Selecting the Motion Ground Truth for Loose-fitting Wearables: Benchmarking Optical MoCap Methods

Lala Shakti Swarup Ray
DFKI, Germany
lala_shakti_swarup.ray@dfki.de

Bo Zhou
DFKI & RPTU Kaiserslautern, Germany
bo.zhou@dfki.de

Sungho Suh
DFKI & RPTU Kaiserslautern, Germany
sungho.suh@dfki.de

Paul Lukowicz
DFKI & RPTU Kaiserslautern, Germany
paul.lukowicz@dfki.de

ABSTRACT

To aid smart wearable researchers in selecting optimal ground truth methods for motion capture (MoCap) across all loose garment types, we introduce a benchmark: DrapeMoCapBench (DMCB). This benchmark is tailored to assess optical marker-based and marker-less MoCap performance. While high-cost marker-based systems are recognized as precise standards, they demand skin-tight markers on bony areas for accuracy, which is problematic with loose garments. Conversely, marker-less MoCap methods driven by computer vision models have evolved, requiring only smartphone cameras and being cost-effective. DMCB employs real-world MoCap datasets, conducting 3D physics simulations with diverse variables: six drape levels, three motion intensities, and six body type-gender combinations. This benchmarks advanced marker-based and marker-less MoCap techniques, identifying the superior approach for distinct scenarios. When evaluating casual loose garments, both methods exhibit notable performance degradation (>10cm). However, for everyday activities involving basic and swift motions, marker-less MoCap slightly surpasses marker-based alternatives. This renders it an advantageous and economical choice for wearable studies.

CCS CONCEPTS

• **Computing methodologies → Model verification and validation**.

KEYWORDS

mocap benchmark, cloth simulation, quantitative characterization

ACM Reference Format:
Lala Shakti Swarup Ray, Bo Zhou, Sungho Suh, and Paul Lukowicz. 2023. Selecting the Motion Ground Truth for Loose-fitting Wearables: Benchmarking Optical MoCap Methods. In *Proceedings of the 2023 International Symposium on Wearable Computers (ISWC '23), October 08–12, 2023, Cancun, Quintana Roo, Mexico.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3594738.3611359

ISWC '23, October 08–12, 2023, Cancun, Quintana Roo, Mexico

ACM ISBN 979-8-4007-0199-3/23/10…$15.00
https://doi.org/10.1145/3594738.3611359

1 INTRODUCTION

Wearable sensing systems have gained a growing interest towards motion tracking, including IMU sensors [13, 16, 49], RFIDs [17], capacitive fabric sensors [50], computational fabrics [21], and multi-modalities [22]. With continuous motion tracking, activity recognition in various scenarios is trivial downstream tasks [7, 15] along with the shared representation with other domains such as computer vision and large language models [30, 37].

However, the widely accepted golden standard for motion capture (MoCap) systems, such as Qualisys (Sweden), Vicon (USA), and OptiTrack (USA), relies on optical markers placed on the body [16, 48]. These systems employ skin-tight marker placement on bony areas and rely on rigid biomechanical models to convert the surface points to inner joints [14, 32]. Optical marker-based MoCap utilizes either active markers [5, 38] with built-in light sources or passive markers [20] with unique visual patterns or retro-reflective properties. These systems capture the surface marker positions using synchronized camera triangulation and infer the joint motion inside the human body using biomechanical models [32]. Optical MoCap systems are generally favored over inertial methods (that lack absolute positioning) due to their simplicity, accuracy, and robustness against external interference [11]. However, when the markers are placed over a loose piece of garment, they are not able to follow the underlying body motion, which results in significant kinematic errors [28]. The development of loose-fitting is crucial in wearable applications [8, 27, 50] to improve user acceptance, comfort, accommodation of various body shapes, and mass adoption. Nevertheless, relying on marker-based MoCap to provide motion ground truth constraints further the development of loose garments. Video-based marker-less MoCap deep learning algorithms map semantic information (e.g., body parts) to pose without explicit markers using deep learning [9, 12, 41] have matured with the rapid advancement of artificial intelligence. But there is a lack of comprehensive comparisons between marker-based and marker-less MoCap systems, especially considering loose garments.

Several studies comparing marker-based and marker-less MoCap in applications such as controlling an endoscopic instrument [40], baseball pitching biomechanics [10], gait analysis [19], and clinical usability [4] have found that while marker-based MoCap generally exhibits slightly higher accuracy, marker-less systems have the potential to serve as a viable alternative, especially in clinical settings where patient comfort and ease of use are crucial factors [31].

