Cooperative Tracking of People Using Networked LiDARs

Marino Matsuba, Ryota Imai
Graduate School of Science and Engineering
Doshisha University
Kyotanabe, Kyoto, 610-0394 Japan
e-mail: {ctwh0137, ctwj0109}@mail4.doshisha.ac.jp

Masafumi Hashimoto, Kazuhiko Takahashi
Faculty of Science and Engineering
Doshisha University
Kyotanabe, Kyoto, 610-0394 Japan
e-mail: {mhashimo, katakaha}@mail. doshisha.ac.jp

***Abstract*—This paper presents a tracking method of people using networked Light Detection And Ranging sensors (LiDARs) set in an environment. Each LiDAR detects people from the LiDAR scan data using a background subtraction method and sends the positions of the people to the neigboring LiDARs. It estimates the people's positions and velocities and exchanges information with the neigboring LiDARs. A Distributed Interacting MultiModel (DIMM)-based method is used to accurately estimate people's positions and velocities under various motion modes, such as stopping, walking, and suddenly running, in a distributed manner without a central server. Simulation experiments of the tracking of 20 people using three Velodyne 32-layer LiDARs are conducted in two different network topologies (ring and line network topologies) to quantitatively evaluate the tracking performance and computation effort of the proposed method. Simulation results show that the tracking performance and computation time of the DIMM-based method are comparable to those of conventional centralized interacting multimodel-based method.**

Keywords—LiDAR; people tracking; cooperative tracking; diffusion strategy; interacting multimodel estimator.

I. INTRODUCTION

Estimating the motion and behavior (*i.e.*, tracking) of moving objects, such as people, cars, and two-wheelers, in an environment is vital in several applications, including Intelligent Transport Systems (ITS), autonomous driving, security, and surveillance. Therefore, many tracking systems based on Light Detection And Ranging sensors (LiDARs) and cameras have been developed [1]–[3]. In this paper, we investigate people tracking using LiDARs allocated in an environment.

In sparsely populated environments, LiDAR-based tracking of people is efficient. However, its performance in crowded environments is poor because of occlusions. An effective method for accurately tracking people in crowded environments is the use of networked multiple LiDARs. Despite occasions where people are occluded or are located outside a surveillance area of a LiDAR, the use of networked multiple LiDARs (referred to as cooperative tracking) improves tracking reliability and accuracy as tracking data are shared among LiDARs [4][5].

For the application of cooperative tracking to ITS domains, we proposed cooperative tracking of people using networked multiple ground LiDARs allocated to different locations in an intersection environment [6]. The cooperative tracking method detects people's positions, velocities, and behavior, such as stopping, walking, and suddenly running. Usual algorithms for people tracking are based on Bayesian filters and assume that people walk or run at an almost constant speed. Therefore, when people suddenly change their motions, such as suddenly running, turning, or stopping, the tracking accuracy decreases. To accurately track people under such conditions, we proposed a multimodel-based approach, which employs an Interacting MultiModel (IMM) estimator [7], instead of the single-model Kalman filter approach.

Most studies on cooperative tracking employ centralized data fusion with a central server, in which sensing data are captured and preprocessed by each sensor, sent to a central server, and then fused in the central server [4][5]. In the first version of our cooperative tracking system [6][8], a Centralized Interacting MultiModel (CIMM) estimator [9] was employed to estimate people's positions, velocities, and behavior by the central server.

Centralized data fusion reduces system robustness and scalability. Recently, various methods for distributed state estimation have been proposed in the field of Bayesian filtering [10][11], in which information processing among multiple sensors is distributed among sensors without using a central server. Thus, in our previous study [12], we proposed cooperative tracking of people that functions in a distributed manner without any central servers, and a Distributed Interacting MultiModel (DIMM) estimator [13] was employed. However, people could be tracked using only two LiDARs.

In this paper, DIMM-based cooperative tracking of people is presented using three LiDARs. The contributions of this paper are as follows:

- A DIMM-based cooperative people tracking method using three LiDARs are designed in two different network topologies (ring and line network topologies). The tracking method is applicable to four or more LiDARs systems in any network topology.
- The tracking performance and computational effort of the presented DIMM-based method are quantitatively evaluated by comparing conventional CIMM-based and Kalman filter-based methods.

The rest of this paper is organized as follows: Section II gives an overview of the experimental system. Section III models people's motion, and Sections IV and V describe the people detection and tracking methods, respectively. Section VI presents the simulation conducted to evaluate the performance of the proposed cooperative tracking system. Section VII concludes the paper.

 ISBN: 978-1-68558-090-2

II. Experimental System

Figure 1 shows our experimental system, which consists of three LiDARs. Each LiDAR has a 32-layer LiDAR (Velodyne HDL-32E) and a computer. The maximum range of the LiDAR is 50 m. The horizontal and vertical viewing angles are 360° and 41.3° with resolutions of 0.16° and 1.33°, respectively. The scanning period is 0.1 s.

