Automated 3D Scene Generation from 2D Visual Inputs Using OpenUSD in NVIDIA Omniverse

Davit Khanjaladze

e-mail: davit.khanjaladze0639@ens.tsu.edu.ge

Department of Computer Science, Faculty of Exact and Natural Sciences, Ivane Javakhishvili Tbilisi State University.

University street 13, Tbilisi, Georgia

Annotation

This project explores the transformation of monocular 2D video sequences into coherent 3D representations by leveraging modern depth estimation techniques and 3D reconstruction pipelines. Traditional 3D scene capture requires complex multi-view setups or specialized sensors like LiDAR. However, recent advancements in computer vision and deep learning have made it feasible to extract depth information from a single camera input.

The goal of this research is to develop a full pipeline that:

- Extracts video frames from monocular video,
- Estimates per-frame depth using state-of-the-art models (MiDaS, DepthAnything),
- Converts depth maps into 3D point clouds using intrinsic camera parameters,
- Aligns multiple point clouds to reconstruct a unified 3D scene, and
- Exports the result to the Universal Scene Description (USD) format for interactive visualization in environments like NVIDIA Omniverse.

This work not only compares various depth estimation models but also demonstrates the viability of scalable, hardware-light 3D reconstruction for robotics, autonomous navigation, and virtual environments.