Robustness of markerless biomechanical analysis using pose2sim: a sensitivity study on transmission and privacy constraints for cloud-based computation

Corentin Hardy*, Erwan Fertray*, Laurent Guillo, Charles Pontonnier and Pierre Hellier
IRISA, INRIA, Université de Rennes,
* denotes equal contribution, corentin.a.hardy@irisa.fr

Abstract

Biomechanical analysis for performance monitoring, fatigue analysis and injury risk prevention has traditionally been confined to laboratory settings using marker-based motion capture. While markerless approaches using multi-view computer vision have emerged, their reliability under "in-the-wild" or broadcast constraints remains underexplored. In the context of cloud computation, where all content has to be sent over the network, we investigate the feasibility of near real-time privacy-preserving performance monitoring by assessing the sensitivity of the Pose2Sim pipeline to degradation in acquisition parameters. Specifically, we evaluate the impact of face deidentification, camera count, framerate, image resolution, and H.264 video compression (QP values) on the recovery of biomechanical quantities. Except for face deidentification, which is required to preserve participant anonymity, all other values influence the total bitrate, which is limited in real-time scenarios. Using the AthletePose3D dataset, we establish a high-fidelity ground truth and measure the joint angle deviations under constrained conditions. Our results provide a guideline for minimum technical requirements to maintain biomechanical fidelity in unconstrained environments. Specifically, we recommend using video compression, while framerate is the most critical factor in acquisition settings. More importantly, we report that face deidentification can degrade the extraction of faithful biomechanical quantities.

1. Introduction

Assessing injury risk in occupational and athletic settings relies heavily on the accurate estimation of biomechanical quantities, including joint angles, velocities, and moments. While traditionally dependent on cumbersome marker-based systems, motion analysis has recently been democratized by advances in markerless pose estimation (e.g., OpenPose [1], RTMPose[3]) paired with biomechanically grounded pipelines such as OpenCap [12] and Pose2Sim [8].

Nevertheless, transitioning these systems from controlled laboratory environments to real-time monitoring in factories or stadiums introduces significant technical and ethical constraints. Because the computationally intensive nature of these pipelines generally necessitates offloading processing to remote cloud servers, real-world deployments utilizing existing surveillance or broadcast infrastructure are inherently limited by wireless transmission bandwidth and storage capacity. Furthermore, transmitting video data to external servers mandates stringent privacy protocols to preserve subject anonymity.

To satisfy these bandwidth limitations and privacy requirements, video streams are often subjected to heavy compression or temporal downsampling prior to cloud transmission, as well as subject deidentification with face blur or masking. Consequently, it is critical to quantify how these alterations (specifically H.264 video compression, reduced framerates, sparse camera configurations, face blur) impact the accuracy of the final biomechanical outputs. A primary objective of this study is to determine the extent to which compression artifacts (governed by the Quantization Parameter, QP) amplify errors in angular measurements, even in the presence of dedicated signal filtering. Furthermore, we aim to evaluate the effects of framerate subsampling, with a specific focus on its impact during highly dynamic athletic movements.

While previous works have evaluated the accuracy of 2D and 3D pose estimation [4, 10, 14], few have propagated these errors through the Inverse Kinematics (IK) pipelines to quantify the error in the actual biomechanical quantities used to analyse motion and prevent injuries, and assess the impact of acquisition parameters on these quantities. Indeed, uncertainties regarding kinematic quantities are largely responsible for uncertainties regarding dynamic quantities such as joint torques [6]. Moreover, external forces estimates, necessary to perform Inverse Dynamic (ID) analyses in the absence of ground reaction forces measurements, are mostly based on kinematic quantities and

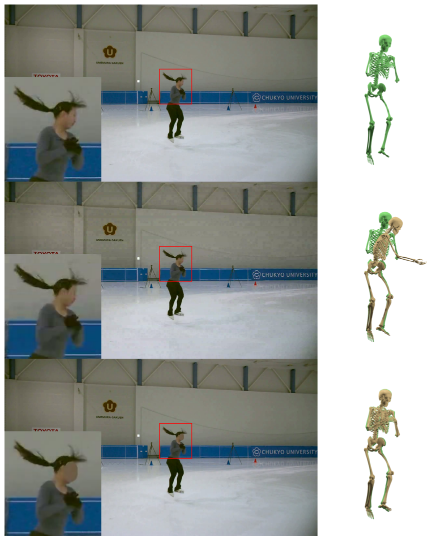

Figure 1. Examples of degradation and impact on the extracted osteoarticular model. Left column: original image, same frame at $QP = 34$, same frame with subject deidentification. Right column: osteoarticular models with ground truth (green) and estimated (orange). While face deidentification introduces almost no difference, severe video compression affects the quality of the model estimation.

contact detection conditions, also affected by the kinematic uncertainties [5].

