Citation for published version:
Needham, L, Evans, M, Wade, L, Cosker, D, McGuigan, P, Bilzon, J & Colyer, S 2021, 'DEVELOPMENT AND EVALUATION OF A DEEP LEARNING BASED MARKERLESS MOTION CAPTURE SYSTEM', Paper presented at International Conference on Biomechanics in Sports, Canberra, Australia, 3/09/21 - 7/09/21.

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

Link to publication

Publisher Rights
Unspecified

University of Bath

Alternative formats
If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 20. Jul. 2024

DEVELOPMENT AND EVALUATION OF A DEEP LEARNING BASED MARKERLESS MOTION CAPTURE SYSTEM

Laurie Needham[1], Murray Evans[1], Logan Wade[1], Darren P. Cosker[1], Polly M. McGuigan[1], James L. Bilzon[1], Steffi L. Colyer[1]
Centre for the Analysis of Motion, Entertainment Research and Applications, University of Bath, Bath, UK[1]

This study presented a deep learning based markerless motion capture workflow and evaluated performance against marker-based motion capture during overground running. Multi-view high speed (200 Hz) image data were collected concurrently with marker-based motion capture (ground-truth data) permitting a direct comparison between methods. Lower limb kinematic data for six participants demonstrated high levels of agreement for lower limb joint angles with average RMSE ranging between 2.5° - 4.4° for hip sagittal and frontal plane motion, and 4.2° - 5.2° for knee and ankle motion. These differences generally fall within the known uncertainties of marker-based motion capture, suggesting that our markerless approach could be used for appropriate biomechanics applications. While there is a need for high quality open-access datasets to further facilitate performance improvements, markerless motion capture technology continues to improve; presenting exciting opportunities for biomechanics researchers and practitioners to capture large amounts of high quality, ecologically valid data both in and out of the laboratory setting.

KEYWORDS: computer vision, pose estimation, validation, running.

INTRODUCTION: Recent developments in computer vision and deep learning research have accelerated the potential for highly accurate and robust markerless motion capture technologies in biomechanics applications. Most notably, pose estimation using deep convolutional neural networks, which estimate a set of sparse key points in 2D images, represent an exciting emerging technology with the potential to compliment and even replace current motion capture technologies (Kanko, Laende, Selbie, & Deluzio, 2021). However, the naive application of pose estimation methods may provide at best, planar estimates of joint centres and vector-based angles which are known to be error prone (Needham, Evans, Cosker, Wade, et al., 2021). To this end, we present an end-to-end automated workflow for computing full body 6 degrees of freedom (6DoF) kinematics from calibrated multi-camera high speed image data. The aim of this study was to present and evaluate the performance of our markerless motion capture system against marker-based motion capture during running activities.

METHODS:
System Development
Camera System - Our system utilises nine TTL-pulse synchronised machine vision cameras (JAI sp5000c, JAI ltd, Denmark) capable of capturing full HD images at 200 Hz. This provides the flexibility to create large capture volumes (~10 x 10 x 3 m) with multiple views providing robustness against occlusions. Camera calibration was achieved using a binary circle pattern calibration board to initialise each camera’s intrinsic parameters before moving the board through the volume ensuring multiple shared observations. Sparse Bundle Adjustment was used to achieve a globally optimised calibration.
2D Pose Estimation - 2D pose estimation of sparse points in each camera view is achieved using an implementation of OpenPose (Cao, Simon, Wei, Sheikh, & Ieee, 2017) as we have shown that this method generalises well to biomechanics image data (Needham, Evans, Cosker, Wade, et al., 2021). However, the modular nature of our workflow means that most pose estimation methods could be utilised at this stage, facilitating new state of the art network architectures or networks specialised to a specific task.
3D Fusion - OpenPose is used to provide keypoint coordinates on the 2D image plane independently for each view at each time instance. To associate detections across views and

track through time, we use our previously developed approach based on occupancy maps which can efficiently associate multi-person detections between viewpoints and time instances (Needham, Evans, Cosker, & Colyer, 2021). Once detections are associated between camera views, a 3D Euclidean space reconstruction of each person's keypoints can be generated. 2D keypoints are back-projected using the appropriate camera calibration resulting in a set of rays (lines in space originating at camera centres) where the 3D point closest to the $n \geq 2$ rays intersect is optimised using linear least squares. To handle common failures of the OpenPose detector such as misplaced limbs or contra-lateral label swaps, a robust RANSAC like approach is used to identify inliers and outliers for each keypoint, with the final reconstruction coming from only the inlier set.

