Taylor & Francis
Taylor & Francis Group

Advanced Robotics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tadr20

Gait-based person identification using 3D LiDAR and long short-term memory deep networks

Hiroyuki Yamada, Jeongho Ahn, Oscar Martinez Mozos, Yumi Iwashita & Ryo Kurazume

To cite this article: Hiroyuki Yamada, Jeongho Ahn, Oscar Martinez Mozos, Yumi Iwashita & Ryo Kurazume (2020) Gait-based person identification using 3D LiDAR and long short-term memory deep networks, Advanced Robotics, 34:18, 1201-1211, DOI: 10.1080/01691864.2020.1793812

To link to this article: https://doi.org/10.1080/01691864.2020.1793812

Published online: 16 Jul 2020.

ADVANCED ROBOTICS
2020, VOL. 34, NO. 18, 1201–1211
https://doi.org/10.1080/01691864.2020.1793812

FULL PAPER

Gait-based person identification using 3D LiDAR and long short-term memory deep networks*

Hiroyuki Yamada[a,b], Jeongho Ahn[a], Oscar Martinez Mozos[c], Yumi Iwashita[d] and Ryo Kurazume[e]

[a]Graduate School of Information Science and Electrical Engineering, Kyushu University, Fukuoka, Japan; [b]Research & Development Group, Hitachi, Ltd., Ibaraki, Japan; [c]Centre for Applied Autonomous Sensor Systems, Örebro University, Örebro, Sweden; [d]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA; [e]Faculty of Information Science and Electrical Engineering, Kyushu University, Fukuoka, Japan

ABSTRACT

Gait recognition is one measure of biometrics, which also includes facial, fingerprint, and retina recognition. Although most biometric methods require direct contact between a device and a subject, gait recognition has unique characteristics whereby interaction with the subjects is not required and can be performed from a distance. Cameras are commonly used for gait recognition, and a number of researchers have used depth information obtained using an RGB-D camera, such as the Microsoft Kinect. Although depth-based gait recognition has advantages, such as robustness against light conditions or appearance variations, there are also limitations. For instance, the RGB-D camera cannot be used outdoors and the measurement distance is limited to approximately 10 meters. The present paper describes a long short-term memory-based method for gait recognition using a real-time multi-line LiDAR. Very few studies have dealt with LiDAR-based gait recognition, and the present study is the first attempt that combines LiDAR data and long short-term memory for gait recognition and focuses on dealing with different appearances. We collect the first gait recognition dataset that consists of time-series range data for 30 people with clothing variations and show the effectiveness of the proposed approach.

ARTICLE HISTORY
Received 31 March 2020
Accepted 26 June 2020

KEYWORDS
Gait recognition; point cloud; convolutional neural network; long short-term memory; data augmentation

1. Introduction

Gait recognition, which identifies individuals based on gait features, is a type of biometric recognition. Other types of biometrics include face, fingerprint, and retina recognition. Although most biometric methods require direct contact between a device and a subject, gait recognition is unique in that this method does not require interaction with the subject and can be performed from a distance, making gait recognition suitable for surveillance.

Gait recognition approaches generally fall into two main categories: (1) model-based methods and (2) model-free (appearance-based) methods. Model-based approaches include the parameterization of gait dynamics, such as the stride length, the cadence, and the joint angles [1–4]. Traditionally, these approaches have not been reported to have high performance on common databases, partly due to the self-occlusion caused by the crossing of the legs and arms of the subject. Therefore, model-free approaches (e.g. using silhouettes or textures) have become mainstream. GEINet [5] is an example of a successful appearance-based approach. In GEINet, the gait energy image (GEI) [6], which is the average of silhouette images for one gait cycle, is input to a convolutional neural network (CNN), and person identification is performed. The GEI is a compressed version of the time-series information of the pedestrian and loses information about time-series changes. Therefore, we herein propose a method that uses long short-term memory-based (LSTM-based) networks as a method by which to more actively use time-series changes in gaits. In other words, the way in which the characteristics of the instantaneous gait changes in the time series (fluctuations or phase shifts in arm and leg speeds, etc.) is learned by the LSTM, and identification is performed. In addition, the LSTM-based network does not need to detect the gait cycle, as in the GEI. Therefore, the LSTM-based method is considered to be more suitable for real-time identification.

Cameras are commonly used for gait recognition, and a number of researchers have used depth information captured by an RGB-D camera, such as the Microsoft

CONTACT Hiroyuki Yamada yamada@irvs.ait.kyushu-u.ac.jp

*Portions of this work were previously presented at the Robotics Symposia 2019 which was held on-line.

Kinect [7,8]. Depth-based gait recognition has several advantages. For example, (i) depth-based gait recognition can be used in the dark, (ii) extracting a subject using simple background subtraction is easy using depth-based gait recognition, and (iii) depth-based gait recognition is robust to appearance variations, such as clothing changes, because, in general, skeletal information extracted from three-dimensional information is used. However, RGB-D cameras have limitations in that they cannot be used outdoors, and the measurement distance of these cameras is limited to approximately 10 meters.

In recent years, real-time multi-line LiDARs, which can obtain three-dimensional range information to a target using multiple laser beams in real time, have been attracting attention for autonomous driving, and the cost of multi-line LiDARs has been decreasing greatly. Some LiDARs, such as Velodyne HDL, PUCK, or SICK TiM7xx, can be used even outdoors, and measurement is possible at ranges of over 25 meters. In the future, LiDAR will be installed widely throughout cities and will be used for indoor and outdoor surveillance and identification systems, as shown in Figure 1. For example, an autonomous driving bus will be able to identify users and determine whether to stop or pass, and the traveling car will be able to detect elderly people with Alzheimer's or dementia wandering in places where security cameras are not installed. Therefore, the present paper focuses on confirming the potential for person identification using LiDAR data. In particular, we focus on the identification of trained subjects with different appearances (mainly due to difference in clothing), which is assumed to identify subjects registered for a specific service, such as gait identification for MaaS.

Despite their rapid spread, multi-line LiDARs have rarely been used in biometrics. To the best of our knowledge, only the works of Benedek et al. exist as an example of gait recognition using LiDAR[1] [9–11]. However, these studies focus on the re-recognition of a person in a short time series with no change in appearance, which is different from identification of a person with different appearances, as is the case examined in the present study. Here, the appearance is a distribution of measured 3D points which changes by the difference in body shape or clothing. One reason may be that, although the spatial resolution of multi-line LiDARs in the horizontal direction is quite high, the spatial resolution in the vertical direction is much lower than that of the camera. For example, the vertical and horizontal resolutions of the Velodyne HDL-32e are 32 and 2170, respectively. Although single-shot range data may not produce a gait feature with a high discrimination capability due to the low resolution in the vertical direction, by using time-series range data obtained by real-time multi-line LiDAR, the possibility of using a sensor to capture biometrics information exists.

