

Motion-Uncertainty-Aware Next-Best-View Planning for Moving Object Reconstruction

Karen Li¹, Mattia Mantovani², Robert J. Wood¹, Lorenzo Sabattini², and Stephanie Gil¹

Abstract—Active 3D reconstruction of moving objects requires selecting informative viewpoints while accounting for object motion uncertainty during the decision-to-execution delay. Existing methods address only parts of this problem: next-best-view (NBV) planners for object reconstruction typically optimize surface coverage but assume static objects, while motion-aware active perception for moving targets accounts for target motion but prioritizes tracking or visibility over reconstruction coverage. This work presents a motion-uncertainty-aware NBV framework for reconstructing an unknown rigid object undergoing planar motion, using noisy planar position measurements of the object and depth observations from a mobile robot. The key idea is to evaluate each candidate viewpoint by its expected observation quality over plausible future object states induced by motion and measurement uncertainty, rather than at a single predicted object pose. To obtain this predictive belief, a fixed-lag Gaussian Process smoother estimates and predicts the object state from noisy position measurements. The resulting belief is used to generate candidate viewpoints around the predicted object location, filter them by reachability, and estimate their expected coverage-driven scores. Simulation and real-world experiments demonstrate improved reconstruction completeness over non-predictive NBV and prediction-only tracking methods, bridging coverage-driven active reconstruction and prediction-driven tracking.

I. INTRODUCTION

Active object reconstruction seeks to build a 3D model of an unknown object by selecting informative sensing viewpoints that maximize reconstruction progress, commonly measured by surface coverage [1–3]. Most next-best-view (NBV) methods in this setting assume static objects [1–6], so a viewpoint evaluated during planning is expected to remain informative when executed. This assumption no longer holds for moving objects: a viewpoint that appears informative during planning can become redundant or poorly aligned by execution time as object motion changes self-occlusion and visible surface regions. Thus, active reconstruction of moving objects requires reasoning not only about which viewpoints are informative, but also about how *uncertainty* in the object’s future state affects the camera-object configuration at the time of execution.

This problem frequently arises in field robotics, where robots must collect informative visual data from moving targets during opportunistic encounters [7]. Whale research, for example, uses visual observations during surfacing events to support health assessment and long-term monitoring [8–12]. While aerial top-down imaging can estimate quantities

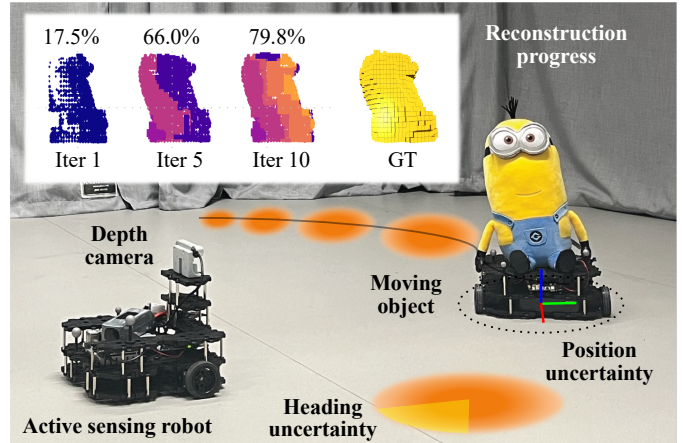


Fig. 1. Overview of the motion-uncertainty-aware NBV planning platform. A mobile robot equipped with a depth camera streams observations to an online offboard planner, which selects reachable viewpoints for reconstructing a rigid object undergoing planar motion. Orange ellipses indicate noisy object position measurements and predictive object-state uncertainty. The inset shows reconstruction progress over planning iterations, with voxel completeness increasing as new viewpoints observe previously unseen object surfaces.

such as body length or mass [10–12], building more complete individual-level 3D models requires broader surface coverage than a limited set of views can provide. This motivates an active perception problem: given coarse target-position measurements from an overhead sensing platform such as a drone [13], a sensing robot must *actively diversify viewpoints* to observe previously unseen surfaces as the target moves.

