# Accelerated GauSS-MI:
# Accelerated Gaussian Splatting Shannon Mutual Information for Active 3D Reconstruction

Yuhan Xie*, Yixi Cai†, Yinqiang Zhang*, Hairuo Wei‡, and Jia Pan*
* School of Computing and Data Science, The University of Hong Kong, Hong Kong SAR, China
Email: {yuhanxie, zyq507}@connect.hku.hk, jpan@cs.hku.hk
† Department of Robotics, Perception, and Learning, KTH Royal Institute of Technology, Stockholm, Sweden
Email: yixica@kth.se
‡ Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Hong Kong SAR, China
Email: hairuowei@cuhk.edu.hk

***Abstract*—While 3D Gaussian Splatting (3DGS) has significantly improved 3D reconstruction quality, deploying it in active reconstruction remains challenging due to the heavy reliance on viewpoint sampling. Our prior framework, GauSS-MI, addressed this by using Shannon Mutual Information to quantify visual uncertainty for next-best-view selection. However, its high computational overhead creates a severe efficiency bottleneck, restricting the number of sampled viewpoints and limiting online quality evaluation, which ultimately degrades active reconstruction performance. This paper extends GauSS-MI to overcome these computational challenges. Specifically, we present a formal problem formulation for active visual reconstruction, introduce optimization techniques that drastically accelerate online 3DGS computing while enhancing rendering quality, and propose an improved reward formulation for superior view selection. Experimental results demonstrate that our extended framework achieves high-fidelity active reconstruction with substantially reduced computational latency.**

## I. Introduction

3D reconstruction is attracting increasing interest across various fields, including computer vision[10, 7], manipulation, robotics[17], construction, etc. Recent advancements, such as Neural Radiance Field (NeRF)[10] and 3D Gaussian Splatting (3DGS)[7], have notably enhanced the visual quality of 3D reconstruction models. However, these techniques necessitate the prior acquisition of a significant number of images, which can be laborious, and the extensive sampling of viewpoints may result in redundancy. Consequently, a challenging issue arises in effectively and efficiently selecting the viewpoints for image capture, which is also a critical problem for active 3D reconstruction. To tackle this, previous studies have primarily relied on geometric metrics, evaluating either volumetric completeness to explore unknown voxels [17] or surface coverage completeness [2]. Because these indirect metrics overlook rendering quality, the visual fidelity of the resulting model cannot be guaranteed. Powered by modern radiance field rendering methods, recent works have instead attempted to quantify visual uncertainty to directly evaluate and guide the visual quality of reconstruction models [4].

Despite recent progress, efficiently assessing and optimizing visual quality in active 3D reconstruction remains challenging. To address this, our prior conference paper, GauSS-MI [15], proposed a novel view selection metric based on visual uncertainty quantification. Specifically, we introduced a probabilistic model integrating measurement data with image loss to quantify the observed information for each spherical Gaussian in 3D Gaussian Splatting (3DGS). By leveraging Shannon Mutual Information, this model evaluates the expected information gain from arbitrary viewpoints without requiring a prior. Integrated into an active 3DGS reconstruction system, GauSS-MI successfully guides a view and motion planner to determine the next best view and optimal motion primitive in real time.

Despite the success of our prior framework in improving active reconstruction performance, we found that its computational efficiency severely bottlenecks the overall system. This limitation restricts several key components of the active reconstruction system, including but not limited to the number of sampling views and the effective evaluation of reconstruction quality. To address these computational efficiency challenges, this manuscript extends our previous work with the following contributions:

- We propose a comprehensive problem formulation for active 3D visual reconstruction systems.
- We significantly accelerate online 3DGS computing speed while simultaneously improving reconstruction quality.
- We introduce an enhanced reward formulation for more effective next-best-view selection.

## II. Problem Formulation

We study the problem of active 3D reconstruction, in which a robot autonomously explores an unknown environment and incrementally builds a high-fidelity 3D Gaussian Splatting (3DGS) of the scene.

### A. Active 3D Reconstruction

Consider a robot operating in a bounded workspace $\mathcal{W}$ with state space $\mathcal{X} \subset SE(3)$. The robot receives exteroceptive measurements (e.g., RGB images) and maintains an online 3D map $\mathcal{G}$.

