

Walking gait dataset: point clouds, skeletons and silhouettes

Technical Report Number 1379

Trong-Nguyen Nguyen* and Jean Meunier†

DIRO, University of Montreal, Montreal, QC, Canada

September 8, 2018

Abstract

Vision techniques are currently playing an important role in healthcare systems. We present here a unique dataset that is helpful for studies on gait analysis. The dataset is expected to be appropriate for the scientific community to evaluate their algorithms. This report presents an overview of our dataset that contains normal (symmetric) walking gaits and 8 simulated abnormal (asymmetric) ones. The dataset was acquired by 9 volunteers and is represented in 3 different data types: point cloud, skeleton, silhouette.

1 Introduction

In recent years, computer vision has been playing an important role in many healthcare systems. For the task of gait analysis, data related to patients are usually acquired and represented as one of the following types: 2D silhouette, 3D skeleton, and 3D model (e.g. point cloud). An obvious drawback of silhouette in the task of gait analysis is the dependence on viewing direction. In other words, the loss of depth dimension significantly reduces the number of possible helpful features describing patient's poses. Therefore, recent studies focused on 3D information instead of a sequence of silhouettes. Skeleton is a popular choice to deal with gait analysis [1, 7]. However, a limitation of skeletal approaches is that the skeleton could be deformed due to self-occlusions of body parts. Such problems, unfortunately, usually occur when working on pathological gaits. In order to overcome that limitation, researchers attempted to employ 3D model as the input of their systems.

In our Image Processing Laboratory (DIRO, University of Montreal), we also focus on such 3D information but replace typical setup of multiple color cameras and/or depth sensors by a cheaper system consisting of a Time-of-Flight (ToF) depth camera (Microsoft Kinect 2 in our work) and two mirrors. The point clouds in our dataset were acquired according to that setup using our proposed algorithm [2, 4]. The acquisition also captured the two other data types (skeleton and silhouette) using provided functions in the camera SDK.

*nguyetn@iro.umontreal.ca

†meunier@iro.umontreal.ca

2 Data description

2.1 System setup

The setup of our system includes:

- a Microsoft Kinect 2 with ToF depth estimation scheme
- two flat mirrors placed behind the subject
- a treadmill placed at the center where each subject performs walking gaits

A depth image that captured the scene is presented in Fig. 1. The point clouds in our dataset were reconstructed based on the 3 subject’s depth maps [2, 4].



Figure 1: A depth image captured by our system (figure from [3]).

2.2 Dataset

There were 9 volunteers in our dataset, each subject performed 9 walking gaits with different levels of symmetry:

| Notation | Description |
|----------|--|
| N | normal walking gait |
| L 5cm | padding a sole with a thickness of 5 <i>cm</i> under left foot |
| L 10cm | padding a sole with a thickness of 10 <i>cm</i> under left foot |
| L 15cm | padding a sole with a thickness of 15 <i>cm</i> under left foot |
| L 4kg | attaching a weight (4 kilograms) to left ankle |
| R 5cm | padding a sole with a thickness of 5 <i>cm</i> under right foot |
| R 10cm | padding a sole with a thickness of 10 <i>cm</i> under right foot |
| R 15cm | padding a sole with a thickness of 15 <i>cm</i> under right foot |
| R 4kg | attaching a weight (4 kilograms) to right ankle |

Each walking gait was acquired in 1200 consecutive frames. Therefore, our dataset includes

$$9 \text{ subjects} \times 9 \text{ gaits} \times (1200 \text{ clouds} + 1200 \text{ skeletons} + 1200 \text{ silhouettes})$$

An example of silhouette, skeleton, and 3D point cloud in our dataset is presented in Fig. 2.



Figure 2: Example of a silhouette, a skeleton, and a point cloud under different view points obtained in our setup for gait analysis. Notice that they correspond to different frames.

In order to give a standard evaluation for researchers that may perform their experiments on our dataset, we suggest default training and test sets. The training set includes walking gaits of 5 volunteers and the test set consists of data of the remaining subjects. The separation was performed by a random selection of 5 over 9 subjects. Beside this default separation, we also recommend researchers evaluating their works using leave-one-out cross validation on the 9 subjects.

Our dataset of skeleton was packed into a single *npz* file and can be easily processed using Python. The silhouette data were packed as 9 *npz* files (corresponding to 9 subjects) because of the large number of data dimensions. Finally, the point clouds in our dataset were stored in *mat* files, in which each one corresponds to a gait sequence, i.e. 1200 consecutive frames. Notice that the point cloud dataset requires a large storage capacity (about 18.6 gigabytes).

