2V-Gait: Gait Recognition using 3D LiDAR Robust to Changes in Walking Direction and Measurement Distance

Jeongho Ahn[1] Kazuto Nakashima[2] Koki Yoshino[1] Yumi Iwashita[3] Ryo Kurazume[2]

***Abstract*— Gait recognition, which is a biometric identifier for individual walking patterns, is utilized in many applications, such as criminal investigation and identification systems, because it can be applied at a long distance and requires no explicit cooperation of the subjects. In general, cameras are used for gait recognition, and several methods in previous studies have used depth information captured by RGB-D cameras. However, RGB-D cameras are limited in terms of their measurement distance and are difficult to access outdoors. In recent years, real-time multi-layer 3D LiDAR, which can obtain 3D range images of a target at ranges of over 100 m, has attracted significant attention for use in autonomous mobile robots, serving as eyes for obstacles detection and navigation. Compared with cameras, such 3D LiDAR has rarely been used for biometrics owing to its low spatial resolution. However, considering the unique characteristics of 3D LiDAR, such as the robustness of the illumination conditions, long measurement distances, and wide-angle scanning, the approach has the potential to be applied outdoors as a biometric identifier. The present paper describes a gait recognition system, called 2V-Gait, which is robust to variations in the walking direction of a subject and the distance measured from the 3D LiDAR. To improve the performance of gait recognition, we leverage the unique characteristics of 3D LiDAR, which are not included in regular cameras. Extensive experiments on our dataset show the effectiveness of the proposed approach.**

I. INTRODUCTION

Gait recognition is a promising biometric identifier that can be utilized as an alternative to other approaches, such as faces, fingerprints, and retinas, based on unique physical and behavioral characteristics. Compared with other biometric modalities, gait, which depicts the walking pattern of an individual, has several advantages in that it can be easily captured at a long distance and requires no explicit cooperation of the subjects of interest. In addition, it is difficult to camouflage because of the gait dynamics. Owing to these advantages, gait recognition has enormous potential in many applications, including criminal investigations and identification systems.

Cameras are generally used for gait recognition [1], and several methods described in previous studies use depth information captured by an RGB-D camera, such as a Microsoft Kinect [2]. In addition, RGB-D cameras have several advantages, such as robustness to the illumination conditions and a simple background subtraction. However, the measurement distance and field of view of RGB-D cameras are limited to approximately 10 m and 70 °, respectively, and such cameras are difficult to be apply outdoors.

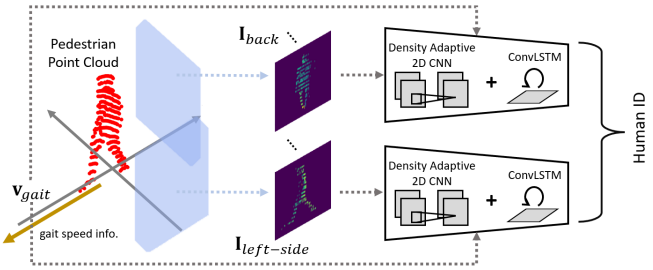


Fig. 1: Overview of 2V-Gait, utilizing the unique characteristics of 3D LiDAR that are not found in regular RGB cameras. After estimating the walking direction of a subject point cloud, the two viewpoint-invariant gait image and gait speed sequences are extracted and fed into the recognition network to identify an individual.

In recent years, real-time multi-layer LiDAR, which can obtain 3D range images of a target at a range of over 100 m, has attracted a great deal of attention in computer vision and has been applied in autonomous mobile robots to serve as eyes for object detection and navigation. In comparison with cameras, however, 3D LiDAR has rarely been used in biometrics. One possible reason for this is the poor spatial resolution compared with RGB cameras, making it difficult to capture the whole human shape. However, considering the unique characteristics of 3D LiDAR, such as the robustness of the illumination conditions, long measurement distances, and a scanning range covering 360 ° in the azimuth, it has a potential outdoor application as a biometric identifier. Furthermore, 3D LiDAR can be used as an alternative to RGB cameras in terms of protecting personal information.

We previously proposed an LSTM-based gait recognition method using 3D LiDAR [10] and showed that 3D LiDAR has a high potential for such recognition. However, in this study [10], both the distance and walking direction from the 3D LiDAR were constant. This situation is valid if the 3D LiDAR is placed in a corridor or narrow street, and people are walking in a single direction. In the future, however, 3D LiDAR will be widely used in practical applications, primarily mobile robots, for person identification. For example, security robots that can be operated 24 h a day and are less conspicuous than humans are becoming increasingly common in malls, offices, and public spaces. A nighttime surveillance system can be achieved by applying

This work was partially supported by JSPS KAKENHI Grant Number JP20H00230.