*1) Initialization:* At time step $k = 0$, the robot visits a small set of predefined setpoints $\mathcal{X}_0 = \{\mathbf{x}_{0,1}, \ldots, \mathbf{x}_{0,N}\}$ and acquires the corresponding observations $\mathcal{Z}_0 = \{Z_{0,1}, \ldots, Z_{0,N}\}$. These initial observations are used to initialize the reconstruction map $\mathcal{G}_0$, as well as the navigation modules.

*2) Active reconstruction loop:* At each active step $k \geq 1$:

**Online reconstruction.** The robot moves to a viewpoint $\mathbf{x}_k \in \mathcal{X}$, acquires a new observation $Z_k$, and updates the reconstruction map online:

$$\mathcal{G}_k \leftarrow \mathrm{Update}(\mathcal{G}_{k-1},\, Z_k,\, \mathbf{x}_k) \tag{1}$$

Based on the updated map $\mathcal{G}_k$, the robot infers the reconstruction status of the scene by identifying regions that remain under-reconstructed and therefore require additional observations.

**Next-best-view selection.** In practice, reconstruction representations are typically non-differentiable; consequently, the next-best viewpoint (NBV) cannot be inferred from the reconstructed map in closed form. Existing NBV methods usually adopt a two-stage paradigm: *(i)* sampling a set of candidate viewpoints $\mathcal{X}_k^{\mathrm{cand}}$, and *(ii)* evaluating each candidate using a utility reward metric $R(\mathbf{x}; \mathcal{G}_k)$. The next viewpoint is then selected by

$$\mathbf{x}_{k+1} = \arg\max_{\mathbf{x} \in \mathcal{X}_k^{\mathrm{cand}}} R(\mathbf{x}; \mathcal{G}_k) \tag{2}$$

This procedure raises two core problems: how to generate the candidate set $\mathcal{X}_k^{\mathrm{cand}}$, and how to design the utility function $R(\mathbf{x}; \mathcal{G}_k)$.

### B. Active 3D Visual Reconstruction on a Robotic System

Deploying vision-based active reconstruction on physical robotic platforms introduces additional requirements that differ from those in traditional geometric frameworks.

*1) Less but diverse viewpoints:* Viewpoints should not be too sparse: sufficient view overlap and multi-view observations from diverse directions are essential to resolve view-dependent photometric appearance, including color and anisotropic reflectance. However, overly dense observations (e.g., a typical RGB-D camera capturing images at 10 Hz) introduces severe inter-frame redundancy and substantial computational cost. Therefore, we aim for fewer, but sufficiently diverse, observations.

Furthermore, when deployed on a robotic system, the robot's motion must be explicitly considered. Rather than executing tortuous trajectories, we aim to exploit the robot's travel cost efficiently, achieving high-fidelity reconstruction with low motion cost.

TABLE I
MAIN NOTATIONS

| Notations | Explanation |
|---|---|
| $\mathcal{G}$ | 3D Gaussian splatting map. |
| $\mathcal{W}$ | The world frame. |
| $\mathcal{N}$ | A series of ordered Gaussians along a camera ray. |
| $\mathbf{x}, \mathcal{X}$ | A viewpoint and a set of viewpoints. |
| $\boldsymbol{\mu}, c, \alpha$ | Position, color, and opacity of a Gaussian. |
| $T$ | Cumulative transmittance of a Gaussian. |
| $C, \hat{C}$ | Rendered color and observed color. |
| $D, \hat{D}$ | Rendered depth and observed depth. |
| $z, Z$ | Random variable and realization of an observation. |
| $P(r)$ | *Real* probability of a Gaussian. |
| $o, l$ | Odds ratio and log odds of a Gaussian. |
| $\delta$ | Inverse sensor model. |
| $L$ | Loss image between the map and the observation. |
| $\lambda$ | Hyperparameters. |
| $H$ | Entropy of the map. |
| $I$ | Mutual Information between the map and the observation. |
| $M$ | Luminance for a pixel. |
| $(\cdot)^{[i]}$ | Iteration of Gaussians. |
| $(\cdot)^{[j]}$ | Iteration of the measurement beams or pixels. |
| $(\cdot)_k$ | Property based on observation at time $k$. |
| $(\cdot)_{1:k}$ | Property based on the observations from start to time $k$. |

