

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

Yuhan Xie¹, Yixi Cai², Yinqiang Zhang¹, Lei Yang³, and Jia Pan¹

Abstract—This research tackles the challenge of real-time active view selection and uncertainty quantification on visual quality for active 3D reconstruction. Visual quality is a critical aspect of 3D reconstruction. Recent advancements such as Neural Radiance Fields (NeRF) and 3D Gaussian Splatting (3DGS) have notably enhanced the image rendering quality of reconstruction models. Nonetheless, the efficient and effective acquisition of input images for reconstruction—specifically, the selection of the most informative viewpoint—remains an open challenge, which is crucial for active reconstruction. Existing studies have primarily focused on evaluating geometric completeness and exploring unobserved or unknown regions, without direct evaluation of the visual uncertainty within the reconstruction model. To address this gap, this paper introduces a probabilistic model that quantifies visual uncertainty for each Gaussian. Leveraging Shannon Mutual Information, we formulate a criterion, Gaussian Splatting Shannon Mutual Information (GauSS-MI), for real-time assessment of visual mutual information from novel viewpoints, facilitating the selection of next best view. GauSS-MI is implemented within an active reconstruction system integrated with a view and motion planner. Extensive experiments across various simulated and real-world scenes showcase the superior visual quality and reconstruction efficiency performance of the proposed system.

I. INTRODUCTION

3D reconstruction is attracting increasing interest across various fields, including computer vision [1], [2], manipulation, robotics [3], construction, etc. Recent advancements, such as Neural Radiance Field (NeRF) [1] and 3D Gaussian Splatting (3DGS) [2], 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 enhance the autonomy of robots and enable them to perform 3D reconstruction tasks in complex environments, there has been a growing focus on active 3D reconstruction in

recent years [3], [4]. In the active 3D reconstruction process, at each decision step, the agent must utilize a series of past observations to actively determine the next viewpoint for capturing new observation, thus gradually accomplishing the reconstruction task. The efficient selection of viewpoints is particularly crucial in this process due to limited onboard resources such as battery power, memory, and computation capability. Previous studies on active 3D reconstruction have primarily relied on evaluating volumetric completeness to explore all unknown voxels in the environment [3]–[5] or assessing surface coverage completeness [6]. These approaches overlook the visual quality. By utilizing these indirect metrics, the resulting visual fidelity of the reconstruction model cannot be guaranteed. Advanced by radiance field rendering methods [1], [2], recent works have attempted to quantify visual uncertainty to directly evaluate the visual quality of reconstruction models [7], [8].

Despite these efforts, effectively and efficiently assessing and optimizing visual quality in active 3D reconstruction remains challenging. To address this, three core issues must be resolved. Firstly, a robust mathematical model is necessary to quantify the information obtained from each measurement, specifically the observed image. This model can serve as a reconstruction completeness metric for visual fidelity. Secondly, a metric is needed to measure the expected information from novel viewpoints without a prior, which can facilitate the selection of the next viewpoint in the active reconstruction process. Lastly, a comprehensive active reconstruction system is required to autonomously identify a reasonable next viewpoint with the highest expected information. The system should then enable the agent to navigate to the selected viewpoint, capture new data, and iteratively advance the reconstruction process.

To overcome the aforementioned challenges, this paper proposes a novel view selection metric based on a visual uncertainty quantification method, from which we develop a novel active 3D reconstruction system. We first introduce a probabilistic model that integrates the measurement model with image loss to quantify the observed information for each spherical Gaussian in 3D Gaussian Splatting. Based on Shannon Mutual Information theory, we leverage the probabilistic model to establish the mutual information between the reconstruction model and observation viewpoint, which measures the expected information gained from an arbitrary viewpoint for the current reconstruction model. This mutual information function is termed Gaussian Splatting Shannon Mutual Information (GauSS-MI), enabling real-time visual quality assessment from novel viewpoints without a prior. The GauSS-MI is implemented and integrated into a novel

¹Yuhan Xie, Yinqiang Zhang, and Jia Pan are with the School of Computing and Data Science, The University of Hong Kong, Hong Kong SAR, China. Email: {yuhanyxie, zyq507}@connect.hku.hk, jpan@cs.hku.hk.

²Y. Cai is with the Division of Robotics and Perception and Learning, KTH Royal Institute of Technology, Stockholm, Sweden. Email: yixica@kth.se.

³Lei Yang is with the Faculty of Engineering, The University of Hong Kong, Hong Kong SAR, China. Email: lyang125@hku.hk

This research is partially supported by the Innovation and Technology Commission of the HKSAR Government under the InnoHK initiative, Hong Kong Research Grants Council under NSFC/RGC Collaborative Research Scheme (CRS.HKU703/24) and Joint Research Scheme (N.HKU705/24). Jia Pan is the corresponding author. Yixi Cai is the project leader.

