Scandinavian Journal of Medicine & Science in Sports

WILEY

ORIGINAL ARTICLE OPEN ACCESS

Automating Video-Based Two-Dimensional Motion Analysis in Sport? Implications for Gait Event Detection, Pose Estimation, and Performance Parameter Analysis

Marion Mundt[1] | Steffi Colyer[2] | Logan Wade[2] | Laurie Needham[2] | Murray Evans[2] | Emma Millett[3,4] | Jacqueline Alderson[1]

[1]UWA Tech & Policy Lab, The University of Western Australia, Crawley, Western Australia, Australia | [2]The Centre for the Analysis of Motion, Entertainment Research and Applications, University of Bath, Bath, UK | [3]New South Wales Institute of Sport, Sydney, New South Wales, Australia | [4]Athletics Australia, Albert Park, Victoria, Australia

Correspondence: Marion Mundt (marion.mundt@uwa.edu.au)

Received: 17 March 2024 | **Revised:** 12 June 2024 | **Accepted:** 25 June 2024

Funding: This research was part-funded by CAMERA, the RCUK Centre for the Analysis of Motion, Entertainment Research and Applications, EP/M023281/1 and EP/T014865/1, the Australian Institute of Sport, AIS Research Grant Number 0003223, and the UWA Tech and Policy Lab at the University of Western Australia.

Keywords: 3D marker trajectory projection | knee angle | OpenPose | running | sampling frequency

ABSTRACT

Background: Two-dimensional (2D) video is a common tool used during sports training and competition to analyze movement. In these videos, biomechanists determine key events, annotate joint centers, and calculate spatial, temporal, and kinematic parameters to provide performance reports to coaches and athletes. Automatic tools relying on computer vision and artificial intelligence methods hold promise to reduce the need for time-consuming manual methods.

Objective: This study systematically analyzed the steps required to automate the video analysis workflow by investigating the applicability of a threshold-based event detection algorithm developed for 3D marker trajectories to 2D video data at four sampling rates; the agreement of 2D keypoints estimated by an off-the-shelf pose estimation model compared with gold-standard 3D marker trajectories projected to camera's field of view; and the influence of an offset in event detection on contact time and the sagittal knee joint angle at the key critical events of touch down and foot flat.

Methods: Repeated measures limits of agreement were used to compare parameters determined by markerless and marker-based motion capture.

Results: Results highlighted that a minimum video sampling rate of 100 Hz is required to detect key events, and the limited applicability of 3D marker trajectory-based event detection algorithms when using 2D video. Although detected keypoints showed good agreement with the gold-standard, misidentification of key events—such as touch down by 20 ms resulted in knee compression angle differences of up to 20°.

Conclusion: These findings emphasize the need for *de novo* accurate key event detection algorithms to automate 2D video analysis pipelines.

1 | Introduction

Automated tools to support motion analysis in sports are commonplace during training and competition. A simple and frequently used modality to capture motion is two-dimensional (2D) video. Videos are currently evaluated by biomechanists, who typically manually (1) determine key events, (2) annotate joint centers, and (3) calculate spatial, temporal, and kinematic parameters to provide performance parameter reports to coaches and athletes. Aided by recent advances in machine learning, computer vision, and the emergence of open-source pose estimation models (PEMs) trained on large image datasets, there exists potential for automating components of this workflow, serving to relieve biomechanists of high-cost labor tasks that impact their availability for direct engagement with coaches and athletes.

Sport practitioners require high-fidelity motion analysis outside of lab settings to analyze technique, and to detect small, but meaningful changes in performance and injury risk factors. Additionally, the measurement tool adopted should not inhibit an athlete's natural performance. While a plethora of studies and commercial smartphone applications purport to analyze motion in 2D video, for example [1–3], there remains limited evidence concerning the agreement of the proposed automated methods and tools compared with the *de facto* standard, three-dimensional (3D) marker-based optical motion capture.

For decades, 3D marker-based motion capture has been the gold-standard technology for motion analysis in biomechanics, which can be ascribed to the high accuracy (error in dynamic trials <2 mm) [4] in determining kinematic parameters. However, extensive experimental setups necessitate laboratory environments reducing ecological validity; extensive marker sets to determine accurate 3D motion which may restrain natural movement patterns and contain clustered markers that become occluded and result in inaccurate tracking; high-cost multiple camera systems for large reconstruction volumes; and marker placement errors and soft tissue artifacts which limit accuracy (for full review see Ref. [5]). These limitations foster the desire for systems with more refined and easier to implement measurement setups, ideally using a single or the most minimally viable number of 2D video cameras.

The use of 2D video cameras combined with manual digitizing of anatomical landmarks has been the standard tool for undertaking kinematic motion analysis in field-based training and competition environments. Using multiple calibrated video cameras, the 3D coordinates of anatomical landmarks have been determined without attaching markers to the athlete. However, the qualitative digitizing process this entails is labor intensive, time-consuming, and prone to human error. To overcome some of these limitations, 2D video-based motion analysis has recently progressed from manual annotation of images to computer vision-based automated marker-less systems [6–9].

Computer vision-based human PEMs are one of the most promising tools to automate the digitizing requirement of current 2D video analyses workflows that are reliant on human labor. These automated models determine keypoints that are intended to represent anatomical-related landmarks in images, with precision and accuracy in some circumstances comparable to 3D marker-based systems [10] but with the flexibility of a video-based system that can be used beyond the confines of the laboratory. Unsurprisingly, these advantages have fuelled biomechanics researchers to rapidly adopt these frameworks, although there is little information surrounding the agreement of 2D PEMs derived outputs against 3D optical motion capture anatomical landmark locations [10] or manually digitized landmarks [11]. Further, crowded scenes and poor video quality can limit the detection of the person of interest in real-world videos.

The use of 2D human pose information is rarely sufficient for a thorough biomechanical assessment which has resulted in research aimed at the accurate estimation of 3D pose. Although there are approaches to directly determine 3D pose from a single 2D image [12], the more common approach uses multiple calibrated cameras (similar to optical motion capture systems) that allow the reconstruction of 3D pose using direct linear transformation [13–15], where 3D pose of rigid segments is fully described by six degrees-of-freedom (DoF); three translations and three orientations. Three non-collinear points per segment are necessary to determine these six DoFs. However, most computer vision-based PEMs are trained on datasets defining a segment using only two keypoints, most commonly the segment endpoints. It is important to note that this information does not provide the orientation of the segment. In computer vision, the reconstruction of 3D pose most commonly refers to the reconstruction of a point in the global 3D volume. From a biomechanists perspective, this can be considered a pseudo-3D approach given that this does not enable the determination of each individual body segment pose in six DoF [5, 15].

Although 3D reconstruction limitations of 2D keypoints determined by multiple cameras are well-established [5, 15], 2D keypoint agreement has primarily been compared with optical motion capture trajectories recorded in 3D. Comparison of three publicly available 2D PEMs (OpenPose, AlphaPose, DeepLabCut) used to reconstruct 3D keypoints with 3D marker-based motion capture found systematic differences in hip and knee joint center identification (30–50 mm), most likely due to mislabelling in the PEM training datasets. Location differences ranging from 1 to 15 mm for ankle joint center dependent on the activity: running, walking, or jumping have also been reported [6]. 3D joint centers determined by OpenPose compared to marker-based motion capture resulted in a mean absolute error of less than 30 mm in 80% of walking, countermovement jumping, and ball throwing trials [16]. These results indicate that marker-less motion capture is a highly promising, emerging technology.

