Validity of Monocular Human Mesh Reconstruction for Estimating Lower-Extremity Joint Kinematics: A Comparison of SAM3D and CameraHMR

Ahmadreza Souri Siddhartha Sikdar Tiphanie E. Raffegeau
George Mason University, Fairfax and Manassas, VA, USA
{asouri, ssikdar, traffege}@gmu.edu

Abstract

Marker-based motion capture is the gold standard for quantifying human movement but is limited by cost and laboratory requirements. This study evaluated two monocular human mesh reconstruction models, SAM3D Body and CameraHMR, for estimating lower-extremity joint kinematics relative to a marker-based motion capture reference using the OpenCap validation dataset. Smartphone videos from nine adults performing walking and sit-to-stand tasks were used to reconstruct 3D meshes, generate virtual markers, and estimate joint angles via OpenSim. Mean absolute error (MAE) during walking was similar for both models (5.98° for SAM3D and 6.09° for CameraHMR), while CameraHMR performed better during sit-to-stand (5.92° vs. 7.95°). A linear mixed-effects model showed that CameraHMR reduced error by approximately 2° during sit-to-stand ($p < 0.001$), with minimal differences during walking. Errors were higher in the frontal and transverse planes, while knee flexion showed lower error. Overall, both methods achieved errors near 6°, indicating that single-camera mesh reconstruction combined with biomechanical modeling can approximate laboratory motion capture, although accuracy remains lower than that of multi-view systems.

1. Introduction

Marker-based three-dimensional motion capture has been the gold standard for measuring human movement for decades. However, its high cost, laboratory requirements, and the need for trained personnel for marker placement limit its use outside specialized facilities [14]. Markerless video-based motion tracking provides a scalable alternative and is increasingly applied in sports analytics and clinical movement assessment [1, 13].

Advances in deep learning since the early 2010s, particularly following the introduction of ImageNet and convolutional neural networks, have significantly improved image- and video-based motion analysis [3, 6]. Methods such as OpenPose enabled reliable two-dimensional pose estimation and accelerated the adoption of computer-vision approaches within biomechanics research [2]. Commercial multi-camera markerless systems demonstrate promising reliability and concurrent validity compared with traditional motion capture, but their hardware requirements and cost remain barriers to widespread adoption [12]. Lower-cost solutions such as OpenCap and Pose2Sim enable three-dimensional motion capture using a small number of calibrated smartphone cameras [10, 15]. However, these systems still rely on multi-camera setups that may be difficult to deploy outside controlled environments. Reducing the hardware requirement to a single camera could therefore substantially expand access to motion analysis.

Recent advances in monocular 3D human pose estimation enable reconstruction of full-body meshes from single images using parametric body models such as the Skinned Multi-Person Linear Model (SMPL) [8]. However, these methods are typically evaluated using computer-vision metrics such as Procrustes-Aligned Mean Per Joint Position Error, which are not directly relevant to biomechanical applications [7]. Biomechanics research instead requires accurate estimation of joint kinematics and physiologically plausible motion.

Recent monocular models reconstruct human body meshes that can potentially be integrated with biomechanical modeling workflows to estimate joint kinematics. While SMPL-based models have previously been evaluated for biomechanical analysis, newer approaches such as Segment Anything Model 3D Body (SAM3D), which predicts meshes using the Momentum Human Rig representation (MHR), have not yet been systematically assessed for biomechanical applications [4, 16]. Therefore, the aim of this study was to compare the validity of two monocular markerless methods, SAM3D Body and CameraHMR, for estimating lower-extremity joint kinematics relative to a marker-based motion capture reference system. We hypothesized that joint kinematics derived from monocular human mesh reconstructions would demonstrate relatively small estimation errors when integrated with a biomechani-

cal modeling workflow.

