The Supplementary Material for LidarGait: Benchmarking 3D Gait Recognition with Point Clouds

Chuanfu Shen[1,2], Fan Chao[2,3], Wei Wu[2], Rui Wang[2], George Q. Huang[4], Shiqi Yu[2,3*]

[1] Department of Industrial and Manufacturing Systems Engineering, The University of Hong Kong

[2] Department of Computer Science and Engineering, Southern University of Science and Technology

[3] Research Institute of Trustworthy Autonomous System, Southern University of Science and Technology

[4] Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University

noahshen@connect.hku.hk, {12131100, 12032501, 12232385}@mail.sustech.edu.cn

gqhuang@hku.hk, yusq@sustech.edu.cn.

A. The Suboptimal Performance on Umbrella Subset

Tab. 2 shows that LidarGait outperforms all other methods in all subsets except for cases where pedestrians carry an umbrella. This suboptimal performance is mainly due to the inclusion of the umbrella in the projection images as illustrated in Fig. 10, which causes misalignment issues decreasing performance. To improve performance in the umbrella subset, we suggest exploring approaches such as umbrella removal or training with random erasing.

B. Multi-view Fusion

In this work, we proposed LidarGait and an extended MV-LidarGait. MV-LidarGait incorporates point clouds from multiple perspectives to generate various depths and fuse multiple depths from different viewpoints into compact multi-view features. As experiments conducted in Tab. 3, we observed that BEV did not improve the performance, and thus, we solely focused on feature fusion for the front-range and right-side views. We explored two methods for combining multiple features: concatenation and sum, and also investigated sequence-level and frame-level feature fusion. Specifically, multi-view features could be fused in a frame-by-frame manner or after the temporal fusion of each viewpoint branch. The results revealed that only frame-level concatenation led to improved performance, which we report in Tab. 3.

C. Effect of Sequence Length

We have studied the impact of sequence length on the inference process. During the inference stage, we examine different frame numbers as input. The frames are randomly selected from the sequence instead of continuous frames sampling. We can observe that: (1) Both camera-based and Lidar-based models obtain better performance with the increasing frames of the sequences. (2) When only given one frame for each sequence of probe and gallery, the LiDAR-based method can surprisingly achieve 25.82 % rank-1 recognition accuracy and 52.29 % rank-5 recognition accuracy, showing the effectiveness of Lidar-based gait recognition in the few-shot setting. (3) The rank-1 accuracy made by LiDAR can be compared to the rank-5 result using camera modality, demonstrating the superior performance of LiDAR in the content of gait recognition.

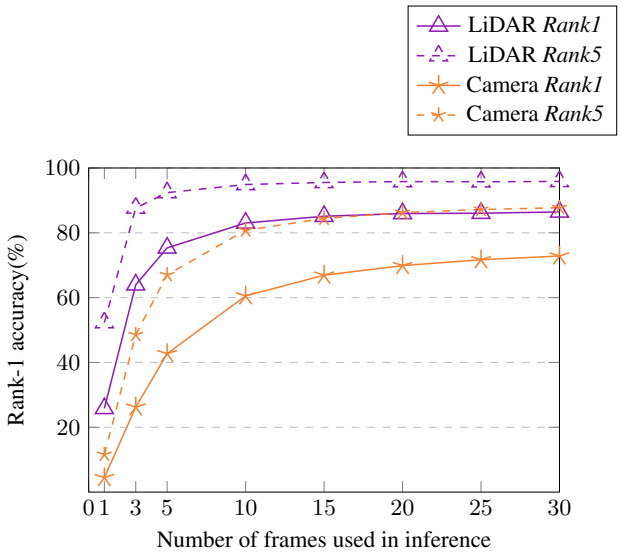


Figure 9. The performance comparison between LiDAR and camera on used frame number in inference.