*2) Plan one step ahead:* Evaluating candidate viewpoints for visual reconstruction is computationally demanding, which limits the practical size of the candidate set $\mathcal{X}_k^{\mathrm{cand}}$. Moreover, to enable continuous robot motion and real-time planning, the robot needs to determine the next viewpoint $\mathbf{x}_{k+1}$ in advance—before completing the motion from $\mathbf{x}_{k-1}$ to $\mathbf{x}_k$—when only observations up to $\mathcal{Z}_{0:k-1}$ are available. Consequently, in practice, $\mathbf{x}_{k+1}$ is selected based on the previous map $\mathcal{G}_{k-1}$. We accordingly reformulate (2) as:

$$\mathbf{x}_{k+1} = \arg\max_{\mathbf{x} \in \mathcal{X}_k^{\mathrm{cand}}} R(\mathbf{x}; \mathcal{G}_{k-1}) \tag{3}$$

## III. MAPPING

For an active reconstruction system, high-quality real-time reconstruction can reveal how the current observations contribute to the reconstruction, thereby enabling the system to infer which regions of the scene have been adequately reconstructed and which remain incomplete. In this chapter, building on 3D Gaussian Splatting (3DGS) and the online method MonoGS[9], we design and optimize an online reconstruction framework tailored to active reconstruction systems.

The main notations for proposed methodolody (including Section III and Section V-A2) are listed in Table I.

### A. 3D Gaussian Splatting Mapping

The proposed system reconstructs the scene by 3DGS, utilizing a collection of anisotropic 3D Gaussians, represented by $\mathcal{G}$. Each 3D Gaussian $i$ contains the properties of mean $\boldsymbol{\mu}_{\mathcal{W}}^{[i]}$ and covariance $\boldsymbol{\Sigma}_{\mathcal{W}}^{[i]}$, representing the geometrical position and ellipsoidal shape in the world frame $\mathcal{W}$, and also optical properties including color $c^{[i]}$ and opacity $\alpha^{[i]}$. By splatting and blending a series of ordered Gaussians $\mathcal{N}$, the color $C^{[j]}$ and depth $D^{[j]}$ for each pixel are synthesized as

$$C^{[j]} = \sum_{i \in \mathcal{N}} c^{[i]} T^{[i]} = \sum_{i \in \mathcal{N}} c^{[i]} \alpha^{[i]} \prod_{n=1}^{i-1} (1 - \alpha^{[n]}) \tag{4}$$

$$D^{[j]} = \sum_{i \in \mathcal{N}} d^{[i]} T^{[i]} = \sum_{i \in \mathcal{N}} d^{[i]} \alpha^{[i]} \prod_{n=1}^{i-1} (1 - \alpha^{[n]}) \tag{5}$$

where $d^{[i]}$ represents the distance from camera pose $\mathbf{x}$ to the position $\boldsymbol{\mu}_{\mathcal{W}}^{[i]}$ of Gaussian $i$ along the camera ray. We denote

$$T^{[i]} = \alpha^{[i]} \prod_{n=1}^{i-1} (1 - \alpha^{[n]}) \tag{6}$$

as the cumulative transmittance of Gaussian $i$ along the ray.

We also render the silhouette image [6] to determine whether a pixel contains information from the current map,

$$S^{[j]} = \sum_{i \in \mathcal{N}} T^{[i]} = \sum_{i \in \mathcal{N}} \alpha^{[i]} \prod_{n=1}^{i-1} (1 - \alpha^{[n]}) \tag{7}$$

### B. Online Reconstruction

At time step $k$, assume we have an existing 3DGS map $\mathcal{G}_k$ that has been trained from a set of obseravtion frames $Z_0$ to $Z_k$. Given a new RGB-D frame $Z_{k+1}$, our online reconstruction module performs the following steps:

*1) Initialization:* The map adds new Gaussians based on the paired RGB-D pixels from new observation. Specifically, for each pixel, a new Gaussian is initialized by the color of the pixel, the center at the 3D location of the projection from pixel depth, the opacity of $0.5$, and the rotation of identity quaternion. The scales for each Gaussian are estimated from local nearest-neighbor distances.

