GRIDDS - A Gait Recognition Image and Depth Dataset

João Ferreira Nunes[1,2][0000−0002−5204−4043], Pedro Miguel Moreira[1][0000−0001−8371−0347], and João Manuel R. S. Tavares[2][0000−0001−7603−6526]

[1] ARC4DigiT - Centro de Investigação Aplicada para a Transformação Digital, Instituto Politécnico de Viana do Castelo, PORTUGAL
{joao.nunes, pmoreira}@estg.ipvc.pt
[2] Instituto de Ciência e Inovação em Engenharia Mecânica e Engenharia Industrial, Departamento de Engenharia Mecânica, Faculdade de Enhenharia, Universidade do Porto, Porto, PORTUGAL
tavares@fe.up.pt

Abstract. Several approaches based on human gait have been proposed in the literature, either for medical research reasons, smart surveillance, human-machine interaction, or other purposes, whose validation highly depends on the access to common input data through available datasets, enabling a coherent performance comparison. The advent of depth sensors leveraged the emergence of novel approaches and, consequently, the usage of new datasets. In this work we present the GRIDDS - A Gait Recognition Image and Depth Dataset, a new and publicly available gait depth-based dataset that can be used mostly for person and gender recognition purposes.

Keywords: Gait Dataset · Person Recognition · Gender Recognition · RGB-D Sensors · GRIDDS.

1 Introduction

Human gait and its underlying dynamics can reveal relevant information for a manifold of applications. For example, gait can be used either as an indicator of a person's health condition [18,7], or to reveal their state of mind [12], or, in another context of use, it may work as a biometric feature, enabling the identification of individuals that are under observation, based on their individual walking styles. In the latest example, when compared to other classical biometric traits, like finger-print, face, iris or retina, gait reveals to be more advantageous in some aspects, since gait can be captured at a distance, through non-intrusive technologies, without the implicit need of the individuals' collaboration [9,17].

The great majority of the dedicated work to this field of research is based on image sequences and on their intrinsic features in order to extract gait characteristics [8,2,19]. However, during the last decade, the dissemination and availability of low cost RGB-D sensors like Microsoft Kinect, Intel RealSense or Asus Xtion,

among others, prompted the development of highly-improved 3D vision systems, as well as the development of new approaches and novel applications. These sensors, with a built-in infrared camera providing additionally depth information, are capable to track in real time at least one human figure and also to extract the coordinates of a set of points, which mostly correspond to human body joints, forming a schematic representation of the human skeleton, without using any kind of markers attached to the human body. Thus, in order to derive gait features, some of the more recently introduced approaches are using the estimated skeleton structure by means of the depth data, [20,21,22].

At the time that we conducted some preliminary work based on human gait for person and gender recognition, we noticed that the number of publicly available datasets including depth data was reduced, as demonstrated in [14]. We also concluded that the existing datasets either did not provide all the sensors' collectable data (color images, depth information, joints' coordinates, etc.), or that they had been acquired with an older version of the sensor that we had (Microsoft Kinect v1 versus Microsoft Kinect v2), varying in data resolution and precision, as well as in the total number of tracked joints. For that reason, we decided to develop a new dataset: the GRIDDS (Gait Recognition Image and Depth Dataset), whose potential applicability is in person and gender recognition.

2 Related Datasets

Most of the published work that is based on human gait uses image sequences of people walking in order to extract a set of features so that gait can be properly characterized. These approaches brought the need for public video-based datasets of people walking under different conditions (e.g.: indoor/outdoor environments, single/multiple views, clothing and carrying variations, etc.). The usage of common input data, available in such datasets, enables a coherent comparison of different approaches and gives an insight into the capabilities of respective methods. A detailed description of some of the must used video-based datasets can be found in [19] and in [15].