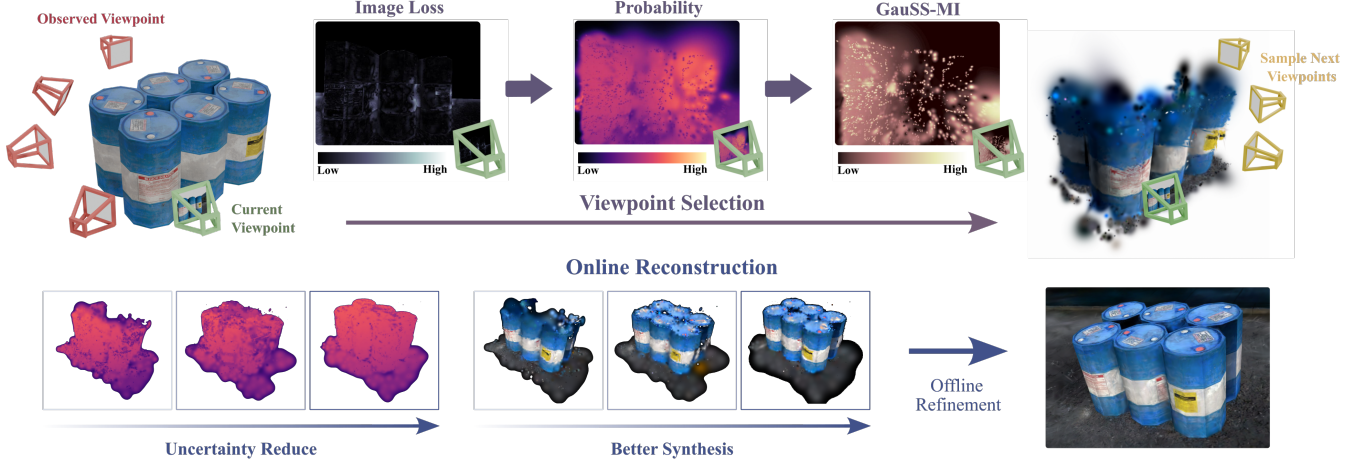


Fig. 1. Illustration of the proposed Gaussian Splatting Shannon Mutual Information (GauSS-MI) method.

active 3D Gaussian splatting reconstruction system featuring a view and motion planner that determines the next best view and optimal motion primitive. Extensive experiments, including benchmark comparisons against state-of-the-art methods, validate the superior performance of the proposed system in terms of visual fidelity and reconstruction efficiency. The implementation of the proposed system is open-sourced on Github¹ to support and advance future research within the community.

The main contributions of our work are summarized below:

- A probabilistic model for the 3D Gaussian Splatting map to quantify the image rendering uncertainty.
- A novel Gaussian Splatting Shannon Mutual Information (GauSS-MI) metric for real-time assessment of visual mutual information from novel viewpoints.
- An active 3D Gaussian splatting reconstruction system implementation based on GauSS-MI.
- Extensive benchmark experiments against state-of-the-art methods demonstrate the superior performance of the proposed system in terms of visual fidelity and reconstruction efficiency.

II. METHODOLOGY

This section presents the probabilistic model for 3D Gaussian Splatting (3DGS) in visual uncertainty quantification, followed by the formulation of Gaussian Splatting Shannon Mutual Information (GauSS-MI) for view selection.

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 $\mu_{\mathcal{W}}^{[i]}$ and covariance $\Sigma_{\mathcal{W}}^{[i]}$, 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

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

$$D^{[j]} = \sum_{i \in \mathcal{N}} d^{[i]} T^{[i]} = \sum_{i \in \mathcal{N}} d^{[i]} \alpha^{[i]} \prod_{n=1}^{i-1} (1 - \alpha^{[n]}) \quad (2)$$

where $d^{[i]}$ represents the distance from camera pose σ to the position $\mu_{\mathcal{W}}^{[i]}$ of Gaussian i along the camera ray. We denote

$$T^{[i]} = \alpha^{[i]} \prod_{n=1}^{i-1} (1 - \alpha^{[n]}) \quad (3)$$

as the cumulative transmittance of Gaussian i along the ray.

At each reconstruction step, the 3D Gaussians are extended and initialized using the collected RGB-D image and estimated camera pose [9]. Then the Gaussians iteratively optimize both their geometric and optical parameters to represent the captured scene with high visual fidelity.

B. 3D Gaussian probability

To model the information obtained from the 3DGS map $\mathcal{G}$ by a random observation z , we first construct a random variable r for each Gaussian. As we are going to optimize the rendering result, we define the probability of a 3D Gaussian i is *reliable* for rendering as $P(r^{[i]}) \in (0, 1)$. Then, the probability of the 3D Gaussian i is *unreliable* for rendering is $P(\bar{r}^{[i]}) = 1 - P(r^{[i]})$. Additionally, we denote the odds ratio $o^{[i]} \in (0, +\infty)$ and log odds $l^{[i]} \in (-\infty, +\infty)$ of a Gaussian by

$$l^{[i]} := \log(o^{[i]}) := \log\left(\frac{P(r^{[i]})}{P(\bar{r}^{[i]})}\right) = \log\left(\frac{P(r^{[i]})}{1 - P(r^{[i]})}\right) \quad (4)$$

We assume each probability of the 3D Gaussian is independent. At the initial stage of the mapping, we assume that the agent has no prior information on the environment, i.e.,

$$P_0(r^{[i]}) = P_0(\bar{r}^{[i]}) = 0.5 \quad \forall i \in \mathcal{G} \quad (5)$$

¹<https://github.com/JohannaXie/GauSS-MI>

Once a new observation Z_k is obtained at time k , the standard binary Bayesian filter can be used to update the probability

$$\begin{aligned} o^{[i]}(Z_{1:k}) &:= \frac{P(r^{[i]}|Z_{1:k})}{P(\bar{r}^{[i]}|Z_{1:k})} \\ &= \frac{P(r^{[i]}|Z_k) P(r^{[i]}|Z_{1:k-1})}{P(\bar{r}^{[i]}|Z_k) P(\bar{r}^{[i]}|Z_{1:k-1})} \\ &= \delta^{[i]}(Z_k) o^{[i]}(Z_{1:k-1}) \end{aligned} \quad (6)$$

where $P(r^{[i]}|Z_k)$ is the *reliable* probability of Gaussian i under the observation Z_k . We refer to $P(r^{[i]}|Z_k)$ as the inverse sensor model, thereby $\delta^{[i]}(Z_k)$ is the odds ratio of the inverse sensor model, which will be constructed and used for updating the *reliable* probability $P(r^{[i]}|Z_{1:k})$. We further use $o_{1:k}^{[i]}$ and $l_{1:k}^{[i]}$ as a shorthand of $o^{[i]}(Z_{1:k})$ and $l^{[i]}(Z_{1:k})$ respectively, referring to the odds ratio for Gaussian i based on the observations from start to time k .

