Hierarchical Sample-Based Joint Probabilistic Data Association Filter for Following Human Legs Using a Mobile Robot in a Cluttered Environment

Yoonchang Sung and Woojin Chung

***Abstract*—Human-following in a cluttered environment is one of the challenging issues for mobile service robot applications. Since a laser range finder (LRF) is commonly installed for autonomous navigation, it is advantageous to adopt an LRF for detection and tracking humans. In this paper, we aim at the reliable human tracking performances in a dynamic cluttered environment. The key idea is to develop a hierarchical sample-based joint probabilistic data association filter (HSJPDAF) by focusing on the leg positions as well as human positions. The proposed HSJPDAF consists of two levels in order to consider the interdependence between targets. Possible locations of multiple human targets can be simultaneously estimated on the basis of Bayesian filtering. Comparison with the general technique was carried out to verify the performance of HSJPDAF in both artificial indoor and real-world environments. Owing to the hierarchical framework, the proposed method shows the improved robustness by reducing the target loss rate significantly in a dynamic cluttered environment.**

***Index Terms*—Data association, human-following, mobile robot, service robot, target tracking.**

I. Introduction

If robots could reliably and safely follow humans, they could support a range of services such as serving as a porter. Tracking consists of estimation and data association [1]. Estimation is the process of estimating unknown parameters through the use of uncertain observations, whereas data association corresponds to an assignment of tracks and measurements. A track indicates the state trajectories of the targets estimated from a set of measurements. The main purpose of target-tracking is to determine which measurement is assigned to which target when multiple targets are being tracked using multiple observed measurements.

There are two representative tracking methods: multiple hypothesis tracking (MHT), proposed by Reid [2], and a joint probabilistic data association filter (JPDAF), proposed by Bar-Shalom *et al.* [3]. MHT generates an alternative data association hypothesis, which is a group of compatible tracks, and provides an optimal solution to the problem of data association [4]. However, MHT has a disadvantage in that its optimality is inevitably reduced through pruning and merging stages owing to its difficult realization and large computational cost. JPDAF computes an assignment probability for each suitable measurement that corresponds to the tracks at the current time step [5]. Many have employed JPDAF because of its simple real-time implementation [6]–[9]. In addition, to estimate nonlinear and non-Gaussian model parameters of human walking motion, the selection of sampling approach can achieve robust performance, as implemented in [10] and [11]. However, in MHT, the filters are copied, which disturbs the effective application of a sample-based approach [6]. Thus, our proposed method is based on JPDAF.

Various sensors fixed in a specific area or mounted onto a mobile robot can support target tracking. In [12] and [13], both methods employ a vision sensor to obtain measurement information of a human. Although they successfully developed a vision-based people tracking system, a vision sensor is sensitive to changes in illumination. Another frequently used sensor for tracking is a laser range finder (LRF). In [14], a stable feature extraction method using an LRF is proposed. In [15], two layers of multiple LRFs are attached to a mobile robot to detect the human chest and legs simultaneously. While only 2-D information can be obtained through the use of an LRF, accurate distance information can be received within a short time period and autonomous navigation can also be conducted. As in [14] and [15], our research therefore employs an LRF. We assumed that an LRF is an appropriate sensor for gathering leg data for human detection.

The objective of a human-following function is to detect the target person of interest and to keep track of the chosen target without confusing the target with other people. The work in [10] defines the concept of following and introduces overall tracking strategies. Similar to [16] and [17], previous research uses an LRF to extract the legs for human recognition. Obstacles for a mobile robot are normally located close to the floor. The leg is the only location where there is no interference with other parts of the human body within a 2-D coordinate frame. For these reasons, it is advantageous to fix an LRF at the height of legs. In [6], [18], and [19], a human target is assigned based on the measurements of the two legs with the shortest distance. This target can cause a failure of human-following because the legs of the target may incorrectly match those of another person. Because a motion model for this target has to consider two different legs simultaneously, it cannot be implemented when only one leg is detected such as when one leg is occluded by the other.

Manuscript received February 8, 2015; revised July 27, 2015; accepted October 11, 2015. Date of publication December 1, 2015; date of current version May 13, 2016. This work was supported by the NRF grant funded by the MSIP(NRF-2014R1A2A1A10049634). This work was also supported by the R&D Convergence Center Support Program (714002-7, Education and Research on agricultural Automation), MAFRA. This paper was recommended by Associate Editor R. A. Hess.

The authors are with the School of Mechanical Engineering, Korea University, Seoul 136-701, Korea (e-mail: yooncs8@vt.edu; smartrobot@korea.ac.kr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/THMS.2015.2501282

In this paper, we address human tracking in a dynamic cluttered environment. A novel tracking method that considers the interdependence between the targets is proposed. Instead of assuming that the targets are exclusively independent of each other [1], our idea is that tracking a grouped target consists of a pair of leg targets. Herein, the target and grouped target refer to human leg and human. Tracking a grouped target while tracking related leg targets simultaneously prevents a possible loss of a grouped target using supplementary leg targets. Although Vadakkepat *et al.* [20] presented an effective leg detection method using an LRF and showed the performance when covering a companion walking and a short period of occlusion, the limits of performance degradation when working in a closely spaced environment cannot be avoided.

To address track confusion or failure during human-following within a densely populated environment, we propose a hierarchical sample-based joint probabilistic data association filter (HSJPDAF) derived from a sample-based JPDAF (SJPDAF) [6]. At a low level, the estimation proceeds using leg information acquired from a sensor model. At a high level, grouping the targets through a data association that regards the interdependence of the targets at a low level is achieved. The key idea of HSJPDAF is to focus on the leg positions as well as human positions. This idea implies that the estimation and data association are separated into a hierarchical structure. The following advantages can be obtained. First, the targets can be grouped into new targets using a probabilistic approach. The tracking reliability can be increased owing to the use of previous grouping information. Second, tracking a grouped target that consists of at most two targets is more robust than tracking a single target in a highly populated environment. Third, since human motion of two legs is complicated, the proposed hierarchical structure yields a simple motion model that considers the motion of only one leg.

The rest of this paper is organized as follows. Section II introduces previous research on SJPDAF. Next, our proposed HSJPDAF is described in Section III. Section IV details the decision problem and state prediction, which are required for tracking. Section V presents the experimental results. Finally, Section VI provides conclusions.

II. Joint Probabilistic Data Association Filter-Related Issues

A. Joint Probabilistic Data Association Filter

JPDAF can be used to solve the data association problem and is an appropriate technique for tracking multiple targets. Considering the tracking of N_T targets, $\mathbf{X}(k) = \{\mathbf{x}_1(k), ..., \mathbf{x}_{N_T}(k)\}$ denotes the state of targets at time k. Each $\mathbf{x}_i(k)$ corresponds to a probability variable regarding the state space of target i. $\mathbf{Z}(k) = \{\mathbf{z}_1(k), ..., \mathbf{z}_{m(k)}(k)\}$ denotes observed measurements at time k. Each $\mathbf{z}_j(k)$ represents measurement j. $\mathbf{z}_0(k)$ is used for a case in which no measurements are observed. $\mathbf{Z}^k = \{\mathbf{Z}(1), ..., \mathbf{Z}(k)\}$ indicates a set of all observed sequential measurements up to and including time k. To detect human legs as measurements using laser sensor data, we adopted the extraction method [20], where we obtained leg data by means of a support vector data description.