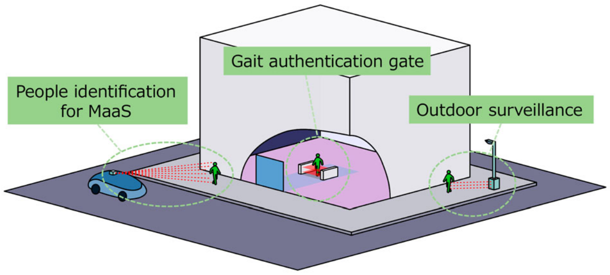

Figure 1. Applications of the present study. In the future, LiDAR will be widely installed throughout cities and can be used for indoor/outdoor surveillance and identification systems.

There are four contributions of the present paper:

- We propose a system for gait recognition using a real-time multi-line LiDAR. To the best of our knowledge, this is the first LiDAR-based identification system that focuses on dealing with different appearances.
- We design an LSTM-based method which has two advantages: (i) robustness to low resolution in the vertical direction and (ii) robustness to clothing variation.
- We develop a data augmentation method called appearance change processing (ACP) in order to improve the performance of gait recognition.
- We design a dataset that consists of time-series range data for 30 people with clothing variations and demonstrate the effectiveness of the proposed approach.

The remainder of the present paper is organized as follows. Section 2 describes related research, including research on laser-based person detection and gait recognition using RGB-D cameras. Section 3 reports a dataset that is newly created using a real-time multi-line LiDAR. Section 4 describes the details of the proposed person identification method and ACP. Section 5 describes experiments performed using the newly developed dataset. Section 6 describes comparative experiments with different methods for evaluating the proposed method. Finally, conclusions are presented in Section 7.

2. Related research

2.1. Person detection in range data

Range sensors have been used to detect and track people in different applications. Pioneering research on person detection using 2D range data has been conducted [12], where point features were learned using 2D laser scans to train a leg classifier. Recent studies have improved the 2D detectors using other techniques, such as deep learning [13–15]. Two-dimensional range detection of persons has been successfully applied in mobile robotics [15] and autonomous cars [16]. A direct extension that combines several 2D range detectors to improve robustness to occlusions has also been proposed [17–19]. Other approaches use 2.5-D multi-layer laser scanners to detect people in range data [20].

More recent approaches use 3D point clouds to detect and track people. In a previous study [21], 3D data from a LiDAR sensor is used for long-range pedestrian detection. A previous study [22] applied robust multi-target tracking in order to avoid the need for data annotation by human experts. Person detection using 3D LiDAR sensors has usually been applied in autonomous vehicles [23] and service robotics [22].

These studies were conducted in order to detect people in range data without identifying specific persons. In the present study, however, we attempt to identify particular persons using range data.

2.2. Person identification

Gait recognition can be separated into two categories: model-free methods and model-based methods [24]. Model-free methods are commonly used for camera-based gait recognition. Several techniques to extract gait features, such as the gait energy image (GEI) [6], affine moment invariants [25], the active energy image [26], the gait flow image [27], and the frame difference frieze pattern [28], were used.

In general, model-free methods are sensitive to changes in appearance, and the correct classification rate is low in cases in which the appearance of the subject is different from that in the database. In order to deal with this problem, several methods have been proposed to reduce the effect of appearance changes [29,30]. Bashir et al. [29] introduced the Gait Entropy Image (GEnI) method to select common dynamic areas among the image of the subject and images in the database. We proposed a part-based gait identification method, which adaptively chooses areas with high discrimination capability [30].

These methods are applicable to depth/range images using these images in the form of silhouette images [31]. However, in general, these approaches require high-resolution images to extract gait features that have high discrimination capability.

The model-based methods from RGB-D cameras, such as the Microsoft Kinect, are based on detecting and tracking the parameters of a human body, such as stride length, cadence, and joint angles, followed by feature extraction and classification [7,8,32]. The quality of the extraction of human parameters depends on the quality of depth information of the human body. For the case in which the depth information of the human body is too sparse, the performance of gait recognition is degraded.

3. Point cloud gait dataset

3.1. Data collection

Since there is no sufficient dataset of gaits by point clouds at present, we created the point cloud gait (PCG) dataset. In order to create this dataset, we collected data from 30 pedestrians. The subjects were in their 20s to 50s, and three of the subjects were women. All subjects were requested to walk as usual along a circular line 5 meters in diameter. There are no other restrictions, such as restrictions related to clothing or walking patterns. Data were acquired twice in spring and summer, and two datasets, PCG1 (spring) and PCG2 (summer), were created with two different conditions of appearance (mainly clothing).

As shown in Figure 2, an omni-directional multi-beam 3D LiDAR, i.e. the HDL-32E (Velodyne), was used to obtain pedestrian data. The sensor measures 360 degrees horizontally with 2170 steps and 41.3 degrees vertically with 32 lines. Each subject walked for approximately 4 minutes along a circular line. The 16 lines from the bottom were discarded because only the ground was acquired, and the dataset was created using the data from the top 16 lines. Hereafter, the top 16 lines are referred to as Lines 1, 2, . . . , 16, in order, from the bottom, as shown in Figure 2. Let $\mathbf{P}^x$ be the point cloud of the time series acquired for subject x:

$$\mathbf{P}^x = \{\mathbf{p}_1^x, \ldots, \mathbf{p}_{N^x}^x\}, \quad \mathbf{p}_t^x \in \mathbb{R}^{L\times S} \tag{1}$$

where N^x is the total number of time series samples for subject x, L is the total number of lines ($= 16$), and S is the total number of horizontal scanning steps ($= 2170$). Moreover, $\mathbf{p}_t^x$ is a matrix of size $L \times S$.

