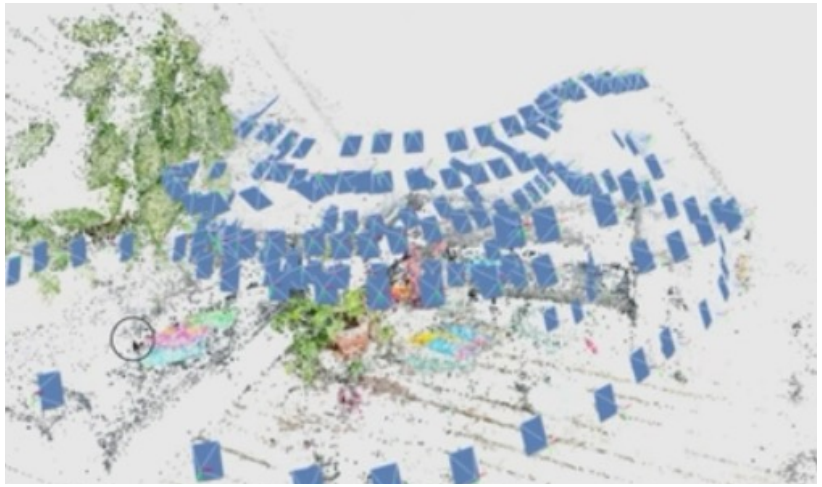
IntelliCap: Intelligent Guidance for Consistent View Sampling

Ayaka Yasunaga Keio University, Japan
Hideo Saito Keio University, Japan
Dieter Schmalstieg University of Stuttgart, Germany
Shohei Mori University of Stuttgart, Keio University

Background

- Novel view synthesis from images has made great progress
- High-quality view synthesis requires uniform and dense sampling

☹ Not easily addressed by human camera operators



Background

- Novel view synthesis from images has made great progress
- High-quality view synthesis requires uniform and dense sampling

  ☹ Not easily addressed by human camera operators

Core Challenge: How can we effectively design a view sampling strategy for novel view synthesis using fewer multi-view images?