Studies comparing 2D keypoint location outputs compared with 3D marker trajectories projected onto the same 2D image frame of reference are warranted. To date, only one validation study [8] has undertaken this like-for-like 2D PEM versus gold-standard keypoint location comparison. This study found that unobstructed OpenPose keypoints used to determine sagittal and frontal plane hip and knee joint angles were equivalent or near-equivalent to those determined using a gold-standard 3D marker-based system (camera-side 7%–9% difference), while obstructed keypoints were outside the author's acceptable

range (occluded-side 14%–15% difference). Alternatively, both obstructed and unobstructed results for the ankle joint angle showed higher errors outside the acceptable range (camera-side 14%, occluded-side 20%) [8].

To automate 2D motion analysis workflows an accurate determination of critical key events such as touch down (TD) and toe off (TO) is necessary to provide information about temporal parameters such as contact time or cadence, and to define the correct frames for the analysis of kinematic parameters such as knee angle at the TD event. The gold-standard method to detect these events is the use of force thresholds determined by ground-embedded force plates in a laboratory environment [17, 18]. However, the reliance on a force plate means this method is only available for a limited number of steps in lab-based settings and is completely absent in outdoor settings. Kinematics-based algorithms (KBAs) that use marker trajectories, windowing, and peak detection to determine TD and TO [17–21] have been suggested as sufficiently accurate alternatives to the force plate-reliant gold-standard method. These KBAs, developed using 3D motion capture marker trajectories, rely on the vertical or anterior–posterior trajectories of markers positioned on the calcaneus (HEEL), the head of the second proximal phalange (TOE), or first and fifth metatarsal heads (MT1, MT5) of the foot. In theory, and ignoring any artifact introduced by the capture medium, a sagittally located camera positioned perfectly orthogonal (without tilt and lens distortion) to the plane of motion being recorded, should provide an equivalent 2D keypoint vertical and anterior–posterior trajectory to its 3D motion capture trajectory counterpart. Although keypoints for the heel, big toe, and small toe (HEEL, MT1, and MT5) are represented in the 25 keypoint 2D OpenPose model [22], no modern off-the-shelf PEM includes a keypoint at the second proximal phalange (TOE). Currently, there is no evidence that 3D gait event detection algorithms can be adapted for use with 2D PEMS, or that using 2D keypoints in place of 3D marker trajectories is a valid substitution. Other computer vision-based algorithms have been proposed for detecting key events in running, but the bulk of these rely on multiple cameras [23, 24], require computationally high-cost and low-accuracy background subtraction techniques [25], or suffer from low-accuracy motion blur [26].

A further challenge encountered in the use of commercially available 2D cameras in lieu of specialized 3D motion capture setups is the available sampling rate. In many video-based applications frame rates of 25–50 Hz are common [27], while biomechanically based 3D motion analysis, applying Nyquist Theorem rules to establish optimal capture rates, most commonly records at 100–200 Hz [28]. Lower sampling rates run the risk of errors in temporal parameter identification as the frame representing the key event is substantially or entirely missed (e.g. the frame of foot to ground contact). This error increases with lower sampling rates and faster movements.

To systematically analyze the steps required to automate motion analysis workflows using 2D video input, this paper analyses: (1) the applicability of KBAs to 2D video data at four sampling rates (200, 100, 50, and 25 Hz); (2) the agreement of 2D keypoints estimated by the OpenPose 2D PEM compared with 3D marker trajectories projected to an equivalent 2D camera field of view; and (3) the influence of event detection frame errors on contact time and the sagittal knee joint angle at the key critical events of TD and foot flat (FF).

2 | Materials and Methods

2.1 | Data Collection

The dataset [13] contains 10 running and 10 walking trials from 15 participants which were performed at self-selected speeds (total trials running $n = 149$, walking $n = 149$) on a straight, level walkway in an indoor laboratory. The participants were healthy volunteers who consented to take part in the study and have their 2D and 3D video, and image data included in a public dataset. The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of the University of Bath (EP1819052 25/07/19).

Data were collected from 7 males (1.82 ± 0.11 m, 85.7 ± 11.1 kg, 27 years) and 8 females (1.65 ± 0.08 m, 63.2 ± 6.0 kg, 25 years). Each participant was equipped with a full-body marker set comprising 38 individual markers and eight 4-marker clusters. This allowed for a full-body six DoF model (bilateral feet, shanks and thighs, pelvis and thorax, upper and lower arms, and hands) [13].

Motion was captured simultaneously by nine high-definition color video cameras (JAI SP-5000C-CXP2-C, 1920 × 1080 px, 200 Hz) (Figure 1), a 15-camera optical motion capture system (Qualisys Oqus, 200 Hz) synchronized to the video camera system, and two ground embedded force plates (Kistler 9287CA, 1000 Hz). The video cameras' intrinsic parameters were determined by capturing observations of a circle-grid calibration board that was moved through the scene to maximize visibility to individual cameras. This creates a graph where every camera is connected to all other cameras through shared board observations. Intrinsic camera parameters were initialised using OpenCV based on [29]. Extrinsic parameters were determined using a Ceres Solver based bundle adjustment [30].

2.2 | Data Processing

The marker data were labeled (Qualisys Track Manager) and exported to Visual 3D (v6, C-Motion Inc). Marker trajectories were low-pass filtered (Butterworth 4th order, cut-off 12 Hz) prior to computing the joint centers as the point 50% between the medial and lateral marker for all joints except the hip joint center, which was computed using regression equations [31], and the shoulder joint center, which was determined using a 25 mm inferior offset from the acromioclavicular shoulder marker. TD and TO events were determined using Visual 3D's in-built "Automatic Gait Events" method, which corresponds to a definition of stance phase at a threshold of 20 N in the vertical ground reaction force. Video data were processed using OpenPose [22] to automatically detect 25 visually approximated joint centers and anatomical landmarks, so-called keypoints. OpenPose is a pre-trained off-the-shelf human PEM that is trained on the MPII Human Pose dataset (ca. 40 000 images) [32], the COCO dataset (ca. 200 000 images) [33], and the COCO+foot dataset (ca. 15 000) [22]. During processing, the u-v-coordinates of each

