

FEATURE SELECTION FOR LIDAR-BASED GAIT RECOGNITION

Bence Gálai and Csaba Benedek

Institute for Computer Science and Control, H-1111 Budapest, Kende u. 13-17, Hungary
E-mail: lastname.firstname@sztaki.mta.hu

ABSTRACT

In this paper, we present a performance analysis of various descriptors suited to human gait analysis in Rotating Multi-Beam (RMB) Lidar measurement sequences. The gait descriptors for training and recognition are observed and extracted in realistic outdoor surveillance scenarios, where multiple pedestrians walk concurrently in the field of interest, their trajectories often intersect, while occlusions or background noise may affect the observation. For the Lidar scenes, we compared the modifications of five approaches proposed originally for optical cameras or Kinect measurements. Our results confirmed that efficient person re-identification can be achieved using a single Lidar sensor, even if it produces sparse point clouds.

Index Terms— Gait recognition, Lidar

1. INTRODUCTION

The use of gait as biometric feature has been extensively examined in the recent decades [1]. Recognizing people through the way they are walking is a highly advantageous approach in video surveillance systems, where the identification of non-cooperative subjects is required. Although various techniques have been already proposed on gait based person identification, their evaluation has been usually limited to strongly controlled indoor or outdoor environments considering illumination, background surfaces and background motions. In the widely used gait recognition benchmarks, such as the *CASIA gait dataset* [2], the *USF database* [3], or the *CMU Motion of Body (MoBo) Database* [4], typically a single pedestrian walks alone in a test video, which simplification facilitates gait print analysis and high quality discriminative feature extraction. On the other hand, in a surveillance scenario, the gait features should be observed in an arbitrary scene, where multiple pedestrians are concurrently present in the field, and they may partially occlude each other. System response should be real time, so that the person assignment is performed during the monitored event.

The clear majority of gait analysis techniques deal with video flows of standard optical cameras, or multi-camera systems. Since in realistic scenarios, it is usually not possible

to ensure that the test subjects walk along a straight line and laterally to the camera [5], establishing view-invariance introduces major challenges. A possible solution is recording training data for each pedestrian from multiple viewpoints [6], however this requirement cannot be fulfilled in a surveillance scene, where the pedestrians have to be recognized during their natural behavior. Another option could be taken by extracting 3D features from multi-camera configurations [7], however installation and continuous calibrations of such systems needs significant efforts, which might be a bottleneck by monitoring customized events.

Applying depth sensors appears as a natural idea for capturing accurate 3D information for gait recognition. The cheap and widely used Kinect sensors have already been investigated for gait analysis in a number of works [8, 9], and a corresponding gait database has already been published [10] for reference. However Kinects are still less efficient for surveillance applications due to their small Field of View (resolvable depth is between 0.8m – 4.0m), and the low quality outdoor performance of the sensor, especially in direct sunlight.

Velodyne's Rotating Multi-Beam (RMB) Lidar system is able to provide point cloud sequences from large outdoor scenes with a frame-rate of 15 Hz, 360° Field of View (FoV), point cloud sizes of around 65K points/frame with a maximal radius of 120m. The RMB Lidar sensor does not need any installation or calibration process after being placed into a new environment. However, the spatial density of the point cloud is quite sparse, showing a significant drop in the sampling density at larger distances from the sensor, and we can also see a ring pattern with points in the same ring much closer to each other than points of different rings. According to our measurement, the point cloud size of a person in a courtyard with 10-20m radius varies between 200-400 points, which is *two orders of magnitude* smaller than the figures of Kinect (10-20K points/person), and also significantly lower than the density of the stereo camera measurements from [11].

In this paper, we investigate the abilities of the RMB Lidar sensor for visual gait analysis, supporting pedestrian re-identification in realistic surveillance environments. Although pedestrian detection and tracking tasks have already been conducted on RMB Lidar measurements [12, 13], to our best knowledge our research [14] has been the first at-