2. Related Work

Previous benchmarking research has compared multiple markerless motion tracking approaches for estimating lower-extremity kinematics relative to marker-based motion capture [7]. These studies evaluated both single-view and multi-view pipelines and consistently reported higher kinematic accuracy for multi-view systems. OpenCap improved accuracy by approximately 1.7° root mean square deviation (RMSD) relative to the best-performing single-view model WHAM, while Theia3D further reduced errors by about 1.3° RMSD compared to OpenCap. Prior research also indicated that applying biomechanical modeling after pose estimation does not consistently improve kinematic accuracy [7].

Additional prior research evaluated a monocular gait analysis workflow based on CameraHMR [5]. In this pipeline, mesh vertices corresponding to anatomical landmarks were used as virtual markers and processed using OpenSim inverse kinematics. The system achieved an average kinematic error of 5.5 ± 1.1° RMSD relative to marker-based motion capture and demonstrated test–retest reliability of 3.0 ± 1.0° RMSD, indicating performance comparable to multi-camera smartphone-based systems while requiring only a single camera.

3. Method

This study evaluated the validity of monocular markerless motion capture pipelines using the OpenCap validation dataset, which includes synchronized smartphone videos and laboratory motion capture recordings collected in a controlled motion capture environment [15]. The dataset consisted of ten healthy adults (6 female and 4 male; age 27.7 ± 3.8 years; body mass 69.2 ± 11.6 kg; height 1.74 ± 0.12 m). Participants performed walking and sit-to-stand tasks under both natural and modified conditions designed to introduce variations in movement patterns. Walking trials included natural walking and a trunk sway condition in which the trunk was leaned laterally toward the stance leg. Sit-to-stand trials included natural sit-to-stand and a modified condition with increased trunk flexion during rising.

One participant did not consent to the public release of identifiable video recordings. Because SAM3D and CameraHMR reconstruction require video input, this participant was excluded from the markerless analysis, resulting in nine participants included in the final dataset.

Marker-based motion capture served as the reference system. An eight-camera optical motion capture system (Motion Analysis Corp., Santa Rosa, CA, USA) recorded the three-dimensional positions of 31 retroreflective markers placed bilaterally on anatomical landmarks representing a full-body marker set. Marker trajectories were recorded at a sampling frequency of 100 Hz. Twenty additional markers were used to improve segment tracking and increase the robustness of kinematic calculations.

Smartphone video recordings were collected simultaneously using five iPhone 12 Pro devices mounted on tripods approximately 1.5 m above the ground around the capture volume. Videos were recorded at 60 frames per second. Although the dataset includes multiple camera views, the present study evaluated a monocular pipeline. Therefore only the camera positioned approximately 45° to the right side of the participant was used for pose estimation.

Monocular human mesh reconstruction was performed using SAM3D Body, which predicts full-body three-dimensional meshes from monocular video frames. The model is based on the MHR and reconstructs meshes consisting of 18,439 vertices using the default configuration. CameraHMR reconstructs a three-dimensional human body mesh by regressing parameters of the SMPL parametric body model, which represents the human body using a mesh containing 6,890 vertices. Both pose estimation models were executed in a Google Colab environment using an NVIDIA A100 GPU.

To convert reconstructed meshes into biomechanically meaningful measurements, 37 mesh vertices corresponding to anatomical landmark locations were selected as virtual markers following previous markerless motion capture studies [5]. These vertices were manually identified using the Blender software environment by visually matching mesh topology to standard anatomical landmark definitions, and their corresponding vertex IDs were then consistently tracked across frames (vertex IDs for each model are provided in Table S1 in the Supplementary Section). Because the focus of the study was lower-extremity joint kinematics, only markers corresponding to the pelvis and lower limbs were used in subsequent analyses.

The coordinate system of the reconstructed mesh does not directly correspond to the laboratory coordinate system used for motion capture analysis. Therefore a rotation matrix transformation was applied to align reconstructed trajectories with the OpenSim coordinate system used for biomechanical modeling. Temporal synchronization between marker-based and markerless kinematic signals was verified for each trial because smartphone recordings were not hardware-synchronized with the motion capture system. Synchronization was performed using cross-correlation of the right knee flexion angle signal, and signals were resampled to a common time grid before alignment.

