TRACE: Ergodic Trajectory Optimization for Active Scene Reconstruction

Ziyue Zheng[1*], Linli Shi[1*], Bingkun He[1], Wen Jiang[2], Ziyun Wang[1†]

[1]Johns Hopkins University, USA
[2]University of Pennsylvania, USA
{zzheng60, lshi42, bhe15}@jh.edu, wenjiang@seas.upenn.edu, claude.w@jhu.edu
[*]These authors contributed equally. [†]Corresponding author.

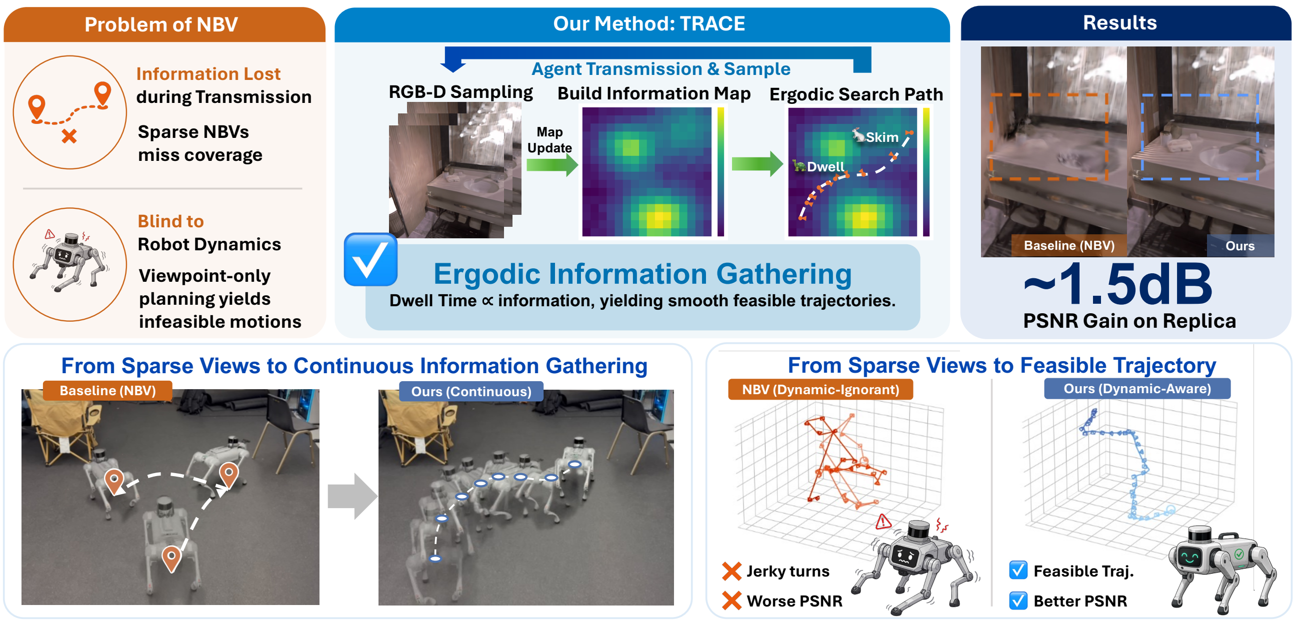

Figure 1: **From discrete viewpoints to ergodic active reconstruction.** Next-Best-View (NBV) planners discard information between sparse viewpoints and ignore robot dynamics, producing jerky, infeasible trajectories. Our method instead constructs an information map from streaming RGB-D observations and plans an ergodic search path that continuously gathers information while respecting the robot's dynamic. This yields smoother trajectories, denser coverage, and 1.5 dB PSNR gain.

Abstract

Existing active reconstruction systems with Gaussian-splatting maps select observations greedily, optimizing a single next-best-view (NBV) at each step and connecting the chosen views by short-horizon path planning. This greedy decoupling disregards the global structure of scene information, producing inefficient trajectories that waste sensing capacity in transit between selected views. In this work, we study active reconstruction as an ergodic coverage problem: the time-averaged spatial statistics of the sensor trajectory should match a target information distribution induced by the current map. Our approach derives this target distribution online from uncertainty and visibility, and calculates ergodic trajectories via a kernel-ergodic horizon planner with gradient flow and footprint depletion, closing the loop between mapping and trajectory optimization. We thoroughly evaluate TRACE on the Replica dataset against the Next-Best-View (NBV) baselines, improving PSNR by 1.5 dB. **Code**: https://github.com/spikelab-jhu/trace-active-reconstruction.

1 Introduction

Active 3D reconstruction recovers the geometry of an unknown scene by moving a sensor along a sequence of informative viewpoints, which has long been a core problem in robotic perception (Zeng et al. 2020; Isler et al. 2016; Bircher et al. 2016). In such tasks, the robot incrementally builds a high-fidelity model of the scene from its own streaming observations. Recent advances in differentiable rendering, particularly 3D Gaussian Splatting (GS) (Kerbl et al. 2023) and its surface-aligned variant 2DGS (Huang et al. 2024), have raised the achievable reconstruction quality dramatically. The bottleneck has shifted from the map representation to the planner (Jin et al. 2025; Chen et al. 2025; Li et al. 2025): under finite budgets of time, energy, and motion, the agent must decide how to spend each meter of motion across an unknown scene. Therefore, the reconstruction quality depends on both map representations and how the agent chooses to distribute its limited motion budget across the scene.

A common paradigm of active reconstruction with Gaussian-splatting maps is to plan view-by-view: at each step, the planner scores candidate viewpoints by an information-gain proxy, commits to the highest-scoring pose, and connects it to the current location by short-horizon path planning (Jin et al. 2025; Chen et al. 2025). Recent work introduces a hierarchy over a topological subgraph, but each step still commits to a single node and routes to it via shortest-path planning (Li et al. 2025). However, this formulation decouples viewpoint selection from trajectory planning, and the motion between successive views serves only to move the sensor toward the next target rather than to acquire additional information. The decoupling has *three main issues*. First, viewpoint scoring is **short-term**: each decision commits to the locally most informative pose without anticipating where the sensor must subsequently travel. Second, the behaviors between selected views do not contribute to **efficient information gathering**. The resulting trajectories are jagged, redundant, and biased toward isolated high-information viewpoints rather than uniform spatial coverage of the scene. Third, the trajectory formed by the sequence of planned views ignores **real-world constraints** such as controllability, actuator limits, and energy efficiency, often producing dynamically infeasible trajectories. In particular, discrete search algorithms used in Next-Best-View (NBV) planning typically optimize over viewpoints rather than executable motions, and therefore cannot directly account for the robot's continuous-time dynamics.

In this work, we aim to address these issues by proposing the first active 3D reconstruction method that plans continuous and control-feasible trajectories using ergodic search. Instead of modeling the problem as a planning over a set of next poses, we directly optimize a trajectory over the constructed information distribution based on the current state of mapping and exploration. Specifically, our planner optimizes the path so that its time-averaged spatial statistics match a target information distribution: the time the trajectory spends in any region is proportional to the information density there. Instantiating this formulation on a live Gaussian-splatting map raises three technical challenges. First, classical ergodic search assumes a target distribution specified a priori, whereas in active reconstruction the information density is implicit in the map state, encoded in per-Gaussian confidence and evolving occupancy, and must be re-derived online as observations are integrated. Second, the informative regions are scene surfaces that the sensor cannot occupy: the trajectory must remain in free space while its viewing footprint, rather than its position, covers the high-information surfaces. Third, the marginal value of observing a surface diminishes under repeated coverage within a horizon, a time-varying effect that a static target distribution cannot capture.

The target information distribution is derived online from per-Gaussian uncertainty and visibility, while the trajectory is computed by a kernel-ergodic horizon planner with gradient-flow updates and a footprint-depletion mechanism that suppresses re-coverage. Our contributions are summarized as:

- To the best of our knowledge, we are the first to formulate active reconstruction with Gaussian-splatting maps as a trajectory-level ergodic coverage problem, replacing greedy viewpoint selection with continuous optimization over dynamically feasible trajectories.
- We adapt kernel-based ergodic search to surface-based reconstruction by diffusing surface information into traversable free space and introducing footprint-aware depletion and gaze objectives for efficient coverage and camera orientation.
- Under the same mapper, budget, and evaluation protocol, TRACE improves over the strongest NBV baseline by 1.5 dB PSNR across eight Replica scenes and supports direct trajectory execution on quadruped and manipulator (supplementary) without an intermediate path planner.

