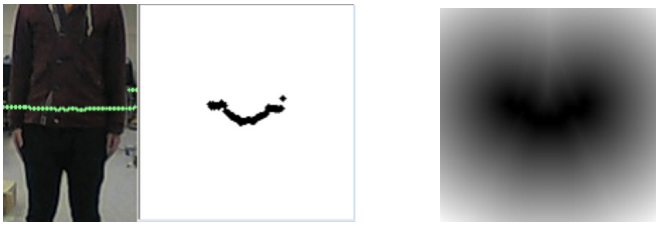
We use a visible part of the section of human body observed by a 2D LIDAR for orientation estimation. We make a set of such data in advance as model data. The set is generated by observing a human body at a certain height from omnidirectional viewpoints. Supposing a robot is following an adult, we put a LIDAR at the height of $85\,cm$ and at $2\,m$ distance. We take data at 36 viewpoints with $10\,deg$ interval. We compare an input shape data with those in the database to estimate the body orientation.

We developed a data acquisition system, which is composed of a motorized turn table with a visual angular position estimation. The accuracy of turn table angle measurement is less than one degree.

B. Body orientation estimation by shape matching

The first step of body orientation estimation is to match an input shape with those in the model database. All models and input data are represented as binary images. Fig. 1(a) shows a scene of obtaining a scan and its image representation, in which one pixel corresponds to a $1\,cm \times 1\,cm$ square region.

An input image is compared with a set of images in the database. Since each model image is specific to a certain



(a) Observation of body shape (green line) and its binary image representation.

(b) Model image after applying distance transform.

Fig. 1. Shape model acquisition and representation.

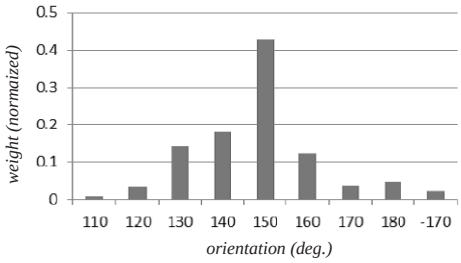
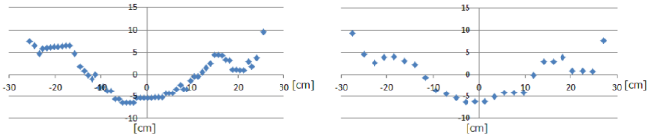


Fig. 2. An example histogram of normalized weights.



(a) LIDAR scan (2m)

(b) LIDAR scan (5m)

Fig. 3. Comparison of LIDAR data for a person at different distances.

viewing direction, only translation is considered in matching the input image and each model image. Here we use its simplified version, that is, calculate the sum of the two distances: d_m is the averaged distance from each input edge to its nearest model edge, and d_i is that from each model edge to its nearest input edge.

After aligning the input and the model image with their center-of-mass positions, we search a certain region (currently, 7×7 region) for the minimum-distance displacement. This can be viewed as a simplified version of the Hausdorff distance-based edge matching strategy [13].

One way to determine the body orientation is to choose the model image with the lowest distance and thus the corresponding viewing angle. This, however, has two drawbacks. One is that the estimated angle become discrete and the other is the sensitivity to noise. We therefore introduce a weighting function which converts a distance to a weight and calculate the weighted mean as an estimate. We also calculate the variance of the estimation; such a variance information is necessary for being used in a statistical framework like Kalman filter. We use the following equations to calculate the estimated orientation θ and its variance σ_θ^2:

$$w_i = e^{-\kappa(d_m+d_i)}, \quad (1)$$

$$\theta = \frac{\sum_{i=1}^{N} w_i i}{\sum_{i=1}^{N} w_i}, \quad (2)$$

$$\sigma_\theta^2 = \frac{\sum_{i=1}^{N} w_i (i-\theta)^2}{\sum_{i=1}^{N} w_i}, \quad (3)$$

where i is a discrete orientation ($i = 1, \cdots, N; N = 36$), w_i is the weight for i. Fig. 2 shows an example histogram of normalized weights for an observation. Note that the weights are calculated for orientations in the $\pm 40\,deg$ range from the previous orientation. At the same time, the periodicity of angles is taken into account to compute a circular mean from directional statistics.