These studies prioritize complexity, ease of use, and overall performance rather than quantitative precision comparison due to the

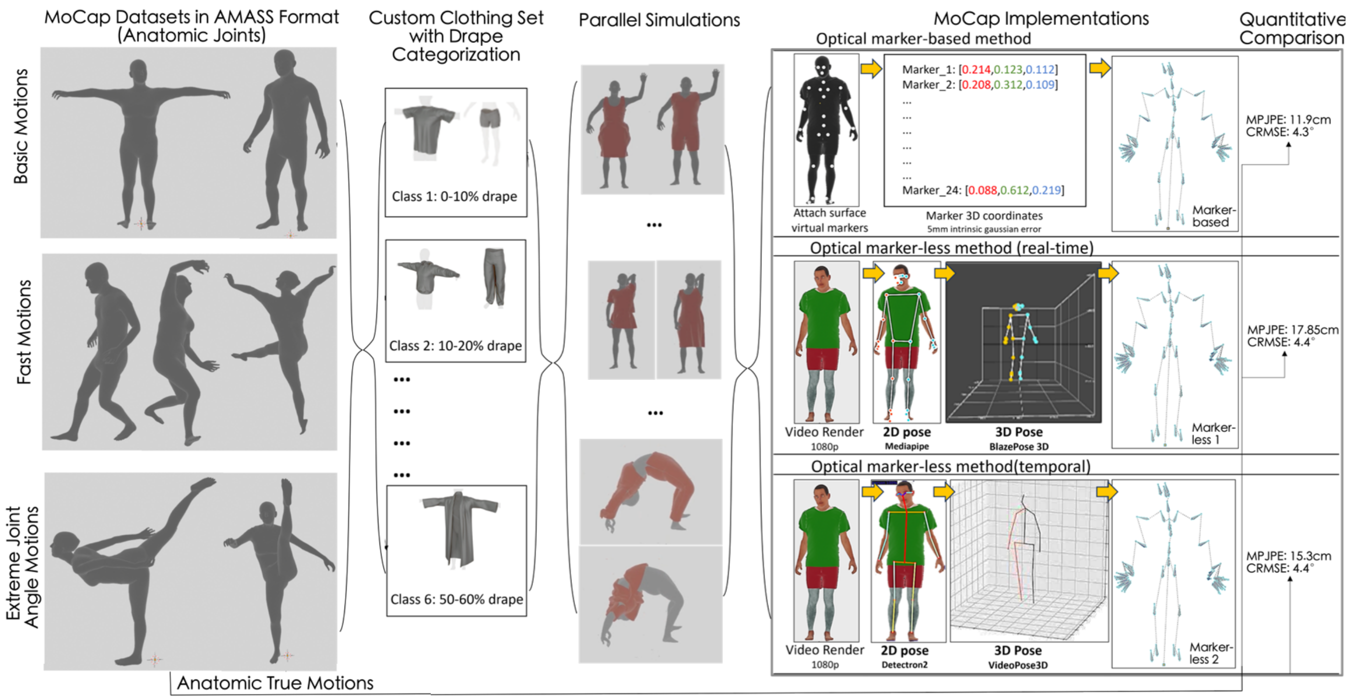

Figure 1: Overall pipeline of DMCB depicting pose estimation and calculation of MPJPE and CRMSE for all MoCap methods for different garment classes using a single motion data sequence.

absence of anatomic motion reference, and no evaluation is done considering loose garments to the level of casual apparel. In practical terms, it's unfeasible to perform an accurate quantitative comparison of motion capture methods for loose garments due to the inability to non-invasively capture true anatomical motion beneath the clothing and replicate precise motion sequences across diverse body shapes and attires. Largely due to these challenges, existing quantitative reviews of marker-less methods use marker-based MoCap as reference [46], which itself has substantial error from the anatomic joints due to the biomechanical approximation. To address these issues, we leverage 3D physics-based simulation to benchmark the impossible. We use real-world captured motion datasets to generate the inputs required for marker-based and marker-less MoCap methods, thus quantitatively comparing them to the common anatomic true motion. In particular, we make the following contributions:

(1) A garment and soft body physics simulation evaluate marker-based and marker-less MoCap performances while persons with different body types repeat the same motions wearing different garments in terms of drape. Using real-world captured motion datasets to generate the input required for both MoCap methods, we quantitatively compare them to the common anatomical true motion.
(2) Our benchmark involves diverse varieties of motion types and garment drape levels which, together with a holistic comparison, can assist practitioners in choosing the optimal MoCap for wearable experiment ground truth for their specific applications balancing aspects such as garment designs, types of motion, cost and time overhead, and precision.

