

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

Yuhan Xie*, Yixi Cai[†], Yinqiang Zhang*, Lei Yang^{‡§}, and Jia Pan^{*§}

* School of Computing and Data Science, The University of Hong Kong, Hong Kong SAR, China

Email: {yuh anxie, zyzq507}@connect.hku.hk, jpan@cs.hku.hk

[†] Division of Robotics, Perception, and Learning, KTH Royal Institute of Technology, Stockholm, Sweden

Email: yixica@kth.se

[‡] Faculty of Engineering, The University of Hong Kong, Hong Kong SAR, China

Email: lyang125@hku.hk

[§] Centre for Transformative Garment Production, Hong Kong SAR, China

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[26, 19], manipulation[38], robotics[23, 43], construction[41], etc. Recent advancements, such as Neural Radiance Field (NeRF)[26] and 3D Gaussian Splatting (3DGS)[19], 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 [43, 17, 34]. 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 [14, 43, 34] or assessing surface coverage completeness [1, 8]. 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[19, 26], recent works have attempted to quantify visual uncertainty to directly evaluate the visual quality of reconstruction models [31, 11, 15].

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

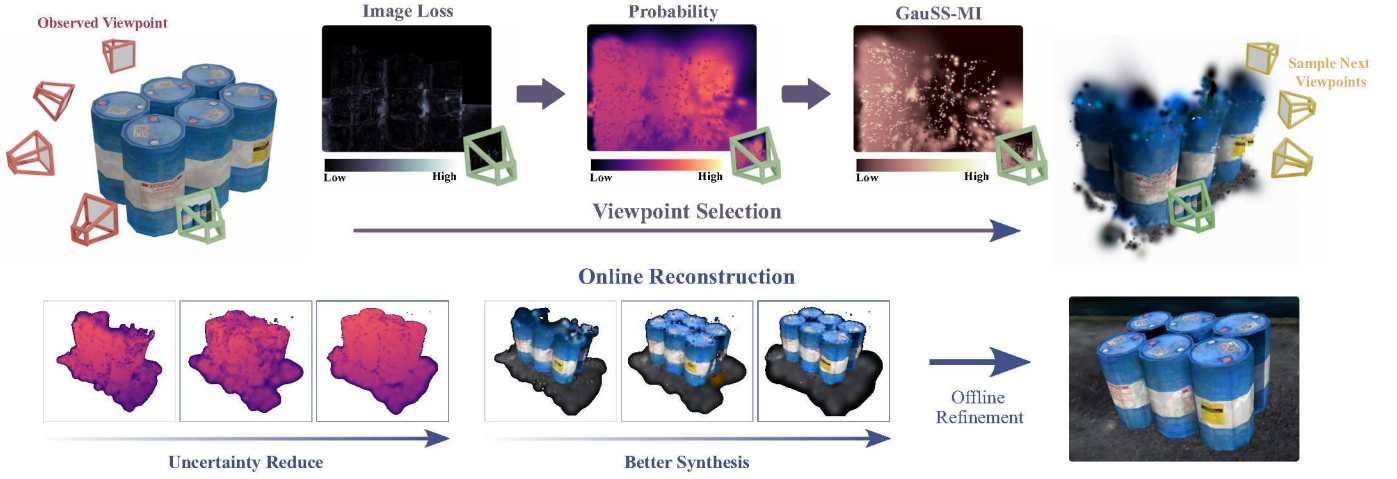


Fig. 1. Illustration of the proposed Gaussian Splatting Shannon Mutual Information (GauSS-MI) method. **Upper part:** At each active reconstruction step, once a new observation is obtained, the 3D Gaussian Splatting (3DGS) map is updated and optimized by minimizing the image loss between observed images and the map. To quantify visual uncertainty, we construct a probabilistic model for each 3D Gaussian ellipsoid by mapping residual image loss onto the 3DGS map. Using this model, we define GauSS-MI, a metric that estimates mutual information between the reconstruction model and a viewpoint. GauSS-MI enables real-time visual quality assessment from novel viewpoints without a prior, facilitating the selection of the next-best-view to effectively reduce map uncertainty. **Lower part:** The active reconstruction process iterates and decreases visual uncertainty, resulting in a high visual fidelity 3D reconstruction result.

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 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. RELATED WORK

The evolution of mapping representations in 3D reconstruction has driven significant advancements in active reconstruction methodologies. A key distinction between active

and passive reconstruction lies in the process of active view selection. In this section, we first review active view selection strategies across various mapping representations. We then present a detailed review of uncertainty quantification techniques employed in information-based approaches.

A. Active View Selection for 3D reconstruction

The first branch of research in active 3D reconstruction focuses on geometric reconstruction, utilizing occupancy-based representations [12]. In this domain, a commonly employed strategy for determining the next best viewpoint involves constructing and evaluating frontiers, which indicate the boundary between mapped and unmapped areas[36, 5, 43]. Additionally, researchers have incorporated information-theoretic approaches, considering metrics such as information gain [14, 22] and mutual information [42, 18, 34], to maximize the information observed at subsequent viewpoints. Recent studies have also explored surface coverage in active reconstruction, with or without prior knowledge of the environment[7, 8, 40], to enhance reconstruction efficiency. However, methods within this branch mainly rely on occupancy-based map information, which does not inherently ensure high visual quality in the resulting reconstructions.

Recent advancements in radiance field representations, such as Neural Radiance Fields (NeRF) [26] and 3D Gaussian Splatting (3DGS) [19], have significantly enhanced visual quality in 3D reconstruction, sparking interest in their application to active reconstruction for even higher visual fidelity. However, early research primarily rely on geometric information derived from occupancy maps for viewpoint selection, rather than directly leveraging the rich visual information inherent in radiance field maps. For instance, Yan et al. [37] explored geometric completeness in NeRF by evaluating infor-

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

mation gain based on volumetric data. Li et al. [21] introduced ActiveSplat, a method that optimizes environment coverage to achieve high visual quality using 3DGS. In NARUTO [9], Feng et al. considered implicit uncertainty in geometric information for active reconstruction with 3DGS.