To express the assignment of N_T targets and $m(k)$ measurements at time k, a joint association event, $\theta(k)$, is introduced. This event is a set of pairs, $(j, i) \in \{0, ..., m(k)\} \times \{1, ..., N_T\}$. Here, $\Theta_{ji}(k)$ represents all possible joint association events of assigning target i and measurement j. The key point of JPDAF is to obtain the following assignment probabilities to determine the data association:

$$\beta_{ji} = \sum_{\theta(k) \in \Theta_{ji}(k)} P\{\theta(k)|\mathbf{Z}^k\}. \tag{1}$$

$P\{\theta(k)|\mathbf{Z}^k\}$ corresponds to the probability of a joint association event at time k given all measurements up to and including time k. According to [6], β_{ji} can be defined through the equation

$$\beta_{ji} = \sum_{\theta(k) \in \Theta_{ji}(k)} [\eta \gamma^{(m(k)-|\theta(k)|)} \prod_{(j,i)\in\theta(k)} \int p(\mathbf{z}_j(k)|\mathbf{x}_i(k)) \times p(\mathbf{x}_i(k)|\mathbf{Z}^{k-1}) d\mathbf{x}_i(k)] \tag{2}$$

where γ denotes the probability of $\mathbf{Z}(k)$ being a false alarm given $\theta(k)$. $p(\mathbf{z}_j(k)|\mathbf{x}_i(k))$ and $p(\mathbf{x}_i(k)|\mathbf{Z}^{k-1})$ are related to the gating and prediction, respectively. Note that η denotes a normalization constant and the different constants used in the remainder of the paper will be subsumed into η.

An association matrix that consists of a series of rows for measurement j and columns for target i can be produced with respect to β_{ji}. The k-best assignment of this matrix can be computed within a polynomial time using the Murty algorithm [21], [23]. Once the best solution is gained, the next-best solution can be found by removing other possible best solutions iteratively.

B. Sample-Based Joint Probabilistic Data Association Filter

SJPDAF [6] eschews a conventional Kalman filter during the estimation process and employs a particle filter. SJPDAF is superior to JPDAF in that multimodal state densities can be described through samples. In addition, the advantage of a particle filter for the nonlinear or non-Gaussian characteristic of the target can be obtained.

The posterior probability of a particle filter is represented through a set of random states of samples, $\{\mathbf{x}_i^l(k), w_i^l(k)\}_{l=1}^N$. Here, $\{\mathbf{x}_i^l(k), l = 1, ..., N\}$ denotes a set of N samples that corresponds to the weights of target i at time k. Prior to considering the weights in the correction step, the assignment probabilities for the sample-based form must be computed in advance to find out which measurement is caused by which target. The measurements $\mathbf{Z}(k)$ can be selectively and proportionally contributed to the weights of the target based on a gating condition and the result of the assignment probabilities. The assignment probabilities of the SJPDAF version proposed by [6] using (2) are in the

Fig. 1. Example of a highly populated and cluttered environment.

following equation:

$$\beta_{ji} = \sum_{\theta(k)\in\Theta_{ji}(k)} [\eta\gamma^{(m(k)-|\theta(k)|)} \prod_{(j,i)\in\theta(k)} \frac{1}{N} \times \sum_{l=1}^{N} p(\mathbf{z}_j(k)|\mathbf{x}_i^l(k))]. \tag{3}$$

The weights of the samples using (3) are

$$w_i^l(k) = \eta \sum_{j=0}^{m(k)} [\beta_{ji} p(\mathbf{z}_j(k)|\mathbf{x}_i^l(k))]. \tag{4}$$

Finally, a new set of N samples for target i at the current time step, $\{\mathbf{x}_i^l(k), w_i^l(k)\}_{l=1}^{N}$, can be obtained through (3) and (4). Acquired set of samples with additional resampling step corresponds to sample importance resampling filter as in [24]–[26] by estimating the posterior density of the state variables.

III. Hierarchical Sample-Based Joint Probabilistic Data Association Filter

A. Hierarchical Structure

Let us consider a case in which the observed measurements are not exclusively independent of each other. Targets that correspond to the legs of one person are dependent when the human legs are observed from an LRF, and the targets have an interrelationship. Particularly in a highly populated and cluttered environment, people have to be sorted precisely among the observed measurements, as shown in Fig. 1. The reason is that both legs of a person might not be observed, or that one leg may be confused with the leg of someone else. For robust tracking, we propose the use of HSJPDAF, which can estimate the state of the target with respect to their legs, and can determine the human positions simultaneously.

As shown in Fig. 2, HSJPDAF consists of two levels. At a low level, the states of the target legs are estimated using the observed measurements. At a high level, a data association is carried out to determine the states of grouped targets using an association matrix of information obtained from the targets and the observed measurements. When a human target (a grouped target) is available through the data association step, a robot is activated to move toward that human.

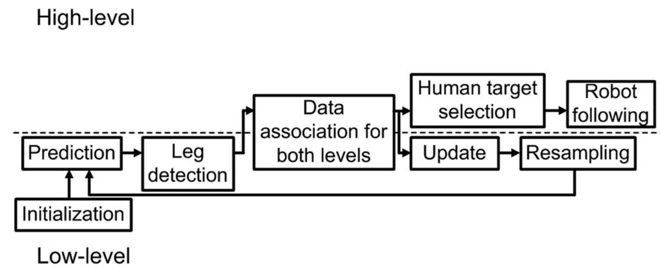


Fig. 2. Diagram for HSJPDAF scheme process.

Consider ${}^hN_T$ grouped targets and ${}^hm(k)$ grouped measurements when N_T targets and $m(k)$ measurements exist, as described in Section II. Note that superscript h is used for a high-level representation, and should not be confused with a low-level notation. The grouped targets and grouped measurements correspond to the targets and observed measurements obtained at a low level, respectively. A high-level joint association event $\varphi(k)$ is used for the assignment between grouped targets and grouped measurements, and is a set of pairs $(q,p) \in \{0,...,{}^h m(k)\} \times \{1,...,{}^h N_T\}$, where p and q correspond to the index for the set of grouped targets and grouped measurements, respectively. In addition, $\Phi_{qp}(k)$ stands for all possible joint association events of assigning grouped target p and grouped measurement q. HSJPDAF-based tracking requires the following two constraints.

1) One measurement at most can be made from a single track. As mentioned in Section II, a measurement obtained from a sensor model corresponds to a human leg, and a track implies the trajectory of leg targets induced from a measurement at each time step. This constraint is relevant to the conventional assumption described in [1].
2) Two targets and two measurements at most can be contained in a grouped target and grouped measurement, respectively, because human position is determined based on the estimation results of the legs. Note that each grouped target cannot share a target with the others because two people cannot both have the same leg.