Identifying these trade-offs allows for an optimal data acquisition strategy under strict bandwidth budgets. Specifically, we determine whether to prioritize spatial fidelity (low QP) or temporal resolution (high framerate) for robust in-the-wild motion capture.

Pose2Sim has previously been assessed in terms of reduced camera counts, applied blur, or calibration errors [7]. In this work, we present a systematic sensitivity analysis of the Pose2Sim pipeline where our contributions are:

1. A quantification of the impact of **video compression (H.264 QP)** on biomechanical output estimates, simulating broadcast conditions and transmission over the network for privacy-preserving cloud computation.
2. An evaluation of the impact of **camera count**, **spatial resolution**, **framerate** using the AthletePose3D dataset [13].
3. An evaluation of subject deidentification through face blur on the accuracy of biomechanical quantities.

2. Methodology

Pose2Sim is a standard markerless motion analysis pipeline [8]. 2D keypoints are detected from multiple views, triangulated into 3D via Pose2Sim - which performs robust triangulation - and then Inverse Kinematics (IK) is computed in OpenSim [2].

2.1. Ground Truth Generation

Since we lack marker-based ground truth for the selected datasets, we adopt a "Silver Standard" approach. The **Baseline Ground Truth (GT)** is defined as the output of the Pose2Sim pipeline using the highest quality input available: maximum number of cameras, original resolution, native framerate, and raw/lossless video data.

2.2. Degradation Protocols

We introduce controlled degradation to the input data to simulate real-time/broadcast constraints:

1. Video Compression (H.264): To simulate bandwidth constraints (e.g., wireless camera transmission), we encode the raw footage using FFmpeg with specific Quantization Parameters (QP). Based on broadcast standards, using the low-delay encoding mode, we test:

- *High Quality:* QP 21 (Visually transparent).
- *Broadcast Standard:* QP 26 (typical HDTV distribution).
- *Low Bandwidth:* QP 34 (noticeable artifacts).

2. Number of Cameras: We iteratively reduce the number of views N. Cameras are selected to maximize angular coverage in the sparse subsets.

3. Framerate: Using high-framerate source data, we temporally subsample the input to simulate lower framerates. We choose the following configurations: *High framerate capture* (60 frame/s), *Standard framerate capture* (30 frame/s) and *Low framerate capture* (15 frame/s).

4. Resolution: Video frame resolution is downscaled to 100%, 50%, and 25% of the native size before pose estimation.

5. Face deidentification: We apply the EgoBlur [9] to de-identify all the faces in the data for privacy context. This step applies a blur on all faces detected in the videos but requires a new encoding process. So, to observe the impact of the deidentification, we use the same compression parameters as the High Quality preset of Video Compression without other degradation. An example of deidentification is shown in Figure 1.

2.3. Metrics

We assess the deviation from the Baseline GT using the **Mean Absolute Error (MAE)**, computed by averaging the absolute differences across Joint Angles and time steps simultaneously. This single-value metric follows the approach of [11] and reflects the deviation from the Pose2Sim silver standard. Upper extremities were excluded due to inherent ambiguities in the Pose2Sim keypoint set, where the lack of constraints makes humeral orientation indeterminate, making direct comparison unreliable.

3. Experimental Setup

3.1. Dataset

AthletePose3D [13]: Proposed by Calvin *et al.*, this dataset contains 3 types of high-speed footage of athletes performing dynamic movements : Ice Skating, Throwing and Running. Each scene was captured with a multi-camera setup hardware-synchronized. The setup characteristics are summarized in Table 1. We use the test dataset containing 3 subjects.

Action type	Number cameras	Resolution	FPS	Number sequences
Ice skating	12	1920x1088	60	70
Throwing	8	1920x1088	60	30
Running	4	1280x768	120	37

Table 1. Characteristics of Athlete3D Pose sequences.

