Advanced Robotics, Vol. 29, No. 22, pp. 1483-1495, 2015.

Original Paper

Specific Person Tracking using 3D LIDAR and ESPAR Antenna for Mobile Service Robots

K. Misu and J. Miura

Department of Computer Science and Engineering, Toyohashi University of Technology

(*v1.0 released January 2013*)

The ability of detecting and following a specific person is indispensable for mobile service robots. Many image-based methods have been proposed for person detection and identification; however they are sometimes vulnerable to illumination changes. This paper therefore proposes a novel approach to the problem, namely, using 3D LIDARs for person detection and identification and a directivity-controllable antenna (called *ESPAR antenna*) for localizing a specific person even under long-term occlusion and/or out-of-view situations. A sensor fusion framework, combined with an adaptive state-based strategy switching, has also been developed for achieving a reliable person following. Experimental results in actual outdoor environments show the effectiveness of the proposed framework.

Keywords: Person detection and tracking, Mobile robot, 3D LIDAR, ESPAR antenna

1. Introduction

There is an increasing demand for mobile service robots that provide services such as attending and guiding to elderly or children. Such a robot has to know the location of a specific person to whom services are provided. Sensor-based person detection and identification are therefore indispensable for mobile service robots.

Vision has frequently been used for person detection and tracking. Various local image features such as HOG (histogram of oriented gradients) [1] and LBP (local binary patterns) [2], or multiple features (e.g., [3, 4]) have been used for establishing classifiers with learning algorithms such as SVM (Support Vector Machines) [5] and AdaBoost [6]. Stereo data [7] or motion data (e.g., optical flow [8]) are also sometimes used. Although vision can provide rich information, it is inherently vulnerable to illumination changes and weak in very dark situations.

LIDAR (laser imaging detection and ranging) has been a primary sensor for mobile robots for its reliability and ranging accuracy. There are many LIDAR-based people detection methods [9–12], which use a single or a few horizontal range scans. Recent development of high-definition LIDARs has realized more reliable people detection even in outdoor scenes [13–15]. Combination with image data further improves the performance [16] as long as the illumination condition is acceptable.

Concerning the person identification, simple methods which utilize color or texture information have often been adopted (e.g., [17]), although they have relatively a low discriminative power. More strong features such as face (e.g., [18]) or gait (e.g., [19]) are sometimes used in vision-based identification; such methods, however, require good observation conditions such as good illumination, static camera, occlusion-free scene and so on, and are not very ready to be applied to mobile robots. RFID tags can also be used; however, it is necessary to physically rotate an

*Corresponding author. Email: jun.miura@tut.jp

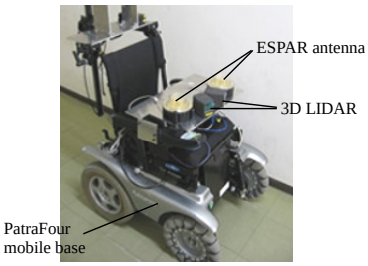


Figure 1. The robot system.

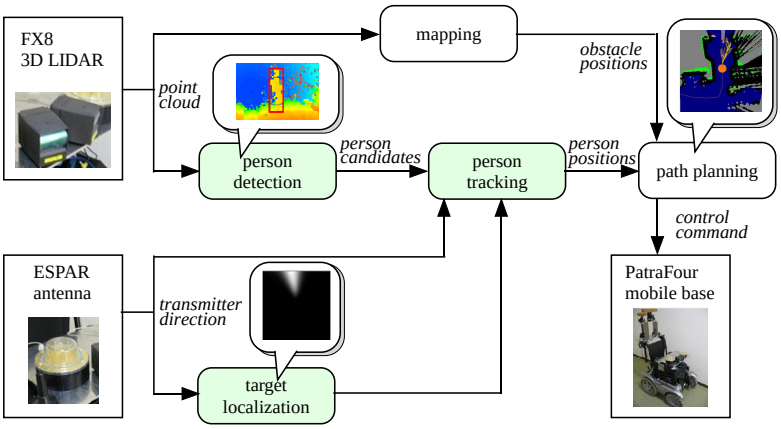


Figure 2. Software configuration. Green-shaded blocks are new modules developed in this research.

antenna [20] or to be equipped with several antennas [21] in order to detect a target in various directions.

In this paper, we propose a new combination of sensors, which is effective in detection and tracking of a specific person. The sensors we use are 3D LIDAR and ESPAR (Electronically-Steerable Passive Array Radiator) antenna [22]. The ESPAR antenna can change its directivity very fast and is suitable for detecting the direction of a radio transmitter in a long range. We mainly use the 3D LIDAR for person detection and tracking, and the ESPAR antenna for identifying a specific person with a radio transmitter but further investigate sensor fusion techniques for reliably continuing a tracking task under difficult situations such as long and full occlusions or out-of-view situations.

The contributions of the paper are: (1) use of a new type of sensor (ESPAR antenna) for identification and localization, (2) a sensor fusion method for the two sensors for a robust specific person detection and tracking, and (3) its implementation and testing on an actual mobile robot in various situations.

The rest of the paper is organized as follows. Sec. 2 describes an overview of our robot system and the proposed method and that of the developed system. Sec. 3 briefly explains a 3D LIDAR-based person detection. Sec. 4 describes a person tracking using the 3D LIDAR and the ESPAR antenna. Sec. 5 describes the target person identification, target person switching, and the detection of lost situation. Sec. 6 describes the localization of the target person using the ESPAR antenna after losing him/her. Sec. 7 shows experimental results of person tracking using the developed system. Sec. 8 summarizes the paper and discusses future work. This paper is an extended version of our previous presentation [23].

2. Overview of the System and the Proposed Method

2.1 *Hardware configuration*

Fig. 1 shows the person following robot system. The mobile base is an electric wheelchair with a four wheel-drive mechanism (PatraFour by Toyota Motor East Japan, Inc.). The wheelchair is modified by the manufacturer so that it can be controlled by a PC by sending commands and receiving odometry data.

We use FX8 by Nippon Signal Co. as 3D LIDAR. This sensor can obtain 97×61 (5197) points with about 4Hz and the maximum detection range is $15\,m$. We use two of this sensor to cover a $120\,deg.$ horizontal field of view.