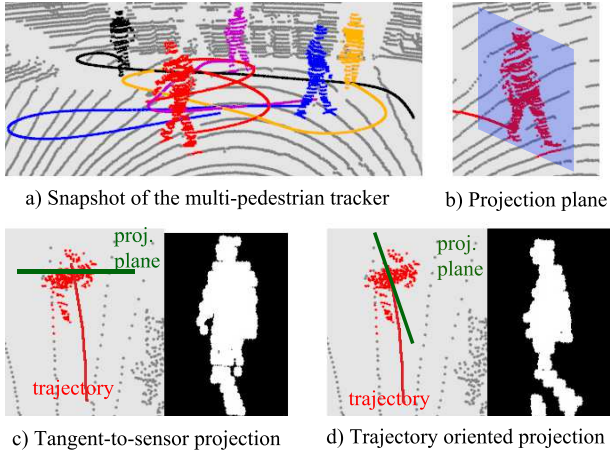


Fig. 1. Silhouette extraction and projection

tempt to involve such sensors in gait recognition. While we demonstrated the relevancy of Lidar based features in [14] at a proof-of-concept level, we focus here on comparing features used in earlier techniques, which have originally been proposed for gait analysis in optical or Kinect data. By each selected method, we had to explore first how their expected input feature maps can be derived from RMB Lidar streams. All methods have been quantitatively evaluated on the pedestrian dataset of [12], which contains Velodyne measurement sequences from multiple concurrently walking people in a courtyard scene.

2. MODEL FRAMEWORK

The proposed gait recognition process is embedded into the Lidar-based surveillance system [12]¹. The framework is able to automatically detect and track multiple moving pedestrians based on measurement sequences of a RMB Lidar sensor, which monitors the Field of Interest (FoI) from a fixed ground position. As output, the 2D ground trajectories of the pedestrians are provided (Fig. 1(a)). However in the basic system [12] several critical issues were related to broken trajectory segments during the person tracking process, caused by frequent occlusions between the people in the scene, or simply by the fact that the pedestrians may temporarily leave the FoI.

In this paper, we compare various 2D silhouette based gait analysis approaches for person re-identification. For ensuring viewpoint invariant features, we interpolate the side view projections of the 3D human silhouettes in the RMB Lidar point clouds (Fig. 1(b)). Since the FoV of the Velodyne sensor is circular, a straightforward projection plane at a given ground position could be taken as the local tangent of the circle around the sensor position, however as Fig. 1(c) demonstrates, this choice would not ensure side view approxima-

¹Demos: <http://web.eee.sztaki.hu/i4d/demos.html>

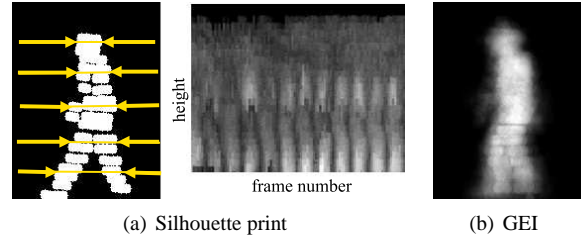


Fig. 2. Extracted features of the (a) Silhouette print [6] and (b) GEI [14, 15] techniques on Lidar data

tions of the silhouettes. Instead, we exploit an assumption that the people walk forwards in the scene, always turning towards the tangent direction of the trajectory. At each time frame, we project the point cloud segment of each person to the plane, which intersects the actual ground position, it is perpendicular to the local ground plane, and it is parallel to the local tangent vector of the Fourier-smoothed trajectory from top view (Fig. 1(d)). The projected point cloud consists of a number of separated points in the image plane, which can be transformed into a connected 2D foreground region by morphological operations (Fig. 1(c)(d)).

3. FEATURE SELECTION

For comparison, we implemented five different model-free silhouette based approaches for our Lidar-based surveillance framework. The first four techniques are Lidar-focused applications of state-of-the-art approaches, proposed earlier for standard optical and Kinect data, while the fifth method is an improved modification of our model from [14].

3.1. Silhouette print

Kale et al. [6] used the width of the outer contour of a binarized silhouette as the basic feature. In this method, a bounding box is placed around the extracted silhouette patch, which is divided into D equal box-parts along the vertical axis. Then the width of the silhouette is stored in each box-part, yielding a D dimensional (used $D = 20$) width-vector at a given time frame. The width-vectors of consecutive frames are combined in an image called silhouette print (SP) image. Brighter pixels refer to larger values in the width vectors. Similarities between the prints are calculated using dynamic time warping (DTW) algorithm [6].