Considering the implicit model in NeRF, researchers have explored neural network-based approaches for evaluating implicit uncertainty in visual quality [20, 29, 30, 31]. However, while these methods successfully evaluate uncertainty in NeRF for next viewpoint selection, their effectiveness depends heavily on the availability of high-quality datasets for training the evaluation neural networks.

In contrast to NeRF’s implicit modeling approach, 3DGS provides an explicit representation of the environment using a collection of spatially distributed spherical Gaussians. This explicit representation facilitates the direct evaluation of uncertainty in visual quality. In GS-Planner [17], image loss is directly incorporated into the occupancy map, enabling the evaluation of both geometric and photometric uncertainty. Jiang et al. [15] introduced FisherRF, which leverages the Fisher information matrix to quantify the parameter uncertainty in radiance maps. Building on this, Xu et al. [35] extended the GS-Planner framework by integrating FisherRF and geometric completeness for more comprehensive uncertainty evaluation. ActiveGS by Jin et al. [16] proposed a model for evaluating the confidence of each Gaussian, which is subsequently used for viewpoint selection in active reconstruction. In a recent study, Chen et al. [4] proposed ActiveGAMER with a silhouette-based information gain to enhance both geometric and photometric reconstruction accuracy. Our method also employs 3DGS as the scene representation for active reconstruction. To enhance the evaluation of observed information in the Gaussian map, we propose a more comprehensive probabilistic model that accounts for both reconstruction loss and sensor measurement noise.

B. Information-theoretic View Uncertainty Quantification

Information theory offers a robust mathematical framework for active 3D reconstruction, enabling the selection of viewpoints that maximize expected information and consequently reduce the map entropy. Occupancy grid maps, typically represented as probability models [18], inherently capture observed information and associated uncertainties. Consequently, information-theoretic approaches can be directly integrated into occupancy-based active reconstruction. One common method constructs information gain from each observation by occupancy probabilities [14, 22]. Alternatively, mutual information has been extensively studied for its ability to quantify the mutual information between a map and observations. Shannon Mutual Information (SMI) has been proven to provide guarantees for comprehensive exploration of the environment [18], demonstrating its theoretical effectiveness and completeness in active reconstruction tasks. However, its practical application is hindered by substantial computational overhead when applied to occupancy mapping, as the computational complexity scales quadratically with the spatial resolu-

tion of the map and linearly with the numerical integration of range measurements. To overcome these limitations, Charrow et al. [3] introduced the Cauchy-Schwarz Quadratic Mutual Information (CSQMI) metric, which enables analytical computation of measurement integration and reduces computational complexity to linear scaling with the map’s spatial resolution. Subsequent studies have demonstrated the efficiency of CSQMI in real-time robotic systems for active reconstruction tasks [2, 28, 33]. To preserve the theoretical guarantee of SMI, Zhang et al. [42] proposed the Fast Shannon Mutual Information (FSMI) algorithm, which significantly enhances the computational efficiency of mutual information evaluation compared to the original SMI algorithm [18] by analytically evaluating integrals.

Information-theoretic uncertainty quantification for radiance field-based approaches can be broadly divided into two categories. The first category learns an implicit probability model to estimate information gain for novel viewpoints [31]. The second employs the Fisher information matrix, derived from the rendering loss, to quantify information gain [11, 15]. Among these, FisherRF, a recent method based on the Fisher information matrix, extends its applicability to 3D Gaussian Splatting (3DGS) [15]. However, FisherRF focuses primarily on next-best-view (NBV) selection, neglecting the real-time demands of active reconstruction. To overcome these limitations, we propose a probabilistic model for 3DGS based on the rendering quality. Utilizing the computationally efficient SMI method, which jointly accounts for uncertainties in the reconstructed map and measurements, we introduce Gaussian Splatting Shannon Mutual Information (GauSS-MI), a novel method for quantifying visual uncertainties. Additionally, we develop an active reconstruction system based on GauSS-MI, which achieves high visual fidelity with real-time performance requirements.

III. OVERVIEW

This paper introduces Gaussian Splatting Shannon Mutual Information (GauSS-MI) as a metric for efficient next best view selection in high-visual fidelity active reconstruction. The proposed method is illustrated in Figure 1. At each active reconstruction step, we assume that we have a set of previous observations and a set of next viewpoint candidates. Our goal is to devise an effective metric for the next best view selection leveraging the available observed information.

During each active reconstruction step, a new observation is obtained, and the 3D Gaussian Splatting (3DGS) map is updated by extending and initializing new Gaussians based on this new observation. Subsequently, the overall 3DGS map undergoes iterative optimization based on the loss between the observed images and the map. To quantify the visual uncertainty, we compute the remaining image loss between the current observation and the optimized map. By mapping the image loss onto the 3DGS map, we construct a probabilistic model for each 3D Gaussian. Subsequently, based on Shannon Mutual Information theory, we leverage the probabilistic

TABLE I
MAIN NOTATIONS FOR GAUSS-MI

Notations	Explanation
$\mathcal{G}$	3D Gaussian splatting map.
$\mathcal{W}$	The world frame.
$\mathcal{N}$	A series of ordered Gaussians along a camera ray.
μ	Position of a Gaussian.
c	Color of a Gaussian.
α	Opacity of a Gaussian.
σ	Camera pose, or a viewpoint.
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.
m, M	Random variable and realization of luminance for a pixel.
$P(r)$	Real probability of a Gaussian.
o, l	Odds ratio and log odds of a Gaussian.
δ	Inverse sensor model.
L	Loss image between the map and the observation.
λ	Hyperparameters.
H	Entropy of the map.
I	Mutual Information between the map and the observation.
$(\cdot)^{[i]}$	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 .

model to establish the mutual information between the reconstruction model and a viewpoint. This mutual information function is referred to as Gaussian Splatting Shannon Mutual Information (GauSS-MI), enabling real-time visual quality assessment from novel viewpoints without a prior. The next best view is then selected using GauSS-MI, to effectively reduce the uncertainty of current map. The iterative process leads to a decrease in visual uncertainty within the reconstruction model, yielding a high visual fidelity 3D reconstruction result.

