arplo

ARPLO's Markerless Motion Capture Engine: Google MediaPipe

Sinan Candan

February 2024

Introduction

ARPLO is a healthcare solution that centers around smart-phone compatible markerless motion capture (MMC) [1] technology which allows a telemedicine pipeline among patients and medical professionals. ARPLO provides joint angles, leading to accurate description of motion geometry and informing relevant medical attributes.

This white paper presents an in-depth literature survey to validate ARPLO's MMC engine, which is built upon the Google MediaPipe software suite. Results of this rigorous literature survey validate ARPLO's MMC engine, Google MediaPipe, by featuring low mean joint positional error. This paper focuses on validation of measuring human joint kinematics only.

We first extend key motivations of MMC technologies, give examples from literature, and provide validation data proving that Google MediaPipe reaches high standards for kinematic human motion capture.

Key Motivations

ARPLO's MMC engine captures limb motion parameters from video recordings. This data can be utilized to evaluate clinically relevant wellness scores such as NIH Stroke – Drift Score 5 [2] or to extrapolate kinetic features of the limbs (torques, speeds, accelerations) to infer joint or muscle strength. Over traditional marker-based systems, the main advantages of smart phone compatible MMCs are ease of use, naturalistic data collection, cost effectiveness and accessibility (Figure 1). Marker based system costs can reach up to $100,000 with days of analysis required (Figure 2) while MMCs only require a personal smartphone and provide data in minutes [3, 19].

Ease of Use

MMC eliminates time consumption of meticulous placement of markers and patient discomfort. It's beneficial in clinical analysis since patient compliance is crucial.

Naturalistic Data Collection

Marker-based solutions confine subjects to artificial laboratory settings. MMC allows naturalistic environments, invaluable for rehabilitation, sports performance, and daily life or habit analysis.

Cost Effectiveness and Accessibility

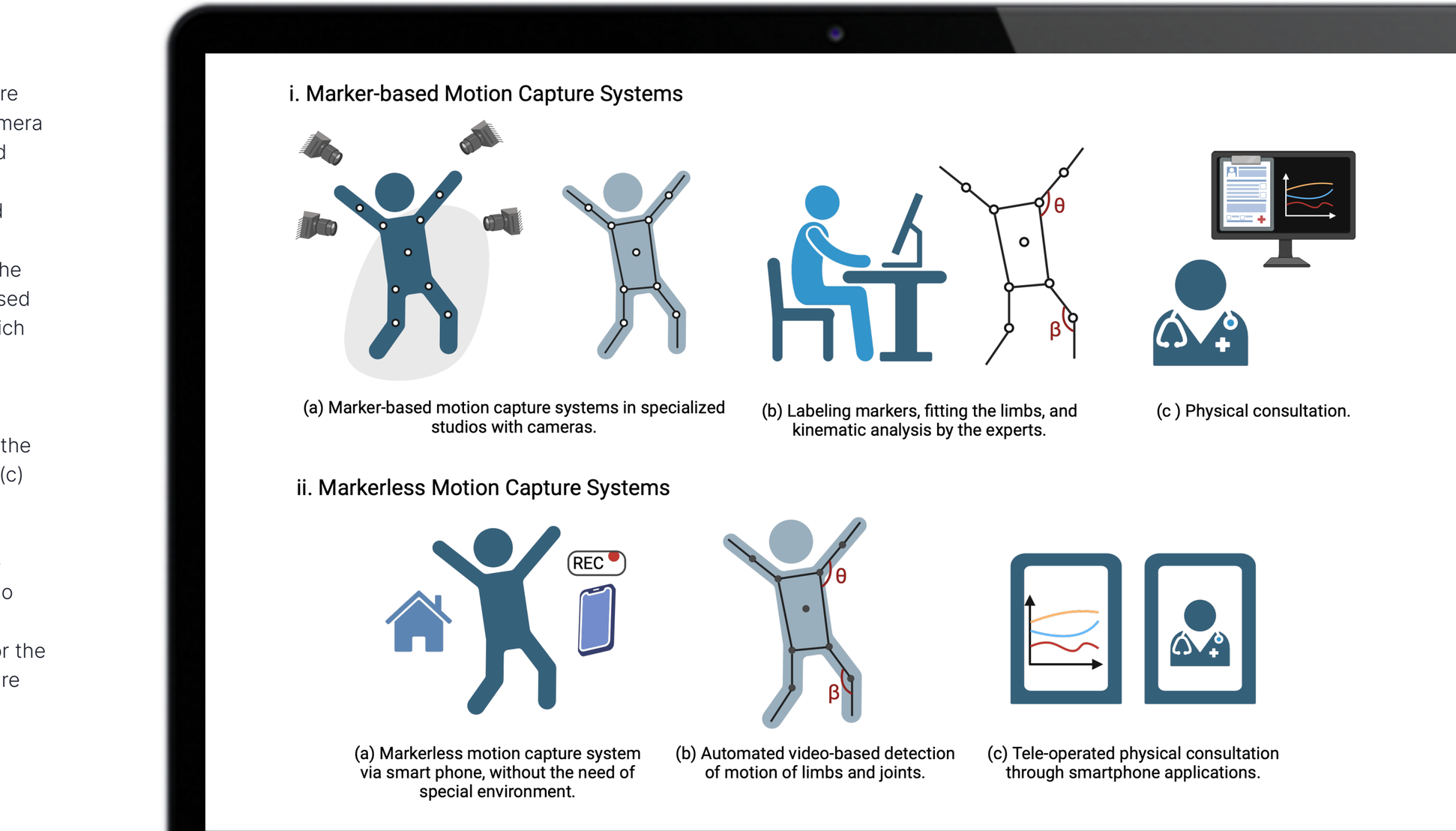
Smart phone based MMC eliminates initial setup and long-term maintenance and operational expenses.