This need for diverse viewpoints under target motion is not fully addressed by existing work. Prior NBV methods [1–6] develop coverage-driven objectives for static object reconstruction, while moving-target active perception [14–17] typically focuses on tracking or visibility maintenance. To address this gap, this work studies online coverage-driven NBV planning for reconstructing an unknown rigid object undergoing planar motion in an open scene, treating the robot pose as known for planning and modeling uncertainty only in the object state over the planning-to-execution interval. This setting is challenging for three reasons: (i) viewpoint quality depends on the future camera-object configuration, so object motion can change self-occlusion and visible surface regions; (ii) coverage-driven viewpoint scores depend on nonlinear visibility and discretized surface representations [6, 18, 19], so their expectation under object-state uncertainty does not admit a closed-form expression; and (iii) the most informative viewpoint under the predictive belief may be unreachable within one planning step due to the robot’s kinematic limits.

¹Harvard University, Cambridge, MA 02134, USA. {karenli@g, rjwood@seas, sgil@seas}.harvard.edu

²University of Modena and Reggio Emilia, 42122 Reggio Emilia, Italy. {name.surname}@unimore.it

*We use target and object interchangeably.

Code can be found at <https://github.com/Harvard-REACT/mua-nbv>

Our key idea is to reformulate coverage-driven NBV planning for moving object reconstruction as a belief-space viewpoint selection problem. Rather than scoring each candidate at a single object pose, candidate viewpoints are evaluated using their expected coverage-driven score under a predictive belief over the execution-time object state. This belief is obtained by modeling the object trajectory with a Gaussian Process (GP) prior and applying a fixed-lag GP smoother [20–22] to noisy position-only measurements, yielding a Gaussian estimate of the object’s planar position and latent velocity. Propagating the belief one step forward gives the predictive distribution used for uncertainty-aware candidate generation around the predicted object location, followed by single-step reachability filtering. Since the expected coverage-driven score has no closed form, each reachable candidate’s score is approximated using Monte Carlo samples from the predictive belief, and the highest-scoring viewpoint is selected.

The proposed framework is evaluated against two baselines representing the paradigms it bridges: *non-predictive NBV*, which omits one-step object-state prediction, and *prediction-only tracking*, which follows the predicted object location without coverage-driven scoring. In both simulation and real-world experiments, the proposed framework improves reconstruction completeness by selecting viewpoints with high expected observation quality under the predicted object-state belief. The improvement over non-predictive NBV shows the value of accounting for object motion during the planning-to-execution interval, while the improvement over prediction-only tracking shows that following the predicted target is insufficient without coverage-driven viewpoint evaluation.

Our main contributions are multifold:

- A motion-uncertainty-aware NBV formulation for moving object reconstruction that integrates GP-based object-state prediction with coverage-driven viewpoint selection under execution-time object-state uncertainty.
- An online receding-horizon planning pipeline that uses the predictive belief for uncertainty-aware candidate generation, reachability filtering, and Monte Carlo estimation of the expected coverage-driven viewpoint score.
- Simulation and real-world validation showing improved reconstruction completeness over baseline methods, with ablations isolating the effects of candidate generation and viewpoint evaluation.

II. RELATED WORK

Our problem lies at the intersection of two lines of work: (i) online NBV planning for model-free object reconstruction, and (ii) predictive perception and view planning for moving targets. We bridge these lines by formulating moving object NBV planning as belief-space viewpoint selection, where a predictive belief over the execution-time object state is used to evaluate expected viewpoint quality. The distinctive challenge is that object motion during the planning-to-execution interval makes the execution-time camera-object configuration uncertain, so the planner must reason about how object-state uncertainty *propagates* into viewpoint quality.