Before starting the evaluation in our Lidar dataset, we validated our implementation on the original CMU MoBo [4] (optical) database, and reproduced similar efficient results to [6]. Thereafter, the adaptation of the method to the more challenging Lidar-scene has been straightforward: we generated 5 prints for every person for gallery (training) data, and during the re-identification step we have chosen the person, whose

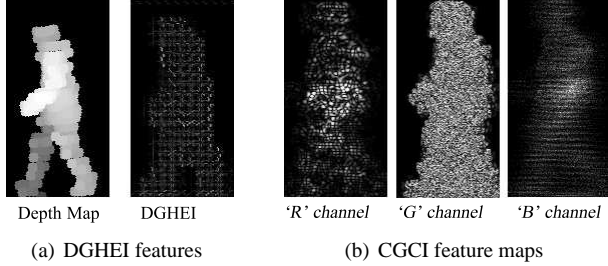


Fig. 3. Feature maps by the (a) DGHEI [16] and (b) CGCI [9] techniques on Lidar data

galleries showed in average the lowest DTW distance from the current probe (test) data.

3.2. Gait Energy Image with Vector Comparison

The Gait Energy Image (GEI) was introduced by Han and Bhanu in 2006 [15] for conventional optical video sequences. GEIs were derived by averaging the binary person silhouettes over the gait cycles:

$$G(x, y) = \frac{1}{T} \sum_{t=1}^T B_t(x, y) \quad (1)$$

where B is the binary silhouette and T is the number of collected frames. Principle Component Analysis (PCA) and Multiple Discriminant Analysis (MDA) were applied to the images for dimension reduction, then the Euclidean distances between the gallery and probe feature vectors were calculated to decide if the GEIs are derived from the same person.

3.3. Depth Gradient Histogram Energy Image

Following the averaging idea of the Gait Energy Image, the Depth Gradient Histogram Energy Image (DGHEI) [16] calculates depth gradients with histogram binning, then it averages the histogram bins of a full gait cycle. PCA+MDA and nearest neighbor classifier were used for dimension reduction and classification. Instead of the binarized silhouettes [16] used depth images captured with a Kinect sensor.

3.4. Color Gait Curvature Image

Tang et al. [9] used point clouds captured with Kinect sensors for gait recognition. They used Gaussian curvature and mean curvature to extract 2.5D gait features and mapped these features and the density of point clouds to a 3-channel *RGB* image. Next, they applied 2D Discrete Cosine Transform and 2D-PCA to the R, G and B components separately. Classification was achieved by calculating a weighted sum of the absolute differences in the R, G and B feature components.

Scenario	SP-DTW	GEI+VC	DGHEI	CGCI	GEI+NN
<i>winter4</i>	0.96	0.98	0.97	0.36	1.00
<i>winter6</i>	0.33	0.83	0.89	0.27	0.85
<i>spring6</i>	0.64	0.94	0.81	0.32	0.95
<i>spring8</i>	0.33	0.57	0.59	0.20	0.70
<i>summer5</i>	0.39	1.00	0.59	0.40	0.99
<i>summer6</i>	0.33	0.67	0.83	0.29	0.79
<i>All</i>	0.50	0.83	0.78	0.31	0.88

Table 1. Evaluation results of the five methods: rates of correct re-identification.

3.5. Gait Energy Image with Neural Networks

As a modification of the original GEI approach, we introduced in [14] the Lidar-based GEI technique, which trained a multilayer perceptron (MLP) for each detected person using the extracted GEI prints. As an improvement, we propose here using an ensemble of a single convolutional neural network (CNN) and a MLP, both ones having N outputs, where N is equal to the number of people in the training scenario, and each output value represents a person. The CNN inputs are the raw 2D GEIs, while the feature vectors of MLP have undergone dimensionality reduction through PCA and MDA compression. For a training sample, the output of the represented person is labeled as 1, the other outputs as -1 . In the test phase, the trained network produces outputs within the range $o \in [-1, 1]$, and any $o > 0$ positive match is considered a successful identification.

4. EXPERIMENTS AND DISCUSSION

We have tested the above listed five methods in six outdoor sequences captured in a courtyard by a Velodyne HDL 64-E RMB Lidar sensor (sequence names in Table 1 refer to the date of capturing and the number of people in the scene) [12, 14]. In all the sequences, the test subjects circulate in the scene, then they leave the FoV for a while, and re-appear later in a different order. The goal is to match the corresponding gait patterns collected in the first (*training*) and second (*test*) parts of each test scenario.