Given the observation Z_k , we construct the $P(r^{[i]}|Z_k)$ as

$$P(r^{[i]}|Z_k) = \frac{1}{(\lambda_L L_k)^{\lambda_T T^{[i]}} + 1} \quad (7)$$

Therefore, the odds ratio of inverse sensor model $\delta^{[i]}(Z_k)$ can be derived as

$$\delta^{[i]}(Z_k) = \frac{P(r^{[i]}|Z_k)}{1 - P(r^{[i]}|Z_k)} = (\lambda_L L_k)^{-\lambda_T T^{[i]}} \quad (8)$$

where $\lambda_L, \lambda_T > 0$ are hyperparameters. L_k denotes the loss between the observation Z_k and the map, i.e., a loss image between the observed groundtruth image and the rendered image. We compute the loss image by

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

where C, D denote the rendered color and depth images from the reconstructed 3DGS map, $\hat{C}_k, \hat{D}_k$ are the groundtruth color and depth images from observation Z_k .

As the 3DGS map optimizes the Gaussians by minimizing the image loss, we use this loss to construct the inverse sensor model, and the cumulative transmittance to regulate the update rate.

To accelerate computation, in implementation, we update the probability $P(r^{[i]}|Z_{1:k})$ by computing log odds $l_{1:k}^{[i]}$. Take log of (6) and substitute (8),

$$l_{1:k}^{[i]} = -\lambda_T T^{[i]} \log \lambda_L L_k + l_{1:k-1}^{[i]} \quad (10)$$

Therefore, the log odds of inverse sensor model can be computed by rasterizing the mapping loss L_k as (1).

C. Gaussian Splatting Shannon Mutual Information

Based on the proposed probability model and Shannon Mutual Information theory, we then construct the Gaussian Splatting Shannon mutual information (GauSS-MI) for visual quality assessment of novel viewpoints.

Given the previous observations $Z_{1:k-1}$, we are interested in minimizing the expected uncertainty, i.e., conditional entropy, of the map after receiving the agent's next observation

z_k . In information theory, the conditional entropy relates to Mutual Information (MI) by

$$H(r|z_k, Z_{1:k-1}) = H(r|Z_{1:k-1}) - I(r; z_k|Z_{1:k-1}) \quad (11)$$

To minimize the conditional entropy $H(r|z_k, Z_{1:k-1})$ is to maximize the MI $I(r; z_k|Z_{1:k-1})$. Note here that we use z_k and Z_k to distinguish random variable and realized variable for the observation at time k .

As we assume that the previous observations $Z_{1:k-1}$ are given and try to compute the MI for the new observation z_k , in the subsequent of this subsection, we omit the probability condition $Z_{1:k-1}$ and simplify z_k into z . Therefore, the (11) can be simplified as

$$H(r|z) = H(r) - I(r; z) \quad (12)$$

As z is a random variable with independence among elements, the total MI can be expressed as the summation of $I(r; z^{[j]})$ between $z^{[j]}$ over all measurement beams $j \in \{1, \dots, n_z\}$ [10].

$$I(r; z) = \sum_{j=1}^{n_z} I(r; z^{[j]}) = \sum_{j=1}^{n_z} \sum_{i \in \mathcal{N}^{[j]}} I(r^{[i]}; z^{[j]}) T^{[i]} \quad (13)$$

Here, the measurement beams $j \in \{1, \dots, n_z\}$ refer to each picture pixel.

From information theory [10], [11], the mutual information between two random variables is defined and can be organized as

$$\begin{aligned} I(r^{[i]}; z^{[j]}) &:= P(r^{[i]}, z^{[j]} = Z) \log \left(\frac{P(r^{[i]}, z^{[j]} = Z)}{P(r^{[i]})P(z^{[j]} = Z)} \right) \\ &= P(z^{[j]} = Z) P(r^{[i]}|z^{[j]} = Z) \log \left(\frac{P(r^{[i]}|z^{[j]} = Z)}{P(r^{[i]})} \right) \\ &= P(z^{[j]} = Z) f(\delta^{[i]}(Z), o_{1:k-1}^{[i]}) \end{aligned} \quad (14)$$

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

$$f(\delta, o) = \frac{o}{o + \delta^{-1}} \log \left(\frac{o + 1}{o + \delta^{-1}} \right) \quad (15)$$

The function $f(\delta, o)$ can be interpreted as an information gain function.

We define the mutual information (14) between the 3DGS map and the observation as Gaussian Splatting Shannon Mutual Information, GauSS-MI.

D. Computation of Expected GauSS-MI

We further derive the computation of the expected mutual information (14) for random viewpoints.