ESPAR antenna is composed of one monopole at the center and the others on a surrounding circle. It can control its directivity very fast by changing the load reactance electronically [22]. We use a product of Tsuryo Technica Corp. based on this principle and can detect the direc-

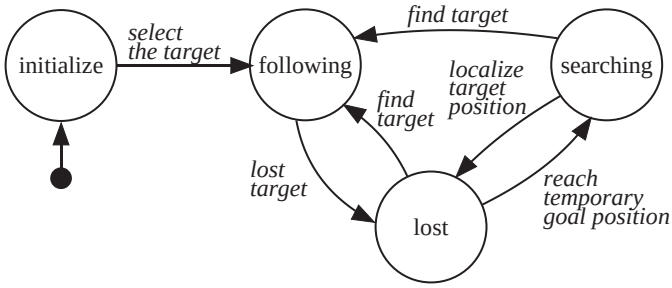


Figure 3. State transition model.

tion of a transmitter with a digitally-encoded signal. This product is designed to return one of twelve discrete directions (i.e., every 30 *deg.*). We use two antennas for increasing reliability with expecting the measurement of the transmitter location by triangulation.

2.2 *Software configuration*

The software configuration of the system is illustrated in Fig. 2. The whole software system is composed of multiple functional modules, shown by rounded rectangles in the figure. Each module is realized as an *RT component* in the *RT-middleware* environment [24], which supports a modularized software development. Brief explanations of the modules are as follows; the first three modules will be explained in detail later.

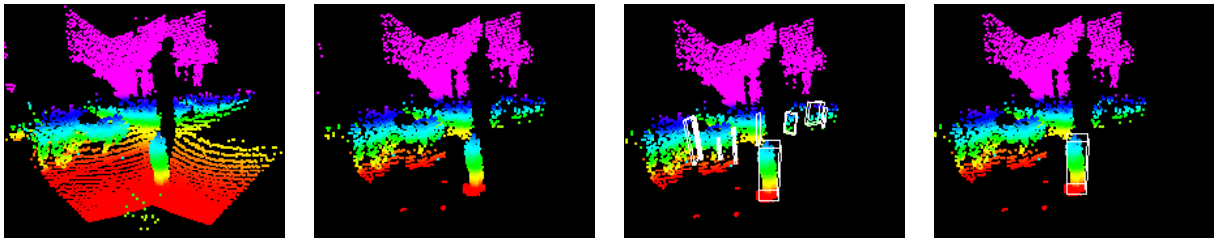
- *Person detection*: Point cloud data is analyzed to detect person candidates using a classifier (see Sec. 3).
- *Person tracking*: Persons are tracked using a set of Kalman filters with target identification and an entropy-based filters management (see Secs. 4 and 5).
- *Target localization*: Possible target positions are determined using ESPAR antenna data in case of lost target (see Sec. 6).
- *Mapping*: A 2D probabilistic occupancy map is constructed and maintained used for path planning.
- *Path planning*: A path to follow the target is generated considering surrounding obstacles and other persons using an RRT-based path planner [25].

2.3 *State transition model for robust tracking*

Robustness is one of the most important aspects in any robotic task execution. In the case of a specific person following task, the robot has to be able to re-detect and track the person after temporarily losing him/her as well as to keep a steady tracking in a normal situation. Since recognition and action operations suitable for each situation may different from each other, a meta-level control such as an event-based behavior selection [26] is necessary.

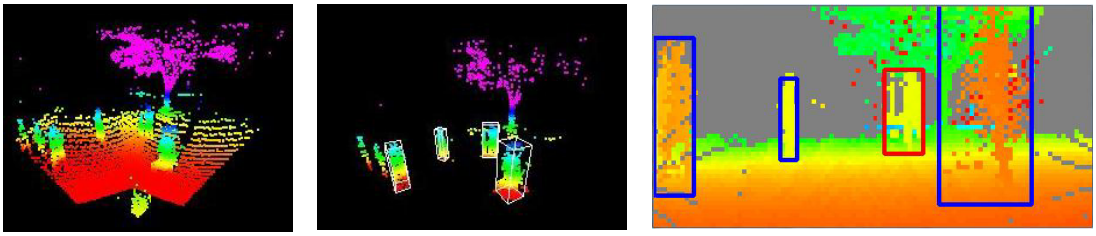
We thus introduce a state-based control with the following four states (see Fig. 3):

- *initialize*: The system starts from this state. It detects person candidates, initiates a Kalman filter for each candidate, calculates the *tracking probability* that each filter is tracking the target using ESPAR antenna data, selects the one which has the largest probability, and transits to the *following* state. This selected Kalman filter is called *tracking filter*.
- *following*: The system is tracking a target person. That is, the tracking filter can localize the target; its positional uncertainty is within a certain range and the target probability is high enough. The robot moves towards the target position. If one of these conditions is violated, move to the *lost* state.



(a) input point cloud (b) ground plane elimination (c) candidate clusters (c) detected person

Figure 4. Person detection steps. Colors indicate the height of 3D point data. While boxes are superimposed on candidates (in (c)) and the detected person (in (d)).



(a) input point cloud (b) detected person (c) detection result in the range image

Figure 5. Multiple person detection.

- *lost*: The system temporarily loses the target person although the tracking filter is still running. The robot moves to the point where the target was detected last while keeping checking all persons' identities. If the target is found again, move back to the *following* state. If the robot reaches that point without re-finding the target, move to the *searching* state.
- *searching*: The system considers the target is completely lost; most probably the target is out of sight of the LIDAR or fully occluded by objects or other persons. The robot stops to calculate a probabilistic target existence grid map, which is used to limit the possible area where the target may exist. If the target is found again in this state, move directly to the *following* state. If not, move back to the *lost* state with an estimate of possible target position.

3. Person Detection

We use a person detection method [15], originally developed for a high-definition LIDAR. The process of person detection is as follows (see Fig. 4).

(1) *ground plane elimination*: 3D points are mapped onto a 2D grid. A cell is classified as an *object cell* if the difference between the highest and the lowest point is larger than a threshold (currently, $0.2\,m$). Otherwise, the cell is classified as the ground and eliminated.
(2) *clustering*: Euclidean clustering is applied to object cells. A parallelepiped is fitted to each cluster by the Rotating Calipers method [27]. The cluster is considered as a person candidate if the dimensions of the parallelepiped satisfy the following conditions (unit is $[m]$):

$$0.8 \leq height \leq 2.0, \quad width \leq 1.2, \quad depth \leq 1.2, \tag{1}$$

(3) *classification*: RealAdaBoost [28] is used for classification. The feature vector is 213-dimensional and includes both shape and reflectance features. Refer to [15] for the details

of the feature.