3.2. Implementation Details

The processing pipeline :

- **2D Pose Estimation:** We use RTMPose model (Large) due to its balance of speed and accuracy for real-time applications. It detects 26 keypoints per person per frame.
- **Triangulation:** The 2D keypoints are triangulated using a robust, weighted method that accounts for the confidence score of each keypoint. The Pose2Sim framework can also exclude specific cameras from the triangulation process when their reprojection error exceeds a predefined threshold.
- **Marker Augmentation:** We use the Stanford LSTM model [12] proposed in Pose2Sim to compute position of 47 virtual markers from the 3D keypoints.
- **Biomechanics Processing:** Before performing the IK, Pose2Sim scales a standard full-body model with 30 joints and 56 degrees of freedom (including 9 not impacted by the IK) using the 3D keypoints detected over the sequence. OpenSim resolves the IK of this model using 52 markers coming from the 3D keypoints and the LSTM virtuals markers across the time.

Results filtering We keep only the dataset sequences if the MAE of the minimal degradation (Video compression with QP21) is under a threshold of 5°. This filtering ensures that we compare only valid sequences, i.e. sequences where the markeless analysis is correctly done when there is no degradation.

4. Results

4.1. Individual impact of the parameters

In the figure 2, we report the MAE of Joints Articulation of each sequence depending on one parameter degradation (Number of views, Framerate, Resolution and Video Compression) . We observe that:

- The MAE varies significantly depending on the sequence type; in particular, the Ice skating sequences, where the movement is the fastest, obtain a large MAE whatever the parameter degradation.
- Framerate degradation has the strongest impact on all sequences.
- Conversely, video compression is the parameter which has the least impact on the MAE and can divide the video bandwidth by 3 with a reasonable error on the result.
- Reducing resolution has lower impact on MAE than reducing the camera count below 4 views. However, the gain on the bandwidth is not as good as that of resolution and video compression.

4.2. Privacy

Figure 3 shows the distribution of MAE in presence of subject deidentification through face blur. For each action type, the distribution of errors is plotted. For the *ice skating* sequences, face blur might have a noticeable impact up to 2.5 degrees. For the other sequences (*running* and *throwing*), the error is limited, demonstrating the feasibility of subject deidentification.

4.3. Optimal choice with bandwidth constraints

The initial motivation of our work was to provide guidance to users, concerning acquisition parameters, when bandwidth constraints are present. Figure 4 plots the optimal pareto frontier for the complete combination of parameters. This figure provides the lower bound of angular error, given the bitrate. Additionally, the optimal configuration is shown in the legend (from the lowest bitrate/higher error on the top to the highest bitrate/lowest error), providing insight about the useful parameters to achieve a given bandwidth. QP parameter, as well as spatial subsampling, are the two primary factors to be tuned.

5. Discussion and conclusion

In this paper, we studied the influence of acquisition and transmission parameters and subject deidentification when the extraction of biomechanical quantities is performed over the network. The purpose was to evaluate the effect of parameters, and provide guidance to limit the bitrate, while ensuring the accuracy of extracted parameters. We have shown that large and complex motions are prone to errors when bandwidth and privacy constraints are introduced. For other motion types, the most sensitive parameters are the framerate and number of views. Moreover, the video compression and the image downscaling can reduce the bandwidth by 3 or 4 maintaining an error below that of the markerless pipeline (usually around 1.5 degrees [7]). We have shown that face deidentification can impact the accuracy.