Jeongho Ahn and Koki Yoshino are with Graduate School of Information Science and Electrical Engineering, Kyushu University, Japan ahn, yoshino@irvs.ait.kyushu-u.ac.jp

Kazuto Nakashima and Ryo Kurazume are with Faculty of Information Science and Electrical Engineering, Kyushu University, Japan k_nakashima@irvs.ait.kyushu-u.ac.jp, kurazume@ait.kyushu-u.ac.jp

Yumi Iwashita is with Jet Propulsion Laboratory, California Institute of Technology, USA yumi.iwashita@jpl.nasa.gov

biometrics to these security robots without the installation of other sensors. Autonomous vehicles can detect and identify specific users while driving. Based on the above cases, it is necessary to design a robust gait recognition model for intra-subject changes, such as the viewing angles and distances measured. To reduce the influence of these variations, we focus on two fixed viewpoints and the walking pace, which may be invariant gait features for the conditions above, as shown in Fig. 1.

Our contributions can be summarized as follows:

- We propose a system for gait recognition using 3D LiDAR that is robust to variations in the walking direction of the subject and the distance measured from a sensor.
- To improve the performance, we utilize the unique characteristics of 3D LiDAR, such as 3D spatial and positional information, which are not included in regular cameras.

II. Related Work

A. Gait Representation

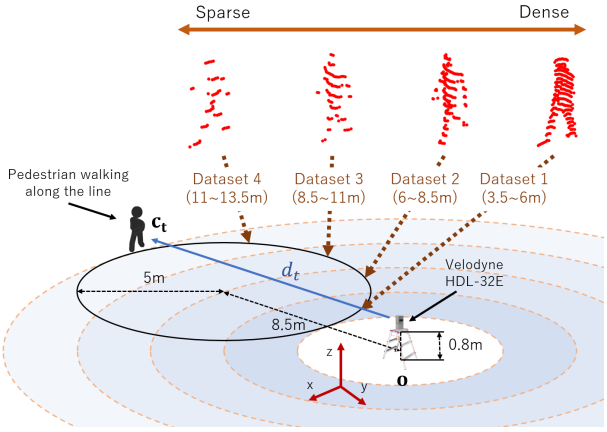
Another recognition problem in computer vision is the extraction of gait-related features from the video frames of a walking person, for which most prior approaches are separated into two types of gait representation: model-based and appearance-based methods. Model-based methods first apply pose estimation and take an articulated body with spatial and temporal parameters, including the skeleton, stride length, cadence, and joint angles [3, 4].

By contrast, appearance-based features use human-shape-related features directly from the original images. A gait energy image (GEI) [5] is an example of a successful appearance-based approach. In a GEI, the spatial-temporal information of a gait cycle is encoded into a single image. Extended approaches, such as the frame difference frieze pattern [6] and gait flow image [7], have been proposed and have shown a better performance than GEI. However, these methods compress the gait sequence into a single frame, which results in a loss of opportunity in applying the temporal changes of gait dynamics in an explicit manner. We focus on the appearance-based approach because it would be difficult to maintain pose estimation accuracy enough due to the sparseness and incompleteness of the point cloud captured by 3D LiDAR.

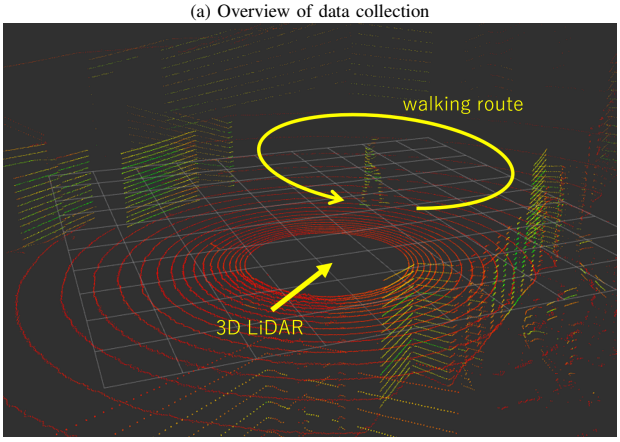
B. Person Identification using 3D LiDAR

To the best of our knowledge, only the studies by Benedek et al. are examples of gait recognition using 3D LiDAR [9]. These studies focus on re-recognition, adopting the GEI-based method, to track a specific person for a short period of time. However, it is difficult to extract the dynamic gait features under temporal changes, which is critical to the performance of appearance-based gait recognition models.