*Corresponding Author

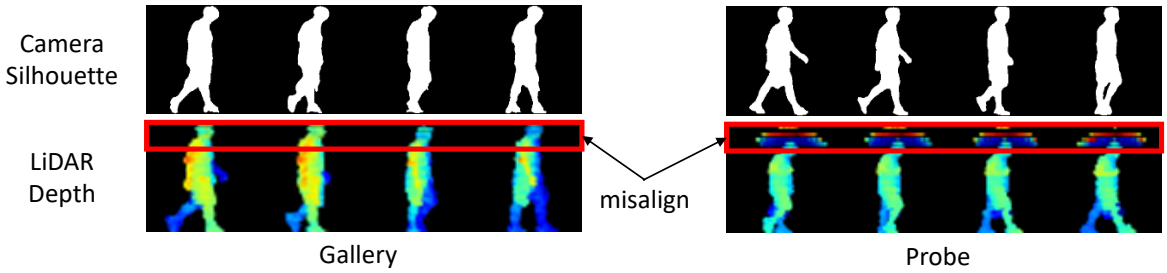
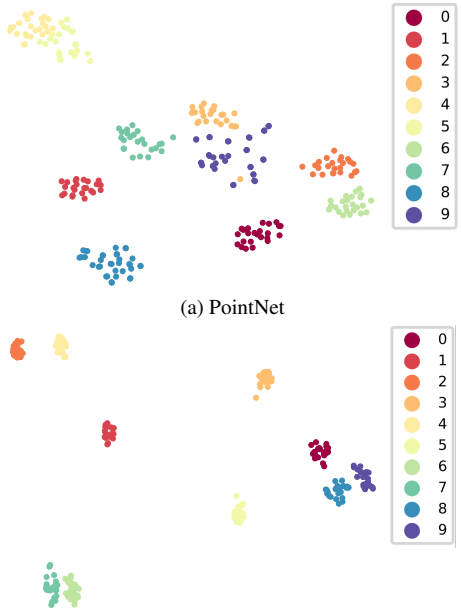


Figure 10. The exemplar depths and silhouettes from LiDAR and camera modality, where LiDAR modality exists misalignment issue when the probe carries an umbrella.

D. Qualitative results

To analyze the performance gap between our LidarGait and representative PointNet, We visualize the feature distribution on the SUSTech1K dataset. We can observe that LidarGait can capture features with clear discrimination. As shown in Fig. 11b, LidarGait prominently learns the inter-class margin and makes the intra-class distribution more compact. However, the representative point-wise model, PointNet, can only obtain global features as shown in Fig. 11a. PointNet captures features with less discrimination, and its intra-class features distribute sparsely.



(a) PointNet

(b) Our LidarGait

Figure 11. Feature distributions are visualized by t-SNE.

E. Evaluation if Variants as Gallery

As reported in Tab. 3, we evaluate when sequences in variation conditions are gallery sets with normal conditions as probe sets. We observe the performance degradation when normal cases are in the role of probes.

Table 3. Performance comparison of different evaluation protocals.

Evaluation Protocol	Gallery: *normal* Probe: ***variation***		Gallery: ***variation*** Probe: *normal*	
	Rank1	*Rank5*	*Rank1*	*Rank5*
Camera	76.12	89.39	74.84	89.28
LiDAR	86.77	96.08	84.31	94.84

F. Exemplar Sequences of SUSTech1K

To demonstrate the necessity of the SUSTech1K dataset, We show several exemplar sequences of the SUSTech1K dataset under normal, occlusion, and poor illumination conditions in Fig. 12 - 14.

Fig. 12 shows that LiDAR provides informative geometry as significant cues that extend gait recognition from 2D to 3D space. The most considerable advantage of LiDAR for gait recognition is that it allows for perspective from another viewpoint, as shown in the bottom row of Fig. 12.

When the pedestrians are occluded, as shown in Fig. 13, the silhouettes obtained by segmentation methods are typical with lower quality. The conventional segmentation methods are based on 2D cameras, but humans live in 3D space, making it difficult to separate the off-the-interest pedestrian from 2D space. LiDAR with precise 3D information can obtain high-quality gait representation under the condition of occlusion.

