People Detection and Tracking Using Distributed LiDAR Network

Ryota Imai, Marino Matsuba
Graduate School of Science and Engineering
Doshisha University
Kyotanabe, Japan
e-mail: {ctwj0109, ctwj0112 }@mail4.doshisha.ac.jp

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

Abstract— **This paper presents a people detection and tracking method using Light Detection And Ranging sensors (LiDARs) set in an environment. Each LiDAR detects people from its own LiDAR measurements, exchanges information of people positions by communicating with its neighboring LiDARs, and then fuses the information of people positions. Thereafter, each LiDAR estimates people's poses from the information of people's positions, and the estimates are fused by exchanging information among neigboring LiDARs. A one-dimensional convolutional neural network is applied to accurately detect people from LiDAR measurements in an environment where various objects, such as people, two-wheelers, and cars, coexist. A distributed interacting multimodel estimator is applied to accurately estimate the poses of people under various motion modes, such as stopping, walking, and suddenly running and stopping, in a distributed manner without a central server. Simulation results of eight people tracked by four Velodyne 32-layer LiDARs in an intersection environment where people and cars coexist show the performance of the proposed method.**

Keywords- LiDAR network; people detection and tracking; one-dimensional convolutional neural network; distributed interacting multimodel estimator.

I. INTRODUCTION

This paper is an extended and improved version of an earlier paper presented at the IARIA Conference on Sensor Technologies and Applications (SENSORCOMM 2023) [1] in Porto, Portugal.

People tracking (motion estimation, such as position and velocity) is an important technology in various fields, including surveillance, security, and Intelligent Transportation Systems (ITS). Consequently, many related studies have been actively performed using Light Detection And Ranging sensors (LiDARs) and cameras [2]–[5]. This paper focuses on people tracking using sensors set in an environment.

In sparse and not crowded environments, the tracking performance of a single sensor is high; however, in dense environments, tracking performance deteriorates due to occlusions. To reduce occlusions and accurately track people in dense environments, a cooperative people tracking method has been proposed in which data from networked sensors set at different locations are shared [6]–[9]. In addition, Bayesian filters, such as Kalman and particle filters, are commonly used to accurately track people without interrupting the tracking process even in occluded conditions.

Most conventional Bayesian-filter-based people tracking methods tend to perform their tasks under the assumption that people move at a nearly constant velocity, and thus tracking performance can be significantly compromised when people's behaviors suddenly change, such as sudden running and sudden turn.

For example, let us consider that in an intersection environment, tracking of people crossing a crosswalk is performed by sensors set on signal lights. In principle, people's motions change according to the signal light conditions, i.e., stop at red lights, walk or run at a nearly constant speed at green lights, and suddenly run at yellow lights. Therefore, it is necessary to develop tracking systems in ITS domains that can accurately track people even in such rapidly changing behaviors.

Multimodel methods, including the Interacting MultiModel (IMM) method [10], are well known for accurately tracking objects exhibiting various behaviors [11]. We previously presented an IMM-based people tracking method [12] using ground LiDAR, and we extended this method to cooperative people tracking using multiple ground LiDARs [13][14].

Most conventional methods for cooperative people tracking, including the one presented in our previous study, are based on a centralized fusion method in which data from multiple sensors, such as LiDARs and cameras, are collected and fused on a central server. As a result, central server failure will unavoidably lead to malfunctions to the entire tracking system. To address this problem and also maintain system robustness even in cases of central server failure, this paper presents a cooperative people tracking method using distributed data fusion (Distributed Interacting MultiModel (DIMM)-based method [15]), in which data among LiDARs are processed without the requirement of a central server.

As a preprocessing people tracking step, it is necessary to accurately detect people from the entire sensor measurements. A simple method for detecting people is based on a background subtraction method [12], in which people are detected by subtracting an environmental map (sensor measurements obtained in advance) from current sensor measurements. However, this approach exhibits a tendency to misidentify objects, such as cars and two-wheelers, that do not exist in the environmental map as people.

To accurately distinguish people from objects in various environments, many studies have been conducted using machine learning methods [16][17]. In this paper, a One-Dimensional Convolutional Neural Network (1D-CNN) method for people detection [18] is implemented in our LiDAR-based cooperative people tracking system.

The performance of cooperative people tracking using a DIMM-based estimator in conjunction with 1D-CNN is

quantitatively evaluated through simulation experiments in an intersection environment where people and cars move simultaneously. The rest of this paper is organized as follows. In Section II, an overview of related works is presented . In Section III, the experimental system is described. In Section IV, people detection and tracking methods are discussed and then described in Sections V and VI, respectively. In Section VII, simulation experiments are conducted to evaluate the performance of the proposed method, followed by our conclusions in Section VIII.

II. Related Work

Compared with centralized data fusion, distributed data fusion enhances system robustness and scalability. Therefore, in the field of Bayesian filtering, distributed data fusion methods have been actively discussed [19][20].

Distributed data fusion methods are classified into consensus and diffusion strategies. In the consensus strategy, sensor nodes iteratively exchange data to reach a consensus. Therefore, this method requires high computation and communication costs due to iterations. In contrast, consensus iteration is not required in the diffusion strategy. For this reason, we are interested in the application of diffusion strategies to LiDAR-based cooperative people tracking. To the best of our knowledge, no studies have been conducted on LiDAR-based cooperative tracking of people moving in various behaviors using distributed fusion methods.

Thus, in our previous study [21], cooperative people tracking was presented using a DIMM-based estimator of a distributed data fusion method [15]. However, cooperative people tracking was performed for only two linked LiDARs, rendering the effectiveness of the tracking method using three or more LiDARs in various network topologies, such as ring and line network topologies, unclear. In this paper, a DIMM-based cooperative people tracking method is presented using four LiDARs in ring and line network topologies.

In addition to tracking algorithms, many studies have used machine learning methods, such as PointNet, VoxelNet, Pointpillars, and CenterPoint, to accurately detect people and objects in various environments [19][20].

In the field of ITS, mechanical 3D LiDARs, which spin laser beams in the horizontal direction to achieve a 360 degree horizontal field of view, are commonly used because of their higher accuracy and reliability compares to those of solid-state 3D LiDARs [22].

Kunisada et al. [18] presented a people detection method for mechanical LiDAR based on 1D-CNN [23]. The 1D-CNN-based method considers all measurements captured from mechanical LiDAR as 1D waveform data to perform a convolution processing. As a result, the time from obtaining LiDAR measurements to the output of detection results has a shorter delay than other machine learning-based methods. For this reason, our study implements a 1D-CNN-based method to accurately detect people.