2 Related Work

2.1 Active 3D Reconstruction

Active 3D reconstruction has been studied for decades under the *next-best-view* (NBV) formulation (Connolly 1985; Isler et al. 2016; Bircher et al. 2016; Delmerico et al. 2018): at each step, the agent selects the viewpoint expected to maximize a chosen information-gain criterion. Sampling-based informative path planning extends NBV to entire trajectories (Hollinger and Sukhatme 2014) and long-horizon tree search (Best et al. 2019), while multi-stage aerial pipelines couple it to multi-view-stereo reconstruction (Hepp, Nießner, and Hilliges 2018). Frontier-based exploration (Yamauchi 1997) and occupancy mapping (Hornung et al. 2013) provide complementary geometric drivers, and SCONE (Guédon, Monasse, and Lepetit 2022) optimizes surface coverage via Monte Carlo volumetric integration.

The emergence of differentiable rendering—NeRF (Mildenhall et al. 2020) and subsequent radiance-field variants (Barron et al. 2022; Müller et al. 2022; Fridovich-Keil et al. 2022; Chen et al. 2022)—extended active reconstruction to learned neural map representations. NARUTO (Feng et al. 2024), ActiveNeRF (Pan et al. 2022), ActiveImplicit (Yan et al. 2023), and NeU-NBV (Jin et al. 2023) score viewpoints by NeRF uncertainty or implicit-occupancy entropy. FisherRF (Jiang, Lei, and Daniilidis 2024) uses Fisher information over radiance-field parameters; NVF (Xue et al. 2024) composites position-based uncertainty into camera-ray uncertainty; Active Neural Mapping (Yan, Yang, and Zha 2023) measures neural variability under weight perturbation; and Siming et al. (2024) maximize mutual information through a generative NeRF model. GenNBV (Chen et al. 2024) trains a generalizable RL policy over a 5-DoF action space, while ACE-NBV (Zhang et al. 2023) selects views that improve grasp quality.

Recent Gaussian-splatting maps (Kerbl et al. 2023; Huang et al. 2024) have prompted dedicated active-GS planners. ActiveGS (Jin et al. 2025), ActiveGAMER (Chen et al. 2025), and ActiveSplat (Li et al. 2025) score viewpoints by per-primitive confidence, rendering-based information gain, or frontier cues; ActiveSplat further introduces a local-vs-global hierarchy over a topological subgraph. GauSS-MI (Xie et al. 2025) introduces Shannon mutual information over Gaussian-splat appearance, POp-GS (Wilson et al. 2025) reframes information gain through P-Optimality, and

Active3D (Li, Li, and Lee 2026) fuses implicit and explicit representations with hierarchical uncertainty quantification. These methods share a common structure: at each step, the planner commits to a best viewpoint and connects it to the current pose by short-horizon path planning, producing trajectories at the viewpoint level rather than the trajectory level.

2.2 Ergodic Coverage for Robotic Exploration

Ergodic search formulates information gathering as a trajectory-level coverage problem: the time-averaged spatial statistics of the trajectory should match a target information distribution (Mathew and Mezić 2011; Miller et al. 2016). This trajectory-centric formulation naturally accommodates finite sensing budgets by optimizing information collection over the entire motion rather than selecting isolated viewpoints. To optimize the ergodic problem, a range of methods optimize the ergodic objective via spectral multiscale coverage with Fourier analysis (Mathew and Mezić 2011), Kullback–Leibler divergence (Abraham, Prabhakar, and Murphey 2021), kernel-based ergodic metrics (Sun et al. 2025), flow matching (Sun, Pinosky, and Murphey 2025), receding-horizon control (Mavrommati et al. 2017), potential field approaches (Ivić, Crnković, and Mezić 2016), and LQR-based optimization (Miller and Murphey 2013). Subsequent work has extended the framework along several axes, including time-optimal ergodic search (Dong, Berger, and Abraham 2024), dynamic sensor footprints (Zheng et al. 2025), probabilistic connectivity (Liu and Ren 2025), manipulation (Shetty, Silvério, and Calinon 2022), target localization (Mavrommati et al. 2017) and coverage in constrained domains (Ayvali, Salman, and Choset 2017). Despite these advances, most of these methods typically assume the target distribution is parametrically given a priori from a model of expected information density. In contrast, we derive the target distribution online from a live 2DGS map, where the information density is implicit in per-Gaussian uncertainty and evolving visibility geometry. To our knowledge, this is the first ergodic trajectory optimization formulation for active reconstruction with Gaussian-splatting maps.

3 Method

We instantiate the formulation of Sec. 1 as a horizon-based ergodic trajectory optimization driven by online 2DGS map. Sec. 3.1 formalizes the problem; Sec. 3.2 constructs the target information distribution from the map state; Sec. 3.3 optimizes a kernel-ergodic trajectory by gradient descent on a footprint-depleting information field.

3.1 Problem Formulation

We consider the problem of active 2D Gaussian-surfel (2DGS) mapping: an autonomous agent equipped with an RGB-D sensor incrementally builds a 2DGS map $\mathcal{M}$ of an unknown bounded scene $\Omega \subset \mathbb{R}^3$ through a sequence of self-selected viewpoints. We adopt the 2DGS representation for its surface-aligned geometry, which directly supports the mesh-quality metrics commonly used to evaluate active reconstruction (Huang et al. 2024). The sensor is modeled as a view cone, rotationally symmetric about its optical axis,

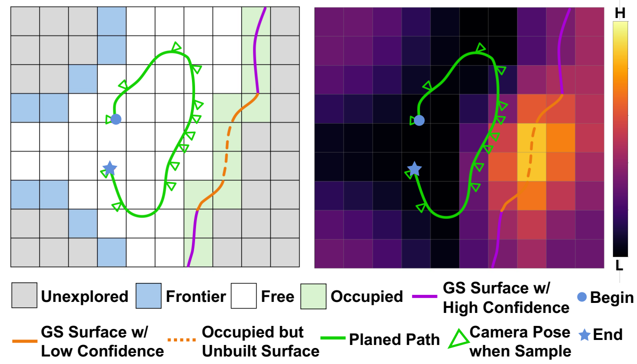

Figure 2: **2D illustration of a single planning horizon. Left:** voxel classification and our planned path. **Right:** the corresponding information-map heatmap (Sec. 3.2). The planner jointly optimizes the trajectory, allocating sampling effort proportional to the information distribution, dwelling longer and sampling more densely in high-information regions.

so its roll is immaterial and held fixed. At each replanning step (horizon) t, the planner selects a K-step trajectory $\tau_t = (\mathbf{p}_t^1, \ldots, \mathbf{p}_t^K) \in SE(3)^K$, with positions $\mathbf{x}_t^k \in \mathbb{R}^3$ and orientations determined by yaw and pitch. The trajectory is parameterized by control $\mathbf{u}_t \in \mathbb{R}^{K \times 5}$ (position velocity, yaw rate, pitch rate) cumulatively integrated under single-integrator dynamics (matching Go2's interface). The supplementary shows a more complex case on FR3.

After executing the full horizon, the agent integrates new RGB-D observations into $\mathcal{M}_t$ and re-plans with $\mathbf{u}$ warm-started from the previous optimization. Although the voxel map is updated during execution, the planning objective remains fixed until the next horizon. Formally, each replanning step solves

$$\tau_t^\star \;=\; \arg\min_{\tau_t \in \mathcal{T}_t} \; J_t(\tau_t; \mathcal{M}_t, \mathcal{V}_t), \tag{1}$$

where $\mathcal{T}_t$ is the feasible-trajectory set in horizon t (collision-free, bounded step size) and J_t is the ergodic optimal-control objective of Sec. 3.3, defined over a target information distribution ϕ_t derived online from the current map (Sec. 3.2). In practice we relax the hard constraints defining $\mathcal{T}_t$ into the soft penalties of Eq. (7) and optimize the resulting objective over the controls.