The derivation of GauSS-MI is elaborated in Section IV, and the system implementation details are presented in Section V.

IV. 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. The main notations of this section 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 $\mu_{\mathcal{W}}^{[i]}$ 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 [24]. 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.,

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

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

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

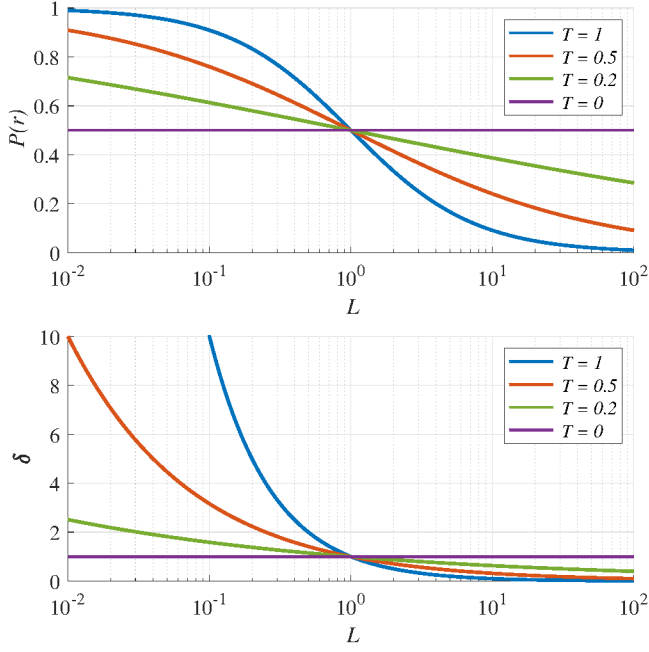


Fig. 2. Inverse sensor model visualization. The hyperparameters λ are omitted in the figure for simplicity.

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. We further visualize the inverse sensor model (7)(8) in Figure 2 to illustrate the probability update. The Gaussians associated with observation Z_k have $T^{[i]} > 0$, resulting in the inverse sensor model $P(r^{[i]}|Z_k)$ and $\delta^{[i]}(Z_k)$ being monotonically decreasing with loss L_k . This suggests that lower loss L_k corresponds to a higher *reliable* probability of Gaussian $P(r^{[i]}|Z_k)$. Additionally, a lower cumulative transmittance $T^{[i]}$ implies less impact of Gaussian i on observation Z_k . Consequently, a smaller $T^{[i]}$ results in less observed information within the inverse sensor model, e.g., when $T^{[i]} = 0$, we have $P(r^{[i]}|Z_k) = 0.5$ and $\delta^{[i]}(Z_k) = 1$.

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

Algorithm 1 Probability Update

Require: 3DGS map $\mathcal{G}$; Observations $\hat{C}_k, \hat{D}_k$ from σ_k

- 1: $\mathcal{N} \leftarrow \text{SortGaussians}(\mu, \sigma_k)$
- 2: $C, D \leftarrow \text{ImageRender}(c, \alpha, \mu, \mathcal{N})$
- 3: $L_k = \lambda_c \|C - \hat{C}_k\| + (1 - \lambda_c) \|D - \hat{D}_k\|$ $\triangleright$ Loss image
- 4: $l_k = 0$
- 5: **for** $j \leftarrow 1$ to n_z **do** $\triangleright$ Per pixel
- 6: $T^{[j]} = 1$
- 7: **for** $i \leftarrow 1$ to $\mathcal{N}^{[j]}$ **do** $\triangleright$ Per gaussian
- 8: $T^{[i]} = \alpha^{[i]} T^{[j]}$ $\triangleright$ Gaussian's transmittance
- 9: $l_k^{[i]} = -\lambda_T T^{[i]} \log \lambda_L L_k$ $\triangleright$ Log odds of (7)
- 10: $T^{[j]} = (1 - \alpha^{[j]}) T^{[j]}$ $\triangleright$ Update pixel transmittance
- 11: **end for**
- 12: **end for**
- 13: $l_{1:k} = l_k + l_{1:k-1}$
- 14: $P_{1:k} \leftarrow \text{InvertLogOdds}(l_{1:k})$ $\triangleright$ Update probability
- 15: **return** $P_{1:k}$

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). The probability update algorithm is summarized in Algorithm 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 $j \in \{1, \dots, n_z\}$ [18].

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

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

From information theory [6, 18], 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 detailed derivation is presented in the Appendix A. 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 [10], 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

Algorithm 2 GauSS-MI

Require: 3DGS map $\mathcal{G}$; Novel view σ

```

1:  $I = 0$ 
2:  $\mathcal{N} \leftarrow \text{SortGaussians}(\mu, \sigma)$ 
3: for  $j \leftarrow 1$  to  $n_z$  do ▷ Per pixel
4:   // Rasterize color and  $f$  (19) on each pixel
5:    $C^{[j]} = 0$ ;  $f^{[j]} = 0$ ;  $T^{[j]} = 1$ 
6:   for  $i \leftarrow 1$  to  $\mathcal{N}^{[j]}$  do ▷ Per gaussian
7:      $T^{[i]} = \alpha^{[i]} T^{[j]}$  ▷ Gaussian's transmittance
8:      $C^{[j]} += c^{[i]} T^{[i]}$  ▷ Rasterize color
9:      $f^{[j]} -= \log(P^{[i]}) T^{[i]}$  ▷ Rasterize (19)
10:     $T^{[j]} = (1 - \alpha^{[i]}) T^{[j]}$  ▷ Update pixel transmittance
11:  end for
12:   $M^{[j]} \leftarrow \text{Color2Luminance}(C^{[j]})$ 
13:   $P(z^{[j]} | M^{[j]}) \leftarrow \text{SensorModel}(M^{[j]})$ 
14:   $I += P(z^{[j]} | M^{[j]}) f^{[j]}$  ▷ Update MI
15: end for
16: return  $I$ 

```

$\mathcal{G}$. 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

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

The computation procedure of GauSS-MI is summarized in Algorithm 2.

