NOVEL COMPUTER VISION AND DEEP LEARNING APPROACHES FOR TRACKING 3-D SPINE MOTION DURING DYNAMIC TRUNK FLEXION USING AN RGB-D CAMERA

Ryan B. Graham[1-2], Wantuir C.R. Junior[1], Kristen Beange[2], Alistair Boyle[1], Matthew P. Mavor[1]
[1]School of Human Kinetics, Faculty of Health Sciences, University of Ottawa, Ottawa, ON, Canada
[2]Ottawa-Carleton Institute for Biomedical Engineering (OCIBME), Ottawa, ON, Canada
Email: rgraham@uottawa.ca

Summary

The overall goal of this work was to develop an inexpensive and portable tool capable of quantifying lumbar spine motion in the field using a red-green-blue-depth (RGB-D) time-of-flight camera. Specifically, we created a novel framework that takes the RGB-D data and manipulates them to measure 3-D lumbar spine kinematics during dynamic trunk flexion. To achieve our goal, we carried out two independent research studies: 1) the development and validation of a custom computer vision method to track infrared (IR) reflective markers from raw depth data to calculate 3-D spine angles; and 2) development and validation of a custom four module convolutional neural network (CNN; SpineNet) to track and automatically segment regions of each participant's back to calculate 3-D kinematics without the use of any markers.

Introduction

In order to better understand motor control and biomechanical differences between low back pain patients and healthy controls, researchers and clinicians often observe movement patterns and quality during pre-defined tasks (e.g., trunk flexion-extension) [1]. Recently, a great amount of work has been undertaken to validate the use of portable inexpensive sensors such as inertial measurement units (IMUs) to measure 3-D spine motion during dynamic tasks in the field [2]; however, an alternative approach is to use RGB-D cameras [3]. While many research studies have been carried out in the biomechanical domain using RGB-D cameras [3], few have attempted to measure spine motion [4], and none have done so during dynamic trunk flexion tasks with or without markers. Through two studies we aimed to address this research gap.

Methods

<u>*Study 1:*</u> 12 healthy young adults (6M, 6F) were recruited to perform repetitive spine flexion-extension, with infrared reflective marker clusters placed over their T_{10}-T_{12} spinous processes and sacrum, and motion capture data were recorded simultaneously by one RGB-D camera and a 10-camera optoelectronic motion capture system. Custom computer vision algorithms were developed to extract spine angles from depth data. Root mean square error *(RMSE)* was calculated for continuous Euler angles, and intraclass correlation coefficients *($ICC_{2,1}$)* were calculated between min, max, and average range of motion angles in all movement planes for both systems.

<u>*Study 2*</u>: A single RGB-D camera captured infrared, depth, and colour image data of 15 male participants performing two batteries of 10 cycles of repetitive trunk flexion-extension under two conditions: marked (i.e., hand drawn markers on key anatomical locations on the back) and unmarked. The collected data were used to create a custom four module convolutional neural network (CNN; SpineNet) (Figure 1 – based on [5]) to segment the back into upper, lower, and spine regions and then to extract lumbar spine kinematics. SpineNet was trained and tested on ten marked participants in a train:test ratio of 80:20. Images of five additional participants without markers were used to evaluate model generalizability.

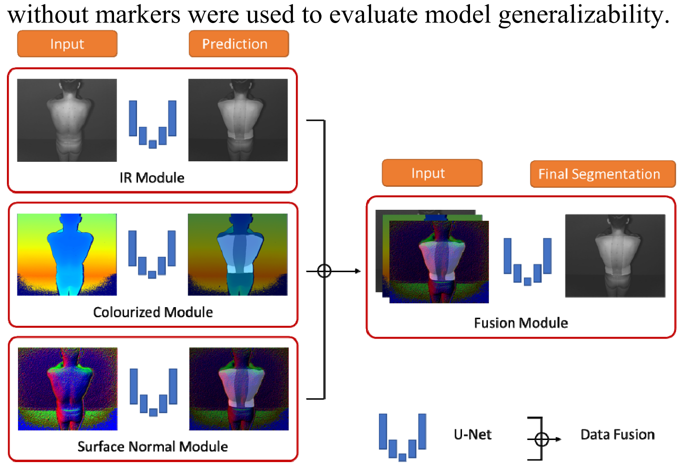

Figure 1. CNN architecture. Each RGB-D camera data stream was used as input for each of the three individualized U-Net modules. Individual modules' anatomical segmentation predictions were then fused and used as input for the final U-Net segmentation module.

Results and Discussion

<u>*Study 1:*</u> *RMSE* in angles between the two calculation methods were very low across all movement axes *($RMSE \leq 2.05° \pm 0.97°$)*, and $ICC_{2,1}$ values were good to excellent across all axes ($0.849 \leq ICC_{2,1} \leq 0.979$). Bland-Altman plots revealed that, on average, the RGB-D camera slightly underestimated lumbar flexion-extension angles (-1.88º) and overestimated lumbar lateral bending and axial twisting angles (≤ 0.58º).

<u>*Study 2:*</u> Quantitative image segmentation analysis on marked data had good similarity (frequency weighted intersection over union ≥ 0.8087) and accuracy (mean pixel accuracy ≥ 0.8855) between the prediction and ground truth across all modules. Qualitative image segmentation analysis on unmarked data showed that colourized and surface normal modules presented a more robust class morphology throughout frames than infrared and fusion modules. Kinematic analysis on unmarked participants showed that flexion-extension angles exhibited movement profiles (i.e., shape, timing, and peaks) that are comparable to similarly collected data from previous research.

Conclusion

This work provides proof for using a single RGB-D camera for assessing lumbar kinematics with or without markers.

Acknowledgments

Funding provided by NSERC and the Ontario ERA program.

References

[1] van Dieën et al. (2019). *JOSPT* **49**: 370-9. [2] Beange et al. (2019). *J Biomech* **97** : 109356. [3] Clark et al. (2019). *Gait Posture* **68** : 193-200. [4] Macpherson et al. (2016). *J Biomech* ***49*** : 474-8. [5] Rahman et al. (2017). *IEEE ICME*, 991-6.