Fig. 5 shows another case where four persons are detected. For a test data set with 800 range images taken from a mobile robot at about 4Hz, the recall and the precision for this test data are 90.1% and 94.6%, respectively.

4. Person Tracking

All detected persons are tracked independently and simultaneously using a set of Kalman filters. Data association is done based on a position proximity and an appearance similarity. In case of occlusion, data from ESPAR antennas are also used only for the target person to follow.

4.1 *State variable and motion model*

State vector $\boldsymbol{x}_t$ is defined as a four dimensional vector with position (X_t, Y_t) and velocity $(\dot{X}_t, \dot{Y}_t)$. We use a linear motion model represented by:

$$\begin{aligned} \boldsymbol{x}_t &= A\boldsymbol{x}_{t-1} + \boldsymbol{u}, \\ A &= \begin{bmatrix} 1 & 0 & \Delta T & 0 \\ 0 & 1 & 0 & \Delta T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \end{aligned} \tag{2}$$

where ΔT is the cycle time. Deviation from the linear motion is considered by the process noise $\boldsymbol{u}$, which represents acceleration terms and whose covariance matrix is empirically determined. Based on this model, the prediction step of the Kalman filter is done by:

$$\hat{\boldsymbol{x}}_t = A\boldsymbol{x}_{t-1}, \tag{3}$$

$$\hat{P}_t = AP_{t-1}A^T + Q_t, \tag{4}$$

where P_t and Q_t are the covariance matrices of the state and the noise, respectively.

4.2 *Data association*

4.2.1 People identification using reflectance values

The histogram of reflectance values is used for person identification. The histogram is constructed by extracting reflectance values from the point data of a detected person, putting them into 25 bins, and normalizing the frequency values to sum up to one. We adopt an on-line boosting with feature selection [29] for acquisition and update of appearance models.

Fig. 6 illustrates an overview of the classifier. It has N *selectors* and each selector has M weak classifiers. At each learning step with input feature vector $\boldsymbol{f}$, all weak classifiers and the weight w_n for the nth selector are updated. The best weak classifier $h_n^{sel}(\boldsymbol{f})$, which has the smallest error, represents the nth selector. The strong classifier is obtained by the selectors:

$$h^{strong}(\boldsymbol{f}) = \text{sign}\left(\sum_{n=1}^{N} w_n \cdot h_n^{sel}(\boldsymbol{f})\right). \tag{5}$$

Refer to [29] for the details of learning algorithm.

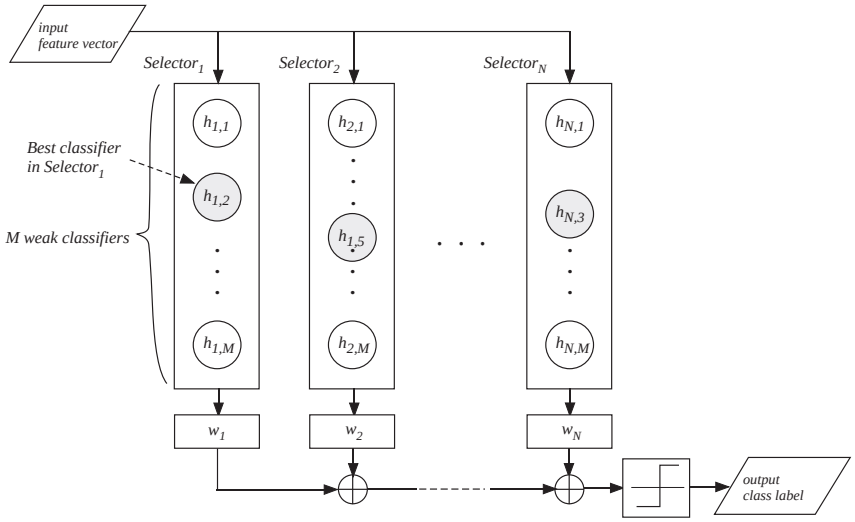


Figure 6. Overview of on-line boosting with feature selection. A modified version from the original one in [29].

We tested this classifier for a data sequence with six persons, with parameters $N = 27$ and $M = 20$. The time for identification is $38.5\,\mu sec.$, that for learning is $1.1\,msec.$ and the averaged reliability is 77.2%.

4.2.2 Data association using proximity and appearance similarity

Two kinds of information are used for data association in the Kalman filter. One is the proximity between a detected person and a filter, evaluated by the Mahalanobis distance using the top-left part (i.e., positional part) of the covariance matrix. The other is the appearance similarity, which is measured by the error of the classifier of the filter.

Let $d_{i,j}$ and e_j be the root-squared Mahalanobis distance and the error estimate of the strong classifier of the jth filter. We define a *joint likelihood* for this combination as:

$$l_{i,j} = N(d_{i,j}; 0, 1)(1 - e_j), \tag{6}$$

where $N(\,\cdot\,; \mu, \sigma^2)$ is the normal distribution with the mean and the variance being μ and σ^2, respectively. A smaller distance and a more reliable classifier give a higher likelihood.

We try to establish one-to-one correspondences among detected persons and filters. First, a set of every feasible correspondence of a person and a filter is enumerated by using the filter's classifier. The best combination with the highest joint likelihood is then chosen to fix, and its person and filter are eliminated from the set. This process is repeated until no persons or no filters exist.

