

# Improving 3D Pose Estimation Accuracy in HighSpeed Sports: A Case Study on Inline Speed Skating

Wataru Kawaguchi<sup>1</sup>, Fumitake Okabe<sup>2</sup>, Junji Yamato<sup>1</sup>  
Kogakuin University<sup>1</sup>, The University of Tokyo<sup>2</sup>

## 1. Background and Objective

- Inline speed skating involves high speeds and close, making marker- or IMU-based motion capture **difficult to use safely**.
- 2D pose estimation **lacks depth** so it cannot reconstruct how the body is tilted in 3D space.
- We compare image-based 3D pose estimation with IMU-based motion capture for Inline speed skating.

## 2. Data Acquisition Environment

- Multi-view video was captured using a 360° camera (Osmo 360) and several monocular cameras (iPhones).
- IMU-based motion capture (Xsens MVN[1]) data served as reference for evaluating the image-based 3D pose estimation (MotionBERT[2]).



Fig1. Data Acquisition Environment and Camera Configuration.

Cam	Device / Lens	Resolution	FPS
1	Osmo 360	6K	60
2	iPhone 15 Pro Max (13 mm)	4K	60
3	iPhone 12 Pro Max (26 mm)	4K	60
4	iPhone 15 Pro Max (24 mm)	4K	60
5	iPhone 15 Pro Max (13 mm)	4K	60
6	iPhone 15 Pro Max (120 mm)	4K	60

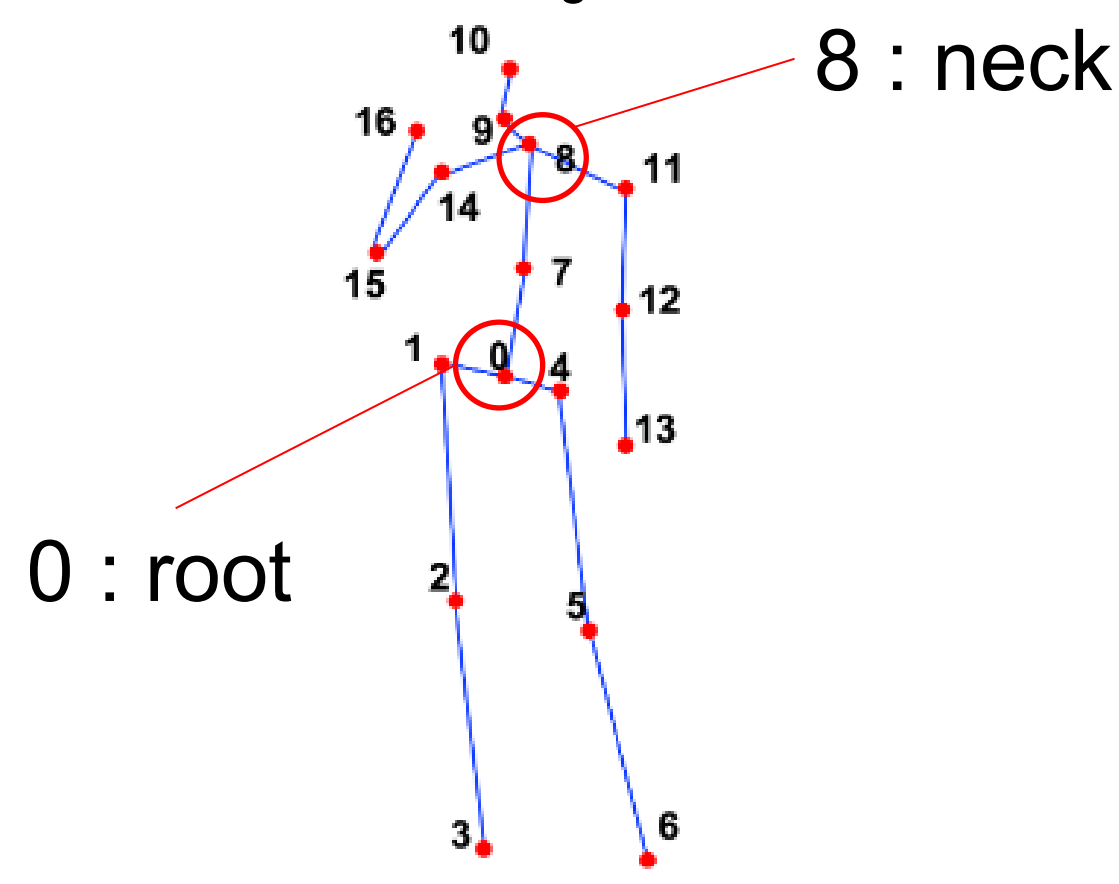


Fig2. Human3.6M  
※The analysis focuses on joint 0 and joint 8.

