Microsoft

Computer Vision Final Project — 3D Reconstruction

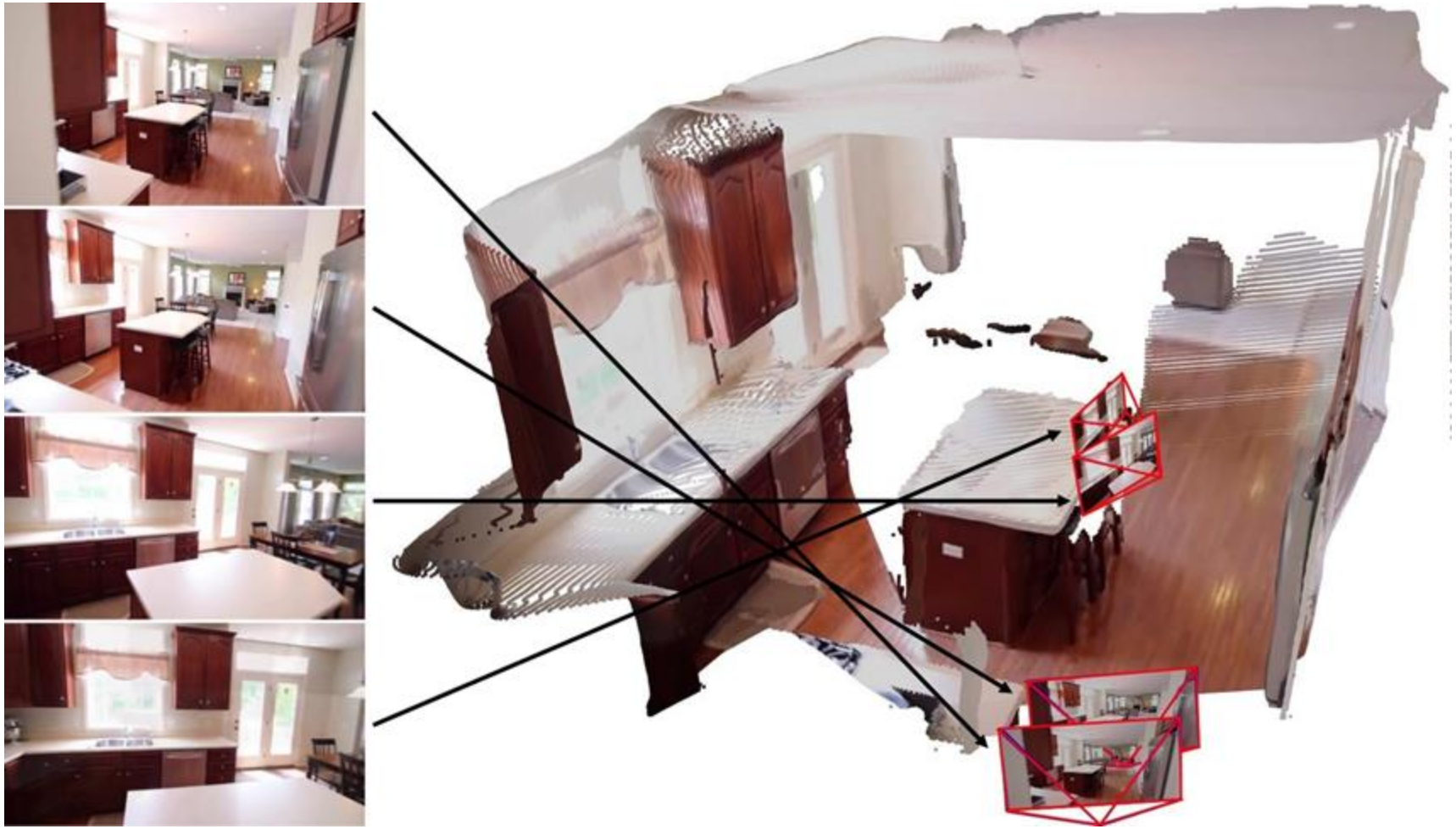
Outline

- Task Description
- Dataset
- Evaluation
- Objective

Task Description

3D Reconstruction

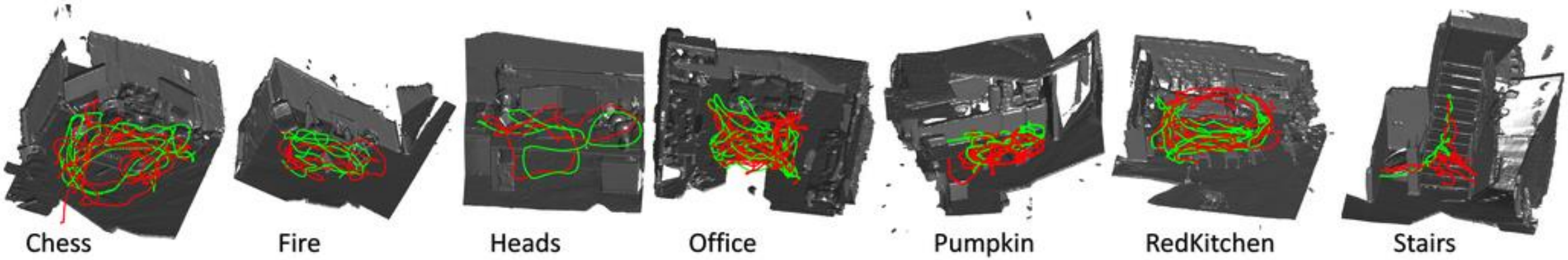
- Input: 2D Multi-view Images