Biomechanical analysis was performed in OpenSim using the musculoskeletal model embedded within the OpenCap workflow. Subject-specific scaling of the model was performed separately for the marker-based and markerless pipelines using anatomical markers or virtual markers from

the static calibration trial. Joint kinematics were estimated using the OpenSim inverse kinematics tool, which determines joint angles by minimizing the error between experimental marker trajectories and model marker locations while respecting joint constraints. Kinematic signals were filtered using a fourth-order zero-lag Butterworth filter with cutoff frequencies of 6 Hz for walking and 4 Hz for sit-to-stand.

Mean absolute error (MAE) was first computed between markerless and marker-based joint kinematics for each degree of freedom and activity. A linear mixed-effects model was then fitted using raw MAE (degrees) as the dependent variable, with model, activity, and their interaction as fixed effects and participant as a random intercept, to evaluate overall differences in accuracy between models and tasks. A second linear mixed-effects model used normalized MAE (nMAE), computed by dividing MAE by the participant-specific peak-to-peak range of motion of the motion capture reference signal (

Table 1. Mean absolute error (degrees) between markerless and marker-based inverse kinematics during walking. Values are mean ± SD across participants (n = 9). Bold indicates the lower error between models.

Joint	SAM3D	CameraHMR
Hip flexion R	6.2 ± 3.9	$\mathbf{5.1 \pm 1.4}$
Hip adduction R	$\mathbf{5.8 \pm 1.6}$	7.6 ± 2.1
Hip rotation R	$\mathbf{7.6 \pm 1.1}$	9.5 ± 1.4
Hip flexion L	8.1 ± 4.0	$\mathbf{6.6 \pm 3.3}$
Hip adduction L	$\mathbf{2.9 \pm 0.5}$	3.5 ± 0.8
Hip rotation L	5.5 ± 1.8	$\mathbf{4.4 \pm 0.7}$
Knee flexion R	4.8 ± 1.2	$\mathbf{4.3 \pm 1.5}$
Knee flexion L	$\mathbf{4.4 \pm 1.9}$	$\mathbf{4.4 \pm 2.0}$
Ankle dorsiflexion R	$\mathbf{5.1 \pm 1.3}$	5.7 ± 1.1
Ankle dorsiflexion L	$\mathbf{4.0 \pm 1.6}$	4.9 ± 1.1
Subtalar angle R	9.2 ± 1.7	$\mathbf{7.7 \pm 1.8}$
Subtalar angle L	$\mathbf{8.1 \pm 2.1}$	9.5 ± 3.5
Average	**6.0**	6.1

4. Results

For walking (54 trials per model, 9 participants), the average MAE was 5.98° for SAM3D and 6.09° for CameraHMR (Table 1). Both models showed similar accuracy for knee flexion (SAM3D: 4.84° and 4.35°; CameraHMR: 4.27° and 4.35° for right and left, respectively). CameraHMR demonstrated lower hip flexion errors (right: 5.12°; left: 6.59°) compared to SAM3D (right: 6.18°; left: 8.10°), while SAM3D showed lower errors for hip adduction (right: 5.78° vs 7.59°) and hip rotation (right: 7.61° vs 9.49°). Subtalar angle errors were mixed, with SAM3D performing better on the right side and worse on the left.

For sit-to-stand (18 trials per model, 9 participants), CameraHMR showed lower overall error (5.92°) compared with SAM3D (7.95°) (Table 2). The largest difference was observed for hip flexion, where SAM3D exhibited errors of 19.43° (right) and 20.69° (left), approximately twice those observed for CameraHMR (11.36° and 11.24°). CameraHMR also showed lower knee flexion errors (right: 5.35° vs 7.28°; left: 5.46° vs 7.78°) and lower subtalar angle errors (right: 4.40° vs 6.88°). SAM3D demonstrated lower ankle dorsiflexion error on the right side (3.62° vs 7.13°).