All the methods (except the silhouette print) were trained using 100 gallery (training) feature maps for each person, extracted from the *training* parts of the sequences. In the evaluation phase, we generated 200 probe (test) maps of each test subject from the *test* segments of the videos. Each probe sample was independently matched to the trained person models, thus we used $200 \cdot N$ test samples in a scenario with N people. For evaluating the performance of the different methods, we calculated the rate of the correct identifications among all test samples, and listed the obtained results in Table 1.

As we can observe in Table 1, the CGCI method showed the most significant drop in performance due to the low point

density of the Lidar sensor. While CGCI has been efficient with Kinect measurements [9], the much sparser Velodyne point clouds caused a notably degraded efficiency of the approach.

By testing the width-vector based SP+DTW approach [6], we experienced that it only favored the first test scene (*winter4*), which included nearly complete silhouettes with noiseless contours. However as the quality of silhouettes decreased due to frequent occlusions, and several holes and discontinuities appeared in more crowded tests scenes, the SP+DTW approach provided quite low recognition rates.

The Depth Gradient Histogram Energy Image (DGHEI) [16] proved to be the second best gait descriptor, outperformed only by the two GEI based methods. The key point in the operation of DGHEI is extracting high-quality depth images. As Fig. 3(a) demonstrates, with the 200-400 points of each person, we can already get a clean depth map, and we expect further quality improvements of the approach if newer generations of Lidar sensors appear in the future.

We found GEI as the most efficient descriptor suited for the Velodyne Lidar measurements due to its robustness: efficient GEIs could be extracted even under challenging circumstances with the presence of low quality or partially missing silhouette parts. However, based on our experiences, we have modified the original approach of [15] at a number of points. To increase the robustness against occlusion and low quality measurement segments, we did not rely on preliminary gait cycle extraction as proposed in [15]. Instead, we selected random seed frames from the recorded gait sequence, and averaged the upcoming 80 consecutive frames to get the person’s GEI. In challenging scenes, such as the ‘crowded’ *spring8* and *summer6* scenarios, the Vector Comparison [15] step did not provide a robust recognition result, here the neural network based solution yielded more efficient performance.

For reducing further artifacts caused of frequent occlusions, we also developed a frame selection algorithm for the best performing GEI+NN approach. A binary mask was created by summing and thresholding the consecutive silhouettes for every person (Fig. 4(a)). For every silhouette we calculated its internal and external area w.r.t. the mask. If the internal area was less then the 40% of the mask’s area (Fig. 4(b)) or the area external area was more then the 30% of the mask’s area (Fig. 4(c)) the frame was discarded for GEI calculation. As a result, we attempted to use the relevant frames only (Fig. 4(d)). In Table 2, we provide comparative recognition rates for the GEI+NN method in four configuration, showing the improvements of the trajectory oriented projection (see column R, and also Fig. 1), and the frame selection filter (F). Note that the R step has already been used in the tests reported in Table 1, since it was not related directly to GEI generation.

The measured computational time requirements of the main steps for the different approaches are listed in Table 3. Although the training of the SP+DTW approach is significantly quicker than the other references, the recognition

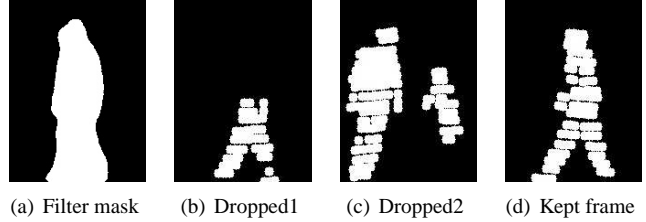


Fig. 4. Demonstration of the automatic frame selection step proposed for the GEI+NN method (not used in Table 1)

Scenario		R	F	F + R
<i>winter6</i>	0.78	0.85	0.81	0.95
<i>spring6</i>	0.80	0.95	0.81	0.98
<i>summer5</i>	0.99	0.99	1.00	1.00
<i>summer6</i>	0.75	0.79	0.83	0.95
<i>average</i>	0.83	0.90	0.86	0.97

Table 2. Performance improvements of rotating (R), filtering (F) and both steps (R+F) using the GEI+NN method.

part is slower due to running DTW comparison between the probe sample and all stored gait print samples. The GEI+NN approach needs relatively significant time for training set generation and NN training, however the recognition step is still very efficient: less then 0.01sec/probe sample.