- Output: 3D Scene Reconstruction



Dataset

Dataset

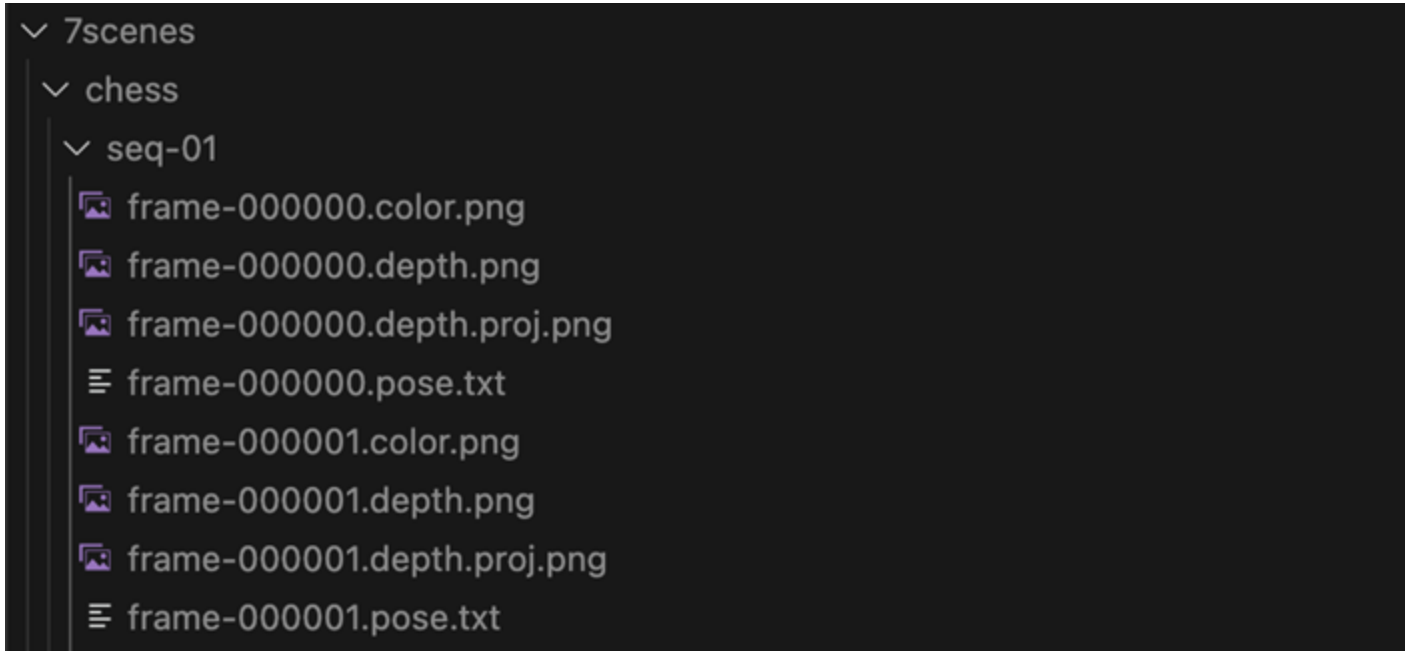
The 7-Scenes dataset is a collection of tracked RGB-D camera frames. All scenes were recorded from a handheld Kinect RGB-D camera at 640×480 resolution.