B. Assignment Probabilities of Hierarchical Sample-Based Joint Probabilistic Data Association Filter

The next step is to obtain assignment probabilities that are contained in a high-level association matrix. With these values, human position can be determined by solving the data association problem. The assignment probabilities of a high level are defined as follows:

$$\alpha_{qp} = \sum_{\varphi(k)\in\Phi_{qp}(k)} P\{\varphi(k)|\mathbf{Z}^k\}. \tag{5}$$

$P\{\varphi(k)|\mathbf{Z}^k\}$ of an individual joint association event can be computed using both the law of total probability theorem and Bayes' theorem as follows:

$$P\{\varphi(k)|\mathbf{Z}^k\} = \sum_{\theta(k)\in\Theta_{ji}(k)} [P\{\varphi(k)|\mathbf{Z}^k, \theta(k)\} \times P\{\theta(k)|\mathbf{Z}^k\}]. \tag{6}$$

Equation (6) can be obtained using the law of total probability. The second term of (6), $P\{\theta(k)|\mathbf{Z}^k\}$, denotes a low-level individual association event and was appeared in (1). The first term of (6) is derived in the following equations:

$$P\{\varphi(k)|\mathbf{Z}^k,\theta(k)\} = P\{\varphi(k)|\mathbf{z}(k),\mathbf{Z}^{k-1},\theta(k)\} \tag{7}$$

$$= \frac{p(\mathbf{z}(k)|\varphi(k),\mathbf{Z}^{k-1},\theta(k))P\{\varphi(k)|\mathbf{Z}^{k-1},\theta(k)\}}{p(\mathbf{z}(k)|\mathbf{Z}^{k-1},\theta(k))} \tag{8}$$

$$= \frac{p(\mathbf{z}(k)|\varphi(k),\theta(k))P\{\varphi(k)|\mathbf{Z}^{k-1},\theta(k)\}}{p(\mathbf{z}(k)|\mathbf{Z}^{k-1},\theta(k))} \tag{9}$$

$$= \frac{p(\mathbf{z}(k)|\varphi(k),\theta(k))P\{\varphi(k)|\theta(k)\}}{p(\mathbf{z}(k)|\mathbf{Z}^{k-1},\theta(k))} \tag{10}$$

$$= \eta p(\mathbf{z}(k)|\varphi(k),\theta(k)). \tag{11}$$

$P\{\varphi(k)|\mathbf{Z}^k,\theta(k)\}$ can be elongated as in (7) so that (8) can be obtained using the Bayes' theorem. Both (9) and (10) are results of the Markovian assumption, supposing that both terms of the numerator of (8) are independent of previous measurements, $\mathbf{Z}^{k-1}$. The second term of the numerator of (10), $P\{\varphi(k)|\theta(k)\}$, represents the probability of a high-level joint association event $\varphi(k)$ given a low-level joint association event $\theta(k)$ at time k. $P\{\varphi(k)|\theta(k)\}$ can be considered as a constant by assuming that the correspondence between a low-level and high-level event occurs with equal probability. All the constants of $P\{\varphi(k)|\theta(k)\}$ and $p(\mathbf{z}(k)|\mathbf{Z}^{k-1},\theta(k))$ of (10) are contained in η and (11) can be consequently derived. $p(\mathbf{z}(k)|\varphi(k),\theta(k))$ of (11) indicates the likelihood function of $\varphi(k)$ and $\theta(k)$ given $\mathbf{z}(k)$. Therefore, $p(\mathbf{z}(k)|\varphi(k),\theta(k))$ is used to determine how the correspondence between low level and high level affects the assignment probabilities of the latter given the observed measurements.

By substituting (6) into (5), the assignment probabilities of a high level, α_{qp}, can be restated for clarity as follows:

$$\alpha_{qp} = \sum_{\varphi(k)\in\Phi_{qp}(k)} \left[\sum_{\theta(k)\in\Theta_{ji}(k)} \left[P\{\varphi(k)|\mathbf{Z}^k,\theta(k)\} \times P\{\theta(k)|\mathbf{Z}^k\} \right] \right]. \tag{12}$$

Given that the first term of (12) is defined in (11) and the second term of (12) is expressed inside the summation of (3), the following equation computes the resultant, α_{qp}:

$$\alpha_{qp} = \sum_{\varphi(k)\in\Phi_{qp}(k)} \left[\sum_{\theta(k)\in\Theta_{ji}(k)} \left[\eta\gamma^{(m(k)-|\theta(k)|)} \times p(\mathbf{z}(k)|\varphi(k),\theta(k)) \prod_{(j,i)\in\theta(k)} \frac{1}{N} \sum_{l=1}^{N} p(\mathbf{z}_j(k)|\mathbf{x}_i^l(k)) \right] \right]. \tag{13}$$

Finally, grouped targets can be tracked through (13). Because a high level proceeds with data association only, weights $w_i^l(k)$ for the target samples are maintained as in (4) for a low level.

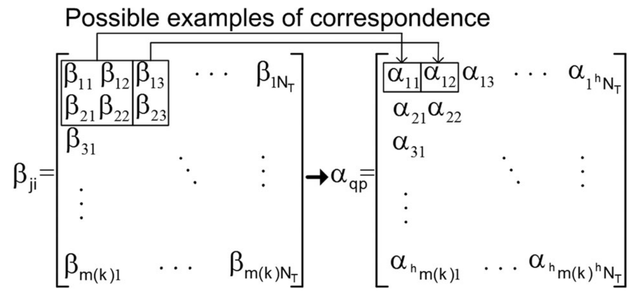


Fig. 3. Low-level and high-level assignment matrix.

C. Correspondence From Low Level to High Level

It is important to define how the correspondence between a low and high level works in an assignment matrix and for the final desirable value α_{qp}. The definition of the correspondence between two levels implies $p(\mathbf{z}(k)|\varphi(k),\theta(k))$ of (13) itself.

Each element of the low-level assignment matrix on the left side of Fig. 3 corresponds to a value from (3). Each element of the high-level assignment matrix on the right side of Fig. 3 is obtained with respect to the value from (13). The left- and right-side matrixes correspond to joint association events $\Theta_{ji}(k)$ and $\Phi_{qp}(k)$, respectively. To determine the link between the two, see [27]. The row and column elements for a joint association event are under the following conditions:

1) *Set q of grouped measurements:* Set q considers all possible groups that consist of n measurements within $m(k)$ measurements at time k. Here, n is 2 because the estimated object is a human leg and is a part of a grouped measurement. Grouped measurements are subsets of the power set of measurements having only two elements of set j. If the number of observed measurements is one, the grouped measurement will be composed of one element. Grouped measurements are candidates for matching with a grouped target. Setting a threshold of the largest distance that two legs can have within the sampling time, the measurements can be divided into several groups with the threshold to remove any unnecessary matching possibilities. This lowers the entire computational cost. Set $c = \{1, ..., c_T\}$ denotes a set of indices of divided measurements using a threshold. This is a similar principle to a cluster, which is explained in [2]. $m_c(k)$ represents the number of measurements given an index in the set c of divided measurements. $m(k)$ is the sum of all c_T divided measurements as follows:

$$\sum_{c=1}^{c_T} m_c(k) = m(k). \tag{14}$$

The total number of grouped measurements regarding c_T divided measurements can be computed according to

$$n(q) = \sum_{c=1}^{c_T} {}_{m_c(k)}C_2 =^h m(k)$$
$$\text{if } m_c(k) = 1, \quad \text{then } {}_{m_c(k)}C_2 \equiv 1 \tag{15}$$

where ${}_{m_c(k)}C_2$ is a combination that gives the number of ways of selecting two unordered results given $m_c(k)$

Algorithm 1 Obtain α_{ef}

$Suppose\ that\ (a,b),\ (c,d),\ e,\ f\ are\ temporal$
$indices\ of\ j,\ i,\ q,\ p,\ respectively.$
if $j \notin q$ $\quad (i.e.,\ \{a,b\} \notin e)$ **then**
  $\alpha_{ef} \leftarrow 0$
else if $j \in q$ $\quad (i.e.,\ \{a,b\} \in e)$ **then**
  if $n(p) = 2$ $\quad (i.e.,\ \{c,d\} \in f)$ **then**
    $\alpha_{ef} \leftarrow 0.25(\beta_{ac} + \beta_{ad} + \beta_{bc} + \beta_{bd})$
  else if $n(p) = 1$ $\quad (i.e.,\ \{c\} \in f)$ **then**
    $\alpha_{ef} \leftarrow 0.5(\beta_{ac} + \beta_{bc})$
  end if
end if
return (α_{ef})

possibilities. We set the value of the threshold to 48.2 cm, which was obtained experimentally.