The contributions of this paper are as follows:

- A DIMM-based cooperative people tracking method using four LiDARs is designed in typical network topologies (ring and line network topologies). The tracking method can then be applicable to many LiDAR systems operating in any network topology.
- The performance of DIMM-based cooperative people tracking in conjunction with 1D-CNN-based people detection is quantitatively evaluated through simulation experiments in an intersection environment where people and cars coexist.

III. Experimental System

In this paper, a system with four networked LiDARs is considered, as shown in Figure 1. Each LiDAR consists of a mechanical LiDAR (Velodyne HDL-32E) and a computer. The LiDAR emits 32 laser beams in the vertical direction, and its maximum range is 50 m. The horizontal viewing angle is 360° with a resolution of 0.16°, and the vertical viewing angle is 41.3° with a resolution of 1.33°. The spinning period is 0.1 s. Approximately 70,000 measurements are acquired during the spinning period.

Two network topologies are considered for exchanging information among LiDARs: a ring network topology (referred to as a ring network) and a line network topology (referred to as a line network). As shown in Figure 1, each LiDAR is connected to two other adjacent LiDARs in a ring network, while LiDARs 1 and 2, LiDARs 2 and 3, and LiDARs 1 and 4 are connected in a line network.

In the case of three or more than five LiDARs, as in the case of four LiDARs, each LiDAR can be connected to two other LiDARs on both sides in a ring network. In a line network, only the LiDARs at both ends of the line can be connected to one adjacent LiDAR, while the other LiDARs can be connected to the two LiDARs.

IV. Overview of People Detection and Tracking

Figure 2 shows the flow of people detection and tracking method. Each LiDAR captures its own measurements and detects people using a 1D-CNN-based method, which, however, requires a high density of LiDAR measurements to accurately recognize people. Our preliminary experiments revealed that the maximum range of people detection could be set to 25 m from the LiDAR, and that the 1D-CNN-based method often failed to detect people located within 1 m from the LiDAR (a reason will be discussed later on). As a result, in the range of 1–25 m from the LiDAR, people can be accurately detected by the 1D-CNN-based method.

Each LiDAR communicates with its neighboring LiDARs and exchanges information on people's positions, including the time stamp, number, and coordinates of their positions. Thereafter, each LiDAR fuses the information.

Based on the DIMM estimator, each LiDAR estimates the positions, moving directions, velocities, and behaviors of people from information regarding their respective positions.

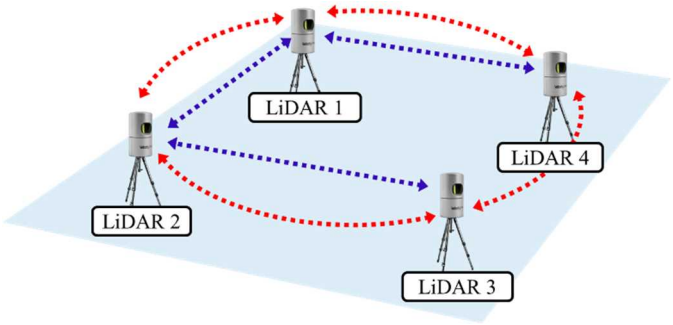


Figure 1. Overview of the network of four LiDARs. The red and blue dotted lines indicate the ring and line network topologies, respectively.

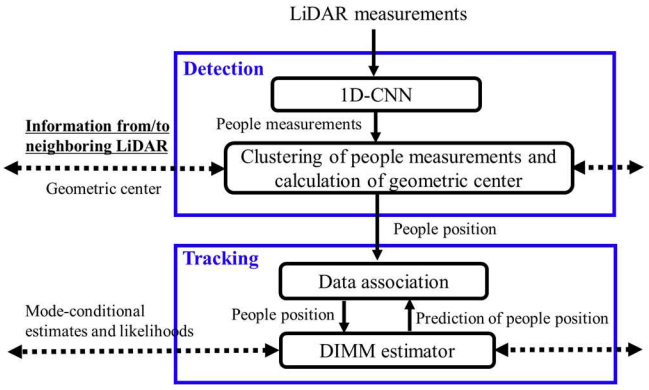


Figure 2. Flow of people detection and tracking.

Their estimates are exchanged among neighboring LiDARs and fused. In this study, three motion modes are considered as people behaviors: stopping, walking/running at an almost constant velocity, and significant acceleration/deceleration, such as suddenly running or suddenly stopping.

V. People Detection Method

Our LiDAR has 32 laser beams in the vertical direction, which are designated as laser IDs 1–32. Because the 32 laser beams are rotated in the horizontal direction, as shown in Figure 3, the distance measurements obtained from the LiDAR are regarded as 1D waveform data for each laser ID.

Outliers (false measurements) of LiDAR obtained from mirror objects or the sky significantly degrade the performance of detection using 1D-CNN. Such outliers are corrected as follows [18]: Outliers captured from a laser beam with a vertical viewing angle of 0° or greater are assigned large values. For outliers captured by a laser beam with a vertical viewing angle of less than 0°, distance measurements to the ground are estimated and set.

Figure 4 shows the structure of people detection using 1D-CNN, which consists of three convolution layers (green and orange blocks) and a fully connected layer (yellow block) [24]. In all convolution layers, the convolution process is performed by moving the 1D convolution filter only in the horizontal direction, and people are then recognized from the feature map for every laser ID.

The 1D waveform data to the input layer are distance measurements obtained by horizontally spinning the 32 laser beams using a 7° window (LiDAR measurements of 32 × 41). This window is moved simultaneously with the LiDAR spinning process, allowing us to identify whether or not a person is present within one scan for the LiDAR measurements.

In the first convolution layer, 32 convolution filters with a size of 1 × 4 are composed, and max spooling with a size of 2 × 2 is performed to obtain feature maps. In the second convolution layer, 32 convolution filters with a size 1 × 3 are convolved with the feature maps obtained in the first layer to obtain new feature maps. In the third convolution layer, 32 convolution filters with a size 1 × 3 are convolved with the feature maps obtained in the second layer to obtain new feature maps. Here, Rectified Linear unit (ReLu) is used as the activation function for all three convolution layers.

The feature maps obtained in the third convolution layer are transformed to feature vectors, which are then provided to a fully connected layer with 1024 units, where the drop-out rate is set to 0.5. The output layer with two units then determines whether or not the LiDAR measurement belongs to a person using the softmax function.