Naturalistic Data Collection, Cost Effectiveness and Accessibility

Figure 1 – Summary on advantages of MMCs, simple and cost-effective limb motion analysis and allow patients to be examined in their natural environments [3]. On the other hand, gold standard marker-based motion capture systems require large numbers of cameras, markers attached on the patient, and operation by experts, which is time consuming and expensive. MMC systems can utilize a video recording obtained with a single camera, without any marker, or a specialized environment.

Figure 2

(i) Steps of marker-based motion capture system which requires (a) high-cost camera system and studio, their calibration, and hours of data collection followed by (b) expert analysis for labeling markers and kinematic analysis to (c) report physiologically relevant information to the physicians. (ii) Steps of smartphone based markerless motion capture systems which (a) begins with a video recording of the relevant motion through a smartphone without any requirement of specialized studio. Then the software performs (b) the kinematic analysis and transferring the (c) physiologically relevant data to the physician. Markerless systems with smartphones significantly decrease the costs associated with the equipment (no external cameras or a studio) and the laborship (experts) and time required for the analysis from days to times [3, 19]. Figure made by BioRender.



What to Validate, and Why?

MMC systems promise accurate description of joint kinematic variables, which are joint angles, joint positions and limb parameters. Furthermore, physicians often require dynamic variables, such as joint/limb velocities, accelerations, and torques. These variables require differentiated joint kinematic variables. However, differentiation amplifies errors in a measurement. Therefore, it is essential to validate the joint kinematics or joint angle measurements of the ARPLO MMC system.

arplo

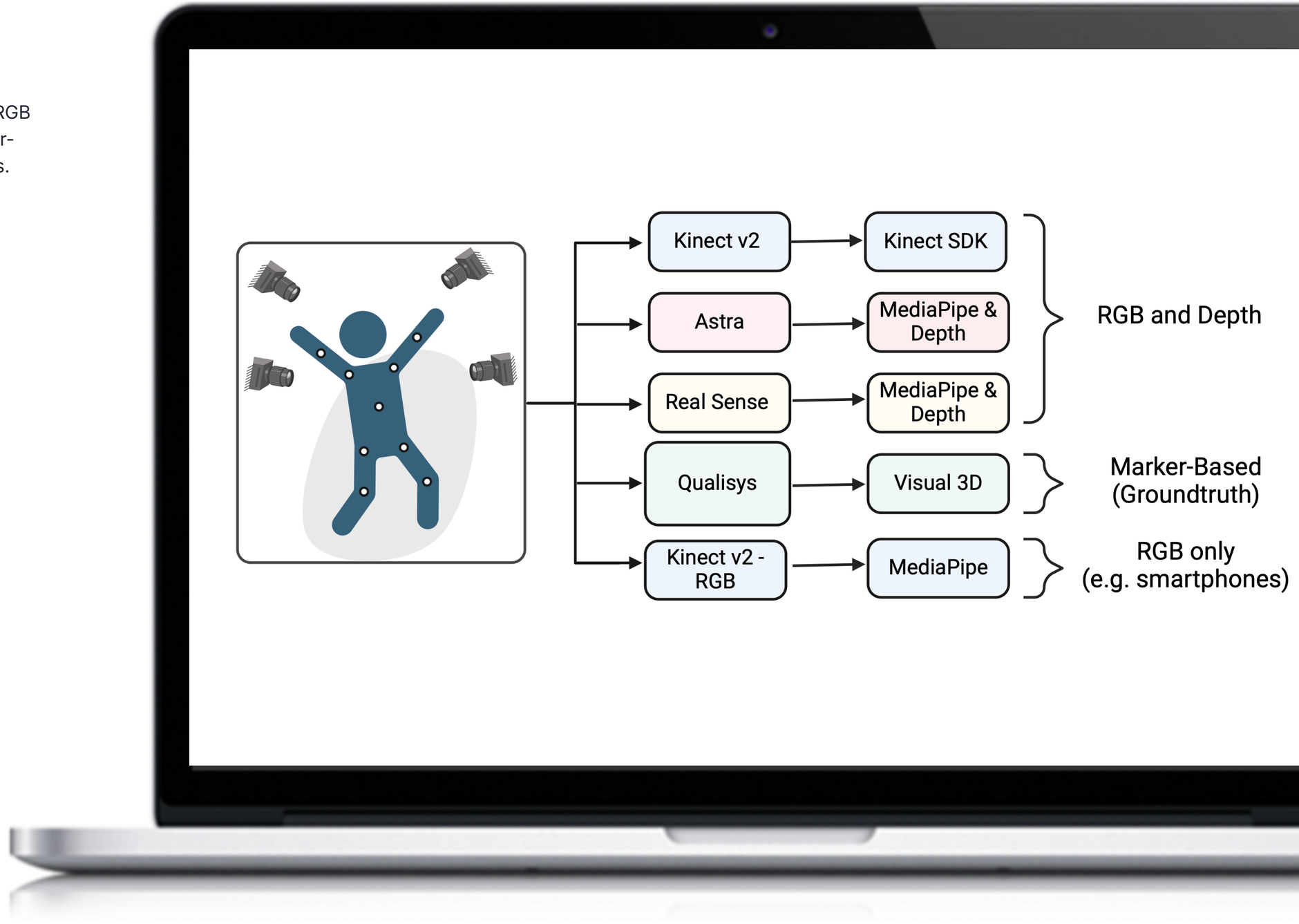
Qualitative Literature Survey to Validate ARPLO's Joint Angle Measurement Engine (Google MediaPipe-BlazePose)