2) *Set p of grouped targets:* Set p stands for grouped targets that consist of at most n targets within N_T targets at time k. Here, n is 2, similar to the case of set q. A grouped target has its own track in a similar manner as a target. Thus, set p maintained up to time k is regarded as the same person until that time. Once the targets of set i are initiated into the grouped targets, two targets at most can be contained in one grouped target. The fact that different people cannot have the same leg makes the subsets of each grouped target unable to share the same element with those of the other grouped targets, as in the following equation:

$$\text{if } \mathbf{A} \text{ and } \mathbf{B} \text{ are elements of } \mathbf{p} \text{ and } \mathbf{A} \neq \mathbf{B},$$
$$\text{then } \mathbf{A} \cap \mathbf{B} = \emptyset. \tag{16}$$

$$n(p) = \begin{cases} \dfrac{N_T}{2} =^h N_T, & \text{if } N_T \text{ is an even number} \\ \dfrac{(N_T+1)}{2} =^h N_T, & \text{if } N_T \text{ is an odd number} \\ 0, & \text{otherwise.} \end{cases} \tag{17}$$

Sets p and q are not general sets, but a family of subsets. This is because both include at most two low-level results. Examples of a correspondence are shown in Fig. 3.

Algorithm 1 describes how to obtain α_{qp} with respect to the correspondence. Here, $n(p) = 2$ indicates the state in which two targets are included, and $n(p) = 1$ indicates when one target is included. According to the second constraint of Section III-A, a two target-state and one target-state can have four and two combinations of the assignment probabilities of a low level, respectively. This is because a grouped measurement can include two measurements at most. For this reason, 0.25 and 0.5 are set for each case to make the assignment probabilities of a high level summed up to 1.

As a result, a high-level assignment matrix is obtained using sets p and q. In addition, k-best assignments can be found through the use of Murty algorithm [23]. Data association of the grouped target, which is a person in this paper, is completed at time k.

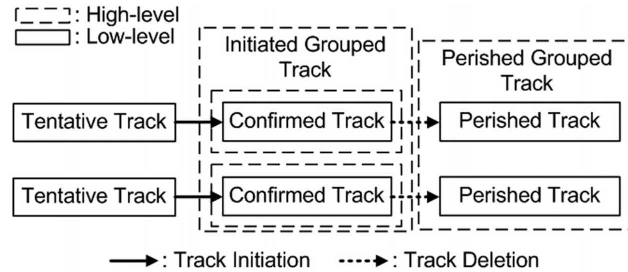


Fig. 4. Hierarchical structure of track decision logic.

IV. Human Tracking

A. Decision Problem

The number of tracks for both a target and a grouped target is changeable as the target either appears or disappears within the sensing range. The number of tracks therefore has to be controlled through a decision-based logic. To classify the states of the track, the authors of [23] and [25] define track initiation, track confirmation, and track deletion.

In this paper, the states of a low-level track and a high-level grouped track are determined, as shown in Fig. 4. In the case of a low level in the solid boxes, a *tentative track* state can be determined as follows. If a new measurement that has yet been matched with existing tracks is observed, this measurement is put into a *tentative track* state. Next, if the track of the new measurement survives during a certain time step, this track changes into a *confirmed track* state through track initiation, which is indicated by the solid arrows. Thus, the track is considered to have the targets of interest. This is due to the prevention of track activation from a false alarm clutter. On the contrary, if a target of the corresponding track is not detected during a certain time interval, this track alters a *perished track* state through track deletion expressed by the dashed arrows. Consequently, the track is eliminated.

Considering an *initiated grouped track* state shown in the dashed box in Fig. 4, two small dashed boxes wrapping a low-level track of a *confirmed track* state are contained in the large box. This implies that a grouped track can be initiated when a single track enters a *confirmed track* state. Furthermore, when two low-level tracks of *tentative track* states within a threshold of the gap between two legs become *confirmed track* states, a grouped track having two low-level tracks can also be activated instantly. If a new confirmed low-level track lies within a threshold of the existing grouped track that was already activated using the one low-level track, it can be included in this existing grouped track.

Unlike an *initiated grouped track* state, a *perished grouped track* state consists of a dashed box without additional boxes. This is because a grouped track is not deleted when one contained low-level track is deleted, but only deleted when two contained low-level tracks are both deleted. In this sense, a grouped track can include at most two tracks at a low level.

B. State Prediction

Prior work on extracting human legs for tracking assumes that both legs of a person are detected [28], [29]. Sometimes

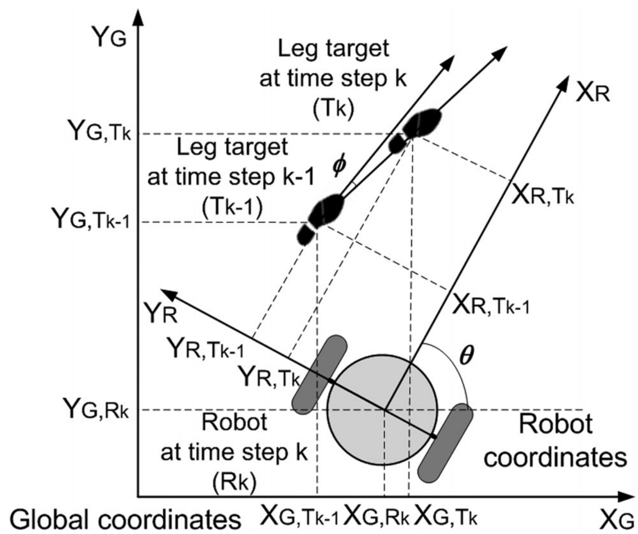


Fig. 5. Coordinate systems for a target and robot with respect to global coordinates.

both legs cannot be extracted such as when one leg is occluded by the other, or when detection failure, clutter or false alarms take place. Our methods require a human leg to be tracked and estimated as a target, and therefore, extracting both legs of the person is not necessary. A motion model for each leg thereby enables a state prediction.

The authors of [30] and [31] point out that modeling the human walking motion is a challenging task as it is neither simple linear nor Brownian motion. We assume that the target gait is composed of either a straight or turning motion in a 2-D plane. To consider both motions together, we not only adopted the parameters of the nearly constant speed horizontal turn model, described in [23] and [32], but also took into account the term of accelerations for both motions to include the changes in velocity. Fig. 5 represents the coordinates of a leg target with respect to the robot and global coordinates. The following equation computes the state variables of a target at time step k:

$$\begin{bmatrix} x(k) \\ y(k) \\ \phi(k) \\ v(k) \\ w(k) \end{bmatrix} = \begin{bmatrix} x(k-1) \\ y(k-1) \\ \phi(k-1) \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} v(k-1)\Delta t \cos(\phi(k-1)) \\ v(k-1)\Delta t \sin(\phi(k-1)) \\ w(k-1)\Delta t \\ v(k-1) \\ w(k-1) \end{bmatrix}$$
$$+ \begin{bmatrix} \frac{1}{2}\Delta t^2 \cos(\phi(k-1)) & 0 \\ \frac{1}{2}\Delta t^2 \sin(\phi(k-1)) & 0 \\ 0 & \frac{1}{2}\Delta t^2 \\ \Delta t & 0 \\ 0 & \Delta t \end{bmatrix} \begin{bmatrix} a_t(k-1) \\ a_r(k-1) \end{bmatrix} + \varepsilon(k-1) \tag{18}$$

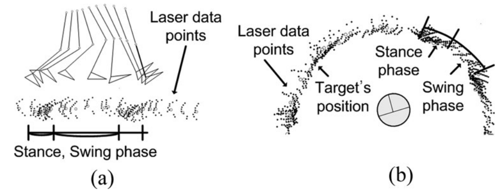


Fig. 6. Raw laser points and leg position while walking. (a) Straight motion. (b) Turning motion.