The LiDAR measurements judged to belong to a person are clustered, as shown in Figure 5, and the geometric center of the clustered measurements is obtained. The position of the

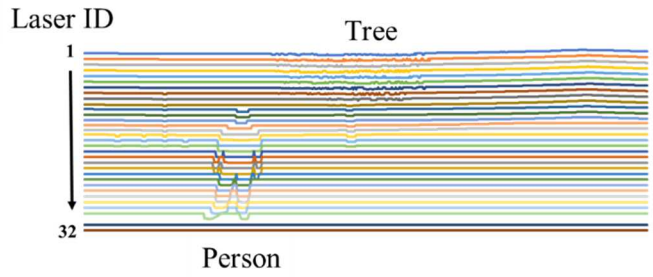


Figure 3. 1D waveform data from LiDAR. Different colored lines indicate different laser IDs.

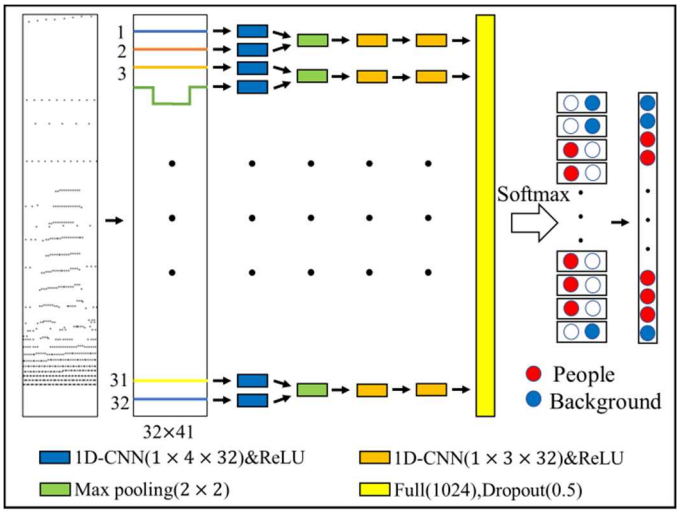
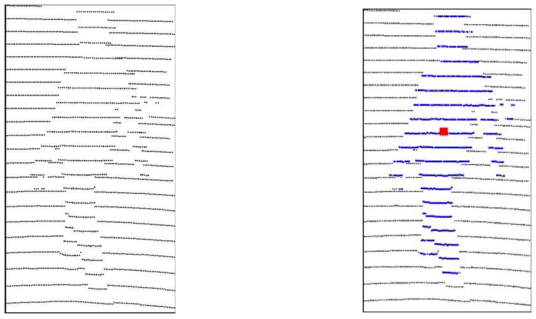


Figure 4. Structure of 1D-CNN.



(a) LiDAR measurements (b) Detection result

Figure 5. Person detection using 1D-CNN. In (b), the blue dots indicate the measurements determined to be taken by a person. The red mark indicates the geometric center of the person (GCP).

geometric center is used as the person's position in people tracking. To accurately cluster LiDAR measurements related to people in the vicinity, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) method [25] is used.

When a person is within 1 m from the LiDAR, detection fails because the LiDAR measurements related to a person do not always fall within the 7° window. Therefore, people are detected whithin 1–25 m from the LiDAR.

VI. People Tracking Method

In this section, people's motions in an intersection environment are firstly modeled. Thereafter, a people tracker is designed based on DIMM estimator in conjunction with the Global-Nearest-Neighbor (GNN)-based data association.

A. Motion Model of People

To accurately estimate people's motions and behaviors in an intersection environment, the following motion modes of a person are used:

a) Stop mode (mode 1): Mode for a person stopping at a red light.

b) Constant velocity mode (mode 2): Mode for a person walking or running at an almost constant translational or rotational velocity under a green light.

c) Sudden motion mode (mode 3): Mode for a person who suddenly runs or stops when a green light turns into a yellow light.

As shown in Figure 6, the position and moving direction of the person are denoted by (x, y) and θ, respectively, in the world coordinate system. The translational and rotational velocities of the person are denoted by v and ω, respectively. The three motion modes are then modeled using the following state equations:

- Mode 1

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

- Mode 2

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

- Mode 3

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

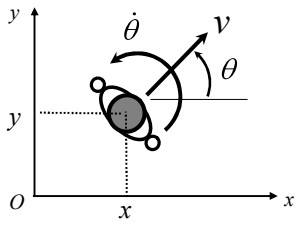


Figure 6. Person state (top view).

where t and t–1 indicate time steps. $\dot{v}$ and $\dot{\omega}$ are the translational and rotational acceleration of a person, respectively. Furthermore, $\Delta\dot{y}$, $\Delta\dot{v}$, $\Delta\ddot{v}$, $\Delta\omega$, and $\Delta\dot{\omega}$ represent the plant disturbances. τ (= 100 ms) is the spinning period of the LiDAR.

The state equation of the m-th mode, where 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 $\boldsymbol{x}^m$ indicates the state vector of the m-th mode, and $\Delta\boldsymbol{v}^m$ indicates the disturbance vector, which is assumed to have a white noise sequence with the covariance matrix $\boldsymbol{Q}^m$.

The LiDAR measurement relating to the person gives the following measurement 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, more specifically, the position of the Geometric Center of the Person (GCP) obtained by the people detection method, and $\Delta\boldsymbol{z}$ 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 given by

$$\boldsymbol{H}^1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \boldsymbol{H}^2 = \boldsymbol{H}^3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

B. People Tracking using DIMM Estimator

The main notations used in this subsection are listed in Appendix A.

It is assumed that any change in the three motion modes is governed by the first-order homogeneous Markov chain as follows:

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

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

where π_{t-1}^m and π_t^n are events that the m-th and n-th modes, where 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 $T_{mn} = 0.05$ for $m \neq n$ (see Figure 7).

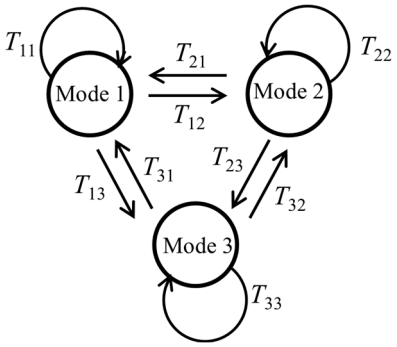


Figure 7. Mode transition.