Related Works: Small Area Coverage

AR can visually indicate

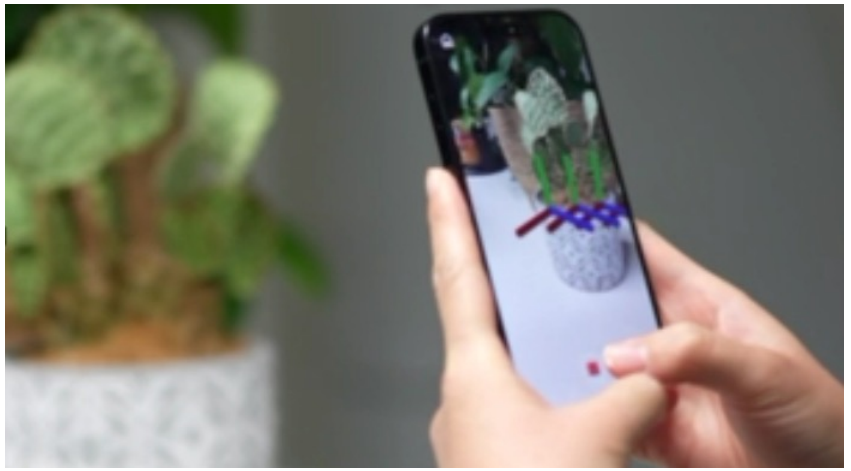
where to collect image sample

3D axes, 2D planes, hemispheres

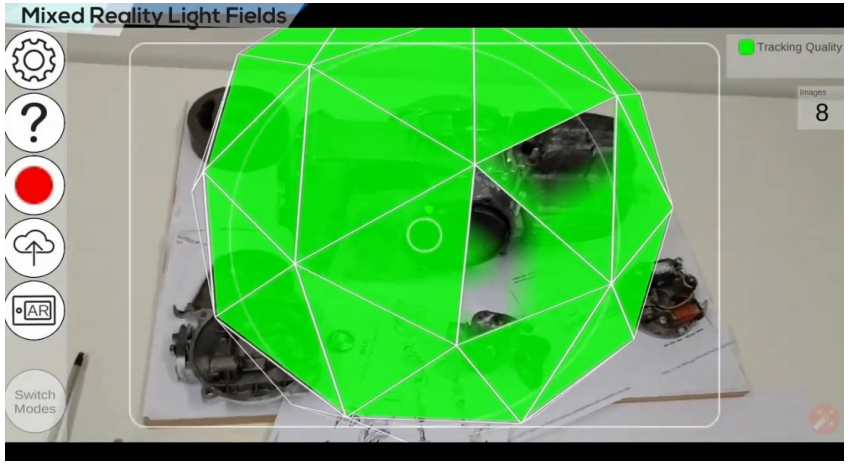
Existing approaches work well for

a single object or forward facing scenes

Solutions that scale to **larger scenes** or **open areas** are lacking



[B. Mildenhall+, TOG, 2019]



[P. Mohr+, CHI, 2020][1]

[1] Mixed Reality Light Fields for Interactive Remote Assistance. (2020). YouTube. https://youtu.be/k77tgEa0LWI?si=Ok496numtRxZPMT7

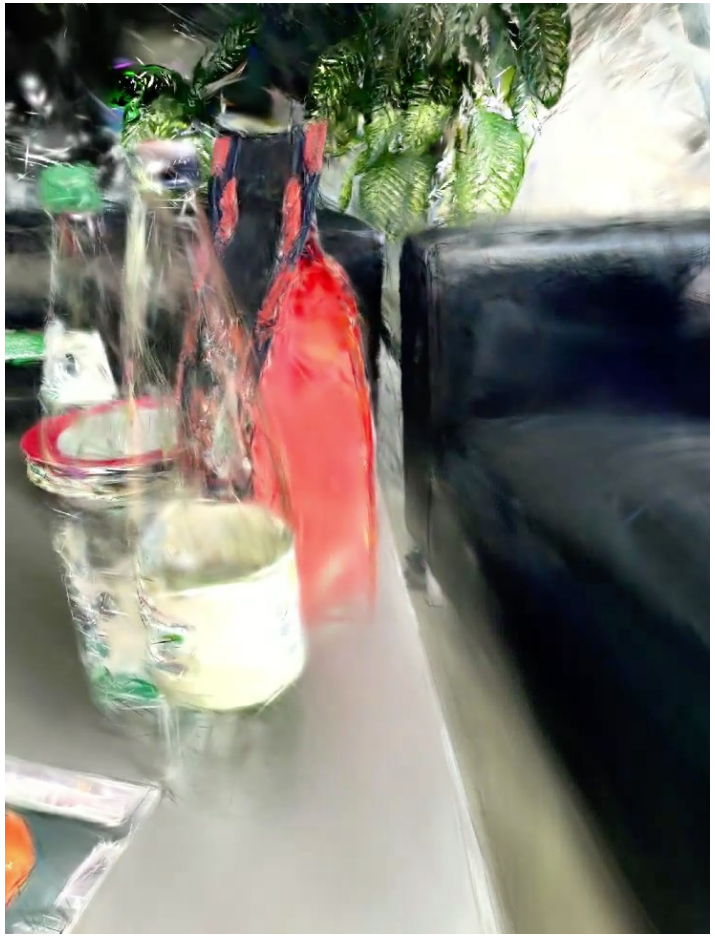
Related Works: Missing Knowledge of Scene Objects

View-dependent effects must be thoroughly captured

specular reflections, transparency, etc.

However, data acquisition should be timely

Therefore, the sampled areas must be **prioritized**



The Solution…?