Algorithm 2 Determination of the leg phase

$N \leftarrow The\ number\ of\ detected\ targets$
for $i \leftarrow 1\ to\ N$ **do**
  $\mathbf{D}(k) \leftarrow \|\mathbf{x}_i(k) - \mathbf{x}_i(k-1)\|$
  if $\mathbf{D} < threshD$ **then**
    $(threshD\ :\ possible\ leg\ motion\ region\ when$
    $a\ foot\ is\ on\ the\ ground)$
    $LegPhase_i() \leftarrow StancePhase()$
    $(a_t)_t, (a_r)_i \leftarrow 0$
    $((a_t)_i\ :\ a_t\ of\ target\ i,\ (a_r)_i\ :\ a_r\ of\ target\ i)$
  else
    $\mathbf{V}(k) \leftarrow (v_i(k) - v_i(k-1)),$
    $\mathbf{W}(k) \leftarrow (w_i(k) - w_i(k-1))$
    if $\mathbf{V} > 0$ **then**
      $LegPhase_i() \leftarrow AccelSwingPhase()$
      $(a_t)_i \leftarrow (\mathbf{V}(k) - \mathbf{V}(k-1))/\Delta t$
      $(\Delta t\ :\ one\ time\ step)$
      $(a_r)_i \leftarrow (\mathbf{W}(k) - \mathbf{W}(k-1))/\Delta t$
    else
      $LegPhase_i() \leftarrow DecelSwingPhase()$
      $(a_t)_i \leftarrow (\mathbf{V}(k) - \mathbf{V}(k-1))/\Delta t,$
      $(a_r)_i \leftarrow (\mathbf{W}(k) - \mathbf{W}(k-1))/\Delta t$
    end if
  end if
end for
return $((a_t)_i,\ (a_r)_i)$

where $x(k-1)$, $y(k-1)$, $\phi(k-1)$, $v(k-1)$, and $w(k-1)$ denote the x and y positions, heading angle, velocity, and angular velocity at time $k-1$, respectively. In addition, Δt represents the sampling time. $a_t(k-1)$ and $a_r(k-1)$ indicate the translational and rotational accelerations, respectively. $\varepsilon(k-1)$ is the process noise term. We must obtain the values of $a_t(k-1)$ and $a_r(k-1)$ at each time step to proceed with the state prediction for the samples of the corresponding target. The term $\varepsilon(k-1)$ should not be set too small to avoid sample degeneracy [33].

As shown in Fig. 6, there are two phases to a human gait [34], [35]. In the stance phase, a leg is connected to the ground and does not move very much. In the swing phase, the leg is detached from the ground and is accelerating or decelerating. Swing phase can appear in both straight and turning walking gaits.

In Algorithm 2, to acquire $a_t(k)$ and $a_r(k)$, a leg phase is separated into three phases: stance, acceleration swing, and deceleration swing. $a_t(k)$ and $a_r(k)$ are obtained from the target's

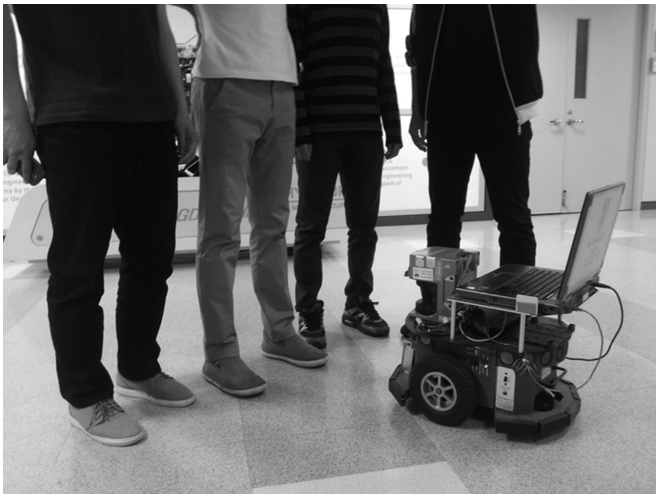


Fig. 7. Robot surrounded by four people.

motion from the previous time step, which is used by the motion model to predict the position of the target for the next time step.

V. Human-Following Evaluation

A. Experimental Setup

We implemented the proposed HSJPDAF and applied the human-following function to a two-wheeled differential mobile robot, a Pioneer 3 DX. A Sick LMS200 LRF was mounted onto the mobile robot to detect the human legs. We set our LRF to an angular resolution of 1 for a field of view of 180. The sensing frequency was 5 Hz, and the LRF was placed at a height of 28 cm above the ground.

First, the human target was selected and found among the initially extracted legs using HSJPDAF. We employed the human-following method for the mobile robot from [20] and [36]. The motion controller in [20] was inspired by [37] and [38]. The particle filter for each target has 1000 samples, and the total number of samples is represented by N. The false alarm probability γ was set to 0.1 as in [6]. For the purpose of the human-following method, the perceptual range of the sensor was up to 1.8 m horizontally. Since detection performance is highly sensitive to the resolution of the sensor data, we limited the perceptual range of the sensor.

B. Hierarchical Sample-Based Joint Probabilistic Data Association Filter

The populated environment, in which a mobile robot is closely surrounded by four people, appears in Fig. 7.

Fig. 8 describes the results of the proposed method at a certain time step in the environment shown in Fig. 7. Fig. 8(a)–(d) is plotted at the same time step and indicates the sequential process of HSJPDAF. Fig. 8(a) indicates the measurements that correspond to the legs obtained from the sensor model according to [20]. Fig. 8(b) shows the results of SJPDAF, where each different color indicates a different target being tracked. Approximately 1000 samples are used for representing the distribution of the target position. Fig. 8(a) and (b) shows the results of the

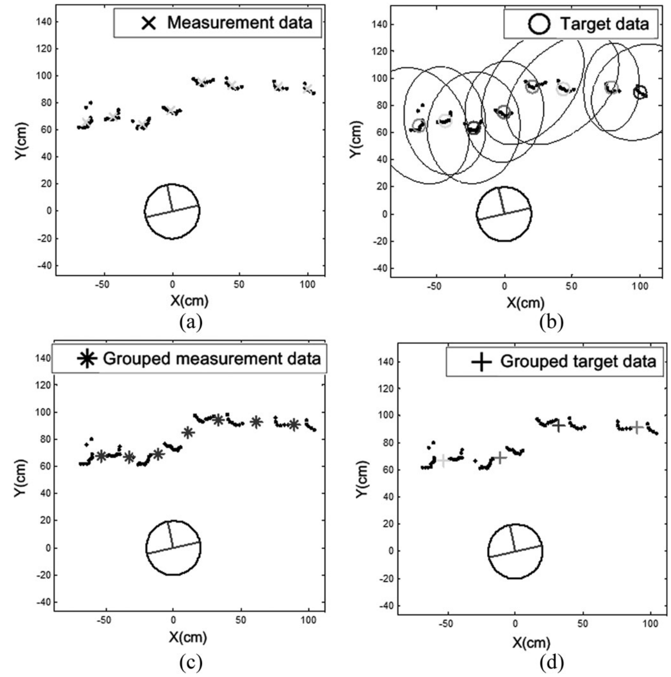


Fig. 8. Process of HSJPDAF. (a) Measurement data obtained from the sensor model. (b) Results of SJPDAF with its targets. (c) Results of creating set q of grouped measurements. (d) Results of HSJPDAF with its grouped targets.