The k-th LiDAR (k = 1–4) estimates people's states in the following five steps:

Step 1) Filter initialization

The probability that the m-th mode occurs at time t–1 is denoted by $\hat{\mu}^m_{k,t-1}$. The m-th mode conditional estimate and its related covariance are denoted by $\hat{\boldsymbol{x}}^m_{k,t-1}$ and $\boldsymbol{P}^m_{k,t-1}$, respectively. These three quantities are mixed as follows:

$$c_{k,mn} = \frac{T_{mn}\hat{\mu}^m_{k,t-1}}{\sum_{m=1}^{3} T_{mn}\hat{\mu}^m_{k,t-1}} \tag{8}$$

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

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

Step 2) Calculation of state estimate and likelihood

The single-model-based Kalman filters for the three modes run, and the prediction and related covariance for each mode at time t are given by

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

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

The people's positions $\boldsymbol{z}_l$ from neighboring LiDARs are combined, and the quantities related to the measurement $\boldsymbol{q}_{k,t}$ and its error covariance $\boldsymbol{S}_{k,t}$ are given as follows:

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

where N_k is the set of neighboring LiDARs, including itself (i.e., k-th LiDAR), given in Table I.

From the quantities in Eqs. (11) and (12), the information filter determines the state estimate $\boldsymbol{\gamma}^m_{k,t}$ and its related error covariance $\boldsymbol{\Gamma}^m_{k,t}$ at time t as follows:

TABLE I. N_k RELATED TO NETWORK TOPOLOGY

k	Ring network	Line network
LiDAR 1: N_1	{1, 2, 4}	{1, 2, 4}
LiDAR 2: N_2	{1, 2, 3}	{1, 2, 3}
LiDAR 3: N_3	{2, 3, 4}	{2, 3}
LiDAR 4: N_4	{1, 3, 4}	{1, 4}

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

In addition, the mode conditional likelihood is calculated by

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

where $\tilde{\boldsymbol{z}}^m_{k,t/t-1}$ and $\boldsymbol{L}^m_{k,t/t-1}$ indicate the predicted measurement error and its associated covariance, respectively, as follows:

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

In a crowded environment where many people coexist, a data association (one-to-one matching of tracked people and positions of GCPs) is required to ensure the accurate process of people tracking. For the data association, a GNN-based method is used (see Appendix B).

Step 3) Exchange of tracking information and likelihood

Each LiDAR communicates with its neighboring LiDARs and exchanges information about the state estimate $\boldsymbol{\gamma}^m_{k,t}$, its related error covariance $\boldsymbol{\Gamma}^m_{k,t}$, and mode conditional likelihood $\phi^m_{k,t}$.

Step 4) Combination of tracking information

By fusing the tracking information exchanged among the LiDARs in Step 3, the m-th mode conditional estimate $\hat{\boldsymbol{x}}^m_{k,t}$ and related covariance $\boldsymbol{P}^m_{k,t}$ at time t are given by

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

where the weight $\alpha^m_{lk,t}$ is set so that the smaller the error covariance $\boldsymbol{\Gamma}^m_{l,t}$ of the state estimate is, the larger the weight as follows:

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

When fusing the tracking information, it is necessary to match the tracking information calculated by Eq. (9) with that obtained from other LiDARs. Matching is performed based on the GNN method by setting a validation region with a certain radius around the tracking information, while treating the tracking information from other LiDARs as measurements. If the tracking information from other LiDARs cannot be matched with individual tracking information, it is assumed that this information is related to person(s) outside the sensing area of the individual LiDAR, which it is subsequently used as it is.

Step 5) Calculation of mode probability and state estimate

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 fused by

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

The weight β_{lk}^m is given by [26]

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

where $|N_l|$ and $|N_k|$ are the dimensions of N_l and N_k, respectively. From Table I, because $|N_1| = |N_2| = |N_3| = |N_4| = 3$ in the ring network, and $|N_1| = |N_4| = 2$ and $|N_2| = |N_3| = 3$ in the line network, the weight β_{lk}^m is set to 0.33 and to values shown in Table II in the ring and line networks, respectively.

The mode probability is therefore calculated as follows:

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

People behavior can be recognized as the mode in which the value of mode probability is maximized.

The state estimate and its related error covariance for tracked people are finally given by

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

The number of people in the sensing areas of LiDARs changes over time, as people continuously enter and leave the sensing area. In addition, people often encounter occlusions in the sensing area. To handle such conditions, a rule-based data-handling method that employs the following track initiation and termination [12] is implemented.

TABLE II. β_{lk}^m RELATED TO THE LINE NETWORK

	$l = 1$	$l = 2$	$l = 3$	$l = 4$
$k = 1$	0.33	0.33	0	0.33
$k = 2$	0.33	0.33	0.33	0
$k = 3$	0	0.33	0.67	0
$k = 4$	0.33	0	0	0.67

a) Track initiation: If a person measurement (position of GCP) cannot be matched with the person being tracked by GNN-based data association, it is assumed that the measurement comes from a new person. As a result, tracking begins. However, the measurement may be an outlier. In that case, it is unlikely that measurements will be obtained continuously. Therefore, if the measurement is not obtained within a threshold of 0.2 s (set in this study) after the tracking start, then the tracking process is terminated; otherwise, the measurement is considered to represent a new person, and tracking is continued.

b) Track termination: If measurements (positions of GCPs) cannot be obtained to associate them with the person being tracked, person tracking is continued using the state prediction (Eq. (11)), and if no measurements are detected after a threshold of 1 s (set in this study), tracking is terminated.

VII. SIMULATION EXPERIMENTS

In this section, the performance of the people detection and tracking method is evaluated through simulation experiments in which people and cars move at the same time. First, the experimental conditions are described, and our results are subsequently shown.

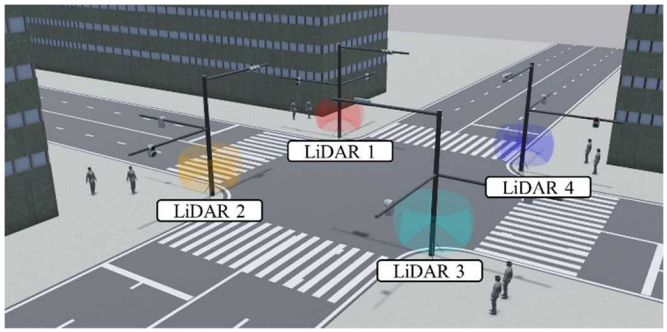
A. Experimental Condition

Simulation experiments in an intersection environment are conducted to evaluate the proposed methods. To generate the motions of both cars and people, and thus the related LiDAR measurements, the Simcenter Prescan (Siemens) [27] is used as a simulator.

