Multi-View 2D to 3D Lifting Optimization-Based Human Pose Estimation with Occluded Joint Prediction

Daniela Rato, Miguel Oliveira, Vítor Santos, Angel Sappa, Bogdan Raducanu

danielarato@ua.pt, mriem@ua.pt, vitor@ua.pt, asappa@cvc.uab.cat, bogdan@cvc.uab.cat

Introduction

Accurate 3D human pose estimation is crucial for enhancing **human-robot collaboration and interaction**. This study introduces a robust optimization-based method designed for video-based 3D human pose estimation, addressing challenges such as 2D joint detection errors, occlusions, and varying camera perspectives.

Objectives

- Propose a multi-camera video-based 3D human pose estimation algorithm.
- Accurately predict the position of occluded 3D joints.
- Compare the performance with other state-of-the-art 3D human pose estimation methods.

Methodology

2D to 3D lifting optimization technique, leveraging information from multiple video frames, integrating movement patterns and **personalized skeletal features,** and using a **nonlinear least squares method**, focusing on minimizing reprojection errors, link length uniformity, and frame-to-frame continuity.

Results

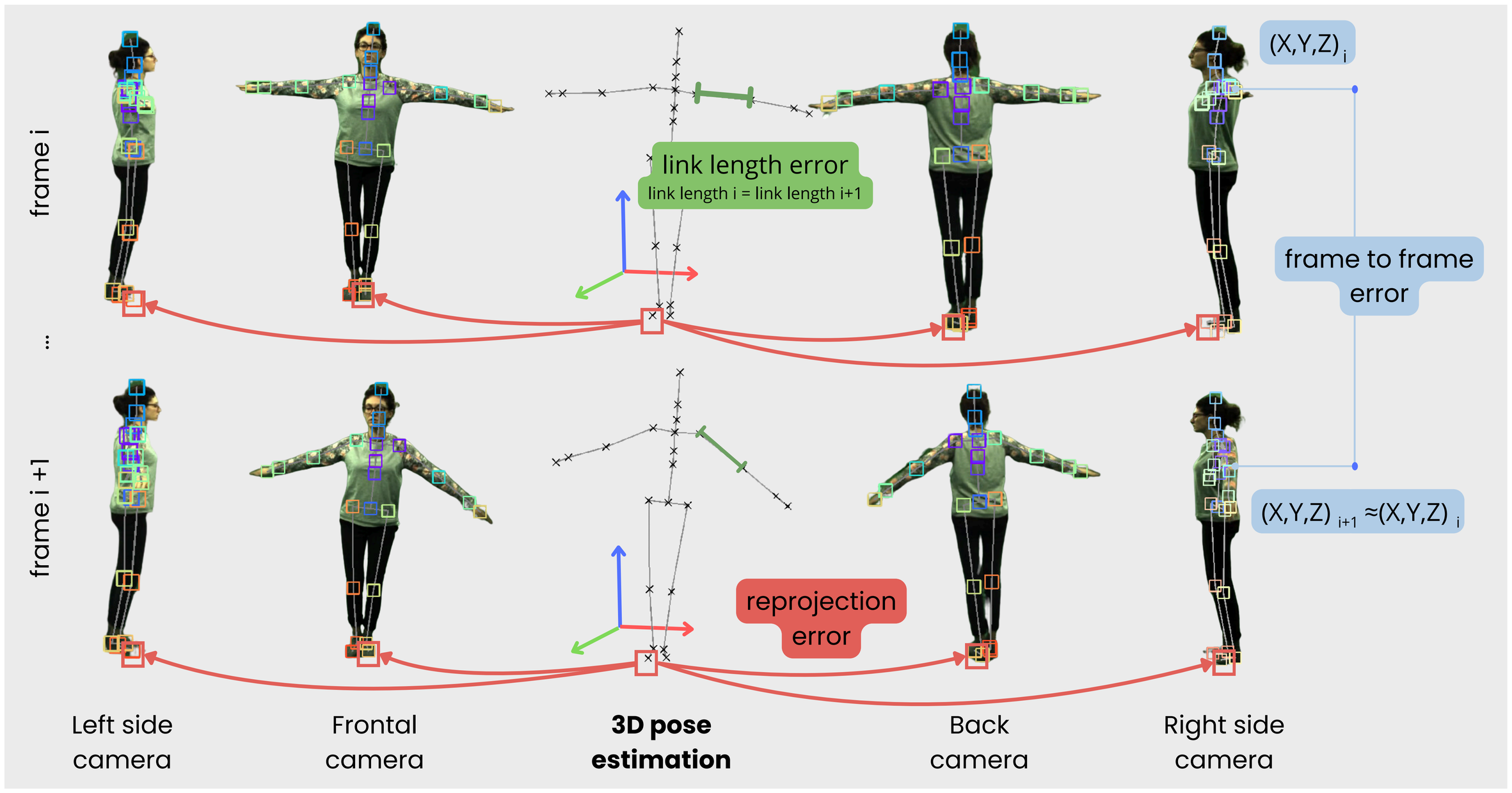
The algorithm was evaluated on the MPI-INF-3DHP dataset, demonstrating robustness to 2D joint errors up to 25 pixels and resilience in scenarios with several occluded joints. Comparative analyses with existing methods showed that our approach performs well in both accuracy and robustness.

Methodology	Optimization	Multi-view	Video	MPJPE ↓
Pavvlo et al.			✓	86.6
Bouazizi et al.		✓		65.9
Jiang et al.	✓			55.2
Zhao et al.			✓	27.8
Yu et al.			✓	27.8
Ours (20px)	✓	✓	✓	36.4
Ours (10px)				18.1

Conclusions

The proposed multi-view video-based optimization method significantly **improves the accuracy of 3D human pose estimation**, especially in the presence of occlusions. This advancement holds substantial potential for enhancing **safety and efficiency in human-robot collaboration**, with applications in diverse fields such as manufacturing, healthcare, and assistive robotics.

Z. Jiang, Z. Zhou, L. Li, W. Chai, C.Y. Yang, and J.N. Hwang, "Back to optimization: Diffusion-based zero-shot 3D human pose estimation," in Proceedings of the IEEE/CVF Winter Conf. on Applications of Computer Vision, 2024.
D. Pavllo, C. Feichtenhofer, D. Grangier, and M. Auli, "3D human pose estimation in video with temporal convolutions and semi-supervised training," in Conf. on Computer Vision and Pattern Recognition, 2019.
A. Bouazizi, U. Kressel, and V. Belagiannis, "Learning temporal 3D human pose estimation with pseudo-labels," in 2021 17th IEEE International Conf. on Advanced Video and Signal Based Surveillance, 2021
B. Yu, Z. Zhang, Y. Liu, S. Zhong, Y. Liu, and C. Chen, "Gla-GNC: Global-local adaptive graph convolutional network for 3D human pose estimation from monocular video," in Proceedings of the IEEE/CVF International Conf. on Computer Vision, 2023
Q. Zhao, C. Zheng, M. Liu, P. Wang, and C. Chen, "PoseFormerV2: Exploring frequency domain for efficient and robust 3d human pose estimation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023



This work was supported by the Foundation for Science and Technology (FCT) under the grant 2021.04792.BD.