1. Progressive visual indicators using 3D mesh and sphere
2. LLM-based scoring of detected objects and their categories

Spatial and angle coverage analysis

3D mesh proxy for **spatial coverage**

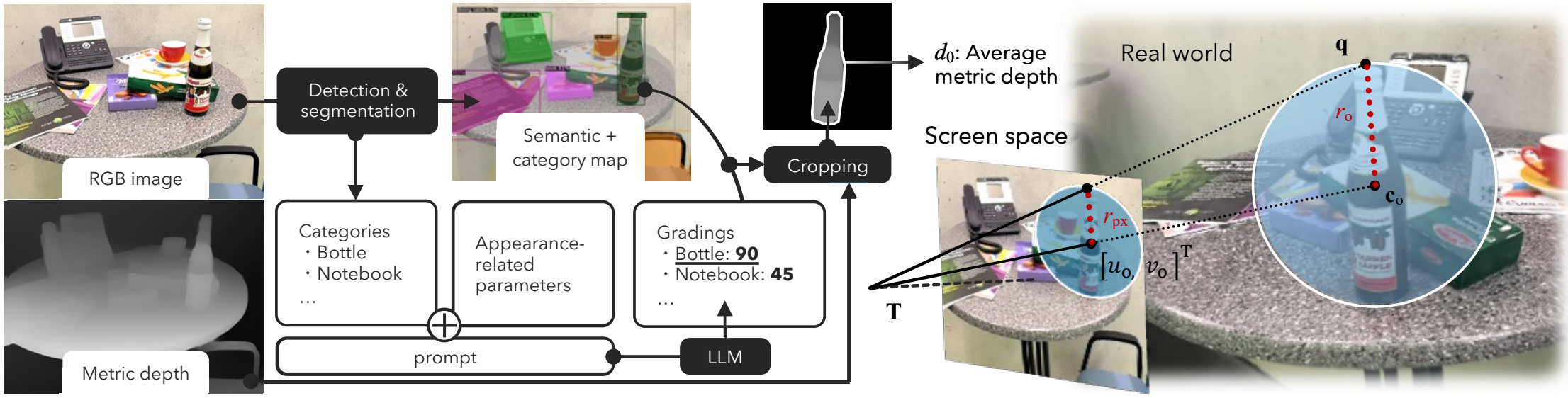
Highlight regions that have not yet been reconstructed, based on the current 3D reconstruction

3D sphere proxy for **angular coverage**

Detect only objects that benefit most from denser view sampling. Only for these, we generate 3D sphere.

3D sphere generation

Our system detects important objects for view sampling using LLM and projects them into real-world space as spheres.



Dynamically adjusting 3D spheres

a. Occlusion handling

b. Distance-based suppression

c. Merging spheres

Scene adaptive 3D sphere visualization

Occlusion handling

Objective : To prevent confusion about occluded areas

Sphere's **alpha value** is dynamically adjusted

based on depth with surrounding scene elements

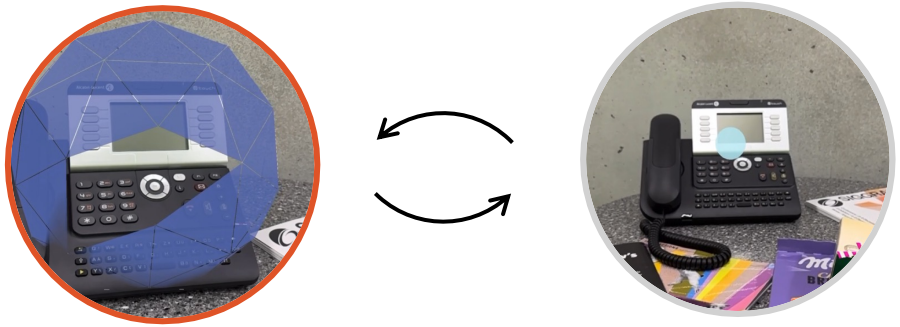
The level of sphere transparency is determined by the offset in depth