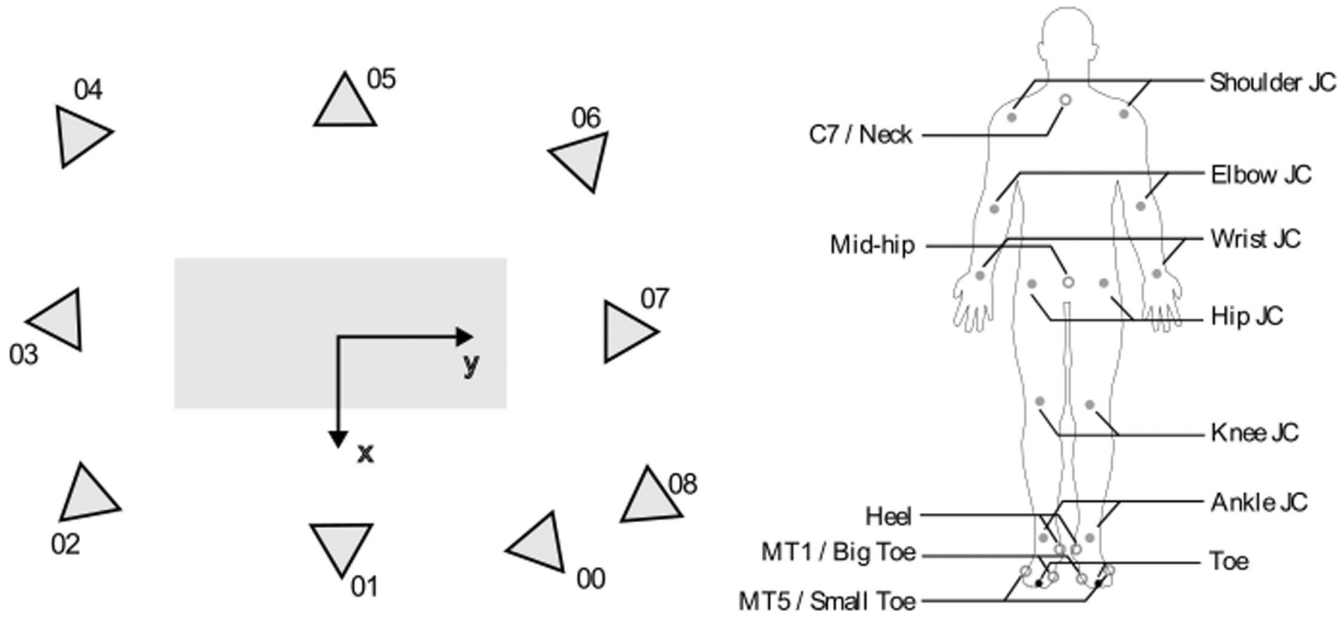

FIGURE 1 | Nine high-definition JAI color video cameras were placed about waist height surrounding the force plate. Participants were approximately 500 pixels tall when standing at the center of the volume. For each camera view, OpenPose and projected keypoints were determined. Keypoints displayed as shaded gray circles correspond to joint centers and should be similar for both OpenPose and projected, while keypoints displayed as hollow gray circles are anatomical landmarks that are likely to exhibit larger differences due to different applied definitions within each respective model. The keypoint displayed in black was used only for gait event detection and does not exist in the OpenPose model.

keypoint are estimated image by image in a video sequence. Estimated keypoints were filtered using a Kalman filter [13] to add a temporal tracking for smoothing the data. This data is referred to as "2D OpenPose keypoints".

Using the extrinsic and intrinsic parameters of the video cameras, motion capture marker trajectories were projected to each 2D video camera field of view using OpenCV's (https://opencv.org) project points method in Python (v3.10). This data is referred to as "2D projected keypoints". Bilateral joint centers of the shoulder, elbow, wrist, hip, knee, and ankle were extracted from both, motion capture and OpenPose data. The neck keypoint was described by the projected C7 marker, and the mid-hip keypoint was calculated as the middle of both hip joint centers. The bilateral heel, big toe, and small toe keypoints were defined as the projected calcaneus marker, and markers of the first and fifth metatarsal bones respectively. Some gait event detection algorithms [19] require a marker to be placed on the second toe, so this 3D marker was also extracted and projected into the 2D image (Figure 1).

2.3 | Data Analysis

Data were analyzed descriptively using mean differences in time of event detection and the contact time. Mean Euclidean distances between OpenPose and projected keypoints and the Euclidean distance in the stance phase time series data were calculated to assess keypoint agreement. The knee angle based on OpenPose and projected keypoints was calculated at the TD frame and with an offset of 20 ms pre and post TD. Repeated measures standard deviation (SD) and limits of agreement (LoA) were calculated for all parameters [8, 34].

2.3.1 | Agreement of 2D OpenPose Keypoints With 2D Projected Keypoints in Walking and Running

Both running and walking trials ($n = 298$) were investigated to examine the influence of different movement speeds on the agreement between OpenPose and projected keypoints. To compare the agreement of keypoints detected by OpenPose, the keypoint $[x, y]$ location in the image frames was compared to the $[x, y]$ location of the projected keypoints. All OpenPose and projected keypoints were extracted during left and right limb stance, the phase most relevant for biomechanical analysis, determined using the force plate signal and normalized to 101 frames representing 100% of the stance phase.

To compare the results of different camera views and to account for individuals of varying stature, the Euclidean distances were normalized to the average distance between the projected left and right shoulder and hip joint center locations for every frame in each trial [9]. The comparison results provided are the best-case scenario, given the stance phase always took place in the center of the image with the least lens distortion.

2.3.2 | Event Detection: Down and TO in Running

A variety of algorithms have been proposed for gait event detection in running [17–21]. Therefore, the analysis of TD and TO detection was restricted to the running condition ($n = 149$). The Foot-Contact-Algorithm (FCA) [19] that used the vertical component of the HEEL, MT5, and TOE marker trajectories to determine TD and TO events was particularly designed for running with different foot strike patterns and reported the best results

with −1 to 1 ms difference in TD and TO detection when compared to a force plate [19].

To investigate the suitability of the FCA with 2D keypoint trajectories as inputs, only the sagittal camera view (Camera 01) was analyzed as this camera view provided comparable vertical and anterior–posterior data to the 3D marker trajectories the algorithm has been designed for. As recommended in the original implementation [19], all marker and keypoint data were low-pass filtered using a 4th-order Butterworth filter (cut-off 50 Hz for sampling rates of 100 and 200, and 20 Hz for sampling rates of 25 and 50 Hz). To be applicable to forefoot and heel strike running styles, the algorithm first detects an approximate TD based on the minimum vertical position of markers positioned on the HEEL and MT5. The earlier event defines TD_{approx}. In a window of $[TD_{approx} - 50\,ms,\ TD_{approx} + 100\,ms]$ the TD event is determined as the time frame of maximum acceleration of the MT5 or HEEL marker for forefoot and heel strike running styles, respectively. TO is then determined in a window of $[TD_{approx} + 100\,ms,\ TD_{approx} + 400\,ms]$ as either the time frame of minimum vertical position or maximum vertical acceleration of the TOE marker; the earlier event defines TO [19].

In a second analysis, the TOE marker/keypoint that is not available in OpenPose has been substituted with the MT1 3D marker, MT1 2D projected keypoint, and MT1 OpenPose keypoint. To investigate the influence of sampling rates, data was downsampled by selecting only the n-th frame (25 Hz: $n = 8$, 50 Hz: $n = 4$, 100 Hz: $n = 2$) of the original sequence sampled at 200 Hz.

2.3.3 | Performance Relevant Parameters in Running

Stance phase contact time and the knee compression angle, the difference in the sagittal knee joint angle at TD and FF, were calculated as secondary parameters that rely on accurate event detection and are relevant to running performance analysis ($n = 149$). The influence of misidentified gait events on the knee angle was analyzed for a detection error of 20 ms pre- and post-TD event, corresponding to a single frame at the commonly used 2D video capture rate of 50 Hz; a threshold we define acceptable. This permutation represents the worst-case scenario where the correct frame is detected but there is no frame depicting the TD key event. The analysis was restricted to right foot contacts on the force plate during running based on the camera 01 field of view (Figure 1).

3 | Results

3.1 | Agreement of 2D OpenPose Keypoints With 2D Projected Keypoints in Walking and Running

The LoA was calculated for all 2D keypoints relevant to the determination of foot strike and secondary parameters. When calculating SD and LoA across all camera views and movement types for individual 2D keypoints, the normalized OpenPose ankle keypoints showed the smallest bias compared to the projected keypoints (bias: 0.04, SD: 0.07, LoA: −0.09 to 0.18), while the MT1/MT5 keypoints showed the largest bias but similar agreement (bias: 0.10, SD: 0.08, LoA: −0.06 to 0.25) to the ankle keypoints. All remaining keypoints (heel, knee, hip, shoulder) returned similar bias and agreement (bias: 0.07, SD: 0.07, LoA: −0.06 to 0.21) (Table S1).

When grouping all keypoints and movements and analyzing single camera views to assess the influence of camera view, SD and LoA between OpenPose and projected keypoints were similar across camera views (bias: 0.07, SD: 0.05, LoA: −0.03 to 0.18), excluding camera 2 and 5 which returned a slightly larger bias and a larger spread (bias: 0.09, SD: 0.12, LoA: −0.14 to 0.32) (Table S2).

SD and LoA for all keypoints and camera views across running and walking conditions showed no differences (running: bias: 0.08, SD: 0.07, LoA: −0.06 to 0.23; walking: bias: 0.07, SD: 0.07, LoA: −0.07 to 0.22). The Euclidean distance was consistent for the joint center keypoints throughout the gait cycle, while a larger difference could be observed towards the end of the stance phase for foot keypoints (Figure 2), potentially impacting TD and TO detection.