In contrast to the above approach, we proposed an LSTM-based gait recognition method [10] that learns the temporal gait changes. However, this method suffers from a degraded recognition performance when the walking direction of an individual or the distance measured from a sensor is not constant. Thus, this method loses the flexibility of the gait recognition under real-world scenarios, including variations such as the viewing angle and measurement distance. To solve these issues, we focus on the robustness against the walking direction of a person and the distance measured by a sensor.



(a) Overview of data collection



(b) Three-dimensional point cloud visualization

Fig. 2: Data acquisition environment. The subjects walked along a circular line with a radius of 5 m, the center of which was 8.5 m away from the sensor. To evaluate the robustness of our approach, the gait data extracted through background subtraction were divided into four datasets according to the distance measured d_t.

III. Datasets

A. Data Collection

Using 32 vertical lasers to verify the effectiveness of our method, we collected the gait data captured using a Velodyne HDL-32E, which creates horizontal 360° 3D range images. These data consist of gait sequences from 31 subjects in a point-cloud format. Requesting the subjects to walk as usual along a circular line with a radius of 5 m, where the center is 8.5 m away from the sensor, the gait data contain the changes in the 360° walking direction and the distance measured from 3.5 to 13.5 m, as shown in Fig. 2.

B. Creating Dataset

To evaluate the robustness of our method for changes in the walking direction or distance measured, we constructed four datasets by dividing the time-series gait data acquired above, and thus the viewing angles and measurement distances are different. Let $\mathbf{P}_t$ be the subject point set for the time step t, extracted by background subtraction processing:

$$\mathbf{P}_t = \{\mathbf{p}_{t,1}, \mathbf{p}_{t,2}, ..., \mathbf{p}_{t,N}\}, \tag{1}$$

where N is the total number of points for each subject, and each point $\mathbf{p}_{t,n} \in \mathbb{R}^3$ indicates its orthogonal coordinates $(p_{t,n,x}, p_{t,n,y}, p_{t,n,z})$. Given a subject point cloud extracted, we define the center of gravity for a subject, $\mathbf{c}_t = (c_{t,x}, c_{t,y}, c_{t,z})$, as follows:

$$\mathbf{c}_t = \frac{1}{N}\sum_{n=1}^{N} \mathbf{p}_{t,n}, \tag{2}$$

where $c_{t,z}$ was set to 0 to simplify the distance calculation. Given a subject central point $\mathbf{c}_t$, the distance from the sensor position $\mathbf{o}$ to a subject d_t can be calculated as follows:

$$d_t = \|\mathbf{c}_t - \mathbf{o}\|_2, \tag{3}$$

Finally, the extracted time-series gait data were divided into four subsets according to the calculated distance d_t: datasets 1, 2, 3, and 4 containing gait sequences with ranges of $3.5-6$ m, $6-8.5$ m, $8.5-11$ m, and $11-13.5$ m, respectively.

IV. Proposed Method

In this section, we propose a pipeline for our proposed gait recognition method: gait direction transformation, input generation, and the use of a recognition network. In particular, to enhance the robustness of changes in the walking direction and the distance measured from a sensor, we take advantage of 3D LiDAR, which is not included in normal cameras.

A. Gait Direction Transformation

We first describe how to estimate the walking direction of a pedestrian and transform the subject point cloud into a new subject point cloud heading in a constant direction. In other words, the walking direction of a subject point set is transformed into the direction toward the $-y$-axis and generated into left-side view gait images from the yz coordinate plane. First, the gait directional angle for a time step t can be calculated from the subject's central points before and after a time step t:

$$\theta_t = \arctan2(c_{t+1,y} - c_{t-1,y}, c_{t+1,x} - c_{t-1,x}), \tag{4}$$

where $\arctan2(\cdot,\cdot)$ is defined as the angle between the positive x-axis in the Euclidean plane. Given a directional angle θ_t, a subject point cloud transformed $\hat{\mathbf{P}}_t = \{\hat{\mathbf{p}}_{t,1}, \hat{\mathbf{p}}_{t,2}, ..., \hat{\mathbf{p}}_{t,N}\}$ is obtained by rotating the original subject point cloud $\mathbf{P}_t$ around $\mathbf{c}_t$ as the z-axis.

$$\hat{\mathbf{p}}_{t,n} = \mathbf{R}_z(-\theta_t - \pi/2) \cdot (\mathbf{p}_{t,n} - \mathbf{c}_t), \tag{5}$$