3.2 Information Map

Ergodic search requires a target information distribution to guide trajectory planning. In our task, ϕ_t (information map at horizon t), evolves as new observations are integrated. To introduce the information map, we first present map representation design. We follow the hybrid map representation of ActiveGS (Jin et al. 2025): a 2DGS Map $\mathcal{M}_t$ is used for rendering, while a voxel map $\mathcal{V}_t$ represents the occupancy probability. GS Map $\mathcal{M}_t$ is trained and updated at the end of each horizon, whereas the voxel map $\mathcal{V}_t$ is updated on the fly following the approach of OctoMap (Hornung et al. 2013).

For brevity we suppress the dependence on $(\mathcal{M}_t, \mathcal{V}_t)$ in the notation below; all fields are evaluated at the current map state. For each voxel v in $\mathcal{V}_t$, the raw information value is

$$\phi_t^{\mathrm{raw}}(v) \;=\; \alpha_{\mathrm{u}}\,\mathbf{1}_{\mathrm{unexp}}(v) \;+\; \alpha_{\mathrm{f}}\,\mathbf{1}_{\mathrm{front}}(v) \;+ \\ \alpha_{\mathrm{b}}\,\mathbf{1}_{\mathrm{unbuilt}}(v) \;+\; \beta\,\big(1-c(v)\big)\,\mathbf{1}_{\mathrm{low}}(v), \qquad (2)$$

Here $\mathbf{1}_{\mathrm{unexp}}(v)$, $\mathbf{1}_{\mathrm{front}}(v)$, and $\mathbf{1}_{\mathrm{unbuilt}}(v)$ indicate, respectively, voxels no depth ray has traversed, free voxels bordering unexplored space, and voxels $\mathcal{V}_t$ deems occupied but holding no Gaussian of $\mathcal{M}_t$. Since $\mathcal{V}_t$ updates on execution but $\mathcal{M}_t$ only at horizon boundaries, known-occupied voxels may still hold no Gaussian; $\mathbf{1}_{\mathrm{unbuilt}}$ draws the sensor back to them and reduces holes in the GS map early in the mission. For the confidence term, $\mathcal{G}_v \subseteq \mathcal{M}_t$ collects the Gaussians whose centers fall in voxel v, $\gamma_g \in [0,1]$ is the per-primitive rendering confidence inherited from ActiveGS (Jin et al. 2025), and $c(v) = \frac{1}{|\mathcal{G}_v|}\sum_{g\in\mathcal{G}_v}\gamma_g$ is their mean. Intuitively, γ_g grows with the number and angular diversity of viewpoints that have already observed Gaussian g, so $1-c(v)$ is large precisely on voxels whose surface fit is still weakly constrained. The gate $\mathbf{1}_{\mathrm{low}}(v) = 1$ on voxels that contain Gaussians but are not yet *well built* (as shown in Fig. 3); well-built voxels contribute no mass and stop attracting the sensor. The non-negative weights $\alpha_{\mathrm{u}}, \alpha_{\mathrm{f}}, \alpha_{\mathrm{b}}, \beta$ control, respectively, exploration of unseen volume, expansion of the frontier, attention to visible-but-unmodeled surfaces, and refinement of low-confidence Gaussians.

We obtain ϕ_t by box-filtering ϕ_t^{raw}, masking it to the observed collision-free region of $\mathcal{V}_t$, and normalizing to a probability distribution. Masking is what makes the ergodic target realizable: it projects information mass from occupied surfaces onto the reachable free space the sensor can actually occupy. The gaze reward instead uses the unmasked ϕ_t^{raw} to preserve absolute scale. Both fields are sampled at continuous locations by trilinear interpolation, so all terms are differentiable in the trajectory.

3.3 Kernel-Ergodic Trajectory with Sensor Footprint Depletion

We plan the trajectory with the kernel-ergodic metric of Sun et al. (2025), which drives a trajectory's time-averaged positions to match a target distribution ϕ. Active reconstruction violates its central premise: the high-ϕ mass lies on scene surfaces the robot cannot occupy, so no collision-free trajectory can match it. We resolve this in two ways. First, we *project* the target into free space—the diffusion and masking of Sec. 3.2, so the body covers reachable space rather than the surfaces themselves. Second, we add two footprint-aware mechanisms on top of the position-based metric: a *footprint-overlap depletion* term (Eq. (3)) that discourages re-observing covered surfaces, and a *footprint gaze reward* (Eq. (5)) that aims the camera at high-ϕ surfaces while the body stays in free space. Unlike Zheng et al. (2025), which replaces the point-sensor delta with a footprint distribution, our footprint enters only through the depletion factor while the kernel-ergodic metric stays position-based, and we derive ϕ online from a live 2DGS map rather than a fixed target. Fig. 3 illustrates the design on an object-centric example.

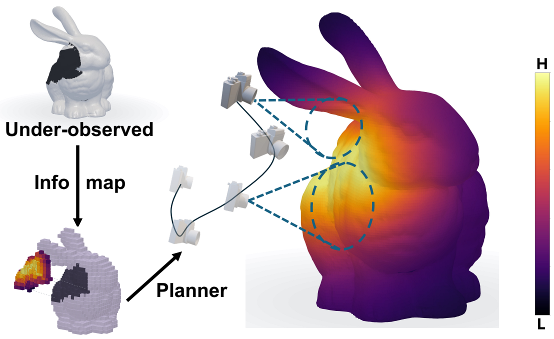

Figure 3: **From reconstruction state to information distribution to ergodic coverage.** The current reconstruction has well-built and under-built regions (top left); the voxel map discretizes this state, and the under-built region carries high information (bottom left); **Right:** the induced information distribution over the object surface and a continuous trajectory whose view cones dwell on high-information regions—the time spent looking at a region is proportional to its information density.

Time-varying ϕ via footprint-overlap depletion. In active sensing, the marginal value of observing a region diminishes once prior visits have already covered it; ϕ should therefore decay at already-covered locations as the trajectory progresses. We model this decay by replacing the static evaluation $\phi_t(\mathbf{x}_t^k)$ with a coverage-discounted value

$$\phi_t^k \;\equiv\; \phi_t(\mathbf{x}_t^k)\prod_{j<k}\big(1-\eta\cdot v_{jk}\big), \qquad (3)$$

where $\eta \in (0,1)$ is the per-visit discount rate and $v_{jk} \in [0,1]$ is a soft footprint-overlap kernel between waypoints j,k. We approximate the sensor footprint at waypoint k by D points $\mathbf{f}_{t,d}^k = \mathbf{x}_t^k + d\,\mathbf{z}_t^k$ sampled along the optical axis $\mathbf{z}_t^k$ at depths $\{d_1,\ldots,d_D\}$, each a candidate surface location at distance d from the camera; the overlap kernel is then

$$v_{jk} \;=\; \max_{d_k,\,d_j}\,\exp\!\left(-\frac{\|\mathbf{f}_{t,d_k}^k - \mathbf{f}_{t,d_j}^j\|^2}{2\,\sigma_{\mathrm{fp}}^2}\right). \qquad (4)$$

This serves as a differentiable proxy for surface co-visibility: two waypoints whose central rays pass through nearby observable points are penalized for re-observing the same region, with σ_{fp} acting as an effective tolerance that softens each ray into a cylinder. The depletion product in Eq. (3) runs over the waypoints of the current horizon; re-coverage *across* horizons is suppressed separately, by re-deriving ϕ_t from the updated map at every horizon boundary, where well-built voxels drop out of Eq. (2).

Footprint-based information attraction. The depletion-aware ergodic metric still evaluates ϕ at waypoint $\mathbf{x}_t^k$, so its gradient attracts the robot toward high-ϕ positions, which in reconstruction lie on or behind surfaces. We complement this with a footprint reward that attracts the camera's footprint to high-ϕ surfaces, while the robot position remains governed

Table 1: Quantitative comparison of rendering quality on the Replica dataset. We report PSNR, SSIM, and LPIPS across all scenes. Best, second-best, and third-best results are highlighted in red, orange, and yellow, respectively.