Kalman Smoothing – 3D reconstructed joint centre trajectories are filtered using optimal state estimation in the presence of additive Gaussian noise. Specifically, a bi-directional Kalman filter which considered both previous and future states of the trajectory was used to provide a smooth output that is robust to outliers that have bypassed the person association and RANSAC processing stages. Previously, we have demonstrated this approach to be more effective for markerless pose estimation data than traditional low-pass filtering (Needham, Evans, Cosker, & Colyer, 2021).

Inverse Kinematics – 3D reconstructed keypoints are used to drive the motion of a constrained rigid-body kinematic model which can describe the position and orientation of each segment as a set of generalised coordinates. The model is scaled to a static calibration trial and the pose globally optimised in each frame using OpenSim's (Delp et al., 2007) Inverse Kinematics (IK) tools accessed via the C++ API. Again, due to the modular nature of our workflow, other IK solvers could be implemented here.

System Evaluation

Six participants (3 males, 3 females [1.70 ± 0.10 m, 71.8 ± 12.8 kg]) provided written informed consent. Each participant performed five running trials at a self-selected speed. Motion data were captured concurrently using our experimental system and a criterion marker-based motion capture system.

Criterion data were captured using a 15-camera marker-based motion capture system (Oqus, Qualisys AB, Gothenburg, Sweden). 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. A right-handed coordinate system was defined for both systems by placing a Qualisys L-Frame in the centre of the capture volume. 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. These marker data provided points with which the spatial alignment could be optimised in a least-squares sense. 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). Following labelling and gap filling of trajectories (Qualisys Track Manager v2019.3, Qualisys, Gothenburg, Sweden) marker-based data were filtered (Butterworth low-pass, cut-off 12 Hz) and used to scale and drive the motion of the same kinematic model that was used for markerless data. To evaluate system performance, the hip (flex/ext, ad/abduction, int/external rotation), knee (flex/ext), and ankle (plantar/dorsiflex, ad/abduction) Euler angles from both systems were compared using root mean squared errors (RMSE).

RESULTS:

For hip joint angles, good agreement was observed between both systems, particularly for hip flexion/extension (mean RMSE = 2.5°) and hip ad/abduction (mean RMSE = 4.4°) (Figure 1). Larger differences were observed during early stance for hip internal/external rotation (mean RMSE = 5.6°). The knee joint exhibited good agreement between systems (mean RMSE = 4.2°) with largest differences observed around touch-down and early stance (Figure 2). Largest joint angle differences were observed for ankle dorsi/plantar-flexion (mean RMSE = 4.5°)

around late-swing and touch-down while lower differences were observed for ankle ad/abduction (mean RMSE = 5.2°) (Figure 2).

