Novelty Assessment Report

Paper: Human3R: Everyone Everywhere All at Once
PDF URL: https://openreview.net/pdf?id=y7duXr0JXF
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2025-12-29

Abstract

We present Human3R, a unified, feed-forward framework for online 4D human-scene reconstruction, in the world coordinate frame, from casually captured monocular videos. Unlike previous approaches that rely on multi-stage pipelines, iterative contact-aware refinement between humans and scenes, and heavy dependencies (i.e., human detection and cropping, tracking, segmentation, camera pose or metric depth estimation, SLAM for 3D scenes, local human mesh recovery, etc.), Human3R jointly recovers global multi-person SMPL-X bodies (“everyone”), dense 3D scene geometry (“everywhere”), and camera trajectories in a single forward pass (“all-at-once”). Our method builds upon the 4D reconstruction foundation model CUT3R, and leverages parameter-efficient visual prompt tuning to preserve its original rich spatiotemporal priors while enabling direct readout of SMPL-X parameters. To further improve the accuracy of global human pose and shape estimation, we introduce a bottom-up (one-shot) multi-person SMPL-X regressor, trained on human-specific datasets. By removing heavy dependencies and iterative refinement, and only training on a relatively small-scale synthetic dataset, BEDLAM, Human3R achieves state-of-the-art performance with remarkable efficiency: it requires just one day of training on a single consumer GPU (NVIDIA RTX 4090) and operates in real time (15 FPS) with a low memory footprint (8 GB). Extensive experiments demonstrate that Human3R delivers state-of-the-art or competitive performance, across all relevant tasks, including global human motion estimation, local human mesh recovery, video depth estimation, and camera pose estimation, with a single unified model. In summary, Human3R achieves one unified model, one-stage inference, one-shot multi-person estimation, and requires just one day of training on one GPU — enabling real-time, online processing of streaming inputs. We hope that Human3R will serve as a simple yet effective baseline, which can be easily extended by other researchers for new applications, such as 6D object pose estimation (“everything”), thereby facilitating future research in this direction. Code and models will be made publicly available.

Core Task Landscape

This paper addresses: **Online 4D Human-Scene Reconstruction from Monocular Video**

A total of **50 papers** were analyzed and organized into a taxonomy with **18 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Human-Centric Reconstruction Methods**
- **Joint Human-Scene Reconstruction**
- **Dynamic Scene Reconstruction**
- **Generative and Diffusion-Based Reconstruction**
- **Specialized Reconstruction Scenarios**

Complete Taxonomy Tree

- Online 4D Human-Scene Reconstruction from Monocular Video Survey Taxonomy
- Human-Centric Reconstruction Methods
  - Template-Based Human Reconstruction
  - Clothing and Deformation Modeling (4 papers)
    - [5] DressRecon: Freeform 4D Human Reconstruction from Monocular Video (Jeff Tan, 2025) View paper
    - [26] Surfel-Based Gaussian Inverse Rendering for Fast and Relightable Dynamic Human Reconstruction From Monocular Videos. (Chenming Wu, 2025) View paper
    - [29] Editable Dynamic Human Scene Reconstruction Using Gaussian Splatting Based on a Skinning Model (Da, 2025) View paper
    - [30] SkinningGS: Editable Dynamic Human Scene Reconstruction Using Gaussian Splatting Based on a Skinning Model (Li Da, 2025) View paper
  - Core Parametric Body Recovery (5 papers)
    - [13] H-NeRF: Neural Radiance Fields for Rendering and Temporal Reconstruction of Humans in Motion (Hongyi Xu, 2021) View paper
    - [42] Monoperfcap: Human performance capture from monocular video (Xu, 2018) View paper
    - [45] Temporal residual neural radiance fields for monocular video dynamic human body reconstruction (Tianle Du, 2024) View paper
    - [46] 4D Facial Avatar Reconstruction From Monocular Video via Efficient and Controllable Neural Radiance Fields (Jeong gi Kwak, 2024) View paper
    - [47] Neural Reconstruction of Relightable Human Model from Monocular Video (Wenzhang Sun, 2023) View paper
  - Non-Parametric Human Reconstruction (3 papers)
  - [17] Single-view RGBD-based reconstruction of dynamic human geometry (Charles Malleson, 2013) View paper
  - [25] Bodyfusion: Real-time capture of human motion and surface geometry using a single depth camera (Tao Yu, 2017) View paper
  - [34] Spatiotemporal Texture Reconstruction for Dynamic Objects Using a Single RGBâ□□D Camera (Hyomin Kim, 2021) View paper
  - Mesh Animation and Rigging (2 papers)