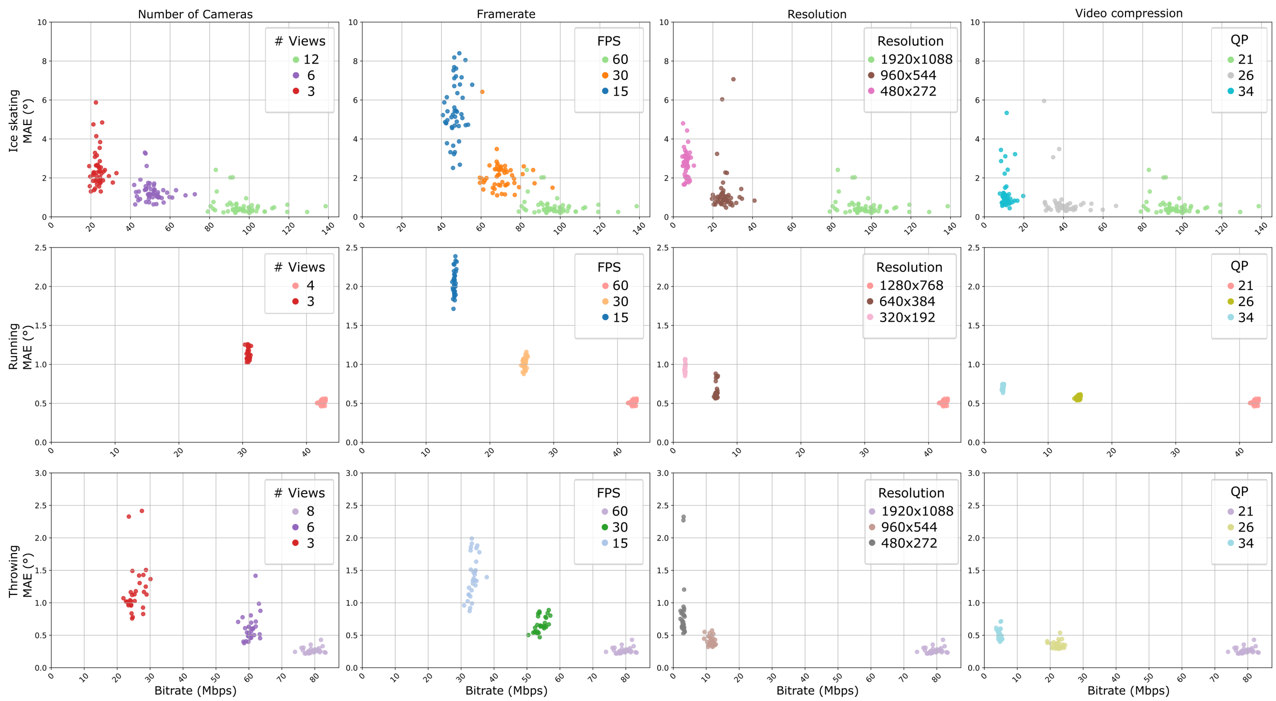

Figure 2. MAE depending of the configuration. For each action type (lines) and for each degradation (columns), the MAE in degrees is plotted against the bitrate in Mbps. Each dot represents a sequence of the dataset with a parameter configuration given by the color. The *ice skating* action is more sensitive to parameters, with average errors up to 8 degrees. In general, the framerate and the number of cameras affect the results more than spatial resolution and QP parameters.

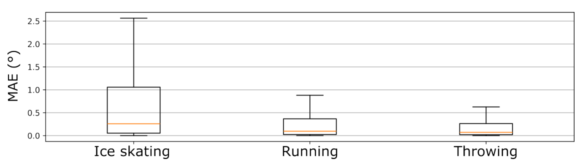

Figure 3. Distribution of the MAE over scene type when the face deidentification is applied. In most cases, the error is very low, demonstrating the usefulness of subject deidentification. However, for the *ice skating* sequence, errors up to 2.5 degrees might be introduced.

Fine-tuning the RTMPose model with datasets containing blurred faces could mitigate this effect. These results provide a bound for parameters in presence of requirements, such as privacy preservation, maximal allowed angular error, or limited bandwidth. All experiments have been compared to the initial Pose2Sim setup, referred to as the silver standard. This is a limitation that could be tackled in the future by capturing high quality data that would be associated with trusted ground truth parameters (joint angles). Additionally, future work should also investigate the spatial distribution of joint angle errors, to have a more precise evaluation of errors related to specific articulations (knee/neck for instance).

