NO DATASET TOO SMALL! ANIMATING 3D MOTION DATA TO ENLARGE 2D VIDEO DATABASES

Marion Mundt[1], Henrike Oberlack[2], Corey Morris[1], Johannes Funken[3], Wolfgang Potthast[3], Jacqueline Alderson[1,4]

Minderoo Tech & Policy Lab, The University of Western Australia, Perth, Australia[1]
Institute of General Mechanics, RWTH Aachen University, Aachen, Germany [2]
Institute of Biomechanics and Orthopaedics, German Sport University Cologne, Cologne, Germany[3]
Auckland University of Technology, Sports Performance Research Institute New Zealand (SPRINZ), Auckland, New Zealand[4]

This study outlines a technique to leverage the wide availability of high resolution three-dimensional (3D) motion capture data for the purpose of synthesising two-dimensional (2D) video camera views, thereby increasing the availability of 2D video image databases for training machine learning models requiring large datasets. We register 3D marker trajectories to generic 3D body-shapes (hulls) and use a 2D pose estimation algorithm to predict anatomical landmark keypoints in the synthesised 2D video views – a novel approach that addresses the limited data available in elite sport settings. We use 3D long jump data as an exemplar use case and investigate the influence of; 1) varying anthropmetrics, and 2) the 2D camera view, on keypoint estimation accuracy. The results indicated that 2D keypoint determination accuracy is affected by body-shape. Frontal plane camera views result in lower accuracy than sagittal plane camera views.
KEYWORDS: machine learning, pose estimation, simulation.

INTRODUCTION: Biomechanical analysis of motion outside the laboratory is becoming increasingly popular thanks to the wide availability of high-resolution cameras and open access 2D and pseudo 3D pose estimation algorithms. These pose estimation algorithms, e.g. DeepLabCut or OpenPose, are based on Convolutional Neural Networks (CNN) that determine keypoints which identify 2D anatomical landmarks (AL), from standard video data. They have been trained on large 2D standard video databases of manually digitised images – a common and longstanding approach of the sports biomechanics community. By applying pre-trained CNN models to newly collected video data, the time burden for AL keypoint determination may be significantly reduced, with no loss to accuracy (Cao, Hidalgo, Simon, Wei & Sheikh, 2019). Importantly, this would result in increased time availability for interpretation, translation and intervention between the sport biomechanist, coach and athlete.
In addition to obtaining information about kinematic motion parameters, the estimation of kinetic parameters using 2D video data inputs into machine learning algorithms is of high interest as it provides insight surrounding an athlete's internal biomechanical load (Morris, Mundt, Mian & Alderson, 2021; Mundt, 2021) and compliments applications of machine learning using 3D optoreflective motion capture trajectories, or inertial sensor data, as inputs (Johnson, Alderson, Lloyd & Mian, 2019; Mundt, Koeppe, David, Witter, Bamer, Potthast & Markert, 2020; Kipp, 2020). Importantly, the ecological validity of unobstructive 2D video as a capture and analysis medium, that is capable of providing high resolution data ouputs, is the holy grail of the sports biomechanics discipline, especially if we are to further our understanding of injury mechanisms on-the-field (Weir, Alderson, Smailes, Elliott & Donnelly, 2019).
Since the late 1970s the high resolution and fidelity of outputs required by the biomechanists has required testing to be undertaken in laboratory settings using 3D motion capture techniques, electromyography or other highly specialised equipment. Computer and technical limitations have seen the vast bulk of this 3D motion data collected without concurrent 2D video streams. Consequently, these datasets are not ideal inputs to train machine learning models

for 2D video pose estimation application. However, historic 3D motion capture data is clearly valuable, especially when considering the small and elite cohort this data originates from.
In this paper we propose a framework to simulate 2D video data images by leveraging existing 3D motion capture trajectory data. We will also assess the sensitivity of the pose estimation to the body-shape used to animate 3D motion capture data across various camera views.