The accuracy and reliability of markerless joint kinematics measurement systems are paramount for their acceptance and widespread use in clinical and research settings. This section delves into the various techniques employed to validate these systems, ensuring their effectiveness and reliability in accurately capturing human motion. The gold standard validation method for MMC systems is comparing their joint angle measurements to the ones measured by marker-based motion capture systems [4] or goniometer for quasistatic measurements [20]. This literature review focuses on validation of Arplo's MMC Engine Google MediaPipe with gold standard marker-based systems and comparison against more expensive markerless systems and marker-based systems.

In a recent study [4], an MMC open-source tool Google MediaPipe, allowing MMC only via RGB images, was compared and validated against commercial systems also using the depth information (RGB-D) Kinect v2, Astra, and Intel Real Sense. RGB-D technologies are relatively cheaper compared to the marker-based systems. This makes it less cost effective and more difficult to integrate with smartphones. Validation of Google MediaPipe only with RGB recordings of the motion is essential to integrate such MMC technologies to smartphones. Also, all systems were validated with a groundtruth measurement through a marker-based system, Qualisys [4]. The data flow of the comparison is summarized in Figure 3 and adapted from reference [4]. Shoulder, elbow, hip, and knee motions were simultaneously captured through RGB video recording, three depth sensors from Kinect v2, Astra, Real Sense, and marker-based system Qualisys. RGB recordings were processed with MediaPipe. They concluded that MediaPipe, using only the RGB recordings, resulted in less joint angle error and overperformed commercial RGB-Ds. This showed how simple RGB based MMC can be useful, such as simple cellphone camera usage. Figure 2 shows the pipeline of validation study. Volunteers with markers perform certain movements, while the systems were collecting data simultaneously for each test. Another study compared BlazePose MMC, a part of Google MediaPipe, via RGB recordings, to a commercial marker-based motion capture system Vicon [18]. Ten subjects were recorded via GoPro RGB camera and Vicion motion capture system using markers simultaneously. BlazePose validated with root mean square error compared to Viconn for knee, hip, and ankle joint angle measurements, with a maximum error of ~14°. Detailed results will be discussed in the last section. In summary, Google MediaPipe-BlazePose is a validated, strong MMC engine, viable for smartphone usage.

Figure 3

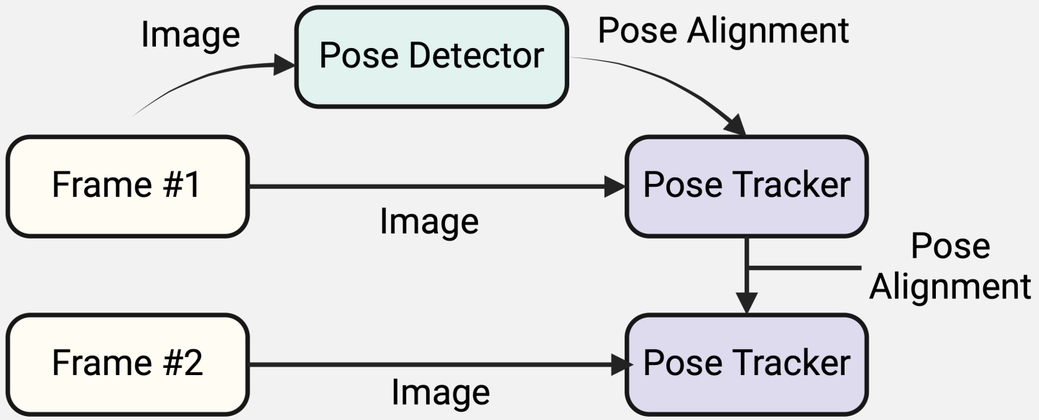
Pipeline for performance evaluation of RGB-D (Kinect v2, Astra, Realsense), RGB (processed via MediaPipe) with marker-based motion capture system Qualisys. Figure was prepared in BioRender and adapted from [4].