Methods	Metrics	Of0	Of2	Of3	Of4	R0	R1	R2	H0
NARUTO	PSNR ↑	32.99	27.60	27.48	30.28	28.41	26.77	29.75	25.49
FisherRF	PSNR ↑	36.21	30.33	28.63	32.80	28.88	30.04	32.26	26.79
ActiveGS	PSNR ↑	36.78	31.68	32.16	34.08	29.93	31.74	32.35	32.04
Ours	PSNR ↑	**38.36**	**32.81**	**33.79**	**34.92**	**31.55**	**32.84**	**34.70**	**33.69**
ActiveGS	SSIM ↑	0.963	0.926	0.921	0.932	0.891	0.902	0.924	0.937
Ours	SSIM ↑	**0.969**	**0.930**	**0.929**	**0.934**	**0.907**	**0.914**	**0.941**	**0.953**
ActiveGS	LPIPS ↓	0.081	0.136	0.152	0.138	0.184	0.170	0.146	0.139
Ours	LPIPS ↓	**0.066**	**0.117**	**0.143**	**0.122**	**0.158**	**0.151**	**0.118**	**0.110**

by the ergodic metric and safety penalties. The reward averages the raw information map ϕ_t^{raw} (Eq. (2)) over the same footprint samples used by the depletion kernel, using the raw version to preserve absolute scale across horizons:

$$L_{\text{gaze}}(\mathbf{u}_t) \;=\; -\frac{1}{KD}\sum_{k=1}^{K}\sum_{d=1}^{D} w_{k,d}\,\phi_t^{\text{raw}}(\mathbf{f}_{t,d}^{k}), \qquad (5)$$

where $w_{k,d} = \exp\big(-\kappa\, d_{\text{unsafe}}(\mathbf{f}_{t,d}^{k})\big)$ attenuates footprint samples in unobservable space. Here $d_{\text{unsafe}}(\mathbf{f})$ is the distance from $\mathbf{f}$ to the observed collision-free region of $\mathcal{V}_t$: zero in free space and growing both inside obstacles *and* in the unobserved region behind them, so the reward never credits pointing at high-ϕ voxels visible only through a wall.

Full cost and finite-horizon trajectory optimization. The depletion-aware ergodic metric is

$$E_{\text{kernel}}^{\text{dep}}(\tau_t) \;=\; -\,\frac{2}{K}\sum_{k=1}^{K}\phi_t^{k} \;+\; \frac{1}{K^2} \sum_{i,j=1}^{K}\exp\left(-\frac{\|\mathbf{x}_t^{i}-\mathbf{x}_t^{j}\|^2}{2\sigma^2}\right), \qquad (6)$$

where ϕ_t^k is the depletion-discounted target value of Eq. (3) (information term) and the pairwise term repels nearby waypoints toward uniform coverage (self-correlation term)—the two terms of the kernel-ergodic metric of Sun et al. (2025), now evaluated on the free-space-projected target of Sec. 3.2. The box-filter diffusion plays the mollifying role of the metric's Gaussian kernel, so evaluating the diffused ϕ_t pointwise realizes the information term on a target the sensor can reach: when the diffusion width equals the kernel bandwidth this is the metric of Sun et al. (2025) exactly, and otherwise a free-space-projected kernel-ergodic objective.

The full per-horizon cost combines the depletion-aware ergodic metric with the gaze reward, soft-barrier penalties L_{safe} for collisions, excessive step size, and out-of-bounds waypoints, and a quadratic control regularizer:

$$\begin{aligned} J_t(\mathbf{u}_t) = E_{\text{kernel}}^{\text{dep}}(\tau_t(\mathbf{u}_t)) + \lambda_{\text{g}}\, L_{\text{gaze}}(\mathbf{u}_t) \\ + \lambda_{\text{s}}\, L_{\text{safe}}(\mathbf{u}_t) + \lambda_{\text{r}}\,\|\mathbf{u}_t\|^2. \end{aligned} \qquad (7)$$

All terms are differentiable in $\mathbf{u}_t$; we optimize with Adam (Kingma and Ba 2017), warm-starting from the previous-horizon solution. The agent executes all K waypoints before re-planning.

4 Experiments

4.1 Implementation Details

Dataset and simulator. We evaluate TRACE on eight indoor scenes from the Replica dataset (Straub et al. 2019), following the active-reconstruction protocol of ActiveGS (Jin et al. 2025). All experiments use the Habitat simulator (Savva et al. 2019) with an RGB-D camera moving freely in $\mathbb{R}^3$ with yaw and pitch control , $[60^\circ \times 60^\circ]$ FOV, 512×512 resolution, $[0.1, 5.0]$ m depth range, and Gaussian depth noise $\sigma = 0.01d$. The world is discretized into a 20 cm voxel grid.
Real Robot Experiments. We deploy the identical planner on two physical platforms: a **Unitree Go2** quadruped that reconstructs a room-scale scene, and an **Franka FR3** arm that reconstructs an object-centric scene. Maps are trained on real RGB-D data. Sec. 4.5 details the quadruped deployment; the manipulator result is in the supplementary material.
Training and planning schedule. All experiments run on a single RTX 5090. Each mission runs for a time budget of 300 s, accumulated as $t_{\text{map}} + t_{\text{plan}} + t_{\text{fly}}$, where t_{fly} is calculated from the executed path length divided by a constant velocity of 1 m/s. Per horizon, the planner optimizes a 10-step trajectory with Adam (40 iterations) warm-started from the previous solution; after execution, the Gaussian map is trained for 12 gradient steps on a mini-batch of 10 frames sampled along the executed path together with 10 frames drawn from the observation history.
Evaluation metrics. We report PSNR, SSIM, and LPIPS on a fixed set of 1000 test viewpoints sampled uniformly in each scene's free space, **averaged over 5 independent runs**. Geometry is examined qualitatively in Sec. 4.3. All evaluation settings match primary baseline (Jin et al. 2025).
Baselines. We compare against ActiveGS (Jin et al. 2025), an NBV planner re-run in our pipeline under matched mapper, budget, and evaluation protocol (same as their setting). For completeness, Table 1 also reports NARUTO (Feng et al. 2024), and FisherRF (Jiang, Lei, and Daniilidis 2024) with numbers cited from Jin et al. (2025), which already established ActiveGS as the strongest of these earlier methods. Sampling-density ablations are described in Sec. 4.4.