V. SYSTEM IMPLEMENTATION

This section details the system implementation of GauSS-MI.

A. System Overview

The proposed active reconstruction system comprises a reconstruction module and a planning module, as illustrated in Figure 3. 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

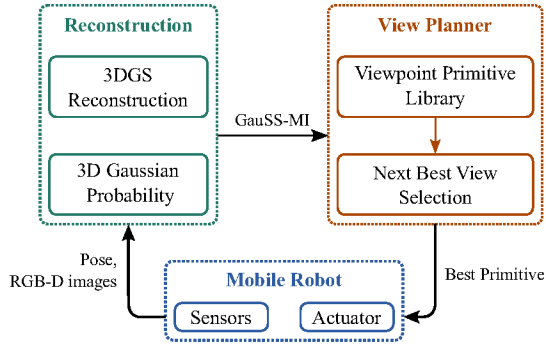


Fig. 3. Overview of proposed active 3D reconstruction system.

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.

We then present the view planner for viewpoint sampling and selection and discuss the autonomous termination condition design for the proposed 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 [39], 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 ,

$$\sigma_f = \sigma_0 + \begin{bmatrix} -v_{xy}T \sin(\psi_0 + \omega_z T) \\ v_{xy}T \cos(\psi_0 + \omega_z T) \\ v_z T \\ \omega_z T \end{bmatrix} \quad (22)$$

$$\sigma_f^{(n)} = \mathbf{0} \quad \text{for } n = 1, 2, 3$$

Algorithm 3 View and Motion Planner

Require: Full Quadrotor States $\mathcal{X}(t)$; Action Space $\mathcal{A}$

```

1:  $R = 0$ 
2: for  $\alpha \in \mathcal{A}$  do
3:    $\sigma(t+T) \leftarrow \text{ForwardPropagate}(\sigma(t), \alpha)$ 
4:    $\sigma_T \leftarrow \text{MotionPrimitive}(\mathcal{X}(t), \sigma(t+T))$ 
5:   if  $\text{SafetyCheck}(\sigma_T)$  then
6:      $I \leftarrow \text{GauSS\_MI}(\sigma(t+T))$  ▷ Algorithm2
7:      $J \leftarrow \text{MotionCost}(\sigma_T)$ 
8:     if  $R < w_I I - w_J J$  then
9:        $R = w_I I - w_J J$ 
10:       $\sigma_T^* = \sigma_T$ 
11:    end if
12:  end if
13: end for
14: return  $\sigma_T^*$ 

```

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 [27] (detailed in the Appendix B).

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

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

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. The complete procedure for view and motion planning is summarized in Algorithm 3.

C. Termination Condition

A spherical Gaussian may be deemed rendering reliable from one perspective, but this reliability may not hold from another perspective. Specifically, the *reliable* probability of a Gaussian $P(r^{[i]})$ should exhibit anisotropic behavior. In the implementation, to address this issue, we update $P(r^{[i]})$ from four orthogonal horizontal perspectives. The visual reconstruction completeness of a Gaussian is quantified based on the average *reliable* probability denoted as μ_P .

At the beginning of the mapping process, we assume no prior information about the environment. Consequently, we initialize the probabilities of all 3D Gaussians as (5). As detailed in Section IV-B, the probability $P(r)$ for a spherical Gaussian decreases if it renders out a relatively large image loss. As the reconstruction process progresses, we expect a decrease in the render-ground truth loss and an increase in $P(r)$.

TABLE II
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%

A Gaussian is considered completely reconstructed when its average probability exceeds a specified threshold, $\mu_P > \tau$. The active reconstruction process is regarded complete and actively terminated once the proportion of fully reconstructed Gaussians reaches a predefined percentage threshold,

$$\frac{N_{\text{done}}}{N_{\text{GS}}} > \varphi \quad (24)$$

where N_{done} represents the number of completely reconstructed Gaussians, and N_{GS} denotes the total number of Gaussians in the map.

VI. 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 VI-A). Subsequently, we conduct experiments to evaluate the proposed GauSS-MI metric from multiple perspectives: the efficiency of next-best-view selection (Section VI-B), real-time computational performance (Section VI-C), and the effectiveness of uncertainty quantification (Section VI-D). Finally, we compare the complete system against baseline methods in Section VI-E and study the termination condition for the system in Section VI-F.

A. System Validation

1) *Simulation Setup*: The simulation environment is created using Flightmare[32], 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[24], 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 II.

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 4. The evaluations of visual quality and efficiency are calculated and summarized in Table III. 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 4 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[15]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 5. 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

²<https://unity.com/>

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

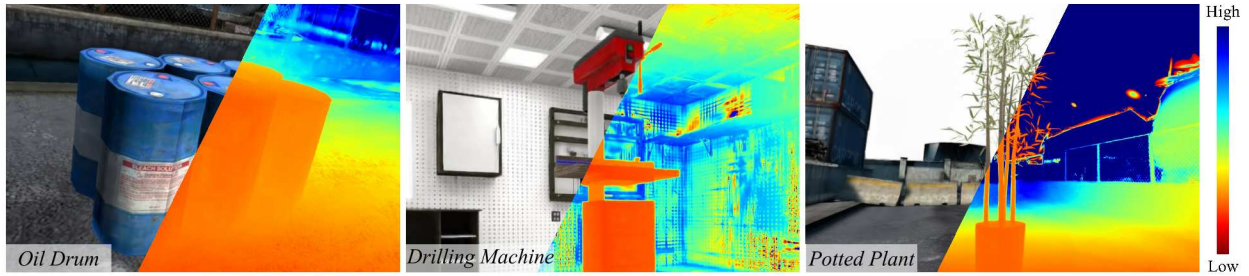


Fig. 4. High-resolution novel view synthesis of the reconstruction result by the proposed system: color rendering against depth rendering.