Averaging across both activities and all 12 degrees of freedom, the overall MAE was 6.97° for SAM3D and 6.00° for CameraHMR. Mean ± SD waveform comparisons for walking (Figure 1 in the Supplementary Section) and sit-to-stand (Figure 2 in the Supplementary Section) illustrate the agreement between markerless and marker-based kinematics.

In the first model using MAE (degrees), CameraHMR was associated with a significantly lower error compared with SAM3D ($\beta = -2.04$, $p < 0.001$). Walking trials exhibited significantly lower MAE than sit-to-stand trials ($\beta = -1.98$, $p < 0.001$). A significant interaction between model and activity was observed ($\beta = 2.14$, $p < 0.001$), indicating that the performance difference between models depended on the movement type. Specifically, for sit-to-stand, CameraHMR reduced MAE by approximately $2.04°$ relative to SAM3D, whereas for walking the difference was minimal and slightly favored SAM3D (approximately $0.11°$).

In the second model using normalized MAE (nMAE), the main effects of model ($p = 0.222$) and activity ($p = 0.205$) were not statistically significant, indicating comparable proportional error between methods and tasks after accounting for differences in movement amplitude. In contrast, joint (degree of freedom) remained a significant factor, demonstrating that estimation accuracy varied across kinematic variables even after normalization. Frontal- and transverse-plane motions, including hip adduction and hip rotation, exhibited significantly higher normalized error (e.g., hip rotation right: $\beta = 49.50$, $p < 0.001$; hip adduction right: $\beta = 30.87$, $p < 0.001$), whereas knee flexion showed significantly lower normalized error (e.g., left: $\beta = -8.45$, $p = 0.001$), indicating comparatively higher proportional accuracy.

5. Discussion

This study evaluated the validity of two monocular markerless motion capture approaches, SAM3D Body and CameraHMR, for estimating lower-extremity joint kinematics relative to a marker-based reference system. Both models achieved similar accuracy during walking, with average joint angle errors close to 6°. However, statistical analysis indicated that CameraHMR produced significantly lower

Table 2. Mean absolute error (degrees) between markerless and marker-based inverse kinematics during sit-to-stand. Values are mean ± SD across participants (n = 9). Bold indicates the lower error between models.

Joint	SAM3D	CameraHMR
Hip flexion R	19.4 ± 5.9	**11.4 ± 4.2**
Hip adduction R	4.7 ± 2.2	**4.0 ± 2.3**
Hip rotation R	6.2 ± 1.9	**5.3 ± 1.8**
Hip flexion L	20.7 ± 5.2	**11.2 ± 4.0**
Hip adduction L	3.8 ± 2.4	**3.7 ± 2.2**
Hip rotation L	3.7 ± 1.0	**3.3 ± 1.8**
Knee flexion R	7.3 ± 2.5	**5.4 ± 2.3**
Knee flexion L	7.8 ± 3.4	**5.5 ± 2.4**
Ankle dorsiflexion R	**3.6 ± 1.9**	7.1 ± 2.6
Ankle dorsiflexion L	4.4 ± 1.8	**3.2 ± 1.0**
Subtalar angle R	6.9 ± 1.7	**4.4 ± 1.7**
Subtalar angle L	7.0 ± 3.0	**6.7 ± 2.0**
Average	8.0	**5.9**

overall MAE than SAM3D, primarily due to improved performance during sit-to-stand trials.

The similarity in error magnitudes suggests that the underlying parametric body representation, MHR versus SMPL, did not strongly influence the final kinematic estimates. Once reconstructed meshes are converted into virtual marker trajectories and processed using inverse kinematics in OpenSim, the biomechanical model likely constrains motion estimates similarly across pipelines.

The magnitude of the observed errors is consistent with previous reports of monocular markerless motion capture [5]. Prior research using CameraHMR-based gait analysis reported average errors around 5.5° RMSD relative to marker-based motion capture. These comparable values suggest that monocular mesh reconstruction combined with musculoskeletal modeling can provide biomechanically meaningful joint kinematics, although accuracy remains lower than multi-camera approaches.