Two network topologies can be considered for exchanging information among LiDARs: a ring network topology (referred to as a ring network) and a line network topology (a line network). As shown in Figure 1, each LiDAR is connected to two other adjacent LiDARs in the ring network, whereas, in the line network, LiDARs 1 and 2 and LiDARs 2 and 3 are connected.

For four or more LiDARs, similar to the case of three LiDARs, each LiDAR is connected to two other LiDARs on both sides in a ring network, whereas, in a line network, only the LiDARs at both ends of the line are connected to one adjacent LiDAR, and other LiDARs are connected to the two LiDARs.

III. Motion and Measurement Models of a Person

To accurately track people in an intersection environment, we consider three motion modes of a person as follows [8]:

- Stop mode (mode 1): A person stops.
- Constant velocity mode (mode 2): A person walks or runs at an almost constant velocity.
- Sudden motion mode (mode 3): A person starts to suddenly run or stops suddenly.

A person's position is denoted by (x, y), and the moving direction of the person is denoted by θ. The translational and turning velocities of the person are denoted by v and $\dot{\theta}$, respectively. The three motion modes are then modeled by the following state equations:

- Mode 1

$$\begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} x_{t-1} + \Delta\dot{x}_{t-1}\tau \\ y_{t-1} + \Delta\dot{y}_{t-1}\tau \end{bmatrix} \tag{1}$$

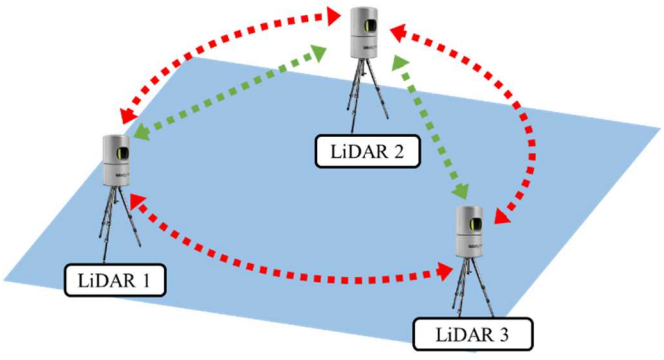


Figure 1. Overview of the networked three LiDARs. The red and green dotted lines indicate ring and line network topologies, respectively.

- Mode 2

$$\begin{bmatrix} x_t \\ y_t \\ \theta_t \\ v_t \\ \dot{\theta}_t \end{bmatrix} = \begin{bmatrix} x_{t-1} + (v_{t-1}\tau + \frac{1}{2}\Delta\dot{v}_{t-1}\tau^2)\cos\theta_{t-1} \\ y_{t-1} + (v_{t-1}\tau + \frac{1}{2}\Delta\dot{v}_{t-1}\tau^2)\sin\theta_{t-1} \\ \theta_{t-1} + \dot{\theta}_{t-1}\tau + \frac{1}{2}\Delta\ddot{\theta}_{t-1}\tau^2 \\ v_{t-1} + \Delta\dot{v}_{t-1}\tau \\ \dot{\theta}_{t-1} + \Delta\ddot{\theta}_{t-1}\tau \end{bmatrix} \tag{2}$$

- Mode 3

$$\begin{bmatrix} x_t \\ y_t \\ \theta_t \\ v_t \\ \dot{v}_t \end{bmatrix} = \begin{bmatrix} x_{t-1} + (v_{t-1}\tau + \frac{1}{2}\dot{v}_{t-1}\tau^2 + \frac{1}{6}\Delta\ddot{v}_{t-1}\tau^3)\cos\theta_{t-1} \\ y_{t-1} + (v_{t-1}\tau + \frac{1}{2}\dot{v}_{t-1}\tau^2 + \frac{1}{6}\Delta\ddot{v}_{t-1}\tau^3)\sin\theta_{t-1} \\ \theta_{t-1} + \Delta\dot{\theta}_{t-1}\tau \\ v_{t-1} + \dot{v}_{t-1}\tau + \frac{1}{2}\Delta\ddot{v}_{t-1}\tau^2 \\ v_{t-1} + \Delta\ddot{v}_{t-1}\tau \end{bmatrix} \tag{3}$$

where t and t-1 indicate time steps. $\dot{v}$ and $\ddot{\theta}$ are the translational and turning accelerations of the person, respectively. $\Delta\dot{x}$, $\Delta\dot{y}$, $\Delta\dot{v}$, $\Delta\ddot{v}$, $\Delta\dot{\theta}$ and $\Delta\ddot{\theta}$ are the plant disturbances. τ (= 100 ms) is the sampling period of the LiDAR.

For simplicity, the state equation of the m-th mode (m = 1, 2, 3) is represented by the following vector form:

$$\boldsymbol{x}_t^m = \boldsymbol{f}^m(\boldsymbol{x}_{t-1}^m, \Delta\boldsymbol{v}_{t-1}^m) \tag{4}$$

where $\boldsymbol{x}^m$ is the state vector, and $\Delta\boldsymbol{v}^m$ is the plant disturbance vector, which is assumed to have a white noise sequence with the covariance matrix $\boldsymbol{Q}^m$.