4.3 *Measurement update*

The measurement update step of the Kalman filter is performed with observation z_t and its associated uncertainty R_t by the following equations:

$$\nu_t = z_t - h(\hat{\boldsymbol{x}}), \tag{7}$$

$$S_t = H_t \hat{P}_t H_t^T + R_t, \tag{8}$$

$$K_t = \hat{P}_t H_t^T S_t^{-1}, \tag{9}$$

$$\boldsymbol{x}_t = \hat{\boldsymbol{x}}_t + K_t \nu_t, \tag{10}$$

$$P_t = (I - K_t H_t)\hat{P}_t, \tag{11}$$

where $h(\boldsymbol{x})$ is the observation equation, H is the Jacobian matrix, ν_t and S_t are the innovation and its uncertainty, K_t is the Kalman gain, and $\boldsymbol{x}_t$ and P_t are the updated estimate and uncertainty. Actual calculations differ depending on the following three cases:

- If the corresponding person is detected, the detected position with uncertainty is used for the measurement update; the observation function $h(\boldsymbol{x})$ becomes linear and is defined as:

$$h(\boldsymbol{x}) = H\boldsymbol{x} + \boldsymbol{v}, \tag{12}$$

$$H = \begin{bmatrix} 1\,0\,0\,0 \\ 0\,1\,0\,0 \end{bmatrix}, \tag{13}$$

  where $\boldsymbol{v}$ is the observation noise, the covariance matrix of which is given by:

$$R_t = \begin{bmatrix} \sigma_r^2 & 0 \\ 0 & \sigma_r^2 \end{bmatrix}, \tag{14}$$

  where σ_r^2 is set to $0.01\,m^2$ in this paper.
- If the corresponding person is not detected and if the filter is for the target person, the direction of the target obtained by one of the ESPAR antennas is used for the measurement update. This is for occlusion cases. The observation function is defined as:

$$h(\boldsymbol{x}) = tan^{-1}\left(\frac{Y_t}{X_t}\right) + v, \tag{15}$$

  where (X_t, Y_t) is the position part of the state vector and v is the observation noise, which follows the uncertainty model of ESPAR antenna readings derived later in Sec. 5.1. We linearize this equation at the predicted position $(\hat{X}_t, \hat{Y}_t)$ to form the following observation equation:

$$h'(\boldsymbol{x}) = H\boldsymbol{x} + v, \tag{16}$$

$$H = \left[-\frac{\hat{Y}_t}{\hat{X}_t^2 + \hat{Y}_t^2} \;\; \frac{\hat{X}_t}{\hat{X}_t^2 + \hat{Y}_t^2} \; 0 \; 0 \right]. \tag{17}$$

- Otherwise, no measurement update is performed. That is, only the prediction step is performed and the uncertainty of the filter increases.

4.4 *Entropy-based filters management*

The management of filters, that is, generation and deletion of filters, is done by an entropy-based method [30]. They use the quadratic Rényi entropy

$$H_2(\boldsymbol{x}) = -\log \int p(\boldsymbol{x})^2 d\boldsymbol{x}. \tag{18}$$

For Gaussian distributions, we have an analytical solution given by

$$H_2(\boldsymbol{x}) = \frac{n}{2}\log 4\pi + \frac{1}{2}\log|\Sigma_t| \tag{19}$$

where Σ_t the covariance matrix and n is the dimension. In the case of the 2D tracking, Σ_t is the positional part of covariance matrix (i.e., the upper left 2×2 submatrix) of a filter and $n = 2$. This entropy is used to estimate the reliability of a filter.

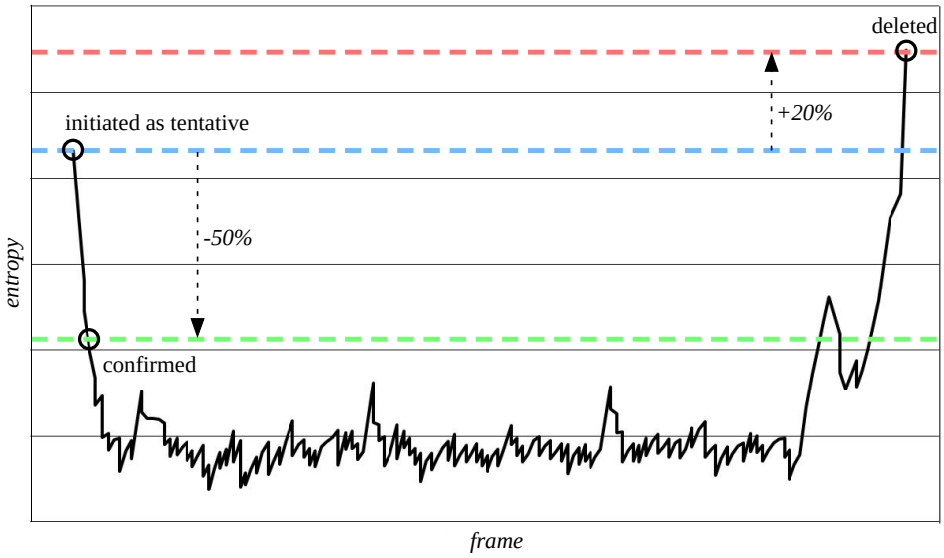


Figure 7. An example history of entropy values and filter management.

Following [30], filters are considered either *tentative* or *confirmed*. When the entropy of a tentative filter drops for 50% from the initial value, it is considered to be confirmed. When the entropy of a confirmed filter becomes 20% higher than the initial value, the filter is deleted. Fig. 7 shows an example history of entropy values of a filter from its generation to deletion.

5. Target Person Identification and Detection of Lost Situation

Tracking using only the range sensor may sometimes fail due to similar location and appearance between the target and others, especially in crowded cases. We therefore use ESPAR antenna data to verify if the filter for the target (i.e., *target filter*) is actually tracking the target person. When this is not the case, we choose another filter as the target filter or judges that the target is lost. For these operations, we associate the probability of tracking the target person with each Kalman filter.

5.1 *The probability of being target*

A Bayesian inference is adopted for assessing the probability that each filter is actually tracking the target person who is with the transmitter. The probability P_t^i that the ith filter is tracking the target at time t is given by

$$P_t^i = \frac{P(Z_t|A_i)P_{t-1}^i}{P(Z_t|A_i)P_{t-1}^i + P(Z_t|\overline{A_i})(1-P_{t-1}^i)}, \tag{20}$$

$$P(Z_t|A_i) = \prod_{j=1}^{2} P(Z_t^j|A_i), \tag{21}$$

where Z_t^j is the direction of the transmitter (i.e., the target person) given by the jth ESPAR antenna at time t and A_i is the event that the ith filter is tracking the target.