Benchmarking studies have consistently shown that multi-view systems outperform monocular pipelines [7]. OpenCap improved accuracy by approximately 1.7° RMSD relative to the best-performing single-view model WHAM, while Theia3D further reduced errors by about 1.3° RMSD. Multi-camera systems benefit from geometric constraints that enable triangulation of joint positions and reduce depth ambiguity and occlusion-related errors.

Joint-specific trends observed in this study were also consistent with previous markerless motion capture research [11]. Lower errors were generally observed for sagittal plane movements such as knee flexion–extension, whereas larger errors occurred for transverse and frontal plane rotations such as hip rotation or subtalar motion. These movements are more difficult to estimate from monocular video because depth information must be inferred from learned priors rather than directly reconstructed.

From a practical perspective, monocular markerless motion capture has the potential to significantly expand access to movement analysis. Single-camera systems could enable remote or home-based assessments using standard smartphone videos, reducing hardware requirements and improving scalability. However, several limitations should be considered. The dataset used in this study was collected in a controlled laboratory environment, included only healthy participants, and relied on a single camera viewpoint. Future research should evaluate monocular mesh reconstruction methods in more diverse environments, movement tasks, and clinical populations.

6. Conclusion

This study evaluated the validity of two monocular markerless motion capture pipelines, SAM3D Body and CameraHMR, for estimating lower-extremity joint kinematics relative to a marker-based motion capture reference system. Both methods produced comparable results, with mean joint angle errors of approximately 6° across the evaluated tasks. Prior work suggests errors $> 5°$ may limit clinical interpretability [9]. The 6° error observed here indicates that further improvements are needed for high-precision clinical use. Nevertheless, these findings suggest that monocular human mesh reconstruction approaches can provide biomechanically meaningful joint kinematic estimates when combined with musculoskeletal modeling workflows. However, the observed error magnitudes remain larger than those typically reported for multi-view markerless systems. Continued improvements in monocular pose estimation, depth inference, and integration with biomechanical models will be necessary before single-camera motion capture can reliably support clinical decision-making and large-scale real-world movement analysis.

References

[1] M. A. Boswell, Ł. Kidziński, J. L. Hicks, S. D. Uhlrich, A. Falisse, and S. L. Delp. Smartphone videos of the sit-to-stand test predict osteoarthritis and health outcomes in a nationwide study. *npj Digital Medicine*, 6(1):32, 2023. 1

[2] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh. OpenPose: Realtime multi-person 2D pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(1):172–186, 2021. 1

[3] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009. 1

[4] A. Ferguson, A. A. Osman, and B. ... Bescos. Mhr: Momentum human rig. *arXiv preprint arXiv:2511.15586*, 2025. 1

[5] B. Horsak, M. Simonlehner, V. Quehenberger, B. Dumphart, P. Wegscheider, A. Kranzl, and D. Slijepcevic. Validity and reliability of monocular 3d markerless gait analysis in simulated pathological gait: A comparative study with opencap. *Journal of Biomechanics*, page 112986, 2025. 2, 4

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2012. 1

[7] Z. Li, S. Shin, V. Phan, E. Meinders, and E. Halilaj. Impact of multi-view fusion and biomechanical modeling on markerless motion tracking. *IEEE Transactions on Biomedical Engineering*, 2025. 1, 2, 4

[8] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black. SMPL: A skinned multi-person linear model. *ACM Transactions on Graphics*, 34(6), 2015. 1

[9] J. L. McGinley, R. Baker, R. Wolfe, and M. E. Morris. The reliability of three-dimensional kinematic gait measurements: a systematic review. *Gait & posture*, 29(3):360–369, 2009. 4

[10] D. Pagnon, M. Domalain, and L. Reveret. Pose2Sim: An open-source Python package for multiview markerless kinematics. *Journal of Open Source Software*, 7(77):4362, 2022. 1

[11] S. Riazati, T. E. McGuirk, E. S. Perry, W. B. Sihanath, and C. Patten. Absolute reliability of gait parameters acquired with markerless motion capture in living domains. *Frontiers in Human Neuroscience*, 16:867474, 2022. 4