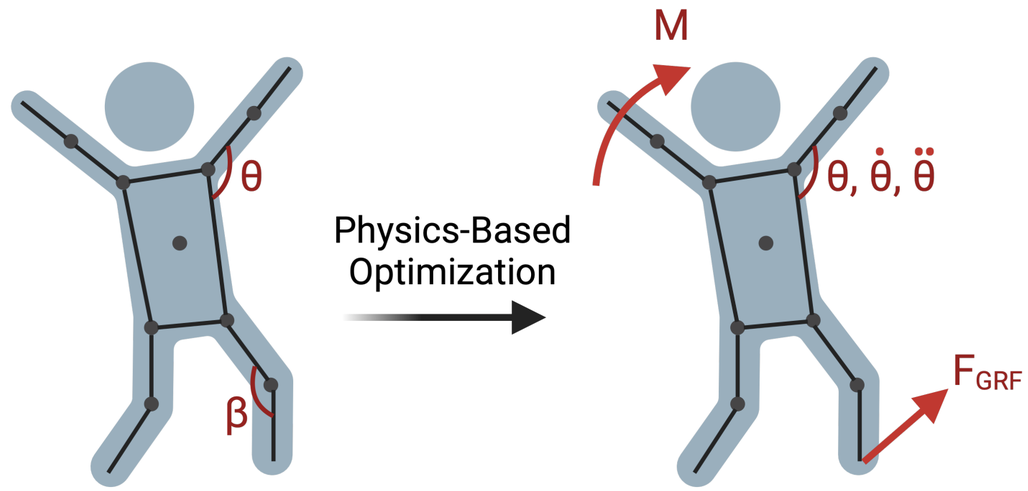
Constituents of Google MediaPipe (GMP): BlazePose and Generative Human Modeling Pipeline (GHUM)

GMP provides flexible development environment for MMC in Android, Python, Web, or iOS platforms. GMP lets us detect certain landmarks of human bodies in a video, which is handled as series of images. It identifies key body locations that can later be used for extraction of kinematic parameters, such as joint angles and eventually transformed into meaningful medical parameters. A smartphone MMC app should be fast, accurate, real-time and on-device output generations. Therefore, GMP used BlazePose infrastructure developed by Bazarevsky et. al. [8], which is based on 33-points body key points topology. While this level of detail provides enough information for our purposes, it keeps computational requirements low, allowing a seamless user experience. BlazePose first runs a detector for the first frame of the video and locates region of interest (ROI) for 33 pose key points by using a machine-learning based pipeline. For the following frames, BlazePose just tracks the detected ROIs, which makes the pipeline computationally lightweight and useful for mobile and real-time applications.

Figure 4 – Human topology of BlazePose – part of GMP (left) performs accurate joint tracking (right). Figure is taken from Google Research, post on pose detection via Google MediaPipe-Blaze Pose [21]



BlazePose is later extended by Xu et. al. via new toolbox named as Generative Human Modeling Pipeline (GHUM) [9], a deep learning framework trained with 60,000 diverse human configurations. Overall, GMP Pose Estimation toolbox is made based on BlazePose and GHUM and designed to be computationally lightweight and feature real-time on- device interface [10].



Dynamic variable estimation is summarized in a study by Stanford university, visualized in Figure 6 [11]. A motion estimation engine provides kinematic variables, which are then processed by a physics-based optimization engine that will calculate the dynamic variables such as joint/limb velocities, accelerations, and forces acting on the body. Eventually, these are processed into clinically relevant scores. Based on all the discussion, general workflow of Arplo is presented in Figure 7 as (1) measurement of joint angles via MMC engine, (2) extraction of kinematic and dynamic parameters, (3) generating reports for physical consultation and calculating clinical wellness scores.

