Sparse-view Pose Estimation and Reconstruction via Analysis by Generative Synthesis

NEURAL INFORMATION PROCESSING SYSTEMS

Qitao Zhao   Shubham Tulsiani

PS: Qitao is applying for PhD!

Why is sparse-view reconstruction in the wild still a problem?

This task typically involves tackling two key sub-tasks: pose estimation and 3D inference. While SoTA pose estimation methods (e.g., Ray Diffusion and DUSt3R) offer **approximate** camera poses, 3D inference methods typically assume these poses to be **perfect**. As a result, a naive combination of SoTA methods from both domains falls short.

Method

Overview of *SparseAGS*: Given estimated camera poses from off-the-shelf models, our method iteratively reconstructs 3D and optimizes poses leveraging diffusion priors.

Ablation Study

Input   Camera Poses w/ 3D   Novel Views   Input   Camera Poses w/ 3D   Novel Views

Comparison of Pose Accuracy

Input   Initial Pose   SPARF   SparseAGS (w/ 3D)   Zoom In

Comparison of Novel View Synthesis

Input   GT   LEAP   SparseAGS (Ray Diff.)   SparseAGS (DUSt3R)   Zoom In

tl;dr Given a set of unposed input images, *SparseAGS* jointly infers the corresponding camera poses and underlying 3D, allowing high-fidelity 3D inference in the wild.

Contributions

(1) We introduce an **analysis-by-generative-synthesis** framework that jointly estimates 3D and camera viewpoints given a sparse set of input images, by integrating a 6-DoF novel-view generative prior in an analysis-by-synthesis approach.

(2) Our approach allows leveraging any **off-the-shelf pose estimation system** and can robustly estimate 3D and viewpoints despite **large errors in the initial estimates**.

(3) We present results across datasets and initializations and show clear improvements over the initializations as well as outperform prior sparse-view 3D reconstruction baselines.