3.2 | Event Detection in Running

As displayed in Figure 3 and outlined in Table S3, TD detected from 3D marker data captured at 200 and 100 Hz showed the most accurate and precise results compared to the ground-truth TD event detected using a force plate. The mean bias was −7 and −13 ms (SD: 23 and 25 ms, LoA: −51 to 38 and −62 to 35 ms). The bias equaled approximately one video frame. For lower sampling rates of 50 and 25 Hz, the bias was −5 and 21 ms, respectively, and displayed a widespread (SD: 43 and 50 ms, LoA: −89 to 79 and −77 to 119 ms). Using the TOE marker as input and a measurement frequency of at least 100 Hz were necessary to achieve a good agreement (bias −8 ms, SD: 44 ms, LoA: −95 to 80 ms) for TO detection.

The use of 2D OpenPose and projected keypoint data as inputs resulted in an overall loss of accuracy and precision. 2D projected keypoints showed a mean bias in TD detection of 38 and 32 ms (SD: 87 and 83 ms, LoA: −132 to 207 and −131 to 195 ms) for a sampling rate of 200 and 100 Hz. OpenPose keypoints showed a smaller mean bias of −3 and −11 ms but a larger spread (SD: 124 and 114 ms, LoA: −245 to 240 and −235 to 213 ms).

The FCA [24] relies on a TOE marker to detect TO. Using the 2D projected keypoints, TO was detected with a mean bias of 44 and 43 ms for frame rates of 200 and 100 Hz (SD: 100 and 97 ms, LoA: −152 to 240 and −147 to 233 ms), which is in a similar range to TD detection. Replacing the TOE marker, which is not available in any open source 2D PEM, with the 2D projected MT1 keypoint resulted in a mean bias of 131 and 127 ms (SD: 92 and 87 ms, LoA: −49 to 311 and −43 to 297 ms), indicating a delay in key event detection but a similar spread. For the 2D OpenPose MT1 keypoint, this delay was smaller with a mean bias of 64 and 59 ms, but increased spread (SD: 126 and 116 ms, LoA: −183 to 311 and −169 to 286 ms).

3.3 | Secondary Parameters: Contact Time and Sagittal Plane Knee Angle in Running

Based on the results for TD and TO detection, the most accurate and precise contact time was determined using 3D markers, specifically a TOE marker at a capture frequency of 100 or 200 Hz

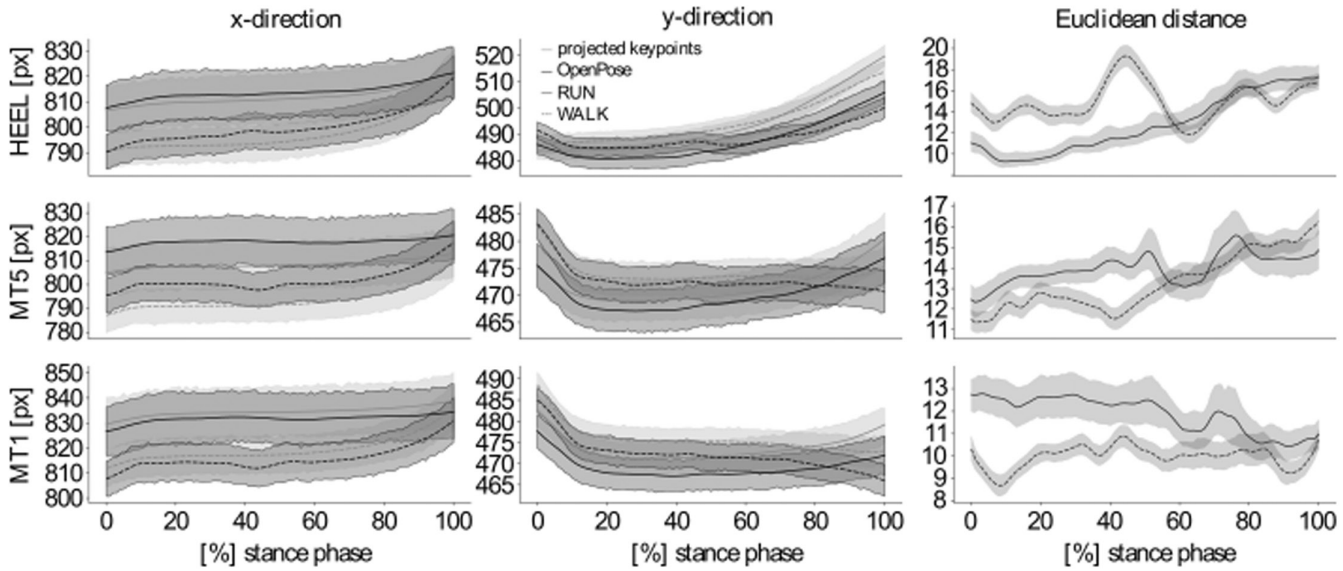

FIGURE 2 | Mean and standard deviation of the foot keypoints and the Euclidean distance between projected and OpenPose keypoints for Camera 08.

returned a mean bias of −1 and 6 ms (SD: 33 and 38 ms, LoA: −65 to 64 and −69 to 81 ms) (Figure 4). For projected 2D keypoints that included a TOE marker the agreement was slightly decreased, with a mean bias of 7 and 11 ms (SD: 41 and 44 ms, LoA: −75 to 88 and −76 to 99 ms). Using the MT1 marker instead of the TOE, contact time was increased for marker data to a mean bias of 100 ms for both measurement frequencies (SD: 47 and 51 ms, LoA: −7 to 193 and 0 to 200 ms) and similarly for projected 2D keypoints with a mean bias of 93 and 95 ms (SD: 54 and 49 ms, LoA: −13 to 200 and −2 to 192 ms). Using the OpenPose MT1 keypoint as input, the mean bias was smaller than when using the projected MT1 keypoint (OpenPose bias 66 ms, projected keypoint bias 70 ms), but the precision was lower (SD: 65 and 63 ms, LoA: −62 to 195 and −53 to 193 ms).

The knee angle from a sagittal view was calculated for TD and FF. The difference in knee angle between these two events is a kinematic parameter that is regularly analyzed in sports applications. It describes the stiffness of the knee joint and thereby how horizontal energy can be transformed to vertical energy which is relevant for jumping tasks [35]. Using projected or OpenPose keypoints for the calculation of this parameter resulted in a mean bias of 2.1° and 1.1° for TD and FF, respectively (SD: 4.0° and 4.1°, LoA: −5.8 to 9.9 and −6.8 to 9.1°). However, as displayed in Figure 5 correct frame identification significantly impacted the findings, where the median angle at TD differed 10° and 7° for projected and OpenPose keypoints when detected one frame pre or post the TD event, and 15° and 13° for FF (Figure 5).

These errors can result in a large variation in the compression angle, the difference between the knee angle at TD and FF. As displayed in Figure 6, an offset in one frame can cause an error exceeding 10°.

4 | Discussion

This study found good agreement between keypoints estimated by an off-the-shelf PEM and 3D marker trajectories projected to 2D for walking and running. The agreement between TD and toe-off events in running calculated from OpenPose keypoints and projected keypoints was overall low, especially for video sampling rates smaller than 100 Hz. The inaccuracies in event detection showed a larger impact on sagittal knee joint angles than differences between OpenPose and projected keypoints.