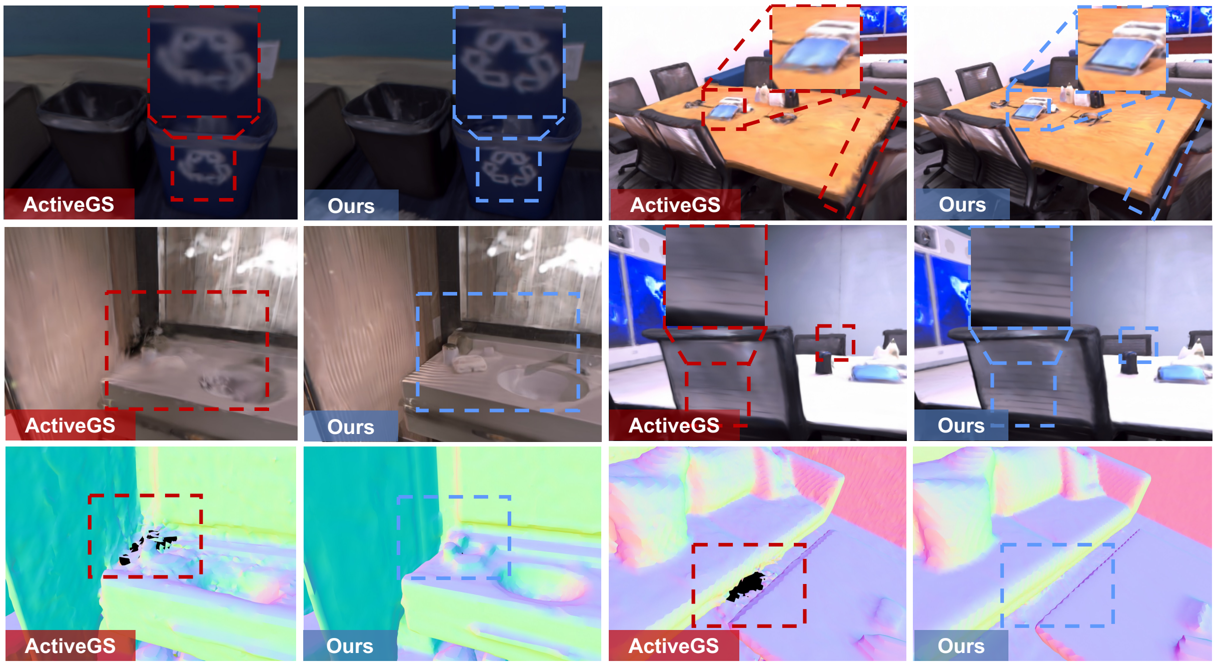

Figure 4: **Qualitative results of ActiveGS and our method on the Replica dataset.** We show RGB renderings (top two rows) and reconstructed surface meshes (bottom row) across four scenes. Red boxes on the ActiveGS results mark regions of low-quality reconstruction, and blue boxes mark the same regions in our results, which appear sharper and more faithful. By replacing greedy NBV selection with ergodic search over an information map, our method achieves higher-quality reconstruction.

4.2 Quantitative Analysis

TRACE achieves higher PSNR than ActiveGS on all scenes (Table 1), with a mean gain of $\mathbf{1.5}$ **dB**; LPIPS and SSIM show the same pattern. The lead arises from two mechanisms: continuous-pose optimization places the camera at any high-ϕ pose along the trajectory rather than the nearest discrete candidate, and depletion suppresses re-coverage, spreading subsequent waypoints over low-confidence surfaces. The PSNR gain is positive on every scene, but its composition differs by scene type. On the office scenes, SSIM is comparable, while LPIPS improves by **6–19**%, indicating that the gain is concentrated in fine-scale appearance. We attribute this to continuous-pose optimization, which reaches poses unavailable to discrete candidates. On the room and hotel scenes, the SSIM gap widens to $\mathbf{0.012}$**–**$\mathbf{0.017}$, indicating that structural differences also emerge. These scenes contain heavier occlusion from furniture, where surfaces missed by discrete viewpoint selection require deliberate coverage, which the ergodic trajectory provides by continuously varying its heading. The gain is not explained by frame count alone: Sec. 4.4 shows that **densifying the baseline's sampling along its path does not close the gap.**

4.3 Qualitative Results

Figure 4 compares reconstructions from TRACE and ActiveGS on Replica scenes. The most pronounced differences lie in surface fidelity: in the highlighted regions, TRACE (blue boxes) recovers sharper texture, consistent lighting, and finer geometric detail on furniture and wall surfaces, while ActiveGS (red boxes) produces flatter, lower-fidelity reconstructions on the same regions. These fidelity gains arise from the same trajectory-level coupling that drives the PSNR lead: depletion keeps the camera looking at under-confident surfaces rather than re-sampling already-covered ones. The regions that differ visually are also where LPIPS separates most, confirming that the gain is concentrated in high-frequency texture rather than spread uniformly over the image. The mesh row shows the same pattern, where ActiveGS leaves holes on surfaces its path merely passes, while TRACE recovers a continuous surface in the same regions.

4.4 Ablation Study

To examine whether TRACE's gain simply comes from denser sampling along the trajectory, and to validate the effectiveness of footprint depletion, we design the following ablations. For the first, we implement two ActiveGS variants: ActiveGS-Random scatters 10 random observations per horizon (a baseline without planner), and ActiveGS-Uniform captures 10 uniformly-interpolated poses along its original path (the information reachable from the trajectory geometry alone). Table 2 reports that ActiveGS-Random matches ActiveGS within noise (**+**$\mathbf{0.08}$ **dB**), and ActiveGS-Uniform drops (**−**$\mathbf{1.7}$ **dB**). Neither closes the gap to TRACE. **Simply densifying observations performs even worse, as redundant views dilute the training weight of informative ones.** For the second, we drop the footprint mechanism in Sec. 3.3, leaving the original kernel-ergodic metric from Sun et al. (2025): This variant loses $\mathbf{2.1}$ **dB** and falls below ActiveGS

Table 2: **Ablation study on Replica (PSNR).** ActiveGS-Random and ActiveGS-Uniform observe at 10 random / uniformly-interpolated poses along ActiveGS's path; *Ours-Kernel ES* removes the footprint mechanism of Sec. 3.3.

Method	**Of0**	**Of2**	**Of3**	**Of4**	**R0**	**R1**	**R2**	**H0**
ActiveGS	36.78	31.68	32.16	34.08	29.93	31.74	32.35	32.04
ActiveGS-Random	36.70	31.37	32.70	34.35	30.09	31.29	32.79	32.11
ActiveGS-Uniform	35.86	29.98	30.49	32.38	27.63	30.95	31.65	28.24
Ours-Kernel ES	35.63	30.78	31.83	34.09	29.91	31.15	33.04	29.39
Ours	**38.36**	**32.81**	**33.79**	**34.92**	**31.55**	**32.84**	**34.70**	**33.69**

on six of eight scenes. NBV replans after every view, and its confidence updates implicitly deplete visited regions. Ergodic search alone optimizes the whole horizon against a frozen information map, yielding redundant views. Depletion is therefore essential to our framework.

4.5 Executing Trajectories in Real World

We deploy TRACE on two physical platforms: **a quadruped (Unitree Go2)** exploring a room-scale scene with the same planner as in Sec. 3.3 and a **Franka arm for object reconstruction (supplementary materials)**. Since the planned trajectory is itself the optimization variable, its waypoints are handed to each platform's controller directly, with no intermediate path planner. This is a practical payoff of trajectory-level planning. An NBV planner commits to discrete viewpoints and delegates the connecting motion to a separate path planner, which can issue **dynamically infeasible** commands on a physical platform. TRACE instead optimizes the trajectory under the platform's own dynamics, so every planned waypoint is **dynamically consistent by construction**. As shown in the supplementary materials, TRACE achieved a 100% **success rate**, whereas our NBV baseline made minor contact with the environment in every trial.

Room-scale exploration (Unitree Go2). A quadruped carries a front-mounted RGB-D camera (Intel D435). The trajectory is constrained to the traversable plane with yaw aligned to the direction of motion, and waypoints are tracked by a velocity-level controller built on the manufacturer's high-level interface, without an intermediate path planner. Over 5-minute missions in a $42\,\text{m}^2$ scene, every planned horizon executes without infeasible commands, and observations are integrated into the 2DGS map. In Fig. 5, the time-colored ergodic path (top) sweeps the room as a single continuous trajectory; the dashed inset marks a cluttered region the **NBV baseline cannot traverse without collision, while TRACE threads the free space safely.** The reconstruction (middle) is clean and complete. The heading traces (bottom) show our heading varying continuously, whereas NBV snaps between discrete targets with rapid reversals, exactly the motion a viewpoint-level planner cannot penalize.

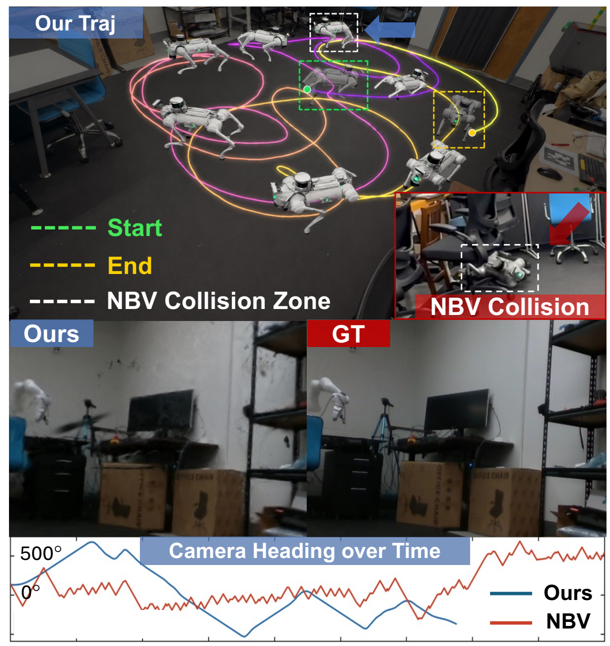

Figure 5: **TRACE on Unitree Go2. Top:** the Unitree Go2 executes a planned ergodic trajectory in a laboratory scene (composite with time-colored path); the dashed boxes mark the start and final poses and a region that the NBV baseline could not traverse safely (inset). **Middle:** Visualization of the reconstruction quality. **Bottom:** the ergodic trajectory turns smoothly and continuously, whereas NBV produces rapid heading reversals between committed viewpoints.