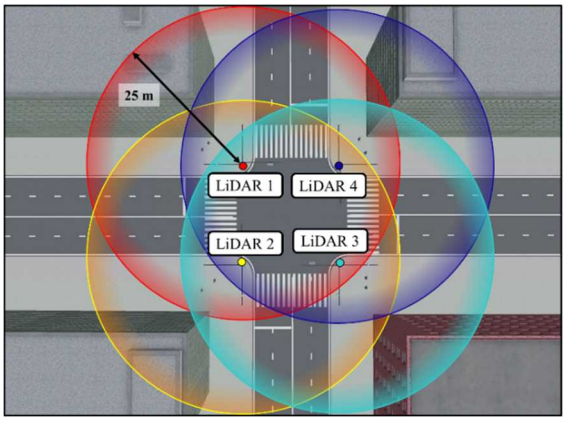
As shown in Figure 8 (a), four LiDARs are set at a height of 1.95 m on signal light posts that are 15 m apart from one another in an intersection environment. The sensing areas of the four LiDARs are shown in Figure 8 (b). Eight people move on crosswalks, and Figures 9 (a) and (b) show the movement paths and motion profiles of these eight people, respectively.

As shown in Figure 10, cars also move in some of our environments. More specifically, there are no cars in environment 1. In environment 2, cars move on two inside lanes at 40 km/h, and their time-headway is 3 s. In environment 3, cars move on four lanes at 40 km/h, and the time-headways in the inside and overtaking lanes are 1.5 s and 3 s, respectively. In environment 4, cars move on four lanes at 40 km/h, and the time-headway is 1.5 s. In our simulation, the tracking duration of eight people is 15 s, in which duration, the total number of cars moving in environments 2, 3, and 4 is 12, 34, and 45, respectively.

Figure 11 shows the people and cars used in the experiments. A person (a-1) and cars (b-1 through 4) are used to generate the

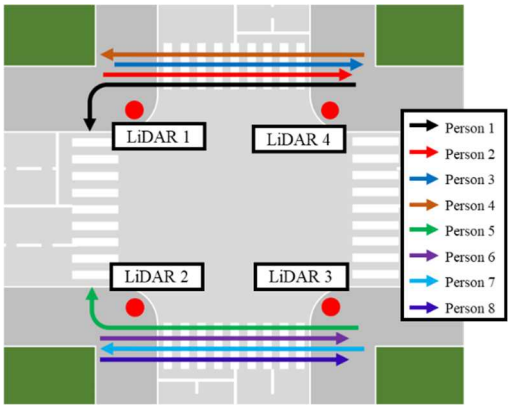


(a) Intersection environment (bird's-eye view)

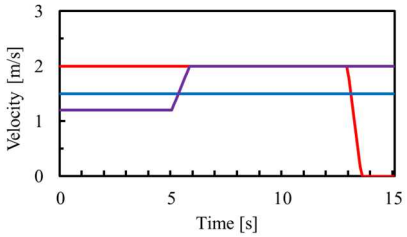


(b) Sensing area of the LiDARs (top view)

Figure 8. Simulation environment.

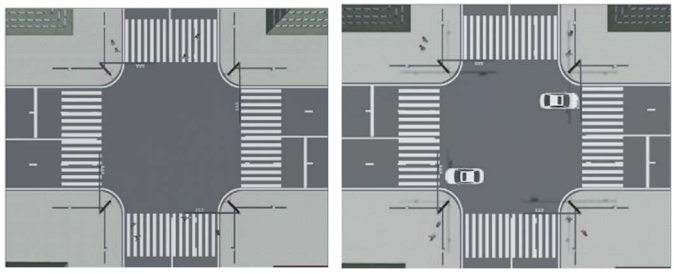


(a) Movement path



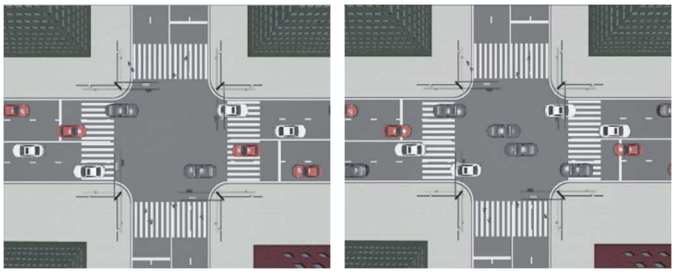
(b) Velocity profile. Persons 1, 5, and 8 (red), persons 2, 4, and 6 (blue), and persons 3 and 7 (purple).

Figure 9. Movement path and velocity profile of the eight people.



(a) Environment 1 (No car)

(b) Environment 2 (Small number of cars)



(c) Environment 3 (Middle number of cars)

(d) Environment 4 (Large number of cars)

Figure 10. Simulation environment (top view). The yellow arrow indicates the movement path and direction of the car.

a-1 a-2 a-3 a-4

(a) People

b-1 b-2 b-3 b-4 b-5

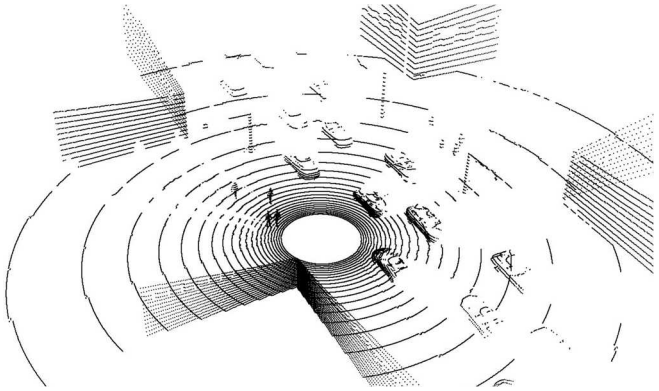
(b) Car

Figure 11. People and cars moving in the environment.

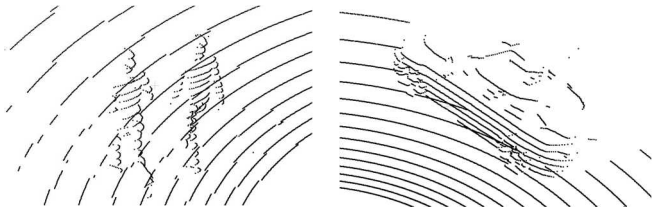
test data set, whereas people (a-2 through 4) and a car (b-5) are used to generate the training data set for 1D-CNN-based people detection. Figure 12 illustrates some examples of the LiDAR measurements generated by the simulator. It is clear from these figures that the simulator can generate good LiDAR measurements.

B. *Evaluation of People Detection*

The dataset used to train the 1D-CNN consists of 8808 background data and 22764 people data. People data are



(a) Environment



(b) People (c) Car

Figure 12. Example of LiDAR measurements generated from Simcenter Prescan (bird's-eye view).

generated from three groups of people (a-2 through 4) in Figure 11(a). Background data are generated from the environment and car (b-5) shown in Figure 11 (b). The 1D-CNN is trained using a mini-batch learning method with a batch size of 20, in which Nadam and binary cross entropy are used as the optimizer and error function, respectively.