When the pedestrians are occluded, as shown in Fig. 13, the silhouettes obtained by segmentation methods are typical with lower quality. The conventional segmentation methods are based on 2D cameras, but humans live in 3D space, making it difficult to separate the of-the-interest pedestrian from 2D space. With precise 3D information, LiDAR can obtain high-quality gait representation under occlusion.

In Fig. 15, we show gait representations in existing in-the-wild datasets, GREW and Gait3D. We can observe that failure gait representations commonly exist because of various factors in the real world. Therefore, it is necessary to investigate a new way to obtain robust gait representation in such complex scenarios.

Figure 12. Exemplar sequences of SUSTech1K dataset under normal conditions. Eight frames in three modalities are visualized. The top three rows show three gait representations in RGB images, silhouettes, and point clouds. The bottom four rows represent gait representations in RGB images, silhouettes, front-view point clouds, and side-view point clouds. It shows LiDAR provides informative 3D geometry. (Best viewed in color.)

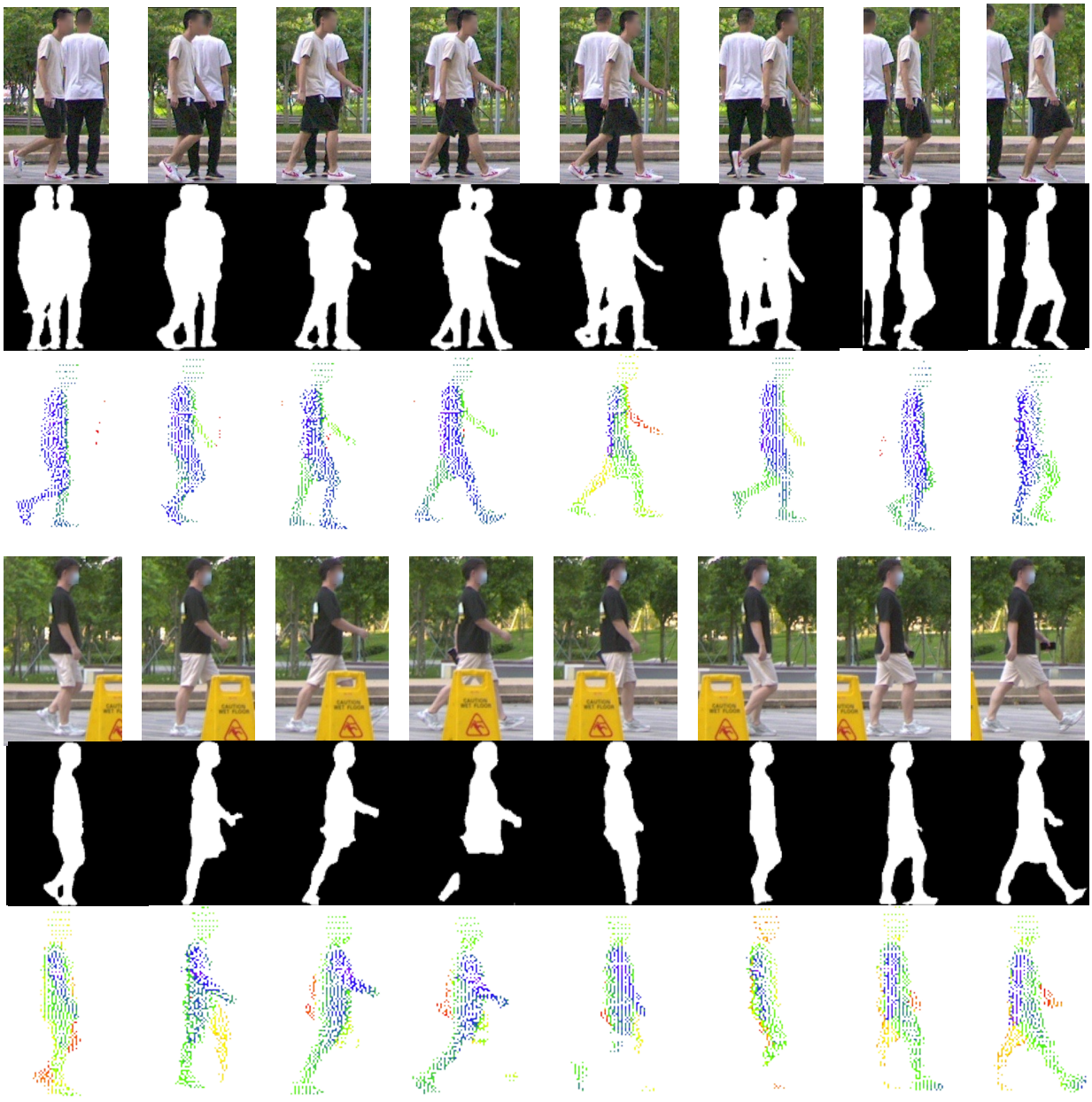


Figure 13. Exemplar sequences of SUSTech1K dataset under occlusions. The top three rows show gait representations when another subject overlaps the of-the-interest pedestrian. The bottom three rows show gait representations occluded by a static obstruction. It indicates that LiDAR can provide robust gait representations under occlusion conditions. (Best viewed in color.)