C. Evaluation of body orientation estimation using one scan of LIDAR data

We here evaluate the accuracy of shape-based body orientation estimation, by considering several factors which may affect to the accuracy, that is, distance, body size, and clothing.

1) Effect of distance to the person: The number of measured 3D points decreases as the distance to the person becomes larger. We compare the accuracy when the person is at either $2\,m$ and $5\,m$ distant from the robot. Fig. 3 shows an example of LIDAR data for these distances.

For the person with whom a set of model data was acquired, we acquired data set for testing for each orientation, ranging from $0\,deg$ to $350\,deg$ with $10\,deg$ interval (as in the case of model data acquisition), and performed scan-to-scan comparisons to estimate the body orientation. For each set of input LIDAR scans, the discrete orientation which is supported by the highest number of estimations is selected as the result. Considering the estimation is sometimes reversed by $180\,deg$ because the body shape looks very similar observed from the front and the back, we classify the difference between the input and the model into the five cases: $0\,deg$, $\pm 10\,deg$, $180\,deg$, $180 \pm 10\,deg$, and otherwise. Table I summarizes the results. Since we can use face recognition for discriminate 0 degree and 180 degree cases, as explained later, we can conclude that even in $5\,m$ cases, the orientation error is expected to be mostly within 10 deg

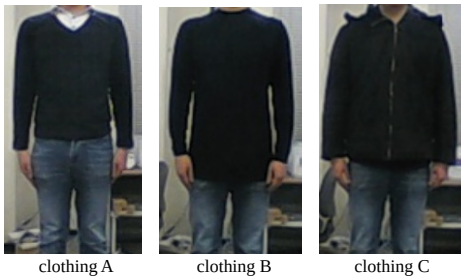
2) Effect of clothing: Changing clothing may change the body shape. We compared three cases shown in Fig. 4. Using clothing A as the model and the others for testing, we estimate the mean and the maximum absolute error. Table II summarizes the results showing clothing change can decrease the accuracy of the body orientation estimation, however, the proposed method still works well.

3) Effect of person to person difference: We take shape data for three persons with the height of $160\,cm$, $170\,cm$ and $180\,cm$. For each person, we collected two sets of shape data for all orientations, one is for making a model and the other for testing. We then calculate the absolute estimation error for every combination of the model person and the test person. Table III summarizes the result showing the height difference between model and test data has a bad influence on the body

TABLE I

CLASSIFICATION OF ESTIMATION RESULTS FOR DIFFERENT DISTANCE TO THE PERSON.

difference (deg)	0	± 10	180	180 ± 10	others
person is at $2\,m$	27	7	2	0	0
person is at $5\,m$	16	10	6	3	1



clothing A clothing B clothing C

Fig. 4. Three different clothing.

orientation estimation. However, the table also include the case where the averaged image of all model images is used as the model, which is shown to reduce the errors in average. The result is similar to the case where the subject wears different clothing.

III. SHAPE-MOTION INTEGRATION FOR BODY ORIENTATION ESTIMATION

A. UKF-based integration

Integration of shape and motion is performed by introducing body orientation terms and information integration rules into a human tracker. The tracker uses the Unscented Kalman Filter (UKF) [14], which is a non-linear statistical filter and known to be more accurate in many cases than Extended Kalman filter. Another superior point of UKF is that it does not require Jacobians, which are sometimes hard to derive.

B. Formulation

We basically follow the formulation by Ardiyanto and Miura [11] but modify it so that the robot-centered coordinate system is used.

1) State transition equation: State vector $\boldsymbol{x}_t$ is defined by:

$$\boldsymbol{x}_t = \left(x_t, y_t, \dot{x}_t, \dot{y}_t, \phi_t^M, \phi_t^S, \theta_t\right)^T, \tag{4}$$

where (x_t, y_t) and $(\dot{x}_t, \dot{y}_t)$ are the position and the velocity of a person in the robot local coordinate, ϕ_t^M and ϕ_t^S are the estimated orientation by the motion information and the shape information, respectively, and θ_t is the orientation estimation result that integrates ϕ_t^M and ϕ_t^S as described later.

Fig. 5 shows the relationships between the robot and the person positions and the robot motion. Supposing a constant velocity model, the person position/velocity part of the state

TABLE II

THE MEAN AND THE MAXIMUM ESTIMATION ERROR FOR DIFFERENT CLOTHING WHEN CLOTHING A IS USED AS MODEL.

