USING COMPUTER VISION AND DEEP LEARNING METHODS TO CAPTURE SKELETON PUSH START PERFORMANCE CHARACTERISTICS

Laurie Needham[1], Murray Evans[1], Darren P. Cosker[1] & Steffi L. Colyer[1]

Centre for the Analysis of Motion, Entertainment Research and Applications, University of Bath, Bath, UK[1]

This study aimed to employ computer vision and deep learning methods in order to capture skeleton push start kinematics. Push start data were captured concurrently by a marker-based motion capture system and a custom markerless system. Very good levels of agreement were found between systems, particularly for spatial based variables (step length error 0.001 ± 0.012 m) while errors for temporal variables (ground contact time and flight time) were within 1.5 frames of the criterion measures. The computer vision based methods tested in this research provide a viable alternative to marker-based motion capture systems. Furthermore they can be deployed into challenging, real world environments to non-invasively capture data where traditional approaches would fail.

KEYWORDS: Winter sports, computer vision, deep learning, pose estimation.

INTRODUCTION: Successful skeleton push start performance has been characterised by the interaction between attaining a high pre-load velocity and executing an effective loading phase (Colyer et al., 2018), where the athlete transitions from pushing the sled to driving the sled in a prone position. Push performance is considered important for overall success in skeleton (Zanoletti et al 2006). Physical abilities (such as lower limb power and sprinting ability) explain a large portion of the variance in the sled velocity attained (Colyer et al., 2018) but factors related to technique are suggested to potentially account for further variation in performance. In order to further understand and enhance push start performance, capturing the underlying kinematics of both the athlete and sled during pushing is likely beneficial.
To date, limited push start kinematic information is available, perhaps due in part to the limitations associated with current motion capture technology and the challenging environment that skeleton takes place in. Advances in computer vision and deep learning are providing viable alternatives to the traditional marker-based motion capture systems that are considered the gold standard for many biomechanical research applications. Evans et al., (2018) presented a computer vision based multi-camera system capable of non-invasively capturing accurate step characteristic data during running. Deep learning based pose estimation aims to identify body landmarks from regular image data and may provide a robust and non-invasive alternative to marker-based motion capture systems to capture additional information such as centre of mass position and velocity. The aim of this study was to compare the performance of a fully-automated computer vision and deep learning based methodology to that of a traditional marker-based motion capture system in challenging real-world environment (skeleton push-track).

METHODS: 12 international skeleton athletes (7 males [1.81 ± 0.05 m, 83.37 ± 2.73 kg], 5 females [1.71 ± 0.03 m, 70.04 ± 1.44 kg]) provided written informed consent. Each athlete performed three maximal effort dry-land push starts at the University of Bath's outdoor push track. Motion data were captured concurrently using two motion capture systems. Criterion data were captured using a 15 camera marker-based motion capture system (Oqus, Qualysis AB, Gothenburg, Sweden) while additional data were captured using a custom 9 camera computer vision system (JAI sp5000c, JAI ltd, Denmark). Motion capture systems were time-synchronised by means of a periodic TTL-pulse generated by the custom system's master frame grabber to achieve a sampling frequency of 200 Hz in both systems. This ensured that frames were captured by all cameras in unison without drift. Both camera systems were positioned around the push track in order to capture the pushing action between 5 m and 15 m from the starting block where there was a constant gradient of ~2%. The Qualysis system

was calibrated as per the manufacture's specifications. The custom camera system used a binary dot matrix to determine each camera's intrinsics and extrinsics. A right-handed coordinate system was defined for both systems by placing a Qualysis L-Frame in the centre of the push track, 10 m from the start block. In order to refine the alignment of each system's Euclidean space, a single marker was moved randomly through the capture volume and tracked by both systems. This marker data provided points with which the spatial alignment could be optimised in a least-squares sense. To assess the reconstruction accuracy of both systems a wand was moved through the capture volume and tracked by both systems before the mean (± SD) resultant vector magnitude was computed and compared to a known value.