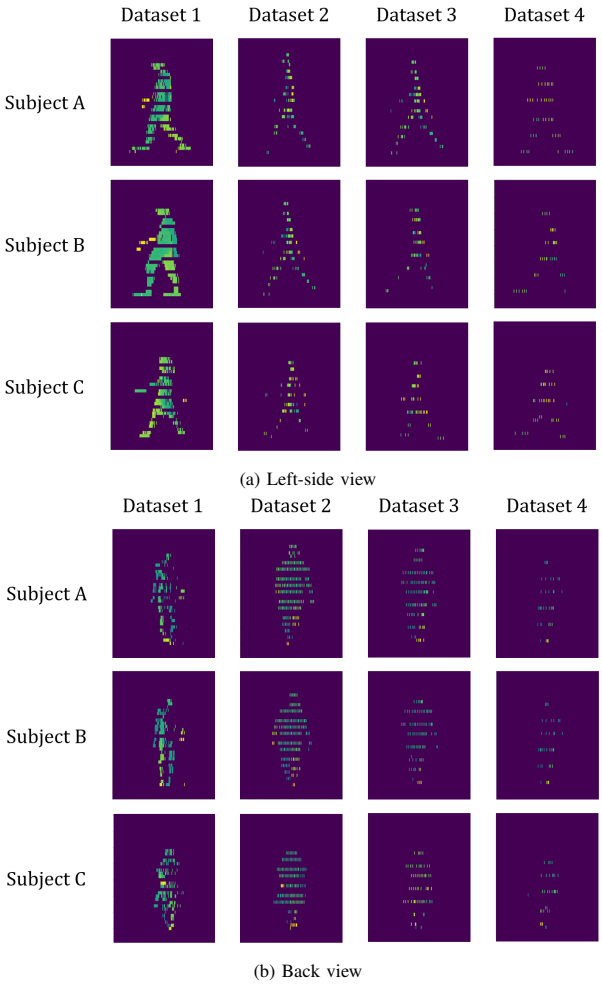


Fig. 3: Example of a generated gait image, which is one of the inputs of the recognition network. The point cloud of the subject is sparse when the target is at a long distance.

where $\mathbf{R}_z$ represents the rotation matrix around the z-axis as follows:

$$\mathbf{R}_z(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{6}$$

The case of generating a back-view gait image follows the same procedure, except that the rotation matrix $\mathbf{R}_z(-\theta_t - \pi/2)$ is replaced with $\mathbf{R}_z(-\theta_t - \pi)$.

B. Input Generation

Next, we describe how to generate the input data, which consist of a left-side view and back-view gait image sequences, and the gait speed sequences from the pedestrian point cloud transformed above for input into a recognition network. The gait image $\mathbf{i}_t \in \mathbb{R}^{V\times H\times 1}$ is generated from the pedestrian heading along the $-y$-axis:

$$\mathbf{i}_t = f_{img}\left(\hat{\mathbf{P}}_t\right), \tag{7}$$

where $f_{img}(\cdot) : \mathbb{R}^{N\times 3} \to \mathbb{R}^{V\times H\times 1}$ extracts a gait image representing the depth information of a pedestrian, as shown in Fig. 3. Here, the depth and size of the height channel are

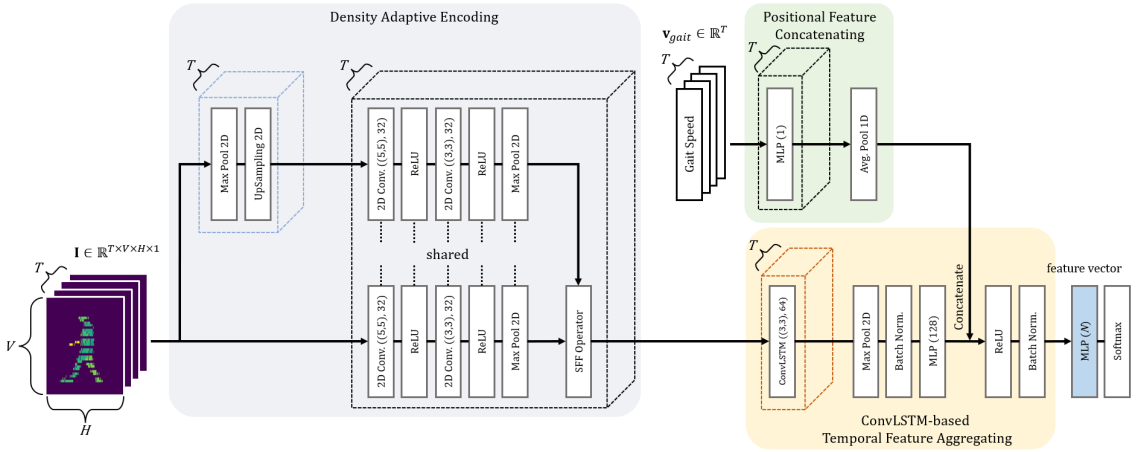


Fig. 4: The overall architecture of our proposed recognition network. The gait image and speed sequence are fed into the network as inputs to learn the spatial-temporal and positional features and improve the discriminative capability of the gait recognition.