METHODS: Eleven long jump trials of five athletes (3 trials of athlete 1, 1 trial of athlete 2, 2 trials of athlete 3, 4 trials of athlete 4 and 1 trial of athlete 5) recorded by an author of this paper in a previous project were used in this study (for details see Willwacher, Funken, Heinrich, Müller, Hobara, Grabowski, Brüggemann & Potthast, 2017). Height, length, width and/or circumference of feet, shanks, thighs, torso, arms, neck and head were measured before motion capture. The 3D motion capture data contained the final three steps of the run-up (approach), the take-off and the first component of the flight phase. The 3D motion capture data was modelled in Vicon Nexus using a standard biomechanical model and joint centres determined. An animated rig was fitted to these (Figure 1) using the scripting language Python with Blender (version 2.79). In a subsequent step, a generic 3D avatar body-shape was fitted to the rig using the MakeHuman plug-in to Blender. The body-shape will now be driven by the motion of the rig.

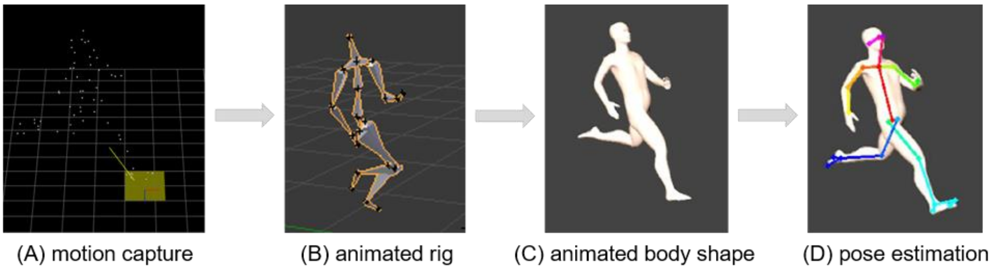

Figure 1 Workflow to estimate AL keypoints from 3D motion capture trials: (A) the motion capture data (.c3d files) are used to calculate joint centres and (B) animate a rig. (C) A human body-shape is morphed to the rig. The animation is able to be captured from varying planar camera views within the software and the 2D views are used to (D) estimate 2D AL keypoints in the reference frame of each image.

This workflow enables the automated creation of video data from motion capture data. The initial personal anthropometric data of the athlete provided the ground-truth 3D body-shape. This initial data along with three varying human body-shapes (Figure 2) based on standard anthropometric data displaying the 5th, 50th and 95th percentile (Fryar, Gu & Ogden, 2012) were morphed to the rig and coupled to the rig's movements. The animated body-shape was captured from eight different planar camera views within the software: frontal, anterior sagittal (AS), true sagittal (TS), posterior sagittal (PS), back, and contralateral to the original view from posterior sagittal (PSL), true sagittal (TSL) and anterior sagittal (ASL). The 2D video views were recorded at a resolution of 960 x 540 px and a frame rate of 100 Hz. Finally, 25 pose estimation AL keypoints were determined in all video frames using OpenPose (Cao et al., 2019). Keypoints are provided as pixel coordinates in each image's (u, v) reference frame with the origin in the top left corner of each image. The difference between the estimated keypoints of the four different body-shapes (in the same camera views) was compared to assess the influence of body-shape on keypoint location accuracy.

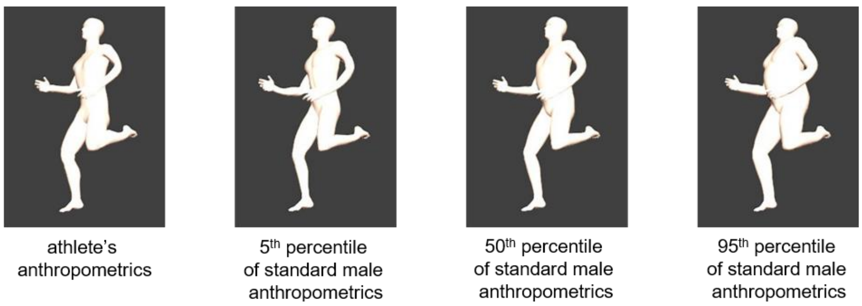

Figure 2 Varying human body-shapes assessed based on anthropometric data (Fryar et al., 2012).

RESULTS: The mean difference between ground truth athlete body-shape and the imposed 5th, 50th and 95th percentile body is displayed in Figure 3. The smallest mean differences and lowest variation were observed between the ground-truth athlete body-shape and the; 5th percentile (mean absolute difference 1.82±2.18 px), the largest for the 95th percentile (2.73±2.53 px) and the medium difference for the 50th percentile (2.34±2.33 px).

