Validation of Single-Camera Pose Estimation for Orthopaedic Motion Analysis

Aniket Pratapneni[1], Ryan Halvorson[1], Pavlos Silvestros[1], Nicholas Harris[1], Jeannie F Bailey[1]
[1]Department of Orthopaedic Surgery, University of California San Francisco, San Francisco, CA
aniket.pratapneni@ucsf.edu

INTRODUCTION: Motion analysis plays a central role in diagnosing and managing musculoskeletal disorders, but its clinical adoption is limited by the technical and financial burden of traditional systems. Multi-camera, marker-based systems provide high spatial fidelity but are expensive, immobile, and require controlled environments. Recently, single-camera, deep learning-based pose estimation models have emerged, offering markerless 3D joint tracking from standard smartphone video, potentially facilitating scalable, accessible assessments. However, their accuracy in orthopedic contexts – particularly across dynamic, multi-joint tasks – remains underexplored. This study evaluates the accuracy of a single-camera pose estimation model (MeTRAbs) across three clinically relevant movement patterns (treadmill gait, sit-to-stand (STS), and trunk flexion-extension) by comparing its performance to a validated multi-camera system (THEIA3D). Single-camera pose estimation is known to be impacted by random errors from joint occlusion, rapid extremity motion, and depth ambiguity inherent to single-camera capture, as well as by systematic biases from model-specific marker geometry and training datasets that lack the types of movement we are assessing. We therefore hypothesized that the single-camera model would overall perform comparably to the multi-camera system, however estimation error would be lowest at proximal joints and during movements in the frontal plane, and highest at distal, dynamic joints or those with motion predominantly along the anteroposterior axis. Validating and understanding the error patterns of single-camera systems in orthopedic tasks could enable low-cost, portable motion analysis for routine clinical assessments, expanding access to quantitative mobility evaluation.

METHODS: 51 participants (24 M, 27 F, average age 51.5 ± 16.7) performed treadmill gait (n=432), STS (n=96), and trunk flexion-extension (n=141) trials while motion was recorded simultaneously with an iPhone (30 Hz) and an 8-camera THEIA3D system (180 Hz). MeTRAbs was applied to the smartphone videos, and rotation, translation, and least-squares optimal similarity scaling were applied to the single-camera data in order to align them with the multi-camera data's coordinate space. Trials were time-windowed based on key phases of each movement, and to address potential desynchronization, dynamic time warping was applied: trials with >0.5s average temporal deviation were excluded (n=16). Sagittal joint angles for the hips, knees, and trunk were computed geometrically, and root-mean-square error (RMSE) was calculated for each joint position and angle. These RMSEs were then stratified per movement task for more granular analysis. For range of motion (ROM) analysis, absolute errors were computed per joint angle per trial, and intraclass correlation coefficients (ICC3,1) were calculated. Bland-Altman plots were used to assess systematic error. We additionally fit harmonic corrections to the worst-performing joint angles to further illustrate that the error was systematic, not random.

RESULTS: Across all tasks and joints, average trajectory RMSE was 5.95 cm, with the lowest error at the hips (5.5 cm) and highest at the ankles (7.5 cm). Joint angle RMSEs ranged from 2.10˚ to 10.98˚, with dynamic joints such as the knees during gait showing the largest discrepancies. Axial plane error contributed to the largest share of overall error (44%), followed by sagittal plane error (35%). Task-specific patterns (Fig. 1) followed intuitive biomechanical trends and were in accordance with our hypothesis: joints undergoing greater displacement or moving predominantly in the anteroposterior axis displayed greater errors. For example, gait trials had elevated ankle (8.85 cm) and knee (6.35 cm) RMSEs, while trunk flexion-extension exhibited higher trunk (6.49 cm) and hip (6.78 cm) errors. Notably, however, when errors were normalized to joint range of motion, dynamic joints like ankles in gait and shoulders in STS exhibited low relative error (12% and 10%, respectively).
For joint ROMs, while there were non-negligible mean ROM errors for each joint angle, the ICCs exceeded 0.93 for all angles, indicating high consistence in relative movement magnitude. Bland-Altman analysis revealed systematic offsets in calculated ROM that remained consistent independent of ROM magnitude, and these offsets closely matched mean ROM errors, suggesting the errors were largely attributable to systematic differences in the two measurement approaches rather than random error, and potentially correctable. Indeed, applying stride-phase-based Fourier correction (Fig. 2) reduced mean joint angle error significantly (e.g., 11.2 cm to 5.3 cm for right knee flexion in gait), confirming that a large portion of observed deviation was systematic.

DISCUSSION: These results suggest that single-camera pose estimation can yield clinically acceptable 3D joint localization and kinematic tracking. Overall accuracy was high, with the strongest agreement in proximal, less dynamic joints such as hips during trunk flexion-extension. Accuracy was lowest at distal joints with high-velocity, anteroposterior motion, such as the ankles during gait. While absolute agreement varies, the systematic nature of the error – as shown in the angle bias plots – supports the feasibility of post hoc correction or calibration. Even without correction, the high ICC coefficients indicate excellent test-retest reliability, suggesting utility for tracking patients over time or comparing between study participants.

SIGNIFICANCE/CLINICAL RELEVANCE: Single-camera pose estimation offers a scalable, low-cost approach to motion capture for clinical motion assessment. While not a full replacement for traditional motion capture, its tracking accuracy and test-retest reliability in measuring joint ROM and movement patterns supports its use in outpatient, telehealth, and rehabilitation settings for mobility screening, progress monitoring, and functional evaluation.

IMAGES AND TABLES:

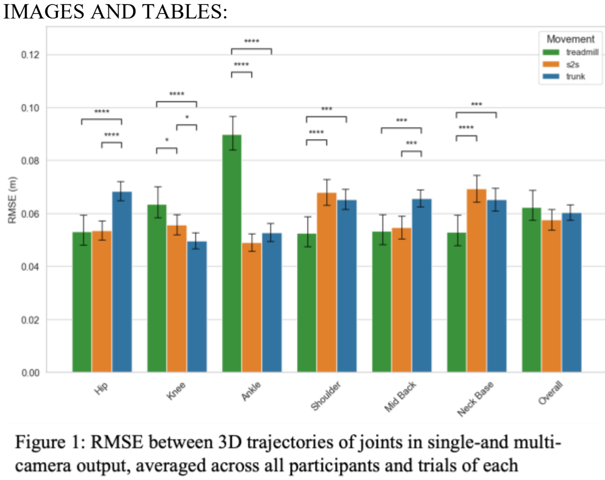

Figure 1: RMSE between 3D trajectories of joints in single-and multi-camera output, averaged across all participants and trials of each movement pattern. Error bars denote 95% confidence intervals.

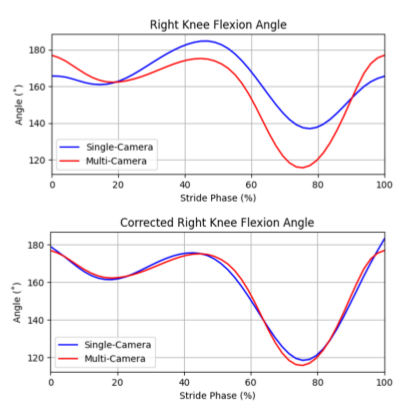

Figure 2: Right knee flexion averaged across all strides of all participants, before and after a systematic correction factor was applied to the single-camera output. 95% confidence bands are present but negligible.