*2) Map Updating:* Unlike SLAM algorithms[9, 6, 8] that is to handle a continuous observation stream, an active reconstruction system can actively choose the next best view for reconstruction. In our system, we believe the viewpoint selected by our proposed viewpoint selection algorithm (presented in the subsequent Section IV), is reliable. Therefore, we do not maintain any keyframes or window of frames anymore; we regard all input frames as 'keyframes' for online reconstruction.

After receiving new Gaussians from the new observation, the system iteratively updates and optimizes the map. At each iteration, one random frame is selected and optimized. The iteration goes on until the next frame is obtained.

*3) Densification and Prune:* To regulate the frequency of densification and pruning as new frames are continuously incorporated, we perform this operation only once per incoming frame, at the $K_d$-th update iteration after the frame is received.

In addition, since online reconstruction allows only a limited number of training iterations, we adopt a more aggressive opacity-control strategy than standard offline 3DGS reconstruction: we increase the opacity-reset ratio and also raise the minimum-opacity threshold used for pruning Gaussians. Therefore, low-quality Gaussians can be removed more rapidly from the reconstruction.

## IV. Viewpoint Selection

This section presents Gaussian Splatting Shannon Mutual Information (GauSS-MI), an information-theoretic metric for selecting the next viewpoint.

### A. 3D Gaussian reliability probability

To model the information obtained from the 3DGS map $\mathcal{G}$ by a random observation $z$, we propose a random variable $r^{[i]}$ for each Gaussian $i \in \mathcal{G}$. We define the probability that a 3D Gaussian $i$ is *reliable* for rendering as $P(r^{[i]}) \in (0, 1)$. The probabilities are assumed to be independent for each 3D Gaussian. We also denote the odds ratio $o^{[i]}$ and log odds $l^{[i]}$,

$$l^{[i]} := \log(o^{[i]}) := \log(\frac{P(r^{[i]})}{1 - P(r^{[i]})}) \tag{8}$$

At the initial, we assume there is no prior information, i.e., $P_0(r^{[i]}) = 0.5, \quad \forall i \in \mathcal{G}$. We then derive the probability $P(r^{[i]}|Z_{1:k})$ based on the observations $Z_1$ to $Z_k$.

Given an observation $Z_k$, the rendering loss between the observed groundtruth image from $Z_k$ and the rendered image from $\mathcal{G}$ can be computed by

$$L_k = \lambda_c \|C - \hat{C}_k\| + (1 - \lambda_c) \|D - \hat{D}_k\| \tag{9}$$

The *reliable* probability of Gaussian $i$ under the observation $Z_k$ is constructed by the rendering loss $L_k$ as

$$P(r^{[i]} \mid Z_k) = \frac{1}{(\lambda_L L_k)^{\lambda_T T^{[i]}} + 1} \tag{10}$$

We refer to $P(r^{[i]}|Z_k)$ as the inverse sensor model. The odds ratio of the inverse sensor model can be derived as

$$\delta^{[i]}(Z_k) = \frac{P(r^{[i]} \mid Z_k)}{1 - P(r^{[i]} \mid Z_k)} = (\lambda_L L_k)^{-\lambda_T T^{[i]}} \tag{11}$$

Based on the constructed inverse sensor model, the probability can be updated by binary Bayesian filter,

$$o^{[i]}(Z_{1:k}) := \delta^{[i]}(Z_k) o^{[i]}(Z_{1:k-1}) \tag{12}$$

In implementation, the probability $P(r^{[i]}|Z_{1:k})$ is updated by log odds $l_{1:k}^{[i]}$,

$$l_{1:k}^{[i]} = -\lambda_T T^{[i]} \log \lambda_L L_k + l_{1:k-1}^{[i]} \tag{13}$$

The computation can be realized by rasterizing the $\log$ of mapping loss $L_k$ as (4).

### B. Gaussian Splatting Shannon Mutual Information

Based on the proposed probability model and Shannon mutual information, we construct Gaussian Splatting Shannon mutual information (GauSS-MI) to assess the visual quality of novel viewpoints.

*1) GauSS-MI:* Given past observations $Z_{1:k-1}$, the next best view is expected to minimize conditional entropy $H(r|z_k, Z_{1:k-1})$. Simplifying the probability condition, the conditional entropy can be computed as follows based on information theory,

$$H(r \mid z) = H(r) - I(r; z) \tag{14}$$

Therefore, minimizing the conditional entropy is equivalent to maximize the mutual information (MI).