One major finding of this study was that accuracy (bias) and especially precision (LoA) in event detection were compromised at a sampling rate of less than 100 Hz, with LoA exceeding 50 ms in the best-case scenario of using 3D marker trajectories as input. Although the requirement for optimal (often high) sampling rates is a foundation tenet of biomechanics [28], in practice, especially in-field servicing situations, 2D cameras with high resolution but low sampling rate are often employed. This hardware setup is a major limitation when it comes to both automatic and manual video analysis as key events can occur at times where no video frames exist. In this study, running was performed at a moderate speed ($2.8 \pm 0.3\,ms^{-1}$) and a sampling rate of 100 Hz was found to be sufficient. For movements at higher speed, a higher sampling rate will certainly be necessary.

The FCA uses vertical marker trajectories and accelerations to determine TD and TO [19]. In theory, data captured from a camera without any tilt and lens distortion, perfectly positioned orthogonal to the plane of motion, will result in the same 2D keypoint vertical trajectory as the vertical component of the equivalent 3D marker trajectory. However, we found that the projected 2D trajectory resulted in decreased detection accuracy (increased bias) and precision (large LoA) for both TD and TO events, suggesting that even in datasets captured in highly controlled laboratory environments, cameras are not perfectly aligned, and small out-of-plane movements occur. The sensitivity of algorithms to small deviations that are inherent in the data capture process is a challenge that requires consideration in ongoing attempts to automate 2D video motion analysis workflows. Compared with projected keypoints, the use

of OpenPose keypoints caused lower precision (larger LoA) and accuracy (higher bias) for event detection throughout the entire dataset potentially attributable to additional sources of error from lighting conditions, inaccuracies of the PEM due to the training dataset that was annotated by workers without biomechanical knowledge, the limited tracking of a keypoint throughout the video sequence resulting in increased noise in the trajectory, and calibration errors. The smaller delay in TO detection using OpenPose keypoints compared with projected keypoints can be explained by differences in the marker/keypoint locations and is also reflected in the keypoint agreement analysis, where MT1 and MT5 markers are placed at the metatarsal heads, while the OpenPose keypoint is detected closer to the distal phalanges. However, the missing TOE marker in the OpenPose model is a major limitation for accurate TO event detection.

Another limitation of this type of threshold and windowing-based event detection algorithm that relies on movement trajectories, is the dependency on homogeneous movement patterns and speed. 2D video analysis showed that the only participant running with a forefoot foot strike technique returned the

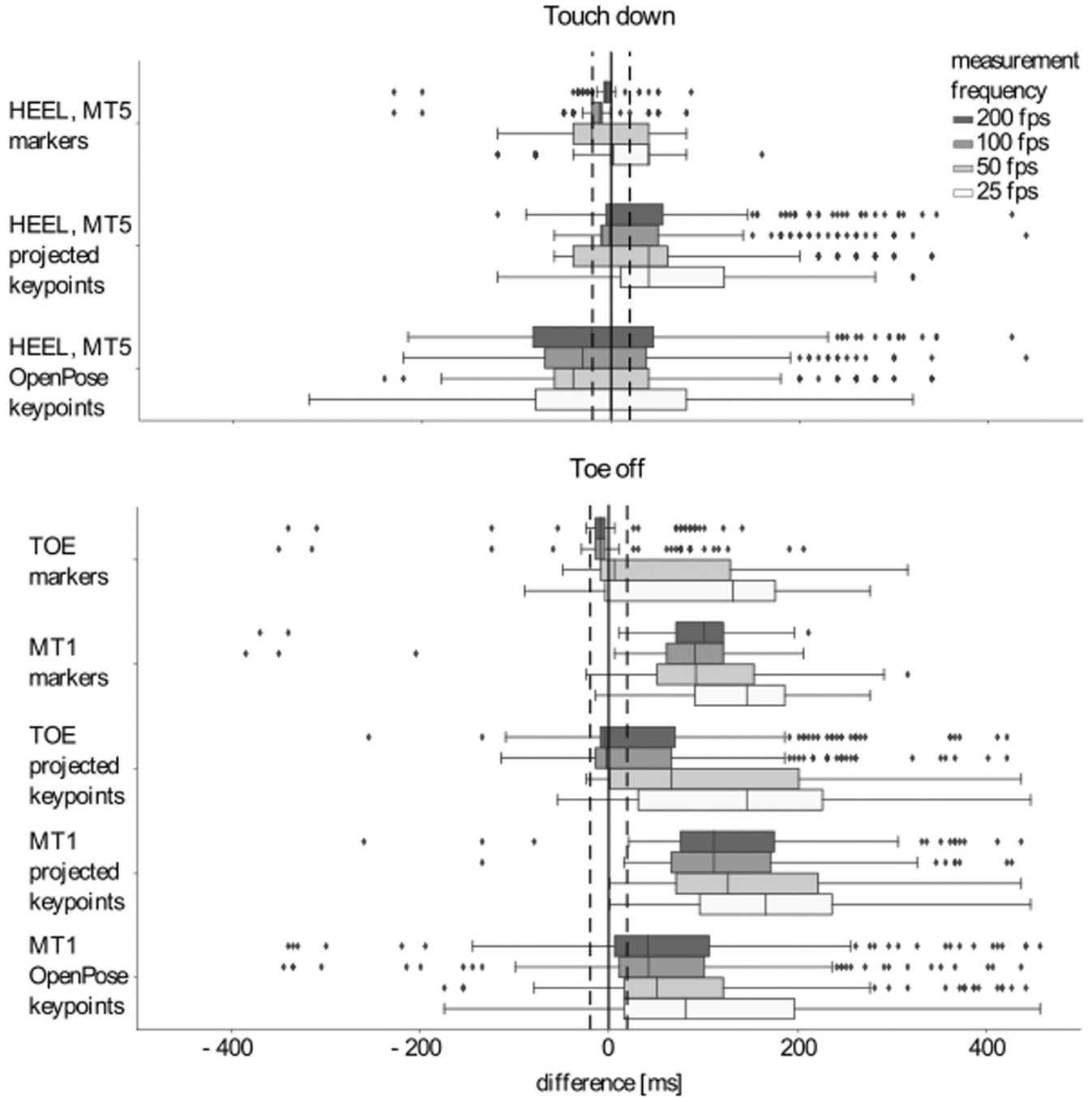

FIGURE 3 | Box plot of the difference between TD and TO detection based on different input parameters and measurement frequencies compared to TD and TO determined by a force plate. The box shows the three quartile values of the distribution along with extreme values. The "whiskers" extend to points that lie within 1.5 interquartile ranges of the lower and upper quartile, while observations that fall outside this range are displayed independently. The vertical lines show a difference of 20 ms, which is a single frame at a sampling rate of 50 Hz.