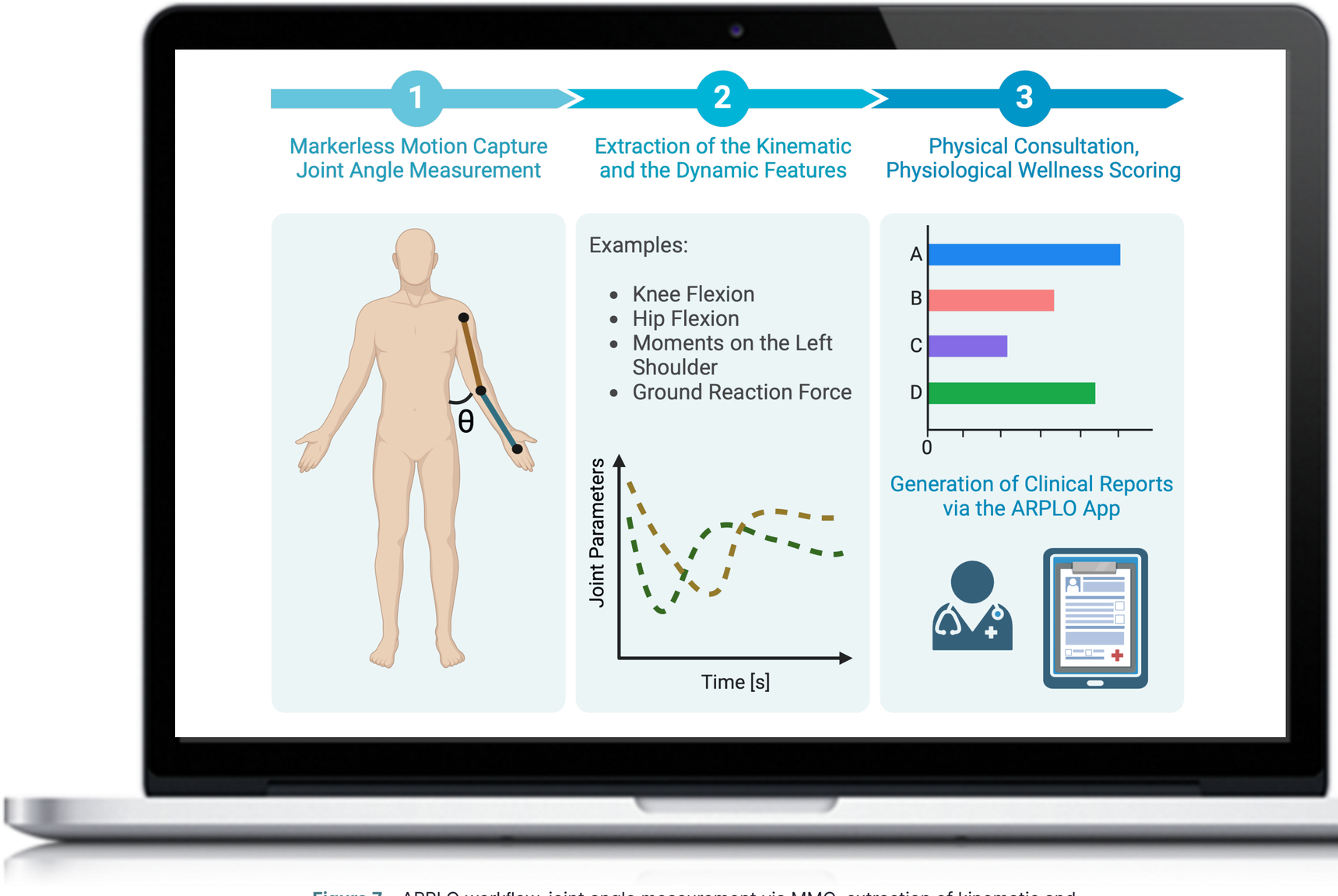


Figure 7 – ARPLO workflow, joint angle measurement via MMC, extraction of kinematic and dynamic features, preparing reports for physicians.

Quantitative Validation of ARPLO Joint Angle Measurement Engine GMP

1. Comparing GMP With Open-Source State of the Art (SOTA) Validated MMC

In this section, the performance of GMP, which is built on BlazePose-GHUM architecture, was compared with other validated SOTA motion capture software [10]: SPIN [12] , HUND [13], and THUNDR [14]. SPIN, HUND, and THUNDR are validated with large datasets of human poses: Human3.6M [15] and 3DPW [16] with millions of human motion frames, represent a strong validation data set as the original data were collected via marker based systems and inertial measurement units (IMUs) respectively.

A renowned motion capture error metric was used for validation: mean per joint positional error (MPJPE) [17]. It is defined as the mean distance between the predicted 3D joint locations and the corresponding ground truth joint locations (mm). As a result, all the architectures showed low MPJPE around 120-150 mm considering the length scales of human body, with GMP showed slightly better performance with 121 mm MPJPE as shown in Table 1 [10].

Table 1 – mean per joint positional error (MPJPE) of SOTA based on challenging data set. Units are in mm.

Method	MPJPE
SPIN	139.5
HUND	156.0
THUNDR (Marker)	138.0
BlazePose- GHUM (GMP)	121.0

2. Validating GMP With Commercial MMC Systems and Golden Standard Marker Based Technologies

Joint angle estimation performance of GMP was compared to commercial MMCs, which also use depth information: Kinect, Astra, and RealSense. The groundtruth data was provided by a commercial marker-based motion capture system Qualisys [4]. Motion was recorded as RGB frames and fed to GMP pipeline. Results of quantitative comparison are presented in Table 2 with the metric, mean absolute joint angle errors. MediaPipe overperformed commercial solutions even though depth data was not being utilized.

Table 2 – Mean absolute joint angle errors and their deviation, evaluated for commercial MMC and MediaPipe. Ground truth is marker based Qualisys motion capture system [4].