Dataset

Each sequence (seq-XX.zip) consists of 500-1000 frames. Each frame consists of three files:

- Color: frame-XXXXXX.color.png (RGB, 24-bit, PNG)
- Depth: frame-XXXXXX.depth.png (depth in millimeters, 16-bit, PNG, invalid depth is set to 65535).
- Pose: frame-XXXXXX.pose.txt (camera-to-world, 4×4 matrix in homogeneous coordinates).
- Depth Projection: frame-XXXXXX.depth.proj.png



RGB-D Dataset 7-Scenes

Evaluation

Evaluation Metrics - Acc

Acc: For each predicted point, find its nearest neighbor in the ground-truth point cloud, compute the Euclidean distance between the two points, and take the median of these distances as the Acc score.

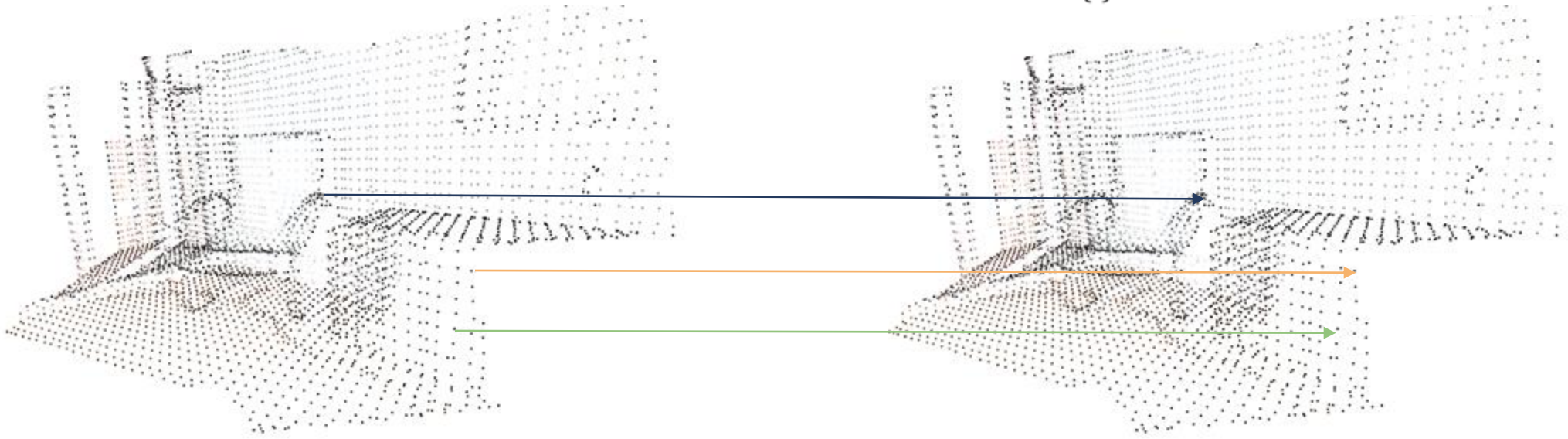
$$\text{Acc}(P, G) = \text{med}_{\mathbf{p}_i \in P} \left[\min_{\mathbf{g}_j \in G} \|\mathbf{p}_i - \mathbf{g}_j\|_2 \right]$$

$P = \{\mathbf{p}_i\}_{i=1}^{N_P}$: Predicted Point Cloud

$G = \{\mathbf{g}_j\}_{j=1}^{N_G}$: Ground-Truth Point Cloud

$\|\cdot\|_2$: Euclidean Distance

$\text{med}\{\cdot\}$: Median number of a set



Predicted Point Cloud

Ground Truth Point Cloud

Evaluation Metrics - Comp

Comp: For each ground-truth point, find its nearest neighbor in the predicted point cloud, compute the Euclidean distance between the two points, and take the median of these distances as the Acc score.