To evaluate the performance of people detection, precision, recall, and Intersection over Union (IoU) are obtained. Its accuracy is not evaluated because the difference in the number of LiDAR measurements for people and backgrounds is numerous. Higher precision indicates fewer misdetections, while higher recall indicates fewer undetections. The larger the IoU value, the higher the detection accuracy for all LiDAR measurements.

The results of people detection are shown in Table III. Compared to environment 1, where there are no cars, both precision and IoU decrease with increasing car congestion. This means that in this case car-related LiDAR measurements are often misdetected as person-related measurements. However, since recall is independent of the respective environment, the degree of car congestion would not affect the performance of people detection.

C. Evaluation of People Tracking

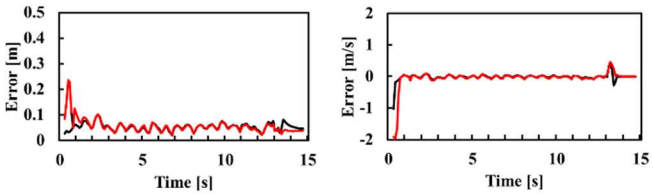
The tracking performance is evaluated for the following three cases.

- Case 1: DIMM-based tracking in a ring network
- Case 2: DIMM-based tracking in a line network
- Case 3: Centralized IMM (CIMM)-based tracking.

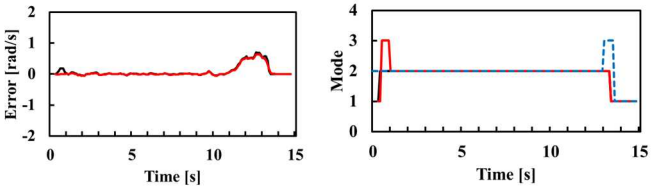
In case 3, LiDAR measurements of people (positions of GCPs) detected by the four LiDARs are collected on a central server, where people are then tracked using a conventional IMM estimator [10][12].

TABLE III. PERFORMANCE OF PEOPLE DETECTION (%)

	Precision	Recall	IoU
Environment 1	92.57	85.58	80.10
Environment 2	83.77	85.58	73.34
Environment 3	73.12	85.57	64.98
Environment 4	70.13	85.60	62.67

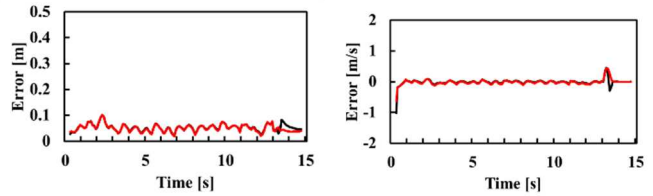


(a) Error of position (b) Error of translation velocity

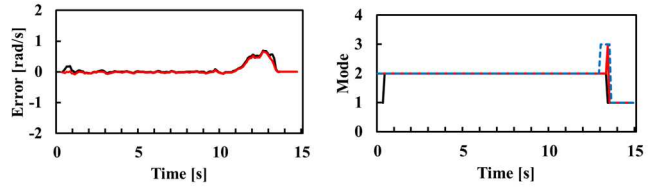


(c) Error of rotational velocity (d) Mode estimate

Figure 13. Tracking error and mode estimate of person 1 in cases 1 (black) and 2 (red). The blue dashed line in (d) indicates the true mode.



(a) Error of position (b) Error of translation velocity



(c) Error of rotational velocity (d) Mode estimate

Figure 14. Tracking error and mode estimate of person 1 in cases 1 (black) and 3 (red). The blue dashed line in (d) indicates the true mode.

Figure 13 shows the tracking results for person 1 in cases 1 and 2 in environment 1, while Figure 14 shows the tracking results for person 1 in cases 1 and 3 in environment 1. Table IV lists the tracking errors for the eight people, which are defined by the following root-mean-squared (RMS) error:

$$J = \sqrt{\frac{1}{8N}\sum_{i=1}^{8}\sum_{t=0}^{N}(\Delta\hat{x}_{it}^2 + \Delta\hat{y}_{it}^2 + \Delta\hat{v}_{it}^2 + \Delta\hat{\theta}_{it}^2)} \qquad (22)$$

where ($\Delta\hat{x}_{it}$, $\Delta\hat{y}_{it}$), $\Delta\hat{v}_{it}$, and $\Delta\hat{\dot{\theta}}_{it}$ are the estimate errors in the position, translational velocity, and rotational velocity, respectively, of the *i*-th person. *N* is the tracking duration.

Note that since the tracking error using the DIMM estimator is slightly different for each LiDAR, cases 1 and 2 in Table IV show the average values of the tracking errors from the four LiDARs. In addition, the ratios of tracking errors in cases 1 and 2 to that in case 3 are shown in brackets in Table IV.

Table IV shows that the tracking error in case 2 (line network) is larger than that in case 1 (ring network). This is because each LiDAR exchanges detection and tracking information with two neighboring LiDARs in case 1, whereas, in case 2, LiDARs 3 and 4 exchange information only with one LiDAR, as shown in Figure 1.

Tables V and VI show the number of falsely tracked cars and untracked people, respectively. As shown in Table V, cars are often tracked as people. However, measurements related to cars were not obtained continuously; thus, the track handing method described in subsection 5.2 terminates the false tracking within 0.2 s (threshold in track initialization). As shown in Table VI, a person is untracked only in case 2 in environment 2. Initially, persons 2 and 3, which were close to each other were tracked as a single person for the first five seconds; subsequently, they were correctly tracked as two people. This is why a person was untracked in case 2 in environment 2. Furthermore, as shown in Table IV, a large tracking error occurs because a person is untracked in the first five seconds in case 2 in environment 2.