The LiDAR measurement related to a person gives the following equation:

$$\boldsymbol{z}_t = \boldsymbol{H}^m\boldsymbol{x}_t^m + \Delta\boldsymbol{z}_t \tag{5}$$

where $\boldsymbol{z} = (z_x, z_y)^T$ is the position of the person. $\Delta\boldsymbol{z}$ is the measurement noise, which is assumed to have a white noise sequence with the covariance matrix $\boldsymbol{R}$. $\boldsymbol{H}^m$ is the measurement matrix.

IV. People Detection Using Background Subtraction Method

Figure 2 shows the sequence of people detection and tracking. Each LiDAR captures its own scan data and maps them onto an elevation map. In the elevation map, a cell containing two or more scan data is called an occupied cell. Each LiDAR extracts the occupied cell related to a person (referred to as a person-cell) based on the background subtraction method. Generally, the LiDAR scan data related to a person occupy two or more cells, and the neighboring person cells are clustered (referred to as person-cell group).

Each LiDAR communicates with the adjacent LiDAR and exchanges the information of clustered person cells. Thereafter, each LiDAR fuses the information of person cells and then determines the geometric center of person cells as the

 ISBN: 978-1-68558-090-2

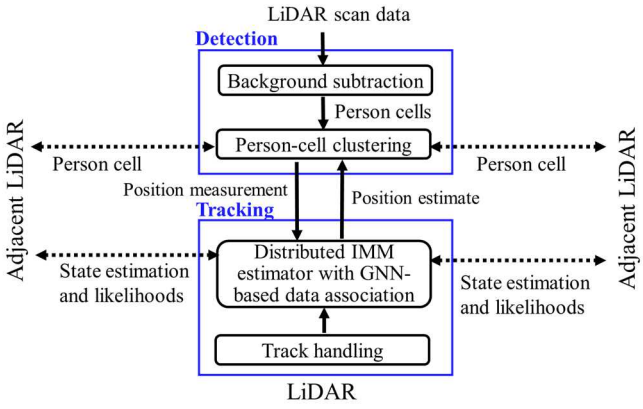
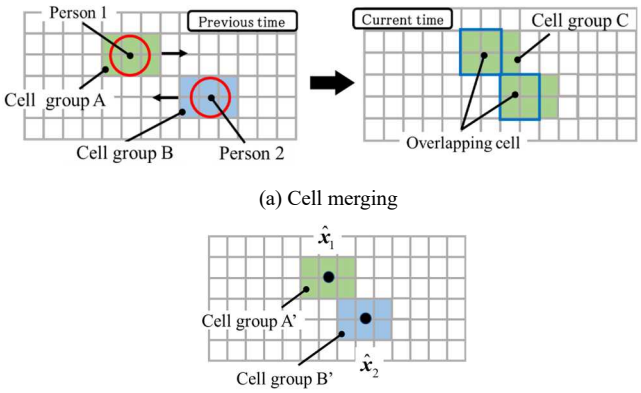


Figure 2. Sequence of people detection and tracking.



(a) Cell merging

(b) Cell splitting

Figure 3. Cell merging and splitting of two passing people (top view).

person's position.

In our simulation, the cell size of the elevation map is set to 0.3 m. Therefore, when the distance of neighboring two people is less than 0.3 m, two person-cell groups are merged, and the people are detected as a person. To address this problem, the merged person-cell groups are split, as described below.

Consider two people passing each other. Figure 3 shows the person cells for such people on the elevation map. Initially, two person-cell groups A and B for persons 1 and 2, respectively, are extracted, and at the current time, the person-cell groups are merged into C. We examine whether the cells in person-cell groups A and B partially overlap those in person-cell group C. If the cells overlap, as those in a blue frame in Figure 3(a), the two person-cell groups are considered to have merged. Then, the merged person-cell group C is split to accurately track the two people.

The coordinates of each cell in the merged person-cell group C are compared with the positions $\hat{\boldsymbol{x}}_1$ and $\hat{\boldsymbol{x}}_2$ of persons 1 and 2, respectively, estimated at the initial time using our tracker. When the coordinate of the cell is near $\hat{\boldsymbol{x}}_1$ ($\hat{\boldsymbol{x}}_2$), the cell is classified as a person-cell group A' (B') related to person 1 (2). Thereafter, the geometrical centers of the person-cell groups A' and B' are obtained as the positions of persons 1 and 2, respectively.

V. DIMM-Based Cooperative Tracking

A sequence of the above-mentioned motion modes is assumed to be governed by the first-ordered homogeneous Markov chain as follows:

$$T_{mn} = \mathrm{Prob}\left\{\pi_t^n \,\middle|\, \pi_{t-1}^m\right\} \tag{6}$$

$$\sum_{n=1}^{3} T_{mn} = 1 \tag{7}$$

where π_{t-1}^m and π_t^n are events that the m-th and n-th modes (m, n = 1, 2, 3) are in effect at times t–1 and t, respectively. T_{mn} is the transition probability that the m-th mode jumps into the n-th mode. In our simulation, the transition probability matrix is set to $T_{mn} = 0.9$ for $m = n$ and 0.05 for $m \neq n$.