1) *Measurement prior*: We refer to the noise model of RGB camera in [12], in which the expectation of the measurement noise is related to luminance. Thus, we construct the measurement prior $P(z)$ for each pixel j as

$$P(z^{[j]}) = \sum_{m^{[j]}=0}^{255} P(z^{[j]}|m^{[j]}) P(m^{[j]}) \quad (16)$$

where $P(z^{[j]}|m^{[j]})$ is the prior probability distribution of the sensor with respect to luminance $m \in \{0, \dots, 255\}$. To compute the expected measurement prior, we define $P(m^{[j]})$ as

$$P(m^{[j]}) = \begin{cases} 1 & \text{for } m^{[j]} = M^{[j]} \\ 0 & \text{otherwise} \end{cases} \quad (17)$$

where $M^{[j]}$ is the pixel's expected luminance, which can be computed from the expected RGB color (R, G, B) by the luminance formula $M = 0.299R + 0.587G + 0.114B$. Thus the measurement prior (16) can be simplified as

$$P(z^{[j]}) = P(z^{[j]}|M^{[j]}) \quad (18)$$

2) *Information gain function*: As there are no observations from random viewpoints, computing the loss image L_k for $\delta(Z_k)$ is infeasible; thus, an expectation of L_k is required. We expect that the rendering result after reconstruction is *reliable*, i.e., there is no loss between groundtruth and the 3DGS map $\mathcal{G}$. Thus, we assume $L_k = 0$ so that $\delta^{-1} = 0$ in $f(\delta, o)$. Then the information gain function (15) can be derived as,

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

The equation shows that when the *reliable* probability $P(r^{[i]})$ is low, the information gain function f will be high, consistent with the intuition of information gain.

Overall, integrating (13)(14)(18)(19), the expected GauSS-MI can be computed as

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

III. SYSTEM IMPLEMENTATION

A. System Overview

The proposed active reconstruction system comprises a reconstruction module and a planning module, as illustrated in Figure 2. In this work, a mobile robot is equipped with sensors that can capture color images and depth images and estimate its pose. Given these messages, the reconstruction module constructs and updates a 3D Gaussian splatting (3DGS) model in real-time, while simultaneously generating the 3D Gaussian probability map. Meanwhile, the planning module creates a library of candidate viewpoints along with the primitive trajectories. The optimal viewpoint and primitive trajectory are subsequently determined by evaluating both the viewpoint's GauSS-MI and the trajectory's motion energy cost. The robot then follows the selected primitive trajectory and captures images from the next-best viewpoint. Given the new observations, the reconstruction module could update the map. The process iterates and results in a high-quality 3D reconstruction with detailed visual representation.

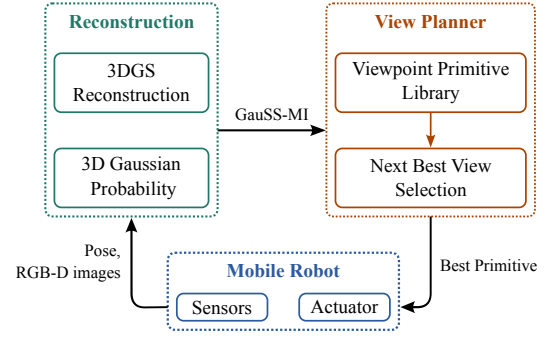


Fig. 2. Overview of proposed active 3D reconstruction system.

B. 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 [13], we design the action to the next viewpoint by

$$\alpha = [v_{xy}, v_z, \omega_z]$$

where v_{xy} and v_z represent the body frame linear velocity in $x_B - y_B$ plane and z_B direction, and ω_z is the body frame angular velocity around the z_B axis. We simplify the action of 2-dimensional horizontal movement into 1 dimension, which can be compensated through the ω_z rotation. The action space is given by sampling each velocity that,

$$\mathcal{A} = \{\alpha | v_{xy} \in \mathcal{V}_{xy}, v_z \in \mathcal{V}_z, \omega_z \in \Omega_z\} \quad (21)$$

In this paper, we assume that the sensor, normally with a limited field of view, is equipped forward, i.e., facing the x_B axis. Thus, in the further forward propagation derivation, we design the horizontal movement action v_{xy} works on the y_B axis.

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

$$\begin{aligned} \sigma_f &= \sigma_0 + \begin{bmatrix} -v_{xy}T \sin(\psi_0 + \omega_z T) \\ v_{xy}T \cos(\psi_0 + \omega_z T) \\ v_z T \\ \omega_z T \end{bmatrix} \\ \sigma_f^{(n)} &= \mathbf{0} \quad \text{for } n = 1, 2, 3 \end{aligned} \quad (22)$$

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 σ_T from current state $\sigma(t) = \sigma_0$ to the next viewpoint $\sigma(t+T) = \sigma_f$ can be derived in closed-form [14].

Overall, by defining the set of actions $\mathcal{A}$, given the current state σ_0 , a library of candidate viewpoints $\Sigma_f = \{\sigma_f\}$ along with the primitive trajectories $\Sigma_T = \{\sigma_T\}$ 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

$$R = w_I I - w_J J \quad (23)$$

TABLE I
PARAMETERS OF THE PROPOSED SYSTEM

Parameter	Value
hyperparameter on loss λ_L	1.7
hyperparameter on cumulative transmittance λ_L	7.0
Primitive duration time T	1.6 s
reward weight on information w_I	0.03
reward weight on motion cost w_J	0.01
probability threshold τ	0.7
reconstruction terminate threshold φ	75%

where $w_I, w_J > 0$ are constant reward weights to balance the range of two components. The motion cost J can be calculated based on the trajectory σ_T with respect to a specific mobile robot. The next best view with primitive σ_T^* is selected by optimizing R over all feasible primitives, which is then assigned to the controller for tracking.