Clothing for testing	clothing B	clothing C
mean absolute error (deg)	3.21	5.96
maximum absolute error (deg)	8.21	17.0

TABLE III

COMPARISON OF ERRORS (IN DEG.) FOR VARIOUS COMBINATIONS OF MODEL AND TEST PERSON.

model person	test person	averaged error	maximum error
A	A	2.96	5.98
	B	3.98	8.48
	C	8.32	15.1
B	A	4.26	10.7
	B	1.70	3.35
	C	6.09	15.3
C	A	6.66	17.9
	B	3.95	7.52
	C	2.86	6.41
Averaged	A	3.33	10.4
	B	2.32	4.41
	C	5.16	12.3

equation is given by:

$$\begin{aligned} \hat{x}_{t+1} &= x_t + \dot{x}_t \Delta t, \\ \hat{y}_{t+1} &= y_t + \dot{y}_t \Delta t, \\ \hat{\dot{x}}_{t+1} &= \dot{x}_t, \\ \hat{\dot{y}}_{t+1} &= \dot{y}_t, \end{aligned} \tag{5}$$

where Δt is the cycle time and $\hat{\cdot}$ indicates the position/velocity in the robot coordinates at time t. Then, applying the coordinate transformation due to the robot motion gives the following relationship:

$$\begin{aligned} x_{t+1} &= (\hat{x}_{t+1} - \Delta x)\cos\Delta\theta + (\hat{y}_{t+1} - \Delta y)\sin\Delta\theta, \\ y_{t+1} &= -(\hat{x}_{t+1} - \Delta x)\sin\Delta\theta + (\hat{y}_{t+1} - \Delta y)\cos\Delta\theta, \\ \dot{x}_{t+1} &= \hat{\dot{x}}_{t+1}\cos\Delta\theta + \hat{\dot{y}}_{t+1}\sin\Delta\theta, \\ \dot{y}_{t+1} &= -\hat{\dot{x}}_{t+1}\sin\Delta\theta + \hat{\dot{y}}_{t+1}\cos\Delta\theta. \end{aligned} \tag{6}$$

Combining eqs. (5) and (6) with additional noise terms, we obtain:

$$\begin{aligned} x_{t+1} &= (x_t + \dot{x}_t\Delta t - \Delta x)\cos\Delta\theta \\ &+ (y_t + \dot{y}_t\Delta t - \Delta y)\sin\Delta\theta + \epsilon_x, \\ y_{t+1} &= -(x_t + \dot{x}_t\Delta t - \Delta x)\sin\Delta\theta \\ &+ (y_t + \dot{y}_t\Delta t - \Delta y)\cos\Delta\theta + \epsilon_y, \\ \dot{x}_{t+1} &= \dot{x}_t\cos\Delta\theta + \dot{y}_t\sin\Delta\theta + \dot{\epsilon}_x, \\ \dot{y}_{t+1} &= -\dot{x}_t\sin\Delta\theta + \dot{y}_t\cos\Delta\theta + \dot{\epsilon}_y. \end{aligned} \tag{7}$$

For orientation elements, we use the following equations:

$$\begin{aligned} \phi_{t+1}^M &= \arctan\left(\frac{-\dot{x}_t\sin\Delta\theta + \dot{y}_t\cos\Delta\theta}{\dot{x}_t\cos\Delta\theta + \dot{y}_t\sin\Delta\theta}\right) + \epsilon_{\phi^M}(v_t), \\ \phi_{t+1}^S &= \phi_t^S - \Delta\theta + \epsilon_{\phi^S}, \\ \theta_{t+1} &= \phi_t^S + \omega\,\frac{1 - e^{-v_t}}{1 + e^{-v_t}}\left(\phi_t^M - \phi_t^S\right) + \epsilon_\theta, \end{aligned} \tag{8}$$