- [3] L4gm: Large 4d gaussian reconstruction model (Ren Jiawei, 2024) View paper
- [6] V2M4: 4D Mesh Animation Reconstruction from a Single Monocular Video (Chen Jian-qi, 2025) View paper

- Joint Human-Scene Reconstruction
  - Optimization-Based Joint Reconstruction (4 papers)
  - [4] HSR: holistic 3d human-scene reconstruction from monocular videos (Lixin Xue, 2024) View paper
  - [28] Joint Optimization for 4D Human-Scene Reconstruction in the Wild (Liu Zhi-zheng, 2025) View paper
  - [38] Prior-based 4D Human-Scene Reconstruction from Monocular Videos (Scene, n.d.) View paper
  - [39] Learning motion priors for 4d human body capture in 3d scenes (Zhang Siwei, 2021) View paper
  - Feed-Forward Joint Reconstruction ★ (3 papers)
  - [0] Human3R: Everyone Everywhere All at Once (Anon et al., 2026) View paper
  - [15] ODHSR: Online Dense 3D Reconstruction of Humans and Scenes from Monocular Videos (Zetong Zhang, 2025) View paper
  - [22] Synergistic global-space camera and human reconstruction from videos (Yizhou Zhao, 2024) View paper
  - Human-Scene Interaction Modeling (5 papers)
  - [19] Visual Imitation Enables Contextual Humanoid Control (Allshire, 2025) View paper
  - [37] SHARE: Scene-Human Aligned Reconstruction (Joshua Li, 2025) View paper
  - [40] CARI4D: Category Agnostic 4D Reconstruction of Human-Object Interaction (Xianghui Xie, 2025) View paper
  - [48] CRISP: Contact-guided Real2Sim from Monocular Video with Planar Scene Primitives (Zihan Wang, 2025) View paper
  - [49] ZeroHSI: Zero-Shot 4D Human-Scene Interaction by Video Generation (Li Hongjie, 2024) View paper
- Dynamic Scene Reconstruction
  - Gaussian Splatting-Based Methods
  - Deformable Gaussian Representations (4 papers)
    - [7] Deformable 3D Gaussians for High-Fidelity Monocular Dynamic Scene Reconstruction (Ziyi Yang, 2023) View paper
    - [31] SplineGS: Robust Motion-Adaptive Spline for Real-Time Dynamic 3D Gaussians from Monocular Video (Jongmin Park, 2024) View paper
    - [44] DGS-LRM: Real-Time Deformable 3D Gaussian Reconstruction From Monocular Videos (Lin, 2025) View paper
    - [50] Real-time Gaussian Splatting for Dynamic Reconstruction in Stationary Monocular Cameras (Minyu Chen, 2024) View paper
  - Specialized Gaussian Applications (2 papers)
    - [8] Endo-4DGS: Endoscopic Monocular Scene Reconstruction with 4D Gaussian Splatting (Huang, 2024) View paper
    - [41] Deblur4DGS: 4D Gaussian Splatting from Blurry Monocular Video (Wu Renlong, 2024) View paper
  - Neural Radiance Field Methods (2 papers)
  - [35] Articulatedfusion: Real-time reconstruction of motion, geometry and segmentation using a single depth camera (Chao Li, 2018) View paper
  - [36] DRSM: efficient neural 4d decomposition for dynamic reconstruction in stationary monocular cameras (Weixing Xie, 2024) View paper
  - Motion and Trajectory Modeling (3 papers)
  - [1] Shape of motion: 4d reconstruction from a single video (Qianqian Wang, 2025) View paper
  - [9] Robot see robot do: Imitating articulated object manipulation with monocular 4d reconstruction (Kerr, 2024) View paper
  - [24] GFlow: Recovering 4D World from Monocular Video (Wang, 2025) View paper
- Generative and Diffusion-Based Reconstruction
  - Video Diffusion Priors (4 papers)
  - [11] ViDAR: Video Diffusion-Aware 4D Reconstruction From Monocular Inputs (Nazarczuk, 2025) View paper
  - [12] Vivid4D: Improving 4D Reconstruction from Monocular Video by Video Inpainting (Huang Jia-xin, 2025) View paper
  - [18] CAT4D: Create Anything in 4D with Multi-View Video Diffusion Models (Rundi Wu, 2024) View paper
  - [20] Geo4d: Leveraging video generators for geometric 4d scene reconstruction (Jiang, 2025) View paper
  - Scene Generation and Synthesis (2 papers)
  - [2] Dreamscene4d: Dynamic multi-object scene generation from monocular videos (Wen-Hsuan Chu, 2024) View paper
  - [27] Feed-forward bullet-time reconstruction of dynamic scenes from monocular videos (Liang, 2024) View paper
  - Foundation Model Integration (3 papers)
  - [14] Uni4D: Unifying Visual Foundation Models for 4D Modeling from a Single Video (Albert J. Zhai, 2025) View paper
  - [23] Feature4x: Bridging any monocular video to 4d agentic ai with versatile gaussian feature fields (Shijie Zhou, 2025) View paper
  - [33] 4DGT: Learning a 4D Gaussian Transformer Using Real-World Monocular Videos (Xu Zhen, 2025) View paper
- Specialized Reconstruction Scenarios
  - Egocentric and First-Person Reconstruction (1 papers)
  - [16] Self-Supervised Monocular 4D Scene Reconstruction for Egocentric Videos (Yuan, 2025) View paper
  - Stationary Camera Reconstruction (1 papers)
  - [32] Building 4D Models of Objects and Scenes from Monocular Videos (Yang, 2023) View paper
  - Fast and Real-Time Methods (1 papers)
  - [10] 4D-Fly: Fast 4D Reconstruction from a Single Monocular Video (Diankun Wu, 2025) View paper
  - Auxiliary Task Applications (2 papers)
  - [21] GeoRecon: Geometric Coherence for Online 3D Scene Reconstruction From Monocular Video (Yanmei Wang, 2024) View paper
  - [43] Unsupervised Monocular Road Segmentation for Autonomous Driving via Scene Geometry (Nasihatkon, 2025) View paper