IV. SIMULATION EXPERIMENTS

In this section, we present a series of simulation experiments designed to validate the proposed method. We begin by detailing the experimental setup and evaluation metrics. Based on this, we initially validate the proposed system (Section IV-A). Subsequently, we conduct experiments to evaluate the proposed GauSS-MI metric from multiple perspectives: the efficiency of next-best-view selection (Section IV-B), real-time computational performance (Section IV-C). Finally, we compare the complete system against baseline methods in Section IV-D.

A. System Validation

1) *Simulation Setup*: The simulation environment is created using Flightmare [15], featuring a configurable rendering engine within Unity² and a versatile drone dynamics simulation. A quadrotor is employed as the agent for active reconstruction, equipped with an image sensor providing RGB-D images at a resolution of 640×480 and a 90 deg Field of View (FOV). The online 3D Gaussian splatting reconstruction is developed based on MonoGS [9], which incorporates depth measurements to enhance the online reconstruction model. Both the proposed active reconstruction system and the simulator operate on a desktop with a 32-core i9-14900K CPU and an RTX4090 GPU. The parameters of the proposed system are summarized in Table I.

2) *Metrics*: The evaluation focuses on assessing the visual quality of the reconstruction results and the efficiency of the active reconstruction process. Visual quality is evaluated using 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. Efficiency is measured by calculating the total length of the reconstruction path P and the number of frames N_f . To provide a quantitative assessment of the efficiency of the reconstruction process, we introduce an efficiency metric that combines visual quality and motion effort, defined as $E = \text{PSNR}/\log N_f$. The logarithmic

transformation of the denominator is applied to align with the PSNR calculation.

3) *Simulation Result*: We actively reconstruct three scenes, the *Oil Drum*, the *Drilling Machine*, and the *Potted Plant*, to validate the proposed system. The offline refinement results, including image rendering and depth rendering, are presented in Figure 3. The evaluations of visual quality and efficiency are calculated and summarized in Table II. The *Oil Drum* is characterized by a relatively simple geometry but detailed texture. The *Drilling Machine* exhibits fine geometric structures, while the *Potted Plant* features a highly cluttered geometric structure. The rendering results in Figure 3 demonstrate a detailed visual fidelity with precise geometric structures, highlighting the system’s capability to capture intricate textures and structures.

B. Comparison Study of Active View Selection

To evaluate the efficiency of the proposed GauSS-MI metric in selecting the next-best-view for high visual quality reconstruction, we conduct comparative experiments on active view selection using a fixed number of frames.

1) *Baselines*: We benchmark our method against FisherRF [8]³, a state-of-the-art radiance field-based active view selection approach that quantifies the expected information gain by constructing the Fisher information matrix. To ensure a fair comparison, FisherRF is integrated into our system by substituting the GauSS-MI evaluation I in (23) with its FisherRF metric. Additionally, a random view selection policy is implemented as a baseline to highlight the benefits of using view selection strategies.

2) *Results*: The comparative experiment is performed across three scenes, with the number of frames limited and gradually increased for each method. We compute the PSNR values for each test and visualize the results by plots in Figure 4. The results show that both GauSS-MI and FisherRF significantly outperform the random policy, demonstrating the methods’ effectiveness in next-best-view selection for enhancing visual quality. While the performance of GauSS-MI and FisherRF is comparable, GauSS-MI achieves higher PSNR values in most tests, validating its superior efficiency in active view selection. The novel view synthesis results for GauSS-MI, FisherRF, and the random policy, with a fixed number of frames $N_f = 200$, are presented alongside the ground truth on the left-hand side of Figure 5. These visualizations further showcase the enhanced visual fidelity reconstruction result of GauSS-MI, particularly in scenes featuring complex geometric or textural details. The efficiency on active view selection is especially advantageous for on-board active reconstruction, where constrained computational and battery resources necessitate minimizing the number of frames and reconstruction time.

C. Computational Efficiency

We analyze the computational complexity of the proposed GauSS-MI method, measure its average runtime, and com-

²<https://unity.com/>

³FisherRF: <https://github.com/JiangWenPL/FisherRF>

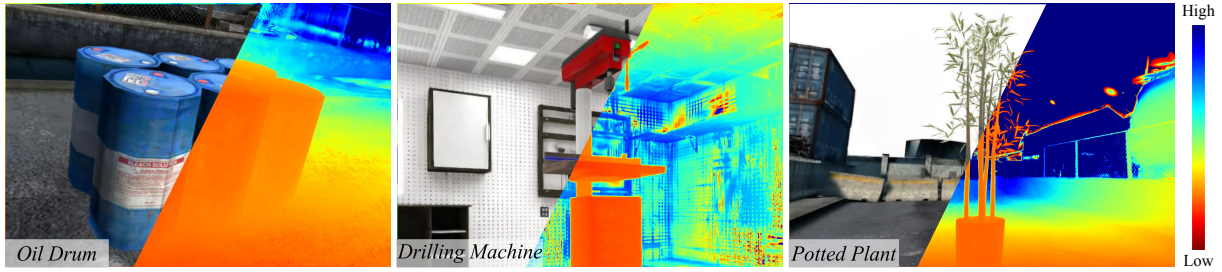


Fig. 3. High-resolution novel view synthesis of the reconstruction result by the proposed system: color rendering against depth rendering.