TABLE III
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↑	SSIM↑	LPIPS↓	N_f ↓	$P(m)$ ↓	E ↑	PSNR↑	SSIM↑	LPIPS↓	N_f ↓	$P(m)$ ↓	E ↑	PSNR↑	SSIM↑	LPIPS↓	N_f ↓	$P(m)$ ↓	E ↑
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 [43]	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 [9]	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 [32].

² Oil drum scene size: 5m × 4m × 3m. Drilling Machine scene size: 4m × 4m × 3m. Potted Plant scene size: 5m × 5m × 5m.

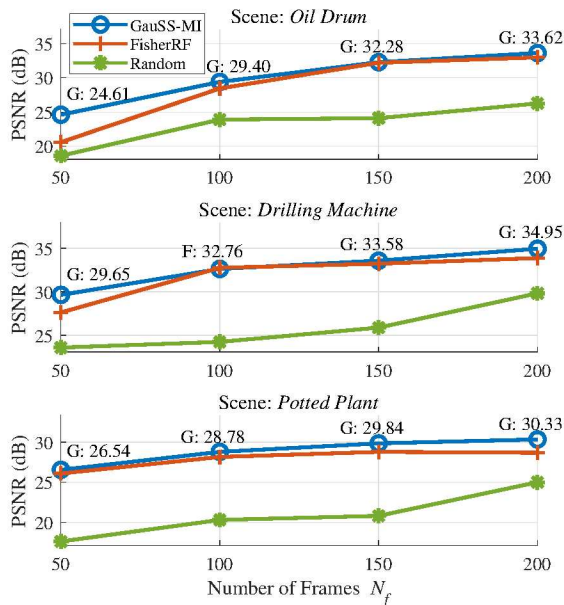


Fig. 5. PSNR results for active view selection with a limited number of frames. The maximum PSNR value for each test is annotated. The abbreviations 'G' and 'F' denote GauSS-MI and FisherRF, respectively.

on the left-hand side of Figure 6. 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 onboard 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 compare it with FisherRF [15], validating the real-time performance of our metric.

1) *Computational Complexity*: The computation of GauSS-MI, as detailed in Algorithm 2, 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 (Algorithm 1). 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 7. GauSS-MI achieves an average runtime of 5.55 ms (182.2 fps), while FisherRF averages 11.66 ms (85.8 fps). These results corrob-

TABLE IV
AREA UNDER SPARSIFICATION ERROR (AUSE) RESULTS

Scene	Oil Drum	Drilling Machine	Potted Plant
Ours	0.264	0.498	0.351
FisherRF [15]	0.276	0.605	0.392

E. 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 [43] and NARUTO [9]. 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 our study, we implemented the comparison using the open-source codes for FUEL⁴ and NARUTO⁵, 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 [19] for offline model reconstruction. Evaluation of reconstruction quality and efficiency was conducted using the metrics outlined in Section VI-A2.

2) *Results*: The quantitative results are presented in Table III, while the qualitative visual comparisons are shown in the right part of Figure 6. 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 strategy. 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.

⁴FUEL: <https://github.com/HKUST-Aerial-Robotics/FUEL>

⁵NARUTO: <https://github.com/oppo-us-research/NARUTO>

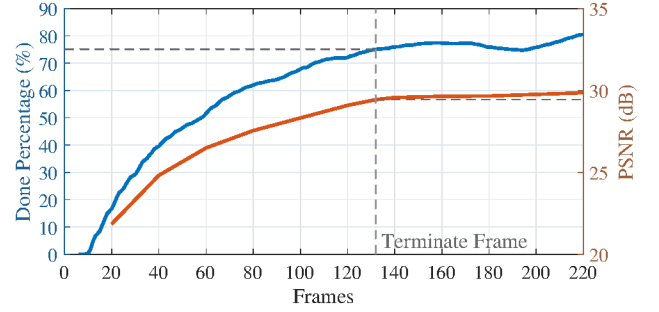


Fig. 9. Terminate Condition Validation

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 6. 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.

F. Termination Study

We additionally discuss our termination condition in the active reconstruction process. We statistic the PSNR and the percentage of reconstructed Gaussians along the active reconstruction process in Figure 9, where the “Done Percentage” refers to $N_{\text{done}}/N_{\text{GS}}$ in (24) and the PSNR is calculated with offline refinement. The figure shows that once 75% of the Gaussians in the map are fully reconstructed, the PSNR reaches 29.44, with minimal further improvement even as the reconstruction process continues. This may result from the limitations of the online 3DGS reconstruction algorithm, which is hard to achieve higher visual quality under the constraints of real-time mapping.

VII. 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 10(a). The active reconstruction system for real-world implementation integrates the online 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 framework⁶, 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.

⁶<https://github.com/moveit/moveit>

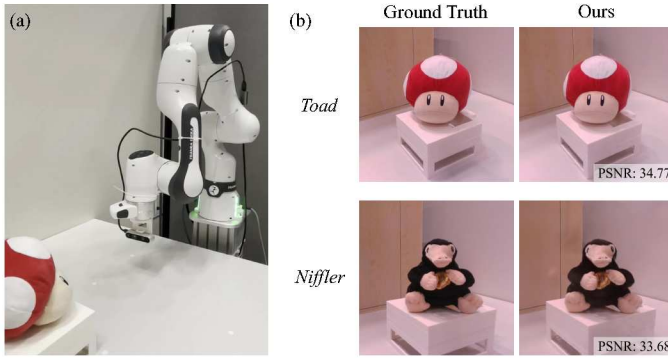


Fig. 10. Active reconstruction experiment with GauSS-MI in the real world. (a) Experiment setup. (b) Novel view synthesis results.