More recently, new gait-based methods have been proposed. These methods use the depth information beyond the objects represented in images, both provided by the RGB-D sensors. Evidently the validation of these new approaches also benefits from the access to common input data by means of public datasets. However, the majority of the existing datasets do not include depth information, or if any, they have been developed with a previous version of the RGB-D sensor that we had at the time, with lower performances in people tracking [23]. Thus, it became necessary to create new datasets that would also include depth information, and, depending on the sensor used, include also some additional data, like the joints' coordinates of the human figures detected and tracked in the scene. In fact, the body-skeleton stream provided by some of the depth sensors has been proven to be significantly accurate [3,4] and it has been used in robust gait-based recognition systems, as evidenced by the number of devoted published

studies in the recent years[6,16]. A review of existing depth-based datasets has been presented in [5], and [14].

When compared to other public depth-based datasets, GRIDDS has the advantage that it was acquired using the latest version of Kinect sensor, thus tracking a bigger number os joints, with grater data resolution and precision. In addition, it includes a greater variety of avalilable data (*color* images, *depth* images and *depth* data, *infrared* images, *joints' coordinates* and the corresponding *timestamps* of each captured frame).

3 GRIDDS

The GRIDDS - Gait Recognition Image and Depth Dataset was recorded at the Polytechnic Institute of Viana do Castelo (IPVC) facilities, in June of 2018. The dataset is publicly available online at [13], and its usage is allowed according to the instructions described on the same web page.

3.1 Participants

For the development of this dataset we had the collaboration of 35 volunteers, among students, teachers and staff from the IPVC. A written informed consent was obtained from all subjects prior to their participation. Besides the recorded walking sessions, some additional data was also collected, including the participant's age, height and gender. This information is also available online and it is summarized in Table 1.

Table 1. Description of the 35 volunteers in numbers, according to their gender, age and height.

	Gender	Age		Height	
		Mean	SD	Mean	SD
♂	11	29.2	9.7	178.2	6.4
♀	24	39.0	10.4	163.0	7.4
Total	35	35.9	11.2	167.8	10.0

3.2 Depth Sensor and Data Specifications

The sensor used to collect the data was the Kinect v2 (also known as the Xbox One Kinect), manufactured by Microsoft. The collected data modalities included the color, depth, infrared and body streams, and their corresponding timestamps for each captured frame. As stated by the sensor specifications, all data modalities were acquired at an approximately frame rate of 30 fps, varying in their content, but also in their format and resolution: both the depth and infrared

images have a resolution of 512×424 pixels, while the color images have a resolution of 1920×1080 pixels. The body data, which was inferred by the sensor's SDK, consists in 2D and 3D coordinates of 25 body points, which mostly correspond to human body joints that are detected and tracked in the scene. The 2D coordinates correspond to the body joints' coordinates on the color images, having its origin (x=0, y=0) located at the top-left corner of every image and each unit corresponds to one pixel. The 3D coordinates are referred to the coordinate system used by the Kinect, whose origin (x=0, y=0, z=0) is located at the center of the sensor, where each unit value corresponds to one meter. The timestamps for every captured stream are expressed in the time unit returned by the Kinect — in order to estimate the time passed between two frames (f_i and f_k, where $k \geq i$), we can refer to the Equation (1).

$$time_{\langle f_i, f_k \rangle} = (timestamp_{fk} - timestamp_{fi})/10000000 \quad (1)$$

3.3 Environment and Data Acquisition Description

The recording sessions occurred in a controlled indoor environment, with a static background and with both natural and artificial lighting. Two trajectories were defined, in a straight line across the room: one starting from the left side of the room to the right side, and the other on the opposite direction. The sensor, supported on a tripod, was fixed at 1.8 meters high, perpendicular to the defined trajectories (see Figure 1). The gray triangle represents the sensor's range, according to its technical specifications provided by the manufacturer.

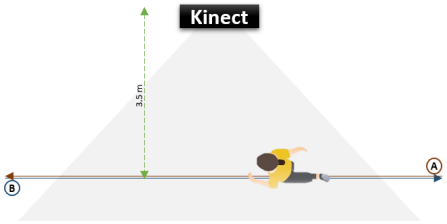

Fig. 1. A graphical representation of the environment where the recording sessions occurred. Letters A and B correspond to the two defined trajectories .