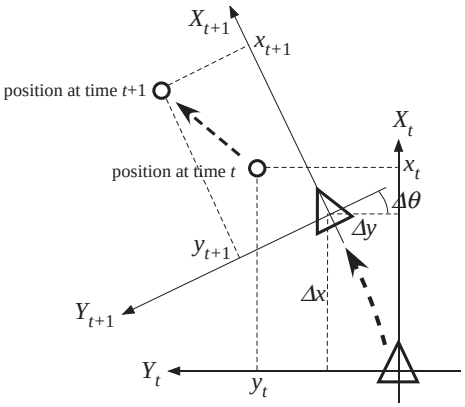
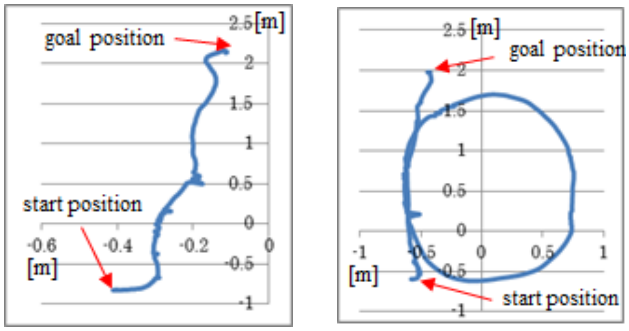


Fig. 5. Robot and person motion and coordinate transformation.

Fig. 6. Accuracy evaluation environment with VICON.



(a) straight movement. (b) circular movement.

Fig. 7. Example motion measurements.

where v_t is the speed of the person calculated from $(\dot{x}_t, \dot{y}_t)$ and ω is a constant. The last line in eq. (8) realizes the shape-motion integration; as the person's speed increases, motion information is used more for orientation estimation. Constant ω controls the influence of the speed to orientation estimation. When $\omega = 0$, only shape information is used, while $\omega = 1$, motion information is maximally utilized. We will determine the value of ω experimentally, as described later. ϵ_x, ϵ_y, $\epsilon_{\dot{x}}$, $\epsilon_{\dot{y}}$, ϵ_{ϕ^S} and $\epsilon_{\dot{theta}}$ are fixed Gaussian noise terms. $\epsilon_{\phi^M}(v_t)$ follows a zero-mean Gaussian with variance $\sigma_v^2 e^{-vt}$; the error in motion orientation prediction becomes smaller as the person speed increase.

2) Observation equation: The robot measures the person position (x_t^L, y_t^L) and the body orientation (ϕ_t^L) from LIDAR data; the observation vector $\boldsymbol{y}_t$ is given by:

$$\boldsymbol{y}_t = \left(x_t^L, y_t^L, \phi_t^L\right)^T. \tag{9}$$

Then the observation equation is defined as:

$$\boldsymbol{y}_t = \begin{pmatrix} x_t + \epsilon_x^L \\ y_t + \epsilon_y^L \\ \phi_t^S + \epsilon_{\phi^L} \end{pmatrix} \tag{10}$$

ϕ_t^S and the variance of ϵ_{ϕ^L} are given by the weighted mean and the weighted variance as explained in Sec. II-B.

IV. Experimental Results

A. Accuracy evaluation with motion capture device

Sec. II-C has described evaluation results for the case where one scan of LIDAR data is used, that is, those for the shape only estimation. To evaluate the effect of shape-motion integration, we used a motion capture device (VICON M2 Camera system) for measuring the body orientation on-line.

Fig. 6 shows the experimental environment. Three markers are attached to a subject at the waist level. From the positions of the markers, we can measure the body orientation as well as body position. Four motion patterns were tested: stand still, turn, straight, and circular. Example motion measurements were shown in Fig. 7.

We first examined the effect of ω in eq. (8), which controls how shape and motion information are integrated, on the estimation accuracy for the straight and circular motion. Fig. 8 summarizes the averaged error for various ω's. Larger ω's are effective for the straight motion while smaller values are effective for the circular one. Based on this result, we determined to use 0.3 as ω.

Table IV compiles the accuracy data for all motion patterns. For the straight and the circular motion, we also compare the estimation only with shape information and that with shape-motion integration. Since the experiments were done in the environment which is different from the one for generating shape models, the averaged error for "stand still" is considered as *offset* between the environments, and subtracted from the averaged errors in the other motion patterns. The results show that shape-motion integration is effective especially when a person walks relatively straightly.

B. On-line tracking and orientation estimation experiments

1) Implementation on the robot: We implemented the proposed method on a real mobile robot, *GRACE-I*. The robot can detect and follow a target person while estimating his/her body orientation. Fig. 9 shows the configuration of processing steps.