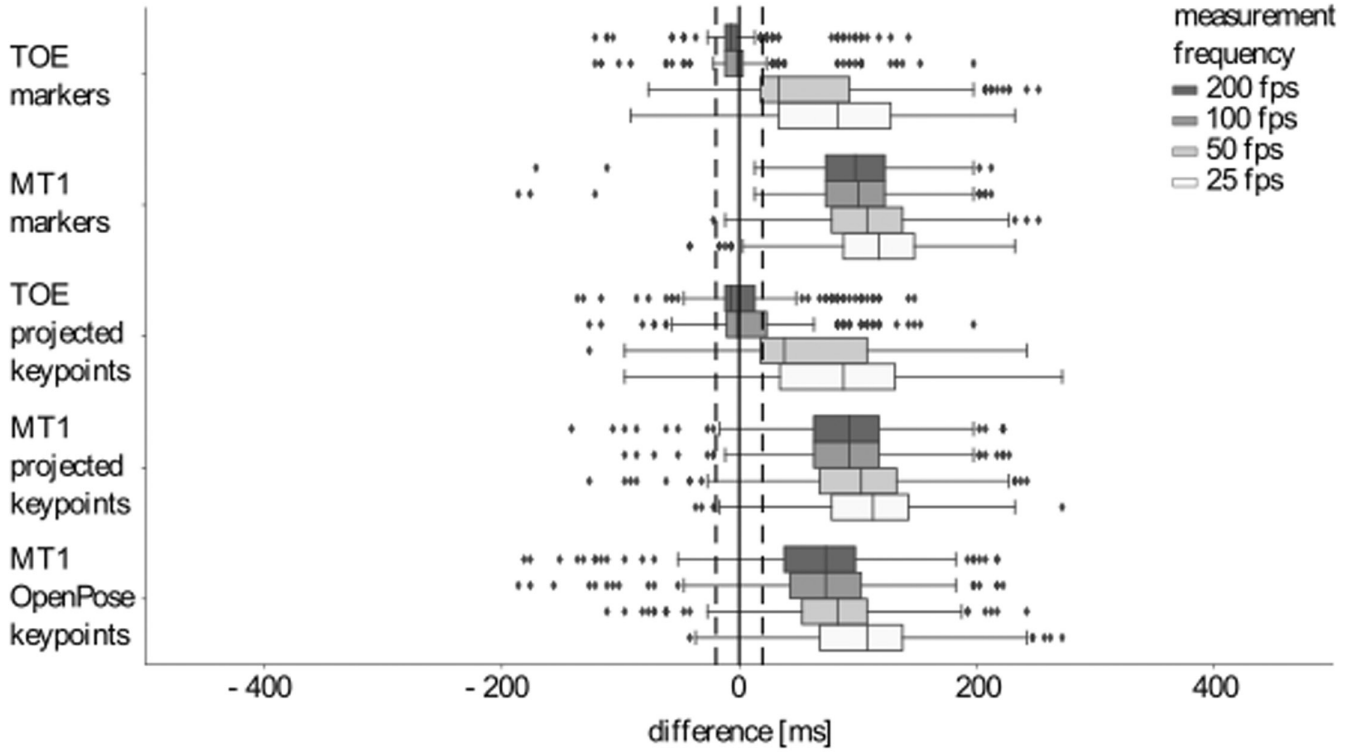

FIGURE 4 | Box plot showing contact time differences based on varying input parameters and measurement frequencies compared to ground-truth force plate contact time. The box displays the three quartile values of the distribution along with extreme values. The "whiskers" extend to points that lie within 1.5 interquartile ranges of the lower and upper quartile, while observations that fall outside this range are displayed independently. The vertical lines show a difference of 20 ms, which is a single frame at a sampling rate of 50 Hz.

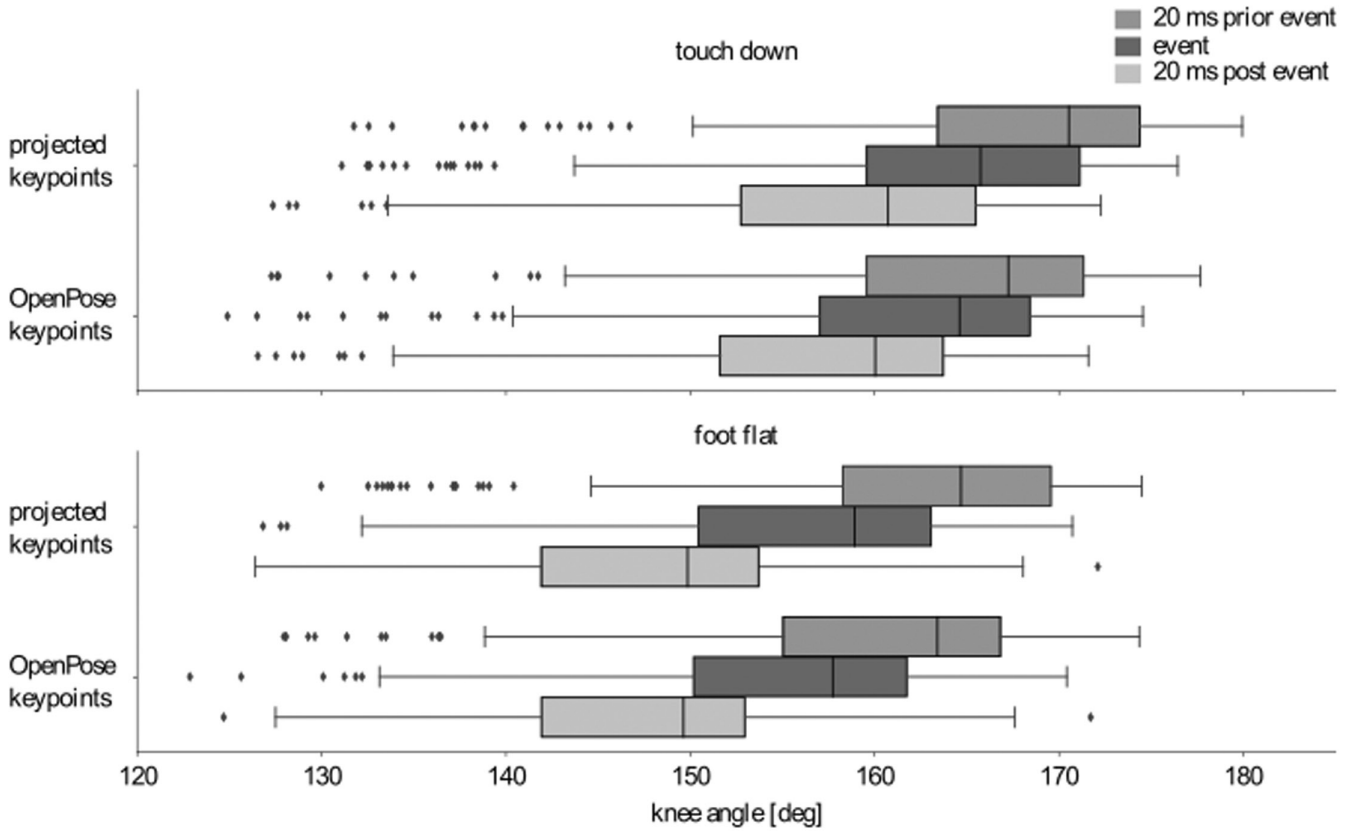

FIGURE 5 | Box plot of the knee angle calculated from the sagittal camera view for TD ±20 ms and FF ±20 ms. The box shows the three quartile values of the distribution along with extreme values. The "whiskers" extend to points that lie within 1.5 interquartile ranges of the lower and upper quartile, while observations that fall outside this range are displayed independently.

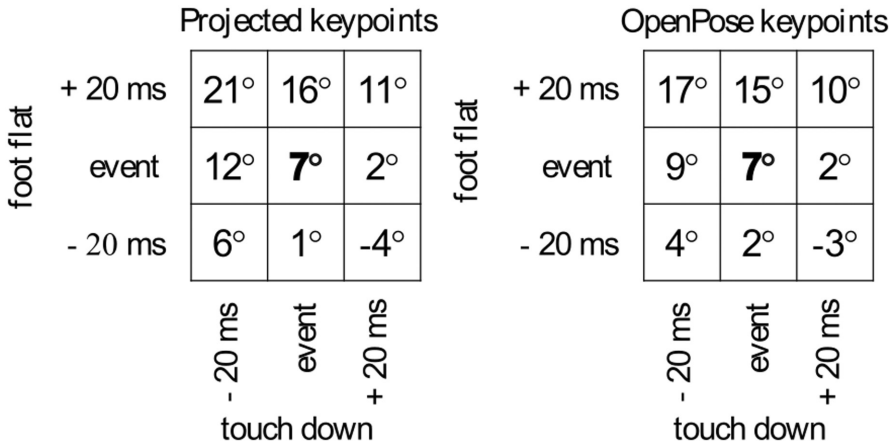

FIGURE 6 | Exemplary compression angle calculated for the correct event and ±20 ms using projected and OpenPose keypoints. The correct compression angle is 7°. The worst-case permutation, where TD is detected one frame early and FF one frame late, leads to an increased compression angle of 21° for projected and 17° for OpenPose keypoints.

poorest TD detection accuracy. Despite ostensibly being designed to accommodate different running styles [19], Leitch et al. [17] found that algorithms performed variably based on the participant's foot strike pattern, which is consistent with the findings of the present study.