[12] S. Scataglini, E. Abts, C. V. Bocxlaer, M. V. den Bussche, S. Meletani, and S. Truijen. Accuracy, validity, and reliability of markerless camera-based 3D motion capture systems versus marker-based 3D motion capture systems in gait analysis: A systematic review and meta-analysis. *Sensors*, 24(11), 2024. 1

[13] K. R. Sheerin, D. Reid, D. Taylor, and T. F. Besier. The effectiveness of real-time haptic feedback gait retraining for reducing resultant tibial acceleration with runners. *Physical Therapy in Sport*, 43:173–180, 2020. 1

[14] J. Stebbins, M. Harrington, and C. Stewart. Clinical gait analysis 1973–2023: Evaluating progress to guide the future. *Journal of biomechanics*, 160:111827, 2023. 1

[15] S. D. Uhlrich et al. OpenCap: Human movement dynamics from smartphone videos. *PLOS Computational Biology*, 19 (10):e1011462, 2023. 1, 2

[16] X. Yang et al. SAM 3D Body: Robust full-body human mesh recovery, 2026. 1

7. Supplementary

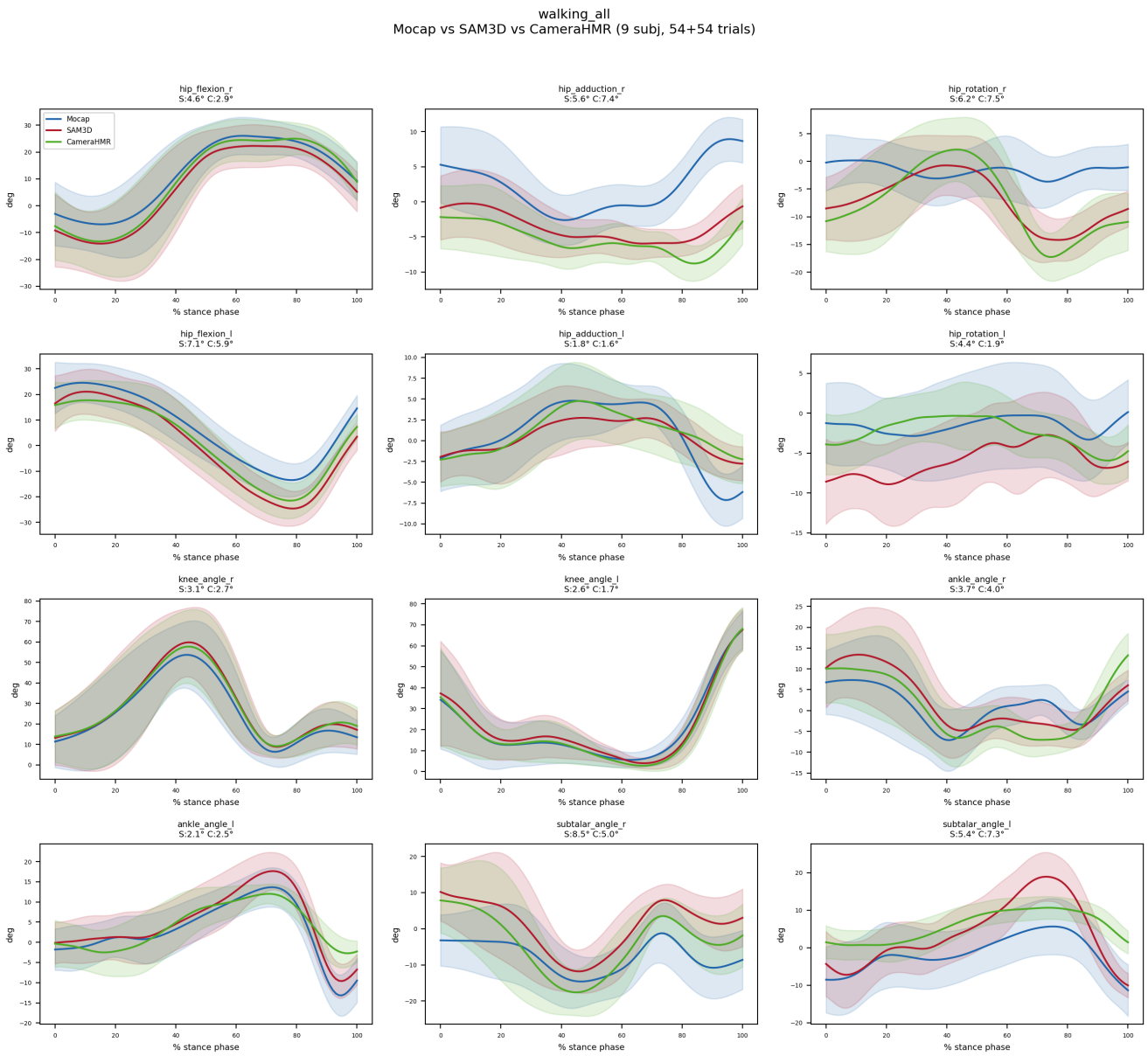

Figure 1. Mean ± SD joint angle waveforms during walking for marker-based motion capture (blue), SAM3D (red), and CameraHMR (green) across 9 participants. The horizontal axis represents the percentage of the stance phase. Per-joint MAE values for SAM3D (S) and CameraHMR (C) are shown in each subplot title.

