AI ENABLED DRONE SYSTEMS: 3D RECONSTRUCTION

Meghana V Gaonkar| Leona E Joseph| M Manishwaran|Dr. R Balamurugan| SCOPE

Introduction

RASTER advances 3D reconstruction by addressing scale ambiguity through a novel integration of projective geometry, Structure from Motion, YOLOv8 with AruCo marker detection and 3D background subtraction. Previous methods struggled with scale precision and point correspondence accuracy. This approach refines detection, reprojection, triangulation, improving overall 3D model quality.

Motivation

RASTER aims to enhance 3D reconstruction, bridging physical and digital realms to improve digital twins for augmented reality, urban planning, and cultural heritage, utilizing AI and computer vision.

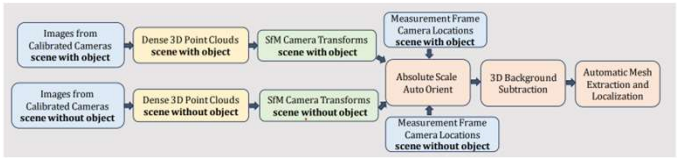
Scope of the Project

The Scope of the Project seeks to enhance 3D reconstruction techniques by resolving scale ambiguity, using a novel integration of YOLOv8 and AruCo marker detection using real life and synthetic datasets. It focuses on advanced object detection, precise triangulation, and 3D background subtraction, aiming to improve applications like urban planning and heritage preservation, streamlining data accuracy and scalability, and optimizing digital twin fidelity and usability, while increasing automation and reducing manual errors.

Methodology

1. RASTER Pipeline

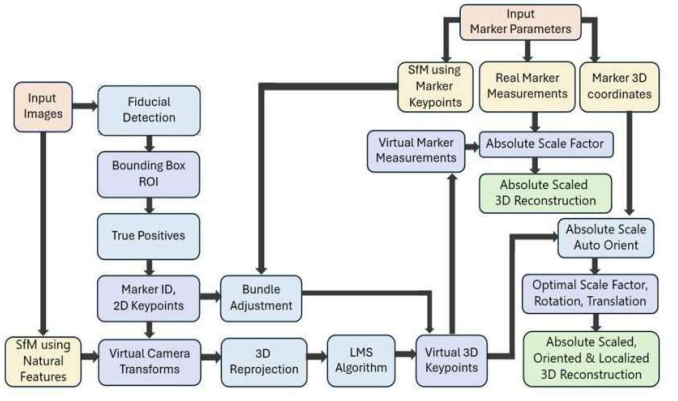
The RASTER (Rapid Absolute Scale Transform Efficient 3D Reconstruction) pipeline for 3D reconstruction with background subtraction and localization. It begins with images taken from calibrated cameras, both with and without the object in the scene. Then we recreate the mesh with and without object using SfM. With the camera locations established, the system determines the absolute scale and orients the measurement frame accordingly. It then performs 3D background subtraction to isolate the object of interest from the scene. The final stage is the extraction and localization of the object within an automatic mesh generation process.



RASTER Pipeline Overview

2. Detection and Scaling Pipeline

The process for 3D reconstruction from input images. Initial stages involve fiducial detection to locate markers, isolating true positives. The procedure continues with Structure from Motion (SfM) using natural features and identified marker key points, proceeding through bundle adjustment and 3D reprojection. The system takes both real and virtual marker measurements to calculate an absolute scale factor. Subsequent stages involve creating an absolute scaled 3D reconstruction, which is automatically oriented. Finally, the optimal scale factor, rotation, and translation parameters are applied to achieve a fully scaled, oriented, and localized 3D reconstruction model.



Detection and Scaling Pipeline Overview

Results

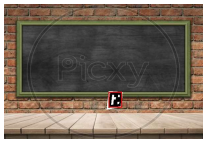
1. **ArUco Marker Detection using YOLOv8:** Results of detection of markers using various models that were trained.



Detection in real world data

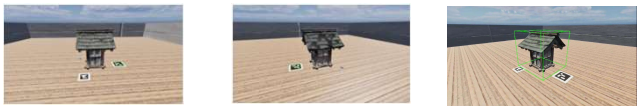


Detection in Synthetic data



Detection of Perspective marker

2. **Automatic Annotation of Synthetic Datasets:** Results obtained while automatically annotating synthetic datasets created using Unreal engine.



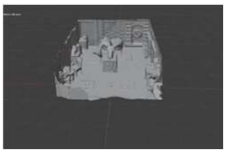
Automatically Annotated ArUco Markers in Synthetic Environment

3. **3D Rescaling:** Results of scale factor estimation using ArUco marker detection used for rescaling scenes to accurate dimensions.



Rescaled Dimensions in 3D Reconstructed Meshes

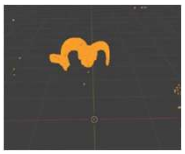
4. **3D Background Subtraction:** Results obtained in the 3D Background Subtraction module are demonstrated below.



Point clouds of Skull in room



Point clouds of room



Result after 3D Background Subtraction

Conclusion

This research project enhances 3D reconstruction using AI-driven drones, integrating YOLOv8 and cv2.aruco for accurate scaling and object detection. Utilizing Unreal Engine, it experiments with diverse datasets to replicate real-world details effectively.

The project aims to streamline processes, reduce errors, and expand future capabilities, targeting improvements in annotation automation, algorithmic accuracy, and environmental adaptability for applications from urban planning to cultural heritage.

References

[1]Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2020). NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. arXiv preprint arXiv:2003.08934.
[2]Chen, A., Xu, Z., Geiger, A., Yu, J., Su, H. (2022). TensoRF: Tensorial Radiance Fields. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T. (eds) Computer Vision – ECCV 2022. ECCV 2022. Lecture Notes in Computer Science, vol 13692. Springer, Cham. https://doi.org/10.1007/978-3- 031-19824-3_20
[3]Maboudi, M., Homaei, M., Song, S., Malihi, S., Saadatseresht, M., & Gerke, M. (2023). A Review on Viewpoints and Path Planning for UAV-Based 3-D Reconstruction. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 16, 5026-5048. https://doi.org/10.1109/JSTARS.2023.3276427