Assuming pixel-wise independent measurements, MI can be decomposed over pixels $j \in \{1, \ldots, n_z\}$ and the visible Gaussians $\mathcal{N}^{[j]}$ [5]:

$$I(r; z) = \sum_{j=1}^{n_z} \sum_{i \in \mathcal{N}^{[j]}} I(r^{[i]}; z^{[j]})\, T^{[i]} \tag{15}$$

From information theory [1, 5], the mutual information between two random variables is defined and can be organized as

$$\begin{aligned} I(r^{[i]}; z^{[j]}) :=& P(r^{[i]}, z^{[j]} = Z) \log(\frac{P(r^{[i]}, z^{[j]} = Z)}{P(r^{[i]}) P(z^{[j]} = Z)}) \\ =& P(z^{[j]} = Z) f(\delta^{[i]}(Z), o^{[i]}_{1:k-1}) \end{aligned} \tag{16}$$

where $P(z^{[j]} = Z)$ is only related to the observation, which is referred to as the measurement prior. $f(\delta^{[i]}(Z), o^{[i]}_{1:k-1})$ can be derived and written in shorthand as

$$f(\delta, o) = \frac{o}{o + \delta^{-1}} \log(\frac{o+1}{o + \delta^{-1}}) \tag{17}$$

The function $f(\delta, o)$ can be interpreted as an information gain function. We define the mutual information (16) between the 3DGS map and the observation as Gaussian Splatting Shannon Mutual Information, GauSS-MI.

*2) Expected GauSS-MI:* For an unobserved candidate viewpoint, we approximate the measurement prior by a luminance-dependent camera noise model [3]:

$$P(z^{[j]}) \approx P(z^{[j]} \mid M^{[j]}) \tag{18}$$

where $M^{[j]}$ is the pixel's expected luminance, which can be computed from the expected RGB color $(R, G, B)$ that $M = 0.299R + 0.587G + 0.114B$.

We expect that the rendering result after reconstruction is *reliable*. Thus, we assume $L_k = 0$ so that $\delta^{-1} = 0$ in $f(\delta, o)$. Then the information gain function (17) can be derived as,

$$f^{[i]} = \log(\frac{o^{[i]} + 1}{o^{[i]}}) = -\log(P(r^{[i]})) \tag{19}$$

Integrating (15)(16)(18)(19), the expected GauSS-MI can be computed as

$$I(r; z) = \sum_{j=1}^{n_z} P(z^{[j]} | M^{[j]}) \sum_{i \in \mathcal{N}^{[j]}} -T^{[i]} \log(P(r^{[i]})) \tag{20}$$

## V. System Implementation

This section details the system implementation. We present the view planner for viewpoint sampling and selection and discuss the autonomous termination condition design for the proposed system.

### A. View Planning

*1) Viewpoint Primitive Library:* To determine the next best viewpoint, we design an action library to generate a set of candidate viewpoints, and choose the next best view within the candidates. Inspired by the action generation method proposed in [16], we design the action to the next viewpoint by

$$\boldsymbol{\alpha} = [v_{\mathrm{xy}}, v_{\mathrm{z}}, \omega_{\mathrm{z}}]$$

where $v_{\mathrm{xy}}$ and $v_{\mathrm{z}}$ represent the body frame linear velocity in $\mathrm{x}_{\mathcal{B}} - \mathrm{y}_{\mathcal{B}}$ plane and $\mathrm{z}_{\mathcal{B}}$ direction, and $\omega_{\mathrm{z}}$ is the body frame angular velocity around the $\mathrm{z}_{\mathcal{B}}$ axis. We simplify the action of 2-dimensional horizontal movement into 1 dimension, which can be compensated through the $\omega_{\mathrm{z}}$ rotation. The action space is given by sampling each velocity that,

$$\mathcal{A} = \{\boldsymbol{\alpha} | v_{\mathrm{xy}} \in \mathcal{V}_{\mathrm{xy}}, v_{\mathrm{z}} \in \mathcal{V}_{\mathrm{z}}, \omega_{\mathrm{z}} \in \Omega_{\mathrm{z}}\} \tag{21}$$

In this paper, we assume that the sensor, normally with a limited field of view, is equipped forward, i.e., facing the $\mathrm{x}_{\mathcal{B}}$ axis. Thus, in the further forward propagation derivation, we design the horizontal movement action $v_{\mathrm{xy}}$ works on the $\mathrm{y}_{\mathcal{B}}$ axis.