The k-th LiDAR (k = 1, 2, 3) estimates people's state in the following five steps [12][13]:

Step 1) Filter initialization: The probability that the m-th mode occurs at time t–1 is denoted by $\hat{\mu}_{k,t-1}^m$. The m-th mode conditional estimate and its related covariance are denoted by $\hat{\boldsymbol{x}}_{k,t-1}^m$ and $\boldsymbol{P}_{k,t-1}^m$, respectively. The three quantities interact with one another as follows:

$$\hat{\mu}_{k,t/t-1}^n = \sum_{m=1}^{3} \boldsymbol{T}_{mn} \hat{\mu}_{k,t-1}^m \tag{8}$$

$$\bar{\boldsymbol{x}}_{k,t-1}^m = \sum_{n=1}^{3} c_{mn} \hat{\boldsymbol{x}}_{k,t-1}^n \tag{9}$$

$$\bar{\boldsymbol{P}}_{k,t-1}^m = \sum_{n=1}^{3} c_{k,mn} [\boldsymbol{P}_{t-1}^n + (\bar{\boldsymbol{x}}_{k,t-1}^m - \hat{\boldsymbol{x}}_{k,t-1}^n)(\bar{\boldsymbol{x}}_{k,t-1}^m - \hat{\boldsymbol{x}}_{k,t-1}^n)^T] \tag{10}$$

where $c_{k,mn} = T_{mn} \hat{\mu}_{k,t-1}^m / \hat{\mu}_{k,t/t-1}^n$

Step 2) State estimation: A bank of the single-model-based Kalman filters runs, and the prediction and its related covariance for each mode are updated:

$$\left.\begin{aligned} \hat{\boldsymbol{x}}_{k,t/t-1}^m &= \boldsymbol{f}^m(\bar{\boldsymbol{x}}_{k,t-1}^m) \\ \boldsymbol{P}_{k,t/t-1}^m &= \nabla \boldsymbol{F}_{t-1} \bar{\boldsymbol{P}}_{k,t-1}^m \nabla \boldsymbol{F}_{t-1}^T + \nabla \boldsymbol{G}_{t-1} \boldsymbol{Q}^m \nabla \boldsymbol{G}_{t-1}^T \end{aligned}\right\} \tag{11}$$

where $\nabla \boldsymbol{F}$ and $\nabla \boldsymbol{G}$ are Jacobian matrices of $\boldsymbol{f}^m$ ((4)) related to $\bar{\boldsymbol{x}}_{k,t-1}^m$ and $\Delta \boldsymbol{v}_{t-1}^m$, respectively.

By blending the measured position of people, $\boldsymbol{z}_k$, the quantities related to the measurement $\boldsymbol{q}_{k,t}$ and its error covariance $\boldsymbol{S}_{k,t}$ are given by

$$\left.\begin{aligned} \boldsymbol{q}_{k,t}^m &= \sum_{l \in N_k} (\boldsymbol{H}_l^m)^T \boldsymbol{R}_l^{-1} \boldsymbol{z}_{l,t} \\ \boldsymbol{S}_{k,t}^m &= \sum_{l \in N_k} (\boldsymbol{H}_l^m)^T \boldsymbol{R}_l^{-1} \boldsymbol{H}_l^m \end{aligned}\right\} \tag{12}$$

The state estimate $\boldsymbol{\gamma}_{k,t}^m$ and its related error covariance $\boldsymbol{\Gamma}_{k,t}^m$ at time t are determined using the information filter as follows:

$$\left.\begin{aligned} \boldsymbol{\gamma}_{k,t}^m &= (\boldsymbol{\Gamma}_{k,t}^m)^{-1} \{(\boldsymbol{P}_{k,t/t-1}^m)^{-1} \hat{\boldsymbol{x}}_{k,t/t-1}^m + \boldsymbol{q}_{k,t}^m\} \\ \boldsymbol{\Gamma}_{k,t}^m &= \{(\boldsymbol{P}_{k,t/t-1}^m)^{-1} + \boldsymbol{S}_{k,t}^m\}^{-1} \end{aligned}\right\} \tag{13}$$

In (12), N_k is the set of neighboring LiDARs, including itself (i.e., k-th LiDAR). In the ring network, N_1=N_2=N_3={1, 2, 3}, and in the line network, N_1={1, 2}, N_2={1, 2, 3}, and N_3={2, 3}.

 ISBN: 978-1-68558-090-2

The model-conditional likelihood is calculated by

$$\phi_{k,t}^m = \prod_{l \in N_k} \frac{1}{\sqrt{2\pi}\left|\boldsymbol{L}_{k,t/t-1}^m\right|} \exp[-\frac{1}{2}(\tilde{\boldsymbol{z}}_{k,t/t-1}^m)^T (\boldsymbol{L}_{k,t/t-1}^m)^{-1} \tilde{\boldsymbol{z}}_{k,t/t-1}^m] \tag{14}$$

where the predicted measurement error $\tilde{\boldsymbol{z}}_{k,t/t-1}^m$ and its associated covariance $\boldsymbol{L}_{k,t/t-1}^m$ are given by

$$\left.\begin{aligned} \tilde{\boldsymbol{z}}_{k,t/t-1}^m &= \boldsymbol{z}_{k,t} - \boldsymbol{H}_k^m \hat{\boldsymbol{x}}_{k,t/t-1}^m \\ \boldsymbol{L}_{k,t/t-1}^m &= \boldsymbol{H}_k^m \boldsymbol{P}_{k,t/t-1}^m (\boldsymbol{H}_k^m)^T + \boldsymbol{R} \end{aligned}\right\} \tag{15}$$