Sensor/Data	All Data	Upper Limbs	Lower Limbs
Astra	11.60 $ 3.71	12.36 $ 4.27	10.84 $ 3.27
Intel	11.56 $ 4.38	11.56 $ 3.74	11.57 $ 5.32
Kinect	12.65 $ 5.99	16.01 ÷ 6.72	9.30 ÷ 2.62
MediaPipe	8.57 $ 3.06	9.98 ± 3.79	7.16 ÷ 1.21

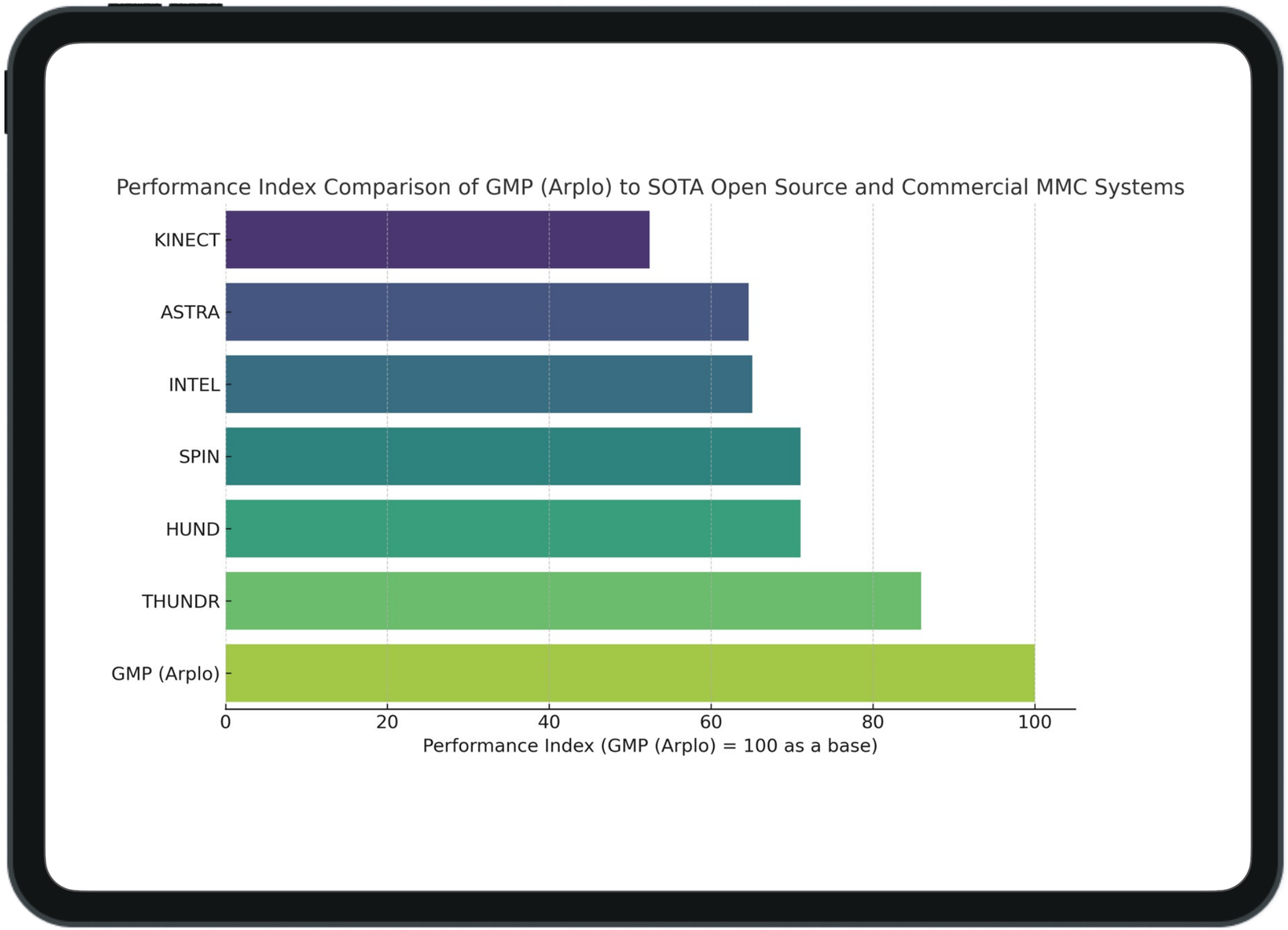
Two independent studies comparing GMP to commercial, validated marker-based and MMC systems based on mean absolute joint angle error, and to SOTA open source MMC systems based on MPJPE were summarized together in Table 3. To set a common baseline, we calculated relative performance index which represents what percentage GMP is better with respect to related comparison parameter. GMP, used by Arplo, was taken as 100 over 100 performance. Relative success is calculated based on the comparison of corresponding metrics that GMP and commercial systems achieved in each study. Results are plotted in Figure 8. All the other tools overperformed by GMP in their corresponding metric and according to the reported data in the literature.

Table 3 – Summary of performance index calculation. It was comparison of common comparison parameters as a percentage.

System	System Type	Performance	Comparison Parameter
GMP (Arplo)	Open-Source SOTA	100.0	-
THUNDR	Open-Source SOTA	85.95	MPJPE
SPIN	Open-Source SOTA	71.07	MPJPE
HUND	Open-Source SOTA	71.07	MPJPE
INTEL	Commercial Product	65.11	Mean Absolute Error
ASTRA	Commercial Product	52.50	Mean Absolute Error
KINECT	Commercial Product	52.39	Mean Absolute Error

Figure 8

Performance Index of GMP compared to other MMC. GMP was taken as baseline, and others were compared with respect to the baseline. Lower values represent lower percentage. GMP was compared to THUNDR, SPIN, and HUND based on MPJPE. GMP was compared to INTEL, ASTRA, and KINECT based on absolute angle error.