5 Conclusion

We presented TRACE, an ergodic-trajectory formulation for active 2D Gaussian-Surfel reconstruction. Instead of committing to discrete next-best-views, TRACE derives a target information distribution online from voxel-level information and per-Gaussian uncertainty, and optimizes a continuous, control-feasible trajectory with a kernel-ergodic horizon planner and footprint-overlap depletion. Across all eight Replica scenes, TRACE improves over our baseline (ActiveGS) by $+1.5$ dB PSNR on average, and its trajectories execute directly on real physical robot without an intermediate path planner. By recasting active reconstruction as ergodic coverage, TRACE plans information gathering and feasible motion jointly within a single objective.

6 Acknowledgments

This work was also supported in part by funding from the Johns Hopkins Data Science and AI Institute. Thanks Rex for insightful discussion on our physical experiments.

References

Abraham, I.; Prabhakar, A.; and Murphey, T. D. 2021. An ergodic measure for active learning from equilibrium. *IEEE Transactions on Automation Science and Engineering*.

Ayvali, E.; Salman, H.; and Choset, H. 2017. Ergodic Coverage In Constrained Environments Using Stochastic Trajectory Optimization. arXiv:1707.04294.

Barron, J. T.; Mildenhall, B.; Verbin, D.; Srinivasan, P. P.; and Hedman, P. 2022. Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5470–5479.

Best, G.; Cliff, O. M.; Patten, T.; Mettu, R. R.; and Fitch, R. 2019. Dec-MCTS: Decentralized Planning for Multi-Robot Active Perception. *The International Journal of Robotics Research*, 38(2–3): 316–337.

Bircher, A.; Kamel, M.; Alexis, K.; Oleynikova, H.; and Siegwart, R. 2016. Receding Horizon "Next-Best-View" Planner for 3D Exploration. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 1462–1468. IEEE.

Chen, A.; Xu, Z.; Geiger, A.; Yu, J.; and Su, H. 2022. TensoRF: Tensorial Radiance Fields. In *Computer Vision – ECCV 2022*, 333–350. Springer.

Chen, L.; Zhan, H.; Chen, K.; Xu, X.; Yan, Q.; Cai, C.; and Xu, Y. 2025. ActiveGAMER: Active GAussian Mapping through Efficient Rendering. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 16486–16497.

Chen, X.; Li, Q.; Wang, T.; Xue, T.; and Pang, J. 2024. GenNBV: Generalizable Next-Best-View Policy for Active 3D Reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 16436–16445.

Connolly, C. 1985. The determination of next best views. In *Proceedings. 1985 IEEE International Conference on Robotics and Automation*, volume 2, 432–435.

Delmerico, J. A.; Isler, S.; Sabzevari, R.; and Scaramuzza, D. 2018. A Comparison of Volumetric Information Gain Metrics for Active 3D Object Reconstruction. *Autonomous Robots*, 42(2): 197–208.

Dong, D. E.; Berger, H.; and Abraham, I. 2024. Time-optimal ergodic search: Multiscale coverage in minimum time. *The International Journal of Robotics Research*, 02783649241273597.

Feng, Z.; Zhan, H.; Chen, Z.; Yan, Q.; Xu, X.; Cai, C.; Li, B.; Zhu, Q.; and Xu, Y. 2024. NARUTO: Neural Active Reconstruction from Uncertain Target Observations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 21572–21583.

Fridovich-Keil, S.; Yu, A.; Tancik, M.; Chen, Q.; Recht, B.; and Kanazawa, A. 2022. Plenoxels: Radiance Fields Without Neural Networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 5501–5510.

Guédon, A.; Monasse, P.; and Lepetit, V. 2022. SCONE: Surface Coverage Optimization in Unknown Environments by Volumetric Integration. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Hepp, B.; Nießner, M.; and Hilliges, O. 2018. Plan3D: Viewpoint and Trajectory Optimization for Aerial Multi-View Stereo Reconstruction. *ACM Transactions on Graphics*, 38(1): 1–17.

Hollinger, G. A.; and Sukhatme, G. S. 2014. Sampling-Based Robotic Information Gathering Algorithms. *The International Journal of Robotics Research*, 33(9): 1271–1287.

Hornung, A.; Wurm, K. M.; Bennewitz, M.; Stachniss, C.; and Burgard, W. 2013. OctoMap: An Efficient Probabilistic 3D Mapping Framework Based on Octrees. *Autonomous Robots*, 34(3): 189–206.

Huang, B.; Yu, Z.; Chen, A.; Geiger, A.; and Gao, S. 2024. 2D Gaussian Splatting for Geometrically Accurate Radiance Fields. In *ACM SIGGRAPH 2024 Conference Papers*. Association for Computing Machinery.

Isler, S.; Sabzevari, R.; Delmerico, J.; and Scaramuzza, D. 2016. An information gain formulation for active volumetric 3D reconstruction. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 3477–3484.

Ivić, S.; Crnković, B.; and Mezić, I. 2016. Ergodicity-based cooperative multiagent area coverage via a potential field. *IEEE transactions on cybernetics*, 47(8): 1983–1993.

Jiang, W.; Lei, B.; and Daniilidis, K. 2024. FisherRF: Active View Selection and Mapping with Radiance Fields Using Fisher Information. In *Computer Vision – ECCV 2024*, 422–440. Springer.

Jin, L.; Chen, X.; Rückin, J.; and Popović, M. 2023. NeU-NBV: Next Best View Planning Using Uncertainty Estimation in Image-Based Neural Rendering. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.

Jin, L.; Zhong, X.; Pan, Y.; Behley, J.; Stachniss, C.; and Popović, M. 2025. ActiveGS: Active Scene Reconstruction Using Gaussian Splatting. *IEEE Robotics and Automation Letters*, 10(5): 4866–4873.

Kerbl, B.; Kopanas, G.; Leimkühler, T.; and Drettakis, G. 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics*, 42(4).

Kingma, D. P.; and Ba, J. 2017. Adam: A Method for Stochastic Optimization. arXiv:1412.6980.

Li, Y.; Kuang, Z.; Li, T.; Hao, Q.; Yan, Z.; Zhou, G.; and Zhang, S. 2025. ActiveSplat: High-Fidelity Scene Reconstruction Through Active Gaussian Splatting. *IEEE Robotics and Automation Letters*, 10(8): 8099–8106.

Li, Y.; Li, Y.; and Lee, G. H. 2026. Active3D: Active High-Fidelity 3D Reconstruction via Multi-Level Uncertainty Quantification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, 6557–6565.

Liu, Y.; and Ren, Z. 2025. A Probabilistic Measure of Multi-Robot Connectivity and Ergodic Optimal Control. In *Proceedings of Robotics: Science and Systems*. Los Angeles, CA, USA.

Mathew, G.; and Mezić, I. 2011. Metrics for ergodicity and design of ergodic dynamics for multi-agent systems. *Physica D: Nonlinear Phenomena*, 240(4): 432–442.

Mavrommati, A.; Tzorakoleftherakis, E.; Abraham, I.; and Murphey, T. D. 2017. Real-time area coverage and target localization using receding-horizon ergodic exploration. *IEEE Transactions on Robotics*, 34(1): 62–80.