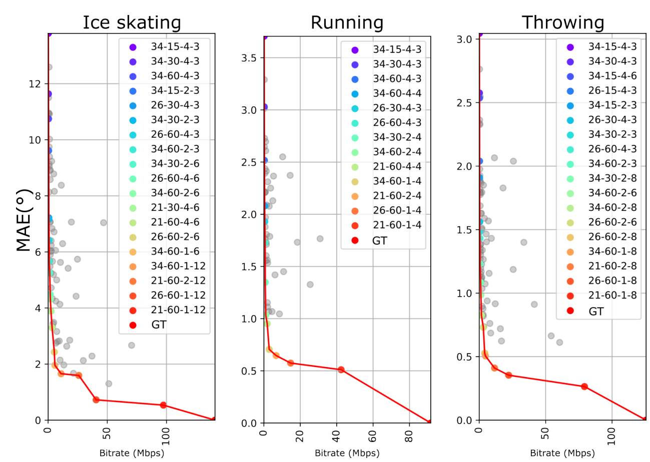

Figure 4. Pareto frontier of configurations based on the mean of the MAE over all sequences. A configuration is described with the following notation [QP]-[FPS]-[Downscaling Factor]-[Number of views]. The ground truth is denoted GT. By exploring the entire set of configurations, the pareto frontier provides the optimal setup in terms of accuracy-bitrate trade-off. The greys dots represents configurations that are not on the pareto frontier. For the *running* and *throwing* sequences, the bitrate can be drastically reduced by a factor 10 with a limited impact on the angle measurement.

References

[1] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. 1

[2] Scott L Delp, Frank C Anderson, Allison S Arnold, Peter Loan, Ayman Habib, Chand T John, Eran Guendelman, and Darryl G Thelen. Opensim: open-source software to create and analyze dynamic simulations of movement. *IEEE transactions on biomedical engineering*, 54(11):1940–1950, 2007. 2

[3] Tao Jiang, Peng Lu, Li Zhang, Ningsheng Ma, Rui Han, Chengqi Lyu, Yining Li, and Kai Chen. Rtmpose: Real-time multi-person pose estimation based on mmpose, 2023. 1

[4] Wu Liu, Qian Bao, Yu Sun, and Tao Mei. Recent advances of monocular 2d and 3d human pose estimation: A deep learning perspective. *ACM Computing Surveys*, 55(4):1–41, 2022. 1

[5] Pauline Morin, Antoine Muller, Georges Dumont, and Charles Pontonnier. Comparison of two contact detection methods for ground reaction forces and moment estimation during sidestep cuts, runs, and walks. *Journal of Biomechanical Engineering*, 146(1):014503, 2024. 2

[6] Antoine Muller, Charles Pontonnier, and Georges Dumont. Uncertainty propagation in multibody human model dynamics. *Multibody System Dynamics*, 40(2): 177–192, 2017. 1

[7] David Pagnon, Mathieu Domalain, and Lionel Reveret. Pose2sim: An end-to-end workflow for 3d markerless sports kinematics—part 1: Robustness. *Sensors*, 21(19), 2021. 2, 3

[8] David Pagnon, Mathieu Domalain, and Lionel Reveret. Pose2sim: An open-source python package for multiview markerless kinematics. *Journal of Open Source Software*, 7(77):4362, 2022. 1, 2

[9] Nikhil Raina, Guruprasad Somasundaram, Kang Zheng, Sagar Miglani, Steve Saarinen, Jeff Meissner, Mark Schwesinger, Luis Pesqueira, Ishita Prasad, Edward Miller, Prince Gupta, Mingfei Yan, Richard Newcombe, Carl Ren, and Omkar M Parkhi. Egoblur: Responsible innovation in aria, 2023. 2

[10] Sofia Scataglini, Eveline Abts, Cas Van Bocxlaer, Maxime Van den Bussche, Sara Meletani, and Steven 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):3686, 2024. 1

[11] Anne Schmitz, Mao Ye, Grant Boggess, Robert Shapiro, Ruigang Yang, and Brian Noehren. The measurement of in vivo joint angles during a squat using a single camera markerless motion capture system as compared to a marker based system. *Gait & posture*, 41(2):694–698, 2015. 2

[12] Scott D Uhlrich, Antoine Falisse, Łukasz Kidziński, Julie Muccini, Michael Ko, Akshay S Chaudhari, Jennifer L Hicks, and Scott L Delp. Opencap: Human movement dynamics from smartphone videos. *PLoS computational biology*, 19(10):e1011462, 2023. 1, 3

[13] Calvin Yeung, Tomohiro Suzuki, Ryota Tanaka, Zhuoer Yin, and Keisuke Fujii. Athletepose3d: A benchmark dataset for 3d human pose estimation and kinematic validation in athletic movements. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 5945–5956, 2025. 2, 3

[14] Ce Zheng, Wenhan Wu, Chen Chen, Taojiannan Yang, Sijie Zhu, Ju Shen, Nasser Kehtarnavaz, and Mubarak Shah. Deep learning-based human pose estimation: A survey. *ACM computing surveys*, 56(1):1–37, 2023. 1