Physical Ergonomics and Human Factors, Vol. 63, 2022, 214–219
https://doi.org/10.54941/ahfe1002615

Developing a Computer-Vision Model to Estimate Anatomical Joint Coordinates During Manual Lifting Tasks

Chih-Yu Hsiao, Chien-Chi Chang, Ting-Yu Chen, and Yi-Ting Lin

Department of Industrial Engineering and Engineering Management, National Tsing Hua University, NO. 101, Sec. 2, Kuang Fu Road, Hsinchu 300034, Taiwan, R.O.C.

ABSTRACT

This study developed a Computer-Vision based anatomical joint coordinates estimation model to predict the 3D joint coordinates with the help of Artificial Intelligence image recognition technology during manual lifting tasks based on single camera video inputs. The workflow of the proposed Computer-Vision model includes 2D joint detection and 3D joint reconstruction. The 3D joint error is calculated based on the Euclidean distance between the predicted 3D joint coordinates from the CV-based method and the corresponding joint coordinates of the ground truth from the Visual3D^{TM} skeletal model. The results indicated that the floor to shoulder lifting height path induced a greater 3D joint error than the floor to knuckle and knuckle to shoulder lifting height paths (p-value = 0.01). The 3D joint error of the hand was the largest than the other estimated joints. This study verified that the proposed Computer-Vision model could predict 3D joint points. Therefore, while the marker-based motion tracking system is inapplicable, the model can be used as an alternative solution for predicting lifting motion.

Keywords: Artificial intelligence image recognition technology, Convolutional neural network, 3D human joint coordinate estimation, Lifting

INTRODUCTION

Manual lifting tasks are closely related to musculoskeletal disorders. In previous studies, Motion Tracking System (MTS) and Force Plate (FP) were often used for biomechanical analysis of lifting task (Faber et al. 2013; Faber et al. 2016; Xu et al. 2012). However, MTS and FP have high deployment costs, and MTS requires sticking reflecting light balls on the skin, which can easily affect people's activities. Therefore, MTS and FP are likely to be inappropriate for human motion analysis in the working field or when used in certain working environments. With the development of Computer Vision (CV) technology, there have been studies on constructing 3D motion models from human body images using CV technology (Sandau et al. 2014; Wang et al. 2021). However, if some limb segments are obscured in the 2D images, it will affect the performance of 3D pose reconstruction. Wang et al. (2021) used VideoPose3D, a temporal convolutional neural network method, to estimate the 3D pose of humans through lifting task images taken by a single camera.

However, most of the output joint positions provided by VideoPose3D had no anatomical meaning, still needed to estimate the joint positions via anthropometric parameters. Hence, if a set of CV methods for estimating human joint based on the anatomical joint positions can be developed, it will provide to facilitate the biomechanics analysis in the future.

The purpose of this study was to develop a CV-based anatomical joint coordinates estimation model to predict the 3D joint coordinates during manual lifting tasks based on single camera video inputs. When MTS is hard to apply, the developed model in this study could serve as an alternative solution to MTS.

METHODS

Participants

30 participants were recruited to perform three symmetric manual lifting tasks in this study. The data of 20 participants were used for CV model development and the data of the other 10 participants were used to verify the accuracy of the developed CV model in predicting joint coordinates. The experimental protocol was approved by the local Institutional Review Board. The informed consent form was obtained from all participants.

Appartus

An optical MTS (OptiTrack, NaturalPoint Inc., Oregon, USA) was used to capture lifting motions. The capture sampling rate is 125 Hz. The raw data were filtered at 10 Hz using a fourth-degree low-pass Butterworth filter. A total of 44 passive reflective markers were placed on the anatomical locations based on Visual 3D^{TM} (C-Motion Inc., Germantown, Maryland, USA) marker placement guidelines. A total of 19 joint centers with anatomical meaning were defined by 44 marker points, including head, C7, L5/S1, left/right shoulders, left/right elbows, left/right wrists, left/right hands, left/right hips, left/right knees, left/right ankles, and left/right foots. To obtain the coordinate of the anatomical position of the 19 joints as reference, the joint coordinate of the ground truth data captured from the MTS were input to Visual 3D^{TM}. The camera was placed 45 degrees to the front left of the participants, with a distance of 1.4 m from the participants and a height of 2.5 m.

Experiment Procedure

Each participant was asked to perform three symmetric manual lifting tasks: floor to knuckle (F-K), knuckle to shoulder (K-S), and floor to shoulder (F-S) at their self-selected speed. The task sequences were arranged randomly. An experimental box (37 $\times$ 33 $\times$ 6 cm) with handles on both sides was used in this study. The weight of the box was set to 10 % of each participant body weight. All participants were instructed to lift the box from an initial location to a final location.