Mildenhall, B.; Srinivasan, P. P.; Tancik, M.; Barron, J. T.; Ramamoorthi, R.; and Ng, R. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *Computer Vision – ECCV 2020*, 405–421. Springer.

Miller, L. M.; and Murphey, T. D. 2013. Trajectory optimization for continuous ergodic exploration. In *2013 American Control Conference*, 4196–4201. IEEE.

Miller, L. M.; Silverman, Y.; MacIver, M. A.; and Murphey, T. D. 2016. Ergodic Exploration of Distributed Information. *IEEE Transactions on Robotics*, 32(1): 36–52.

Müller, T.; Evans, A.; Schied, C.; and Keller, A. 2022. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. *ACM Transactions on Graphics*, 41(4).

Pan, X.; Lai, Z.; Song, S.; and Huang, G. 2022. ActiveNeRF: Learning Where to See with Uncertainty Estimation. In *Computer Vision – ECCV 2022*, 230–246. Springer.

Savva, M.; Kadian, A.; Maksymets, O.; Zhao, Y.; Wijmans, E.; Jain, B.; Straub, J.; Liu, J.; Koltun, V.; Malik, J.; Parikh, D.; and Batra, D. 2019. Habitat: A Platform for Embodied AI Research. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 9339–9347.

Shetty, S.; Silvério, J.; and Calinon, S. 2022. Ergodic Exploration Using Tensor Train: Applications in Insertion Tasks. *IEEE Transactions on Robotics*, 38(2): 906–921.

Siming, H.; Hsu, C. D.; Ong, D.; Shao, Y. S.; and Chaudhari, P. 2024. Active Perception using Neural Radiance Fields. In *2024 American Control Conference (ACC)*, 4353–4358.

Straub, J.; Whelan, T.; Ma, L.; Chen, Y.; Wijmans, E.; Green, S.; Engel, J. J.; Mur-Artal, R.; Ren, C.; Verma, S.; et al. 2019. The Replica Dataset: A Digital Replica of Indoor Spaces. *arXiv preprint arXiv:1906.05797*.

Sun, M. M.; Gaggar, A.; Trautman, P.; and Murphey, T. 2025. Fast Ergodic Search With Kernel Functions. *IEEE Transactions on Robotics*, 41: 1841–1860.

Sun, M. M.; Pinosky, A.; and Murphey, T. 2025. Flow Matching Ergodic Coverage. arXiv:2504.17872.

Wilson, J.; Almeida, M.; Mahajan, S.; Labrie, M.; Ghaffari, M.; Ghasemalizadeh, O.; Sun, M.; Kuo, C.-H.; and Sen, A. 2025. POp-GS: Next Best View in 3D-Gaussian Splatting with P-Optimality. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3646–3655.

Xie, Y.; Cai, Y.; Zhang, Y.; Yang, L.; and Pan, J. 2025. GauSS-MI: Gaussian Splatting Shannon Mutual Information for Active 3D Reconstruction. In *Proceedings of Robotics: Science and Systems (RSS)*. Los Angeles, CA, USA.

Xue, S.; Dill, J.; Mathur, P.; Dellaert, F.; Tsiotras, P.; and Xu, D. 2024. Neural Visibility Field for Uncertainty-Driven Active Mapping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 18122–18132.

Yamauchi, B. 1997. A Frontier-Based Approach for Autonomous Exploration. In *Proceedings of the IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA)*, 146–151. IEEE.

Yan, D.; Liu, J.; Quan, F.; Chen, H.; and Fu, M. 2023. Active Implicit Object Reconstruction Using Uncertainty-Guided Next-Best-View Optimization. *IEEE Robotics and Automation Letters*, 8(10): 6395–6402.

Yan, Z.; Yang, H.; and Zha, H. 2023. Active Neural Mapping. In *IEEE/CVF International Conference on Computer Vision (ICCV)*.

Zeng, R.; Wen, Y.; Zhao, W.; and Liu, Y.-J. 2020. View planning in robot active vision: A survey of systems, algorithms, and applications. *Computational Visual Media*, 6(3): 225–245.

Zhang, X.; Wang, D.; Han, S.; Li, W.; Zhao, B.; Wang, Z.; Duan, X.; Fang, C.; Li, X.; and He, J. 2023. Affordance-Driven Next-Best-View Planning for Robotic Grasping. In *Conference on Robot Learning (CoRL)*.

Zheng, Z.; Liu, Y.; Wang, H.; and Ren, Z. 2025. Ergodic Trajectory Planning with Dynamic Sensor Footprints. arXiv:2512.08661.

TRACE: Ergodic Trajectory Optimization for Active Scene Reconstruction Supplementary Material

A Per-Horizon Algorithm

Algorithm 1 summarizes one plan-and-execute horizon of TRACE; the notation follows Sec. 3.

Algorithm 1: TRACE per-horizon plan-and-execute

Input: Live 2DGS map $\mathcal{M}_t$, voxel map $\mathcal{V}_t$, current pose $\mathbf{p}_t$, warm-start $\mathbf{u}_{\text{prev}}$

```
φ_t ← BuildInfoMap(M_t, V_t) ;                // (2)
u ← WarmStart(u_prev);
for i = 1, ..., N_iter do
    τ ← Rollout(p_t, u);
    Compute J(u) via Eqs. (3)–(7);
    u ← AdamStep(u, ∇_u J);
end
τ_t ← Rollout(p_t, u);
Execute τ_t; integrate RGB-D into M_{t+1}, V_{t+1};
u_prev ← u;
```

B Joint-Space Ergodic Search on the FR3

On the Franka FR3 arm, we plan directly in joint space. A single integrator sufficed for the Go2 because it has a velocity interface; the arm does not. The trajectory variable is instead a sequence of joint configurations, and a differentiable forward-kinematics chain maps each to the camera pose, which determines where the world-frame information map ϕ is sampled. This makes reachability intrinsic: with joint limits as box constraints, every planned view is reachable, and the plan runs with **no inverse kinematics, no candidate viewpoint set, and no feasibility repair.** Collisions with the table, wall, and base column are penalized in the same objective, on forward-kinematic body points. Here the ergodic objective of Sec. 3.3 is optimized through the nonlinear forward kinematics rather than over camera poses directly, yet every executed waypoint is dynamically consistent.

Setup. The arm's state is a joint configuration $q \in \mathcal{Q} \subset \mathbb{R}^7$. Working within a single horizon, we drop the horizon index t; the planner optimizes a K-step joint trajectory $\mathbf{q} = (q^1, \ldots, q^K)$ from joint-velocity controls $\mathbf{u}$, with $q^k = q^{k-1} + \Delta t\, u^k$. The camera is mounted on the end-effector, so its pose follows from the configuration through forward kinematics.

Definition 1 (Kinematic Sensor Map)**.** Let $\mathrm{FK} : \mathcal{Q} \to SE(3)$ be the forward-kinematics map to the camera pose, $\mathrm{FK}(q) = T_{\text{ee}}(q)\, T_{\text{cam}}$, with T_{cam} the fixed hand–eye transform. We write $\mathrm{FK}(q) = (R(q), \mathbf{x}(q))$, with $R(q) \in SO(3)$ the camera orientation and $\mathbf{x}(q) \in \mathbb{R}^3$ its position.

A configuration fixes both where the camera sits, through $\mathbf{x}(q)$, and where it looks, through the optical axis $\mathbf{z}(q) = R(q)\mathbf{e}_3$. We model its view as samples along this axis.