3.2. Creating dataset

In the present study, $\mathbf{p}_t^x$ is a horizontal 360-degree point cloud, in which most of the data are unrelated to the subjects. For efficient learning, only the point cloud indicating the subjects should be extracted, and the input size should be reduced. Therefore, a point cloud for fixed number of s steps was extracted in the horizontal direction around the position of the foot of the pedestrian. This also means that the input size to the neural network is fixed. The procedure is as follows:

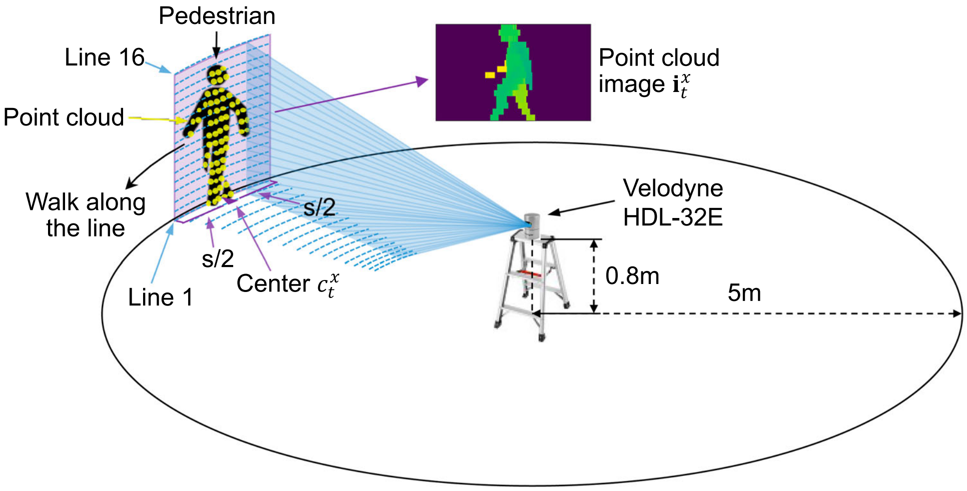

Figure 2. Data acquisition environment. Subjects walk a fixed distance around the 3D LiDAR, and the point cloud data were acquired for the sides of the subjects. The point cloud of the subjects was extracted as a point cloud image I of $16 \times s$ in size.

(1) The background subtraction processing $\mathcal{B}$ is applied to the acquired point cloud $\mathbf{p}_t^x$ for a time step t, and the point cloud $\tilde{\mathbf{p}}_t^x$ only for pedestrians is created:

$$\tilde{\mathbf{p}}_t^x = \mathcal{B}(\mathbf{p}_t^x) = \begin{cases} \mathbf{p}_{t,l,h}^x & (\mathbf{p}_{t,l,h}^x < \mathbf{B}_{l,h} - d_{th}) \\ 0 & (\text{otherwise}) \end{cases} \tag{2}$$

In the background subtraction processing $\mathcal{B}$, the measurement data $\mathbf{p}_t^x$ is compared with background data $\mathbf{B}$ obtained without a subject in advance. Then, if the distance between the point of each line l and horizontal step h of the measurement data and the corresponding background data is equal to or less than the threshold d_{th}, the point is set as a background. In this processing, the background is set to 0.

(2) The horizontal center position c_t^x of a pedestrian x at time step t is calculated as follows:

$$c_t^x = f_{index-mean}(\tilde{\mathbf{p}}_t^x, l_c) \tag{3}$$

which is a function that finds the center of the index of the point cloud existing for line l_c. Except where noted, $l_c = 1$ was used in the experiments in the present paper, which means that the horizontal center position of the pedestrian is determined based on the foot position. In this dataset, Line 1 was approximately at the height of the ankle and there was no clear situation where Line 1 was empty.

(3) A point cloud image $\mathbf{i}_t^x$ (see Figure 1) is extracted from $\tilde{\mathbf{p}}_t^x$ by c_t^x:

$$\mathbf{i}_t^x = \tilde{\mathbf{p}}_{t,1:L,(c_t^x - s/2):(c_t^x + s/2 - 1)}^x \tag{4}$$

Although this is similar to an image, elements are depths rather than color channels.

(4) Repeat procedures (1) through (3) for all measurement time steps N^x of subject x. Let $\mathbf{I}^x$ be the dataset of subject x:

$$\mathbf{I}^x = \{\mathbf{i}_1^x, \ldots, \mathbf{i}_{N^x}^x\}, \quad \mathbf{i}_t^x \in \mathbb{R}^{L \times s} \tag{5}$$

We obtained data of two appearance conditions, $c1$ and $c2$, for 30 subjects. Here, let PCG1 and PCG2 be the point cloud gait (PCG) datasets for conditions $c1$ (spring)

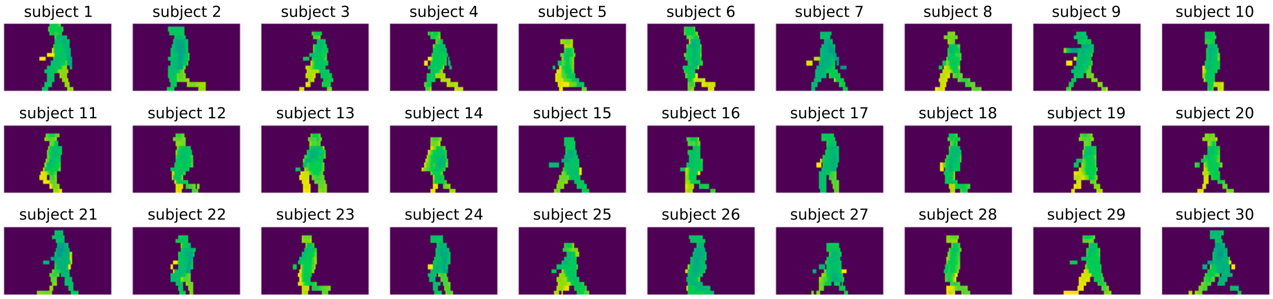

Figure 3. Example of a PCG dataset. This figure shows the point cloud images $\mathbf{I}^{x,c1}$ ($x = 1, \ldots, 30$) at time step $t = 0$. In this example, the size of each image is 16×128 ($L = 16, s = 128$). Color represents depth.

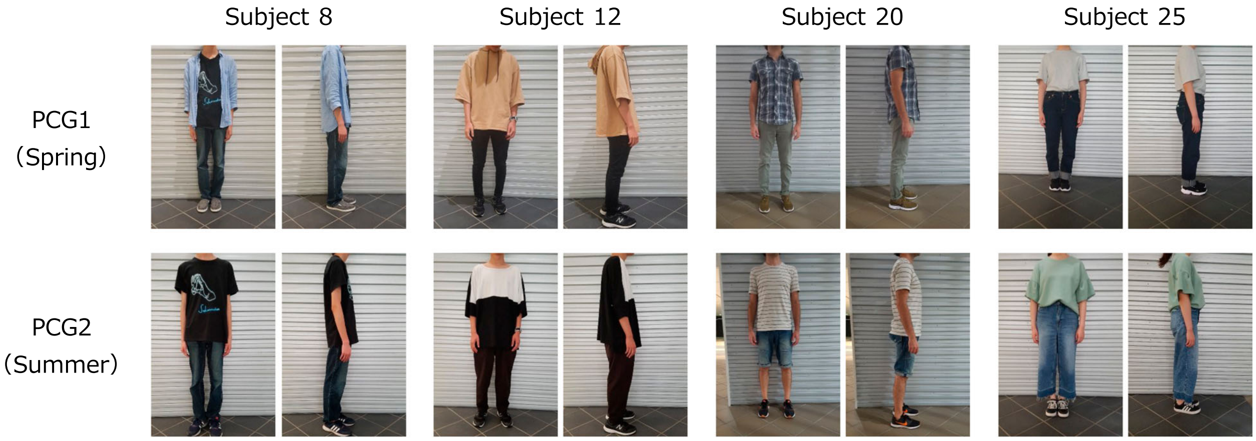

Figure 4. Example of a difference in appearance between the two datasets. PCG1 dataset is composed of 3D point cloud data measured for subjects wearing spring clothing, and PCG2 is for subjects wearing summer clothing.