GRACE-I is equipped with four LIDARs (Hokuyo UTM-30LX) at the heights of $30\,cm$ and $85\,cm$ so that $360\,deg$ scans at two levels are obtained. These two heights are chosen to correspond to the body and the feet positions. The robot has cameras for person identification based on face, clothing, and other appearance features.

We measured the actual processing time in the case where

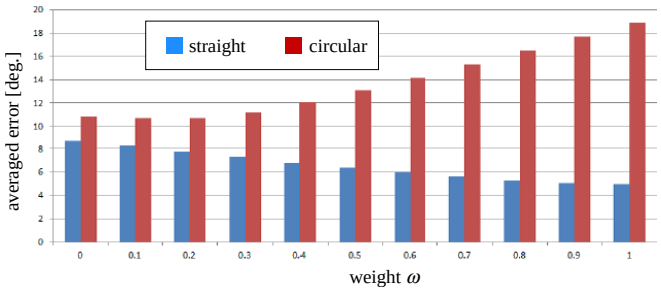


Fig. 8. Effect of ω on the body orientation estimation accuracy.

TABLE IV
ERRORS IN VARIOUS MOTION PATTERNS (IN DEG.).

motion pattern		averaged error	std. dev.
stand still		6.34	0.83
turn		7.40	4.89
forward	shape only	8.70	7.01
	shape-motion integration	7.28	5.90
circular	shape only	12.2	8.88
	shape-motion integration	11.2	8.74

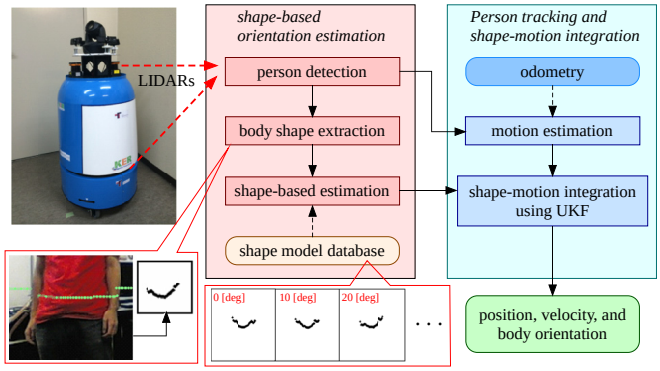


Fig. 9. Outline of tracking and orientation estimation.

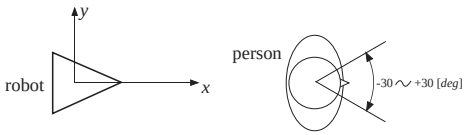


Fig. 10. Orientation range for reversing the measured orientation.

three to four persons are exist in the environment. The average processing time for one frame was 37.6 $msec$ and that for the shape matching and the UKF update per person was 2.39 $msec$.

2) Face recognition and orientation correction: As described in Sec. II-C, we use a face recognition method to discriminate 0 degree and 180 degree cases. We apply the Viola-Jones face detector [15] with our illumination normalization method [16] for detecting faces in various illumination conditions. When a face is detected in the predicted head region, and when the estimated body orientation is within the range between $-30\,deg$ and $30\,deg$ as shown in Fig. 10, we reverse the orientation, that is, add $180\,deg$ to the estimation result.

3) Tracking and orientation estimation: Fig. 11 shows snapshots of tracking a person with estimating his body orientation. The left column indicates the images from a camera on the robot, in which cylinders and arrows indicate the person positions and body orientations, respectively. The right column indicate the maps with person information. Circles, red one among them, and arrows indicate detected persons, the target person, and the target's body orientation, respectively. In Fig. 11(a), the target person is standing still and facing to the robot. In Fig. 11(b)-(d), the robot is following the target person and his body orientations are well estimated.

Fig. 12 shows the results of tracking and body orientation estimation of multiple persons. The top-right corner of the images shows the 2D mapping of persons and obstacles. The results show the body orientations are well estimated.