3. Validating GMP With Gold Standard Marker Based Motion Capture Systems

Here we summarize the comparison of GMP with Qualisys [4] and another marker-based motion capture system Vicon [18]. Comparison with Vicon in terms of lower limbs: knee, hip, and ankle are summarized in Figure 9. Whereas Qualisys resulted in ~7° mean absolute joint angle error for lower limbs [4] compatible with validation by Vicon which gives mean absolute joint angle error as ~6.7°.

While literature reinforces validity of Arplo's engine GMP, we aim to validate finalized POC with gold standard goniometer and marker-based motion capture systems.

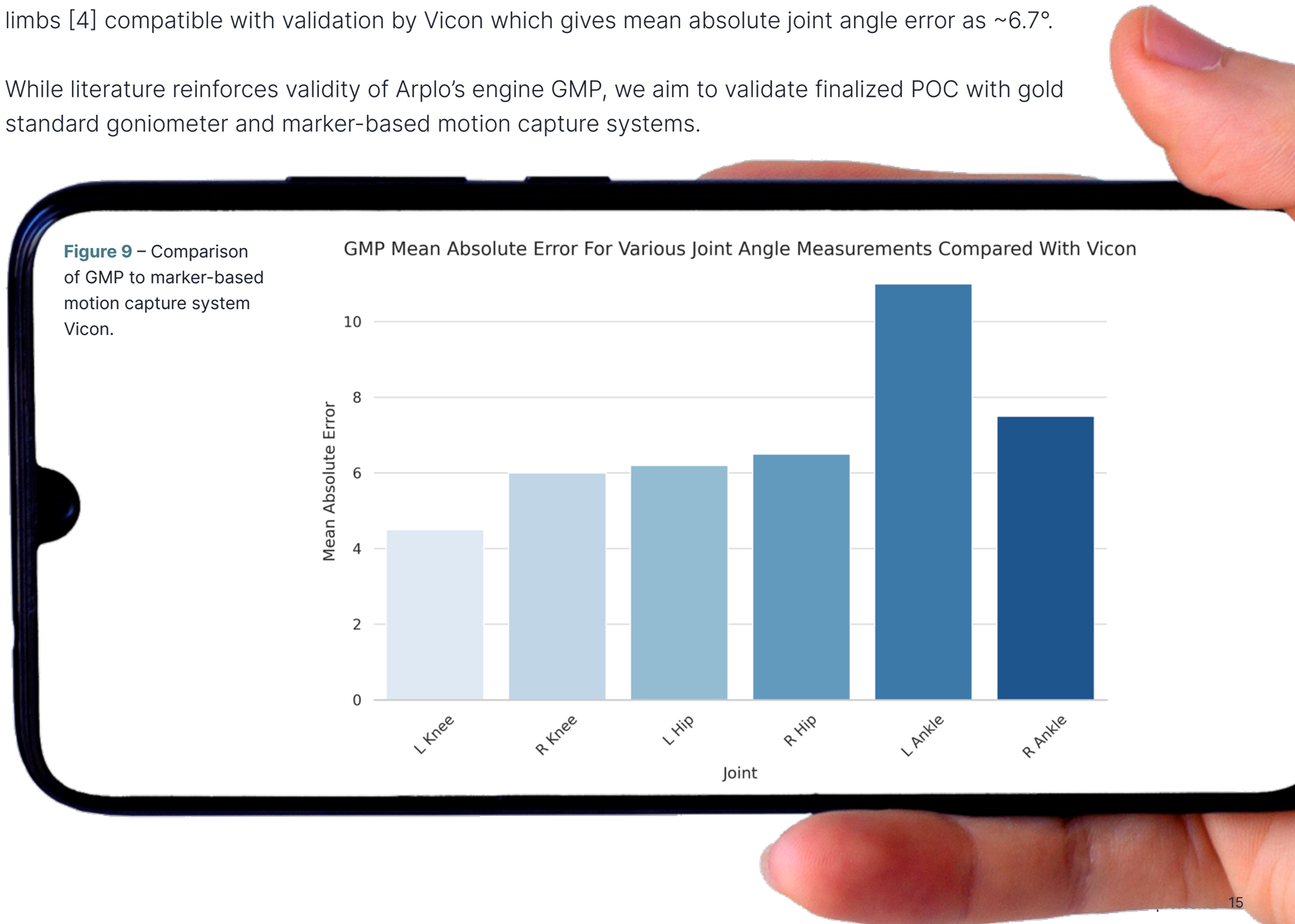


Figure 9 – Comparison of GMP to marker-based motion capture system Vicon.