and $c2$ (summer):

$$\mathrm{PCG1} = \{\mathbf{I}^{1,c1}, \ldots, \mathbf{I}^{X,c1}\}, \quad \mathrm{PCG2} = \{\mathbf{I}^{1,c2}, \ldots, \mathbf{I}^{X,c2}\} \tag{6}$$

where $\mathbf{I}^{x,c}$ is the dataset of subject x under condition c, and X is the total number of subjects ($X = 30$). In the experiment, PCG1 and PCG2 are used as training or test data. Figure 3 shows an example of the PCG dataset and Figure 4 shows an example of a difference in appearance between the two datasets.

4. Proposed network

4.1. Input to the network

The network for person classification by gait using point clouds that we have constructed is roughly divided into an encoding part and a classification part. The encoding part is expected to encode information from the shape of the instantaneous gait and appearance of the pedestrian using CNNs, and the classification part is expected to classify pedestrians from the time series change of the encoded information using LSTM networks.

The PCG dataset is made up of point cloud images of L lines and s horizontal steps ($\mathbf{i}_t^x \in \mathbb{R}^{L\times s}$). On the other hand, if the data can be classified with fewer lines of data, there are many practical advantages, such as lower training costs and a simpler sensor configuration. Therefore, we tried to classify subjects using only the point clouds of some lines. Let l be the number of lines used for network input. If $l = 1$, then only data near the foot are used. If $l = 1{:}4$ (Line 1, 2, 3, 4), then data from the foot to the knee are used. Finally, if $l = 1{:}8$ (Line $1, 2, \ldots, 8$), then data near the lower body are used. Then, input data $\mathbf{i}_{t:t+T,l}^x$ for T time steps are extracted at any time step t and for any subject x, which are randomized with respect to the order of input to the network.

Here, T is the number of inputs to the LSTM. In other words, a person is classified according to changes in T time steps. In this way, the input to the network is generated as shown in Figure 5.

4.2. Appearance change processing (ACP)

In order to improve the classification performance for unknown appearances, we proposed ACP as a data augmentation method. Appearance change processing is applied to the data for each input datum $\mathbf{i}_{t:t+T}^x$ in order to achieve high generalization performance that can cope with changes in appearance, such as differences in clothing. The processing procedure is as follows:

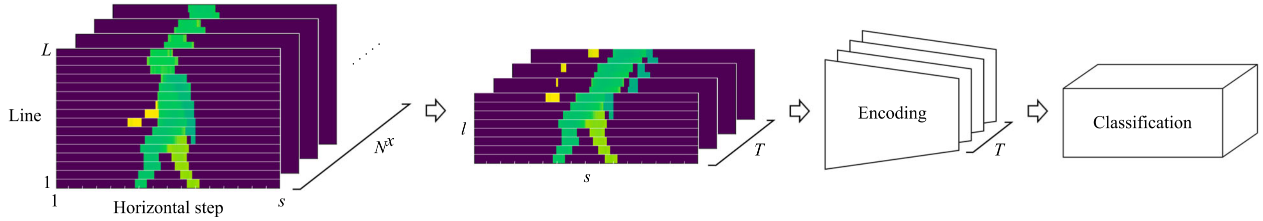

Figure 5. Inputs to the network. Subject x and time step t are selected in a random order from the dataset, and l lines of data for T time steps are used as the input.

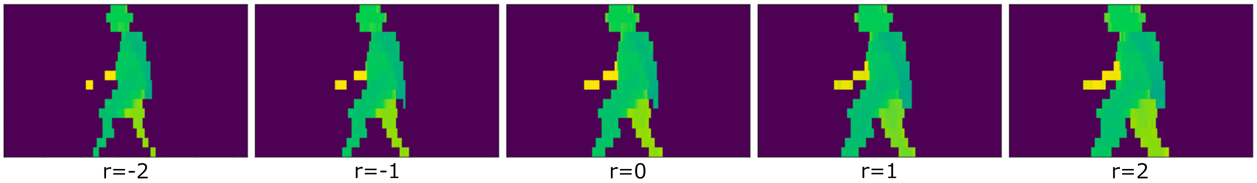

Figure 6. Examples of a point cloud image using ACP (when R is 2). The case of $r = 0$ is the original case, and the appearance becomes thick when $r > 0$ and thin when $r < 0$.

(1) Cluster points consecutive in the horizontal step direction within each line in the data $\mathbf{i}_t^x$ of time step t.
(2) Generate a uniform random variable r between -R and R (where R is a variable indicating how many steps to change).
(3) Copy data from both ends to the outside for each cluster in each line by r steps (when $r > 0$) or delete the inner data (when $r < 0$). Add noise according to a Gaussian distribution when copying.
(4) Repeat processing for T time steps with the same r.

Figure 6 shows example results of applying ACP. The purpose of this process is to virtually create a change in appearance from light clothing or clothing that fits the body to thick clothing, such as coats or down jackets. It is expected that learning with data from one type of appearance condition by applying ACP is equivalent to learning with data for multiple appearance conditions.

4.3. Network structure

As mentioned above, the network consists of an encoding part and a classification part. The encoding part consists of four layers of a 2D-CNN, and the classification part consists of four layers of the LSTM. We conducted experiments using two types of networks, herein referred to as Network 1 and Network 2.

Figure 7 shows the basic structure of the proposed network (Network 1). The encoding part of Network 1 consists of four layers of a 2D-CNN and the 2D-Max pooling layer. The batch normalization layer does not depend on the depth direction (left-right shift of the pedestrian) during laser measurement, and a dropout layer to avoid overlearning is inserted after each CNN layer. The parameters of each layer in Network 1 are as shown in Figure 7. In the encoding part, a total of T CNN-based layers are prepared and processed in parallel. These layers at the same level share their weights. Eventually, T encoded data $\mathbf{f}_t$, which are made one dimensional by a flattening layer, are calculated by encoder $\mathcal{E}$:

$$\{\mathbf{f}_1, \ldots, \mathbf{f}_T\} = \mathcal{E}(\mathbf{i}_t^x, \ldots, \mathbf{i}_{t+T}^x) \tag{7}$$

which is an encoding function that consists of all layers of the encoding part.