The gait event detection results in this study were slightly poorer than those previously reported. This might be caused by differences in experimental design: Maiwald et al. [19] upsampled kinematic data from 240 to 960 Hz to match the force plate measurement frequency, which will improve the accuracy and calculated TD based on a vertical force threshold of 10 N and TO based on 5 N, which is half of the 20 N threshold adopted for both events in the present study. However, force thresholds used for gait event detection in the literature are inconsistent [17] but the impact is likely to be negligible for overground running (no difference between a threshold of 5 N and 10 N, detection of TD might be one frame late (+1 ms) and TO on frame earlier (−1 ms) when using a threshold of 20 N). To overcome the problem of inaccurate marker/keypoint trajectories, the implementation of machine learning methods has been suggested as a potential solution to event detection across a range of camera views and should be investigated in future work [36–41].

The agreement between OpenPose keypoints and projected keypoints was consistently high (<0.1 normalized pixels) across all camera views, however, the sagittal camera view returned the highest number of keypoint dropouts (i.e. missing data points) due to body part occlusion. This is a particularly consequential finding for applied biomechanists given sagittal camera views are considered best practice for data collection, event detection algorithms usually rely on this view, and secondary parameters are preferentially determined from data derived from this motion plane. For the foot keypoints, the agreement difference between projected and OpenPose keypoints increased towards terminal stance, serving to further impact TO detection accuracy when a motion trajectory-based algorithm is implemented. The only other study assessing 2D keypoint agreement between PEM and projected marker trajectories reported similar trends [8].

A striking finding was the high failure rate of keypoints detected from a perfectly orthogonal sagittal camera view, the best-practice preferential camera location for on-field biomechanical analysis. In this view, keypoints are rarely detected when joints are occluded by other body segments [8], while left and right limb keypoints are commonly flipped when left and right limbs cross each other in the field of view. These errors can be corrected using tracking algorithms [13] as a post-processing step but can potentially be avoided when using a PEM that labels keypoints of interest more robustly and tracks body parts continuously across frames, as opposed to traditional frame by frame estimation.

The detailed limitations of current automated 2D video analysis tools emphasize that the adoption of high-speed cameras in training and competition, alongside more robust techniques to detect TD and TO events in 2D videos from a variety of camera angles is paramount. One potential solution for improved gait event detection is the use of machine learning models trained on videos from a variety of camera views rather than a KBA focusing on single keypoint trajectory components. Further examination is required to assess whether this approach can overcome the need for a TOE keypoint or whether 2D PEMs containing these keypoints are required.

5 | Perspective

This study investigated the influence of keypoint accuracy and event detection on the accuracy of performance relevant parameters such as contact time and knee compression angle. The findings support previous studies that reported good agreement between kinematic parameters determined by 2D pose estimation and 3D marker-based motion analysis [6–9]. However, the extended investigation highlights the need for high video sampling rates and accurate event detection tools to automate the full 2D video analysis pipeline, as these have shown to have a large impact on the agreement of performance relevant parameters such as contact time and knee compression angle.

The results of this study present the best-case data collection scenario with 2D videos captured specifically for validation purposes. Consequently, these findings may not be generalisable to 2D videos captured in-field, those containing multiple people

in the camera field of view, or in less ideal lighting conditions. Despite these limitations this study is one of the first to investigate the potential impact of adopting automated pipeline procedures in 2D biomechanical analyses and offers some cautions and benefits of utilizing these emerging technologies.

Acknowledgments

Open access publishing facilitated by The University of Western Australia, as part of the Wiley - The University of Western Australia agreement via the Council of Australian University Librarians.

Ethics Statement

The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of the University of Bath (EP1819052 25/07/19).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The dataset is currently under review to be published.

References

1. Y. Aoyagi, S. Yamada, S. Ueda, et al., "Development of Smart-Phone Application for Markerless Three-Dimensional Motion Capture Based on Deep Learning Model," *Sensors* 22 (2022): 5282, https://doi.org/10.3390/s22145282.

2. L. M. Reimer, M. Kapsecker, T. Fukushima, and S. M. Jonas, "Evaluating 3D Human Motion Capture on Mobile Devices," *Applied Sciences* 12 (2022): 4806, https://doi.org/10.3390/app12104806.

3. M. T. Parks, Z. Wang, and K. C. Siu, "Current Low-Cost Video-Based Motion Analysis Options for Clinical Rehabilitation: A Systematic Review," *Physical Therapy* 99 (2019): 1405–1425, https://doi.org/10.1093/ptj/pzz097.

4. P. Merriaux, Y. Dupuis, R. Boutteau, P. Vasseur, and X. Savatier, "A Study of Vicon System Positioning Performance," *Sensors* 17 (2017): 1591, https://doi.org/10.3390/s17071591.

5. S. L. Colyer, M. Evans, D. P. Cosker, and A. I. Salo, "A Review of the Evolution of Vision-Based Motion Analysis and the Integration of Advanced Computer Vision Methods Towards Developing a Markerless System," *Sports Medicine—Open* 4 (2018): 24, https://doi.org/10.1186/s40798-018-0139-y.

6. L. Needham, M. Evans, D. P. Cosker, et al., "The Accuracy of Several Pose Estimation Methods for 3D Joint Centre Localisation," *Scientific Reports* 11 (2021): 20673, https://doi.org/10.1038/s41598-021-00212-x.

7. R. M. Kanko, E. K. Laende, E. M. Davis, W. S. Selbie, and K. J. Deluzio, "Concurrent Assessment of Gait Kinematics Using Marker-Based and Markerless Motion Capture," *Journal of Biomechanics* 127 (2021): 110665, https://doi.org/10.1016/j.jbiomech.2021.110665.

8. L. Wade, L. Needham, M. Evans, et al., "Examination of 2D Frontal and Sagittal Markerless Motion Capture: Implications for Markerless Applications," *PLoS One* 18 (2023): e0293917, https://doi.org/10.1371/journal.pone.0293917.

9. B. Van Hooren, N. Pecasse, K. Meijer, and J. M. N. Essers, "The Accuracy of Markerless Motion Capture Combined With Computer Vision Techniques for Measuring Running Kinematics," *Scandinavian Journal of Medicine & Science in Sports* 33 (2023): 966–978, https://doi.org/10.1111/sms.14319.

10. B. Sheng, L. Chen, J. Cheng, Y. Zhang, Z. Hua, and J. Tao, "A Markerless 3D Human Motion Data Acquisition Method Based on the Binocular Stereo Vision and Lightweight Open Pose Algorithm," *Measurement* 225 (2024): 113908, https://doi.org/10.1016/j.measurement.2023.113908.

11. N. Cronin, T. Rantalainen, J. P. Ahtiainen, E. Hynynen, and B. Waller, "Markerless 2D Kinematic Analysis of Underwater Running: A Deep Learning Approach," *Journal of Biomechanics* 87 (2019): 75–82, https://doi.org/10.1016/j.jbiomech.2019.02.021.

12. V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, "BlazePose: On-Device Real-Time Body Pose Tracking," *arXiv* (2020), https://doi.org/10.48550/arXiv.2006.10204.

13. L. Needham, M. Evans, L. Wade, et al., "The Development and Evaluation of a Fully Automated Markerless Motion Capture Workflow," *Journal of Biomechanics* 144 (2022): 111338, https://doi.org/10.1016/j.jbiomech.2022.111338.