Step 3) Exchange of tracking information and likelihood: All LiDARs communicate with one another and exchange information about the state estimate $\boldsymbol{\gamma}_{k,t}^m$, its related error covariance $\boldsymbol{\Gamma}_{k,t}^m$, and the model-conditional likelihood $\phi_{k,t}^m$.

Step 4) Integration of tracking information: By integrating the tracking information exchanged among LiDARs in Step 3, the m-th model-conditional estimate $\hat{\boldsymbol{x}}_{k,t}^m$ and its related covariance $\boldsymbol{P}_{k,t}^m$ at time t are given by

$$\left.\begin{aligned} \hat{\boldsymbol{x}}_{k,t}^m &= \boldsymbol{P}_{k,t}^m \{\sum_{l \in N_k} \alpha_{lk,t}^m (\boldsymbol{\Gamma}_{k,t}^m)^{-1} \boldsymbol{\gamma}_{k,t}^m\} \\ (\boldsymbol{P}_{k,t}^m)^{-1} &= \sum_{l \in N_k} \alpha_{lk,t}^m (\boldsymbol{\Gamma}_{k,t}^m)^{-1} \end{aligned}\right\} \tag{16}$$

where the weight $\alpha_{lk,t}^m$ is set so that the smaller the state estimation error covariance $\boldsymbol{\Gamma}_{l,t}^m$ is, the larger the weight:

$$\alpha_{lk,t}^m = \begin{cases} \dfrac{\dfrac{1}{\mathrm{Tr}(\boldsymbol{\Gamma}_{k,t}^m)}}{\displaystyle\sum_{l \in N_k} \frac{1}{\mathrm{Tr}(\boldsymbol{\Gamma}_{l,t}^m)}} & \text{for } l \in N_k \\ 0 & \text{for } l \notin N_k \end{cases} \tag{17}$$

Step 5) Update of mode probability: Based on the likelihood $\phi_{k,t}^m$ exchanged among LiDARs in step 3, the likelihood function of the m-th mode, $\Lambda_{k,t}^m$, is integrated by

$$\log \Lambda_{k,t}^m = \sum_{l \in N_k} \beta_{lk}^m \log \phi_{k,t}^m \tag{18}$$

The weight β_{lk}^m is given by [14]

$$\beta_{lk}^m = \begin{cases} \dfrac{1}{\max\left(\left|N_l\right|, \left|N_k\right|\right)} & \text{for } l \in N_k, l \neq k \\ 1 - \displaystyle\sum_{l \in N_k, l \neq k} \beta_{lk}^m & \text{for } l = k \\ 0 & \text{for } l \notin N_k \end{cases} \tag{19}$$

where $\left|N_l\right|$ and $\left|N_k\right|$ are the dimensions of N_l and N_k, respectively.

The mode probability is therefore calculated as follows:

$$\hat{\mu}_{k,t}^m = \frac{\hat{\mu}_{k,t/t-1}^m \Lambda_t^m}{\displaystyle\sum_{m=1}^{3} \hat{\mu}_{k,t/t-1}^m \Lambda_{k,t}^m} \tag{20}$$

Using the mode probability, we recognize the motion mode that occurs.

Step 6) Calculation of state estimates: The state estimate and its related error covariances of tracked people are given by

$$\left.\begin{aligned} \hat{\boldsymbol{x}}_{k,t} &= \sum_{m=1}^{3} \hat{\mu}_{k,t}^m \hat{\boldsymbol{x}}_{k,t}^m \\ \boldsymbol{P}_{k,t} &= \sum_{m=1}^{3} \hat{\mu}_{k,t}^m [\boldsymbol{P}_{k,t}^m + (\hat{\boldsymbol{x}}_{k,t} - \hat{\boldsymbol{x}}_{k,t}^m)(\hat{\boldsymbol{x}}_{k,t} - \hat{\boldsymbol{x}}_{k,t}^m)^T] \end{aligned}\right\} \tag{21}$$

In steps 2 and 4, to accurately track many people, each LiDAR sets a validation region around the predicted position of each tracked person [15]. LiDAR measurements (measurements of person's position) within the validation region, which are obtained from the tracked person, are applied to the track update. In crowded environments, multiple LiDAR measurements are within a validation region, and several validation regions overlap. To achieve reliable data association (matching of tracked people and LiDAR measurements), the global-nearest-neighbor-based data association [16] is exploited. The number of people in the sensing areas of LiDARs changes over time. They often encounter occlusions. To handle such conditions, a rule-based data-handling method, which employs track initiation and termination [15], is implemented.