2 PROPOSED METHOD

The overall framework of our benchmark methodology is shown in Figure 1. By leveraging 3D physics simulation, we solve the reality challenge that the exact motion cannot be perfectly reproduced to establish quantitative comparisons of different scenarios.

2.1 DrapeMoCapBench Pipeline

The simulation pipeline strictly adheres to reality, as the inputs to all MoCap methods are true to their specifications: 3D surface marker locations for marker-based kinematic methods and 1080p image sequences for marker-less vision models.

2.1.1 3D physics simulation. With Blender3D [2], motion sequences from Section 2.1.2 were converted to volumetric human bodies of different builds with the help of SMPL-X blender addon [34], then dressed in garments described in Section 2.1.3. All simulated garments are assigned cloth properties equivalent to that of woven cotton (un-stretchable) with vertex mass of 0.05 kg, stiffness tension and compression of 15, stiffness and damping bending of 0.5, damping tension, compression, and shear 5, and stiffness shear of 10. Doubled layered cloth mesh and improved body-cloth collision provided in Simplycloth [1] along with soft tissue dynamics over captured skeletal motions using Mosh++[23] enabled us to introduce realistic deformation of the garments over volumetric human models performing dynamic activities while having minimal artifacts. Then inputs for optical MoCap were derived from the 3D scenes of parallel simulations of the same underlying motion as described in Section 2.2.

2.1.2 Motion Source Dataset. We used the AMASS framework [25] for converting MoCap data from various sources and formats to a standardized format based on the SMPL [24] body model, a 3D

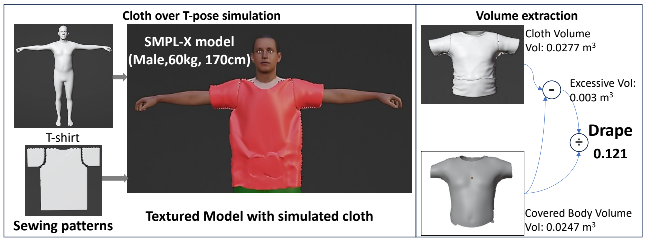

Figure 2: Quantifying drape for t-shirt and trousers.

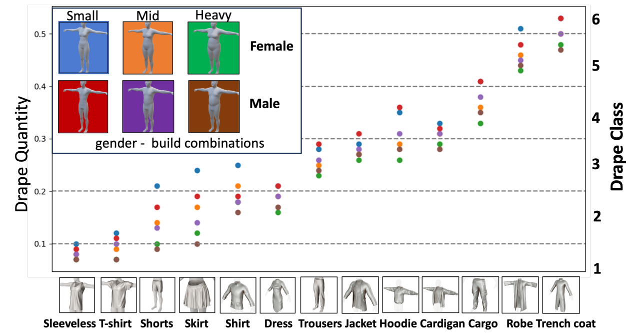

Figure 3: Drape class for different garments, gender & builds.

model that accurately represents the human body. We considered three types of motion. First, the basic motions, such as walking and interaction, are derived from the HumanEva [42] and TotalCapture [43] dataset containing 20 minutes of motion sequences. Second, the fast motions, including sports, dancing, etc., are derived from DanceDB [44] and Totalcapture containing 40 minutes of motion. Finally, the motions with extreme joint bending, like Yoga and gymnastics, are derived from PosePrior [3], and PresSim [39] dataset containing 37 minutes of motion.

2.1.3 Quantifying drape of loose garments. We used 3D assets for a broad selection of apparel from commonly available categories for both genders using the Simplycloth[1] plugin with garments ranging from skin-tight (minimal drape) to very loose (maximal drape). They are assigned into one of six classes based on the drape percentage, as depicted in Figure 3. Drape amount is calculated as the percentage of the extra volume occupied by the garment as compared to another garment that fits the body perfectly and has the volume underlying the covered body as visualized in Figure 2: $Drape = \frac{Volume_{\text{garment}} - Volume_{\text{CoveredBody}}}{Volume_{\text{CoveredBody}}}$. For all garments except uni-cloths, a piece for the upper body and lower body for a particular build sharing combined drape class from 1 to 6 is selected to dress the SMPL body mesh.