5. CONCLUSION

We showed that various silhouette based gait recognition techniques can efficiently adopted to the measurement sequences of a RMB Lidar sensor. The methods were tested on challenging simulated surveillance scenarios, and the Gait Energy Image + Neural Network solution proved to be the most robust against degraded silhouette shapes.

Acknowledgment: This work was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences

Method	Training set generation	NN training phase	Recognition of 100 test samples
SP+DTW	8.43	–	43.7
GEI+VC	142.7	–	0.39
DGHEI	110.3	–	0.98
CGCI	108.9	–	0.26
GEI+NN	142.7	46.9	0.98

Table 3. Computational time (in seconds) of the main steps of the different approaches for the Lidar based gait database

6. REFERENCES

- [1] Z. Zhang, M. Hu, and Y. Wang, "A survey of advances in biometric gait recognition," in *Biometric Recognition*, vol. 7098 of *Springer LNCS*, pp. 150–158. 2011.
- [2] S. Zheng, J. Zhang, K. Huang, R. He, and T. Tan, "Robust view transformation model for gait recognition," in *International Conference on Image Processing (ICIP)*, 2011.
- [3] S. Sarkar, P.J. Phillips, Z. Liu, IR. Vega, P. Grother, and K.W. Bowyer, "The humanID gait challenge problem: data sets, performance, and analysis," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 2, pp. 162–177, Feb 2005.
- [4] R. Gross and J. Shi, "The CMU Motion of Body (MoBo) Database," Tech. Rep. CMU-RI-TR-01-18, Robotics Institute, Pittsburgh, PA, June 2001.
- [5] Y. Li, Y. Yin, L. Liu, S. Pang, and Q. Yu, "Semi-supervised gait recognition based on self-training," in *International Conf. Advanced Video and Signal-Based Surveillance (AVSS)*, Beijing, China, Sept 2012, pp. 288–293.
- [6] A. Kale, N. Cuntoor, B. Yegnanarayana, A.N. Rajagopalan, and R. Chellappa, "Gait analysis for human identification," in *Audio- and Video-Based Biometric Person Authentication*, Josef Kittler and MarkS. Nixon, Eds., vol. 2688 of *Lecture Notes in Computer Science*, pp. 706–714. Springer Berlin Heidelberg, 2003.
- [7] R. Bodor, A. Drenner, D. Fehr, O. Masoud, and N. Papanikolopoulos, "View-independent human motion classification using image-based reconstruction," *Image Vision Comput.*, vol. 27, no. 8, pp. 1194–1206, July 2009.
- [8] M. Gabel, E. Renshaw, A. Schuster, and R. Gilad-Bachrach, "Full body gait analysis with kinect," in *Proceedings of EMBC 2012*. August 2012, Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).
- [9] J. Tang, J. L., T. Tjahjadi, and Y. Gao, "2.5D multi-view gait recognition based on point cloud registration," *Sensors*, vol. 14, no. 4, pp. 6124, 2014.
- [10] M. Hofmann, J. Geiger, S. Bachmann, B. Schuller, and G. Rigoll, "The TUM gait from audio, image and depth (GAID) database: Multimodal recognition of subjects and traits," *J. Vis. Comun. Image Represent.*, vol. 25, no. 1, pp. 195–206, Jan. 2014.
- [11] J. Ryu and S. Kamata, "Front view gait recognition using spherical space model with human point clouds," in *IEEE International Conf. Image Processing (ICIP)*, Brussels, Belgium, Sept 2011, pp. 3209–3212.
- [12] C. Benedek, "3D people surveillance on range data sequences of a rotating Lidar," *Pattern Recognition Letters*, vol. 50, pp. 149–158, 2014, Special Issue on Depth Image Analysis.
- [13] L. Spinello, M. Luber, and K.O. Arras, "Tracking people in 3D using a bottom-up top-down detector," in *IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, China, 2011, pp. 1304–1310.
- [14] C. Benedek, B. Nagy, B. Gálai, and Z. Jankó, "Lidar-based gait analysis in people tracking and 4D visualization," in *European Signal Processing Conference*, Nice, France, 2015.
- [15] J. Han and B. Bhanu, "Individual recognition using gait energy image," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 28, no. 2, pp. 316–322, Feb 2006.
- [16] M. Hofmann, S. Bachmann, and G. Rigoll, "2.5D gait biometrics using the depth gradient histogram energy image," in *IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, Sept 2012, pp. 399–403.