To capture criterion data a full body marker set comprising of 44 individual markers and four clusters were attached to each participant to create a full body six degrees of freedom (6DoF) model (bilateral feet, shanks and thighs, pelvis and thorax, upper and lower arms, and hands). Four additional markers were placed on the sled to track position and orientation. Following labelling and gap filling of trajectories (Qualysis Track Manager v2019.3, Qualysis, Gothenburg, Sweden) data were exported to Visual 3D (v6, C-Motion Inc, Germantown, USA) where raw trajectories were low-pass filtered (Butterworth 4^{th} order, cut-off 12 Hz) and a 6DoF inverse kinematics (IK) constrained model was computed. Athlete mass centres were computed using the model described by de Leva (1996). Additionally, the sled was modelled as a rigid object with uniformly distributed mass. Filtered marker data and mass centre locations were exported to a custom Python script (v3.7, Python Software Foundation, USA) where mass centre derivatives were computed using a finite central differences method and touch-down (TD) and toe-off (TO) events were computed according to the method described by Handsaker et al. (2016). Computing TD and TO events permitted the calculation of step characteristics including step length (SL), step frequency (SF), step time (ST), ground contact time (GCT) and flight time (FT).

This research utilised a modified version of the computer vision based foot contact detection algorithm presented by Evans et al., (2018). Challenging lighting conditions caused severe problems for typical background subtraction methods, preventing robust segmentations of the athlete. As such, background subtraction was replaced with CDCL-human-part-segmentation (Lin et al., 2019) which implements a convolutional neural network (CNN) to detect and segment body parts from the background. This approach proved robust to challenging lighting conditions and had the added advantage of not segmenting the sled as foreground. To detect approximate foot contact locations and timings, foreground segmentations were fused and occupancy maps of the ground plane were computed. The resulting occupancy maps represent the additive projection of each camera's foreground mask to the ground plane, which in this case was set at 0.025 m above the ground to avoid partial occlusions created by the sled.

Foot position was further refined by initialising an approximately foot-sized 3D bounding box along the axis representing the direction of travel. The position of the bounding box was initialised using foot contact information from the ground plane occupancy maps before the position was optimised to fit the foot. Further refinement of TD and TO event timings was achieved by tracking the foot in individual camera views. The foot-sized bounding box was projected into each camera view and each 2D image was split into vertical slices. Colour and gradient based image features were computed for each slice permitting the tracking of vertical displacement. Tracking begins at the frame where the contact has the largest area in the ground plane occupancy map. A forwards and backwards pass detects the last frame where the foot is in contact with the ground and the first frame where the foot is in the air thus providing TD and TO event timings. Again, ascertaining TD and TO locations and timings permitted the computation of step characteristics (SL, SF, GCT, ST and FT).

Sled motion was tracked using an occupancy map at the level of the hand which was holding the sled handle and the centroid of this binary topological structure was computed to provide sled displacement. To acquire the location of the athlete's mass centre without markers, an open-source deep learning based pose estimation algorithm (OpenPose, Cao et al., 2017) was implemented. OpenPose utilises a multi-stage CNN for fast multi-person key point detection and was applied to all 2D camera views. 3D reconstruction and tracking of joint centres was

achieved by back-projecting the key points on the image plane into the 3D space and finding the intersect of these vectors. Joint centre data were then parsed to compute the athlete's mass centre location using de Leva's model (1996).
Sled displacement coordinates and joint centre coordinates were low-pass filtered (Butterworth 4th order, cut-off 12 Hz) and centre of mass velocities were computed using a finite central differences method before being averaged across each step. In order to evaluate system performance, results were compared using linear regression and Bland-Altman analysis.

RESULTS: Motion capture system mean reconstruction accuracy was 0.91 ± 0.76 mm for the criterion system and 0.74 ± 0.68 mm for the computer vision system. Overall for step characteristic timings, agreement between the proposed computer vision system and the criterion system (Table 1) was within 1.5 frames. The best agreement was observed for SL with mean differences of 0.001 ± 0.012 m.