## 3. Analysis Method

- We aligned the pose data and computed **trunk-lean and bank angles**.
- Estimation accuracy was evaluated using MAE, RMSE, and CCC.

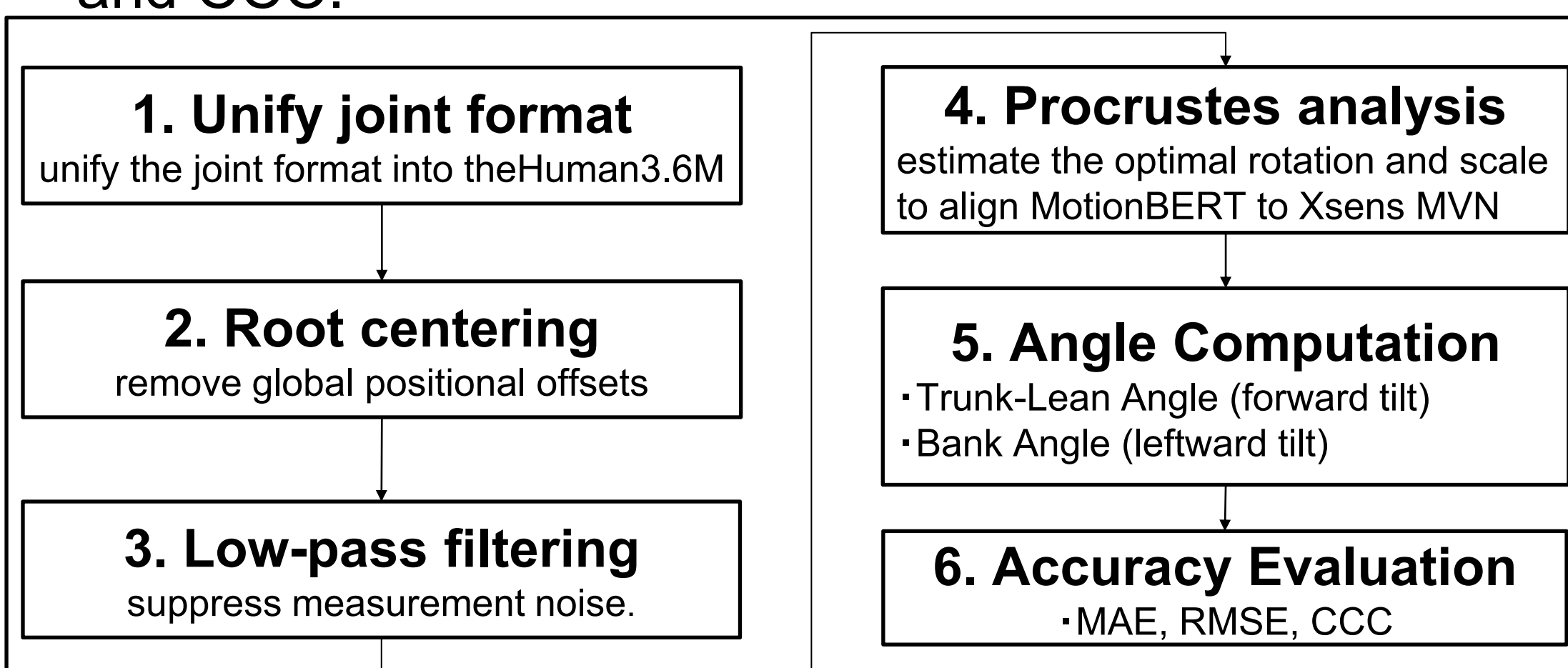


Fig3. Processing flow

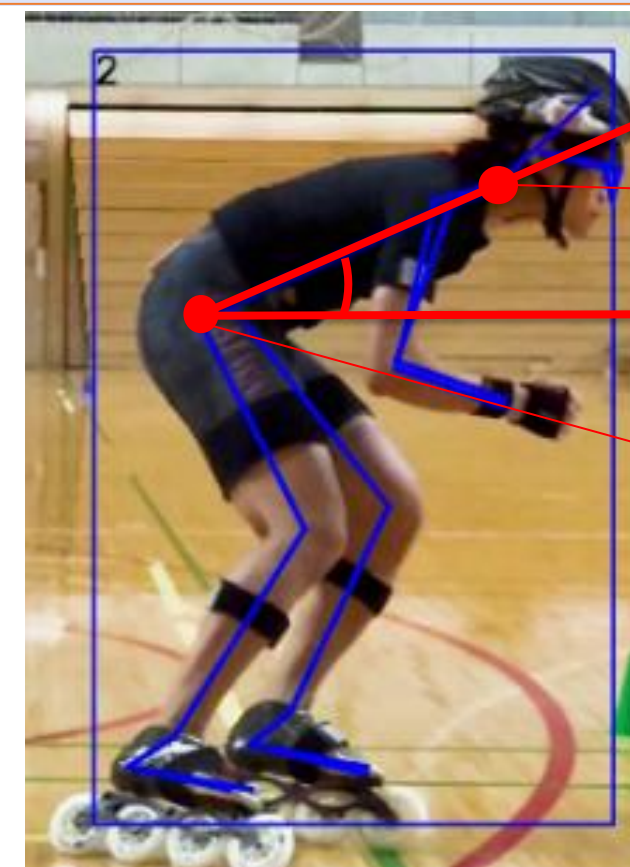