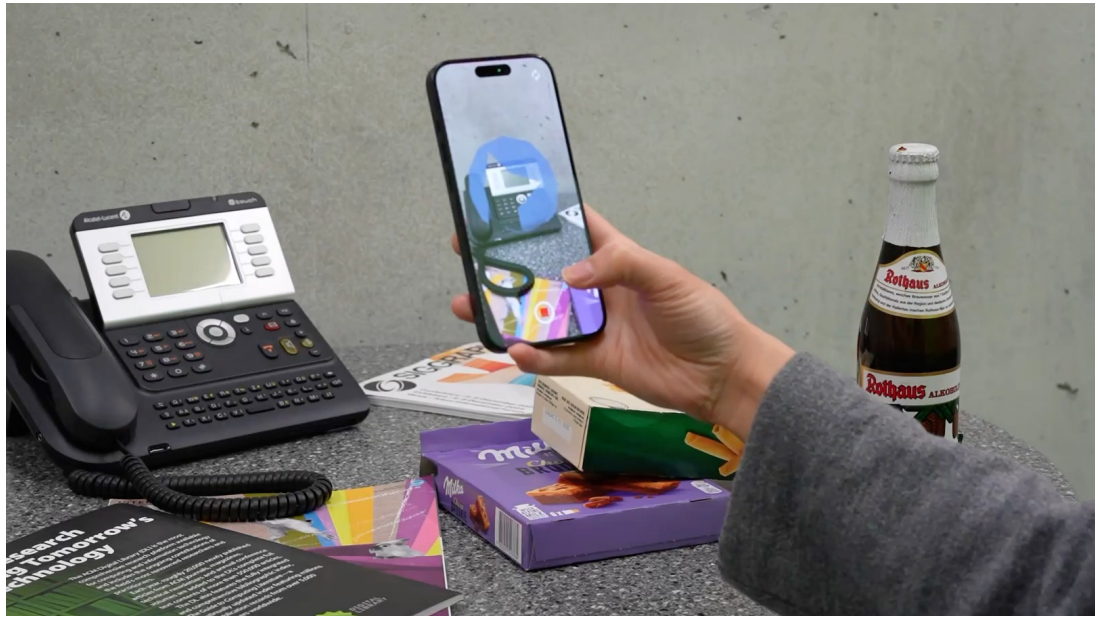
- The depth difference between the sphere and the scene
- Tolerance threshold set to 5.0 cm by default

Distance-based suppression

Objective: View sampling requires approaching within a set distance, encouraging closer interaction