Table 1: Comparison of computed step characteristics

Variable	Mean Difference (Bias)	± SD	Bias ± 1.96 SD	R^2
GCT (s)	0.008	0.015	0.037	0.10
FT (s)	-0.008	0.016	0.023	0.28
ST (s)	0.001	0.017	0.033	0.20
SL (m)	0.001	0.012	0.021	0.99
SF (m)	-0.022	0.148	0.278	0.20
Athlete CoM Velocity (m.s^{-1})	-0.015	0.020	0.331	0.81
Sled Velocity (m.s^{-1})	0.029	0.161	0.296	0.77

Mean differences between the criterion and computer vision based sled velocities were -0.015 ± 0.02 m.s^{-1}, while OpenPose derived athlete mass centre velocities were -0.029 ± 0.161 m.s^{-1} (Figure 1).

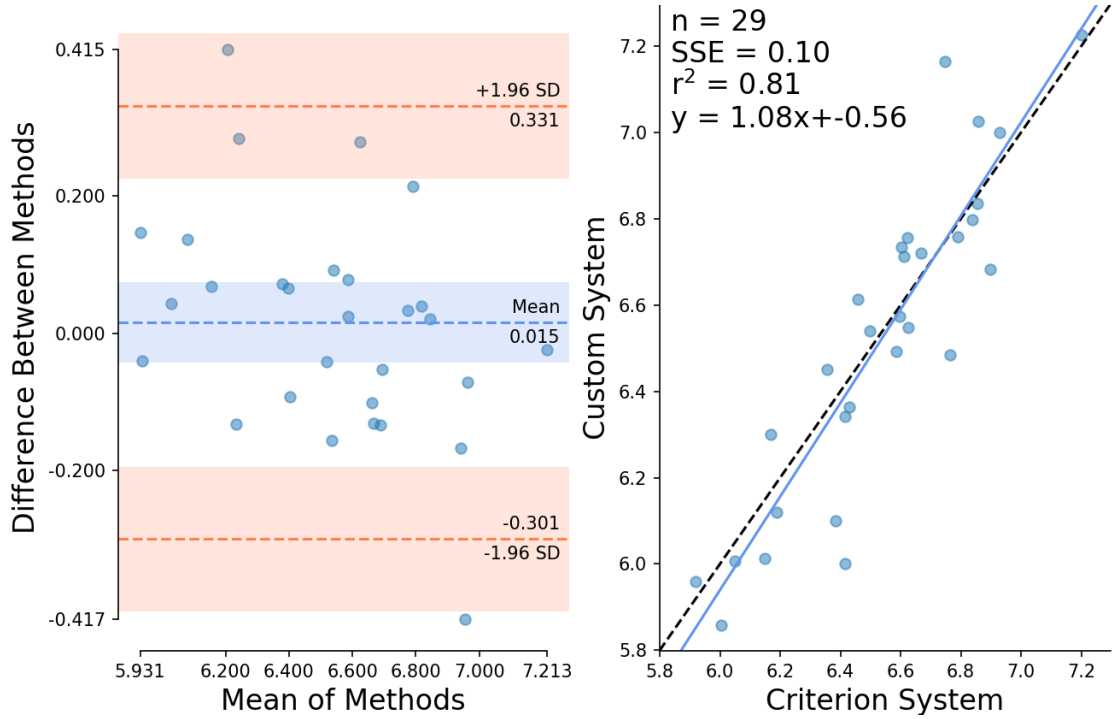

Figure 1. Bland-Altman and linear regression plots comparing step averaged athlete CoM velocity between systems. Confidence intervals are given around the mean difference and 95% limits of agreement.