Narrative

Core task: online 4D human-scene reconstruction from monocular video. This field aims to recover dynamic 3D geometry and motion of both humans and their surrounding environments from single-camera footage, often in real-time or near-real-time settings. The taxonomy reveals several complementary research directions. Human-Centric Reconstruction Methods focus primarily on capturing detailed human body shape and motion, often leveraging parametric models or learned priors. Joint Human-Scene Reconstruction tackles the coupled problem of simultaneously modeling people and their environments, addressing challenges like occlusion handling and consistent spatial alignment. Dynamic Scene Reconstruction emphasizes general non-rigid or articulated scene motion without necessarily privileging human subjects, while Generative and Diffusion-Based Reconstruction explores synthesis-driven approaches that can hallucinate plausible geometry from limited observations. Specialized Reconstruction Scenarios address domain-specific constraints such as endoscopic imaging, robotic manipulation, or aerial capture.

Within Joint Human-Scene Reconstruction, a particularly active line of work explores feed-forward architectures that predict 4D representations in a single pass, balancing speed and fidelity. Human3R[0] exemplifies this feed-forward joint reconstruction approach,

aiming for efficient inference without iterative optimization. Nearby methods like ODHSR[15] and Synergistic Global-Space[22] similarly pursue real-time or online processing but may differ in their scene representation choices—some favor Gaussian splatting primitives while others use neural radiance fields or hybrid schemes. Compared to optimization-heavy pipelines such as HSR[4] or DressRecon[5], which refine geometry over many frames, Human3R[0] prioritizes immediacy and generalization across diverse scenes. This trade-off between reconstruction quality and computational efficiency remains a central open question, with recent works exploring how much geometric detail can be recovered from a single forward pass versus how much benefit iterative refinement truly provides in dynamic human-scene settings.