VI. Simulation Experiments

As shown in Figure 4 (a), three LiDARs are placed in an intersection environment, and 20 people are tracked. People's motions and LiDAR scan data are generated by a simulator (Siemens, Simcenter Prescan). Figure 4 (b) shows the paths taken by 20 people. People that moved along the blue paths walked at 1.5 m/s and then stopped. People that moved along the red paths walked at 1.2 m/s and then stopped. People that moved along the purple paths walked at 1.5 m/s from a stop, ran at 3.0 m/s, and then stopped. As an example, Figure 5 shows the velocity profiles of seven of the 20 people.

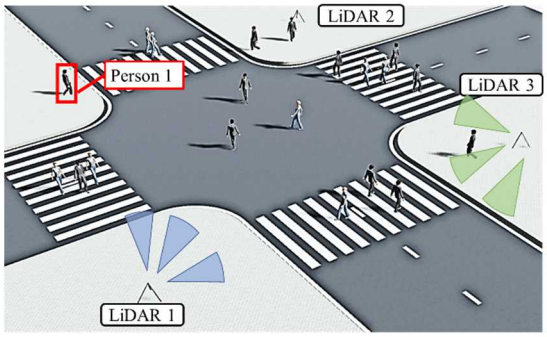
The tracking performance is evaluated for the following four cases.

- Case 1: DIMM-based tracking in a ring network,
- Case 2: DIMM-based tracking in a line network,
- Case 3: CIMM-based tracking,
- Case 4: Distributed Kalman filter (DKF)-based tracking in a ring network.

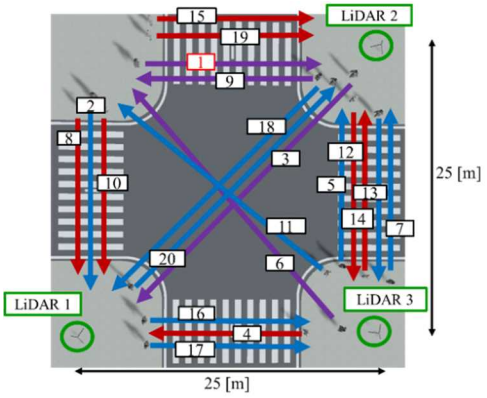
In case 3, the person cells detected by the three LiDARs are collected on a central server, and the central server tracks people using a conventional IMM estimator [6][8]. In case 4, only the constant speed model ((2)) is used as a motion model of a person.

Figure 6 shows the tracking results for person 1 in cases 1 and 2, and Figure 7 shows those in cases 1 and 3. Table I lists the tracking errors for the 20 people. In the table, the result for case 1 shows the following root-mean-squared error J_i for the i-th person (i = 1 to 20):

$$J_i = \sqrt{\frac{1}{N}\sum_{t=0}^{N} (\Delta \hat{x}_{it}^2 + \Delta \hat{y}_{it}^2 + \Delta \hat{v}_{it}^2 + \Delta \hat{\theta}_{it}^2)} \tag{22}$$

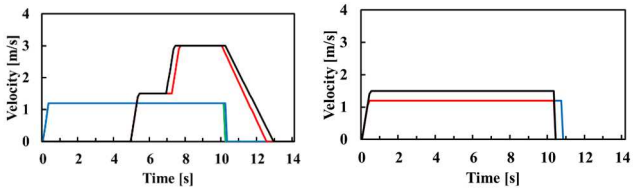


(a) Bird's-eye view



(b) Top view

Figure 4. Simulation environment and paths taken by 20 people.



(a) Persons 1(black), 9 (red), 15 (blue) and 19 (green).

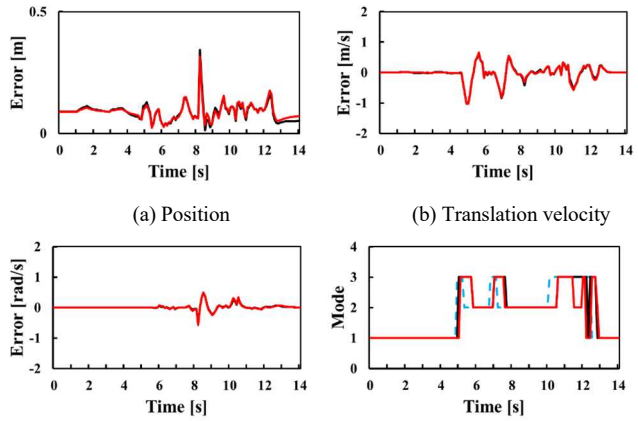
(b) Persons 2 (black), 8 (red), and 10 (blue).

Figure 5. Velocity profile of people. The blue and green lines in (a) significantly overlap, and the red and blue lines in (b) significantly overlap.

where ($\Delta\hat{x}_{it}$, $\Delta\hat{y}_{it}$), $\Delta\hat{v}_{it}$, and $\Delta\hat{\dot{\theta}}_{it}$ are estimate errors in a position, translational velocity, and turn velocity, respectively. N is the tracking duration.