2.3 Data structure

2.3.1 Point cloud

The point cloud corresponding to a frame was reconstructed according to the method proposed in [2, 4]. This is represented by a matrix of size $n \times 3$, where n is the number of points. Each *mat* file is a sequence of 1200 consecutive point clouds.

```

1  '''Example of Python code reading the point clouds in a mat file'''
2  import numpy as np
3  import scipy.io as sio
4  loaded = sio.loadmat('subject_9/L4kg.mat')
5  #get point cloud data
6  data = loaded['seq_cloud']
7  #check data's shape, expected (1, 1200)
8  print(data.shape)
9  #get point cloud of the first frame
10 cloud = data[0,0]
11 #check cloud's shape, expected (n, 3)
12 print(cloud.shape)

```

2.3.2 Skeleton

The skeleton was estimated frame-by-frame using the Kinect 2 [5, 6] in our setup. Each skeleton is represented as a collection of 25 joints in 3D space. The indices of these joints in each frame are the default indices indicated by Microsoft on their website¹. Notice that a skeleton and its corresponding point cloud are in different 3D coordinate systems. If the readers want to match the two representations, they should perform a rotation on the skeleton according to the following rotation matrix:

$$\begin{bmatrix} 0.999695775540 & -0.024644093990 & 0.000934195170 \\ 0.024028207187 & 0.964782261798 & -0.261950057413 \\ 0.005563603415 & 0.262269299066 & 0.964978684264 \end{bmatrix} \quad (1)$$

```
1  '''Example of Python code reading the skeletons'''
2  import numpy as np
3  loaded = np.load('DIRO_skeletons.npz')
4  #get skeleton data of size (n_subject, n_gait, n_frame, 25*3)
5  data = loaded['data']
6  #get joint coordinates of a specific skeleton
7  skel = data[0,0,0,:]
8  x = [skel[i] for i in range(0, len(skel), 3)]
9  y = [skel[i] for i in range(1, len(skel), 3)]
10 z = [skel[i] for i in range(2, len(skel), 3)]
11 #get default separation
12 separation = loaded['split']
13 #print information
14 print(data.shape)
15 print(separation)
16 #expected results:
17  #(9, 9, 1200, 75)
18 #[ 'train' 'test' 'train' 'test' 'train' 'train' 'test' 'test' 'train']
```

2.3.3 Silhouette

The silhouette in our dataset was extracted based on depth maps captured by the Kinect 2 under frontal view. Each binary silhouette was enhanced by applying a procedure of noise removal. Since the silhouette data were stored as pixel values, combining the whole set into one file is restricted in Numpy due to the large number of elements. Therefore, this dataset contains 9 files corresponding to 9 subjects.

```
1  '''Example of Python code reading the silhouette of LAST SUBJECT'''
2  import numpy as np
3  loaded = np.load('subject_9.npz')
4  #get silhouette data of size (n_gait, n_frame, n_height, n_width)
```

¹<https://msdn.microsoft.com/en-us/library/microsoft.kinect.jointtype.aspx>

```

5 data = loaded['data']
6 #get a specific silhouette
7 silh = data[0,0,:,:]
8 #print information
9 print(data.shape)
10 #expected results:
11 #(9, 1200, 424, 512)

```

The dataset belongs to the Image Processing Lab, DIRO, University of Montreal. If the researchers use our dataset in their works, please refer this technical report as

T.-N. Nguyen and J. Meunier, “Walking gait dataset: point clouds, skeletons and silhouettes”, Technical report 1379, DIRO, University of Montreal, April 2018.

References

- [1] T.-N. Nguyen, H.-H. Huynh, and J. Meunier. Skeleton-based abnormal gait detection. *Sensors*, 16(11):1792, 2016.
- [2] T.-N. Nguyen, H.-H. Huynh, and J. Meunier. 3d reconstruction with time-of-flight depth camera and multiple mirrors. *IEEE Access*, 6:38106–38114, 2018.
- [3] T. N. Nguyen, H. H. Huynh, and J. Meunier. Assessment of gait normality using a depth camera and mirrors. In *2018 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, pages 37–41, Las Vegas, NV, USA, March 2018.
- [4] T.-N. Nguyen, H.-H. Huynh, and J. Meunier. Using tof camera and two mirrors for 3d reconstruction of dynamic objects. Technical Report 1380, DIRO, University of Montreal, April 2018. Available at www.iro.umontreal.ca/~labimage/GaitDataset/reconstruct3D.pdf.
- [5] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from single depth images. In *CVPR 2011*, pages 1297–1304, June 2011.
- [6] J. Shotton, R. Girshick, A. Fitzgibbon, T. Sharp, M. Cook, M. Finocchio, R. Moore, P. Kohli, A. Criminisi, A. Kipman, and A. Blake. Efficient human pose estimation from single depth images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(12):2821–2840, Dec 2013.
- [7] L. Tao, A. Paiement, D. Damen, M. Mirmehdi, S. Hannuna, M. Camplani, T. Burghardt, and I. Craddock. A comparative study of pose representation and dynamics modelling for online motion quality assessment. *Computer Vision and Image Understanding*, 148:136–152, 2016.