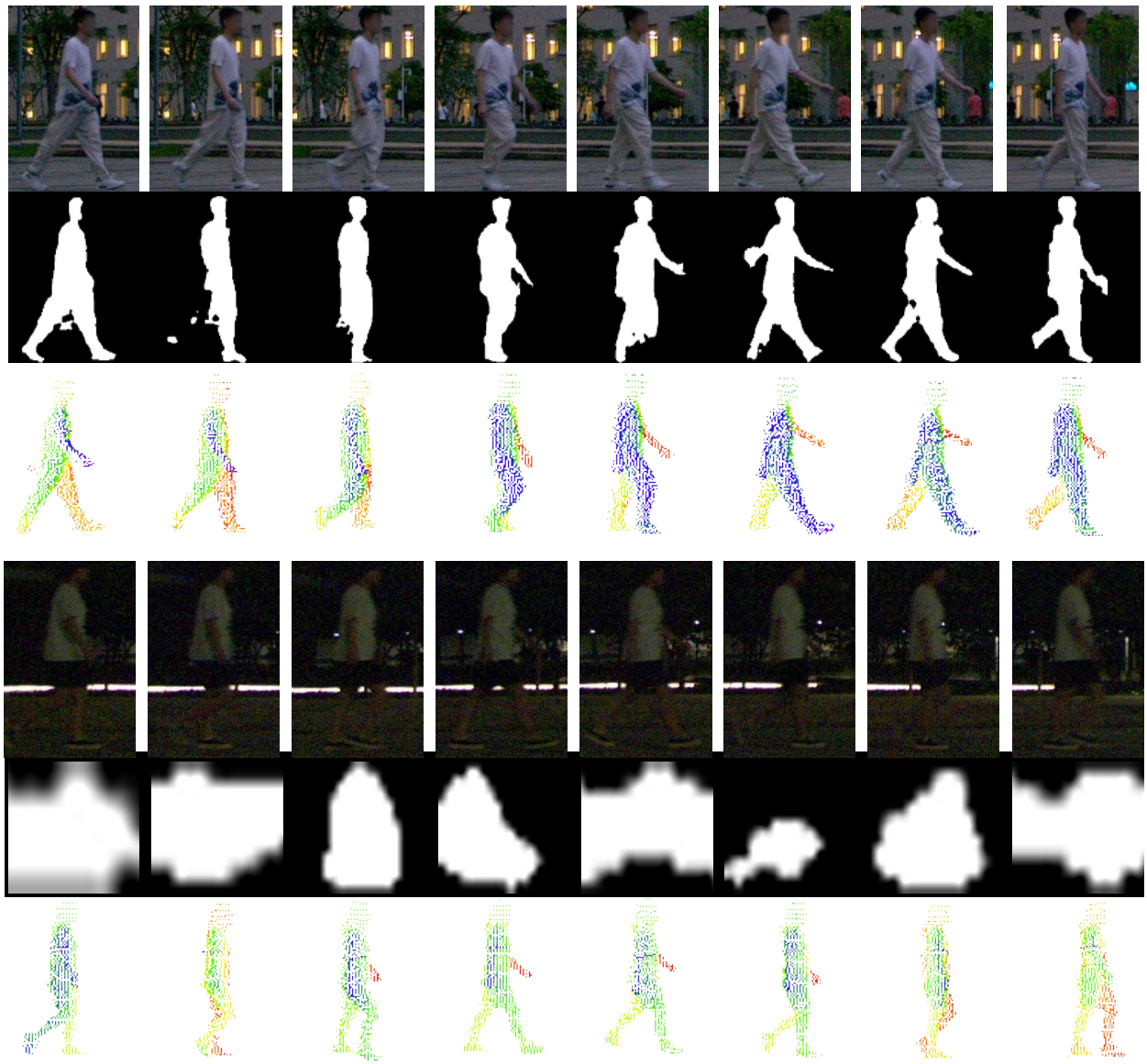
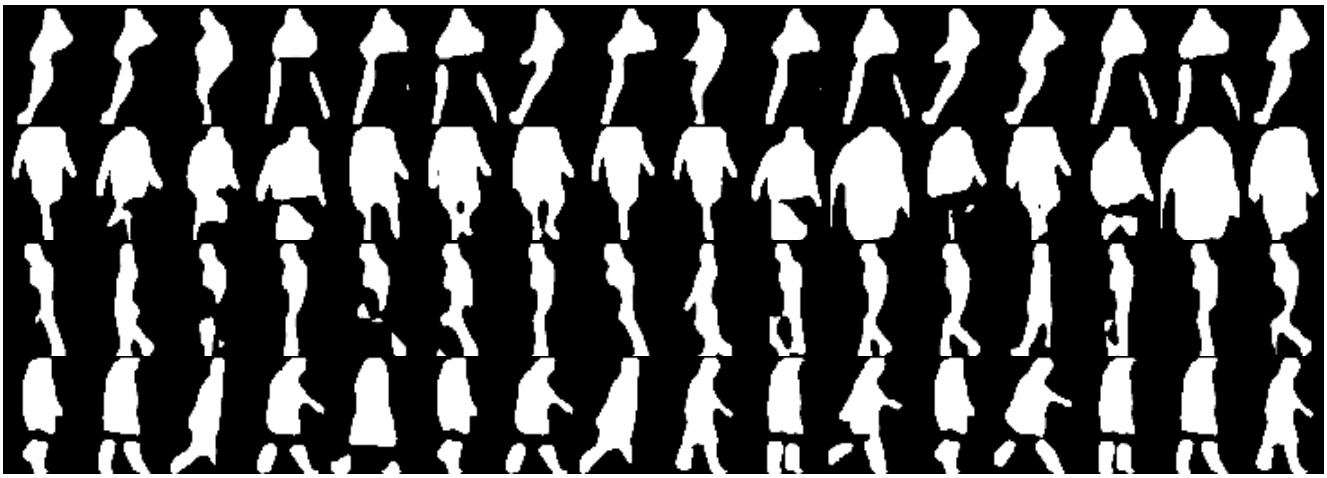


Figure 14. Exemplar sequences of SUSTech1K dataset under poor illumination. When illumination is extremely low, human segmentation is barely performed. In contrast, LiDAR provides robust gait representation with point clouds regardless of lighting. (Best viewed in color.)



(a) GREW

(b) Gait3D

Figure 15. Failure silhouettes in the existing in-the-wild datasets. The existing in-the-wild datasets face the issues that segmentation methods fail to provide gait representations with high quality by the effect of many real-world factors.