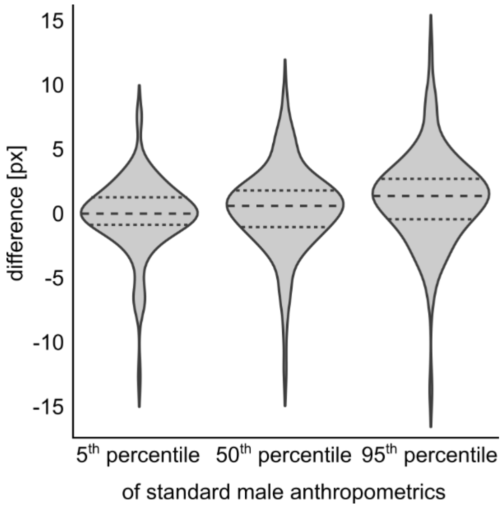

Figure 3 Violin plots showing the distribution of the mean differences in keypoint estimation between the athlete ground-truth body-shape and the three imposed percentile body-shapes (5th, 50th, 95th) across all trials. Mean and standard deviation are represented by horizontal dashed and dotted lines respectively.

Figure 4 shows the mean difference between the estimated AL keypoints of all individual trials across the eight camera views, for all five athletes. The largest mean differences can be found in the frontal plane view (rear aspect) (mean absolute difference 4.54±3.59 px), and the smallest mean differences in the anterior sagittal view (0.95±0.59 px).

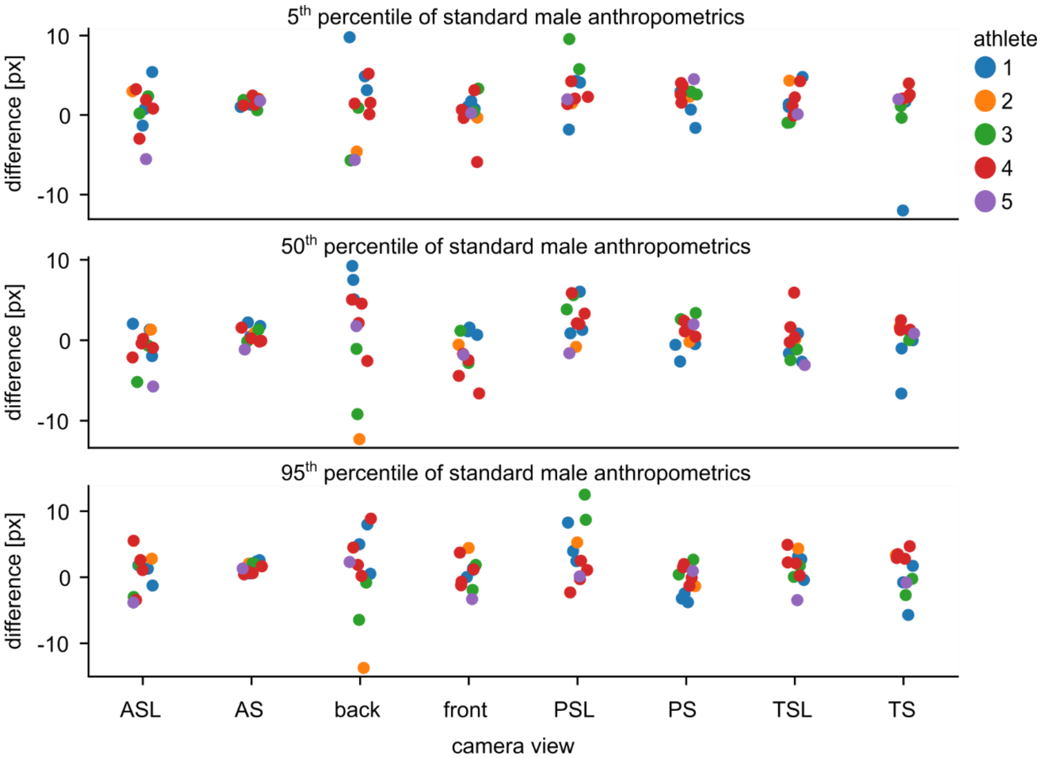

Figure 4 Distribution of mean AL keypoint estimation differences for varying body-shapes across eight camera views; anterior sagittal from the left (ASL), anterior sagittal from the right (AS), back, front, posterior sagittal from the left (PSL), posterior sagittal from the right (PS), true sagittal from the left (TSL) and true sagittal from the right (TS), for all five athletes.