Given the action $\boldsymbol{\alpha} = [v_{\mathrm{xy}}, v_{\mathrm{z}}, \omega_{\mathrm{z}}] \in \mathcal{A}$ , the next viewpoint is designed by forward propagation with duration time $T$,

$$\begin{aligned} \mathbf{x}_f &= \mathbf{x}_0 + \begin{bmatrix} -v_{\mathrm{xy}} T \sin(\psi_0 + \omega_{\mathrm{z}} T) \\ v_{\mathrm{xy}} T \cos(\psi_0 + \omega_{\mathrm{z}} T) \\ v_{\mathrm{z}} T \\ \omega_{\mathrm{z}} T \end{bmatrix} \\ \mathbf{x}_f^{(n)} &= \mathbf{0} \quad \text{for } n = 1, 2, 3 \end{aligned} \tag{22}$$

where $(\cdot)^{(n)}$ denotes the $n$-th derivatives, which constraints the final state to ensure a stable picture taking on the next viewpoint. A motion primitive trajectory $\mathbf{x}_T$ from current state $\mathbf{x}(t) = \mathbf{x}_0$ to the next viewpoint $\mathbf{x}(t+T) = \mathbf{x}_f$ can be derived in closed-form [11].

Overall, by defining the set of actions $\mathcal{A}$, given the current state $\mathbf{x}_0$, a library of candidate viewpoints $\mathcal{X}_f^{\mathrm{cand}} = \{\mathbf{x}_f\}$ along with the primitive trajectories $\mathcal{X}_T = \{\mathbf{x}_T\}$ can be formed as a viewpoint primitive library.

*2) Next Best View Selection:* The total reward for the next best view evaluation includes the mutual information $I$ (20) and the motion cost $J$ as

$$R = w_I I + w_U U - w_J J \tag{23}$$

where $w_I, w_U, w_J > 0$ are constant reward weights to balance the range of two components. $U$ denotes the unknown ratio gain from the viewpoint, calculated by the zero pixel ratio in silhouette image (7). The unknown ratio gain $U$ helps to aware unknown region in the scene. The motion cost $J$ can be calculated based on the trajectory $\mathbf{x}_T$ with respect to a specific mobile robot. The next best view with primitive $\mathbf{x}_T^*$ is selected by optimizing $R$ over all feasible primitives, which is then assigned to the controller for tracking.

## VI. Experiments

In this section, we present the experiments designed to validate the proposed method.

### A. Experiment Setup

The simulation environment is created using Habitat simulator[1] [12] and Replica dataset[2] [14], The drone motion planner

[1]https://github.com/facebookresearch/habitat-sim
[2]https://github.com/facebookresearch/replica-dataset

TABLE II
PARAMETERS OF THE PROPOSED SYSTEM

| Parameter | Value |
|---|---|
| hyperparameter on loss $\lambda_L$ | 1.7 |
| hyperparameter on cumulative transmittance $\lambda_L$ | 7.0 |
| Primitive duration time $T$ | 1.6 s |
| reward weight on information $w_I$ | 0.03 |
| reward weight on unknown ratio $w_U$ | 5.0 |
| reward weight on motion cost $w_J$ | 0.01 |

is realized by SUPER[3] [13]. A quadrotor is employed as the agent for active reconstruction, equipped with an image sensor providing RGB-D images at a resolution of $640 \times 480$ and a $90\,\mathrm{deg}$ Field of View (FOV). Both the proposed active reconstruction system and the simulator operate on a desktop with a 24-core i9-14900K CPU and an RTX5080 GPU. The parameters of the proposed system are summarized in Table II.

### B. Metrics

We evaluate our method by assessing the visual quality of the online reconstruction results and counting the number of frames $N_f$ within the active reconstruction process. Visual quality assessment includes Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS) to quantitatively compare rendered images from the 3DGS model with a testing dataset of ground-truth images.

### C. Benchmarks

We consider the following methods for comparison:

- *Ours*: The proposed method in this paper.
- *GauSS-MI*[4] [15]: Our prior work, which proposed active reconstruction framework with previous online reconstruction algorithm.