References

References

1. W. W. T. Lam, Y. M. Tang, and K. N. K. Fong, “A systematic review of the applications of markerless motion capture (MMC) technology for clinical measurement in rehabilitation,” J. NeuroEngineering Rehabil., vol. 20, no. 1, p. 57, May 2023.
2. “NIH Stroke Scale | National Institute of Neurological Disorders and Stroke.” Accessed: Jan. 07, 2024. [Online]. Available: https://www.ninds.nih.gov/health-information/public-education/know-stroke/health- professionals/nih-stroke-scale
3. Kidziński, Ł., Yang, B., Hicks, J.L. et al. Deep neural networks enable quantitative movement analysis using single-camera videos. Nat Commun 11, 4054 (2020).
4. T. B. de G. Lafayette et al., “Validation of Angle Estimation Based on Body Tracking Data from RGB-D and RGB Cameras for Biomechanical Assessment,” Sensors, vol. 23, no. 1, p. 3, Dec. 2022.
5. G. Amprimo, G. Masi, and G. Pettiti, “Hand tracking for clinical applications: validation of the Google MediaPipe Hand (GMH) and the depth-enhanced GMH-D frameworks.”.
6. B. Horsak et al., “Concurrent validity of smartphone-based markerless motion capturing to quantify lower-limb joint kinematics in healthy and pathological gait,” J. Biomech., vol. 159, p. 111801, Oct. 2023.
7. Lawin FJ, Byström A, Roepstorff C, Rhodin M, Almlöf M, Silva M, Andersen PH, Kjellström H, Hernlund E. Is Markerless More or Less? Comparing a Smartphone Computer Vision Method for Equine Lameness Assessment to Multi-Camera Motion Capture. Animals. 2023; 13(3):390.
8. V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, “BlazePose: On-device Real- time Body Pose tracking.” arXiv, Jun. 17, 2020.
9. H. Xu, E. G. Bazavan, A. Zanfir, W. T. Freeman, R. Sukthankar, and C. Sminchisescu, “GHUM & GHUML: Generative 3D Human Shape and Articulated Pose Models,” in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA: IEEE, Jun. 2020, pp. 6183–6192.
10. I. Grishchenko et al., “BlazePose GHUM Holistic: Real-time 3D Human Landmarks and Pose Estimation.” arXiv, Jun. 23, 2022.
11. D. Rempe, L. J. Guibas, A. Hertzmann, B. Russell, R. Villegas, and J. Yang, “Contact and Human Dynamics from Monocular Video,” in Computer Vision – ECCV 2020, vol. 12350, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds., in Lecture Notes in Computer Science, vol. 12350. , Cham: Springer International Publishing, 2020, pp. 71–87.
12. N. Kolotouros, G. Pavlakos, M. J. Black, and K. Daniilidis, “Learning to Reconstruct 3D Human Pose and Shape via Model-fitting in the Loop.” arXiv, Sep. 27, 2019. Accessed: Jan. 21, 2024.
13. A. Zanfir, E. G. Bazavan, M. Zanfir, W. T. Freeman, R. Sukthankar, and C. Sminchisescu, “Neural Descent for Visual 3D Human Pose and Shape.” arXiv, Jun. 14, 2021. Accessed: Jan. 21, 2024.
14. M. Zanfir, A. Zanfir, E. G. Bazavan, W. T. Freeman, R. Sukthankar, and C. Sminchisescu, “THUNDR: Transformer- based 3D HUmaN Reconstruction with Markers.” arXiv, Jun. 17, 2021.
15. C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, “Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 36, no. 7, pp. 1325– 1339, Jul. 2014.
16. T. Von Marcard, R. Henschel, M. J. Black, B. Rosenhahn, and G. Pons-Moll, “Recovering Accurate 3D Human Pose in the Wild Using IMUs and a Moving Camera,” in Computer Vision – ECCV 2018, vol. 11214, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds., in Lecture Notes in Computer Science, vol. 11214., Cham: Springer International Publishing, 2018, pp. 614–631.
17. “Mean Per Joint Position Error (MPJPE).” Accessed: Jan. 22, 2024. [Online]. Available: https://oecd.ai/en/tool- use-cases/mean-per-joint-position-error-(mpjpe)
18. A. A. Hulleck et al., “Accuracy of Computer Vision-Based Pose Estimation Algorithms in Predicting Joint Kinematics During Gait,” In Review, preprint, Aug. 2023.
19. Uhlrich SD, Falisse A, Kidziński Ł, Muccini J, Ko M, et al. (2023) OpenCap: Human movement dynamics from smartphone videos. PLOS Computational Biology 19(10): e1011462
20. Lam WWT, Fong KNK. Validity and Reliability of Upper Limb Kinematic Assessment Using a Markerless Motion Capture (MMC) System: A Pilot Study. Arch Phys Med Rehabil. 2024 Apr;105(4):673-681.e2
21. “On-device, Real-time Body Pose Tracking with MediaPipe BlazePose.” Accessed: Jul. 22, 2024. [Online]. Available: http://research.google/blog/on-device-real-time-body-pose-tracking-with-mediapipe-blazepose/

Email: info@arplo.com • Phone: 504-215-8337
Web: www.arplo.com