Fig4. Trunk-Lean Angle

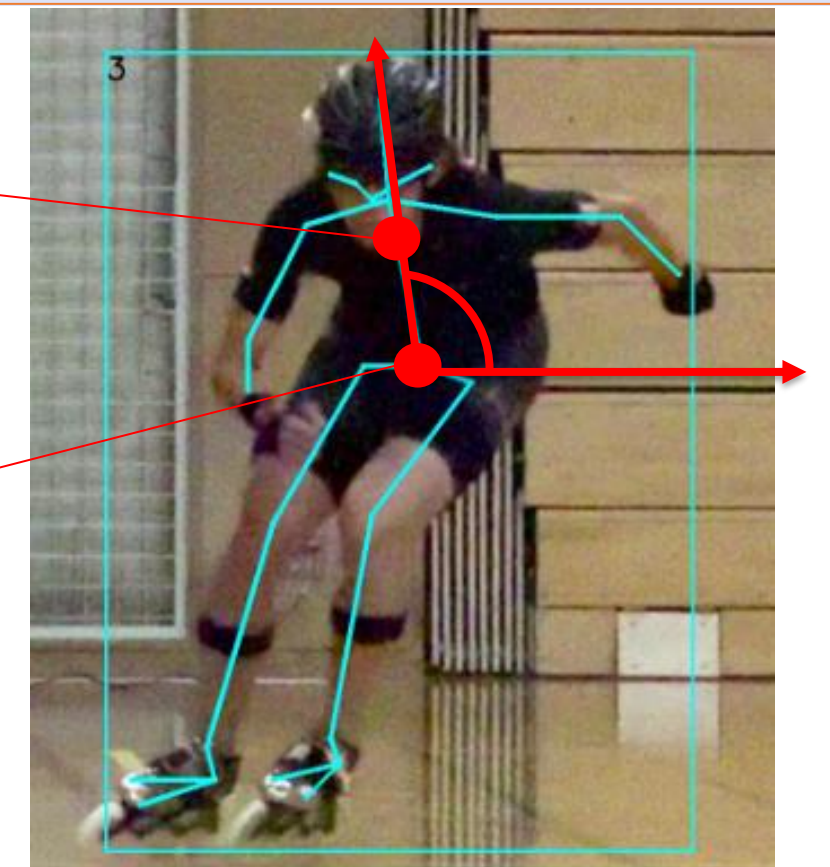


Fig5. Bank Angle

## 4. Results and Discussion

Tab 2. Accuracy Evaluation Results. (Xsens vs. MotionBERT)

evaluation metrics	MAE[°]	RMSE[°]	CCC
Trunk-Lean	4.11	6.50	0.09
Bank Angle	19.6	24.6	0.38