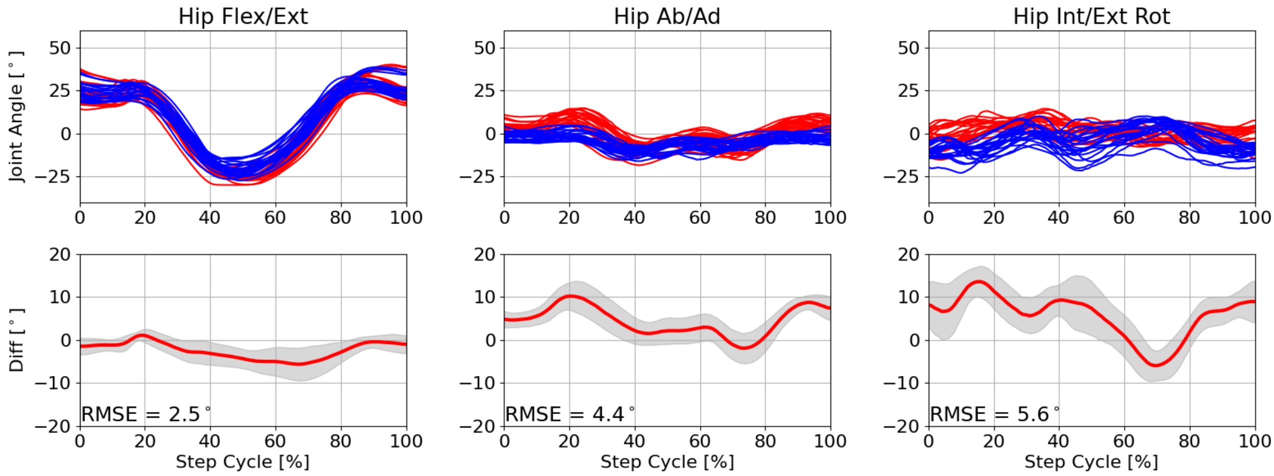

Figure 1. Upper - Hip joint angles (top) for all participant's running trials, computed from marker-based (red) and markerless (blue) data. Lower – Mean ± SD differences between both systems. Step cycles normalised from toe-off to toe-off.

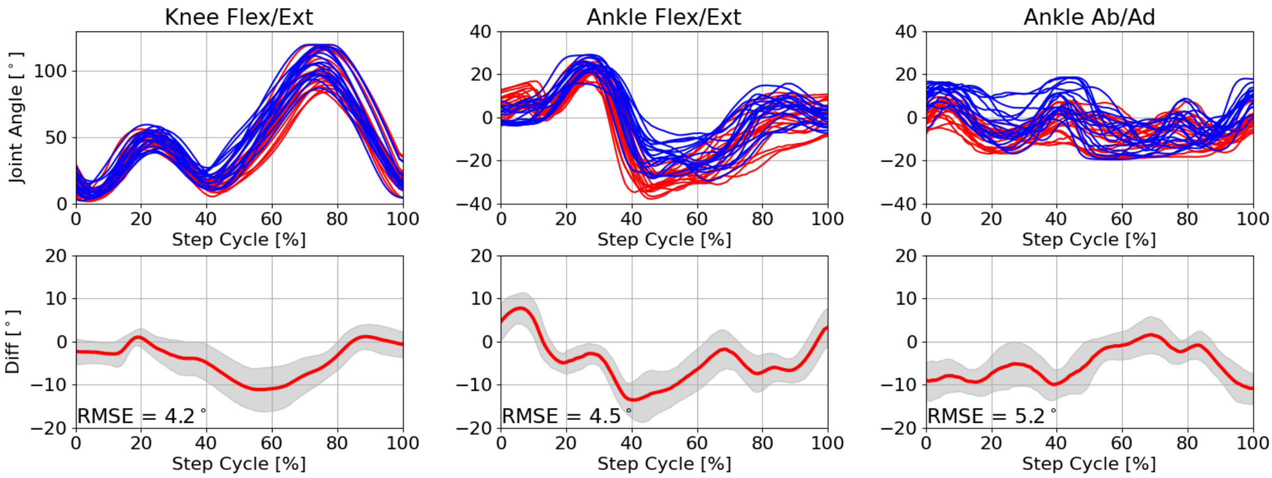

Figure 2. Upper – Knee & ankle joint angles (top) for all participant's running trials, computed from marker-based (red) and markerless (blue) data. Lower – Mean ± SD differences between both systems. Step cycles normalised from toe-off to toe-off.

DISCUSSION:

This study presents a workflow for markerless motion capture and provides a sub-sample of our evaluation work, focusing on lower limb joint angles during overground running. Our markerless system was able to capture running kinematics with a high level of agreement to marker-based motion capture, most notably for hip flexion/extension, hip ad/abduction and for knee flexion/extension (mean RMSE 2.5 - 4.4°; Figure 1 & 2). Larger differences observed for knee flexion/extension around touch-down and early stance may have been due in part to systematic differences in joint centre locations computed at the 2D pose estimation stage (Needham, Evans, Cosker, Wade, et al., 2021). However, marker-based kinematic errors due to soft tissue artefact (STA) have been documented to be highest during this period (Miranda, Rainbow, Crisco, & Fleming, 2013) which may contribute to an unknown proportion of the differences here. Differences exhibited for ankle dorsi/plantar flexion angles were likely the result of a) sub-optimal pose estimation accuracy and reliability when detecting distal foot keypoints in 2D images, b) systematic differences in knee and hip joint centre locations which have propagated down the IK chain, and c) errors in marker-based data due to an unknown proportion of STA and marker-misplacement. The results presented here are primarily limited by the performance of pose estimation methods which were pre-trained on crowd sourced datasets (e.g., COCO, MPII etc.). To facilitate improved pose estimation performance for biomechanics applications, there is a need for large open-access image datasets which are annotated with high anatomical accuracy, reliability and objectivity. By ensuring that each body segment of interest is annotated with at least three non-colinear points, methods such as the one presented in this study will most likely be able to determine segment poses with greater accuracy, robustness, and additional degrees of freedom. Furthermore, such datasets would facilitate reduced reliance on global pose optimisation methods (e.g., IK solvers) and allow independent segment pose optimisation solutions to be implemented giving users greater flexibility.