TABLE II
EVALUATION RESULTS AND COMPARISON OF SIMULATION EXPERIMENTS

Scene ¹	Oil Drum ²						Drilling Machine ²						Potted Plant ²					
Metric	Visual Quality			Efficiency			Visual Quality			Efficiency			Visual Quality			Efficiency		
Method	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	$N_f \downarrow$	$P(m) \downarrow$	$E \uparrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	$N_f \downarrow$	$P(m) \downarrow$	$E \uparrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	$N_f \downarrow$	$P(m) \downarrow$	$E \uparrow$
Ours	34.35	0.986	0.068	141	61.04	16.0	33.99	0.995	0.040	122	36.16	16.3	30.33	0.986	0.084	200	79.60	13.2
FUEL [3]	22.82	0.915	0.186	165	15.21	10.3	21.08	0.967	0.116	145	11.16	9.8	25.39	0.963	0.149	205	17.28	11.0
NARUTO [16]	31.84	0.976	0.072	3000	116.34	9.2	31.50	0.992	0.047	3000	92.35	9.1	30.83	0.988	0.057	4000	157.75	8.6

¹ Simulation scenes are built by Flightmare [15].

² Oil drum scene size: $5m \times 4m \times 3m$. Drilling Machine scene size: $4m \times 4m \times 3m$. Potted Plant scene size: $5m \times 5m \times 5m$.

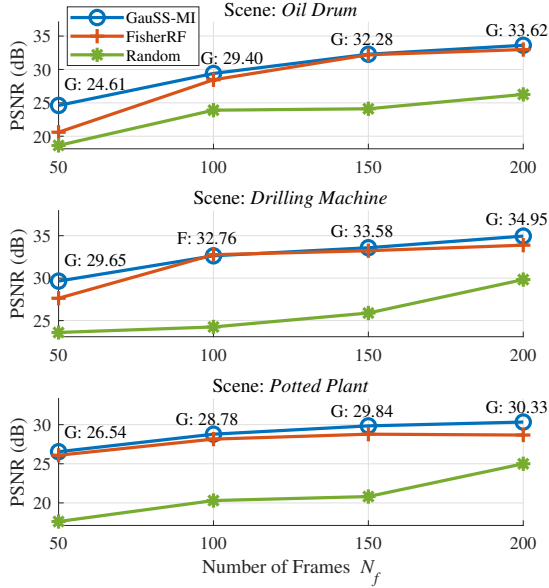


Fig. 4. PSNR results for active view selection with a limited number of frames. The maximum PSNR value for each test is annotated. The abbreviations 'G' and 'F' denote GauSS-MI and FisherRF, respectively.

pare it with FisherRF [8], validating the real-time performance of our metric.

1) *Computational Complexity*: The computation of GauSS-MI is similar to 3DGS rasterization in that Eq. (20) projects the information gain function (19) onto an image. The algorithm is implemented in parallel using CUDA. Assuming the current 3DGS map with N_g Gaussians, the image with N_p pixels, and N_c candidate viewpoints to be evaluated, the computational complexity of GauSS-MI is $O(N_p N_g N_c)$. In contrast, FisherRF's complexity depends on both candidate and observed views. With N_o observed views, FisherRF requires a complexity of $O(N_p N_g (N_o + N_c))$ to evaluate all candidates, as it has to compute the

information from both observed and candidate views at each decision step. Consequently, the computational cost scales linearly with $N_o + N_c$, indicating the increasing runtime as active reconstruction progresses. GauSS-MI, however, maintains consistent computation, scaling linearly with only N_c . This efficiency stems from our probabilistic model, which quantifies the information from prior observations with a low computational overhead of $O(2N_p N_g)$ during the map update process. As a result, in the next-best-view decision step, GauSS-MI evaluates only candidate views, achieving low and stable computational complexity, making it ideal for real-time applications.

2) *Runtime*: We conducted a complete active reconstruction experiment to measure the runtime of each method at each planning timestep, as shown in Figure 6. GauSS-MI achieves an average runtime of 5.55 ms (182.2 fps), while FisherRF averages 11.66 ms (85.8 fps). These results corroborate the computational complexity analysis that GauSS-MI maintains consistent runtime throughout the reconstruction process, whereas FisherRF's runtime increases due to its dependence on the growing number of observed views.

D. Comparison Study of Active Reconstruction

This section evaluates and compares the complete system, including the proposed view planning and active termination condition. We select the state-of-the-art baselines employing different map representations and uncertainty quantification techniques to validate our system's efficiency on visual quality.

1) *Baselines*: To evaluate the efficacy of our proposed method, we conducted a comparative analysis between our active reconstruction system and existing systems, FUEL [3] and NARUTO [16]. FUEL is a volumetric-based active reconstruction system with no consideration of visual quality, while NARUTO is a NeRF-based framework that addresses radiance field uncertainty with a focus on geometry. For

