Multimodal Pose Estimation with IMU and Camera Fusion vs. Deep Inertial Poser on 3DPW

Assignee Research

June 2, 2026

Abstract

This report synthesises findings from 4 peer-reviewed papers addressing the following research question: How do multimodal pose estimation models integrating IMU and camera data compare to Deep Inertial Poser in terms of reconstruction accuracy (MSE) and robustness to occlusions on the 3DPW outdoor. 4 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Stereo-Inertial Poser: Towards Metric-Accurate Shape-Aware Motion Capture Using Sparse IMUs and a Single Stereo Camera. Research question: How do multimodal pose estimation models integrating IMU and camera data compare to Deep Inertial Poser in terms of reconstruction accuracy (MSE) and robustness to occlusions on the 3DPW outdoor benchmark?.

2 Methodology

Systematic literature search across multiple databases yielded 4 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 5.5/10.

3 Results

4 papers retrieved. 4 claims extracted; 0 independently verified. Quality review score: 5.5/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
Human motions can be estimated from monocular videos but are vulnerable to occlusion, extreme light conditions, and dept	×	0.04
Pure inertial systems suffer from the drifting issue due to sensor noise.	×	0.02
Recent hybrid solutions introduce visual-inertial fusion, combining a single RGB camera with six sparse IMUs.	×	0.13
Visual-inertial fusion systems demonstrate reduced drifting effects and improved robustness against occlusion by decoupl	×	0.13

References

- http://arxiv.org/abs/2003.11163v2
- http://arxiv.org/abs/2603.02130v1
- http://arxiv.org/abs/2411.15127v3