V. CONCLUSIONS AND FUTURE WORK

This paper has described a body orientation estimation method by shape-motion information. The shape information is extracted from a 2D LIDAR scan and then compared with the shape models to generate an orientation estimation with its variance. Considering the fact that the body orientation and the motion direction are correlated to some extent, the shape-based orientation information is combined with motion information in a UKF framework to produce a better orientation estimate. We evaluated the method with a motion capture system as a ground truth and have shown the shape-motion integration increases the accuracy of body orientation estimation. The method is also implemented on a real robot and successfully applied to a following robot scenario.

Although we have examined the effect of body size and clothing on the estimation accuracy, the effects of larger disturbances such as belongings (e.g., shoulder bag or back pack) have not been evaluated. In such a case, if the change of the body shape at the height of 2D LIDAR scans is large, the estimation could be much worse. We are also planning to use a 3D LIDAR so that informative parts of scans can be selectively utilized for a better estimation. Application of body orientation estimate to recognition of various human interactions such as two persons talking closely is also future work.

ACKNOWLEDGMENT

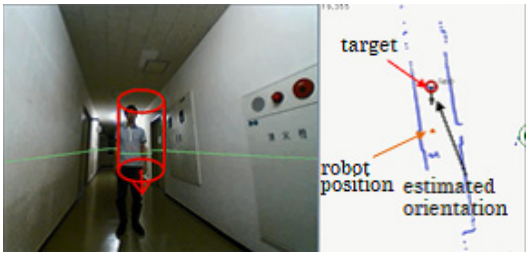
This work was supported by JSPS KAKENHI Grant Number 25280093.

REFERENCES

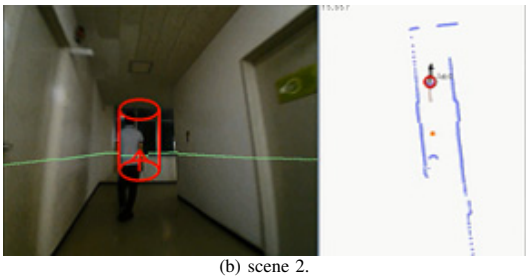
[1] K.O. Arras, O.M. Mozos, and W. Burgard. Using Boosted Features for the Detection of People in 2D Range Data. In *Proceedings of the 2007 IEEE Int. Conf. on Robotics and Automation*, pp. 3402–3407, 2007.

[2] Z. Zainudin, S. Kodagoda, and G. Dissanayake. Torso Detection and Tracking using a 2D Laser Range Finder. In *Proceedings of Australasian Conf. on Robotics and Automation 2010*, 2010.

[3] N. Dalal and B. Briggs. Histograms of Oriented Gradients for Human Detection. In *Proceedings of 2005 IEEE Conf. on Computer Vision and Patttern Recognition*, pp. 886–893, 2005.

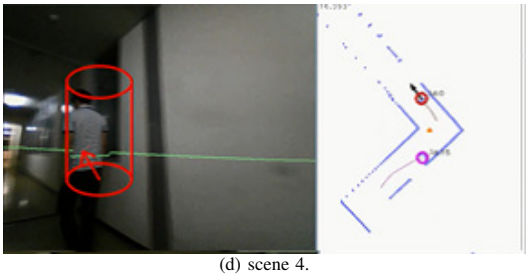


(a) scene 1.



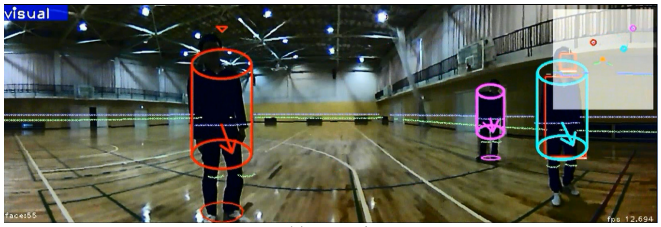
(b) scene 2.

(c) scene 3.

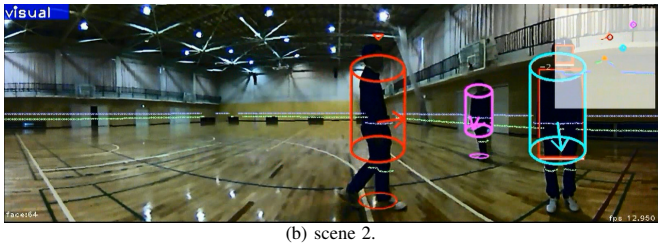


(d) scene 4.