The likelihood function $P(Z_t^j|A_i)$ is modeled as a Gaussian, calculated in advance by being fitted to an actual distribution of ESPAR antenna readings. The data were taken in the following setting; the transmitter was put in the right front direction (i.e., the angle is equal to zero degree)

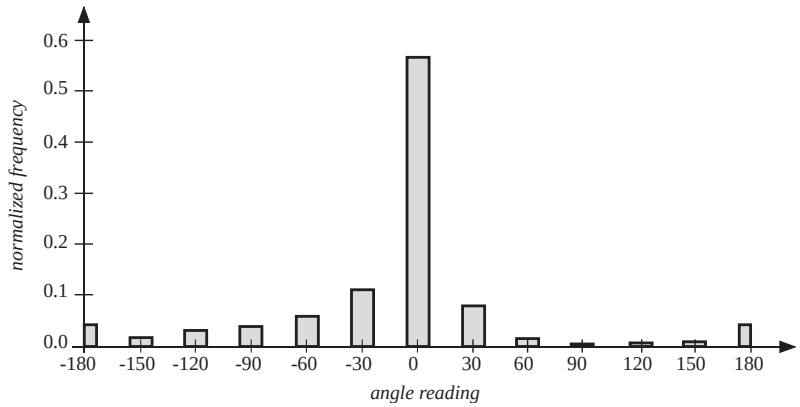


Figure 8. Histogram of antenna readings.

and the distance to the transmitter from the antenna was changed from one to ten meters with one meter interval (i.e., ten positions); data were collected for two minutes at each position. Fig. 8 shows the normalized histogram of antenna readings. There are actually two peaks in the histogram; one is at zero degree while the other at the 180 (-180) degrees. Since the latter one is small and considered due to the reflection from the robot body, we focus to the values ranging between $-120\,deg.$ and $120\,deg.$ and fit a Gaussian to these values. We get $36.6\,deg$ as the standard deviation and use zero as the mean value.

Concerning $P(Z_t^j|\overline{A_i})$, which is the probabilistic distribution of antenna readings when no target exist nearby, we model it with a uniform distribution, that is, given by $P(Z_t^j|\overline{A_i}) = 1/12$. Note that the antenna returns one of twelve directions.

The updated probabilities and the other probability P_{lost} that all filters are not tracking the target are normalized to sum up to one using the following expressions:

$$SP = \sum_i P_t^i + P_{lost}, \tag{22}$$

$$P_t^i = P_t^i/SP, \tag{23}$$

$$P_{lost} = P_{lost}/SP. \tag{24}$$

When a filter j is deleted, P_{lost} increases as follows:

$$P_{lost} = P_{lost} + P_t^j. \tag{25}$$

This probability decreases by the probability normalization (see eq. (24)) when a new filter is added or when the probabilities of other filters increase .

5.2 *Switching of the target filter and detection of lost situation*

The current target filter i_T is judged to have lost the target when one of the following conditions holds:

- The probability that that filter is tracking the target is less than a threshold (currently, 0.1), or
- The entropy of that filter is above the threshold for deletion (see Sec. 4.4).

(a) Robot is stopping. (b) Robot is moving towards the target. (c) Robot is turning.

Figure 9. ESPAR-based mapping results for three robot motion patterns.

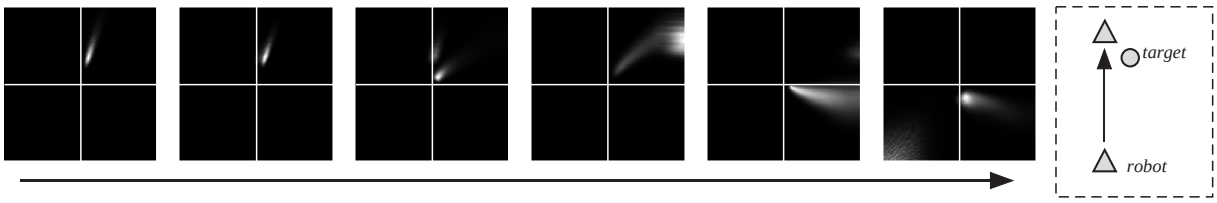


Figure 10. Change of the ESPAR-based map for a straight robot motion. The chronological order of the maps are from left to right.

When this happens, P_t^i's are normalized with excluding filter i_T:

$$SP' = \sum_{i \neq i_T} P_t^i + P_{lost}, \tag{26}$$

$$\hat{P}_t^i = P_t^i / SP'. \tag{27}$$

If the maximum probability is larger than a threshold (currently, 0.7), the corresponding filter is chosen as a new target filter. Otherwise the system transits to the *lost* state (see Fig. 3).

6. Localizing the Target Person using ESPAR Antenna

Completely lost situations usually arise when the target person is fully occluded and/or out of view of the LIDARs. The ESPAR antenna is quite suitable for finding the target in such situations. To localize the target person, we construct a probabilistic target existence grid map.

The probability update for each cell in the map is done by a Bayesian inference, which is similar to the one used in updating the probability of a Kalman filter being tracking the target (see eq. (20)). We use the following equation:

$$P_{i,j}^t = \frac{P(Z_t|E_{i,j})P_{i,j}^{t-1}}{P(Z_t|E_{i,j})p_{i,j}^{t-1} + P(Z_t|\overline{E}_{i,j})(1 - P_{i,j}^{t-1})}, \tag{28}$$

where $P_{i,j}^t$ and is the (unnormalized) probability that the target exists at cell (i, j), $E_{i,j}$ is the event that the target is at (i, j), and Z_t is the current ESPAR antenna reading. After updating the values of all cells using this equation, they are normalized to sum up to one. This update is performed once for each ESPAR antenna.

We tested this approach for a static target in several cases. Fig. 9 shows the mapping results for three robot motion patterns. The center of the map is the robot position and the upward direction of the map is the aligned with the latest heading of the robot. In cases (a) and (b), since the direction of the target does not change, the correct direction is estimated. In case (c), however, the target existence region spreads widely due to a large robot heading change. Fig. 10 shows the result when the robot moves slowly near the static target. Since the target direction changes gradually, and since this setting actually realizes triangulation with a large baseline, the

(a) Scene 1. (b) Scene 2.

Figure 11. The probabilistic target existence maps and corresponding scenes. The upward direction in the map is aligned to the heading of the robot. Red boxes in the maps indicate the highest probability cells.

target is relatively well localized in the map.