2.2 MoCap methods

State-of-the-art marker-based and marker-less MoCap methods were implemented with the benchmark pipeline.

2.2.1 Marker-based. Regardless of the marker principle, they return a 3D coordinate. A marker set of 24 marker pairs (48 markers) associated with the 24 joints from the SMPL skeleton is either attached over the garments or the skin, depending on whether their original position is covered by the garment in T-pose. This represents the optimal real-world marker placement, and 24 marker pairs are sufficient, as the simulation retrieves the coordinates directly without accounting for occlusion and triangulation from multiple cameras. 5 mm error was added as Gaussian noise to the coordinate according to the best-performing solutions [26, 29, 45]. The surface markers are converted to Bio-Vision Hierarchy (BVH) MoCap files having SMPL skeleton hierarchy using forward kinematics to approximate joints' absolute position and angle.

2.2.2 Marker-less. Two marker-less models were considered: a temporal semi-supervised 3d pose estimation model VideoPose3D [35] and a lightweight real-time 3d pose estimation model BlazePose3D [6]. Videos (1920×1080) were rendered from the simulation scene, then fed into Detectron2[47] + VideoPose3D or BlazePose3D to extract multi-joint poses relative to the video frame. They are then rescaled to the original size of the body (170cm height) and converted to BVH files.

2.3 Evaluation Metrics

We consider two frequently used metrics in the MoCap field: absolute Mean Per Joint Position Error (MPJPE), which provides a quantitative measure of the accuracy of 3D joint positions and is more often used in animation, and Circular Root Mean Squared Error (CRMSE), that assesses the performance of pose joint angle estimation used primarily in sports and medical research on joint angles. They are defined as follows: $MPJPE = \frac{1}{n}\sum_{i=1}^{n} ||\mathbf{P}_i - \hat{\mathbf{P}}_i||$ where n is the number of joints, $\mathbf{P}_i$ is the ground truth position of the i-th joint, $\hat{\mathbf{P}}_i$ is the estimated position of the i-th joint, and $||\cdot||$ denotes the Euclidean distance. $CRMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(1-\cos(\theta_i - \hat{\theta}_i)\right)}$ Here, N represents the total number of joint angles, θ_i represents the ground truth angle for the i-th joint, and $\hat{\theta}_i$ represents the corresponding predicted angle. Due to the availability of comprehensive measurements about human models and garments obtained from the simulation, it is straightforward to calculate MPJPE using the 3D estimated joints in Euler space. On the other hand, the CRMSE involves estimating joint angles by applying forward kinematics and then computing the error.

3 RESULTS AND DISCUSSION

With the unclothed body, the MPJPE between marker-based and marker-less implementations of basic motion tested on the TotalCapture dataset is 4.7 cm, and extreme angle motions tested on the PosePrior dataset is 8.2 cm. These results quantitatively align with the literature comparing marker-less models with marker-based mocap as reference [18, 33, 36, 46]. The MPJPE and CRMSE between different MoCap implementations and the anatomic joints are detailed in Figure 4. We calculated MPJPE for both markers placed on cloth only as well as the entire marker set for marker-based MoCap method. The minimum joint-position error is unsurprisingly from drape class 1 with marker-based methods, which is still >10 cm. Such comparison has only been possible before with our simulation pipeline, as the anatomic joint coordinates cannot be derived in reality with non-invasive superficial methods like surface markers or video analysis as explained in Section 1. The term 'looseness' is highly subjective, and its interpretation depends on the relative volume of the garment as compared to the wearer. To account for this variability, we employed a quantitative measurement of drape and organized our findings into drape classes. Everyday loose garments that effectively follow the wearer's body motion typically belong to drape class 2 or 3. In this range, either marker-based or marker-less gives 15cm to 35cm MPJPE and 6° to 11° CRMSE. Absolute MPJPE