TABLE IV. RMS VALUE AND RATIO OF THE TRACKING ERROR

	Case 1	Case 2	Case 3
Environment 1	0.142 (3.65 %)	0.158 (15.26 %)	0.137
Environment 2	0.142 (3.58 %)	0.183 (33.75 %)	0.137
Environment 3	0.142 (0.28 %)	0.164 (16.78 %)	0.142
Environment 4	0.143 (2.80 %)	0.161 (15.57 %)	0.139

TABLE V. THE NUMBER OF FALSELY TRACKED CARS

	Case 1	Case 2	Case 3
Environment 1	0	0	0
Environment 2	9	7	6
Environment 3	11	12	10
Environment 4	11	11	10

TABLE VI. THE NUMBER OF UNTRACKED PEOPLE

	Case 1	Case 2	Case 3
Environment 1	0	0	0
Environment 2	0	1	0
Environment 3	0	0	0
Environment 4	0	0	0

VIII. CONCLUSION AND FUTURE WORK

This paper presented a people detection and tracking method using a distributed LiDAR network. People were detected using the 1D-CNN-based method, and the detected people were tracked using the DIMM-based estimator. Simulation experiments of tracking eight people were conducted using four LiDARs allocated in an intersection environment. The performance of people detection was evaluated in an intersection environment in which people and cars moved simultaneously. The tracking performance was also evaluated in two different network topologies, namely ring and line networks, and compared with the tracking performance of the conventional CIMM-based estimator.

A mechanical 32-layer LiDAR (Velodyne HDL-32E) was used in this paper. In future works, we will evaluate the proposed method through testbed/real-world experiments using the Velodyne HDL-32E LiDAR.

In this paper, a small-scale system (eight people and four LiDARs) was evaluated through simulation experiments. To better validate the performance of the proposed method, we plan to evaluate it on large-scale systems. In addition, we will compare the proposed method with other state-of-the-art methods for people detection and tracking.

APPENDIX A: NOTATION

$\boldsymbol{T}_{mn}$: transition probability matrix that the *m*-th mode jumps into the *n*-th mode

$\boldsymbol{x}_{t-1}^{m}$: true state of the *m*-th mode at time *t*–1

$\hat{\mu}_{k,t-1}^{m}$: *m*-th mode probability estimate obtained by the *k*-th LiDAR at time *t*–1

$\hat{\boldsymbol{x}}_{k,t-1}^{m}$, $\boldsymbol{P}_{k,t-1}^{m}$: *m*-th mode conditional estimate of $\boldsymbol{x}_{t-1}^{m}$ and its estimation error covariance obtained by the *k*-th LiDAR at time *t*–1

$c_{k,mn}$: mixing probability obtained by the *k*-th LiDAR

$\bar{\boldsymbol{x}}_{k,t-1}^{m}$, $\bar{\boldsymbol{P}}_{k,t-1}^{m}$: mixing state estimate and its estimation error covariance of the *m*-th mode obtained by the *k*-th LiDAR at time *t*–1

$\hat{\boldsymbol{x}}_{k,t/t-1}^{m}$, $\boldsymbol{P}_{k,t/t-1}^{m}$: *m*-th mode conditional prediction of $\boldsymbol{x}_{t}^{m}$ and its prediction error covariance obtained by the *k*-th LiDAR at time *t*

$\boldsymbol{\gamma}_{k,t}^{m}$, $\boldsymbol{\Gamma}_{k,t}^{m}$: updated conditional estimate and its estimation error covariance of the *m*-th mode obtained by the *k*-th LiDAR at time *t*

$\hat{\boldsymbol{x}}_{k,t}^{m}$, $\boldsymbol{P}_{k,t}^{m}$: combined conditional estimate of $\boldsymbol{x}_{t}^{m}$ and its estimation error covariance of the *m*-th mode obtained by the *k*-th LiDAR at time *t*

$\hat{\mu}_{k,t}^{m}$: updated probability of the *m*-th mode obtained by the *k*-th LiDAR at time *t*

$\phi_{k,t}^{m}$: *m*-th mode conditional likelihood obtained by the *k*-th LiDAR at time *t*

$\Lambda_{k,t}^{m}$: combined likelihood of the *m*-th mode conditional likelihood obtained by the *k*-th LiDAR at time *t*

$\alpha_{lk,t}^{m}$: nonnegative scalar weight for combination of the *m*-th mode conditional estimate and its estimation error at time *t*

β_{lk}^{m} : nonnegative scalar weight for combination of the *m*-th mode conditional likelihood at time *t*

APPENDIX B: DATA ASSOCIATION

To briefly explain data association, the case of tracking two people is considered. As shown in Figure A(a), a validation region (black dashed line) is set around the predicted position of the tracked person (red circle). A measurement (GCP, black triangle) obtained within the validation region is considered to originate from person 1 being tracked, and the state of the tracked person 1 is updated using this measurement. In contrast, the GCP obtained outside the validation region is considered to originate from another person, and person 2 is tracked using the prediction.

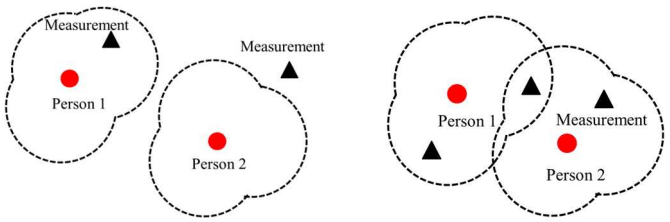
The validation region is set as follows: Based on the *m*-th mode (m = 1, 2, 3) of the person, the position of the tracked person is predicted by

$$\hat{\boldsymbol{p}}_t^m = \boldsymbol{H}^m \hat{\boldsymbol{x}}_{k,t/t-1}^m \quad \text{(A)}$$

where $\hat{\boldsymbol{x}}_{k,t/t-1}^m$ is the prediction of the *m*-th mode shown in Eq. (11), and $\boldsymbol{H}^m$ is the measurement matrix shown in Eq. (5).

As shown in Figure B, three circular regions with constant radii r (0.5 m in this study) are set around the predicted positions, $\hat{P}_t^1$, $\hat{P}_t^2$, and $\hat{P}_t^3$, of the tracked person, and their union is set as the validation region.

The method of state update mentioned above is effective when a GCP exists within the validation region, as shown in Figure A(a). However, in an environment where two people are in close proximity, as shown in Figure A (b), multiple GCPs are often obtained within the validation region or the validation regions for two people overlap.



(a) Case 1 (b) Case 2

Figure A. Data association between tracked people and LiDAR measurements (GCPs). The red circles and black triangles indicate the tracked people and GCPs, respectively. The black dashed lines indicate the validation region.

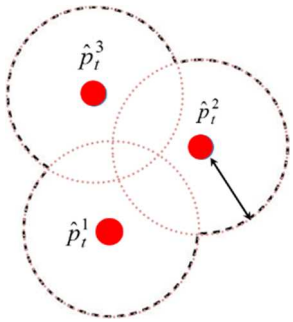


Figure B. Validation region. The red circles indicate the positions of the tracked person predicted by the three motion modes. The orange dotted circles indicate the regions with a constant radius around the predicted positions. The black dashed lines indicate the validation region for data association.