Spheres are visualized only when they fall within a valid distance range





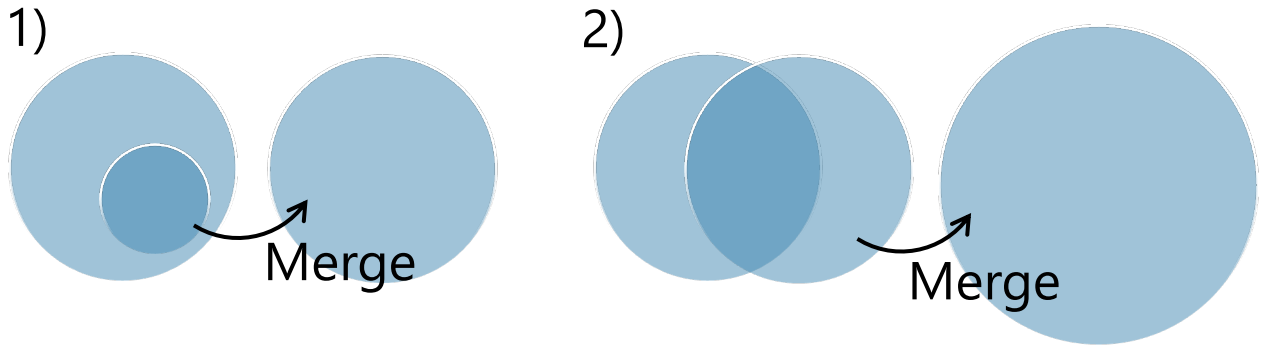
Merging spheres

☹ As the number of spheres increases,

the workload for capturing becomes higher

Objective : To minimize the number of spheres

Intelligently **merging expendable spheres**

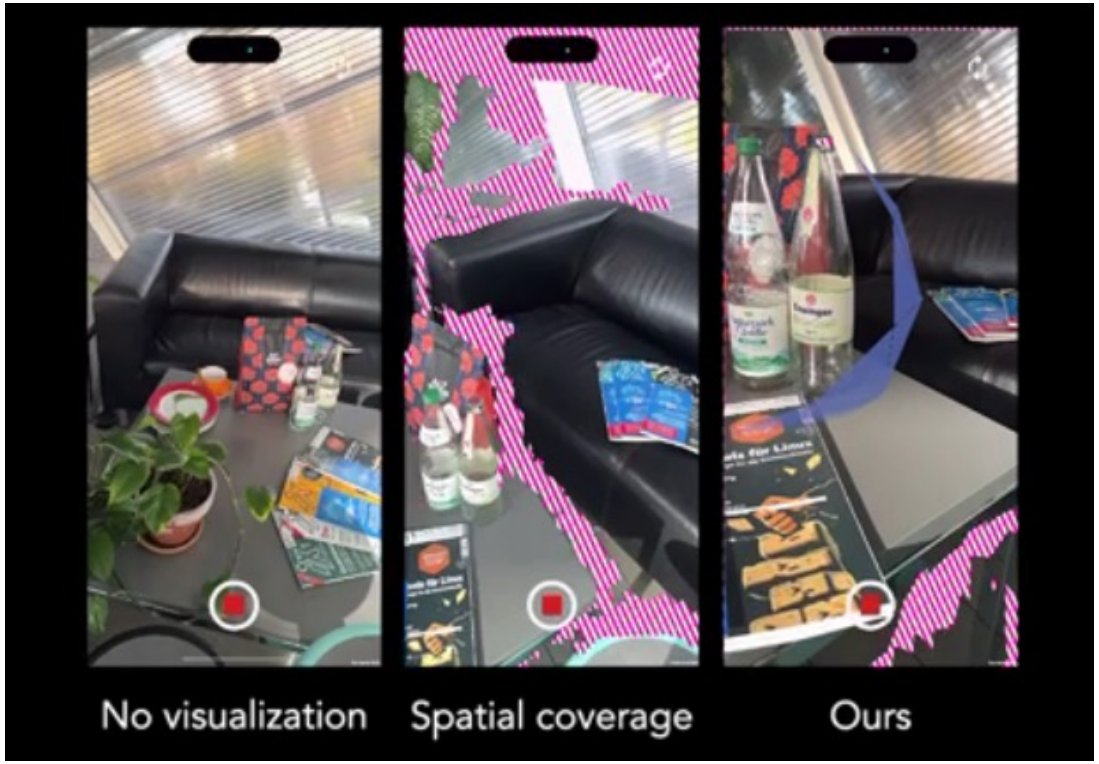


Evaluation: Study Setup

Gaol: We evaluate our system in terms of usability, task load, and the resultant view synthesis quality from images collected.

Baseline approach:

- **NV**: The video stream is presented without any guidance
- **SC:** Only spatial coverage is visualized



Evaluation: Study Setup

Participants: 12 participants (two female and 10 male, $\bar{X}$ =29.2 (SD=4.2) years old, all right-handed and corrected vision).