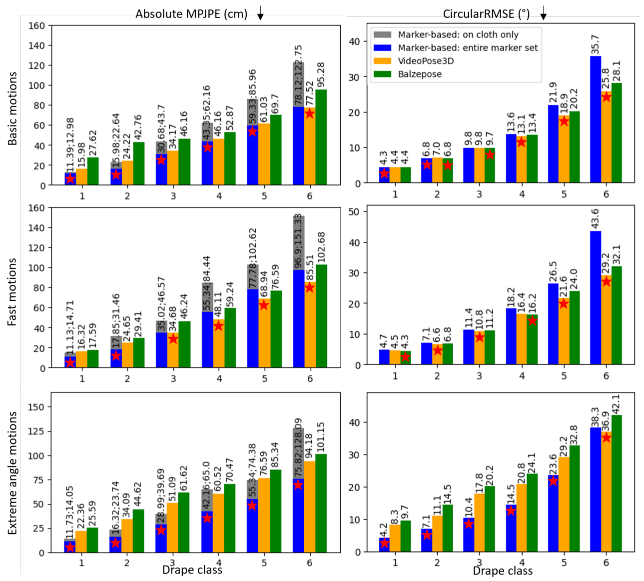

Figure 4: Benchmark results of marker-based & marker-less MoCap methods in different drape & motion combinations.

is susceptible to joint hierarchy alignment, including shifting and rotation errors, while CRMSE is affected by the relative angle of bone in terms of its parent. As expected bio-mechanical constraints the MPJPE overall marker set as compared to where it is calculated for markers placed only over the garment in marker-based MoCap. However, there are extreme cases, such as robes and trench coats, where the garments exhibit limited adherence to the wearer's motion. All methods see gradually increasing errors with more drape of the garment.Marker-less methods (especially VideoPose3D) are more stable as the drape increases, which might be attributed to the fact that they work on detecting semantic segments of body parts. DL-based marker-less models are expected to perform better on basic and fast motions than on extreme joint angles since they are primarily trained on datasets consisting of the first two. On the other hand, the marker-based method does not have such limitations, as it uses forward kinematics with biomechanical constraints from the markers. Apart from the quantitative benchmark results, we also compare both methods holistically as listed in Figure 5.

In our simulated benchmark, we aimed to maintain realism as much as possible; however, there are certain limitations that could be addressed in future research. For the MoCap implementations, the marker-based method excluded occlusions in real-world triangulation and human setup errors as marker positions and marker-less methods did not consider optical confusions which may lead to suboptimal image quality. However, these aspects are also challenges in real-world experiments and are often actively eliminated by practitioners. Our simulation only considered one standard cloth material although draping is significantly affected by cloth material as it determines the weight, stretch, and rigidity which influences how the fabric hangs, folds, and holds a shape when draped over a 3D form. To further enhance realism, future work should incorporate support for external collisions, such as garments interacting with the floor. The variety of clothing options and motion types could also be expanded. It is also crucial to introduce support for other

	Marker-based MoCap (e.g. Qualisys, Vicon, OptiTrack)	Marker-less MoCap (e.g. Detectron2+VidPose3D, MediaPipe+BlazePose3D)
Requirements	Dedicated volume space, multiple (typical 6-12) special cameras and synchronization systems, high throughput computer, active camera or active marker, multiple markers (typical 39-57) of tight placement for full body pose.	Single common digital camera (e.g. smartphone camera), AI-capable computing hardware for model inference, Sufficient subject/background contrast and lighting conditions
Pros	Highly precise for rigid object tracking and MoCap with skin-tight clothing, stable for all types of motions	Better angle accuracy especially with extreme loose cloths, more stable with increasing drape, computation can be outsourced (e.g. to cloud services), can be used in the wild, little setup time
Cons	High cost, setup time of attaching markers and calibrations, setup required before every recording, multiple markers need to be placed all over the body, markers can be uncomfortable, restricted to dedicated space	Worse with certain motions due to DL model limitations, sensitive to lighting conditions
Remarks	Preferred if the marker placement requirements specified by the producer's manual can be met (e.g. skin-tight clothing), for example medical or sports evaluations.	Sufficient in most daily activities with loose casual apparels for wearable technology research. The performance on extreme angle motions may be improved in time with ongoing computer vision research

Figure 5: Holistic comparisons for two MoCap methods

modalities in our benchmarking multi-camera and RGBD markerless MoCap methods, which could provide even better precision with still less cost than marker-based MoCap.