determined as $V(=l_z/l_{z\text{-res}}) \times H(=l_y/l_{y\text{-res}})$ pixels, where l_z, l_y, $l_{z\text{-res}}$, and $l_{y\text{-res}}$ are the height of the z-axis, the width of the y-axis, the vertical resolution of the image, and the horizontal resolution of the image, respectively. As the criteria by which each pixel value of the depth channel is determined, the value is set to 0 when none of the points correspond to the pixel. Otherwise, it is set to the largest x-coordinate value in the candidate set $\{\hat{\mathbf{p}}_{t,1}, \hat{\mathbf{p}}_{t,2}, ..., \hat{\mathbf{p}}_{t,\tilde{N}}\} \subset \hat{\mathbf{P}}_t$, which are the points corresponding to the pixels in (v,h). In addition, for a clear distinction between a pedestrian and a background, the constant $l_x/2$ is added to the pixel value $i_{t,v,h}$ corresponding to

$$i_{t,v,h} = \begin{cases} \max\limits_{\tilde{n}\in\{1,\ldots,\tilde{N}\}} (\hat{\mathbf{P}}_{t,x}) + \frac{l_x}{2} & (\exists \hat{\mathbf{p}}_{t,\tilde{n}}) \\ 0 & (\text{otherwise}) \end{cases} \tag{8}$$

Here, the candidate point $\hat{\mathbf{p}}_{t,\tilde{n}}$ should satisfy the following formulation:

$$\hat{\mathbf{p}}_{t,\tilde{n}} = \{\hat{\mathbf{p}}_{t,\tilde{n}} \in \hat{\mathbf{P}}_t \mid (v=\tilde{v}) \wedge (h=\tilde{h})\}, \tag{9}$$

where $\tilde{v}$ and the $\tilde{h}$ are defined as follows:

$$\tilde{v} = \left\lfloor \frac{1}{l_{z\text{-res}}} \cdot \left(\hat{p}_{t,\tilde{n},z} + \frac{l_z}{2} \right) \right\rfloor, \tag{10}$$

$$\tilde{h} = \left\lfloor \frac{1}{l_{y\text{-res}}} \cdot \left(\hat{p}_{t,\tilde{n},y} + \frac{l_y}{2} \right) \right\rfloor \tag{11}$$

It is similar to Z-buffer method which is one of the commonly used methods for hidden surface detection in terms of comparing surface depths at each pixel position on the projection plane. Thus, we can obtain the gait image sequence $\mathbf{I} \in \mathbb{R}^{T\times V\times H\times 1}$ with T frames:

$$\mathbf{I} = (\mathbf{i}_1, \ldots, \mathbf{i}_T) \tag{12}$$

As another input for the recognition network, the gait speed can be obtained from the pedestrian central points before and after time step t and the rotation rate f_{rps} of the sensor:

$$v_t = \frac{f_{rps}}{2} \cdot \|\mathbf{c}_{t+1} - \mathbf{c}_{t-1}\|_2 \tag{13}$$

Based on the calculation above, we can obtain the gait speed sequence $\mathbf{v}_{gait} \in \mathbb{R}^T$ with T frames:

$$\mathbf{v}_{gait} = (v_1, \ldots, v_T) \tag{14}$$

Finally, we can obtain the left-side view gait image sequence $\mathbf{I}_{left-side}$, the back-viewed gait image sequence $\mathbf{I}_{back}$, and the gait speed sequence $\mathbf{v}_{gait}$ for the proposed recognition network. In this study, l_z, l_y, $l_{z\text{-res}}$, $l_{y\text{-res}}$, T, and f_{rps} are set as 2.4 m, 1.6 m, 0.06 m, 0.01 m, 10, and 10 Hz, respectively.

C. Recognition Network

In this section, we describe the recognition network for learning the discriminative information from the inputs generated above. The overall architecture, which consists of four key components, i.e., if density adaptive encoding (**DAE**), temporal feature aggregating (**TFA**), positional feature concatenating (**PFC**), and viewpoint-informed feature aggregating (**VFA**), is visualized in Fig. 4.