To accurately track people in such situations, data association (one-to-one matching of multiple GCPs and tracked people) is performed using the GNN method.

As shown in Figure A (b), the entire validation region where the validation regions for two people overlap is called the overlapping validation region. In such a case, we consider that I people exist and J GCPs are received, where I is not necessarily equal to J. Here, the one-to-one matching of I people and J GCP s is considered.

The *m*-th mode-based predicted position of the *i*-th tracked person is denoted by $\hat{\boldsymbol{p}}_{it}^m$, where m = 1, 2, 3, and $i = 1, 2, \cdots I$. The *j*-th GCP is denoted by z_{jt}, where $j = 1,2,\cdots,J$. The Mahalanobis distance related to $\hat{\boldsymbol{p}}_{it}^m$ and z_{jt} is then defined as follows:

$$\lambda_{ij}^m = \sqrt{(\boldsymbol{z}_{jt} - \hat{\boldsymbol{p}}_{it}^m)^T (\boldsymbol{L}_{it}^m)^{-1} (\boldsymbol{z}_{jt} - \hat{\boldsymbol{p}}_{it}^m)} \quad \text{(B)}$$

where $\boldsymbol{L}_{it}^m$ is the covariance of the prediction error ($\boldsymbol{z}_{jt}$ - $\hat{\boldsymbol{p}}_{it}^m$).

The cost function is defined by $d_{ij} = \min(\lambda_{ij}^1, \lambda_{ij}^2, \lambda_{ij}^3)$, where i= 1, 2,$\cdots I$, and j= 1, 2,$\cdots J$, and the following cost matrix $\boldsymbol{D}$ is defined:

$$\boldsymbol{D} = \begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1J} \\ d_{21} & d_{22} & \cdots & d_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ d_{I1} & d_{I2} & \cdots & d_{IJ} \end{pmatrix} \quad \text{(C)}$$

As shown in Figure A (b), of the two GCPs within the validation region of person 2, the GCP on the right side is not considered to be that of person 1 because it does not lie within the validation region of person 1. Thus, if the GCP z_{jt} is not in the validation region of the *i*-th tracked person, the cost function is set to $d_{ij} = \infty$.

Let $a(i)$ be the number of GCPs corresponding to the *i*-th person being tracked. Then, the data association can be performed by finding $a(i)$ that minimizes $\sum_{i=1}^{I} d_{ia(i)}$ [28][29].

ACKNOWLEDGMENT

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

REFERENCES

[1] M. Matsuba, R. Imai, M. Hashimoto, and K. Takahashi, "Cooperative Tracking of People using Networked LiDARs," Proc. of Int. Conf. on Sensor Technologies and Applications, pp. 1–6, 2023.

[2] 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.

[3] F. Camara et al., "Pedestrian Models for Autonomous Driving Part I: Low-level Models from Sensing to Tracking," IEEE Trans. on Intelligent Transportation Systems, vol. 22, pp. 6131–6151, 2021.

[4] Md. H. Sharif, "Laser-based Algorithms Meeting Privacy in Surveillance: A Survey," IEEE Access, vol. 9, pp. 92394 –92419, 2021.

[5] 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.

[6] T. Wu, J. Hu, L. Ye, and K. Ding, "A Pedestrian Detection Algorithm Based on Score Fusion for Multi-LiDAR Systems," Sensors, vol. 21, pp. 1–16, 2021.

[7] K. Nakamura, H. Zhao, X. Shao, and R. Shibasaki, "Human Sensing in Crowd using Laser Scanners," Laser Scanner Technology, pp. 15–32, 2012.

[8] A. T. Kamel, J. A. Farrell, and A. K. Roy-Chowdhury, "Information Weighted Consensus Filters and Their Application in Distributed Camera Networks," IEEE Trans. on Automatic Control, vol. 58, pp. 3112–3125, 2013.

[9] Y. C. Wu, C. H. Chen, Y. T. Chiu, and P. W. Chen, "Cooperative People Tracking by Distributed Cameras Network," Electronics, vol. 10, pp. 1–15, 2021

[10] 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, pp. 780–783, 1988.

[11] 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, pp. 103–123, 1988.

[12] 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.

[13] 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.

[14] T. Nakahira, M. Hashimoto, and K. Takahashi, "Cooperative People Tracking with Multiple Ground Laser Scanners," Proc. of Int. Symp. on Flexible Automation, 2018.

[15] 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.

[16] L. Liu et al., "Deep Learning for Generic Object Detection: A Survey," Int. J. of Computer Vision, vol. 123, pp. 261–318, 2020.

[17] Y. Li et al., "Deep Learning for LiDAR Point Clouds in Autonomous Driving: A Review," IEEE Trans. on Neural Networks and Learning Systems, vol. 32, pp. 3412–3432, 2020.

[18] Y. Kunisada, T. Yamashita, and H. Fujiyoshi, "Pedestrian-Detection Method Based on 1D-CNN during LiDAR Rotation," Proc. of 2018 21st Int. Conf. on Intelligent Transportation Systems, pp. 2692–2697. 2018.

[19] 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.

[20] 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.

[21] 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.

[22] T. Raj, F. H. Hashim, A. B. Huddin, M. F. Ibrahim, and A. Hussain, "A Survey on LiDAR Scanning Mechanisms," Electronics, vol.9, pp. 1–25, 2020.

[23] S. Kiranyaz et al., "1D Convolutional Neural Networks and Applications: A Survey," Mechanical Systems and Signal Processing 151, 2021.

[24] M. Matsuba, M. Hashimoto, and K. Takahashi, "People Detection and Tracking Using Ground LiDAR," Proc. of 2022 3rd Asia Symp. on Signal Processing, pp. 56–61, 2022.

[25] M. Ester, H. P. Kriegel, J. Sander, and X. Xu, "A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," Proc. of 2nd Int. Conf. on Knowledge Discovery and Data Mining, pp. 226–231, 1996.

[26] L. Xiao and S. Boyd, "Fast Linear Iterations for Distributed Averaging," Systems & Control Letters, vol. 53, pp. 65–78, 2004.

[27] Simcenter Prescan, https://www.plm.automation. siemens. com/ global/en/ products/ simcenter/ prescan. html (access: 9 November, 2023).

[28] P. Konstantinova, A. Udvarev, and T. Semerdjiev, "A Study of a Target Tracking Algorithm using Global Nearest Neighbor Approach," Proc. of Int. Conf. on Systems and Technologies, 2003.

[29] H. W. Kuhn, "The Hungarian Method for the Assignment Problem," Naval Research Logistics Quarterly, vol. 2, pp. 83–98, 1955.