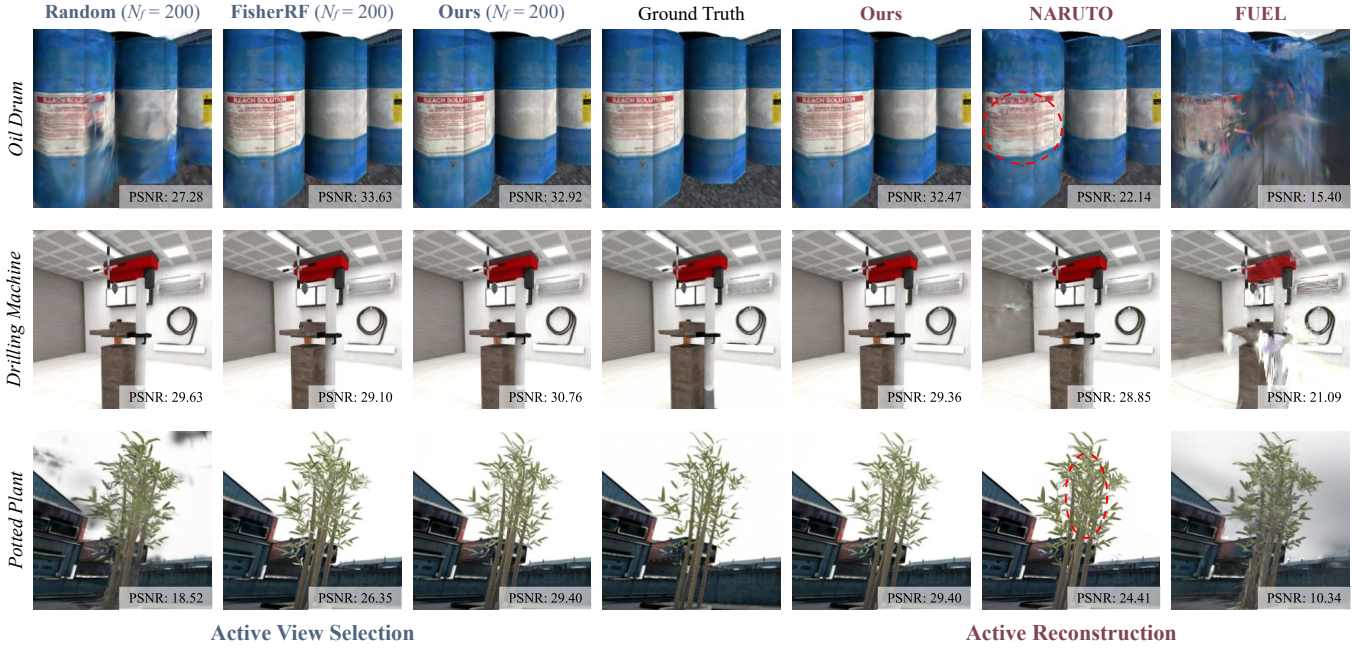


Fig. 5. Novel view synthesis results compared to ground truth.

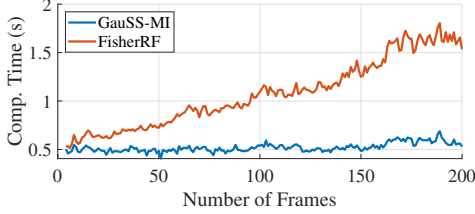


Fig. 6. Comparison of computation time. Statistics in a complete active reconstruction process.

our study, we implemented the comparison using the open-source codes for FUEL and NARUTO, employing their default parameter settings. Each system, including the next best view selection and path planning algorithm, captured color images in the three simulation scenes, which are subsequently employed to 3D Gaussian Splatting [2] for offline model reconstruction. Evaluation of reconstruction quality and efficiency was conducted using the metrics outlined in Section IV-A.2.

2) *Results*: The quantitative results are presented in Table II, while the qualitative visual comparisons are shown in the right part of Figure 5. Our system demonstrates superior efficiency across all scenes and attains the highest visual quality in the *Oil Drum* and *Drilling Machine*. In the *Potted Plant* scene, NARUTO slightly outperforms our system by a small margin. However, it is worth noting that NARUTO completed its reconstruction process after capturing thousands of images, which contributed to its commendable reconstruction performance. The extensive collection of images is attributed to NARUTO’s continuous high-frequency image capture throughout its movement. The abundance of images with significant overlap resulted in a lower active viewpoint selection efficiency, indicating an inadequate assessment of observed information and a suboptimal reconstruction strat-

egy. In contrast, our system efficiently selects viewpoints guided by GauSS-MI. As a result, we achieve comparable or even superior visual quality to NARUTO while maintaining consistently high efficiency.

In terms of total path length for active reconstruction, FUEL stands out for completing the process with a notably shorter trajectory compared to both our system and NARUTO. This outcome aligns with expectations, given that FUEL focuses solely on geometric completeness during active reconstruction. However, despite its efficiency in path length, FUEL consistently yields the lowest visual quality and the reconstruction results exhibit poor texture quality, as illustrated in Figure 5. This indicates the inadequacy of relying solely on geometric evaluation for high-quality visual reconstructions.

Overall, our system excels in active efficiency while simultaneously delivering high visual quality across all scenes. This demonstrates the effectiveness of our probabilistic model in evaluating observed information and the capability of GauSS-MI in identifying optimal viewpoints to enhance efficiency.

V. REAL-WORLD EXPERIMENTS

To validate the efficacy of the proposed method in practical settings, we conduct real-world experiments using a Franka Emika Panda robotic arm equipped with an Intel RealSense D435 depth camera for capturing RGB-D images. The real-world setup is shown in Figure 7(a). The active reconstruction system for real-world implementation integrates the on-line 3DGS reconstruction algorithm with the proposed active view sampling and selection method. Motion planning and control for the Franka arm are implemented using the MoveIt

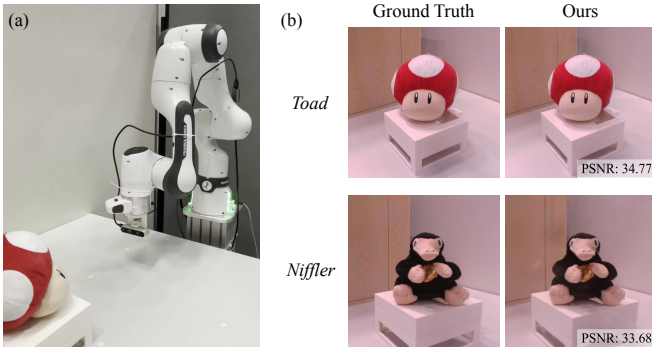


Fig. 7. Active reconstruction experiment with GauSS-MI in the real world. (a) Experiment setup. (b) Novel view synthesis results.