4 CONCLUSIONS AND OUTLOOK

We propose the DrapeMoCapBench as a benchmark methodology based on 3D physics simulations to compare marker-based and marker-less MoCap systems, mainly when dealing with loose garments. The simulation allows us to quantitatively compare different scenarios of precisely reproduced motions against the anatomic true motion under draped garments and body skin. It incorporates physics-based simulation of the human body and garment models with real-world motion datasets, generating input data for marker-based and marker-less MoCap methods. Through the benchmark, we provide a comprehensive comparison of the MoCap methods with a benchmark table that quantifies precision for different motion types and levels of garment drape and holistic considerations. Our findings indicate that, while in line with the literature for skin-tight clothing, both marker-based and marker-less MoCap suffer significant performance loss in casual loose garments like shirts. For daily activities, marker-less MoCap slightly outperforms marker-based MoCap and could be a preferable choice of reference for wearable studies. DMCB can be a valuable resource for wearable practitioners seeking to select the most suitable MoCap method for their own applications, considering scenario-specific precision and holistic factors. The marker-less methods are closer to the marker-based MoCap than the anatomic joints in terms of MPJPE, which could be explained by the fact that the DL models were trained mostly using marker-based MoCap as the ground truth. DMCB can also be used in future work as a data augmentation tool to improve vision-based pose estimation models. Furthermore, Wearable developers can consider the specific errors identified for each motion and garment type rather than relying solely on marker tracking errors, which through our findings, do not accurately represent the pose estimation error often not specified explicitly by the MoCap system and to evaluate the expected MoCap performance for new garment designs for tailored assessment.

ACKNOWLEDGMENTS

The research reported in this paper was supported by the BMBF (German Federal Ministry of Education and Research) in the project VidGenSense (01IW21003).

REFERENCES

[1] 2022. SimplyCloth. Retrieved May 23, 2023, from https://blendermarket.com/products/simply-cloth.

[2] 2023. Blender. Computer software. https://www.blender.org

[3] Ijaz Akhter and Michael J Black. 2015. Pose-conditioned joint angle limits for 3D human pose reconstruction. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1446–1455.

[4] A. Ancans. 2021. Wearable Sensor Clothing for Body Movement Measurement during Physical Activities in Healthcare. *Sensors (Basel, Switzerland)* (2021).

[5] Jan Carlo Barca, Grace Rumantir, and Raymond Koon Li. 2006. A new illuminated contour-based marker system for optical motion capture. In *2006 Innovations in Information Technology*. IEEE, 1–5.

[6] Valentin Bazarevsky, Ivan Grishchenko, Karthik Raveendran, Tyler Zhu, Fan Zhang, and Matthias Grundmann. 2020. Blazepose: On-device real-time body pose tracking. *arXiv preprint arXiv:2006.10204* (2020).

[7] Ardhendu Behera, Zachary Wharton, Alexander Keidel, and Bappaditya Debnath. 2020. Deep cnn, body pose, and body-object interaction features for drivers' activity monitoring. *IEEE Transactions on Intelligent Transportation Systems* 23, 3 (2020), 2874–2881.

[8] Hymalai Bello, Bo Zhou, Sungho Suh, and Paul Lukowicz. 2021. Mocapaci: Posture and gesture detection in loose garments using textile cables as capacitive antennas. In *Proceedings of the 2021 ACM International Symposium on Wearable Computers*. 78–83.

[9] Theocharis Chatzis, Andreas Stergioulas, Dimitrios Konstantinidis, Kosmas Dimitropoulos, and Petros Daras. 2020. A comprehensive study on deep learning-based 3D hand pose estimation methods. *Applied Sciences* 10, 19 (2020), 6850.

[10] Glenn S Fleisig, Jonathan S Slowik, Derek Wassom, Yuki Yanagita, Jasper Bishop, and Alek Diffendaffer. 2022. Comparison of marker-less and marker-based motion capture for baseball pitching kinematics. *Sports Biomechanics* (2022), 1–10.

[11] Martin Kokholm Fleron, Niels Christian Hauerbach Ubbesen, Francesco Battistella, David Leandro Dejtiar, and Anderson Souza Oliveira. 2019. Accuracy between optical and inertial motion capture systems for assessing trunk speed during preferred gait and transition periods. *Sports biomechanics* 18, 4 (2019), 366–377.

[12] Miniar Ben Gamra and Moulay A Akhloufi. 2021. A review of deep learning techniques for 2D and 3D human pose estimation. *Image and Vision Computing* 114 (2021), 104282.

[13] Jian Gong, Xinyu Zhang, Yuanjun Huang, Ju Ren, and Yaoxue Zhang. 2021. Robust inertial motion tracking through deep sensor fusion across smart earbuds and smartphone. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 2 (2021), 1–26.