The classification part of Network 1 consists of four layers of the LSTM. Again, the dropout layer is inserted after each LSTM layer in order to reduce overlearning. The first three LSTM layers of this network use the output of the hidden layer for T time steps as the input to the next LSTM layer, and the last LSTM layer uses the output of the Tth hidden layer as the input to the fully connected layer. Finally, the probability p^x of each subject x is output by the Softmax layer of classifier $\mathcal{C}$:

$$\{p^1, \ldots, p^X\} = \mathcal{C}(\mathbf{f}_1, \ldots, \mathbf{f}_T) \tag{8}$$

which is a classification function that consists of all layers of the classification part.

The proposed Network 1 with ACP appears to work well for subjects with different appearances. However, the classifier should preferably use only the gait features that

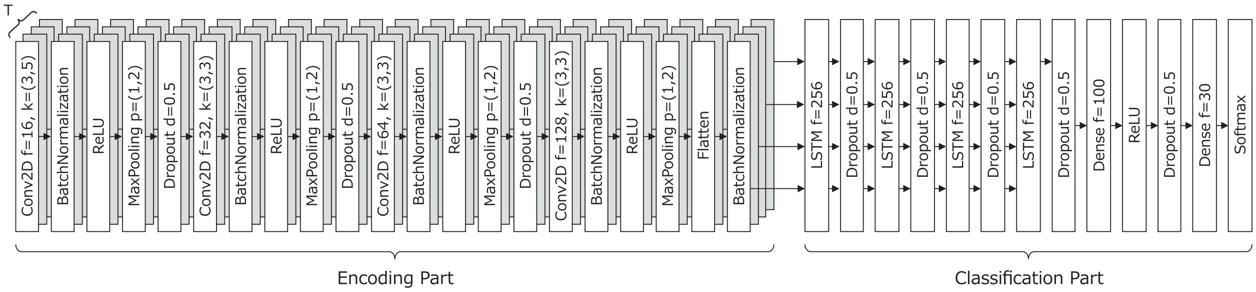

Figure 7. Structure of Network 1 for person classification by gait using point clouds. Here, f is the filter size, k is the kernel size, p is the pooling size, and d is the dropout coefficient. The smaller number is selected for the kernel size of each CNN layer, and the pooling size of each max pooling layer when the input size I is smaller than the numbers in the figure.

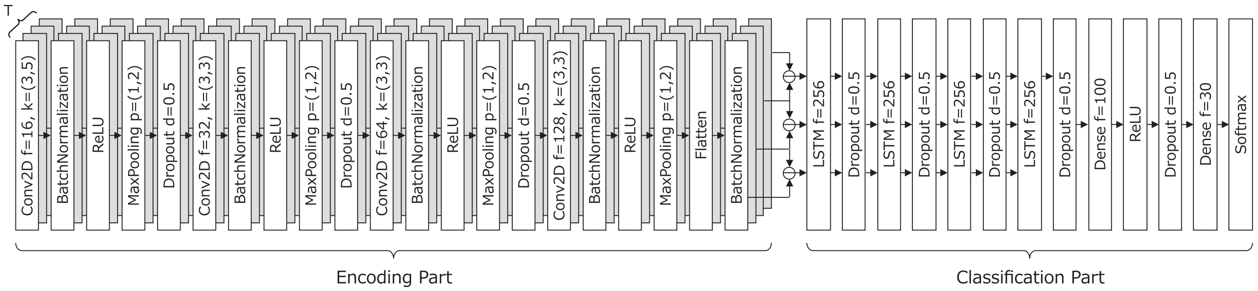

Figure 8. Structure of Network 2 for person classification by gait using point clouds. Only the input method of the classifier $\mathcal{C}$ is different from that of Network 1. ⊖ is an element-wise subtraction operator.

are unique to the individual with the appearance features removed. Therefore, we came up with Network 2, in which the input to the classifier of Network 1 is modified. Figure 8 shows the structure of Network 2. In this network, the input to the classifier $\mathcal{C}$ is changed from $\mathbf{f}_i$ in Equation (8) to $\mathbf{f}_{i+1} - \mathbf{f}_i$, as follows:

$$\{p^1, \ldots, p^X\} = \mathcal{C}(\mathbf{f}_2 - \mathbf{f}_1, \ldots, \mathbf{f}_T - \mathbf{f}_{T-1}) \qquad (9)$$

The input to the classifier is the difference between time steps for the encoded feature $\mathbf{f}_t$. This calculation assumes that the appearance features are canceled out by subtraction and that more individual relative change features are extracted.

5. Experiments

5.1. Examining the difference in classification accuracy of each line

First, in order to investigate the effect of point cloud lines on classification accuracy, we performed classification using only one line from Lines 1 to 16. When Line 1 is not used, which means that l does not include 1, the horizontal center position c_t^x is recalculated using $l_c = min(l)$ according to Equation (3). Here, PCG1 and Network 1 were used as the dataset and network, respectively. Training and test data were used by dividing PCG1 by two. Appearance change processing was not used in training. Although the accuracy tends to be high because the training and test data are from the same time series, the purpose of this experiment is to determine the relative potential for the classification of each line.

Table 1. Overall accuracy for each line.

l	Accuracy	l	Accuracy
1	0.756	9	0.482
2	0.837	10	0.609
3	0.828	11	0.381
4	0.824	12	0.299
5	0.726	13	0.359
6	0.729	14	0.386
7	0.789	15	0.159
8	0.643	16	0.071

Table 1 shows the experimental results. The accuracies of $l = 1$ to 4 were high, and the accuracy decreased as the line moved from the foot to the head. This indicates that the leg movement is most suitable for gait recognition in the proposed network. The movement of the leg has a large difference between individuals, dynamic behavior, and a small periodic difference. Therefore, the data of this part is considered to contribute to the accuracy improvement. On the other hand, swinging arms have fewer individual differences and dynamics than leg movements. In addition, since the arms are more flexible than the legs, there were movements with no periodicity, such as crossing of arms and touching of the face in the datasets. These are thought to be factors that reduce classification accuracy when using the data of the middle part of point cloud lines. Since each line is normalized at the horizontal center of the point clouds, classification is easily understood to be difficult only for the line near the head.

Based on these results, Lines 9 through 16 do not contribute significantly to the classification performance and may even have an adverse influence. Thus, Lines 1 through 8 are used for classification in the following experiments.

5.2. Evaluation of proposed networks and ACP

We evaluated four patterns of the two proposed networks, i.e. with and without ACP. In addition, we investigated three patterns of input data ($l = 1, l = 1{:}4, l = 1{:}8$). Adam [33] was used as an optimizer, and a training coefficient of 0.001 was used in training. The batch size was 32, and the training data were 1504 and 1672 batches per epoch in PCG1 and PCG2, respectively. The experiments were conducted using $T = 10$ and $s = 128$, and $R = 2$ when ACP was enabled.