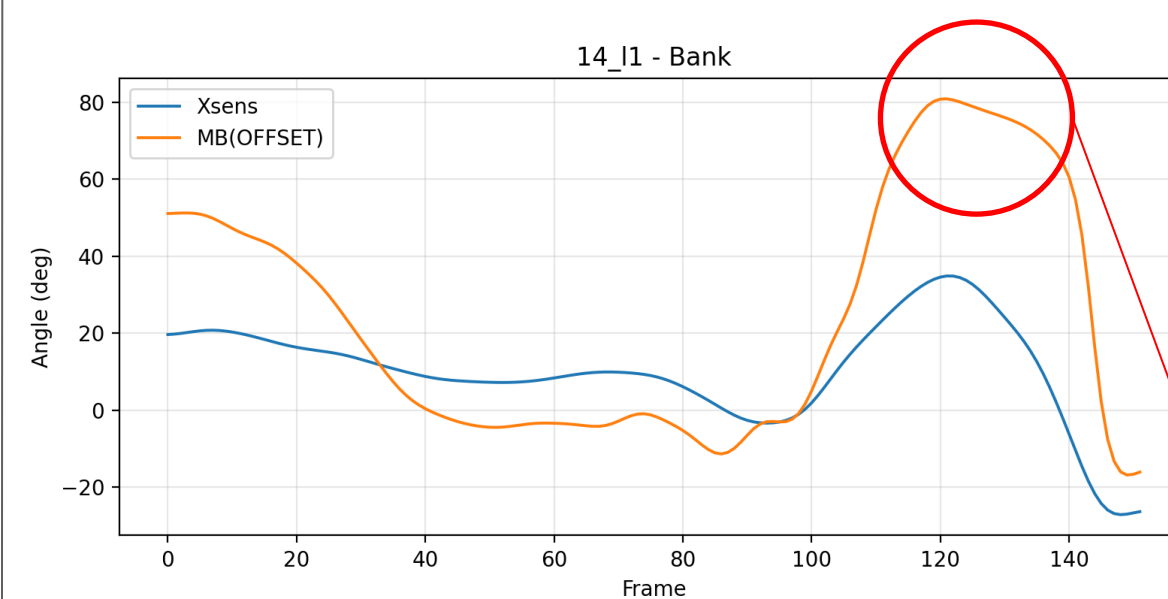


Fig6. Comparison of bank-angle waveforms.