Each one of the 35 volunteers completed 5 walking sequences per trajectory, at a distance of approximately 3.5 meters from the sensor, resulting in 10 sequences per participant, and a total of 350 recorded sequences.

3.4 Data Availability

The dataset is composed by 35 folders (one per participant), each one containing the following collected data: *color* images, *depth* images and *depth* data, *infrared*

images, *joints' coordinates* and the corresponding *timestamps* of each captured frame of the ten recorded sessions per participant. Additionally, we included the *body silhouettes* images, cropped, facing all to the same side, and normalized in size, with a resolution of 80×120 pixels. The information that is made available inside of each folder, is in either one of the following two formats:

- `vvv_ss_stream_nnn.fmt`, for the *color*, *depth*, *silhouette* and *infrared* streams;
- `vvv_ss_stream.fmt`, for the *timestamp* and *joints' coordinates* streams;

where *vvv* corresponds to the volunteers' id, *ss* to the session number, *stream* to the different available streams, *nnn* to the frame number and *fmt* to the different file formats (PNG or CSV). For example, the file named `003_09_depth_021.csv` corresponds to the '*depth*' stream from the volunteer number '003', captured during session number '09', at the frame number '21', saved in the 'CSV' file format. All image files are in the PNG format, varying only in the bit-depth color information, where the color images are in 24-bit, depth images in gray-scale 16-bit, body silhouettes in 1-bit and the infrared images in 16-bit. The depth data files (which are in CSV format) have the same resolution as the depth images, however, in this case, each cell contains the distance between the Kinect device and the objects in front of the device, in millimeters. The coordinates files are also in CSV format and have a resolution of $6 \times N_{frames} \times N_{joints}$, where N_{frames} is the number of captured frames, N_{joints} is the number of tracked joints (25 joints) and the 6 columns correspond to: the frame number, the 3D coordinates and the 2D coordinates of the joints, both in meters. Figure 2 illustrates some of the extracted data during a recorded session.

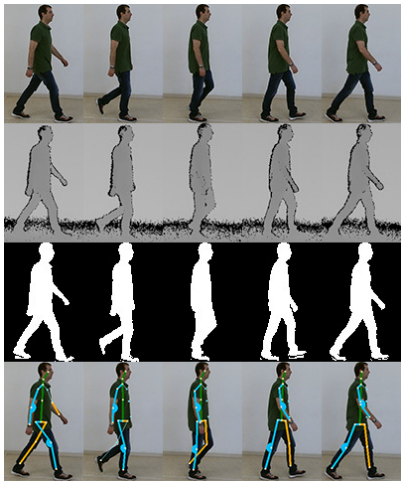

Fig. 2. Examples of normalized and aligned captured streams. First row: color images; Second row: depth images; Third row: body silhouettes; Fourth row: skeleton representation on top of color images.

Furthermore, we include the `body-viwer` tool [1] that we developed to visualize the gait sequences with a graphical representation of all tracked joints and '*bones*', as well as a representation of the angles between hip-knee-ankle joints and shoulder-elbow-wrist joints from the body side closest to the sensor.

3.5 Summary

In conclusion, GRIDDS can be briefly described in form of a table (see Table 2), based on set of characteristics, namely: (i) its applicability, referring to potential application fields (person recognition (PR), gender recognition (GR)); (ii) number of subjects that participate in the recorded sessions, indicating also their distribution in respect to gender; (iii) type of sensor used and how it was placed on scene; (iv) number and type of defined trajectories; (v) number of sequences recorded per participant; (vi) list of the collected data; and (vii) list of additional data provided (e.g.: source-code or applications to manipulate data, documentation, etc.).

Table 2. Proposed framework to summarily describe our dataset.

identification	year	applicability[1]	#subjects	#sensors	#trajectories[2]	#sequences/subject[2]	collected data[3]	extra data
GRIDDS	2018	PR + GR	35 11♂+24♀	kinect v2 fixed at 1.8m high	2 S	10 (5S+5S)	C, D, SK, T, DS, S	start-end frame/gait cycle + matlab scripts

1 PR: Person Recognition; GR: Gender Recognition.
2 S: Side (Left-to-Right and/or Right-to-Left).
3 C: RGB data; D: Depth data; SK: 3D Skeleton Coordinates; T: Time; S: Silhouettes; DS: Depth Silhouettes.