Scene: Arranged objects with reflective and transparent materials

Lounge

Shelf

Workspace

Result: View synthesis quality

✓ Our method outperforms the baselines in **all image quality metrics**

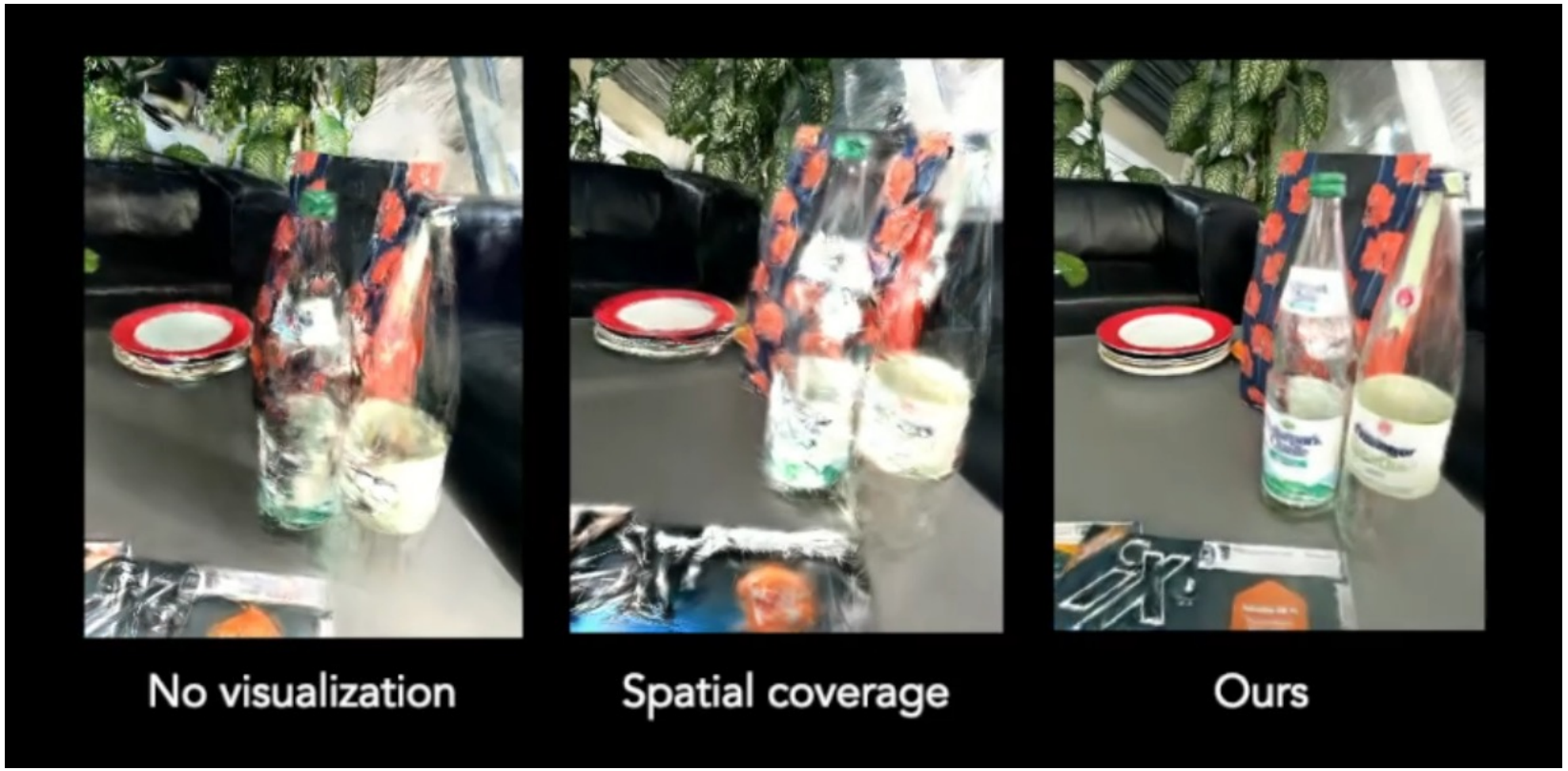
- Performance: Ours > SC > NV

Table: Quantitative comparisons against baseline data collection strategies in view synthesis quality.