DISCUSSION: This study aimed to compare the performance of a computer vision and deep learning based motion capture methodology to that of a traditional marker-based motion capture system in a challenging real-world environment (skeleton push-track). Both systems presented excellent spatial accuracy with reconstruction errors of less than 1 mm. Spatial

accuracy was further demonstrated by the low mean differences for SL (0.001 ± 0.012 mm, Table 1) with excellent agreement between systems.
Temporal variables such as GCT and FT presented higher differences (Table 1) but these were still within approximately 1.5 frames of the criterion system. Due to sled motion, the ground plane with which footfall events were detected was raised by 0.025 m which caused a systematic overestimation of GCT, underestimation of FT and thus, reduced R^2 values (Table 1). However, mean differences were similar to those reported for the widely used OptoJump™ system (Microgate, Bolzano, Italy) (0.005 ± 0.004 s, Healey et al., 2015). Furthermore, ST (GCT + FT) differences were low (0.001 ± 0.017 s) as the errors from overestimated GCT and underestimated FT effectively cancel each other out, further indicating that errors in temporal variables were likely due to the raised ground plane.
Sled velocities (Table 1) and athlete mass centre velocities derived from OpenPose key points (Figure 1) performed surprisingly well. The deep learning based approach does appear to be susceptible to some larger outliers, likely due to errors in joint centre locations. However, the OpenPose derived method showed comparable performance to other field based methods for measuring athlete running velocities such as laser distance measurement which exhibits mean errors of up to 0.41 ± 0.18 $m.s^{-1}$ (Bezodis et al., 2012). Within the skeleton push start application it appears that deep learning based pose estimation provides an alternative, non-invasive way to capture important technique related information (mass centre velocities) where more conventional systems would not be viable. More work is required to reduce measurement errors before deep learning based pose estimation methods are comparable to marker-based motion capture. However, this field of research is advancing quickly and such improvements are likely to emerge in the near future. It is also important to note that the ability of deep learning based pose estimation algorithms such as OpenPose to accurately and reliably locate joint centres has yet to be established in a biomechanics context.

CONCLUSION: A computer vision and deep learning based approach to non-invasively collect kinematic data was validated for skeleton push start performance. The novel method was applied in a challenging real world environment and application (skeleton push starts) and was able to capture representative kinematic data including step characteristics and mass centre velocities of the athlete and sled. Such an approach could be employed by coaches and practitioners to monitor technique where traditional motion capture techniques may not.

REFERENCES

Bezodis, N. E., Salo, A. I., & Trewartha, G. (2012). Measurement error in estimates of sprint velocity from a laser displacement measurement device. *IJSM*, *33*(06), 439-444.
Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2017). Realtime multi-person 2D pose estimation using part affinity fields. In *Proceedings of the IEEE (CVPR),* (pp. 7291-7299).
Colyer, S. L., Stokes, K. A., Bilzon, J. L., Holdcroft, D., & Salo, A. I. (2018). The effect of altering loading distance on skeleton start performance: Is higher pre-load velocity always beneficial? *JSS*, *36*(17), 1930-1936.
Evans, M., Colyer, S.L., Cosker, D., Salo, A.I.T., 2018. Foot contact timings and step length for sprint training. In: 2018 IEEE Vision (WACV), pp. 1652–1660.
Handsaker, J. C., Forrester, S. E., Folland, J. P., Black, M. I., & Allen, S. J. (2016). A kinematic algorithm to identify gait events during running at different speeds and with different footstrike types. *JoB*, *49*(16), 4128-4133.
Healy, R., Kenny, I. C., & Harrison, A. J. (2015). Estimating step parameters using photoelectric cells. *Proceedings of the 33rd International Conference of Biomechanics in Sports.*
Lin, K., Wang, L., Luo, K., Chen, Y., Liu, Z., & Sun, M. T. (2019). Cross-Domain Complementary Learning with Synthetic Data for Multi-Person Part Segmentation. *arXiv preprint arXiv:1907.05193*.
Zanoletti, C., La Torre, A., Merati, G., Rampinini, E., & Impellizzeri, F. M. (2006). Relationship between push phase and final race time in skeleton performance. *JSCR*, *20*(3), 579.

ACKNOWLEDGEMENT

This investigation was part-funded by CAMERA, the RCUK Centre for the Analysis of Motion, Entertainment Research and Applications, EP/M023281/1. Thank you to British Skeleton for their time and support with this project.