On the other hand, the results for cases 2, 3, and 4 represent the percentage of the tracking error to that in case 1. Thus, the positive sign (+) indicates that the tracking error is larger than that for case 1, whereas the negative sign (-) indicates the opposite.

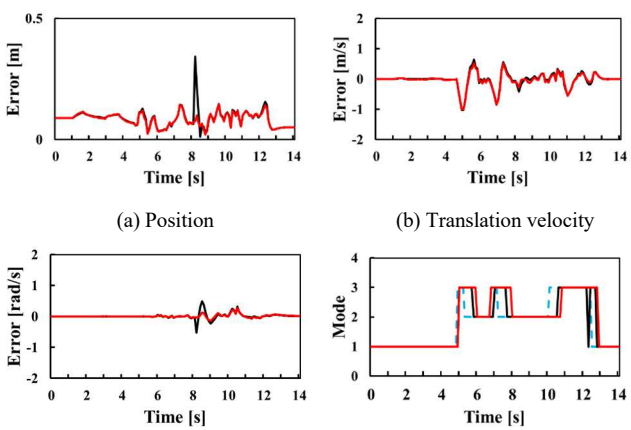
As listed in Table I, the tracking error in case 2 (line network) is approximately 8 % larger than that in case 1 (ring network). This is because in case 1, each LiDAR exchanges detection and tracking information with two LiDARs located at both sides, whereas, in case 2, LiDARs 1 and 3 exchange information only with LiDAR 2. Since the difference in the tracking error between case 1 (DIMM-based tracking) and case 3 (CMM-based tracking) is approximately 1 %, both methods can track people at almost the same degree of accuracy.



(a) Position (b) Translation velocity

(c) Turn velocity (d) Mode estimate

Figure 6. Tracking error in cases 1 (black) and 2 (red). The blue dashed line in (d) indicates the true mode.



(a) Position (b) Translation velocity

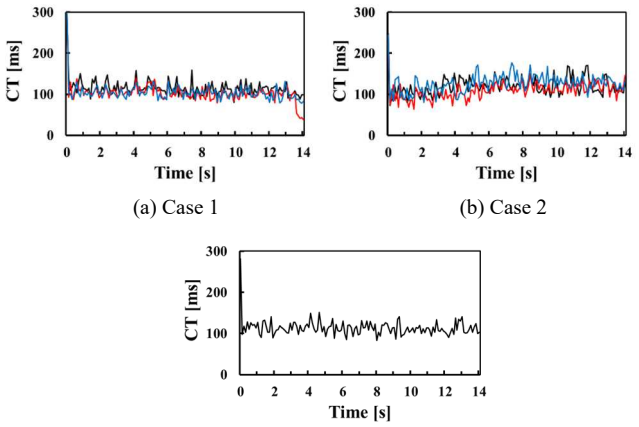
(c) Turn velocity (d) Mode estimate

Figure 7. Tracking error in cases 1 (black) and 3 (red). The blue dashed line in (d) indicates the true mode.

TABLE I TRACKING PERFORMANCE IN CASES 1 – 4

Person #	Case 1	Case 2 [%]	Case 3 [%]	Case 4 [%]
1	0.28	+1.6	-9.0	+77.0
2	0.35	-7.0	-4.7	+37.0
3	0.50	+4.1	+36.3	+48.1
4	0.21	+2.5	+0.4	+22.3
5	0.31	+18.5	-3.8	+3.1
6	0.29	+1.7	-6.1	+110.4
7	0.24	+6.3	-5.3	+16.4
8	0.29	+10.3	-1.9	+55.3
9	0.25	+0.4	+4.7	+521.9
10	0.38	-8.4	-6.3	+38.9
11	0.24	+93.6	-11.7	+31.7
12	0.21	+6.1	-0.7	+45.6
13	0.34	+14.1	-4.1	+14.5
14	0.25	+7.2	-0.2	+11.3
15	0.38	+2.9	0.0	+11.6
16	0.27	-0.2	-2.4	+101.3
17	0.23	+4.1	+0.2	+19.9
18	0.32	-0.6	-12.9	+0.1
19	0.38	-10.7	-7.1	+35.2
20	0.45	+30.8	-8.5	+89.4
Mean	0.31	+8.1	-1.2	+61.2

 ISBN: 978-1-68558-090-2



(a) Case 1 (b) Case 2

(c) Case 3

Figure 8. Computation time. The black, red, and blue lines in (a) and (b) indicate the results of LiDARs 1, 2, and 3, respectively.

TABLE II ROOT MEAN SQUARES OF THE COMPUTATION TIME

	LiDAR 1	LiDAR 2	LiDAR 3
Case 1	114.5 ms	103.9 ms	104.5 ms
Case 2	119.7 ms	109.1 ms	126.5 ms
Case 3	113.4 ms		

The tracking performance in case 4 (DKF-based tracking) is approximately 61% worse than that in case 1 (DIMM-based tracking). This is because, in case 4, a constant-velocity model is employed as the motion model of a person, and the tracking error increases when the person performs a sudden acceleration motion.