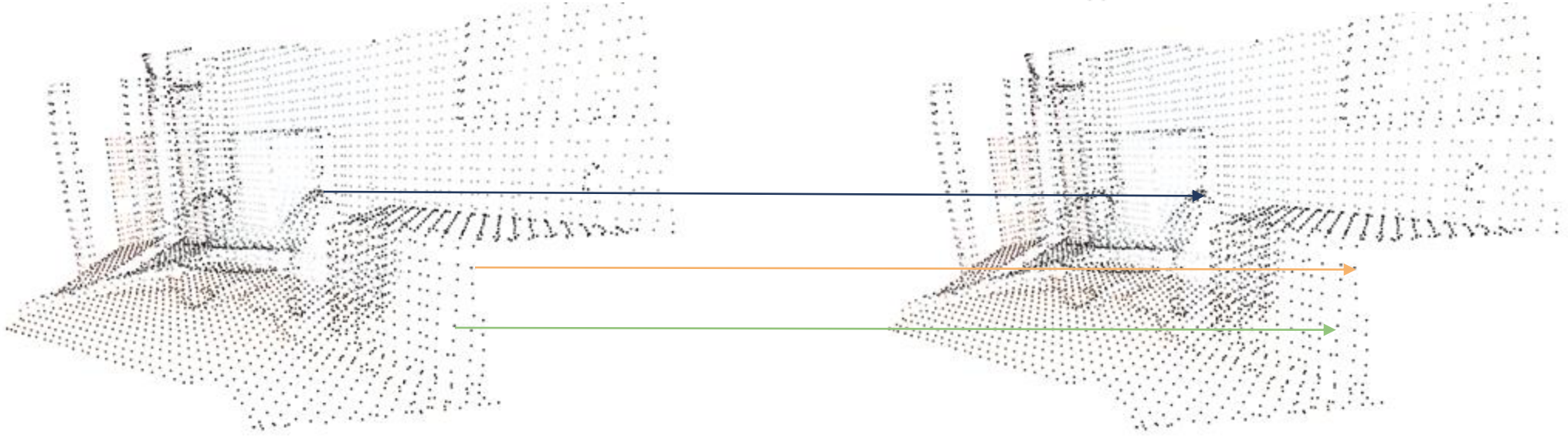
$$\text{Comp}(P, G) = \text{med}_{\mathbf{g}_j \in G} \left[\min_{\mathbf{p}_i \in P} \|\mathbf{g}_j - \mathbf{p}_i\|_2 \right]$$

$P = \{\mathbf{p}_i\}_{i=1}^{N_P}$: Predicted Point Cloud

$G = \{\mathbf{g}_j\}_{j=1}^{N_G}$: Ground-Truth Point Cloud

$\|\cdot\|_2$: Euclidean Distance

$\text{med}\{\cdot\}$: Median number of a set



Ground-truth Point Cloud

Predicted Point Cloud

Objective

Objective

The goal of this project is to **reconstruct accurate 3D scenes** using the **7-Scenes Dataset** and to be ranked on a **Leaderboard** based on two metrics:

- **Accuracy (Acc)**: Median distance between reconstructed points and ground-truth points.
- **Completeness (Comp)**: How well the reconstructed scene covers the ground-truth geometry.

Students are encouraged to:

✅ Use **any 3D reconstruction method or pretrained model**

✅ Train from scratch or **fine-tune on external datasets**

✅ **Experiment freely**, as long as testing data from 7-Scenes remains unseen during training

However, the following are **strictly prohibited**:

🚫 Directly using pretrained models already trained on the 7-Scenes dataset

🚫 Fine-tuning any models on the 7-Scenes **testing set**

Grading

Grading

- Performance (60%)
  - Average **Acc** (30%)
  - Average **Comp** (30%)
- Report (40%) (For Top 10 Teams)
  - Novelty and technical contribution (15%)
  - Experiment completeness (15%)
  - Oral Presentation (10%)
- Report (40%) (For Others)
  - Novelty and technical contribution (20%)
  - Experiment completeness (20%)
- Bonus (10%)
  - Reconstruction with sparse sequence

Points (For each Metric)	# of Teams
30%	1
29%	2
28%	2
26%	The rest teams / 4
24%	The rest teams / 4
22%	The rest teams / 4
20%	The rest teams / 4

Thanks