14. E. D'Antonio, J. Taborri, I. Mileti, S. Rossi, and F. Patane, "Validation of a 3D Markerless System for Gait Analysis Based on OpenPose and Two RGB Webcams," *IEEE Sensors Journal* 21 (2021): 17064–17075, https://doi.org/10.1109/JSEN.2021.3081188.

15. L. Wade, L. Needham, P. McGuigan, and J. Bilzon, "Applications and Limitations of Current Markerless Motion Capture Methods for Clinical Gait Biomechanics," *PeerJ* 10 (2022): e12995, https://doi.org/10.7717/peerj.12995.

16. N. Nakano, T. Sakura, K. Ueda, et al., "Evaluation of 3D Markerless Motion Capture Accuracy Using OpenPose With Multiple Video Cameras," *Frontiers in Sports and Active Living* 2 (2020): 50, https://doi.org/10.3389/fspor.2020.00050.

17. J. Leitch, J. Stebbins, G. Paolini, and A. B. Zavatsky, "Identifying Gait Events Without a Force Plate During Running: A Comparison of Methods," *Gait and Posture* 33 (2011): 130–132, https://doi.org/10.1016/j.gaitpost.2010.06.009.

18. N. Zahradka, K. Verma, A. Behboodi, B. Bodt, H. Wright, and S. C. Lee, "An Evaluation of Three Kinematic Methods for Gait Event Detection Compared to the Kinetic-Based 'Gold Standard'," *Sensors* 20 (2020): 5272, https://doi.org/10.3390/s20185272.

19. C. Maiwald, T. Sterzing, T. A. Mayer, and T. L. Milani, "Detecting Foot-To-Ground Contact From Kinematic Data in Running," *Footwear Science* 1 (2009): 111–118, https://doi.org/10.1080/19424280903133938.

20. R. Nagahara and K. Zushi, "Biomechanics Determination of Foot Strike and Toe-Off Event Timing During Maximal Sprint Using Kinematic Data," *International Journal of Sport and Health Science* 11 (2013): 96–100, https://doi.org/10.5432/ijshs.201318.

21. J. C. Handsaker, S. E. Forrester, J. P. Folland, M. I. Black, and S. J. Allen, "A Kinematic Algorithm to Identify Gait Events During Running at Different Speeds and With Different Footstrike Types," *Journal of Biomechanics* 49 (2016): 4128–4133, https://doi.org/10.1016/j.jbiomech.2016.10.013.

22. Z. Cao, G. Hidalgo, T. Simon, S. E. Wei, and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," in IEEE Transactions on Pattern Analysis and Machine Intelligence (2019), https://doi.org/10.1109/CVPR.2017.143.

23. M. Evans, S. Colyer, A. Salo, and D. Cosker, "Automatic High Fidelity Foot Contact Location and Timing for Elite Sprinting," *Machine Vision and Applications* 32 (2021): 1–20, https://doi.org/10.1007/s00138-021-01236-z.

24. L. Needham, M. Evans, D. P. Cosker, and S. L. Colyer, "Development, Evaluation and Application of a Novel Markerless Motion Analysis System to Understand Push-Start Technique in Elite Skeleton Athletes," *PLoS One* 16 (2021): e0259624, https://doi.org/10.1371/journal.pone.0259624.

25. R. Harle, J. Cameron, and J. Lasenby, "Foot Contact Detection for Sprint Training," in *Computer Vision—ACCV 2010 Workshops*, eds. R. Koch and F. Huang (Berlin, Heidelberg: Springer, 2010), 297–306.

26. W. Zhu, B. Anderson, S. Zhu, and Y. Wang, "A Computer Vision-Based System for Stride Length Estimation Using a Mobile Phone Camera," ASSETS 2016—Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility (Association for Computing Machinery, Inc, 2016, 121–130), https://doi.org/10.1145/2982142.2982156.

27. F. Fallahtafti, S. R. Wurdeman, and J. M. Yentes, "Sampling Rate Influences the Regularity Analysis of Temporal Domain Measures of Walking More Than Spatial Domain Measures," *Gait and Posture* 88 (2021): 216–220, https://doi.org/10.1016/j.gaitpost.2021.05.031.

28. D. Winter, *The Biomechanics and Motor Control of Human Gait: Normal, Elderly, and Pathological*, 2nd ed. (Waterloo, ON: University of Waterloo Press, 1991).

29. Z. Zhang, "A Flexible New Technique for Camera Calibration," IEEE Trans. Pattern Anal. Mach. Intell. (2000), https://api.semanticscholar.org/CorpusID:1150626.

30. B. Triggs, P. F. Mclauchlan, R. I. Hartley, and A. F. Fitzgibbon, "Bundle Adjustment—A Modern Synthesis," in *Vision Algorithms: Theory and Practice: International Workshop on Vision Algorithms*, eds. B. Triggs, A. Zisserman, and R. Szeliski (Berlin, Heidelberg: Springer, 2000), 298–372.

31. A. L. Bell, R. A. Brand, and D. R. Pedersen, "Prediction of Hip Joint Centre Location From External Landmarks," *Human Movement Science* 8, no. 1 (1989): 3–16.

32. M. Andriluka, L. Pishchulin, P. Gehler, and B. Schiele, "2D Human Pose Estimation: New Benchmark and State of the Art Analysis," IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2014).

33. T. Y. Lin, M. Maire, S. Belongie, et al., "Microsoft COCO: Common Objects in Context," in *Computer Vision—ECCV 2014. Lecture Notes in Computer Science*, vol. 8693, eds. D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars (Cham: Springer, 2014), 740–755, https://doi.org/10.1007/978-3-319-10602-1_48.

34. J. M. Bland and D. G. Altman, "Agreement Between Methods of Measurement With Multiple Observations per Individual," *Journal of Biopharmaceutical Statistics* 17, no. 4 (2007): 571–582, https://doi.org/10.1080/10543400701329422.

35. V. Hojka, R. Bačáková, and P. Kubový, "Differences in Kinematics of the Support Limb Depends on Specific Movement Tasks of Take-Off," *Acta Gymnica* 46 (2016): 82–89, https://doi.org/10.5507/ag.2016.005.

36. K. Sato, Y. Nagashima, T. Mano, A. Iwata, and T. Toda, "Quantifying Normal and Parkinsonian Gait Features From Home Movies: Practical Application of a Deep Learning–Based 2D Pose Estimator," *PLoS One* 14 (2019): e0223549, https://doi.org/10.1371/journal.pone.0223549.

37. A. Jamsrandorj, M. D. Nguyen, M. Park, K. S. Kumar, K. R. Mun, and J. Kim, "Vision-Based Gait Events Detection Using Deep Convolutional Neural Networks," Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS, 2021, 1936–1941), https://doi.org/10.1109/EMBC46164.2021.9630431.

38. A. Jamsrandorj, D. Jung, K. S. Kumar, et al., "View-Independent Gait Events Detection Using CNN-Transformer Hybrid Network," *Journal of Biomedical Informatics* 147 (2023): 104524, https://doi.org/10.1016/j.jbi.2023.104524.

39. I. Akhter, A. Jalal, and K. Kim, "Adaptive Pose Estimation for Gait Event Detection Using Context-Aware Model and Hierarchical Optimization," *Journal of Electrical Engineering and Technology* 16 (2021): 2721–2729, https://doi.org/10.1007/s42835-021-00756-y.

40. R. J. Cotton, E. Mcclerklin, A. Cimorelli, A. Patel, and T. Karakostas, "Transforming Gait: Video-Based Spatiotemporal Gait Analysis," Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS, 2022, 115–120), https://doi.org/10.1109/EMBC48229.2022.9871036.

41. A. Rivadulla, X. Chen, G. Weir, et al., "Development and Validation of Footnet; A New Kinematic Algorithm to Improve Footstrike and Toe-Off Detection in Treadmill Running," *PLoS One* 16 (2021): e0248608, https://doi.org/10.1371/journal.pone.0248608.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.