### D. Results

*1) NBV Metric Runtime Tests:* We test and compare the runtime for both methods. GauSS-MI[15] achieves an average runtime of $3.039\,\mathrm{ms}$ ($329.0\,\mathrm{fps}$). The proposed Accelerated GauSS-MI reaches the average runtime of $1.512\,\mathrm{ms}$ ($661.2\,\mathrm{fps}$), demonstrating the improvement of the proposed framework.

*2) Active Reconstruction System Tests:* The quantitative results is shown in Table III. The novel view synthesis results for two scenes are shown in Figure 1 and Figure 2, respectively. The testing 3DGS map are reconstructed online, and the testing dataset are collected from novel views.

The results indicate that the proposed method and framework substantially enhance online reconstruction performance, thereby improving autonomous reconstruction systems. Moreover, the $N_f$ results and the holes visible in Figure 1 suggest that *GauSS-MI* fails to adequately explore the room and becomes trapped in a local minimum during viewpoint selection.

[3]https://github.com/hku-mars/SUPER
[4]GauSS-MI: https://github.com/JohannaXie/GauSS-MI

TABLE III
ACTIVE RECONSTRUCTION RESULTS AND COMPARISON

| Scene | Metric | PSNR↑ | SSIM↑ | LPIPS↓ | $N_f$ ↓ |
|---|---|---|---|---|---|
| Office 0 | Ours | 33.46 | 0.9669 | 0.0758 | 78 |
| | *GauSS-MI* | 24.24 | 0.9008 | 0.2117 | 48 |
| Office 2 | Ours | 30.18 | 0.9578 | 0.1198 | 82 |
| | *GauSS-MI* | 22.60 | 0.8796 | 0.3013 | 75 |

In contrast, within the proposed framework, the unknown-ratio gain successfully guides the viewpoint toward previously unknown regions, leading to more complete reconstructions.

## VII. CONCLUSION AND FUTURE WORK

This paper addresses the computational bottlenecks of active 3D reconstruction systems using 3D Gaussian Splatting (3DGS). Building upon our prior work, GauSS-MI, which introduced a real-time visual uncertainty metric based on Shannon Mutual Information, this manuscript introduces an accelerated, highly efficient framework. By establishing a comprehensive problem formulation, significantly increasing the speed and quality of online 3DGS computation, and optimizing the next-best-view reward formulation, we successfully mitigate system overhead, enabling dense view sampling and real-time visual quality assessment.

## REFERENCES


[1] Thomas M Cover and Joy A Thomas. Information theory and the stock market. *Elements of Information Theory. Wiley Inc., New York*, pages 543–556, 1991.

[2] Chen Feng, Haojia Li, Mingjie Zhang, Xinyi Chen, Boyu Zhou, and Shaojie Shen. Fc-planner: A skeleton-guided planning framework for fast aerial coverage of complex 3d scenes. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 8686–8692. IEEE, 2024.

[3] Alessandro Foi, Mejdi Trimeche, Vladimir Katkovnik, and Karen Egiazarian. Practical poissonian-gaussian noise modeling and fitting for single-image raw-data. *IEEE transactions on image processing*, 17(10):1737–1754, 2008.

[4] Wen Jiang, Boshu Lei, and Kostas Daniilidis. Fisherrf: Active view selection and mapping with radiance fields using fisher information. In *European Conference on Computer Vision*, pages 422–440. Springer, 2025.

[5] Brian J Julian, Sertac Karaman, and Daniela Rus. On mutual information-based control of range sensing robots for mapping applications. *The International Journal of Robotics Research*, 33(10):1375–1392, 2014.

[6] Nikhil Keetha, Jay Karhade, Krishna Murthy Jatavallabhula, Gengshan Yang, Sebastian Scherer, Deva Ramanan, and Jonathon Luiten. Splatam: Splat, track & map 3d gaussians for dense rgb-d slam. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

[7] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time

Fig. 1. Novel view synthesis results.