4 Application Examples

In order to demonstrate some of the potential usefulness of GRIDDS, we have conducted two sample applications. The first one consists in extracting valid sequences of gait cycles, based on the joints' coordinates. The second is a demonstration of developing some state-of-the-art gait image representations, which are commonly used for person recognition purposes.

[1] available at `https://github.com/joaofnunes/gridds`

4.1 Gait Cycle Detection and Validation

Human gait is considered to be as a periodic activity and a single gait cycle can be regarded as the time passed between two identical events that occurred during the human walking sequence. The proposed method consists in detecting 'valid' sequences of one or more gait cycles by ensuring an effective feet side alternation. The method is exclusively based on the joints' coordinates, and regardless the availability of different approaches to extract joints' coordinates, for practical reasons we have decided to work with the coordinates provided by the sensor's SDK. Typically, there are two techniques for estimating gait cycles: the double support method, based on the local maxima, in which both legs are farthest from each other; and the mid-stance method, based on local minima, when both legs are in the rest (standing) position at the minimum distance from each other. In our proposal we used the double support phase to determine gait cycles. This validation can reveal to be quite useful, since the Kinect's tracking system may tend to confuse the left and right sides of the body joints, particularly when the sensor is placed perpendicular to the defined trajectories.

The first step consists in identifying all the double support positions of the walking sequence (i.e., when both feet are at a maximum distance from each other), knowing that three consecutive double support positions represent a gait cycle. Therefore, the Euclidean distance between ankle joints is computed, and local maxima (peaks) of the computed distance are identified, with a minimum separation criterion between peaks. Figure 3 illustrates the Euclidean distance signal between ankle joints (left side), and the same distance signal smoothed with a moving average filter with a fixed window length, determined heuristically (right side).

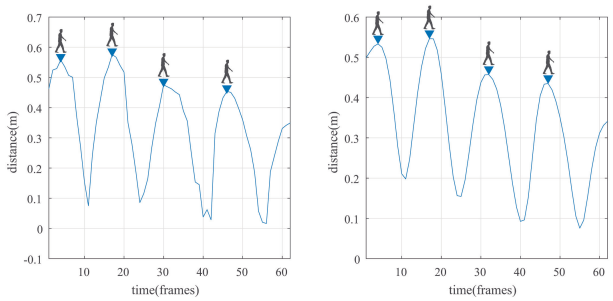

Fig. 3. Peaks detection in the the Euclidean distance between the ankle joints (left side), and a moving average filter applied to the same signal (right side).

Then, in order to verify that the side of the ahead ankle joint at each peak has been alternating (left-right-left or reverse), a characters sequence is build indicating the side of the ahead ankle joint at each detected peak (either 'L' for

left side or 'R' for right side). Whenever a "LRL" or "RLR" characters sequence is detected, it means that a valid gait cycle has been detected.

4.2 Gait Cycle Representations

In this experiment the depth images in form of silhouettes were used to create different state-of-the-art gait image representations, which are commonly used for person recognition purposes [8,11,24,10]. Firstly some additional image processing operations were needed, specifically: images' flipping, ensuring that the human silhouettes were all facing to the same side; and then the implementation of some basic morphological operations (e.g.: dilation and erosion). Then, following each of the selected five representations' descriptions, we developed the code necessary to generate each of the gait image representation, which is also available to download. Figure 4 illustrates the selected representations: Motion-Energy Image [2], Motion-History Image [2], Gait Energy Image [8], Active Energy Image [24] and Gait Entropy Image [1].