TABLE V
EVALUATION RESULTS OF REAL-WORLD EXPERIMENTS

Metric	Visual Quality			Efficiency	
	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

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 10(b), demonstrate the high visual fidelity achieved by our method. Quantitative evaluation of visual quality and reconstruction efficiency is summarized in Table V. The results demonstrate the effectiveness and efficiency of the proposed system in the real world, highlighting its robustness across different scene complexities.

VIII. LIMITATIONS

In this section, we discuss the limitations of our study.

Firstly, the performance of our active reconstruction system is closely tied to the rendering quality of the online 3DGS framework. Due to real-time computation constraints, the online 3DGS algorithm is hard to match the high rendering quality achievable in offline settings. Future research could focus on enhancing real-time 3DGS reconstruction algorithms for active systems, potentially leading to significant improvements in reconstruction results.

Furthermore, the determination of the termination threshold φ for achieving high visual quality was made heuristically in our study, without a systematic investigation. A more thorough analysis of how the termination condition, scene complexity, and reconstruction fidelity are interrelated would be valuable for refining termination criteria in future studies.

Lastly, although we have implemented the proposed algorithm on an agile and versatile quadrotor, the viewpoints are currently limited to yaw variations (rotation on the vertical

z body axis) only. Roll and pitch movements have not been taken into account. The whole-body movements requires comprehensive planning on the observation viewpoints and the quadrotor motion.

IX. 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.

Future research will first focus on addressing the limitations outlined in Section VIII. Beyond this, we aim to extend our work from simulation to real-world deployment on drones. This transition introduces additional challenges, such as constrained onboard computational resources.

ACKNOWLEDGEMENT

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. The authors gratefully acknowledge Ruixing Jia and Rundong 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] Chao Cao, Ji Zhang, Matt Travers, and Howie Choset. Hierarchical coverage path planning in complex 3d environments. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3206–3212. IEEE, 2020.
- [2] Benjamin Charrow, Gregory Kahn, Sachin Patil, Sikang Liu, Ken Goldberg, Pieter Abbeel, Nathan Michael, and Vijay Kumar. Information-theoretic planning with trajectory optimization for dense 3d mapping. In *Robotics: Science and Systems*, volume 11, pages 3–12, 2015.
- [3] Benjamin Charrow, Sikang Liu, Vijay Kumar, and Nathan Michael. Information-theoretic mapping using cauchy-schwarz quadratic mutual information. In *2015*

- IEEE International Conference on Robotics and Automation (ICRA)*, pages 4791–4798. IEEE, 2015.
- [4] Liyan Chen, Huangying Zhan, Kevin Chen, Xiangyu Xu, Qingan Yan, Changjiang Cai, and Yi Xu. Activegamer: Active gaussian mapping through efficient rendering. *arXiv preprint arXiv:2501.06897*, 2025.
 - [5] Titus Cieslewski, Elia Kaufmann, and Davide Scaramuzza. Rapid exploration with multi-rotors: A frontier selection method for high speed flight. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2135–2142. IEEE, 2017.
 - [6] 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.
 - [7] Chen Feng, Haojia Li, Fei Gao, Boyu Zhou, and Shaojie Shen. Predrecon: A prediction-boosted planning framework for fast and high-quality autonomous aerial reconstruction. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1207–1213. IEEE, 2023.
 - [8] 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.
 - [9] Ziyue Feng, Huangying Zhan, Zheng Chen, Qingan Yan, Xiangyu Xu, Changjiang Cai, Bing Li, Qilun Zhu, and Yi Xu. Naruto: Neural active reconstruction from uncertain target observations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21572–21583, 2024.
 - [10] 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.
 - [11] Lily Goli, Cody Reading, Silvia Sellán, Alec Jacobson, and Andrea Tagliasacchi. Bayes’ rays: Uncertainty quantification for neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20061–20070, 2024.
 - [12] Armin Hornung, Kai M Wurm, Maren Bennewitz, Cyrill Stachniss, and Wolfram Burgard. Octomap: An efficient probabilistic 3d mapping framework based on octrees. *Autonomous robots*, 34:189–206, 2013.
 - [13] Eddy Ilg, Ozgun Cicek, Silvio Galesso, Aaron Klein, Osama Makansi, Frank Hutter, and Thomas Brox. Uncertainty estimates and multi-hypotheses networks for optical flow. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 652–667, 2018.
 - [14] Stefan Isler, Reza Sabzevari, Jeffrey Delmerico, and Davide Scaramuzza. An information gain formulation for active volumetric 3d reconstruction. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3477–3484. IEEE, 2016.
 - [15] 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.
 - [16] Liren Jin, Xingguang Zhong, Yue Pan, Jens Behley, Cyrill Stachniss, and Marija Popović. Activegs: Active scene reconstruction using gaussian splatting. *IEEE Robotics and Automation Letters*, 2025.
 - [17] Rui Jin, Yuman Gao, Yingjian Wang, Yuze Wu, Haojian Lu, Chao Xu, and Fei Gao. Gs-planner: A gaussian-splatting-based planning framework for active high-fidelity reconstruction. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 11202–11209. IEEE, 2024.
 - [18] 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.
 - [19] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), July 2023. URL <https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/>.
 - [20] Keifer Lee, Shubham Gupta, Sungyoung Kim, Bhargav Makwana, Chao Chen, and Chen Feng. So-nerf: Active view planning for nerf using surrogate objectives. *arXiv preprint arXiv:2312.03266*, 2023.
 - [21] Yuetao Li, Zijia Kuang, Ting Li, Guyue Zhou, Shaohui Zhang, and Zike Yan. Activesplat: High-fidelity scene reconstruction through active gaussian splatting. *arXiv preprint arXiv:2410.21955*, 2024.
 - [22] Liang Lu, Yinqiang Zhang, Peng Zhou, Jiaming Qi, Yipeng Pan, Changhong Fu, and Jia Pan. Semantics-aware receding horizon planner for object-centric active mapping. *IEEE Robotics and Automation Letters*, 2024.
 - [23] Mehdi Maboudi, MohammadReza Homaei, Soohwan Song, Shirin Malihi, Mohammad Saadatseresht, and Markus Gerke. A review on viewpoints and path planning for uav-based 3-d reconstruction. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 16:5026–5048, 2023.
 - [24] 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.
 - [25] Daniel Mellinger and Vijay Kumar. Minimum snap trajectory generation and control for quadrotors. In *2011 IEEE international conference on robotics and automation*, pages 2520–2525. IEEE, 2011.
 - [26] 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.
 - [27] Mark W Mueller, Markus Hehn, and Raffaello D’Andrea. A computationally efficient motion primitive for quadro-