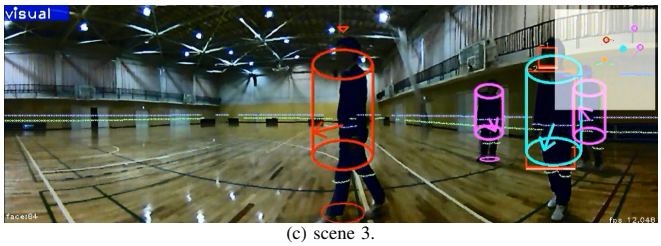
Fig. 11. Tracking and body orientation estimation by the robot.



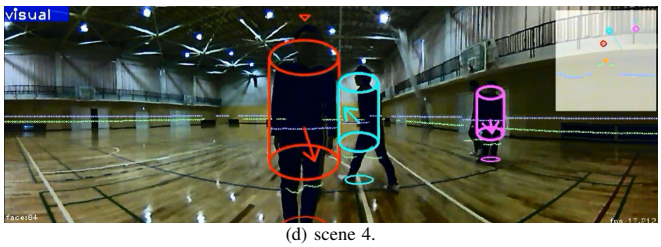
(a) scene 1.



(b) scene 2.



(c) scene 3.



(d) scene 4.

Fig. 12. Tracking and body orientation estimation of multiple persons.

[4] J. Satake and J. Miura. Robust Stereo-Based Person Detection and Tracking for a Person Following Robot. In *Proceedings of ICRA-2009 Workshop on Person Detection and Tracking*, 2009.

[5] L. Spinello and K.O. Arras. People Detection in RGB-D Data. In *Proceedings of the 2011 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 3838–3843, 2011.

[6] M. Munaro and E. Menegatti. Fast RGB-D People Tracking for Service Robots. *Autonomous Robots*, Vol. 37, No. 3, 2014.

[7] D.F. Glas, T. Miyashita, H. Ishiguro, and N. Hagita. Laser-Based Tracking of Human Position and Orientation using Parametric Shape Modeling. *Advanced Robotics*, Vol. 23, No. 4, pp. 405–428, 2009.

[8] T. Matsumoto, M. Shimosaka, H. Noguchi, T. Sato, and T. Mori. Pose Estimation of Multiple People using Contour Features from Multiple Laser Range Finders. In *Proceedings of 2009 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 2190–2196, 2009.

[9] C. Weinrich, C. Vollmer, and H.-M. Gross. Estimation of Human Upper Body Orientation Estimation for Mobile Robotics using an SVM Decision Tree on Monocular Images. In *Proceedings of 2012 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 2147–2152, 2012.

[10] C. Chen, A. Heili, and J.-M. Odobez. Combined Estimation of Location and Body Pose in Surveillance Video. In *Proceedings of 2011 IEEE Int. Conf. on Advanced Video and Signal-Based Surveillance*, pp. 5–10, 2011.

[11] Igi Ardiyanto and J. Miura. Partial Least Squares-based Human Upper Body Orientation Estimation with Combined Detection and Tracking. *Image and Vision Computing*, Vol. 32, No. 11, pp. 904–915, 2014.

[12] M. Svenstrip, S. Tranberg, H.J. Andersen, and T. Bak. Pose Estimation and Adaptive Robot Behavior for Human-Robot Interaction. In *Proceedings of 2009 IEEE Int. Conf. on Robotics and Automation*, pp. 3571–3576, 2009.

[13] D.P. Huttenlocher, G.A. Klanderman, and W.J. Rucklidge. Comparing Images using the Hausdorff Distance. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 9, pp. 850–863, 1993.

[14] S.J. Julier and J.K. Uhlmann. Unscented Filtering and NonLinear Estimation. *Proceedings of IEEE*, Vol. 92, No. 3, pp. 401–422, 2004.

[15] P. Viola and M.J. Jones. Robust Real-Time Face Detection. *Int. J. of Computer Vision*, Vol. 57, No. 2, pp. 137–154, 2004.

[16] Bima Sena Bayu Dewantara and J. Miura. Fuzzy-based Illumination Normalization for Face Recognition. In *2013 IEEE Workshop on Advanced Robotics and its Social Impacts*, pp. 131–136, 2013.