Related Works in Same Category

The following **2 sibling papers** share the same taxonomy leaf node with the original paper:

1. ODHSR: Online Dense 3D Reconstruction of Humans and Scenes from Monocular Videos

Authors: Zetong Zhang, Manuel Kaufmann, Lixin Xue, Jie Song, Martin R. Oswald | **Year/Venue**: 2025 | **URL**: View paper

Abstract

Creating a photorealistic scene and human reconstruction from a single monocular in-the-wild video figures prominently in the perception of a human-centric 3D world. Recent neural rendering advances have enabled holistic human-scene reconstruction but require pre-calibrated camera and human poses, and days of training time. In this work, we introduce a novel unified framework that simultaneously performs camera tracking, human pose estimation and human-scene reconstruction in an online fashion. ...

Relationship Analysis

Both papers belong to the Feed-Forward Joint Reconstruction category, performing online joint human-scene reconstruction from monocular video in a single-pass manner without iterative refinement. They overlap in jointly estimating global human meshes, dense scene geometry, and camera poses in real-time from RGB video streams. The key difference is that Human3R builds upon CUT3R with visual prompt tuning for direct SMPL-X readout and achieves 15 FPS with minimal dependencies, while ODHSR uses 3D Gaussian Splatting within a SLAM framework with occlusion-aware silhouette rendering and requires human segmentation preprocessing, achieving 85 FPS rendering but with different architectural foundations.

2. Synergistic global-space camera and human reconstruction from videos

Authors: Yizhou Zhao, Bhiksha Raj, Tuanfeng Y. Wang, Min Xu, Jimei Yang, et al. (6 authors total) | **Year/Venue**: 2024 | **URL**: View paper

Abstract

Remarkable strides have been made in reconstructing static scenes or human bodies from monocular videos. Yet, the two problems have largely been approached independently, without much synergy. Most visual SLAM methods can only reconstruct camera trajectories and scene structures up to scale, while most HMR methods reconstruct human meshes in metric scale but fall short in reasoning with cameras and scenes. This work introduces Synergistic Camera and Human Reconstruction (SynCHMR) to marry the be...

Relationship Analysis

Both papers belong to the Feed-Forward Joint Reconstruction category, performing online human-scene reconstruction from monocular video without iterative refinement. They overlap in jointly estimating global human meshes, camera poses, and scene geometry in a unified framework. However, Human3R uses visual prompt tuning on CUT3R with a bottom-up multi-person approach for real-time inference (15 FPS), while SynCHMR employs a two-phase pipeline that calibrates SLAM with camera-frame HMR priors and applies scene-aware denoising, focusing on synergistic integration of HMR and SLAM components.

Contributions Analysis

Overall novelty summary. Human3R proposes a unified feed-forward framework for online 4D human-scene reconstruction from monocular video, jointly recovering multi-person SMPL-X bodies, dense scene geometry, and camera trajectories in a single forward pass. The paper sits within the 'Feed-Forward Joint Reconstruction' leaf of the taxonomy, which contains only three papers total. This represents a relatively sparse research direction compared to more crowded areas like Template-Based Human Reconstruction (nine papers) or Gaussian Splatting-Based Methods (six papers), suggesting the feed-forward joint reconstruction paradigm remains an emerging approach rather than a saturated field.

The taxonomy structure reveals that Human3R's closest neighbors are optimization-based joint reconstruction methods (four papers) and human-scene interaction modeling approaches (five papers). While optimization-based methods like HSR and DressRecon emphasize iterative refinement for quality, Human3R diverges by prioritizing single-pass efficiency. The broader Joint Human-Scene Reconstruction branch (twelve papers total) sits between purely human-centric methods (sixteen papers across multiple leaves) and general dynamic scene reconstruction (thirteen papers), positioning Human3R at the intersection of human-specific modeling and holistic scene understanding. The taxonomy's scope notes clarify that feed-forward methods explicitly exclude iterative optimization, distinguishing Human3R's architectural philosophy from refinement-heavy alternatives.