1) Density Adaptive Encoding: A high-resolution image, which is transformed from the point set of a subject, is expected to represent fine-grained patterns of the gait. However, this method lacks the ability to recognize the whole human body in sparse data. Consequently, the spatial features learned in dense data acquired at short distances may not generalize to long distances because a point set is generally accompanied with a non-uniform density at different distances. To alleviate this issue, we designed a density adaptive encoding (**DAE**) module that leverages the low resolution, which may be robust to sparse data and better recognize coarse-grained patterns, learning to combine multi-scale features extracted from different resolutions.

First, the double-reduced-resolution image sequence $\mathbf{I}_{low-res} \in \mathbb{R}^{T\times V/2\times H/2\times 1}$ is obtained by feeding a gait image sequence $\mathbf{I}$ into the max pooling layer:

$$\mathbf{I}_{low-res} = \text{Maxpool2D}(\mathbf{I}) \tag{15}$$

Then, $\mathbf{I}_{low-res}$ is upsampled to meet the height and width dimensions of $\mathbf{I}$:

$$\hat{\mathbf{I}}_{low-res} = \text{Upsampling2D}(\mathbf{I}_{low-res}) \tag{16}$$

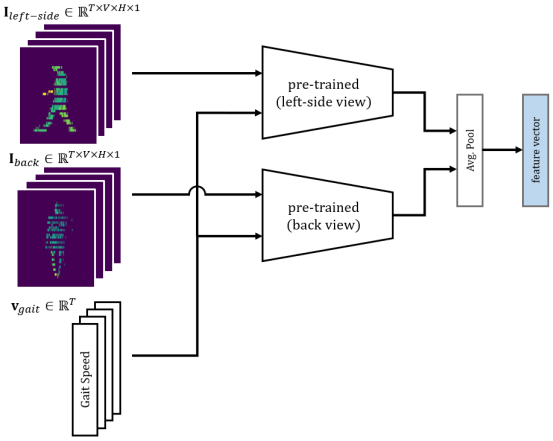


Fig. 5: Detailed structure of VFA module, which aggregates the gait features of the recognition networks pre-trained from two different viewpoints: the left-side and back views.

Next, two gait image sequences with different resolutions, $\mathbf{I}$ and $\hat{\mathbf{I}}_{low-res}$, are fed into the 2-D convolution network module Conv2D$(\cdot)$, which shares the weights and kernel optimizations, and obtains the spatial features, $\mathbf{F}, \mathbf{F}_{low-res} \in \mathbb{R}^{T\times V/2\times H/2\times 16}$, respectively. Finally, $\mathbf{F}$ and $\mathbf{F}_{low-res}$ are combined into one feature sequence $\hat{\mathbf{F}}$ using the spatial feature fusion (**SFF**) operator as follows:

$$\hat{\mathbf{F}} = \frac{1}{2} \cdot (\text{Conv2D}(\mathbf{I}) \oplus \text{Conv2D}(\hat{\mathbf{I}}_{low-res})), \tag{17}$$

where $\oplus$ represents an element-wise addition.

2) Temporal Feature Aggregating: The TFA module is composed of a convolutional LSTM (ConvLSTM) network [11], which is an extension of LSTM and is responsible for modeling the spatial-temporal representations of the gait, and a multi-layer perceptron (MLP) used to aggregate the temporal features of $\hat{\mathbf{F}}$. ConvLSTM is expected to outperform LSTM-based approaches because it captures spatio-temporal correlations simultaneously. The spatial-temporal feature $\hat{f}_{aggr} \in \mathbb{R}^{128}$ can be extracted by the TFA module from $\hat{\mathbf{F}}$ as follows:

$$\hat{f}_{aggr} = \text{TFA}(\hat{\mathbf{F}}) \tag{18}$$

3) Positional Feature Concatenating: The positional feature concatenating (**PFC**) module takes advantage of the walking speed information, which can be one of the invariant attributes for changes in distance measured from 3D LiDAR. As shown in Fig. 4, the PFC takes the average of all frames T in a re-weighted gait speed sequence $\mathbf{v}_{gait}$, and the extracted gait speed feature $f_{speed} \in \mathbb{R}$ is concatenated into $\hat{f}_{aggr}$ as follows:

$$\hat{f}^{re}_{aggr} = \text{Concat}(\hat{f}_{aggr}, f_{speed}) \tag{19}$$

4) Viewpoint-informed Feature Aggregating: The VFA module aims to obtain more discriminative features from two viewpoints: the left-side view and back view, by aggregating the outputs of two networks, $f_{left-side}, f_{back} \in \mathbb{R}^N$, which are pre-trained from two different viewpoints.

$$f_{gait} = \text{Avgpooling1D}(f_{left-side}, f_{back}), \tag{20}$$

where N is the number of subjects trained, and the one-dimensional average pooling operator is adopted for effectively aggregating two of the feature vectors, which is the same as that of RotationNet [12], as shown in Fig. 5.