In the experiment, PCG1 and PCG2 were exchanged for gallery and probe. Table 2 shows the overall accuracies of classification. Network 2 outperformed Network

Table 2. Overall accuracies. Here, two datasets were used alternately for gallery and probe, and two cases were averaged.

		Gallery	Probe	$l = 1$	$l = 1{:}4$	$l = 1{:}8$
Network 1	without ACP	PCG1	PCG2	0.127	0.214	0.475
		PCG2	PCG1	0.207	0.289	0.475
		Average		0.167	0.252	0.475
	with ACP	PCG1	PCG2	0.266	0.423	0.577
		PCG2	PCG1	0.271	0.338	0.508
		Average		0.267	0.381	0.543
Network 2	without ACP	PCG1	PCG2	0.195	0.366	0.527
		PCG2	PCG1	0.231	0.306	0.526
		Average		0.213	0.336	0.527
	with ACP	PCG1	PCG2	0.271	0.421	0.621
		PCG2	PCG1	0.275	0.370	0.616
		Average		**0.273**	**0.396**	**0.619**

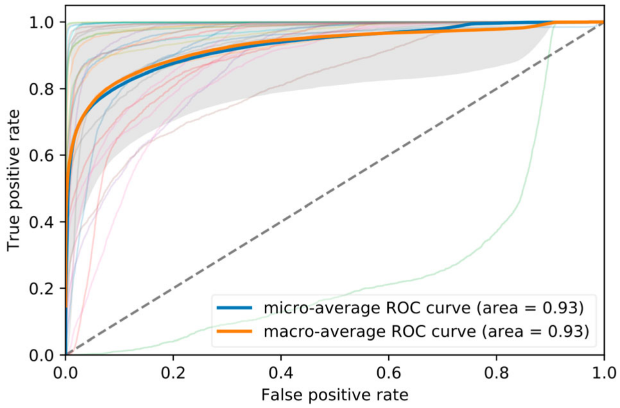

Figure 9. Average ROC curve of Network 2 with ACP using eight lines. Each thin line is the ROC curve of 30 subjects, and the gray area is their standard deviation.

1 in almost all patterns, and the case with ACP outperformed the case without ACP for all patterns. On average of $l = 1$, $l = 1{:}4$ and $l = 1{:}8$, Network 2 achieved a 4.6% higher accuracy than Network 1, and ACP contributed to an improvement in accuracy of 8.5%. Even under the difficult condition of using only one line, Network 2 with ACP had an accuracy of 27.3%.

Figure 9 shows the average ROC curves for Network 2 with ACP using $l = 1{:}8$. The AUC on the average ROC curves was 0.93, which indicates that the proposed network has high classification performance.

5.3. Appearance change processing evaluation experiment

In order to confirm the effectiveness of ACP, a third dataset, PCG3, was prepared and experiments were performed. Here, PCG3 is composed of data measured in different clothing for 20 of the 30 subjects included in PCG1 and PCG2. Here, PCG1, PCG2, and PCG3 were measured in the spring, summer, and winter of the same year, respectively. Therefore, PCG2 contains data for the lightest clothing, and PCG3 contains data for the thickest clothing.

Table 3. Overall accuracies of the ACP evaluation experiment.

		Gallery	Probe	$l = 1$	$l = 1{:}4$	$l = 1{:}8$
Network 2	without ACP	PCG1,2	PCG3	0.573	0.609	0.569
		PCG2,3	PCG1	0.466	0.609	0.818
		PCG1,3	PCG2	0.376	0.496	0.788
		Average		0.472	0.590	**0.725**
	with ACP	PCG1,2	PCG3	0.488	0.708	0.547
		PCG2,3	PCG1	0.525	0.701	0.815
		PCG1,3	PCG2	0.444	0.540	0.792
		Average		**0.486**	**0.649**	0.718

Note: Two datasets are used for gallery, and the other dataset is used for probe.

We conducted an experiment in which 20 subjects included in PCG3 were trained on two datasets and tested using the other dataset. In this experiment, only Network 2 was used among the proposed networks.

The results are shown in Table 3. When the network was trained using multiple datasets, the difference with and without ACP was smaller than for cases that use a single dataset for training in Section 5.2. In addition, the results revealed an improvement over the accuracies in Table 2. These results indicate that ACP is effective when training data restricted to only one type of appearance are used, and is not so effective when training with data containing two or more types of appearances. It was also found that the use of actual clothing data contributes to higher accuracy than data augmentation using virtual clothing by ACP.

6. Discussions

6.1. Comparison with GEINet

As a comparison with existing methods, we performed an identification experiment using GEINet [5], which is one of the most successful gait recognition methods among the model-free methods. Figure 10 shows an example of a GEI realized using the dataset that we created from LiDAR data. GEINet inputs an 88 × 128-sized GEI, while we input 1, 4, 8, 16 × 128-sized GEIs. Therefore, the kernel size of the CNN in GEINet was changed and used in the experiment. Table 4 shows the results of an experiment using GEINet. The proposed network outperformed all other identification methods using one, four, and eight lines. In this experiment, although evaluation using sixteen lines (whole body) was also performed, it did not exceed the accuracy of the proposed network using eight lines. Since the LiDAR data has a lower spatial resolution and a lower time density than general camera images, the results indicate that the GEI is rough and difficult to identify. Considering these results, gait recognition using LiDAR data is more difficult than using

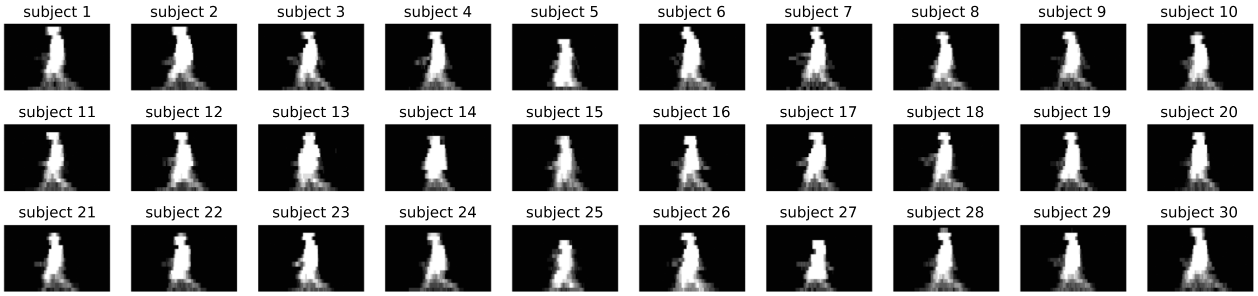

Figure 10. Example of GEIs for all subjects.

Table 4. Overall accuracies of the GEINet experiment.