This ESPAR-based localization is used when the target is lost, as mentioned above. In such a situation, the target is most probably moving (in an unknown direction) thereby making it difficult to choose a robot motion which sufficiently suppresses the target direction changes. The robot, therefore, stops and accumulates antenna data for target localization. The number of frames for accumulation is currently set to ten. Note that using two ESPAR antennas enables the robot to do triangulation, although the baseline is small. The robot can therefore at least roughly localize the target even while it is stopping.

Fig. 11 shows examples of probabilistic target existence grid map and the corresponding scenes in a person following scenario. The red boxes in the figures show the highest probability cells, used as the tentative position of the target. The peakiness of the distribution may change from case to case; even in a large uncertainty case, however, the direction to the target is correct enough for heading the robot towards the target person.

7. Experimental Results

7.1 *Comparison of data association methods in tracking*

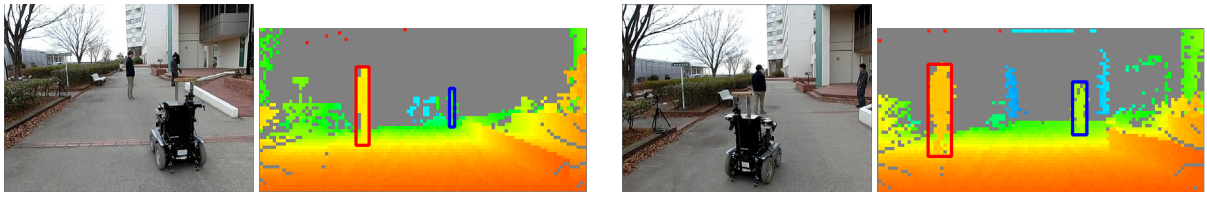
The proposed state transition model has two main states: *following* state and *searching* state (see Fig. 3). In the *searching* state, since the LIDAR data are not available, only ESPAR data are used, and ESPAR antennas are definitely effective and indispensable. In the *following* state, three kinds of information, LIDAR position data, LIDAR appearance data (i.e., reflectance value histogram), and ESPAR data, are available. To assess which kinds of data are effective in the data association step of the tracking, we performed a comparison of four data association strategies, each of which uses a different set of data as follows:

- LIDAR position data and LIDAR appearance data. This is the one we use in the proposed system. Both data are combined using the joint likelihood defined as eq. (6).
- LIDAR position data only. This uses only the proximity part (the first term) in eq. (6).
- ESPAR data only. This relies only on the probability of a filter tracking the target, defined as eq. (20).
- LIDAR data (position and appearance) and ESPAR data. In this case, we extend the joint likelihood by further multiplying the probability of being the target to the original one.

The test data was collected in a quite difficult situation as follows. Four persons including the target person walked (and sometimes stopped) in front of the robot, within the robot's FOV. For the tracking of the target using LIDAR data, the tracking filter is correctly initialized using the ESPAR data. We collected 500 frames and the target was visible from the robot in 275 frames. For each set of data, we counted the number of the reported and the correct detection of the target, as summarized in Table 1.

Table 1. Comparison of data association methods.

	LIDAR position/appearance	LIDAR position	ESPAR	LIDAR+ESPAR
# of detection reported	173	400	435	312
# of correct detection	156	237	213	179
recognition rate (%)	90.2	59.3	45.0	57.4



(a) Correct target person (left one) is being tracked.

(b) Another correct tracking case.

Figure 12. Normal tracking. Red boxes indicate the target person.

(a) A person is passing between the robot and the target.

(b) The target is correctly identified after a short occlusion.

(c) Another person is passing the left side of the robot.

Figure 13. Short occlusion case.

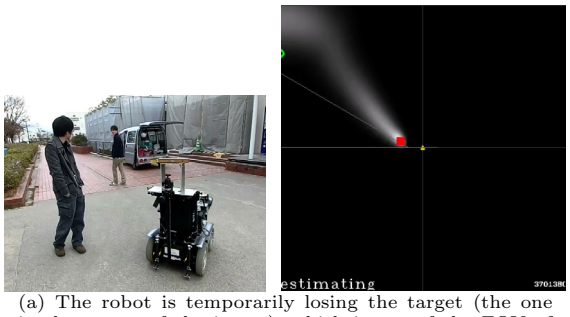
The results show that the LIDAR appearance feature is quite effective in increasing the precision, which is the most important factor for realizing a robust following behavior. On the other hand, the use of ESPAR data for data association is not effective, probably because of its large uncertainty in direction estimation. Based on these results, we have decided to rely only on LIDAR data (position and appearance) in the data association step as long as they are available, as described in Sec. 4.2.

7.2 *Specific person following experiments*

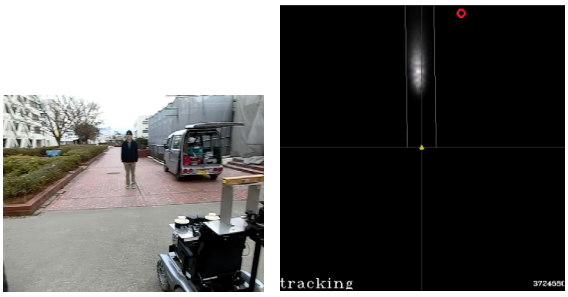
We performed specific person following experiments in our campus. This section introduces an experimental run and explains the behaviors of the system in detail.

The distance traveled in this run is about 340 m and the number of total frames is 4020. The system was at *searching* state, where the ESPAR antenna is used for searching for the target, ten times, for 271 frames in total. The robot followed a wrong person once for 30 frames and came back to the correct person after that. Person following using only LIDARs or cameras could probably had failed in one of these cases, while our system was able to continue the following behavior.

We here explain several typical behaviors of the system. Fig. 12 shows a normal tracking situation. The person on the left is the correct target person and red boxes indicate the target

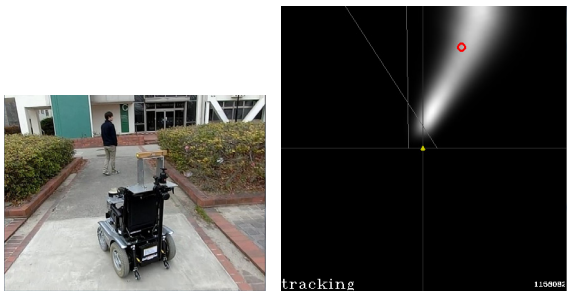


(a) The robot is temporarily losing the target (the one in the center of the image), which is out of the FOV of LIDAR.