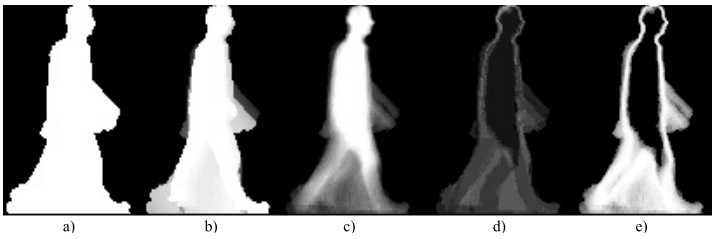

Fig. 4. Five different gait image representations: a) Motion-Energy Image [2], b) Motion-History Image [2], c) Gait Energy Image [8], d) Active Energy Image [24], and e) Gait Entropy Image [1].

All referred representations convert a sequence of silhouettes into a two-dimensional image, varying in the way the resulting image is computed. The first representation, Motion Energy Image, is a binary image describing where the motion has occurred in an image sequence. The second, Motion History Image, is a scalar-valued image where the intensity is a function of recency of motion. Both representations, together, can be considered as a two component version of a temporal template, a vector-valued image where each component of each pixel is a function of the motion at that pixel location. The third representation, Gait Energy Image, is one of the most adopted model free representations, representing gait in a single gray scale image obtained by averaging the silhouettes of a complete gait cycle. The fourth representation, Active Energy Image, has the advantage of reducing the influence of carrying or clothing conditions. By calculating the difference between two adjacent silhouettes in the gait sequence, it aims to extract only the active regions. The last representation, Gait Entropy

Image, captures mostly motion information, and that is why it is also robust to covariate condition changes that affect appearance (e.g.: clothing or carrying).

5 Conclusions and Future Work

The emergence of low cost depth sensors gave a new impetus to human motion studies. This was mainly due to the fact that these sensors enabled a relative accurate, real time, and markerless tracking, even without the collaboration or awareness of the people under study. At the same time, there was also a need for public data for the validation and performance comparison of newly proposed methodologies. Given the lack of public datasets containing more information, we decided to create our own data and make it public for the scientific community.Thus, we have developed a new depth-based gait dataset, publicly available, especially devoted for person and gender recognition purposes: the GRIDDS - A Gait Recognition Image and Depth Dataset. The dataset was acquired using the Microsoft Kinect v2 depth sensor, and unlike other datasets, it makes available all the collected data. Despite its contribution to the scientific community, we have identified some limitations of the dataset, which we hope to overcome some of them in a timely manner. One of them is related to the nonexistence of covariates. We expect to acquire more sessions in the near future, repeating some of the volunteers, thereby ensuring at least a variation in time and clothing. Another limitation is related to the reduced number of trajectories proposed, and for this reason we plan to gather new sessions in which participants move in front of the sensor, towards to it. And finally, another limitation identified is the scenario used for the recordings, however in this aspect we do not have much room for maneuver due to the limitation of the power supply to the sensor.

References

1. Bashir, K., Xiang, T., Gong, S.: Gait recognition using gait entropy image. In: Proceedings of the 3rd International Conference on Imaging for Crime Detection and Prevention. ICDP, IET (2009)
2. Bobick, A.F., Davis, J.W.: The recognition of human movement using temporal templates. IEEE Transactions on Pattern Analysis and Machine Intelligence **23**(3), 257–267 (2001)
3. Bonnechère, B., Jansen, B., Salvia, P., Bouzahouene, H., Omelina, L., Moiseev, F., Sholukha, V., Cornelis, J., Rooze, M., Jan, S.V.S.: Validity and reliability of the kinect within functional assessment activities: Comparison with standard stereophotogrammetry. Gait & Posture **39**(1), 593–598 (2014)
4. Clark, R.A., Pua, Y.H., Fortin, K., Ritchie, C., Webster, K.E., Denehy, L., Bryant, A.L.: Validity of the microsoft kinect for assessment of postural control. Gait & Posture **36**(3), 372–377 (2012)
5. Firman, M.: RGBD datasets: Past, present and future. In: Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition - Workshops. IEEE (2016)
6. Gabel, M., Gilad-Bachrach, R., Renshaw, E., Schuster, A.: Full body gait analysis with kinect. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 1964–1967. IEEE (2012)