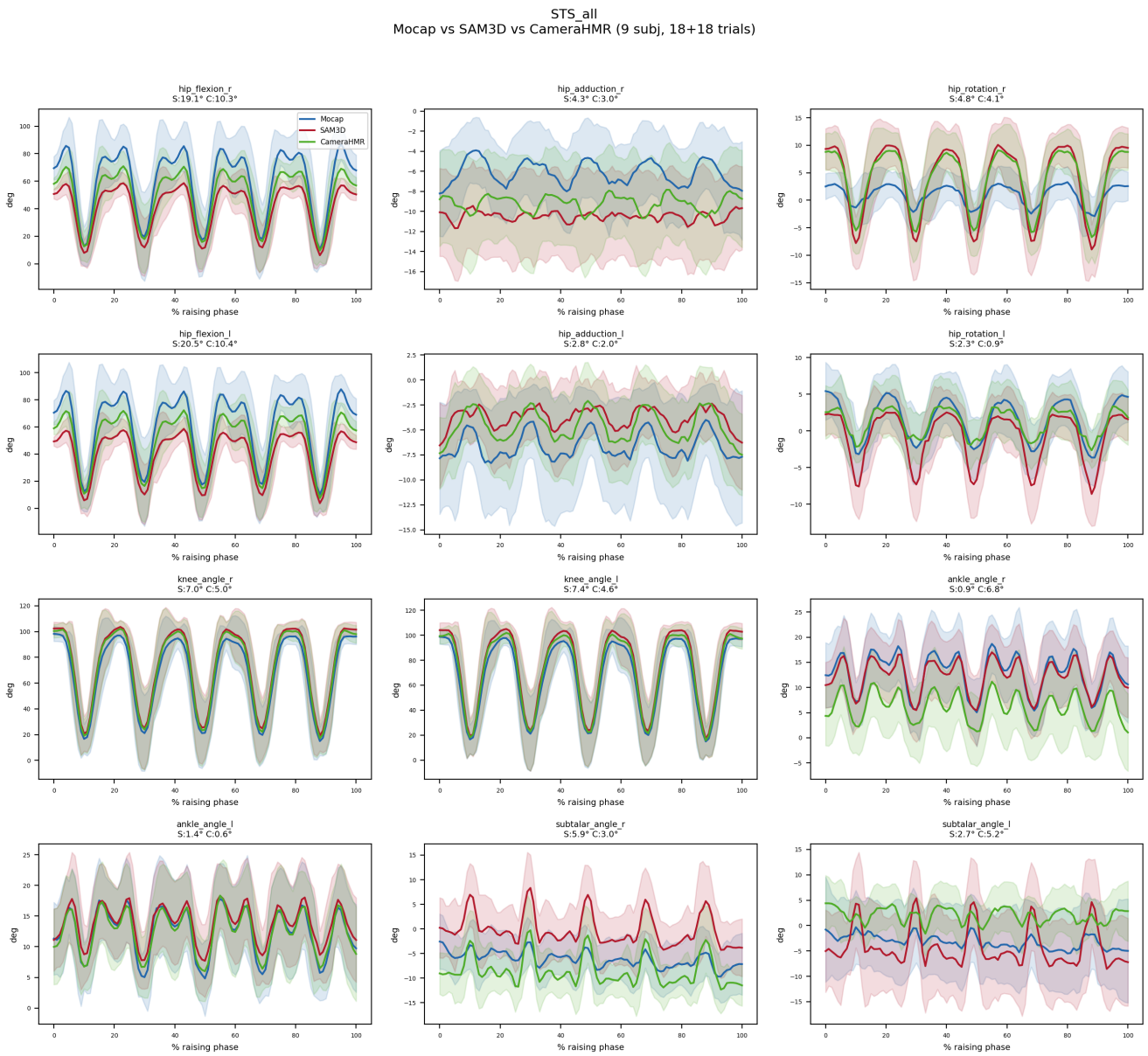

Figure 2. Mean ± SD joint angle waveforms during sit-to-stand for marker-based motion capture (blue), SAM3D (red), and CameraHMR (green) across 9 participants. The horizontal axis represents the percentage of the raising phase. Per-joint MAE values for SAM3D (S) and CameraHMR (C) are shown in each subplot title.

Table S1. Vertex IDs corresponding to anatomical landmark definitions for each mesh reconstruction model.

Marker	Full Name	SMPL (CameraHMR)	MHR (SAM3D Body)
RASI	Right Anterior Superior Iliac Spine	6573	7834
LASI	Left Anterior Superior Iliac Spine	3156	6664
SACR	Sacrum (S1)	3159	5740
RKNE	Right Knee (Lateral Epicondyle)	4495	16744
RKNM	Right Knee (Medial Epicondyle)	4634	16790
RTT	Right Tibial Tuberosity	4664	16923
RANK	Right Ankle (Lateral Malleolus)	6727	17419
RANM	Right Ankle (Medial Malleolus)	6833	17434
RHEE	Right Heel	6786	17486
RTOE	Right Toe (2nd Metatarsal Head)	6741	18082
RD1M	Right 1st Metatarsal Head	6750	18034