A. NBV for Model-Free Object Reconstruction

NBV planning has a long history in model-free object reconstruction, where a robot incrementally builds an object representation and selects viewpoints that improve reconstruction completeness or surface coverage [1–5]. Many methods in this setting maintain a voxel-based object representation and evaluate frontier or visibility cues at the boundary between observed and unobserved regions, since this provides a natural representation of reconstruction progress [1, 18]. A direct way to score candidate viewpoints is to ray-cast through this map and count visible frontier, occupied, occluded, or unknown voxels as proxies for expected reconstruction gain [18, 23, 24].

These approaches are effective for static objects, where the camera-object configuration used for scoring remains valid at execution time. For moving objects, however, object motion can change self-occlusion and surface visibility between planning and execution, making a score computed at a single object pose unreliable. Thus, we retain the model-free, object-centric reconstruction representation, but evaluate viewpoints under a predictive distribution over the future object state.

Direct voxel-level scoring can be costly for online planning when many viewpoints must be evaluated repeatedly, especially under multiple samples of a predictive belief. We therefore instantiate the deterministic per-view score using PB-NBV [6], a projection-based surrogate for efficient coverage-driven viewpoint evaluation. PB-NBV is used as an existing deterministic score; our contribution is to evaluate such scores in expectation under a predictive object-state belief.

B. State Estimation, Prediction via Sparse Gaussian Processes

The belief-space NBV formulation above requires a predictive distribution over the object state at viewpoint execution time. In our setting, only noisy planar object-position measurements are available, while velocity is latent and is used to define object heading for viewpoint selection. Thus, the planner must infer both latent velocity and predictive uncertainty from recent observations. Filtering and smoothing methods are commonly used for this type of state-estimation problem [25–27], providing state estimates and predictive uncertainty from noisy partial measurements. We use a fixed-lag smoother to estimate position and velocity jointly over a short measurement window, a common strategy for online state estimation [28, 29].

Because the planner needs not only a point prediction but also uncertainty over the object’s execution-time state, we model the object trajectory probabilistically. Sparse GP trajectory priors [20–22, 30] provide a lightweight continuous-time model for this short-horizon prediction problem. Continuous-time trajectory estimation supports prediction at arbitrary execution times within a replanning loop [31]. These methods exploit the Markov structure induced by stochastic differential equation priors for efficient inference and uncertainty queries in continuous time, making them attractive for bounded online prediction. We use this machinery in fixed-lag form as an enabling prediction module, where the resulting predictive belief over object state is then consumed by the NBV planner.