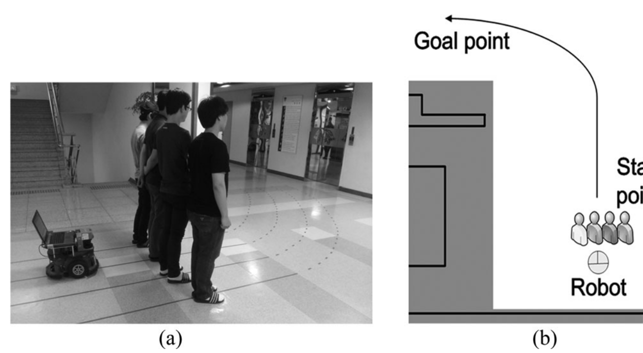


Fig. 9. Closely spaced environment. (a) Four people walking along a predefined path. (b) Schematic diagram.

leg estimation at a low level. The assignment probabilities of a low level, β_{ji}, can be computed. In Fig. 8(b), the validation gate for each target is formed based on the squared normalized innovation between the measurement and target, according to [39].

Fig. 8(c) shows set q of grouped measurements at a high level. Set q is used to match the candidate to the grouped targets to determine the human positions using low-level information. We assumed a threshold of the possible width between two legs with respect to the sensing frequency. Two measurements within the set threshold from each other are thus paired into grouped measurements. In total, ten grouped measurements were obtained under this scenario.

Fig. 8(d) shows that the grouped targets of the current time step can be found using the constructed assignment matrix and Murty algorithm. The assignment matrix that consists of the assignment probabilities at a high level, α_{qp}, can be formed

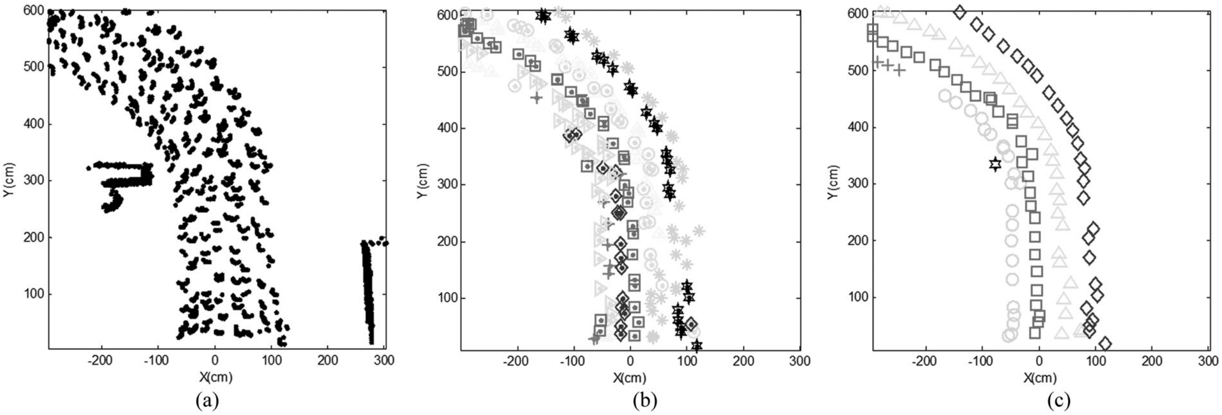


Fig. 10. Trajectories of targets and grouped targets. (a) Accumulated raw laser data. (b) Results of SJPDAF. (c) Results of the proposed HSJPDAF.

regarding the grouped measurements with the grouped targets of the previous time step. "+" marks for the grouped targets show that the positions of the four individuals were detected correctly, as shown in Fig. 7. In this manner, we can obtain their positions at each time step using the interdependence of their legs through HSJPDAF.

C. Comparison of Hierarchical Sample-Based Joint Probabilistic Data Association Filter With Sample-Based Joint Probabilistic Data Association Filter

The performance of SJPDAF and HSJPDAF were compared in a closely spaced environment. Three people walked along a predetermined path in the vicinity of the target person, as shown in Fig. 9.

It is difficult to obtain the actual ground truth of a person walking, and we therefore produced the path manually. The path consists of straight and turning sections, as shown in Fig. 9(a). A distance of 45 cm was initially set between two individuals. We assumed the ground truth based on the path with 10 cm intervals from side to side. The errors were computed from the normal distance to the base line. The mean error of the experiment was 8 cm during 80 time steps. However, this error is not critical to the following performance because it is not large enough to cause confusion about the persons being tracked while carrying out the human-following function.

From the same results obtained from the experiment shown in Fig. 9, the accumulated raw laser data show the footprints of people walking and are provided in Fig. 10(a). Fig. 10(b) and (c) shows the results of the conventional SJPDAF and the results of the proposed HSJPDAF, respectively. Both results show trajectories of low-level targets and high-level grouped targets. The same tracks from the experiment are marked with the same color and mark in the figures. For the purpose of clarity, all targets that are plotted in the figures are obtained every five time steps. As shown in Fig. 10(b), each distinctive target is not consistently maintained and changes frequently. Forty-eight of the total 640 time steps were changed, which shows its instability in a cluttered environment. This result shows the difficulty in maintaining the target tracks due to confusion with the neighboring tracks.

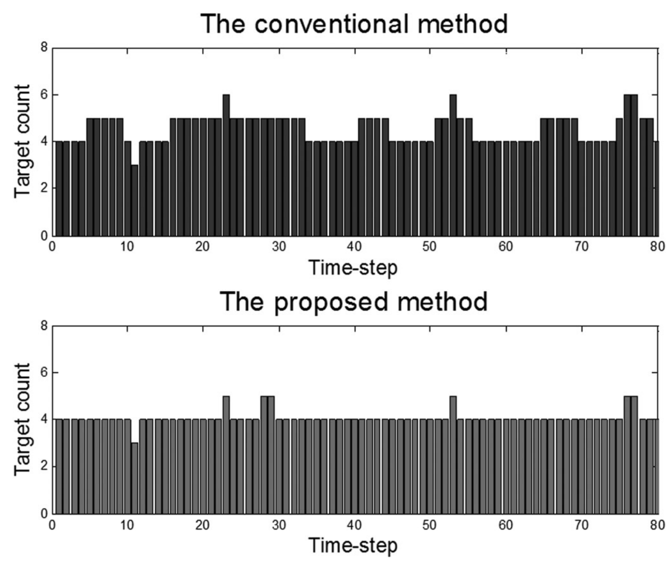
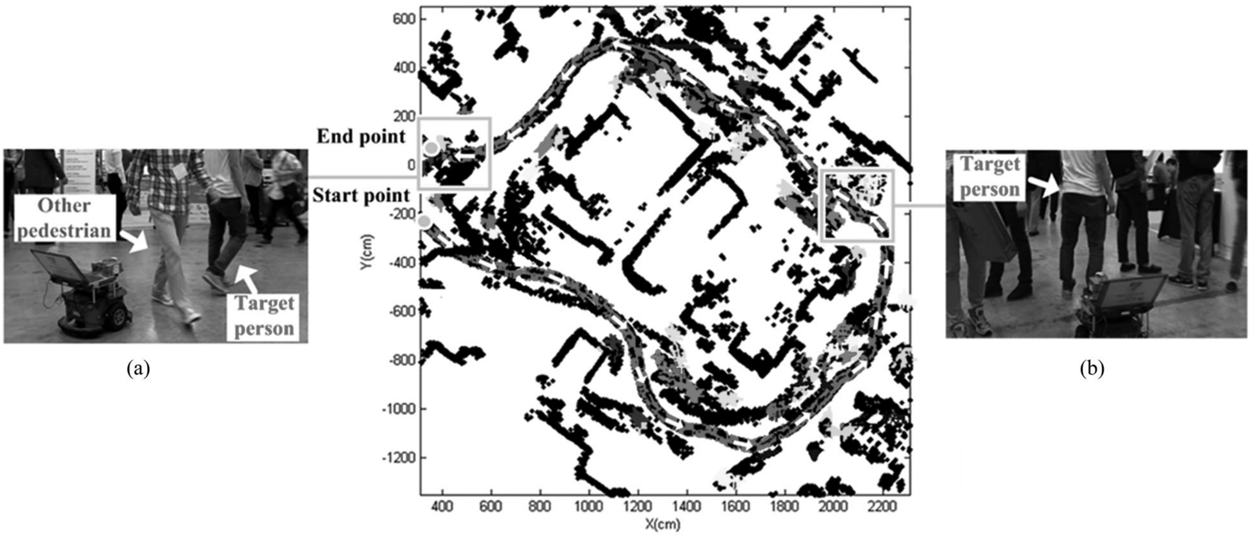