- copter trajectory generation. *IEEE transactions on robotics*, 31(6):1294–1310, 2015.
- [28] Erik Nelson and Nathan Michael. Information-theoretic occupancy grid compression for high-speed information-based exploration. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4976–4982. IEEE, 2015.
- [29] Xuran Pan, Zihang Lai, Shiji Song, and Gao Huang. Activeness: Learning where to see with uncertainty estimation. In *European Conference on Computer Vision*, pages 230–246. Springer, 2022.
- [30] Yunlong Ran, Jing Zeng, Shibo He, Jiming Chen, Lincheng Li, Yingfeng Chen, Gimhee Lee, and Qi Ye. Neurar: Neural uncertainty for autonomous 3d reconstruction with implicit neural representations. *IEEE Robotics and Automation Letters*, 8(2):1125–1132, 2023.
- [31] Jianxiong Shen, Adria Ruiz, Antonio Agudo, and Francesc Moreno-Noguer. Stochastic neural radiance fields: Quantifying uncertainty in implicit 3d representations. In *2021 International Conference on 3D Vision (3DV)*, pages 972–981. IEEE, 2021.
- [32] Yunlong Song, Selim Naji, Elia Kaufmann, Antonio Loquercio, and Davide Scaramuzza. Flightmare: A flexible quadrotor simulator. In *Proceedings of the 2020 Conference on Robot Learning*, pages 1147–1157, 2021.
- [33] Wennie Tabib, Micah Corah, Nathan Michael, and Red Whittaker. Computationally efficient information-theoretic exploration of pits and caves. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3722–3727. IEEE, 2016.
- [34] Wennie Tabib, Kshitij Goel, John Yao, Curtis Boirum, and Nathan Michael. Autonomous cave surveying with an aerial robot. *IEEE Transactions on Robotics*, 38(2):1016–1032, 2021.
- [35] Zijun Xu, Rui Jin, Ke Wu, Yi Zhao, Zhiwei Zhang, Jieru Zhao, Fei Gao, Zhongxue Gan, and Wenchao Ding. Hgs-planner: Hierarchical planning framework for active scene reconstruction using 3d gaussian splatting. *arXiv preprint arXiv:2409.17624*, 2024.
- [36] Brian Yamauchi. A frontier-based approach for autonomous exploration. In *Proceedings 1997 IEEE International Symposium on Computational Intelligence in Robotics and Automation CIRA'97. Towards New Computational Principles for Robotics and Automation*, pages 146–151. IEEE, 1997.
- [37] Dongyu Yan, Jianheng Liu, Fengyu Quan, Haoyao Chen, and Mengmeng Fu. Active implicit object reconstruction using uncertainty-guided next-best-view optimization. *IEEE Robotics and Automation Letters*, 2023.
- [38] Linhan Yang, Lei Yang, Haoran Sun, Zeqing Zhang, Haibin He, Fang Wan, Chaoyang Song, and Jia Pan. One fling to goal: Environment-aware dynamics for goal-conditioned fabric flinging. *arXiv preprint arXiv:2406.14136*, 2024.
- [39] 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.
- [40] Yichen Zhang, Xinyi Chen, Chen Feng, Boyu Zhou, and Shaojie Shen. Falcon: Fast autonomous aerial exploration using coverage path guidance. *IEEE Transactions on Robotics*, 2024.
- [41] Yinqiang Zhang, Liang Lu, Xiaowei Luo, and Jia Pan. Global bim-point cloud registration and association for construction progress monitoring. *Automation in Construction*, 168:105796, 2024.
- [42] Zhengdong Zhang, Theia Henderson, Sertac Karaman, and Vivienne Sze. FSMI: Fast computation of shannon mutual information for information-theoretic mapping. *The International Journal of Robotics Research*, 39(9):1155–1177, 2020.
- [43] 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.

APPENDIX

A. Information Gain Function Derivation

The information gain function $f(\delta(Z), o_{1:k-1})$ is defined by (14) that,

$$f(\delta(Z), o_{1:k-1}) = P(r|z_k = Z, Z_{1:k-1}) \log\left(\frac{P(r|z_k = Z, Z_{1:k-1})}{P(r|Z_{1:k-1})}\right) \quad (25)$$

Based on the definition of $o_{1:k-1}$ in (4), we have

$$P(r|Z_{1:k-1}) = \frac{1 + o_{1:k-1}}{o_{1:k-1}} \quad (26)$$

Substitute the probability update (6) into (26), we can derive

$$P(r|z_k = Z, Z_{1:k-1}) = \frac{1 + \delta(Z)o_{1:k-1}}{\delta(Z)o_{1:k-1}} \quad (27)$$

Substitute (26) and (27) into (25), the information gain function can be derived and simplified as,

$$f(\delta, o) = \frac{o}{o + \delta - 1} \log\left(\frac{o + 1}{o + \delta - 1}\right) \quad (28)$$

B. Closed-form Minimum-snap motion primitive

Here we present the detailed motion planner used in our system, associated with the motion cost derivation. The planner is designed for a quadrotor, which is agile and versatile in cluttered environments. The main notations of this section are listed in Table VI.

