2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Scaling of 2D gait skeleton data for quantitative assessment of movement disorders from freehand single camera video recordings

Wei Tang*, Peter M.A. van Ooijen, *Senior Member, IEEE*, Deborah A. Sival and Natasha M. Maurits

***Abstract*— Five different methods to scale 2D skeleton keypoint data as derived from pose estimation deep neural networks for freehand video recording of gait were evaluated. Scaling using the bounding box of left shoulder and right hip keypoints best maintained original angles and provided the expected stable distance between hips and between shoulders.**

***Clinical Relevance*— Quantifying gait features from freehand single camera video recordings will allow for more objective and easy-to-use clinical evaluation of abnormal gait.**

I. Introduction

Gait can clinically be assessed semi-quantitatively using rating scales. However, such assessment requires extensive clinical knowledge and expertise, and is subjective and time-consuming. Therefore, wearable inertial sensors have been considered for objective and more accurate assessment. Yet, this approach still requires special hardware, preparation time and can by itself not be used to derive important distance-based features, such as step width. Recent advances in video-based pose estimation and the ubiquity of video cameras in outpatient clinics, have motivated us to design a 2D skeleton-based method for quantitative gait analysis using video images taken in the coronal plane. To allow calculation of (variability in) distance-based features, individual video frames need to be scaled first. Therefore, we evaluated different scaling methods to preprocess 2D skeleton keypoint data as derived from pose estimation deep neural networks.

II. Methods

Six children (2 ataxia, 2 developmental coordination disorder and 2 controls) walked in a straight line, with a single 2D camera placed in front of them. A deep neural network model trained on the MSCOCO dataset based on Alphapose was first used to extract 17 skeleton keypoints[1]. Then a PoseFlow framework was used to match the skeleton to the same participant in a recording[2]. We analyzed five different scaling methods, that each used different distances to scale every video frame with, according to:

$$x' = \frac{(x - x_0)}{w + e}, \quad y' = \frac{(y - y_0)}{h + e} \tag{1}$$

Here (x,y) and (x',y') are the current and scaled coordinates, respectively, (x_0,y_0) are the coordinates of the top left corner of the bounding box of all keypoints, e =1 (pixel; used to avoid division by zero) and w and h are width and height defined by the specific scaling method. Four of the assessed scaling methods used width and height of the bounding boxes of: 1) all keypoints (Box scale or BS method), 2) the two shoulder keypoints (S method), 3) the left shoulder and right hip (LS/RH method) and 4) the two hip keypoints (H method). The 5th method used the distance between left shoulder and right hip for both w and h (LS/RH-d method). Performance was assessed by the mean absolute angle error (MAAE; angles should remain the same after scaling) and mean variance of distance between certain keypoints. An example of scaling is at *https://github.com/jiudaa/Scaling.*

III. Results

The mean likelihood for all keypoints was 0.82 (std=0.14), indicating high reliability. The MAAE was 4-5 times as high for the S and H compared to the other methods; these methods were therefore omitted from further consideration. Mean variance of distance between certain pairs of keypoints is illustrated in Fig. 1.

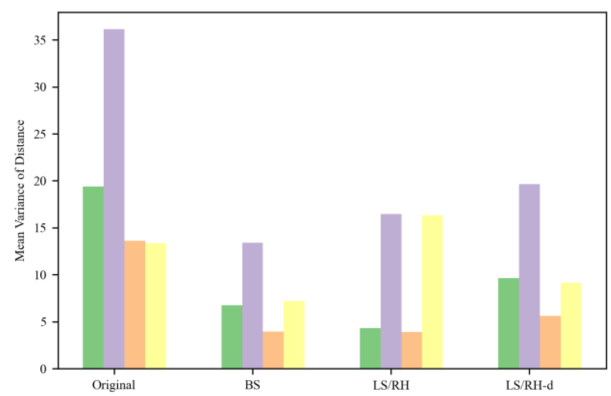

Figure 1. Mean variance of distance between shoulder (green), wrist (purple), hip (orange) and ankle (yellow) keypoints for the original data and BS, LS/RH and LS/RH-d methods.

The LS/RH method performed best because it maintained expected low variability in shoulder and hip distance, while allowing for expected variability in hands and ankle distance.

IV. Conclusion

This study is a first step towards quantitative analysis of clinically observed gait using single camera freehand video recording, that may be used to differentiate movement disorders with machine learning in the future.

References

[1] H.S. Fang, S. Xie, Y.W. Tai, et al. *Rmpe: Regional multi-person pose estimation*, Proceedings of the IEEE international conference on computer vision. 2017: 2334-2343.

[2] Y. Xiu, J. Li, H. Wang, et al. *Pose Flow: Efficient online pose tracking*. arXiv preprint arXiv:1802.00977, 2018.

*Author supported by a grant from the China Scholarship Council.
Wei Tang is with the University of Groningen, University Medical Center Groningen, The Netherlands; e-mail: w.tang@umcg.nl.