	3DGS			Nerfacto		
	PSNR (↑)	SSIM (↑)	LPIPS (↓)	PSNR (↑)	SSIM (↑)	LPIPS (↓)
NV	16.338 (3.346)	0.752 (0.142)	0.292 (0.120)	16.115 (2.592)	0.558 (0.088)	0.589 (0.072)
SC	17.086 (3.435)	0.800 (0.132)	0.253 (0.112)	16.372 (2.488)	0.574 (0.075)	0.570 (0.069)
Ours	**18.891 (3.195)**	**0.848 (0.111)**	**0.201 (0.102)**	**18.134 (1.859)**	**0.616 (0.068)**	**0.520 (0.065)**

Result: View synthesis quality

✓ Our approach **effectively captures** challenging aspects such as **specular reflections** and **transparency**, which baseline methods often miss.

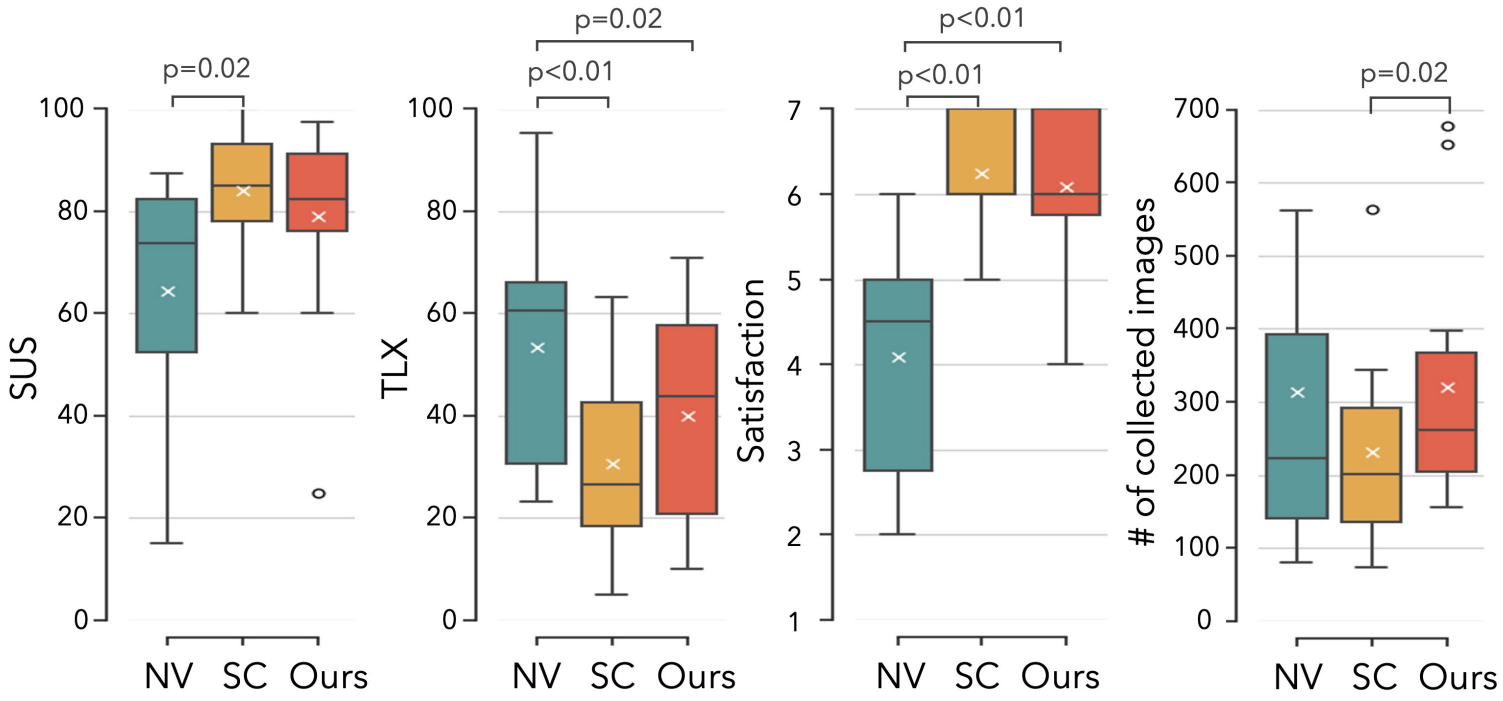


3DGS view synthesis results using data collected by ours and baseline approaches

Result: Usability and Task Load

✓ SUS, TLX, Satis:

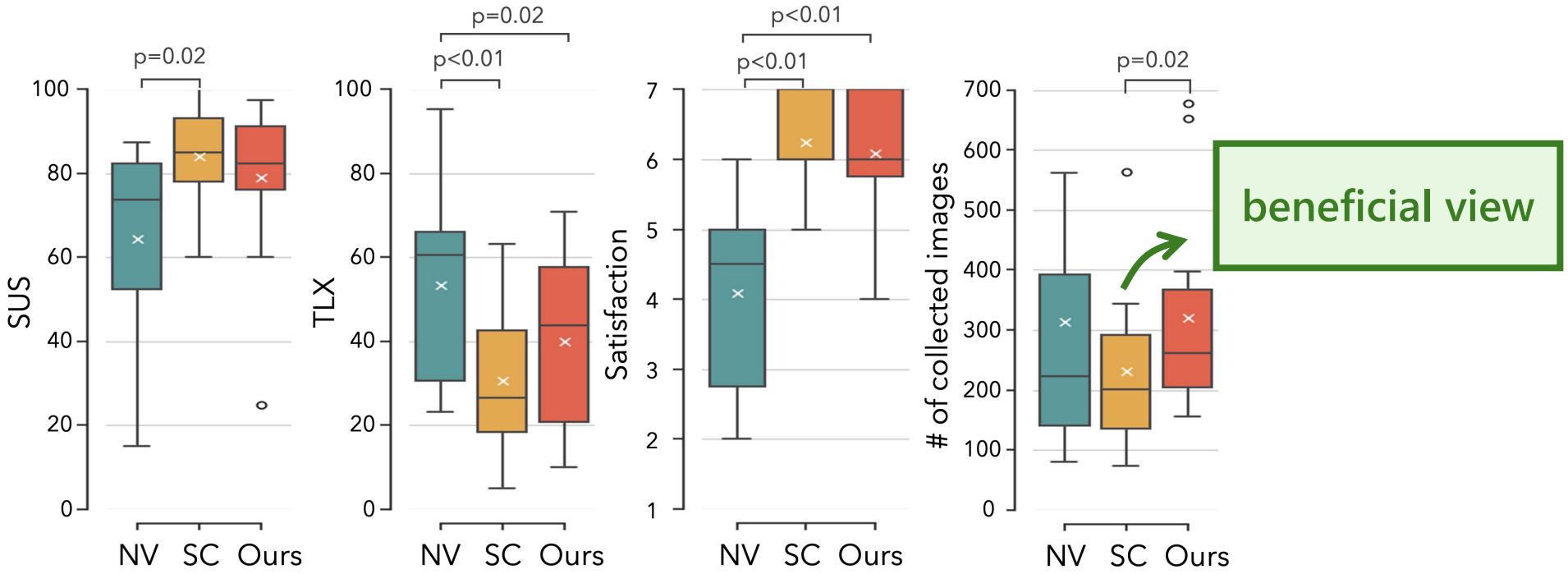
No statistically significant difference was observed for Ours in SUS



Results of the user study

Result: Usability and Task Load

✓ Ours encourages the addition of beneficial viewpoints, enhancing view synthesis quality.



Results of the user study

Conclusion

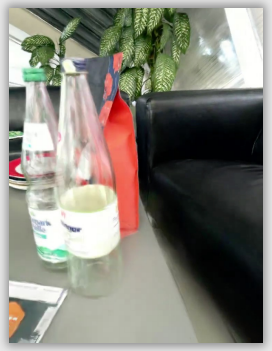
Novel view sampling approach for high-quality view synthesis

- Spatial and angle coverage analysis
- 3D sphere generation
- Dynamically adjusting 3D spheres

Future directions

- Large scale and outdoor scenes
- Efficient task completion
- Misidentification and misalignment

Our approach



Project page