TABLE III

EVALUATION RESULTS OF REAL-WORLD EXPERIMENTS					
Metric	Visual Quality			Efficiency	
Scene	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	$N_f \downarrow$	$E \uparrow$
<i>Toad</i>	32.53	0.9336	0.2693	24	23.6
<i>Niffler</i>	28.20	0.9273	0.3020	36	18.1

framework⁴, facilitating precise pose control and reliable feedback. All algorithms are executed on a desktop equipped with a 32-core Intel i9-13900K CPU and an NVIDIA RTX 4090 GPU.

For the real-world demonstration, we actively reconstructed two scenes: the *Toad* and the *Niffler*. The *Toad* scene features relatively smooth surfaces, whereas the *Niffler* exhibits intricate geometric details. Constrained by the robot arm’s workspace and the minimal detection range of the depth camera, the experimental scenes are limited to the size of $0.2\text{m} \times 0.2\text{m} \times 0.2\text{m}$. The novel view synthesis results, presented in Figure 7(b), demonstrate the high visual fidelity achieved by our method. Quantitative evaluation of visual quality and reconstruction efficiency is summarized in Table III. The results demonstrate the effectiveness and efficiency of the proposed system in the real world, highlighting its robustness across different scene complexities.

VI. CONCLUSION AND FUTURE WORK

This paper addresses a critical challenge in active reconstruction—active view selection—with a focus on enhancing visual quality. We first introduce an explicit probabilistic model to quantify the uncertainty of visual quality, leveraging 3D Gaussian Splatting as the underlying representation. Building on this, we propose Gaussian Splatting Shannon Mutual Information (GauSS-MI), a novel algorithm for real-time assessment of mutual information between measurements from a novel viewpoint and the existing map. GauSS-MI is employed to facilitate the active selection of the next best viewpoints and is integrated into an active reconstruction system to evaluate its effectiveness in achieving high visual fidelity in 3D reconstruction. Extensive experiments across various simulated environments and real-world scenes demonstrate the system’s ability to deliver superior visual

quality over state-of-the-art methods, validating the effectiveness of the proposed approach.

ACKNOWLEDGEMENT

The authors gratefully acknowledge Ruixing Jia and Run-dong Li for their insightful and valuable discussions. We also thank the anonymous reviewers for their constructive and thoughtful feedback, which greatly enhanced this manuscript.

REFERENCES

- [1] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, “Nerf: Representing scenes as neural radiance fields for view synthesis,” *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.
- [2] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, “3d gaussian splatting for real-time radiance field rendering,” *ACM Transactions on Graphics*, vol. 42, no. 4, July 2023. [Online]. Available: <https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/>
- [3] B. Zhou, Y. Zhang, X. Chen, and S. Shen, “FUEL: Fast uav exploration using incremental frontier structure and hierarchical planning,” *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 779–786, 2021.
- [4] W. Tabib, K. Goel, J. Yao, C. Boirum, and N. Michael, “Autonomous cave surveying with an aerial robot,” *IEEE Transactions on Robotics*, vol. 38, no. 2, pp. 1016–1032, 2021.
- [5] S. Isler, R. Sabzevari, J. Delmerico, and D. Scaramuzza, “An information gain formulation for active volumetric 3d reconstruction,” in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 3477–3484.
- [6] C. Cao, J. Zhang, M. Travers, and H. Choset, “Hierarchical coverage path planning in complex 3d environments,” in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 3206–3212.
- [7] L. Goli, C. Reading, S. Sellán, A. Jacobson, and A. Tagliasacchi, “Bayes’ rays: Uncertainty quantification for neural radiance fields,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 20 061–20 070.
- [8] W. Jiang, B. Lei, and K. Daniilidis, “Fisherrf: Active view selection and mapping with radiance fields using fisher information,” in *European Conference on Computer Vision*. Springer, 2025, pp. 422–440.
- [9] H. Matsuki, R. Murai, P. H. J. Kelly, and A. J. Davison, “Gaussian Splatting SLAM,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.
- [10] B. J. Julian, S. Karaman, and D. Rus, “On mutual information-based control of range sensing robots for mapping applications,” *The International Journal of Robotics Research*, vol. 33, no. 10, pp. 1375–1392, 2014.
- [11] T. M. Cover and J. A. Thomas, “Information theory and the stock market,” *Elements of Information Theory*. Wiley Inc., New York, pp. 543–556, 1991.
- [12] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, “Practical poissonian-gaussian noise modeling and fitting for single-image raw-data,” *IEEE transactions on image processing*, vol. 17, no. 10, pp. 1737–1754, 2008.
- [13] X. Yang, K. Sreenath, and N. Michael, “A framework for efficient teleoperation via online adaptation,” in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 5948–5953.
- [14] M. W. Mueller, M. Hehn, and R. D’Andrea, “A computationally efficient motion primitive for quadcopter trajectory generation,” *IEEE transactions on robotics*, vol. 31, no. 6, pp. 1294–1310, 2015.
- [15] Y. Song, S. Naji, E. Kaufmann, A. Loquercio, and D. Scaramuzza, “Flightmare: A flexible quadrotor simulator,” in *Proceedings of the 2020 Conference on Robot Learning*, 2021, pp. 1147–1157.
- [16] Z. Feng, H. Zhan, Z. Chen, Q. Yan, X. Xu, C. Cai, B. Li, Q. Zhu, and Y. Xu, “Naruto: Neural active reconstruction from uncertain target observations,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 21 572–21 583.

⁴<https://github.com/moveit/moveit>