2026-4 | 51th congress of the Société de Biomécanique
27-29 October 2026, Montpellier, France

Multidisciplinary Biomechanics Journal

# 3D Kinematics Using a Fast Extended Kalman Filter Relying on Multiview 2D Pose Observations: A preliminary Implementation

Mickaël Begon[a*], Léa Drolet-Roy[b], Bérenger Chedal-Anglay[b], Eve Charbonneau[c], Lama Seoud[b]

a Université de Montréal, Montréal (Qc) Canada
b Polyechnique Montréal, Montréal (Qc) Canada
c Université de Sherbrook, Sherbrooke (Qc) Canada
*Corresponding author: mickael.begon@umontreal.ca

**Received date:** 01/04/2026
**Accepted date:** 25/06/2026
**Publication date:** 27/10/2026

**Keywords:** markerless, kinematics, twisting somersault, trampoline

© 2026 The Authors
Licence CC-BY 4.0
Published by Société de Biomécanique

## 1. Introduction

Markerless biomechanical analysis has progressed rapidly with the development of deep-learning-based pose estimators and synchronized multi-camera workflows. Most pipelines follow a sequential strategy: 2D pose estimation in each view, triangulation, and filter or biomechanical model fitting (Pagnon et al., 2021). Pose errors, occlusions, and view-dependent ambiguities degrade the reconstructed 3D kinematics. This limitation is particularly important for acrobatic movements, for which the body undergoes rapid changes in orientation, frequent self-occlusions, and occasional front/back ambiguities in the image plane. We need to move part of the robustness strategy upstream, at the level of the original 2D detections, while keeping the final state variables expressed as generalized coordinates of an articulated model.

Multibody kinematics optimization, and Kalman-based approaches have all been proposed to regularize the reconstruction problem and reduce non-physical solutions (Begon et al., 2018). Rather than relying exclusively on a 3D triangulate data, we focus here on an image-space extended Kalman filter (EKF) where the measurement is updated directly from synchronized 2D detections ($EKF_{2D}$). Confidence assigned to each observation should reflect not only the score provided by the pose estimator, but also its multiview geometric consistency. A secondary objective is to explicitly address front/back inversions caused occasionally by the pose estimator. Frames associated with abnormally large epipolar inconsistencies are treated as potentially ambiguous and are re-evaluated under left/right-swapped hypotheses.

## 2. Methods

### 2.1 Pose estimator and 2D data processing

ViTPose estimator was fine tuned for acrobatic moves using 3624 synthetic images generated from motion capture data, improving the "mean average precision" as defined in COCO from 55.8% to 73.1% (Drolet-Roy, 2026). As filtering 3D trajectories is not the most appropriate treatment (Begon et al., 2010), aberrant 2D poses were deleted, when the Euclidean distance between raw and moving-average smoothed 2D-points exceed 10% of the 2D-point range. Moreover, in camera-frames with abnormally high temporal and epipolar inconsistencies ($\overline{e}_c$: median of all available camera pairs using Sampson error), left and right 2D-points were inverted if it reduced the inconsistencies.

### 2.2 Kalman filter with multiview 2D observations

From $\overline{e}_{c,p,t}$ (camera $c$, 2D-point $p$, and frame $t$) a consistency score was calculated (with $\tau$=15 px): $\gamma_{c,p,t} = \exp\left(-1/2\left(\overline{e}_{c,p,t}/\tau\right)^2\right)$ which was combined with the confidence score returned by the 2D pose estimator to define the effective confidence in the 2D-point observations ($\tilde{\gamma}$). This model does not require time-consuming triangulation and reprojection error. The measurement variance was then defined as: $\sigma^2_{c,p,t} = \lambda_R\left(4/\max\left(\tilde{\gamma}_{c,p,t}, 10^{-3}\right)\right)^2$, with $\lambda_R = 1$ is the global

**To cite this article:** Mickaël Begon, Léa Drolet-Roy, Bérenger Chedal-Anglay, Eve Charbonneau, Lama Seoud, 3D Kinematics Using a Fast Extended Kalman Filter Relying on Multiview 2D Pose Observations: A preliminary Implementation, *Multidisciplinary Biomechanics Journal*, 2026, vol. 4, p. 237–239, DOI: 10.46298/mbj.19188.

measurement-noise scale; the constant 4 defines the nominal pixel standard deviation for a fully trusted observation, while the lower bound $10^{-3}$ prevents divergence when the effective confidence becomes small.

The $EKF_{2D}$ state $[\mathbf{q}, \dot{\mathbf{q}}, \ddot{\mathbf{q}}]^{\top}$ includes the generalized coordinates and derivates of the multibody model (10 segments with 25 degrees-of-freedom). The prediction step applies a constant joint acceleration. When airborne, the root-segment acceleration follows a free-floating dynamics (Puchaud et al., 2023). The correction step is performed in image space. The model markers are projected into each camera and compared with the observed 2D-points. The observation model was linearized with respect to $\mathbf{q}$, yielding a measurement Jacobian obtained by chaining the model-marker Jacobian with the Jacobian of the epipolar constraints. The measurement covariance is adapted frame by frame from the effective confidence defined above.

### *2.3 Data collection and analysis*

Trampolinists were recorded by 8 synchronized cameras (1080p; 120 Hz) during a training. We used a trial with five twisting somersaults and a vertical jump (14.55 s) as a proof-of-concept. Standard triangulation (lasting 5.60 s) failed reconstructing the 3D kinematics (Fig. 1; 2 missing twists).