	Gallery	Probe	$l = 1$	$l = 1{:}4$	$l = 1{:}8$	$l = 1{:}16$
GEINet	PCG1	PCG2	0.200	0.340	0.455	0.514
	PCG2	PCG1	0.211	0.360	0.476	0.573
	Average		**0.205**	**0.350**	**0.466**	**0.543**

Table 5. Overall accuracies of the 3d convolutional network.

		Gallery	Probe	$l = 1$	$l = 1{:}4$	$l = 1{:}8$
3D-CNN Network	without ACP	PCG1	PCG2	0.124	0.257	0.439
		PCG2	PCG1	0.188	0.254	0.393
		Average		**0.156**	0.256	0.416
	with ACP	PCG1	PCG2	0.153	0.280	0.410
		PCG2	PCG1	0.137	0.243	0.439
		Average		0.145	**0.262**	**0.425**

camera images, and the proposed method deals with this data relatively well.

6.2. Comparison with the 3D-CNN

The proposed network is a combination of a 2D-CNN and the LSTM and attempts high accuracy classification by training time-series changes of instantaneous gait patterns encoded by a CNN using the LSTM. In previous studies [34], a 3D-CNN was used as a method for training temporal changes in 2D images. However, we built such a network in an attempt to reduce the dependence on appearance features and improve generalization performance. In order to verify this performance, we constructed a network that uses a 3D-CNN to simultaneously convolve the time series direction (t direction). Figure 11 shows the network using a 3D-CNN for evaluation.

The results are shown in Table 5. The overall accuracy was lower than in the proposed network. It was also shown that ACP does not contribute much to accuracy improvement in a 3D-CNN network. This appears to be because, even if the data is augmented by ACP, the data is different from the original data. In other words, the network using only the CNN strongly depends on the appearance, and the classification ability for unknown clothing is relatively low.

These results are thought to be the difference between using the shape change directly and using information encoded with gait patterns. Although the 3D-CNN network is effective for data in which appearance does not change greatly, the proposed network is more suitable for gait recognition, and the proposed LSTM-based network works effectively, as expected.

7. Conclusions

The present paper introduces the first system of LSTM-based gait recognition using a real-time multi-line LiDAR. We realized a dataset consisting of the time-series range data of 30 people with clothing variations and proposed two types of LTSM-based CNN networks for gait recognition. The experimental results revealed

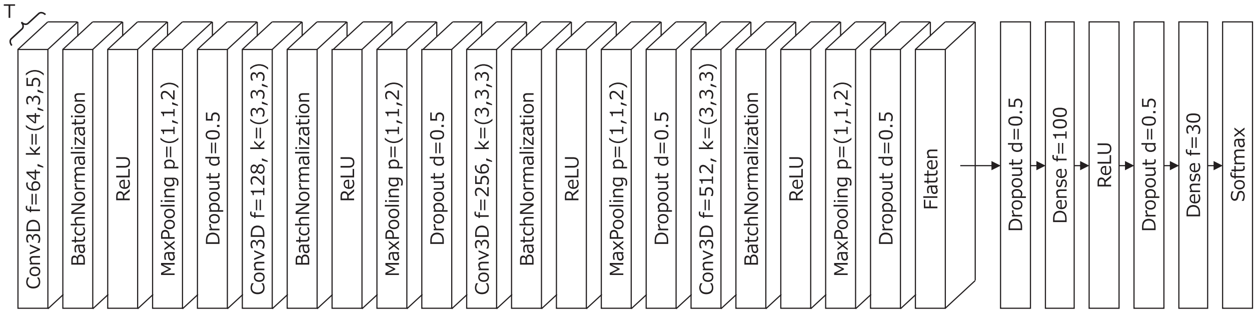

Figure 11. Structure of 3D-CNN Network for person classification by gait using point clouds.

that the proposed approach achieved a high classification performance of approximately 60% and successfully verified the effectiveness of gait recognition using a real-time multi-line LiDAR.

Note

1. Sadeghzadehyazdi et al. also state in their literature [32] that Benedek et al.'s work is the only existing lidar-based person identification methods.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Hiroyuki Yamada received his MS degree from Graduate School of Information Science and Electrical Engineering, Kyushu University, Japan, in 2008. He is currently a senior researcher of Research & Development Group, Hitachi Ltd., Japan, and also a PhD student at Graduate School of Information Science and Electrical Engineering, Kyushu University, Japan. His research interests mainly include robotics, computer vision and machine learning. He is a member of the RSJ adn JSME.

Jeongho Ahn received his B.S. degree from Undergraduate School of Electronic Engineering, Gachon University, Republic of Korea, in 2019. He is currently a M.S. student at Graduate School of Information Science and Electrical Engineering, Kyushu University. His research interests mainly include computer vision, machine learning and biometrics.

Oscar Martinez Mozos is a WASP-AI Senior Lecturer at Örebro University, Sweden. Before, he was a senior researcher at the Technical University of Cartagena, Spain; a senior lecturer at the University of Lincoln, UK; and a postdoctoral fellow of the Japan Society for the Promotion of Science (JSPS) at Kyushu University, Japan. He got his MSc (2005) and PhD (2008) from the University of Freiburg, Germany. His areas of research include artificial intelligence, robotics, and assistive technologies.

Yumi Iwashita received her M.S. degree and her Ph.D. from the Graduate School of Information Science and Electrical Engineering, Kyushu University in 2004 and 2007, respectively. In 2007, she was a PostDoc at Imperial College London under Professor Maria Petrou. From 2007 to 2014, she was an assistant professor at Kyushu University. Between 2011 and 2013, she was a visiting researcher at NASA's Jet Propulsion Laboratory. From 2014 to 2016, she was an associate professor at Kyushu University. Since 2016, she has been a research technologist at Jet Propulsion Laboratory and a visiting associate professor at Kyushu University. Her current research interests include computer vision and machine learning for robotics and security systems.

Ryo Kurazume is a Professor at the Graduate School of Information Science and Electrical Engineering, Kyushu University. He received his M.Eng. and PhD in Mechanical Engineering from Tokyo Institute of Technology in 1989 and 1998. He was a director of the Robotics Society of Japan (RSJ) from 2009 to 2011 and 2014 to 2015 and the Society of Instrument and Control Engineers (SICE) from 2013 to 2015, and a chairman of the Japan Society of Mechanical Engineers (JSME) Robotics and Mechatronics Division in 2019. He received JSME Robotics and Mechatronics Academic Achievement Award in 2012, RSJ Fellow in 2016, SICE System Integration Division Academic Achievement Award in 2017, JSME Fellow in 2018, and SICE Fellow in 2019. His current research interests include legged robot control, computer vision, multiple mobile robots, service robots, care technology, and biometrics.