Fig. 11. Comparison of the estimated number of persons.

In Fig. 10(c), each grouped target was maintained from the start to the end points. Even though one of the low-level targets within the grouped target was lost, HSJPDAF was able to track the grouped target using another low-level target. However, HSJPDAF could also fail to track a grouped target when a failure in the tracking of both low-level targets occurs at the same time step, as shown in the upper left corner. As the dissipation of two low-level targets at the same time step was rare, the failure cases for a single target were compensated from the view point of a grouped target. This result shows that the proposed HSJPDAF is more effective when the targets are closely gathered together.

Fig. 11 shows the number of estimated persons using the conventional SJPDAF and the proposed HSJPDAF. The SJPDAF method uses the nearest-neighbor decision rule for determining the human positions by grouping two legs. HSJPDAF determines the human positions based on the interdependence of the targets. Since there were four people, it is desirable to maintain four targets during the experiment in Fig. 11. Fig 11 shows that the proposed HSJPDAF outperformed the conventional SJPDAF during the test. Furthermore, our technique is more



(a) (b) (c)

Fig. 12. Human-following at WIS. (a) Other person occlusion case. (b) Cluttered case with six people. (c) Positions using HSJPDAF.

robust than the conventional method in dealing with a track coalescence, which is a well-known problem in JPDAF, and a failure in the observation measurements. The reason is that the proposed method is capable of considering two auxiliary targets for the goal of tracking a grouped target.

Both SJPDAF and HSJPDAF suffer from exponential complexity when increasing the number of tracked targets. Similar to [6] and [40], the computational cost can be reduced as the sensor's local field of view can only contain a few targets, and different targets are tracked independently by using individual particle filters. In Fig. 9 where there were eight low-level targets being tracked, the average frequency of SJPDAF and that of HSJPDAF were 24.4 and 14.9 Hz, respectively. While the proposed technique requires more computational resources due to additional high-level process computations, the requisite frequency is still higher than the sensing frequency, which is admissible for real-time implementation.

D. Human-Following in a Real-World Environment

Fig. 12 shows results of human-following at World IT Show (WIS). Different colored marks indicate the different positions of humans using HSJPDAF, and the red line with white arrows indicates the movement of a target person. A total of 1200 time steps, which continued for 240 s, were used to accumulate the data, and the average walking speed of the user was 0.53 m/s. Fig. 12(a) shows the occlusion situation. We initially set out to dissipate the track when the samples for the target did not appear for more than eight time steps sequentially, and therefore, a short period occlusion could be handled. A similar period occlusion could also be overcome in our previous research [20]; however, working in a crowded place induces the tracking failure due to the difficulty in dealing with the data association.

Overall, in the low-level case corresponding to the general SJPDAF, the ID of targets of interest was misidentified by others for 107 time steps out of total 1200 time steps. The high-level case, on the other hand, did not have a tracking failure of the user under this particular scenario even though a trace of the user was sometimes biased because one low-level target was missed. The red line shows that the target person was not lost during the test, and thus, our method is robust in a real-world environment. A video of experimental results in Fig. 12 can be found at the following URL: https://youtu.be/CjUWeuvQ5lQ.

VI. Conclusion

In this paper, we have addressed a novel human-tracking technique with a mobile robot using HSJPDAF. Our proposed method strengthens the existing alternatives in the manner of hierarchical consideration of targets of interest.

The proposed HSJPDAF consists of two levels. The low level is for the process of estimation and the high level is for the process of data association. The information on a person, which is a grouped target, can be obtained based on the estimated information of the targets with respect to their legs. The proposed scheme allows legs to be grouped based on a probabilistic approach. Moreover, a simple motion model can be implemented because a target is considered to be a leg, not two legs. Various experiments are carried out to verify that our proposed technique shows a better performance than the conventional scheme depending on merely a single tracking in a highly populated environment.

References

[1] Y. Bar-Shalom and X. R. Li, *Multitarget-Multisensor Tracking: Principles and Techniques*. Storrs, CT, USA: YBS Publishing, 1995.

[2] D. Reid, "An algorithm for tracking multiple targets," *IEEE Trans. Autom. Control*, vol. AC-24, no. 6, pp. 843–854, Dec. 1979.

[3] Y. Bar-Shalom, F. Daum, and J. Huang, "The probabilistic data association filters," *IEEE Control Syst. Mag.*, vol. 29, no. 6, pp. 82–100, Dec. 2009.

[4] S. Blackman, "Multiple hypothesis tracking for multitarget tracking," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 40, no. 1, pp. 5–18, Jan. 2004.

[5] I. J. Cox, "A review of statistical data association techniques for motion correspondence," *Int. J. Comput. Vision*, vol. 10, no. 1, pp. 5366, Feb. 1993.

[6] D. Schulz, W. Burgard, D. Fox, and A. B. Cremer, "People tracking with a mobile robot using sample-based joint probabilistic data association filters," *Int. J. Robot. Res.*, vol. 22, no. 2, pp. 99–116, Feb. 2003.

[7] P. Kondaxakis, S. Kasderidis, and P. Trahanias, "A multi-target tracking technique for mobile robots using a laser range scanner," in *Proc. IEEE Int. Conf. Robot. Autom.*, Nice, France, Sep. 2008, pp. 3370–3377.

[8] C. Rasmussen and G. Hager, "Joint probabilistic techniques for tracking multi-part objects," in *Proc. Comput. Vision Pattern Recog.*, Jun. 1998, pp. 16–21.

[9] D. Musicki and R. Evans, "Joint integrated probabilistic data association: JIPDA," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 40, no. 3, pp. 1093–1099, Jul. 2004.

[10] E. A. Topp and H. I. Christensen, "Tracking for following and passing persons, in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, Aug. 2005, pp. 2321–2327.