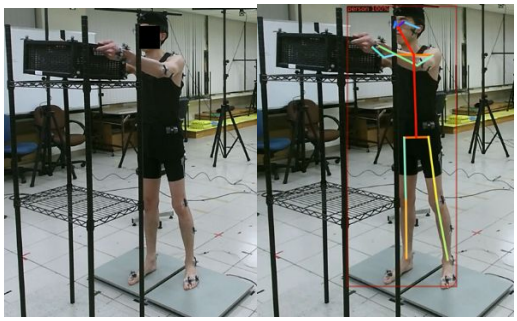

Figure 1: 17 joint points of the lifting image acquired by Detectron2. Left: single-view image, Right: 17 joint points in human images.

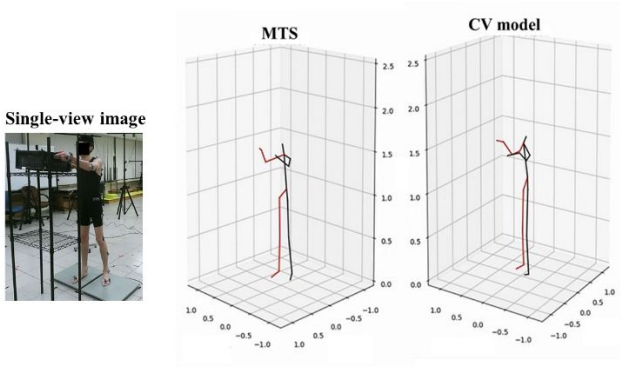

Figure 2: 3D joints predicted by the proposed CV method. Left: single-view image, Middle: corresponding ground-truth joints, and Right: 3D joint reconstruction based on CV model.

Computer-Vision Model

The workflow of the proposed CV model includes two major steps: 2D joint detection and 3D joint reconstruction.

The 2D human joint points in each frame of lifting images were detected by Detectron2, a 2D joint point detector that has been pre-trained using the COCO 2017 dataset (Lin et al., 2014) and can detect 17 joint points in human images, as shown in Figure 1. After the detection is completed, the 2D human joint points in each frame of lifting images are stored as datasets in the "NumPy" format, which can be used as an input of VideoPose3D.

In this study, VideoPose3D was used to reconstruct 3D human joint points. VideoPose3D uses the detected 2D joint points as the input and estimates the 3D joint points through time convolution, i.e. to extract joint points with each time interval from time series and combine these points after all the extractions were completed, as shown in Figure 2. In this study, all of the model parameters in VideoPose3D were set according to Pavllo et al. (2019). The architecture of VideoPose3D is constructed by stacking of ResNet-style blocks. In the VideoPose3D architecture, there are four blocks (B) in total, and in each block contains two convolution layers, one with kernel size = 3 and the another with kernel size = 1.

Most of the 3D human joint points generated by original VideoPose3D model do not have practical anatomical meaning (Wang et al., 2021). In order to develop a CV model that can predict joint with anatomical meaning, in this study, we transformed the ground truth joint based on the MTS coordinate system into the camera coordinate system. And used ground truth joints on the camera coordinate system as CV model targets. Because MTS and cameras have different coordinate systems, many previous studies aligned different camera coordinate systems through calibrators. Therefore, this study used the calibrator and calibration process proposed by Liu et al. (2021) to align the MTS and camera coordinate systems. The relationship between the MTS and camera coordinate systems can be expressed by the following formula.

$$(X, Y, Z)^{\text{Motion}} = (x, y, z)^{\text{Camera}} x R_{3,3} + T_{3,1}$$

For the ground truth joint in the MTS coordinate system, the coordinate system was transposed to the camera coordinate system based on the rotation matrix (R), which is a 3x3 matrix, and translation matrix (T), which is a 3x1 matrix in the above formula.

The supervised training method was used into VideoPose3D, and 80% of the data were used for model training to adjusted the parameters and hyperparameters, while the remaining 20% of the data were used for model testing.

Statistical Analysis

The ground truth joints were adopted as the benchmark. 3D joint error was calculated based on the Euclidean distance between the predicted 3D joint coordinates from the CV-based method and the corresponding joint coordinates of the ground truth from the Visual3D^{TM} skeletal model. In addition, univariate analysis of variance (ANOVA) was adopted to analyze the effect of the lifting height path on the prediction performance of the CV model. In the analysis, three different lifting height paths were the independent variables, whereas the error of the 3D joint point was the dependent variable. Duncan's multiple range post-hoc tests were performed as post-hoc tests. The statistical significance level was set at $\alpha = 0.05$. All statistical analyses were performed using SPSS Statistics software v.22.

RESULTS

The results indicated that the average Euclidean distance between predicted 3D joints coordinates and corresponding ground truth joint coordinates was 14 cm. In addition, the study calculated the prediction error of all estimated joints, among which that of the hands was the largest, as shown in Figure 3. ANOVA result indicated that the lifting height path exhibited a significant effect on the 3D joint error [$F(2, 327) = 4.65$, p-value $= 0.01$]. The F-S lifting path induced a greater 3D joint error than the other lifting paths, as shown in Table 1.