[14] Brenda E Groen, Marjolein Geurts, Bart Nienhuis, and Jacques Duysens. 2012. Sensitivity of the OLGA and VCM models to erroneous marker placement: Effects on 3D-gait kinematics. *Gait & posture* 35, 3 (2012), 517–521.

[15] Bart Jansen, Frederik Temmermans, and Rudi Deklerck. 2007. 3D human pose recognition for home monitoring of elderly. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 4049–4051.

[16] Yifeng Jiang, Yuting Ye, Deepak Gopinath, Jungdam Won, Alexander W Winkler, and C Karen Liu. 2022. Transformer Inertial Poser: Real-time Human Motion Reconstruction from Sparse IMUs with Simultaneous Terrain Generation. In *SIGGRAPH Asia 2022 Conference Papers*. 1–9.

[17] Haojian Jin, Zhijian Yang, Swarun Kumar, and Jason I Hong. 2018. Towards wearable everyday body-frame tracking using passive RFIDs. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (2018), 1–23.

[18] Angjoo Kanazawa, Michael J Black, David W Jacobs, and Jitendra Malik. 2018. End-to-end recovery of human shape and pose. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 7122–7131.

[19] Robert M Kanko, Elise K Laende, Elysia M Davis, W Scott Selbie, and Kevin J Deluzio. 2021. Concurrent assessment of gait kinematics using marker-based and markerless motion capture. *Journal of biomechanics* 127 (2021), 110665.

[20] Yeonkyung Lee and Hoon Yoo. 2017. Low-cost 3D motion capture system using passive optical markers and monocular vision. *Optik* 130 (2017), 1397–1407.

[21] Ruibo Liu, Qijia Shao, Siqi Wang, Christina Ru, Devin Balkcom, and Xia Zhou. 2019. Reconstructing human joint motion with computational fabrics. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 1 (2019), 1–26.

[22] Shiqiang Liu. 2020. A wearable motion capture device able to detect dynamic motion of human limbs. *Nature Communications* (2020).

[23] Matthew Loper, Naureen Mahmood, and Michael J Black. 2014. MoSh: motion and shape capture from sparse markers. *ACM Trans. Graph.* 33, 6 (2014), 220–1.

[24] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. 2015. SMPL: A Skinned Multi-Person Linear Model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)* 34, 6 (Oct. 2015), 248:1–248:16.

[25] Naureen Mahmood, Nima Ghorbani, Nikolaus F Troje, Gerard Pons-Moll, and Michael J Black. 2019. AMASS: Archive of motion capture as surface shapes. In *Proceedings of the IEEE/CVF international conference on computer vision*. 5442–5451.

[26] Lorin P Maletsky, Junyi Sun, and Nicholas A Morton. 2007. Accuracy of an optical active-marker system to track the relative motion of rigid bodies. *Journal of biomechanics* 40, 3 (2007), 682–685.

[27] Eric McAdams, Asta Krupaviciute, Claudine Géhin, Etienne Grenier, Bertrand Massot, André Dittmar, Paul Rubel, and Jocelyne Fayn. 2011. Wearable sensor systems: The challenges. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 3648–3651.

[28] Ciarán McFadden, Katherine Daniels, and Siobhán Strike. 2020. The sensitivity of joint kinematics and kinetics to marker placement during a change of direction task. *Journal of Biomechanics* 101 (2020), 109635.

[29] Pierre Merriaux, Yohan Dupuis, Rémi Boutteau, Pascal Vasseur, and Xavier Savatier. 2017. A study of vicon system positioning performance. *Sensors* 17, 7 (2017), 1591.

[30] Seungwhan Moon, Andrea Madotto, Zhaojiang Lin, Alireza Dirafzoon, Aparajita Saraf, Amy Bearman, and Babak Damavandi. 2022. IMU2CLIP: Multimodal Contrastive Learning for IMU Motion Sensors from Egocentric Videos and Text. *arXiv preprint arXiv:2210.14395* (2022).

[31] Nobuyasu Nakano, Tetsuro Sakura, Kazuhiro Ueda, Leon Omura, Arata Kimura, Yoichi Iino, Senshi Fukashiro, and Shinsuke Yoshioka. 2020. Evaluation of 3D markerless motion capture accuracy using OpenPose with multiple video cameras. *Frontiers in sports and active living* 2 (2020), 50.