ORCID

Oscar Martinez Mozos http://orcid.org/0000-0002-3908-4921
Ryo Kurazume http://orcid.org/0000-0002-4219-7644

References

[1] Bouchrika I, Nixon MS. People detection and recognition using gait for automated visual surveillance. 2006 IET Conference on Crime and Security; 2006 Jun; London. p. 576–581.

[2] Cunado D, Nixon MS, Carter JN. Automatic extraction and description of human gait models for recognition purposes. Comput Vis Image Und. 2003;90(1):1–41.

[3] Tafazzoli F, Safabakhsh R. Model-based human gait recognition using leg and arm movements. Eng Appl Artif Intell. 2010;23(8):1237–1246.

[4] Yam C, Nixon MS, Carter JN. Automated person recognition by walking and running via model-based approaches. Pattern Recogn. 2004;37(5):1057–1072.

[5] Shiraga K, Makihara Y, Muramatsu D. Geinet: View-invariant gait recognition using a convolutional neural network. 2016 International Conference on Biometrics (ICB); 2016 Jun; Halmstad, Sweden. p. 1–8.

[6] Han J, Bhanu B. Individual recognition using gait energy image. IEEE Trans Pattern Anal Mach Intell. 2006 Feb;28(2):316–322.

[7] Balazia M, Sojka P. You are how you walk: Uncooperative mocap gait identification for video surveillance with incomplete and noisy data. Proceedings of the IEEE/IAPR International Joint Conference on Biometrics (IJCB); 2017 Oct; Denver, CO.

[8] Kozlow P, Abid N, Yanushkevich S. Gait type analysis using dynamic bayesian networks. Sensors. 2018;18(10):3329.

[9] Benedek C, Gálai B, Nagy B, et al. Lidar-based gait analysis and activity recognition in a 4d surveillance system. IEEE Trans Circuits Syst Video Technol. 2018 Jan;28(1):101–113.

[10] Benedek C, Nagy B, Gálai B. Lidar-based gait analysis in people tracking and 4d visualization. 2015 23rd European Signal Processing Conference (EUSIPCO); 2015 Aug; Nice, France. p. 1138–1142.

[11] Galai B, Benedek C. Feature selection for lidar-based gait recognition. 2015 International Workshop on Computational Intelligence for Multimedia Understanding (IWCIM); 2015 Oct; Prague, Czech Republic. p. 1–5.

[12] Arras KO, Mozos OM, Burgard W. Using boosted features for the detection of people in 2d range data. Proceedings 2007 IEEE International Conference on Robotics and Automation; 2007; Roma, Italy. p. 3402–3407.

[13] Beyer L, Hermans A, Leibe B. Drow: real-time deep learning-based wheelchair detection in 2-d range data. IEEE Robot Autom Lett. 2017 Apr;2(2):585–592.

[14] Beyer L, Hermans A, Linder T, et al. Deep person detection in two-dimensional range data. IEEE Robot Autom Lett. 2018;3(3):2726–2733.

[15] Sun L, Yan Z, Mellado SM, et al. 3dof pedestrian trajectory prediction learned from long-term autonomous mobile robot deployment data. 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE; 2018. p. 1–7.

[16] Wang DZ, Posner I, Newman P. Model-free detection and tracking of dynamic objects with 2d lidar. Int J Robot Res. 2015;34(7):1039–1063.

[17] Mozos OM, Kurazume R, Hasegawa T. Multi-part people detection using 2D range data. Int J Soc Robot. 2010 Mar;2(1):31–40.

[18] Spinello L, Arras KO, Triebel R. A layered approach to people detection in 3d range data. Twenty-Fourth AAAI Conference on Artificial Intelligence; 2010; Atlanta, GA.

[19] Spinello L, Luber M, Arras KO. Tracking people in 3d using a bottom-up top-down detector. 2011 IEEE International Conference on Robotics and Automation; 2011 May; Shanghai, China. p. 1304–1310.

[20] Kim B, Choi B, Park S, et al. Pedestrian/vehicle detection using a 2.5-d multi-layer laser scanner. IEEE Sens J. 2016 Jan;16(2):400–408.

[21] Li K, Wang X, Xu Y, et al. Density enhancement-based long-range pedestrian detection using 3-d range data. IEEE Trans Intell Transp Syst. 2016 May;17(5):1368–1380.

[22] Yan Z, Duckett T, Bellotto N. Online learning for human classification in 3d lidar-based tracking. 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS); 2017 Sep; Vancouver, BC, Canada. p. 864–871.

[23] Wang H, Wang B, Liu B, et al. Pedestrian recognition and tracking using 3d lidar for autonomous vehicle. Robot Auton Syst. 2017;88:71–78.

[24] Matovski DS, Nixon MS, Carter JN. Gait recognition. Boston (MA): Springer US; 2014. p. 309–318.

[25] Iwashita Y, Kurazume R. Person identification from human walking sequences using affine moment invariants. IEEE International Conference on Robotics and Automation; 2009; Kobe, Japan. p. 436–441.

[26] Zhang E, Zhao Y, Xiong W. Active energy image plus 2dlpp for gait recognition. Signal Process. 2010;90(7): 2295–2302.

[27] Lam T, Cheung K, Liu J. Gait flow image: A silhouette-based gait representation for human identification. Pattern Recogn. 2011;44(4):973–987.

[28] Shinzaki M, Iwashita Y, Kurazume R. Gait-based person identification method using shadow biometrics for robustness to changes in the walking direction. IEEE Winter Conference on Applications of Computer Vision; 2015; Waikoloa, HI. p. 670–677.

[29] Bashir K, Xiang T, Gong S. Gait recognition without subject cooperation. Pattern Recognit Lett. 2010;31(13): 2052–2060. Meta-heuristic Intelligence Based Image Processing.

[30] Iwashita Y, Uchino K, Kurazume R. Gait-based person identification robust to changes in appearance. Sensors. 2013;13(6):7884–7901.

[31] Nunes JF, Moreira PM, Tavares JMRS. Gridds – a gait recognition image and depth dataset. In: Tavares JMRS, Jorge RMN, editors, VipIMAGE 2019. Cham: Springer International Publishing; 2019. p. 343–352.

[32] Sadeghzadehyazdi N, Batabyal T, Glandon A. Glidar3dj: a view-invariant gait identification via flash lidar data correction. 2019 IEEE International Conference on Image Processing (ICIP); 2019; Taipei,Taiwan. p. 2606–2610.

[33] Kingma D, Ba J. Adam: A method for stochastic optimization. International Conference on Learning Representations; 2015 May; San Diego, CA.

[34] Tran D, Bourdev L, Fergus R. Learning spatiotemporal features with 3d convolutional networks. 2015 IEEE International Conference on Computer Vision (ICCV); 2015; Santiago, Chile. p. 4489–4497.