The results of this study are highly promising and the differences between systems typically fell within the known uncertainties of marker-based motion capture. For example, errors of up to ~9° have been reported for marker-based motion capture when compared to biplanar videoradiography for knee flexion/extension (Miranda et al., 2013) and ankle joint rotations (Kessler et al., 2019). As such our system could be employed for applications where the present accuracy meets the required minimum detectable change. Marker-based motion capture represents the current de-facto standard in many biomechanics applications and its limitations including STA, marker placement reliability, and joint axis crosstalk have been well documented (Kessler et al., 2019; Miranda et al., 2013). It is important that the same rigorous evaluation processes be applied to markerless motion capture so that researchers and practitioners can make informed decisions about their strengths and weaknesses. Ultimately, as markerless motion capture technologies continue to develop, we should ask ourselves if benchmarking against marker-based technologies remains the best option or whether we should seek to establish new ways of acquiring high fidelity ground-truth information that can be used to advance measurement technologies and most importantly allow us to answer new biomechanics research questions.

CONCLUSION:

We have presented a novel markerless motion capture workflow that can take calibrated multi-camera images and produce 6DoF segment data. We evaluated lower limb joint angles during overground running against established marker-based techniques and found a high level of agreement between systems. Markerless motion capture technologies are maturing at a rapid pace and present a promising approach to collect ecologically valid, real world data.

ACKNOWLEDGEMENT:

This research was funded by CAMERA, the RCUK Centre for the Analysis of Motion, Entertainment Research and Applications, EP/M023281/1 and EP/T014865/1

REFERENCES:

Cao, Z., Simon, T., Wei, S. E., Sheikh, Y. (2017, Jul 21-26). *Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields.* Paper presented at the 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI.

Delp, S. L., Anderson, F. C., Arnold, A. S., Loan, P., Habib, A., John, C. T., . . . Thelen, D. G. (2007). OpenSim: open-source software to create and analyze dynamic simulations of movement. *IEEE transactions on biomedical engineering, 54*(11), 1940-1950.

Kanko, R. M., Laende, E., Selbie, W. S., & Deluzio, K. J. (2021). Inter-session repeatability of markerless motion capture gait kinematics. *Journal of Biomechanics*, 110422.

Kessler, S. E., Rainbow, M. J., Lichtwark, G. A., Cresswell, A. G., D'Andrea, S. E., Konow, N., & Kelly, L. A. (2019). A Direct Comparison of Biplanar Videoradiography and Optical Motion Capture for Foot and Ankle Kinematics. *Frontiers in Bioengineering and Biotechnology, 7*. doi:10.3389/fbioe.2019.00199

Miranda, D. L., Rainbow, M. J., Crisco, J. J., & Fleming, B. C. (2013). Kinematic differences between optical motion capture and biplanar videoradiography during a jump–cut maneuver. *Journal of Biomechanics, 46*(3), 567-573.

Needham, L., Evans, M., Cosker, D. P., & Colyer, S. L. (2021). Can Markerless Pose Estimation Algorithms Estimate 3D Mass Centre Positions and Velocities during Linear Sprinting Activities? *Sensors, 21*(8), 2889.

Needham, L., Evans, M., Cosker, D. P., Wade, L., McGuigan, P. M., Bilzon, J. L., & Colyer, S. L. (2021). Human Movement Science in The Wild: Can Current Deep-Learning Based Pose Estimation Free Us from The Lab? *bioRxiv*.