Among the nineteen candidates examined across three contributions, no clearly refuting prior work was identified. The unified feed-forward framework contribution examined nine candidates with zero refutations, suggesting limited direct overlap in the constrained search scope. The parameter-efficient visual prompt tuning method examined only one candidate without refutation, indicating either genuine novelty or insufficient search coverage in this specific technical dimension. The real-time multi-person reconstruction contribution also examined nine candidates with no refutations. These statistics reflect a top-K semantic search rather than exhaustive coverage, meaning the absence of refutations indicates no obvious overlaps within the limited candidate pool examined, not definitive novelty across all prior work.

Based on the limited search scope of nineteen candidates, Human3R appears to occupy a relatively underexplored position within feed-forward joint reconstruction, though the small candidate pool prevents strong conclusions about absolute novelty. The sparse population of its taxonomy leaf and absence of refutations among examined papers suggest the specific combination of feed-forward architecture, joint human-scene modeling, and SMPL-X parameter readout may represent a less-traveled path. However, the analysis does not cover exhaustive literature in related areas like optimization-based methods or human-centric reconstruction, where overlapping ideas might exist outside the semantic search radius.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: Human3R unified feed-forward framework for online 4D human-scene reconstruction

Description: Human3R is a unified model that jointly recovers global multi-person SMPL-X bodies, dense 3D scene geometry, and camera trajectories from monocular video in a single forward pass, eliminating multi-stage pipelines and heavy dependencies such as human detection, depth estimation, and SLAM preprocessing.

This contribution was assessed against **9 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. 4DGT: Learning a 4D Gaussian Transformer Using Real-World Monocular Videos

URL: View paper

Brief Assessment

4DGT[33] focuses on general dynamic scene reconstruction from monocular videos using 4D Gaussian representations, not specifically on human-scene reconstruction with SMPL-X body recovery. The technical approaches and problem domains differ fundamentally.

---

2. Dynamic neural radiance fields for monocular 4d facial avatar reconstruction

URL: View paper

Brief Assessment

Dynamic Neural Radiance[53] focuses on monocular 4D facial avatar reconstruction with neural radiance fields for faces, not general human-scene reconstruction with SMPL-X bodies and camera trajectories in a unified feed-forward framework.

---

3. MoSca: Dynamic Gaussian Fusion from Casual Videos via 4D Motion Scaffolds

URL: View paper

Brief Assessment

MoSca[52] focuses on general dynamic scene reconstruction from monocular videos using motion scaffolds and Gaussian splatting, without explicit human mesh recovery or SMPL-X body modeling. The candidate does not address multi-person reconstruction or human-specific parametric models.

---

4. MoVieS: Motion-Aware 4D Dynamic View Synthesis in One Second

URL: View paper

Brief Assessment

MoVieS[51] focuses on general dynamic scene synthesis from monocular video without explicit human body modeling (SMPL-X), whereas Human3R specifically targets joint reconstruction of parametric human meshes, dense scenes, and camera trajectories with human-specific components.

---

5. Tensor4D: Efficient Neural 4D Decomposition for High-Fidelity Dynamic Reconstruction and Rendering

URL: View paper

Brief Assessment

Tensor4D[55] focuses on efficient 4D tensor decomposition for dynamic scene rendering from multi-view cameras, not on unified feed-forward monocular human-scene reconstruction with SMPL-X body recovery and camera trajectory estimation.

---

6. CAT4D: Create Anything in 4D with Multi-View Video Diffusion Models

URL: View paper

Brief Assessment

CAT4D[18] focuses on creating 4D scenes from monocular video using multi-view video diffusion models for novel view synthesis, not on unified feed-forward human-scene reconstruction with SMPL-X body recovery and camera trajectory estimation in a single forward pass.

---

7. L4gm: Large 4d gaussian reconstruction model

URL: View paper

Brief Assessment

L4gm[3] focuses on reconstructing animated 3D objects from monocular videos using 4D Gaussian representations, not human-scene reconstruction. The technical domains are distinct: L4gm targets general object animation while Human3R specifically addresses multi-person SMPL-X body recovery with scene context.