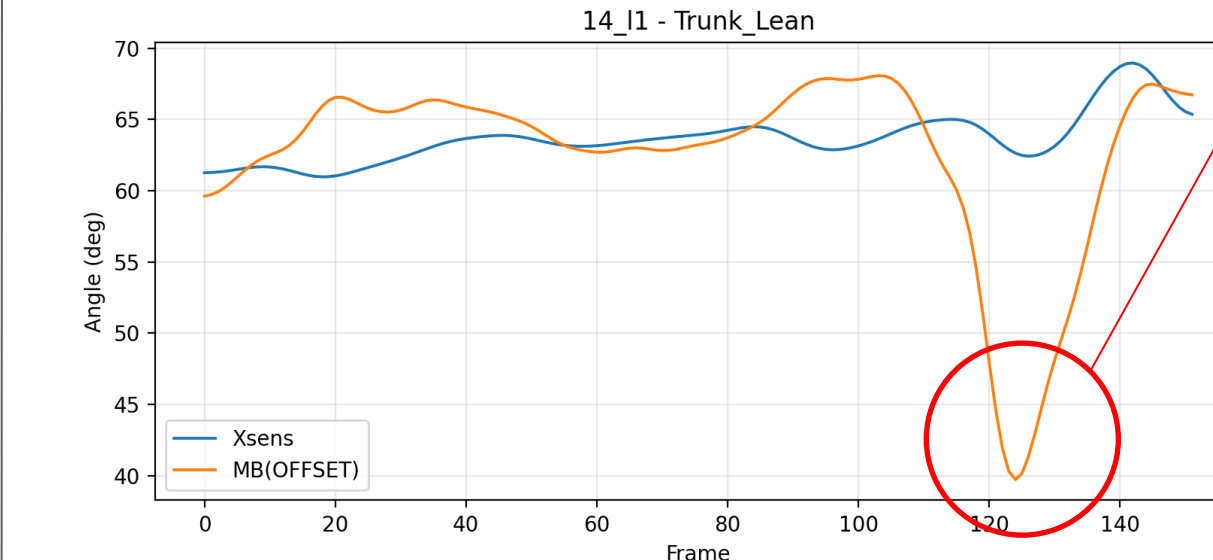


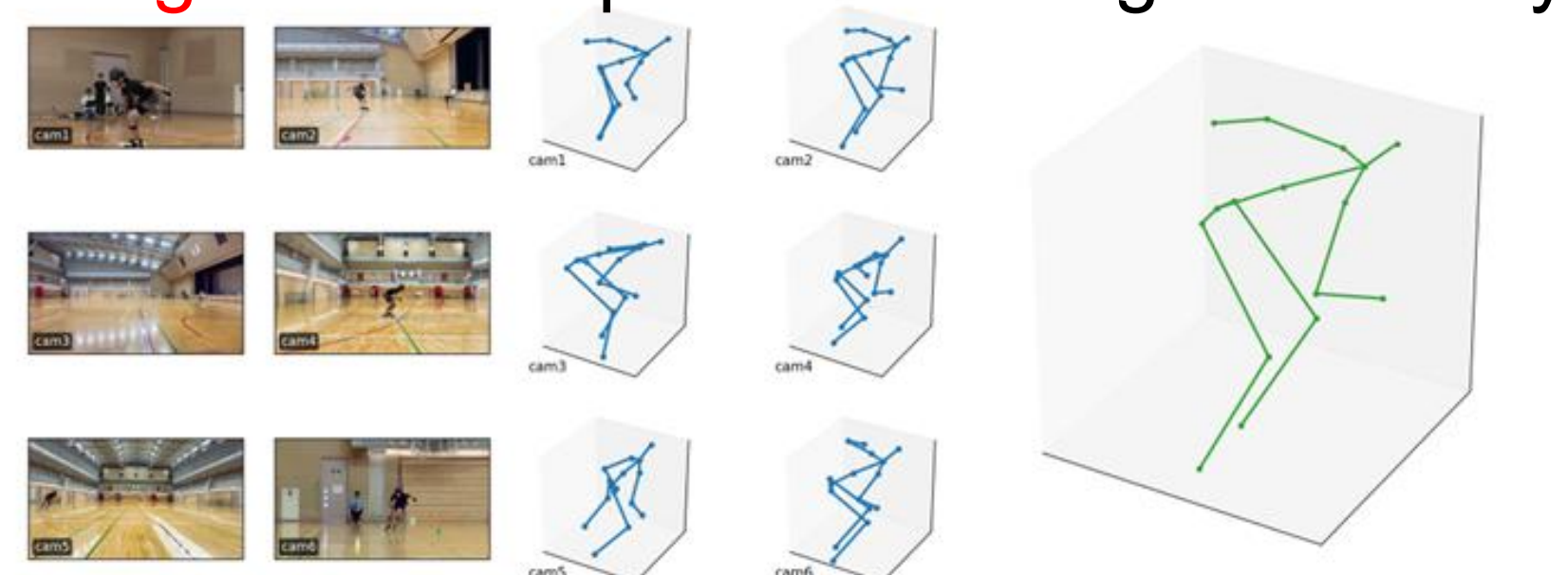
Fig7. Comparison of trunk-lean angle waveforms. Fig8. Frames with large angle errors.



- The trunk-lean angle **showed small errors** because the participant was mainly captured from the left side.
- In contrast, the bank angle had **larger errors** due to its strong dependence on depth information.
- Keypoint detection failed when the left arm covered the torso center, reducing angle accuracy.

## 5. Conclusion and Future Work

- We evaluated trunk-lean and bank angles estimation by comparing image-based 3D pose estimation with IMU-based motion capture.
- Trunk-lean angle showed small errors, whereas bank angle errors were larger.
- Future work includes **IMU-based calibration** and **multi-view integration** to improve bank-angle accuracy.



Input videos Per-camera 3D pose estimation Multi-view fused 3D pose  
Fig 9. Multi-view videos and 3D pose estimations

## References

- [1] M. Schepers, M. Giuberti, and G. Bellusci, "Xsens MVN: Consistent Tracking of Human Motion Using Inertial Sensing," Xsens Technologies B.V., Enschede, The Netherlands, White Paper MV0424P.A, 2018. [Online]. Available: <https://www.xsens.com>
- [2] W. Zhu, X. Ma, Z. Liu, L. Liu, W. Wu, and Y. Wang, "MotionBERT: A Unified Perspective on Learning Human Motion Representations," arXiv preprint arXiv:2210.06551, Aug. 2023. [Online]. Available: <https://motionbert.github.io>