As quadrotor dynamics has been demonstrated differentially flat [25], any smooth trajectory with physically bounded derivatives in the space of flat outputs can be followed by the under-actuated quadrotor. Aiming at view planning, we choose the flat outputs as $\sigma = [x, y, z, \psi]^T$, which includes the drone's position $\mathbf{p} = (x, y, z)$ and yaw ψ . Then, the full quadrotor states $\mathcal{X}$ can be expressed by the algebraic functions of flat outputs and their derivatives that

$$\mathcal{X} = [\mathbf{p}, \mathbf{v}, \mathbf{a}, \mathbf{j}, \psi, \dot{\psi}] \quad (29)$$

TABLE VI
MAIN NOTATIONS FOR MOTION PLANNING

Notations	Explanation
$\mathcal{X}$	Full quadrotor state.
$\mathcal{B}$	The quadrotor body frame.
t, T	Current time and trajectory duration.
σ	Flat outputs of a quadrotor, a viewpoint.
$\mathbf{p}$	Position in the world frame.
$\mathbf{v} := \dot{\mathbf{p}}$	Linear velocity in the world frame.
$\mathbf{a} := \dot{\mathbf{v}}$	Linear acceleration in the world frame.
$\mathbf{j} := \dot{\mathbf{a}}$	Jerk.
$\mathbf{s} := \dot{\mathbf{j}}$	Snap.
ψ	Yaw, rotation around $z_{\mathcal{B}}$ axis.
$\alpha, \beta, \gamma, \delta$	Primitive trajectory coefficients.
$(\cdot)^*$	Optimal state.
$(\cdot)_0$	Initial state, current state.
$(\cdot)_f$	Final state, objective state.
$(\cdot)_T$	A state transition with duration T .

where the velocity $\mathbf{v}$, acceleration $\mathbf{a}$, jerk $\mathbf{j}$ are derivatives of position $\mathbf{p}$.

As the control input, thrust and torque, can be formulated by at-most the fourth derivatives, we design trajectories that minimize the snap $\mathbf{s}$,

$$\begin{aligned} \min \quad & \frac{1}{T} \int_0^T \|\mathbf{s}(t)\|^2 dt \\ \text{s.t.} \quad & \mathcal{X}(t) = \mathcal{X}_0 \\ & \mathcal{X}(t+T) = \mathcal{X}_f \end{aligned} \quad (30)$$

where $\mathcal{X}_0$ and $\mathcal{X}_f$ represent the initial state and final state respectively.

To efficiently generate the trajectory from current state to the next viewpoint, we extend the result of the work [27] and derive the closed-form minimum-snap primitives. The optimal trajectory generation problem (30) can be decoupled into three orthogonal axes that,

$$J = \frac{1}{T} \int_0^T \|\mathbf{s}(t)\|^2 dt = \sum_{k \in \{x, y, z\}} \frac{1}{T} \int_0^T s_k^2(t) dt \quad (31)$$

The axis subscript k for vectors will be simplified for the remainder of this section. Employing Pontryagin's minimum principle, the optimal control input for each axis can be solved for,

$$s^*(t) = \frac{1}{2}\alpha t^3 + \frac{3}{2}\beta t^2 + 3\gamma t + 3\delta \quad (32)$$

from which the optimal position for each axis follows from integration,

$$\begin{aligned} p^*(t) = & \frac{1}{1680}\alpha t^7 + \frac{1}{240}\beta t^6 + \frac{1}{40}\gamma t^5 + \frac{1}{8}\delta t^4 \\ & + \frac{1}{6}j_0 t^3 + \frac{1}{2}a_0 t^2 + v_0 t + p_0 \end{aligned} \quad (33)$$

The minimum cost for each axis can be derived by substituting

optimal snap (32) into (30),

$$\begin{aligned} J_k = & \frac{1}{28}\alpha^2 T^6 + \frac{1}{4}\alpha\beta T^5 + \left(\frac{9}{20}\beta^2 + \frac{3}{5}\alpha\gamma\right)T^4 \\ & + \left(\frac{3}{4}\alpha\delta + \frac{9}{4}\beta\gamma\right)T^3 + (3\gamma^2 + 3\beta\delta)T^2 \\ & + 9\gamma\delta T + 9\delta^2 \end{aligned} \quad (34)$$

The constant coefficients α, β, γ and δ can be expressed by algebraic functions of initial state $\mathcal{X}_0$ and final state $\mathcal{X}_f$ as

$$\begin{bmatrix} \alpha \\ \beta \\ \gamma \\ \delta \end{bmatrix} = \frac{1}{T^7} \begin{bmatrix} -33600 & 16800T & -3360T^2 & 280T^3 \\ 16800T & -8160T^2 & 1560T^3 & -120T^4 \\ -3360T^2 & 1560T^3 & -280T^4 & 20T^5 \\ 280T^3 & -120T^4 & 20T^5 & -\frac{4}{3}T^6 \end{bmatrix} \begin{bmatrix} \Delta p \\ \Delta v \\ \Delta a \\ \Delta j \end{bmatrix} \quad (35)$$

where the constants $\Delta p, \Delta v, \Delta a, \Delta j$ are defined as,

$$\begin{bmatrix} \Delta p \\ \Delta v \\ \Delta a \\ \Delta j \end{bmatrix} = \begin{bmatrix} p_f - p_0 - v_0 T - \frac{1}{2}a_0 T^2 - \frac{1}{6}j_0 T^3 \\ v_f - v_0 - a_0 T - \frac{1}{2}j_0 T^2 \\ a_f - a_0 - j_0 T \\ j_f - j_0 \end{bmatrix} \quad (36)$$

Therefore, given the current state $\mathcal{X}_0$ and the desired state $\mathcal{X}_f$, the optimal motion primitive and control input for the fully defined end state can be solved by (32)(33)(35)(36).