---

8. Feature4x: Bridging any monocular video to 4d agentic ai with versatile gaussian feature fields

URL: View paper

Brief Assessment

Feature4x[23] focuses on general 4D scene reconstruction and feature field distillation from monocular video, not specifically on human-scene reconstruction with SMPL-X body recovery. The technical approaches differ fundamentally in their objectives and methods.

---

9. 4dnex: Feed-forward 4d generative modeling made easy

URL: View paper

Brief Assessment

4dnex[54] focuses on general 4D scene generation from single images using video diffusion models, not specifically on human-scene reconstruction from monocular video streams. The technical approaches and problem formulations differ fundamentally.

Contribution 2: Parameter-efficient visual prompt tuning method for human reconstruction

Description: The authors introduce a parameter-efficient finetuning approach that uses visual prompt tuning on CUT3R, detecting human head tokens and transforming them into human prompts via learnable projection layers, while keeping the CUT3R backbone frozen to preserve its spatiotemporal priors.

This contribution was assessed against **1 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. PromptHMR: Promptable Human Mesh Recovery

URL: View paper

Brief Assessment

PromptHMR[56] uses spatial and semantic prompts (bounding boxes, masks, language) for human pose estimation, not visual prompt tuning on a 4D reconstruction foundation model. The original paper's approach of detecting human head tokens and transforming them via learnable projection layers while freezing the CUT3R backbone represents a distinct technical contribution focused on preserving spatiotemporal priors from a 4D reconstruction model.

Contribution 3: Real-time one-shot multi-person reconstruction with minimal training

Description: Human3R achieves efficient training and inference by requiring only one day of training on a single GPU using the BEDLAM dataset, while enabling real-time reconstruction at 15 FPS with low memory usage and supporting bottom-up multi-person reconstruction in a single forward pass.

This contribution was assessed against **9 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. XNect: Real-time Multi-Person 3D Motion Capture with a Single RGB Camera

URL: View paper

Brief Assessment

XNect[62] focuses on real-time multi-person 3D pose estimation from RGB cameras but does not address the minimal training aspect (one day on single GPU) or bottom-up reconstruction in a single forward pass that Human3R emphasizes. XNect uses a multi-stage pipeline rather than the unified one-shot approach.

---

2. Real-Time 3D Multi-Person Pose Estimation Using an Omnidirectional Camera and mmWave Radars

URL: View paper

Brief Assessment

Omnidirectional mmWave[66] focuses on 3D pose estimation using omnidirectional cameras and mmWave radars for outdoor/indoor scenarios, not on one-shot multi-person mesh reconstruction with minimal training requirements. The technical approaches and problem domains are fundamentally different.

---

3. Multi-Person 3D Pose Estimation in Mobile Edge Computing Devices for Real-Time Applications

URL: View paper

Brief Assessment

Mobile Edge 3D[65] focuses on 2D/3D pose estimation for mobile edge devices using depthwise separable convolutions, not on unified 4D human-scene reconstruction with SMPL-X meshes, camera trajectories, and dense scene geometry as in the original paper.

---

4. Exploring Novel Methods for Real-Time Multi-Camera People Tracking in Machine Learning

URL: View paper

Brief Assessment

Multi-Camera People Tracking[63] focuses on multi-camera tracking systems for people detection and tracking, not on 3D human mesh reconstruction or scene reconstruction from monocular video. The candidate's emphasis on tracking across camera views differs fundamentally from Human3R's unified 4D reconstruction framework.

---

5. Real-time omnidirectional 3D multi-person human pose estimation with occlusion handling

Brief Assessment

Omnidirectional Multi-Person[61] focuses on multi-person 3D pose estimation using radar sensing and 2D-3D lifting models, not on one-shot reconstruction with minimal training requirements. The candidate does not address training efficiency or the one-day training paradigm claimed by the original paper.

---

6. Multi-hmr: Multi-person whole-body human mesh recovery in a single shot

URL: View paper

Brief Assessment

Multi-hmr[58] focuses on multi-person whole-body mesh recovery from single RGB images without scene reconstruction or camera trajectory estimation, which are core components of the original paper's contribution. The candidate does not address 4D human-scene reconstruction or online video processing.