RD5M	Right 5th Metatarsal Head	6715	17924
LKNE	Left Knee (Lateral Epicondyle)	1010	11545
LKNM	Left Knee (Medial Epicondyle)	1148	11594
LTT	Left Tibial Tuberosity	1178	11750
LANK	Left Ankle (Lateral Malleolus)	3327	12223
LANM	Left Ankle (Medial Malleolus)	3433	12236
LHEE	Left Heel	3387	12338
LTOE	Left Toe (2nd Metatarsal Head)	3340	12919
LD1M	Left 1st Metatarsal Head	3350	12881
LD5M	Left 5th Metatarsal Head	3348	12737
CLAV	Clavicle	3078	2855
STRN	Sternum	3076	5704
T10	10th Thoracic Vertebra	3015	5753
C7	7th Cervical Vertebra	828	5770
RSHO	Right Shoulder (Acromion)	4724	7950
RELB	Right Elbow (Lateral Epicondyle)	5129	13495
RUPA	Right Upper Arm	6282	13327
RWRA	Right Wrist (Radial Styloid)	5573	13967
RWRB	Right Wrist (Ulnar Styloid)	5608	13958
RFIN	Right Finger	5595	14401
LSHO	Left Shoulder (Acromion)	1239	6737
LELB	Left Elbow (Lateral Epicondyle)	1658	8345
LUPA	Left Upper Arm	1505	8148
LWRA	Left Wrist (Radial Styloid)	2112	8817
LWRB	Left Wrist (Ulnar Styloid)	2108	8826
LFIN	Left Finger	2135	9222