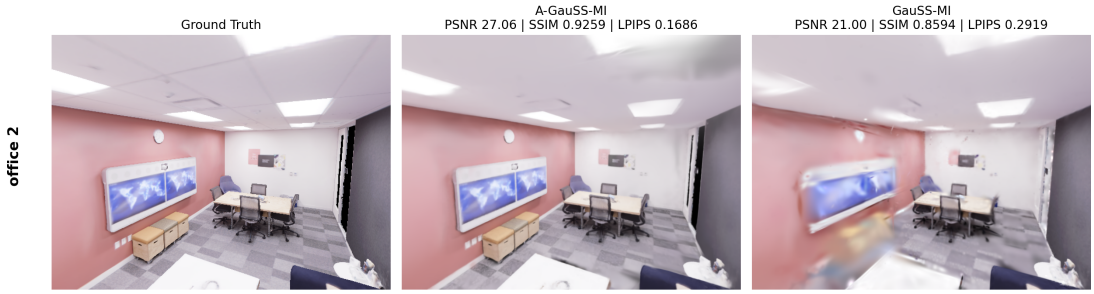


Fig. 2. Novel view synthesis results.

radiance field rendering. *ACM Transactions on Graphics*, 42(4), July 2023. URL https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/.

[8] Saswat Subhajyoti Mallick, Rahul Goel, Bernhard Kerbl, Markus Steinberger, Francisco Vicente Carrasco, and Fernando De La Torre. Taming 3dgs: High-quality radiance fields with limited resources. In *SIGGRAPH Asia 2024 Conference Papers*, SA '24, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400711312. doi: 10.1145/3680528.3687694. URL https://doi.org/10.1145/3680528.3687694.

[9] Hidenobu Matsuki, Riku Murai, Paul H. J. Kelly, and Andrew J. Davison. Gaussian Splatting SLAM. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.

[10] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.

[11] Mark W Mueller, Markus Hehn, and Raffaello D'Andrea. A computationally efficient motion primitive for quadrocopter trajectory generation. *IEEE transactions on robotics*, 31(6):1294–1310, 2015.

[12] Xavi Puig, Eric Undersander, Andrew Szot, Mikael Dallaire Cote, Ruslan Partsey, Jimmy Yang, Ruta Desai, Alexander William Clegg, Michal Hlavac, Tiffany Min, Theo Gervet, Vladimir Vondrus, Vincent-Pierre Berges, John Turner, Oleksandr Maksymets, Zsolt Kira, Mrinal Kalakrishnan, Jitendra Malik, Devendra Singh Chaplot, Unnat Jain, Dhruv Batra, Akshara Rai, and Roozbeh Mottaghi. Habitat 3.0: A co-habitat for humans, avatars and robots, 2023.

[13] Yunfan Ren, Fangcheng Zhu, Guozheng Lu, Yixi Cai, Longji Yin, Fanze Kong, Jiarong Lin, Nan Chen, and Fu Zhang. Safety-assured high-speed navigation for mavs. *Science Robotics*, 10(98):eado6187, 2025.

[14] Julian Straub, Thomas Whelan, Lingni Ma, Yufan Chen, Erik Wijmans, Simon Green, Jakob J. Engel, Raul Mur-Artal, Carl Ren, Shobhit Verma, Anton Clarkson, Mingfei Yan, Brian Budge, Yajie Yan, Xiaqing Pan, June Yon, Yuyang Zou, Kimberly Leon, Nigel Carter, Jesus Briales, Tyler Gillingham, Elias Mueggler, Luis Pesqueira, Manolis Savva, Dhruv Batra, Hauke M. Strasdat, Renzo De Nardi, Michael Goesele, Steven Lovegrove, and Richard Newcombe. The Replica dataset: A digital replica of indoor spaces. *arXiv preprint arXiv:1906.05797*, 2019.

[15] Yuhan Xie, Yixi Cai, Yinqiang Zhang, Lei Yang, and Jia Pan. GauSS-MI: Gaussian Splatting Shannon Mutual Information for Active 3D Reconstruction. In *Proceedings of Robotics: Science and Systems*, LosAngeles, CA, USA, June 2025. doi: 10.15607/RSS.2025.XXI.030.

[16] Xuning Yang, Koushil Sreenath, and Nathan Michael. A framework for efficient teleoperation via online adaptation. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5948–5953. IEEE, 2017.

[17] Boyu Zhou, Yichen Zhang, Xinyi Chen, and Shaojie Shen. FUEL: Fast uav exploration using incremental frontier structure and hierarchical planning. *IEEE Robotics and Automation Letters*, 6(2):779–786, 2021.