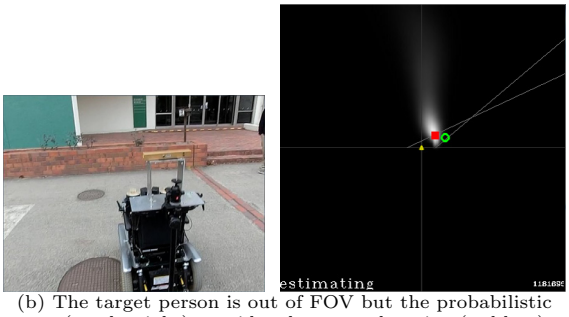


(b) The robot turned to the target using the probabilistic map.

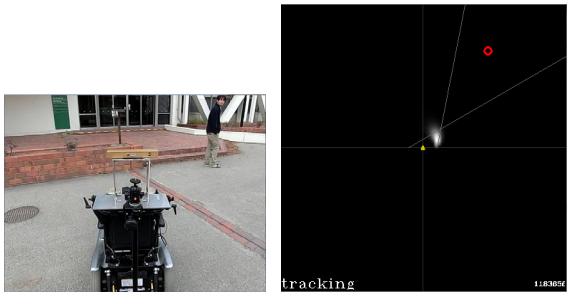
Figure 14. Recovery from an out-of-sight situation.



(a) The robot is following the target in the FOV of LIDAR.



(b) The target person is out of FOV but the probabilistic map (on the right) provides the target location (red box). The target filter is still running (green circle).



(c) The robot detects the target using LIDAR and re-initiates the Kalman filter tracking.

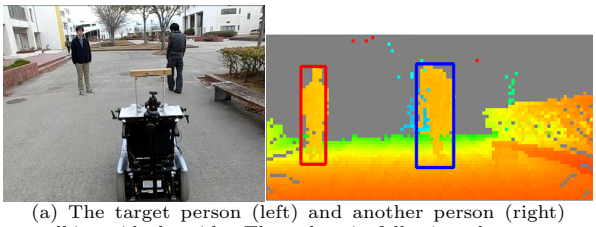
Figure 15. Recovery from another out-of-sight situation.

person identified by the system. Fig. 13 shows a tracking scene with a short occlusion. All persons are detected and the target person is correctly identified. Such short occlusion cases can be handled by many of previous works. Figs. 14 and 15 show the cases where the robot loses the target out of the FOV of LIDAR and the system transits to *searching* state. Once the target location is estimated using ESPAR antenna data, the robot moves towards there to find the target again and the system transits to *following* state. Fig. 16 shows the case where the robot tracks a wrong person for a while but recovers to tracking the correct target using ESPAR antenna data.

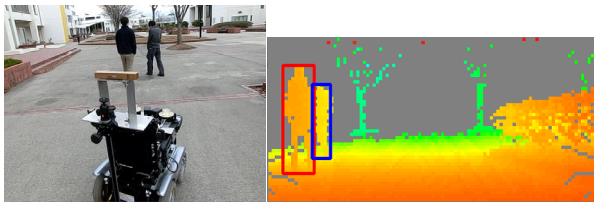
The visual sensors such as cameras and LIDARs can provide rich information for detecting people and for identifying a specific person but inherently weak to out-of-sight and misidentification situations. Adding some exploratory search strategy could happen to find the target again; however, such a search may not be efficient. On the other hand, the ESPAR antenna can provide useful data for quickly directing the robot to the correct direction in such a situation (see Figs. 14, 15 and 16).

We also tested the system at night time. In a test run at night, the distance traveled was about 300 m and the number of total frames is 2244. The robot encountered eleven persons other than the target but never failed to identify the target. Fig. 17 shows example scenes during the following experiment. Since the system does not rely on vision at all, it works at night as well.

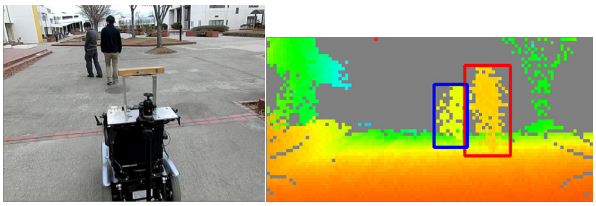
These experimental results show the effectiveness of the proposed combination of 3D LIDAR



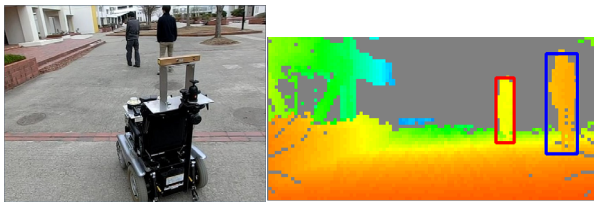
(a) The target person (left) and another person (right) walking side by side. The robot is following the target correctly.



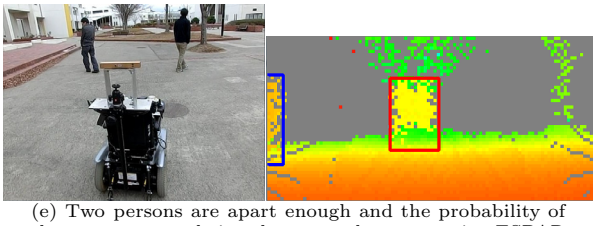
(b) Two persons comes very near to each other but the robot correctly follows the target.



(c) Two persons cross but the target is still correctly identified.



(d) The robot is temporarily tracking another person.



(e) Two persons are apart enough and the probability of the wrong person being the target decreases using ESPAR antenna data; the robot starts tracking the correct target again.

Figure 16. Recovery from a wrong tracking.

(a) (b) (c) (d)

Figure 17. A person following sequence at night.

and the ESPAR antenna.

8. Conclusions and Future Work

This paper has described a method of specific person detection and tracking using 3D LIDAR and ESPAR antenna. With the antenna that can detect the direction of the radio transmitter very fast, the target person can be tracked reliably and can also be localized even when he/she is occluded or out of sight of the LIDAR. A state transition model-based tracking framework has also been developed. Combination of these methods makes the person following behavior of our robot very robust even in difficult situations.