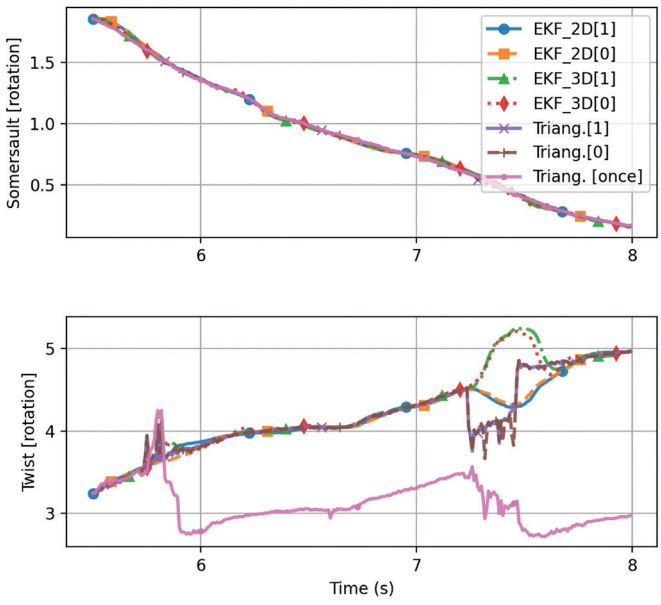


**Figure 1.** Body rotation in somersaults and twist for $EKF_{2D}$, $EKF_{3D}$, exhaustive triangulation with [1] and without [0] 2D-point processing, as well as standard triangulation. Credit: Mickaël Begon

To assess its performance, the trunk kinematics obtained with $EKF_{2D}$ was compared with 1) iterative triangulation with a 15 px threshold (Pagnon et al., 2021), the trunk orientation being decomposed into somersault, tilt, and twist angles (Triang.); 2) a 3D Kalman-filter-based approach ($EKF_{3D}$). Each method was run with and without 2D data processing (exclusion and left/right inversion). Average reprojection distances, proportion of 2D-point flip, and processing times (Apple M2) were reported.

## 3. Results and discussion

The 3D multibody model (generated once for each athlete) took 5.75 s. Between 1.2% and 11.1% of frame were considered inverted (4.62 s) with very few false positives. $EKF_{2D}$ was fast (<6 s; Table 1). The 2D-processing hastened the triangulation (37.56 vs 58.35 s) since it required less iterations. $EKF_{3D}$ added only ~1.11 s (C++ code only).

**Table 1.** Performance of the three methods with and without 2D processing (average reprojection distance, processing time).

| Method | 2D proc. | Repro. (px) | Time (s) |
|---|---|---|---|
| $EKF_{2D}$ | no | 26.54 | 4.14 |
| $EKF_{2D}$ | yes | 21.60 | 5.71 |
| $EKF_{3D}$ | no | 26.41 | 59.49 |
| $EKF_{3D}$ | yes | 19.80 | 38.73 |
| Triang. | no | 22.64 | 58.35 |
| Triang. | yes | 15.65 | 37.56 |

Reprojection distances varied between methods (Table 1) with a reduction of >5 px with 2D-point processing. Rigidity constraints of the multibody model led to increased reprojection distances. Trunk kinematics (Fig. 1) were more unstable without model especially in twist and less affected by 2D-point processing. Repeated left–right swaps in the 2D detections caused discontinuities in the triangulated orientation (leading to π jumps) and still affected EKF-based reconstructions. Between 7.2 and 7.6 s, the EKF variants estimate different twist kinematics.

To further assess $EKF_{2D}$ performance, it will be applied to several series of jumps including either manually labelled or synthetic (ground truth) data and then packaged into an App for automatic judging of trampoline routines.

## 4. Conclusions

Our original extended Kalman filter that relies on multiview 2D-points is promisingly fast with smoother kinematic output.

## Acknowledgements

We acknowledge the support of INS Québec and Gymnastics Canada especially the athletes of Karina Kosko.

## Contributor Roles

MB: Conceptualization, Funding acquisition, Investigation, Methodology, Software, Writing original draft; LDR: Methodology, Data curation, Investigation, Writing – review & editing; BCA: Methodology, Writing – review & editing; EC: Methodology, Software, Writing – review & editing; LS: Funding acquisition, Investigation, Supervision, Writing – review & editing.

## Funding

The project was funded by MITACS and OTP.

## Data, software, code availability

https://github.com/s2mLab/EKF_multicamera

## References

Begon, M., Andersen, M. S., & Dumas, R. (2018). Multibody kinematics optimization for the estimation of upper and lower limb human joint kinematics: a systematized methodological review. *Journal of biomechanical engineering, 140*(3), 030801. https://doi.org/10.1115/1.4038741

Begon, M., & Lacouture, P. (2010). Accuracy of 3-D reconstruction with occlusions. *Journal of applied biomechanics*, 26(1), 104–108. https://doi.org/10.1123/jab.26.1.104

Drolet-Roy, L., Nogues, V., Gaudet, S., Charbonneau, E., Begon, M., & Séoud, L. (2026). *Human pose estimation in trampoline gymnastics: Improving performance using a new synthetic dataset. arXiv*. https://doi.org/10.48550/arXiv.2604.01322

Pagnon, D., Domalain, M., & Reveret, L. (2021). Pose2Sim: an end-to-end workflow for 3D markerless sports kinematics—part 1: robustness. *Sensors, 21*(19), 6530. https://doi.org/10.3390/s21196530

Puchaud, P., Charbonneau, E., Michaud, B., & Begon, M. (2023). Optimality equivalence and computational advantages of free-floating base dynamics compared to full-body dynamics. *Mechanism and Machine Theory*, *181*, 105164. https://doi.org/10.1016/j.mechmachtheory.2022.105164