We compare the computation time of DIMM-based tracking (cases 1 and 2) and CIMM-based tracking (case 3). The specifications of the computer are Windows 10 Pro OS, Intel(R) Core (TM) i7-8565U@1.80 GHz CPU, 16 GB RAM, and C++ software language. Figure 8 shows the results, and Table II lists the root mean squares of the computation time. The computation time indicates the time required to detect and track people from the LiDAR scan data obtained within a scan. Note that the computation time in case 3 (CIMM-based tracking) is the sum of the computation times in LiDAR 1 and the central server.

The computation time is almost the same for all the cases. Although the computation time should be less than 100 ms of the LiDAR scan period, herein, it is slightly higher than 100 ms. This can be reduced by optimizing the program code and using a graphical processing unit for real-time operations.

VII. CONCLUSION AND FUTURE WORK

This paper presented a cooperative tracking of people using networked LiDARs based on a DIMM estimator. Simulation experiments of tracking of 20 people were conducted using three Velodyne 32-layer LiDARs set in an intersection environment. The tracking performance of the presented method in two different network topologies (ring and line network topologies) was evaluated by comparing the tracking performance of the CIMM and DKF estimators.

DIMM- and CIMM-based tacking showed comparable performance, and the performance of DIMM-based tracking was 61% lower than that of DKF-based tracking. In addition, the computation times for DIMM- and CIMM-based tacking were almost the same.

In our future studies, we will evaluate the presented method through real experiments. In addition, we will employ a machine-learning-based method to improve the performance of people detection in crowded environments.

ACKNOWLEDGMENT

This study was partially supported by the KAKENHI Grant #20H00589, the Japan Society for the Promotion of Science (JSPS).

REFERENCES

[1] A. Brunetti, D. Buongiorno, G. F. Trotta, and V. Bevilacqua, “Computer vision and deep learning techniques for pedestrian detection and tracking: A survey,” Neurocomputing, vol. 300, pp. 17–33, 2018.

[2] E. Marti, J. Perez, M. A. Miguel, and F. Garcia, “A review of sensor technologies for perception in automated driving,” IEEE Intelligent Transportation Systems Magazine, pp. 94–108, 2019.

[3] Md. H. Sharif, “Laser-based algorithms meeting privacy in surveillance: A survey,” IEEE Access, vol. 9, pp. 92394 –92419, 2021.

[4] K. Nakamura, H. Zhao, X. Shao, and R. Shibasaki, “Human sensing in crowd using laser scanners,” Laser Scanner Technology, pp. 15–32, 2012.

[5] T. Wu, J. Hu, L. Ye, and K. Ding, “A pedestrian detection algorithm based on score fusion for multi-LiDAR systems,” Sensors 2021, 21, 1159, 2021.

[6] M. Hashimoto, M. Yuminaka, and K. Takahashi, “Laser-based people tracking system using multiple ground laser scanners,” Proc. of the first IASTED Int. Conf. on Intelligent Systems and Robotics, pp. 95–101, 2016.

[7] E. Mazor, A. Averbuch, Y. Bar-Shalom, and J. Dayan, “Interacting multiple model methods in target tracking: A survey,” IEEE Trans. on Aerospace and Electronic Systems, vol. 34, no. 1, pp. 103–123, 1988.

[8] T. Nakahira, M. Hashimoto, and K. Takahashi, “Cooperative people tracking with multiple ground laser scanners,” Proc. of Int. Symp. on Flexible Automation, 2018.

[9] H. A. P. Blom and Y. Bar-Shalom, “The interacting multiple model algorithm for systems with Markovian switching coefficient,” IEEE Trans. on Automatic Control, vol.33, no.8, pp. 780–783, 1988.

[10] F. F. C. Rego, A. M. Pascoal, P. A. Aguiar, and C. N. Jones, “Distributed state estimation for discrete-time linear time invariant systems: A survey,” Annual Reviews in Control, vol. 48, pp. 36–56, 2019.

[11] C. Y. Chong, K. C. Chang, and S. Mori, “A review of forty years of distributed estimation,” Proc. of 2018 21st Int. Conf. on Information Fusion, 2018.

[12] T. Nakahira, M. Hashimoto, and K. Takahashi, “Cooperative people tracking using multiple ground Lidars based on distributed interacting multimodel estimator,” Proc. of the IEEE 6th 2019 Int. Conf. on Control, Decision, and Information Technologies, 2019.

[13] W. Li and Y. Jia, “Distributed estimation for Markov jump systems via diffusion strategies,” IEEE Trans. on Aerospace and Electronic Systems, vol. 53, pp. 448–460, 2017.

[14] L. Xiao and S. Boyd, “Fast linear iterations for distributed averaging,” Systems & Control Letters, vol. 53, pp. 65–78, 2004.

[15] M. Hashimoto, T. Konda, Z. Bai, and K. Takahashi, “Laser-based tracking of randomly moving people in crowded environments,” Proc. of IEEE Int. Conf. on Automation and Logistics, pp. 31–36, 2010.

[16] H. W. Kuhn, “The Hungarian method for the assignment problem,” Naval Research Logistics Quarterly, vol. 2, no.1–2, pp. 83–98, 1955.

 ISBN: 978-1-68558-090-2