[32] OptiTrack. 2019. *Skeleton Tracking*. https://docs.optitrack.com/motive/skeleton-tracking

[33] Mirela Ostrek, Helge Rhodin, Pascal Fua, Erich Müller, and Jörg Spörri. 2019. Are existing monocular computer vision-based 3D Motion capture approaches ready for deployment? A methodological study on the example of alpine skiing. *Sensors* 19, 19 (2019), 4323.

[34] Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed AA Osman, Dimitrios Tzionas, and Michael J Black. 2019. Expressive body capture: 3d hands, face, and body from a single image. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10975–10985.

[35] Dario Pavllo, Christoph Feichtenhofer, David Grangier, and Michael Auli. 2019. 3d human pose estimation in video with temporal convolutions and semi-supervised training. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 7753–7762.

[36] Haibo Qiu, Chunyu Wang, Jingdong Wang, Naiyan Wang, and Wenjun Zeng. 2019. Cross view fusion for 3d human pose estimation. In *Proceedings of the IEEE/CVF international conference on computer vision*. 4342–4351.

[37] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*. PMLR, 8748–8763.

[38] Ramesh Raskar, Hideaki Nii, Bert Dedecker, Yuki Hashimoto, Jay Summet, Dylan Moore, Yong Zhao, Jonathan Westhues, Paul Dietz, John Barnwell, et al. 2007. Prakash: lighting aware motion capture using photosensing markers and multiplexed illuminators. *ACM Transactions on Graphics (TOG)* 26, 3 (2007), 36–es.

[39] Lala Shakti Swarup Ray, Bo Zhou, Sungho Suh, and Paul Lukowicz. 2023. PresSim: An End-to-end Framework for Dynamic Ground Pressure Profile Generation from Monocular Videos Using Physics-based 3D Simulation. In *2023 IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops)*. IEEE, 484–489.

[40] Rob Reilink, Stefano Stramigioli, and Sarthak Misra. 2013. 3D position estimation of flexible instruments: marker-less and marker-based methods. *International journal of computer assisted radiology and surgery* 8 (2013), 407–417.

[41] Leonid Sigal. 2021. Human pose estimation. In *Computer Vision: A Reference Guide*. Springer, 573–592.

[42] Leonid Sigal, Alexandru O Balan, and Michael J Black. 2010. Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated human motion. *International journal of computer vision* 87, 1-2 (2010), 4.

[43] Matthew Trumble, Andrew Gilbert, Charles Malleson, Adrian Hilton, and John Collomosse. 2017. Total capture: 3d human pose estimation fusing video and inertial sensors. In *Proceedings of 28th British Machine Vision Conference*. 1–13.

[44] University of Cyprus. 2023. Dance Motion Capture Database. http://dancedb.cs.ucy.ac.cy. Motion capture data used in this work were obtained from the Dance Motion Capture Database of the University of Cyprus.

[45] Eline Van der Kruk and Marco M Reijne. 2018. Accuracy of human motion capture systems for sport applications; state-of-the-art review. *European journal of sport science* 18, 6 (2018), 806–819.

[46] Jinbao Wang, Shujie Tan, Xiantong Zhen, Shuo Xu, Feng Zheng, Zhenyu He, and Ling Shao. 2021. Deep 3D human pose estimation: A review. *Computer Vision and Image Understanding* 210 (2021), 103225.

[47] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. 2019. Detectron2. https://github.com/facebookresearch/detectron2.

[48] Chunzhi Yi, Baichun Wei, Zhen Ding, Chifu Yang, Zhiyuan Chen, and Feng Jiang. 2022. A self-aligned method of IMU-based 3-DoF lower-limb joint angle

estimation. *IEEE Transactions on Instrumentation and Measurement* 71 (2022), 1–10.

[49] Xinyu Yi, Yuxiao Zhou, Marc Habermann, Soshi Shimada, Vladislav Golyanik, Christian Theobalt, and Feng Xu. 2022. Physical inertial poser (pip): Physics-aware real-time human motion tracking from sparse inertial sensors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 13167–13178.

[50] Bo Zhou, Daniel Geissler, Marc Faulhaber, Clara Elisabeth Gleiss, Esther Friederike Zahn, Lala Shakti Swarup Ray, David Gamarra, Vitor Fortes Rey, Sungho Suh, Sizhen Bian, et al. 2023. MoCaPose: Motion Capturing with Textile-integrated Capacitive Sensors in Loose-fitting Smart Garments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 1 (2023), 1–40.