[11] R. Kirby, J. Forlizzi, and R. Simmons, "Natural person-following behavior for social robots," in *Proc. Human-Robot Interaction*, Mar. 2007, pp. 17–24.

[12] P. Vadakkepat, P. Lim, L. C. De Silva, L. Jing, and L. L. Ling, "Multimodal approach to human-face detection and tracking," *IEEE Trans. Ind. Electron.*, vol. 55, no. 3, pp. 1385–1393, Mar. 2008.

[13] J. Garcia, A. Gardel, I. Bravo, J. L. Lazaro, M. Martinez, and D. Rodriguez, "Directional people contour based on heads tracking," *IEEE Trans. Ind. Electron.*, vol. 60, no. 9, pp. 3991–4000, Sep. 2013.

[14] J. Cui, H. Zha, H. Zhao, and R. Shibasaki, "Laser based detection and tracking of multiple people in crowds," *Comput. Vis. Image Und.*, vol. 106, no. 2/3, pp. 300–312, May 2007.

[15] A. Carballo, A. Ohya, and S. Yuta, "Multiple people detection from a mobile robot using double layered laser range finders," in *Proc. IEEE Int. Conf. Robot. Autom.*, Kobe, Japan, May 2009, pp. 1–7.

[16] K. O. Arras, S. Grzonka, M. Luber, and W. Burgard, "Efficient people tracking in laser range data using a multi-hyporhesis leg-tracker with adaptive occlusion probabilities," in *Proc. IEEE Int. Conf. Robot. Autom.*, Pasadena, CA, USA, May 2008, pp. 1710–1715.

[17] T. Horiuchi, S. Thompson, S. Kagami, and Y. Ehara, "Pedestrian tracking from a mobile robot using a laser range finder," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2007, pp. 931–936.

[18] M. Luber, G. D. Tipaldi, and K. O. Arras, "Place-dependent people tracking," *Int. J. Robot. Res.*, vol. 30, no. 3, pp. 280–293, Jan. 2011.

[19] G. Taylor and L. Kleeman, "A multiple hypothesis walking person tracker with switched dynamic model," presented at the Australasian Conf. Robot. Autom., Canberra, Australia, Dec. 2004.

[20] W. Chung, H. Kim, Y. Yoo, C. B. Moon, and J. Park, "The detection and following of human legs through inductive approaches for a mobile robot with a single laser range finder," *IEEE Trans. Ind. Electron.*, vol. 59, no. 8, pp. 3156–3166, Aug. 2012.

[21] J. D. C. Little, K. G. Mury, D. W. Sweeney, and C. Karel, "An algorithm for the traveling salesman problem," *Oper. Res.*, vol. 11, pp. 972–989, Dec. 1963.

[22] I. Cox, M. Miller, R. Danchick, and G. Newnam, "On finding ranked assignments with application to multitarget tracking and motion correspondence," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, no. 1, pp. 486–489, Jan. 1995.

[23] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Norwood, MA, USA: Artech House, 1999.

[24] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Feb. 2002.

[25] A. Almeida, J. Almeida, and R. Araujo, "Real-time tracking of multiple moving objects using particle filters and probabilistic data association," *Automatika*, vol. 46, nos. 1/2, pp. 39–48, Dec. 2005.

[26] Y. M. Baek, S. Tanaka, K. Harada, N. Sugita, A. Morita, S. Sora, and M. Mitsuishi, "Robust visual tracking of robotic forceps under a microscope using kinematic data fusion," *IEEE/ASME Trans. Mechatronics*, vol. 19, no. 1, pp. 278–288, Feb. 2014.

[27] S. Y. T. Lin and Y. F. Lin, *Set Theory: An Intuitive Approach*. Boston, MA, USA: Houghton Mifflin, 1974.

[28] X. Shao, H. Zhao, K. Nakamura, K. Katabira, R. Shibasaki, and Y. Nakagawa, "Detection and tracking of multiple pedestrians by using laser range scanners," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, Oct. 2007, pp. 2174–2179.

[29] J. H. Lee, T. Tsubouchi, K. Yammamoto, and S. Egawa, "People tracking using a robot in motion with laser range finder," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, Oct. 2006, pp. 2936–2942.

[30] D. F. Glas, T. Miyashita, H. Ishiguro, and N. Hagita, "Laser-based tracking of human position and orientation using parametric shape modeling," *Adv. Robot.*, vol. 23, no. 4, pp. 405–428, Apr. 2009.

[31] N. Bellotto and H. Hu, "Multisensor-based human detection and tracking for mobile service robots," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 1, pp. 167–181, Feb. 2009.

[32] J. L. Gertz, "Multisensor surveilance for improved aircraft tracking," *The Lincoln Lab. J.*, vol. 2, no. 3, pp. 381–396, 1989.

[33] A. Doucet, J. F. G. de Freitas, and N. J. Gordon, *Sequential Monte Carlo Methods in Practice*. New York, NY, USA: Springer-Verlag, 2001.

[34] K. Wakita, J. Huang, P. Di, K. Sekiyama, and T. Fukuda, "Human-walking-intention-based motion control of an omnidirectional-type cane robot," *IEEE/ASME Trans. Mechatronics*, vol. 18, no. 1, pp. 285–296, Feb. 2013.

[35] M. P. Murray, A. B. Drought, and R. C. Kory, "Walking patterns of normal men," *J. Bone Joint Surg.*, vol. 46, no. 2, pp. 335–360, Mar. 1964.

[36] K. Morioka, J. -H. Lee, and H. Hashimoto, "Human-following mobile robot in a distributed intelligent sensor network," *IEEE Trans. Ind. Electron.*, vol. 51, no. 1, pp. 229–237, Feb. 2004.

[37] M. Cheuh, Y. L. W. A. Yeung, K. P. C. Lei, and S. S. Joshi, "Following controller for autonomous mobile robots using behavioral cues," *IEEE Trans. Ind. Electron.*, vol. 55, no. 8, pp. 3124–3132, Aug. 2008.

[38] J. S. Hu, J. J. Wang, and D. M. Ho, "Design of sensing system and anticipative behavior for human following of mobile robots, *IEEE Trans. Ind. Electron.*, vol. 61, no. 4, pp. 1916–1927, Apr. 2014.

[39] M. Adams, S. Zhang, and L. Xie, "Particle filter based outdoor robot localization using natural features extracted from laser scanners," in *Proc. IEEE Int. Conf. Robot. Autom.*, New Orleans, LA, USA, Apr. 2004, pp. 854–859.

[40] Z. Khan, T. Balch, and F. Dellaert, "MCMC-based particle filtering for tracking a variable number of interacting targets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 11, pp. 1805–1819, Nov. 2005.

Yoonchang Sung received the B.S. and M.S. degrees from the Department of Mechanical Engineering, Korea University, Seoul, Korea, in 2011 and 2013, respectively. He is currently working toward the Ph.D. degree with the Department of Mechanical Engineering, Virginia Tech, Blacksburg, VA, USA.

His main research interests include object detection, search and tracking, and cooperative robotics.

Woojin Chung received the B.S. degree from the Department of Mechanical Design and Production Engineering, Seoul National University, Seoul, Korea, in 1993, and the M.S. and Ph.D. degrees from the Department of Mechano-Informatics, The University of Tokyo, Tokyo, Japan, in 1995 and 1998, respectively.

Since 2005, he has been with the Department of Mechanical Engineering, Korea University, Seoul. His research interests include the design and control of nonholonomic underactuated mechanical systems, trailer system design and control, and mobile robot navigation.