---

7. Few-Shot Multi-Human Neural Rendering Using Geometry Constraints

URL: View paper

Brief Assessment

Few-Shot Multi-Human[59] focuses on multi-view neural rendering from sparse static images (5-20 views), not real-time online reconstruction from monocular video streams. The candidate requires multi-view camera setups and does not address the one-day training efficiency or 15 FPS real-time performance claims.

---

8. Light3DPose: Real-time Multi-Person 3D Pose Estimation from Multiple Views

URL: View paper

Brief Assessment

Light3DPose[64] focuses on multi-view 3D pose estimation from calibrated cameras, not monocular video reconstruction. The candidate requires multiple synchronized camera views and does not address the minimal training or single-day GPU training aspects that characterize the original contribution.

---

9. 3D real-time human reconstruction with a single RGBD camera

URL: View paper

Brief Assessment

3D Real-time RGBD[60] focuses on single-person reconstruction using RGBD cameras with a parametric model approach, not multi-person reconstruction from monocular RGB videos with minimal training as in the original paper.

Appendix: Text Similarity Detection

Textual similarity detection checked 23 papers and found 2 similarity segment(s) across 1 paper(s).

The following **1 paper(s)** were detected to have high textual similarity with the original paper. These may represent different versions of the same work, duplicate submissions, or papers with substantial textual overlap. Readers are advised to verify these relationships independently.

1. Multi-hmr: Multi-person whole-body human mesh recovery in a single shot

Detected in: Contribution: contribution_3

⚠ **Note**: This paper shows substantial textual similarity with the original paper. It may be a different version, a duplicate submission, or contain significant overlapping content. Please review carefully to determine the nature of the relationship.

References

- [0] Human3R: Everyone Everywhere All at Once View paper
- [1] Shape of motion: 4d reconstruction from a single video View paper
- [2] Dreamscene4d: Dynamic multi-object scene generation from monocular videos View paper
- [3] L4gm: Large 4d gaussian reconstruction model View paper