A possible improvement is to additionally use image data, which usually provide richer information at least under moderate illumination conditions. Fusion with image data is expected to further increase the accuracy of detection, tracking, and identification. Long-term experiments in more crowded situations are also desired in the future. For such an experiment, hardware improvements including using more long-life batteries will be necessary. Development of smaller antennas and transmitter is also desired for daily use.

Acknowledgment

The authors would like to thank Prof. Takashi Ohira of TUT for giving advice on the ESPAR antenna. They would also like to thank the members of Active Intelligent Systems Laboratory at TUT for their supports in implementing the system.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work is supported by the Japan Society for the Promotion of Science (JSPS) [grant number 25280093].

References

[1] Dalal N, Briggs B. Histograms of oriented gradients for human detection. In: Proceedings of 2005 ieee conf. on computer vision and patttern recognition. 2005. p. 886–893.
[2] Corvee E, Bak S, Bremond F. People detection and re-identification for multi surveillance cameras. In: Proceedings of int. conf. on computer vision theory and applications. 2012.
[3] Ess A, Leibe B, Schindler K, Gool L. A mobile vision system for robust multi-person tracking. In: Proceedings of the 2008 ieee conf. on computer vision and pattern recognition. 2008.
[4] Ardiyanto I, Miura J. Partial least squares-based human upper body orientation estimation with combined detection and tracking. Image and Vision Computing. 2014;32(11):904–915.
[5] Vapnik V. Statistical learning theory. New York: John Wiley & Sons. 1998.
[6] Freund Y, Schapire R. A deicision-theoretic generalization of on-line learning and an application to boosting. J of Computer and System Sciences. 1997;55(1):119–139.
[7] Satake J, Miura J. Robust stereo-based person detection and tracking for a person following robot. In: Proceedings of icra-2009 workshop on person detection and tracking. 2009.
[8] Tsutsui H, Miura J, Shirai Y. Optical flow-based person tracking using multiple cameras. In: Proceedings of the 2001 int. conf. on multisensor fusion and integration for intelligent systems. 2001. p. 91–96.
[9] Schulz D, Burgard W, Fox D, Cremers A. People tracking with a mobile robot using sample-based joint probabilistic data association filters. Int J of Robotics Research. 2003;22(2):99–116.
[10] Arras K, Mozos O, Burgard W. Using boosted features for the detection of people in 2d range data. In: Proceedings of the 2007 ieee int. conf. on robotics and automation. 2007. p. 3402–3407.
[11] Premebida C, Ludwig O, Nunes U. Exploiting lidar-based features on pedestrian detection in urban scenarios. In: Proceedings of the 12th ieee int. conf. on intelligent transportation systems. 2009. p. 18–23.
[12] Zainudin Z, Kodagoda S, Dissanayake G. Torso detection and tracking using a 2d laser range finder. In: Proceedings of australasian conf. on robotics and automation 2010. 2010.
[13] Spinello L, Arras K, Triebel R, Siegwart R. A layered approach to people detection in 3d range data. In: Proceedings of the 24th aaai conf. on artificial intelligence. 2010. p. 1625–1630.
[14] Navarro-Serment L, Mertz C, Hebert M. Pedestrian detection and tracking using three-dimensional ladar data. Int J of Robotics Research. 2010;29(12):1516–1528.
[15] Kidono K, Miyasaka T, Watanabe A, Naito T, Miura J. Pedestrian recognition using high-definition lidar. In: Proceedings of 2011 ieee intelligent vehicles symp.. 2011. p. 405–410.
[16] Kidono K, Naito T, Miura J. Pedestrian recognition combining high-definition lidar and vision data. In: Proceedings of 15th ieee int. conf. on intelligent transportation systems. 2012. p. 1783–1788.
[17] Satake J, Chiba M, Miura J. A sift-based person identification using a distance-dependent appearance model for a person following robot. In: Proceedings of the 2012 ieee int. conf. on robotics and biomimetics. 2012. p. 962–967.

[18] Belhumer P, Hespanha J, Kriegman D. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Trans on Pattern Analysis and Machine Intelligence. 1997;19(7):711–720.
[19] Sugiura K, Makihara Y, Yagi Y. Gait identification based on multi-view observations using omnidirectional camera. In: Proceedings of 8th asian conf. on computer vision. Vol. 1. 2007. p. 452–461.
[20] Kim M, Chong N, Yu HS. Rfid-enabled target tracking and following with a mobile robot using direction finding antennas. In: Proceedings of 2007 ieee int. conf. on automation science and engineering. 2007. p. 1014–1019.
[21] Germa T, Lerasle F, Ouadah N, Cadenat V. Vision and rfid data fusion for tracking people in crowds by a mobile robot. Computer Vision and Image Understanding. 2010;114(6):641–651.
[22] Kawakami H, Ohira T. Electrically steerable passive array radiator (espar) antennas. IEEE Antennas and Propagation Magazine. 2005;47(2):43–50.
[23] Misu K, Miura J. Specific person detection and tracking by a mobile robot using 3d lidar and espar antenna. In: Proceedings of the 13th int. conf. on intelligent autonomous systems. 2014. 14 pages.
[24] Ando N, Suehiro T, Kotoku T. A software platform for component based rt system development: Openrtm-aist. In: Proceedings of the 1st int. conf. on simulation, modeling, and programming for autonomous robots (simpar '08). 2008. p. 87–98.
[25] Ardiyanto I, Miura J. Real-time navigation using randomized kinodynamic planning with arrival time field. Robotics and Autonomous Systems. 2012;60(12):1579–1591.
[26] Košeckà J, Bajcsy R. Discrete event systems for autonomous mobile agents. Robotics and Autonomous Systems. 1994;12(3/4):187–198.
[27] Toussaint G. Solving geometric problems with the rotating calipers. In: Ieee mediterranean electrotechnical conf.. 1983. p. 1–8.
[28] Schapire R, Singer Y. Improved boosting algorithms using confidence-rated projections. Machine Learning. 1999;37(3):297–336.
[29] Grabner H, Bischof H. On-line boosting and vision. In: Proceedings of ieee conf. on computer vision and pattern recognition. Vol. 1. 2006. p. 260–267.
[30] Jurić-Kavelj S, Marković I, Petrović I. People tracking with heterogeneous sensors using jpdaf with entropy based track management. In: Proceedings of the 5th european conf. on mobile robots. 2011. p. 31–36.