V. Experiments

In this section, extensive experiments conducted on our point cloud gait datasets, which contain variations in the gait direction and distance measured from a sensor, are described, the results of which demonstrate that our proposed method is robust to such changes. The details of the dataset and training are first described. An evaluation was then conducted to verify the effectiveness of each component in 2V-Gait. Finally, the performance of the proposed method was compared with that of other prior methods.

A. Implementation Details

Four point-cloud gait datasets with different gait directions and measurement distances were used in this experiment. Each dataset contains 31 subjects and each subject contains 210 sequences. To train the recognition network, we divided each of the four datasets into training, validation, and test sets with 140, 35, and 35 sequences, respectively. The first 16 subjects were grouped into a training set, and the remaining 15 subjects were reserved as probes for testing. In addition, we used the four training sets during the training phase. In other words, the total training data consist of $16\times4\times140 = 8960$ point set sequences. During the test, the three training sets of the 15 subjects, which were not used in the training phase, were regarded as galleries. Here, there are four combinations in the three training sets used for learning. During the test phase, the network identifies subjects through the nearest neighbor algorithm (rank-1), which computes the cosine similarity between the gallery and probe.

For optimization, categorical cross-entropy loss was employed to train the network without any metric learning in this study. In addition, we trained the networks based on backpropagation using RMSProp with a learning rate of 0.001, and the training batch size was set to 20 for each person. We employ an early stopping, which is a form of regularization to avoid an overfitting during the training phase, where the patience is set to 20.

B. Evaluation of the effectiveness in 2V-Gait

The experiment result is shown in Tab. I. First, an evaluation was conducted on our dataset to verify the effectiveness of each component in 2V-Gait, following the method of adding modules individually. Comparing the results within the range of 11–13 m, we achieved a better performance by gradually adding the proposed modules. It is worth noting that the overall accuracy was greatly improved when applying the VFA module. These results indicate that the proposed approach is effective and robust for recognizing gait features in sparse data by leveraging spatial representations of two different resolutions, the walking speed information, and two invariant viewpoints. By contrast, the average accuracy decreased when the PFC module was applied. One possible

TABLE I: Averaged rank-1 accuracies on our dataset. The recognition accuracy in which the range of the test set is not included in range of the training sets is shown in bold.

Network	Gallery				Probe				mean	
	3.5−6m	6−8.5m	8.5−11m	11−13.5m	3.5−6m	6−8.5m	8.5−11m	11−13.5m	included	non-included
2V-Gait (ours) → TFA	✓	✓	✓		89.90	91.40	88.57	**62.67**	87.39	**72.60**
	✓	✓		✓	89.33	91.59	**73.52**	81.71		
	✓		✓	✓	88.76	**77.44**	86.10	80.57		
		✓	✓	✓	**76.76**	91.01	86.48	83.24		
2V-Gait (ours) → TFA + DAE	✓	✓	✓		89.71	91.59	89.52	**68.00**	87.80	**74.04**
	✓	✓		✓	89.52	89.87	**71.62**	82.48		
	✓		✓	✓	88.95	**85.47**	87.81	81.90		
		✓	✓	✓	**71.05**	91.01	87.62	83.62		
2V-Gait (ours) → TFA + DAE + PFC	✓	✓	✓		81.33	89.29	83.62	**69.71**	84.26	**71.65**
	✓	✓		✓	82.86	89.29	**66.86**	83.05		
	✓		✓	✓	81.14	**77.44**	83.05	82.48		
		✓	✓	✓	**72.57**	86.04	82.10	84.76		
2V-Gait (ours) → TFA + DAE + PFC + VFA	✓	✓	✓		92.95	95.22	94.86	**76.57**	93.57	**84.27**
	✓	✓		✓	91.81	95.41	**89.71**	91.43		
	✓		✓	✓	92.38	**89.29**	95.81	90.67		
		✓	✓	✓	**81.52**	95.22	95.62	91.43		
GEINet [8] (Shiraga et al.)	✓	✓	✓		87.05	88.72	85.71	**64.38**	84.34	**73.08**
	✓	✓		✓	87.81	88.53	**72.76**	75.43		
	✓		✓	✓	87.43	**78.59**	83.24	79.81		
		✓	✓	✓	**76.57**	87.19	84.95	76.19		
LSTMNet [10] (Yamada et al.)	✓	✓	✓		74.48	76.29	70.67	**51.43**	70.53	**61.78**
	✓	✓		✓	73.14	73.23	**59.62**	64.57		
	✓		✓	✓	74.10	**69.02**	69.14	65.33		
		✓	✓	✓	**67.05**	71.89	68.00	65.52		

reason for this is the negative effect of the walking speed information on the spatial feature of the gait among the dense data.