7. Givon, U., Zeilig, G., Achiron, A.: Gait analysis in multiple sclerosis: Characterization of temporal–spatial parameters using GAITRite functional ambulation system. Gait & Posture **29**(1), 138–142 (2009)
8. Han, J., Bhanu, B.: Individual recognition using gait energy image. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(2), 316–322 (2006)
9. Kale, A., Sundaresan, A., Rajagopalan, A.N., Cuntoor, N.P., Roy-Chowdhury, A.K., Kruger, V., Chellappa, R.: Identification of humans using gait. IEEE Transactions on Image Processing **13**(9), 1163–1173 (2004)
10. Lam, T., Cheung, K.H., Liu, J.N.K.: Gait flow image: A silhouette-based gait representation for human identification. Pattern Recognition **44**(4), 973–987 (2011)
11. Liu, Z., Sarkar, S.: Simplest representation yet for gait recognition: averaged silhouette. In: Proceedings of the Int. Conf. on Pattern Recognition. IEEE (2004)
12. Michalak, J., Troje, N.F., Fischer, J., Vollmar, P., Heidenreich, T., Schulte, D.: Embodiment of sadness and depression—gait patterns associated with dysphoric mood. Psychosomatic Medicine **71**(5), 580–587 (2009)
13. Nunes, J.F., Moreira, P.M., Tavares, J.M.R.S.: Gridds - a gait recognition image and depth dataset (2018), `http://gridds.ipvc.pt`
14. Nunes, J.F., Moreira, P.M., Tavares, J.M.R.S.: Benchmark RGB-D gait databases: A systematic review. In: VipIMAGE 2019 / VII ECCOMAS Thematic Conference on Computational Vision and Medical Image Processing (2019)
15. Poppe, R.: A survey on vision-based human action recognition. Image Vision Computing **28**(6), 976–990 (2010)
16. Preis, J., Kessel, M., Werner, M., Linnhoff-Popien, C.: Gait recognition with kinect. In: Proceedings of the Int. Workshop on Kinect in Pervasive Computing (2012)
17. Reid, D.A., Samangooei, S., Chen, C., Nixon, M.S., Ross, A.: Soft biometrics for surveillance: An overview. In: Handbook of Statistics - Machine Learning: Theory and Applications, pp. 327–352. Elsevier (2013)
18. Salarian, A., Russmann, H., Vingerhoets, F.J.G., Dehollain, C., Blanc, Y., Burkhard, P.R., Aminian, K.: Gait assessment in parkinson's disease: Toward an ambulatory system for long-term monitoring. IEEE Transactions on Biomedical Engineering **51**(8), 1434–1443 (2004)
19. Sarkar, S., Phillips, P.J., Liu, Z., Vega, I.R., Grother, P.J., Bowyer, K.W.: The humanID gait challenge problem: Data sets, performance, and analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence **27**(2), 162–177 (2005)
20. Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., Blake, A.: Real-time human pose recognition in parts from single depth images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1297–1304. CVPR, IEEE Computer Society (2011)
21. Wang, J., Liu, Z., Wu, Y., Yuan, J.: Mining actionlet ensemble for action recognition with depth cameras. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. CVPR, (IEEE) (2012)
22. Wei, X., Zhang, P., Chai, J.: Accurate realtime full-body motion capture using a single depth camera. ACM Transactions on Graphics - Proceedings of ACM SIGGRAPH Asia **31**(6), 188:1–188:12 (2012)
23. Zennaro, S., Munaro, M., Milani, S., Zanuttigh, P., Bernardi, A., Ghidoni, S., Menegatti, E.: Performance evaluation of the 1st and 2nd generation kinect for multimedia applications. In: Proceedings of the IEEE International Conference on Multimedia and Expo. pp. 1–6. IEEE (jun 2015)
24. Zhang, E., Zhao, Y., Xiong, W.: Active energy image plus 2DLPP for gait recognition. Signal Processing **90**(7), 2295–2302 (2010)