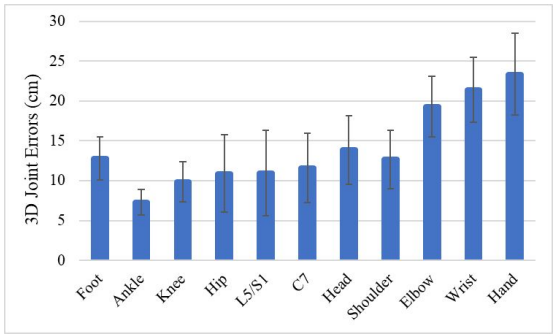

Figure 3: Average error between the predicted 3D joint and the corresponding ground truth joint point.

Table 1. Average 3D joint error (mm) for each lifting height path and Duncan's multiple range post-hoc test results on lifting height paths.

Source	Mean ± SD
Lifting height path (mm)	
F-K	14.49 ± 6.46 [a, b]
K-S	13.123 ± 6.68 [a]
F-S	15.28 ± 6.63 [b]

DISCUSSION

This study developed a CV model that could estimate the coordinates of anatomical joints during manual lifting tasks. The results showed that the CV model was capable for predicting 3D joint points. Previous studies have used CV approach to estimate joint positions (Mehrizi et al., 2019; Wang et al., 2021). Mehrizi et al. (2019)'s research utilized multi-view images for 3D pose estimation. In Wang et al. (2021)'s research, part of predicted 3D joints that had no anatomical meaning. This study was completely different with these studies, the CV model developed in this study could predict and reconstruct anatomically based 3D joint points using images from a single camera.

The results showed that the effect of the lifting height path on the estimated 3D joint point errors were significant. The estimation error reached its maximum when predicting the lifting path from floor to shoulder height. This is likely happening due to the large change in the posture when lifting the experimental crate from floor to shoulder height.

In addition, the CV model developed in this study demonstrated a larger error in predicting the hand joint, which was because the participants' hand was covered by the crate when lifting the experimental crate using both hands. Wang et al. (2021) reported that the partial block of limb segments could compromise the performance of the CV in detecting joints, which was possibly why the study showed larger errors in predicting the hand joint.

CONCLUSION

This study developed a model that could predict 3D anatomical joints based on the CV method. The model utilized lifting images acquired using a single camera as the input data to predict 3D joint points with the AI image recognition technology. The results showed that the estimation joint error of the model reached its maximum when predicting the lifting path from floor to shoulder height. In addition, the model demonstrated a larger error in predicting the hand joint. This study verified that the proposed CV model could predict human 3D joints. However, larger predicting errors seemed to appear in some of the joints. The authors plan to introduce further lifting data for model training or to conduct experiments with different predicting algorithms in efforts to improve the accuracy of human 3D joints prediction.

ACKNOWLEDGMENT

This study was partially supported by the Ministry of Science and Technology, Taiwan, R. O. C. (MOST 109-2221-E007-062-MY3).

REFERENCES

C-Motion. (August 15, 2017) Marker Set Guidelines. C-Motion Product Documentation Website: https://www.c-motion.com/v3dwiki/index.php/Marker_Set_Guidelines.

Faber, G. S., Chang, C. C., Kingma, I., Dennerlein, J. T., Van Dieën, J. H. (2016) "Estimating 3D L5/S1 moments and ground reaction forces during trunk bending using a full-body ambulatory inertial motion capture system", Journal of biomechanics, 49(6), pp. 904–912.

Faber, G. S., Chang, C. C., Rizun, P., Dennerlein, J. T. (2013) "A novel method for assessing the 3-D orientation accuracy of inertial/magnetic sensors", Journal of biomechanics, 46(15), pp. 2745–2751.

Liu, P. L., Chang, C. C., Lin, J. H., Kobayashi, Y. (2021) "Simple benchmarking method for determining the accuracy of depth cameras in body landmark location estimation: Static upright posture as a measurement example", Plos one, 16(7), pp. e0254814.

Mehrizi, R., Peng, X., Xu, X., Zhang, S., Li, K. (2019) "A Deep Neural Network-based method for estimation of 3D lifting motions", Journal of biomechanics, 84, pp. 87–93.

Pavllo, D., Feichtenhofer, C., Grangier, D., Auli, M. (2019) "3D human pose estimation in video with temporal convolutions and semi-supervised training", In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7753–7762.

Sandau, M., Koblauch, H., Moeslund, T. B., Aanæs, H., Alkjær, T., Simonsen, E. B. (2014) "Markerless motion capture can provide reliable 3D gait kinematics in the sagittal and frontal plane", Medical engineering & physics, 36(9), pp. 1168–1175.

Wang, H., Xie, Z., Lu, L., Li, L., Xu, X. (2021) "A computer-vision method to estimate joint angles and L5/S1 moments during lifting tasks through a single camera", Journal of Biomechanics, 129, pp. 110860.

Xu, X., Chang, C. C., Faber, G. S., Kingma, I., Dennerlein, J. T. (2012) "Estimation of 3-D peak L5/S1 joint moment during asymmetric lifting tasks with cubic spline interpolation of segment Euler angles. Applied ergonomics, 43(1), pp. 115–120.