C. Comparison with Prior Methods

Second, the performance of 2V-Gait is compared with that of the other prior methods, i.e., GEINet [8] and an active LSTM-based network (LSTMNet) [10]. GEINet, where a GEI image is fed into a convolution neural network as an input, is one of the most successful gait recognition methods, and is equivalent with the architecture of LGEI [9]. However, LSTMNet is an LSTM-based network that uses depth image sequences as inputs for identifying individuals based on 3D LiDAR. The left-side view gait image sequence $\mathbf{I}_{left-side}$ was used as an input in both GEINet and LSTMNet. Here, $\mathbf{I}_{left-side}$ is transformed into a GEI image when GEINet is implemented. Compared with GEINet and LSTMNet, 2V-Gait clearly presented a better performance when all components were applied. In particular, our proposed approach was confirmed to be beneficial for improving the discriminative performance for individual recognition at long distances. From the results of the quantitative evaluations, we showed that 2V-Gait, which utilizes unique characteristics of 3D LiDAR, such as two different spatial resolutions, the gait speed information, and two viewpoints, improves the discriminative capability for gait recognition.

VI. Conclusion

In this paper, we presented a gait recognition method using 3D LiDAR to improve the performance of the discrimination capability and robustness of variations in the walking direction and distance measured. The gait image and speed sequences were generated as gait representations from the pedestrian point set, whose directional angle was estimated and transformed, and fed into the recognition network for identification. Finally, experiments conducted on our collected dataset demonstrate the effectiveness of the proposed method.

References

[1] Z. Zhang, L. Tran, X. Yin, Y. Atoum, X. Liu, J. Wan, and N. Wang, "Gait recognition via disentangled representation learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 4710–4719.

[2] P. Kozlow, N. Abid, and S. Yanushkevich, "Gait type analysis using dynamic Bayesian networks," *Sensors*, 2018, vol. 18, no. 10, p. 3329.

[3] T. Whytock, A. Belyaev, and N. M. Robertson, "Dynamic distance-based shape features for gait recognition," *Journal of Mathematical Imaging and Vision*, 2014, vol. 50, no. 3, pp. 1–13.

[4] F. Tafazzoli and R. Safabakhsh, "Model-based human gait recognition using leg and arm movements," *Engineering Applications of Artificial Intelligence*, 2010, vol. 23, no. 8, pp. 1237–1246.

[5] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, vol. 28, no. 2, pp. 316–322.

[6] M. Shinzaki, Y. Iwashita, R. Kurazume, and K. Ogawara, "Gait-based person identification method using shadow biometrics for robustness to changes in the walking direction," in *IEEE Winter Conference on Applications of Computer Vision*, 2015, pp. 670–677.

[7] Toby H. W. Lam, K. H. Cheung and James N. K. Liu, "Gait flow image: A silhouette-based gait representation for human identification," *Pattern Recognition*, 2011, vol. 44, no. 4, pp. 973–987.

[8] K. Shiraga, Y. Makihara, D. Muramatsu, T. Echigo, and Y. Yagi, "GEINet: related View-invariant gait recognition using a convolutional neural network," in *2016 International Conference on Biometrics (ICB)*, 2016, pp. 1–8.

[9] C. Benedek, B. Galai, B. Nagy and Z Janko, "Lidar-based gait analysis and activity recognition in a 4D surveillance system," *IEEE Transactions on Circuits and Systems for Video Technology*, 2018, vol. 28, no. 1, pp. 101–113.

[10] H. Yamada, J. Ahn, O. M. Mozos, Y. Iwashita, and R. Kurazume, "Gait-based person identification using 3D LiDAR and long short-term memory deep networks," *Advanced Robotics*, 2020, vol. 34, no. 18, pp. 1–11.

[11] X. SHI, Z. Chen, H. Wang, D. Yeung, W. Wong and W. WOO, "Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting" *Advances in Neural Information Processing Systems (NIPS)*, 2015, vol. 28.

[12] A. Kanezaki, Y. Matsushita, and Y. Nishida, "RotationNet: Joint object categorization and pose estimation using multiviews From unsupervised viewpoints," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 5010–5019.