DISCUSSION: Eighty-eight animated videos and four different body-shapes applied to five athletes were used to estimate keypoints for eleven long jump trials using OpenPose. The results showed that off-the-shelf 2D pose estimation models do appear to be sensitive to human body-shape, although shapes with a higher similarity, like the athelete specific shape and a generic 5th percentile shape in this study, lead to a higher agreement in keypoint location. As the predefined body-shape is scaled and morphed to the rig, the height of all hulls is constant and only the influence of differences in girths are assessed. Based on our findings, when using 3D motion capture data to generate 2D camera views the body-shape selection can be simplified as it is not necessary to develop personalised human body-shapes to achieve accurate results for the estimation of keypoints.

We also found that differences between AL keypoint estimation occur more frequently for some camera views, especially the one from the rear. In this camera view, the pose estimation algorithm confuses the right and left leg often towards the end of the trial — at beginning of the flight phase. The impact of this finding is likely dependent on application, e.g. the estimation of ground reaction forces during long jump take-off where only the approach and take-off frames are relevant would not be adversely impacted by pose estimation errors during the flight phase. Dependent on the application, a concise hygiene check of the AL keypoint data from the pose estimation model is necessary.

CONCLUSION: This study outlined a workflow pipeline to enable to use of 3D motion capture data for the synthesis of 2D video data from a near-infinite number of camera views. When using these views to estimate AL keypoints using off-the-shelf pose estimation models the keypoint location did not appear to be sensitive if a similar body-shape is applied. Based on the motion analysed, the accuracy of pose estimation algorithms needs to be reviewed carefully and flawed trials need to be excluded.

The recreation of animated 3D athletes and volumes within software facilitated the synthesis of 2D camera views that did not initially exist. The creation of these views and their use in increasing the size of 2D video databases, especially in cases where limited data is available as is commonly the case in elite sport, represents a novel and exciting approach to leveraging historical 3D data sets for field-based applications.

ACKNOWLEDGEMENTS

This work was partially funded by the Australian Intstitute of Sport.

REFERENCES

Cao, Z., Hidalgo, G., Simon, T., Wei, S. E., & Sheikh, Y. (2019). OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Fryar C.D., Gu Q. & Ogden C.L. (2012). Anthropometric reference data for children and adults: United States, 2007–2010. National Center for Health Statistics. *Vital Health Stat* 11(252).

Johnson, W. R., Alderson, J. A., Lloyd, D., & Mian, A. S. (2019). Predicting Athlete Ground Reaction Forces and Moments from Spatio-Temporal Driven CNN Models. *IEEE Transactions on Biomedical Engineering,* 66(3), 689–694

Kipp, K. (2020). Relative importance of lower extremity net joint moments in relation to bar velocity and acceleration in weightlifting, *Sports Biomechanics*.

Morris, C., Mundt, M., Mian, A. & Alderson, J. (2021). Estimation of 3D ground reaction force from 2D video using artificial neural networks. Submitted to *XXVIII Congress of the International Society of Biomechanics (ISB)*.

Mundt, M. (2021). Machine learning and motion analysis: Wearable and video applications. *Contribution to the ISBS 2021 Mid-year Symposium*.

Mundt, M., Koeppe, A., David, S., Witter, T., Bamer, F., Potthast, W., & Markert, B. (2020). Estimation of Gait Mechanics Based on Simulated and Measured IMU Data Using an Artificial Neural Network. *Frontiers in Bioengineering and Biotechnology*, 41(8).

Weir, G., Alderson, J., Smailes, N., Elliott, B. & Donnelly, C.A. (2019). Reliable Video-based ACL Injury Screening Tool for Female Team Sport Athletes. *International Journal of Sports Medicine* 40(3):191-199.

Willwacher, S., Funken, J., Heinrich, K., Müller, R., Hobara, H., Grabowski, A.M., Brüggemann, G.P. & Potthast, W. (2017). Elite long jumpers with below the knee prostheses approach the board slower, but take-off more effectively than non-amputee athletes. *Scientific Reports,* 7.