- [4] HSR: holistic 3d human-scene reconstruction from monocular videos View paper
- [5] DressRecon: Freeform 4D Human Reconstruction from Monocular Video View paper
- [6] V2M4: 4D Mesh Animation Reconstruction from a Single Monocular Video View paper
- [7] Deformable 3D Gaussians for High-Fidelity Monocular Dynamic Scene Reconstruction View paper
- [8] Endo-4DGS: Endoscopic Monocular Scene Reconstruction with 4D Gaussian Splatting View paper
- [9] Robot see robot do: Imitating articulated object manipulation with monocular 4d reconstruction View paper
- [10] 4D-Fly: Fast 4D Reconstruction from a Single Monocular Video View paper
- [11] ViDAR: Video Diffusion-Aware 4D Reconstruction From Monocular Inputs View paper
- [12] Vivid4D: Improving 4D Reconstruction from Monocular Video by Video Inpainting View paper
- [13] H-NeRF: Neural Radiance Fields for Rendering and Temporal Reconstruction of Humans in Motion View paper
- [14] Uni4D: Unifying Visual Foundation Models for 4D Modeling from a Single Video View paper
- [15] ODHSR: Online Dense 3D Reconstruction of Humans and Scenes from Monocular Videos View paper
- [16] Self-Supervised Monocular 4D Scene Reconstruction for Egocentric Videos View paper
- [17] Single-view RGBD-based reconstruction of dynamic human geometry View paper
- [18] CAT4D: Create Anything in 4D with Multi-View Video Diffusion Models View paper
- [19] Visual Imitation Enables Contextual Humanoid Control View paper
- [20] Geo4d: Leveraging video generators for geometric 4d scene reconstruction View paper
- [21] GeoRecon: Geometric Coherence for Online 3D Scene Reconstruction From Monocular Video View paper
- [22] Synergistic global-space camera and human reconstruction from videos View paper
- [23] Feature4x: Bridging any monocular video to 4d agentic ai with versatile gaussian feature fields View paper
- [24] GFlow: Recovering 4D World from Monocular Video View paper
- [25] Bodyfusion: Real-time capture of human motion and surface geometry using a single depth camera View paper
- [26] Surfel-Based Gaussian Inverse Rendering for Fast and Relightable Dynamic Human Reconstruction From Monocular Videos. View paper
- [27] Feed-forward bullet-time reconstruction of dynamic scenes from monocular videos View paper
- [28] Joint Optimization for 4D Human-Scene Reconstruction in the Wild View paper
- [29] Editable Dynamic Human Scene Reconstruction Using Gaussian Splatting Based on a Skinning Model View paper
- [30] SkinningGS: Editable Dynamic Human Scene Reconstruction Using Gaussian Splatting Based on a Skinning Model View paper
- [31] SplineGS: Robust Motion-Adaptive Spline for Real-Time Dynamic 3D Gaussians from Monocular Video View paper
- [32] Building 4D Models of Objects and Scenes from Monocular Videos View paper
- [33] 4DGT: Learning a 4D Gaussian Transformer Using Real-World Monocular Videos View paper
- [34] Spatiotemporal Texture Reconstruction for Dynamic Objects Using a Single RGBâ□□D Camera View paper
- [35] Articulatedfusion: Real-time reconstruction of motion, geometry and segmentation using a single depth camera View paper
- [36] DRSM: efficient neural 4d decomposition for dynamic reconstruction in stationary monocular cameras View paper
- [37] SHARE: Scene-Human Aligned Reconstruction View paper
- [38] Prior-based 4D Human-Scene Reconstruction from Monocular Videos View paper
- [39] Learning motion priors for 4d human body capture in 3d scenes View paper
- [40] CARI4D: Category Agnostic 4D Reconstruction of Human-Object Interaction View paper
- [41] Deblur4DGS: 4D Gaussian Splatting from Blurry Monocular Video View paper
- [42] Monoperfcap: Human performance capture from monocular video View paper
- [43] Unsupervised Monocular Road Segmentation for Autonomous Driving via Scene Geometry View paper
- [44] DGS-LRM: Real-Time Deformable 3D Gaussian Reconstruction From Monocular Videos View paper
- [45] Temporal residual neural radiance fields for monocular video dynamic human body reconstruction View paper
- [46] 4D Facial Avatar Reconstruction From Monocular Video via Efficient and Controllable Neural Radiance Fields View paper
- [47] Neural Reconstruction of Relightable Human Model from Monocular Video View paper
- [48] CRISP: Contact-guided Real2Sim from Monocular Video with Planar Scene Primitives View paper
- [49] ZeroHSI: Zero-Shot 4D Human-Scene Interaction by Video Generation View paper
- [50] Real-time Gaussian Splatting for Dynamic Reconstruction in Stationary Monocular Cameras View paper
- [51] MoVieS: Motion-Aware 4D Dynamic View Synthesis in One Second View paper
- [52] MoSca: Dynamic Gaussian Fusion from Casual Videos via 4D Motion Scaffolds View paper
- [53] Dynamic neural radiance fields for monocular 4d facial avatar reconstruction View paper
- [54] 4dnex: Feed-forward 4d generative modeling made easy View paper
- [55] Tensor4D: Efficient Neural 4D Decomposition for High-Fidelity Dynamic Reconstruction and Rendering View paper
- [56] PromptHMR: Promptable Human Mesh Recovery View paper
- [57] TranSG: Transformer-based skeleton graph prototype contrastive learning with structure-trajectory prompted reconstruction for person re-identification View paper
- [58] Multi-hmr: Multi-person whole-body human mesh recovery in a single shot View paper
- [59] Few-Shot Multi-Human Neural Rendering Using Geometry Constraints View paper
- [60] 3D real-time human reconstruction with a single RGBD camera View paper
- [61] Real-time omnidirectional 3D multi-person human pose estimation with occlusion handling
- [62] XNect: Real-time Multi-Person 3D Motion Capture with a Single RGB Camera View paper
- [63] Exploring Novel Methods for Real-Time Multi-Camera People Tracking in Machine Learning View paper
- [64] Light3DPose: Real-time Multi-Person 3D Pose Estimation from Multiple Views View paper
- [65] Multi-Person 3D Pose Estimation in Mobile Edge Computing Devices for Real-Time Applications View paper
- [66] Real-Time 3D Multi-Person Pose Estimation Using an Omnidirectional Camera and mmWave Radars View paper