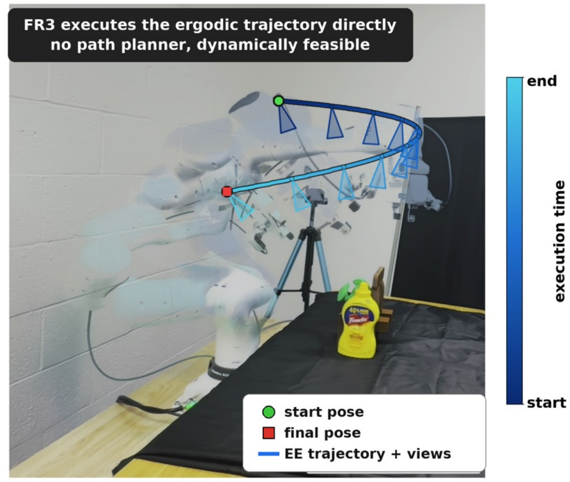

Figure 6: **TRACE on a Franka FR3.** The arm executes the planned ergodic trajectory directly, with no path planner. We overlay the executed end-effector (EE) path, its camera frustums, and the start and final poses (green circle, red square). Planning in joint space makes every waypoint dynamically consistent. **The supplementary video shows this in full.**

Definition 2 (Sensor Footprint)**.** The footprint at q is the set of world points

$$\mathbf{f}_d(q) = \mathbf{x}(q) + d\,\mathbf{z}(q), \qquad d \in \{d_1, \ldots, d_D\}, \tag{8}$$

sampled along the optical axis at depths $d_1, \ldots, d_D$.

Objective. We keep the position-space kernel-ergodic metric of Sun et al. (2025), which scores a trajectory by how closely its time-averaged occupancy matches the target distribution and splits into an information and a uniform-coverage term. Evaluated on the camera positions $\mathbf{x}(q^k)$, it reads

$$E^{\text{dep}}_{\text{kernel}}(\mathbf{q}) = -\frac{2}{K}\sum_{k=1}^{K}\phi^k + \frac{1}{K^2} \sum_{i,j=1}^{K} \exp\left(-\frac{\|\mathbf{x}(q^i) - \mathbf{x}(q^j)\|^2}{2\sigma^2}\right). \tag{9}$$

The first term is the information term: it evaluates the depletion-discounted map $\phi^k = \phi(\mathbf{x}(q^k)) \prod_{j<k}(1 - \eta \cdot v_{jk})$ at each camera position and pulls the trajectory toward informative regions, where v_{jk} is the footprint-overlap kernel (Eqs. (3)–(4)). The second is the coverage term: a Gaussian kernel between camera positions that penalizes revisiting and spreads the trajectory out. They reproduce the information-and-coverage decomposition of Sun et al. (2025), now on the camera positions the configuration produces.

Position alone does not fix what the camera sees, so orientation enters through the footprint. The gaze term evaluates

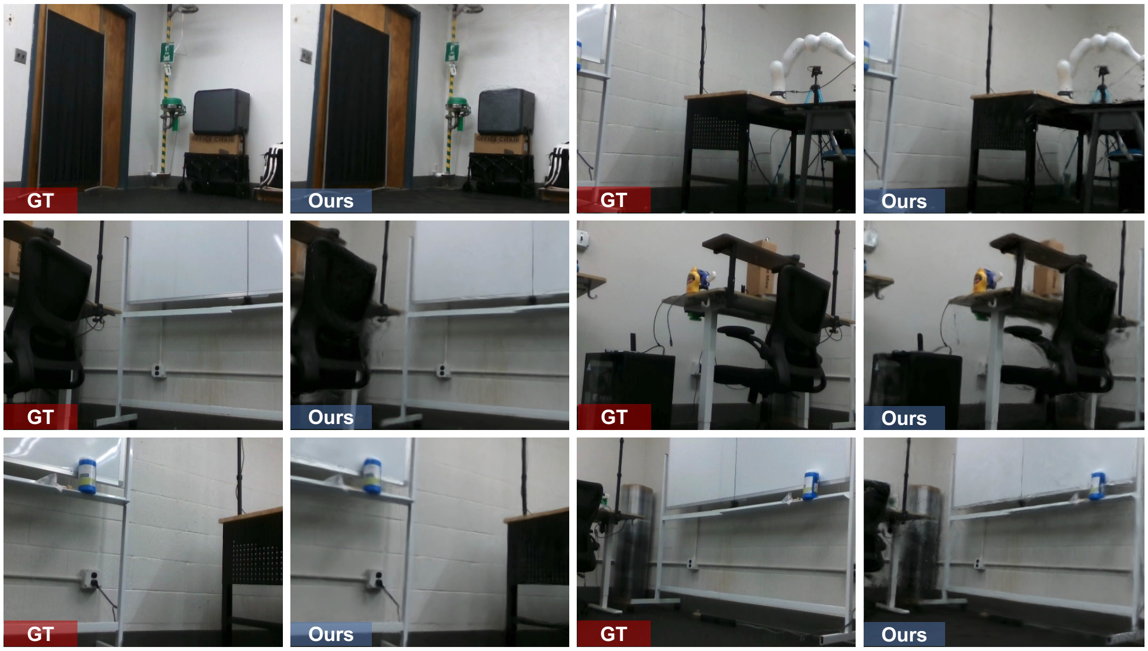

Figure 7: **Real scene versus TRACE reconstruction on the Unitree Go2.** The room-scale reconstruction, produced from the directly executed ergodic trajectory, recovers the scene geometry and appearance.**The supplementary video shows this in full.**

the raw map ϕ^{raw} over the footprint samples,

$$L_{\text{gaze}}(\mathbf{q}) = -\frac{1}{KD}\sum_{k=1}^{K}\sum_{d=1}^{D} w_{k,d}\,\phi^{\text{raw}}\big(\mathbf{f}_d(q^k)\big), \qquad (10)$$

with $w_{k,d}$ down-weighting occluded samples; its gradient turns $R(q)$ toward high-information surfaces. Position and orientation are therefore optimized together, both as functions of the single map FK.

Problem 1 (Joint-Space Ergodic Search)**.** Given an information map ϕ, minimize the objective over joint-velocity controls subject to the arm's kinematics:

$$\begin{aligned} \min_{\mathbf{u}}\ & E^{\text{dep}}_{\text{kernel}}(\mathbf{q}) + \lambda_{\text{g}} L_{\text{gaze}}(\mathbf{q}) + \lambda_{\text{s}} L_{\text{safe}}(\mathbf{q}) + \lambda_{\text{r}}\|\mathbf{u}\|^2 \\ \text{s.t. }\ & q^k = q^{k-1} + \Delta t\, u^k, \qquad q^k \in [q_{\min}, q_{\max}], \end{aligned}$$

where L_{safe} is the soft-barrier penalty of the full cost (7), evaluated on forward-kinematic body points.

Proposition 1 (Feasibility and Generality)**.** *The joint-space problem has three properties. First, the objective is differentiable in* $\mathbf{u}$*: the forward kinematics is differentiable and the kernel-ergodic and gaze terms are differentiable in the camera pose, so the arm reuses the quadruped's gradient-based optimizer unchanged. Second, every trajectory that respects the joint limits* $q^k \in [q_{\min}, q_{\max}]$ *maps to an achievable camera pose* $\text{FK}(q^k)$*, so every planned waypoint is executable without inverse kinematics or feasibility repair. Third, when the decision variable is the camera pose itself, the forward-kinematics map drops out and the problem reduces to the camera-pose objective of the main paper, recovering the quadruped case.*

Table 3: **Executing ActiveGS viewpoints directly on the FR3 (simulation).** Across 10 paired trials with diverse start poses under a shared mapper and budget, we command each planner's output on the arm without a path planner. Every ActiveGS trial collides with the scene and knocks the object off the table; TRACE stays collision-free.

Method	Collision-free trials
ActiveGS	$0/10$
TRACE (Ours)	$\mathbf{10/10}$

Figure 6 shows a representative mission on the real arm. TRACE follows the joint-space trajectory in one continuous sweep around the object, and every view executes without collision or intervention. Table 3 quantifies the contrast in simulation, where an unrealizable plan can be run safely. Across 10 paired trials under a shared mapper and budget, the camera poses ActiveGS selects **collide with the scene on every trial** and knock the object off the table, while TRACE, planning in joint space, **stays collision-free throughout.**

C Room-Scale Reconstruction on the Go2

Figure 7 compares the reconstruction TRACE produces on the Unitree Go2 with photo of the same scene. Over a 5-minute mission the quadruped executes the planned ergodic trajectory directly, and the online 2DGS map fills in furniture and wall surfaces across the room. The reconstruction is clean and complete, recovering geometry and texture on the surfaces the trajectory sweeps, which confirms the executed path is both **feasible and information-rich at room scale.**