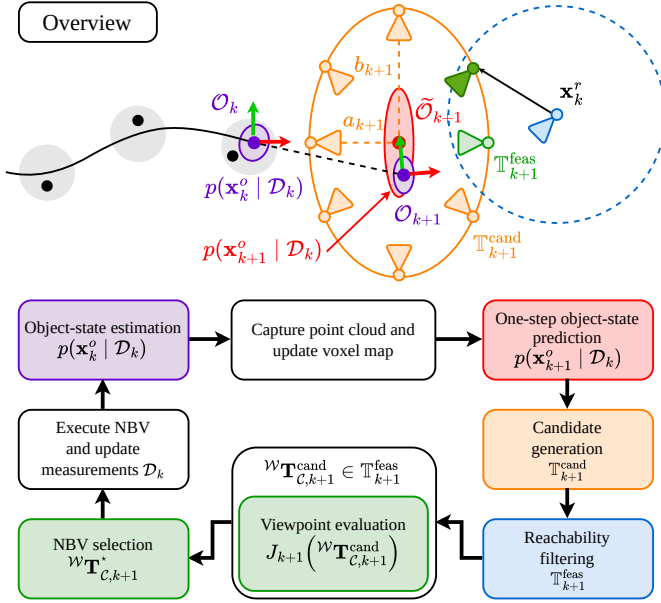


Fig. 2. **Motion-uncertainty-aware NBV online planning loop.** Given noisy object-position measurements (black dots), the GP-based smoother computes the **current belief** and propagates it forward to the **predictive belief**. The **candidate generation** module samples viewpoints on an uncertainty-adaptive ellipse around the predicted object-position mean. The **reachability filtering** module keeps only candidates reachable within one planning step. The **viewpoint evaluation** module estimates each reachable candidate's expected coverage-driven score under the predictive belief. The selected action is the reachable candidate with the highest expected score; after execution, the new position measurement and point cloud are incorporated, and the loop repeats.

C. View Planning for Moving Targets

For moving targets, many active perception systems [14–17] select viewpoints to maintain visibility or improve observability rather than maximize reconstruction coverage. For example, ActiveMoCap [14] selects viewpoints to improve 3D human pose and motion-capture accuracy from a moving camera. Aerial tracking systems [15, 32, 33] similarly plan viewpoints to maintain observability of moving targets. Visibility-aware tracking methods also plan under target-state uncertainty in cluttered environments [34], but the objective remains target tracking. These methods account for target motion or state uncertainty, but they do not optimize viewpoints for coverage-driven reconstruction of previously unseen object surfaces.

Closer to our setting, prior work on moving object 3D reconstruction considers object motion during viewpoint selection [35]. However, viewpoint quality is evaluated at a single nominal object pose, which is fragile when delay and noisy measurements induce uncertainty over the execution-time object state. In contrast, our proposed motion-uncertainty-aware NBV framework uses a predictive object-state belief to evaluate coverage-driven viewpoint quality, explicitly accounting for object-state uncertainty induced by object motion.

III. PROBLEM SETUP

We consider active reconstruction of an unknown rigid object undergoing planar motion. At each discrete planning time t_k , the planner receives two inputs: a depth observation from the sensing robot's calibrated camera and a noisy planar object-position measurement from a separate sensing source.

A. Notation and Reference Frames

Let $\mathbf{x}^o(t)$ denote the continuous-time object state, with $\mathbf{x}_k^o := \mathbf{x}^o(t_k)$ its value at planning time t_k . Measurement incorporation, estimation updates, viewpoint selection, and execution are indexed by the discrete planning times $\{t_k\}$.

The world, robot base, and camera optical frames are denoted by $\mathcal{W}$, $\mathcal{R}$, and $\mathcal{C}$, respectively. The reconstruction proxy (Sec. IV-C) is maintained in an object-centric map frame $\mathcal{O}_k$ at t_k . During planning at time t_k , a temporary predicted object frame $\tilde{\mathcal{O}}_{k+1}$ is created for the next planning time t_{k+1} . This frame is derived from the one-step predictive belief and is used only to select viewpoints for execution at t_{k+1} .

After the robot executes the selected motion and the planner incorporates the new object-position measurement at t_{k+1} , the next object-centric map frame $\mathcal{O}_{k+1}$ is defined and the voxel map is updated in that frame. Because $\tilde{\mathcal{O}}_{k+1}$ is defined from the prediction before the measurement update, whereas $\mathcal{O}_{k+1}$ is defined after incorporating the measurement at t_{k+1} , the two frames generally do not coincide.

The rigid-body transform that maps coordinates expressed in frame $\mathcal{B}$ into frame $\mathcal{A}$ is denoted by ${}^{\mathcal{A}}\mathbf{T}_{\mathcal{B}} \in SE(3)$, with rotation ${}^{\mathcal{A}}\mathbf{R}_{\mathcal{B}} \in SO(3)$ and translation ${}^{\mathcal{A}}\mathbf{t}_{\mathcal{B}} \in \mathbb{R}^3$.

B. Robot model

At planning time t_k , the robot's planar pose is denoted by

$$\mathbf{x}_k^r := [\mathbf{p}_k^r, \theta_k^r]^\top \in \mathbb{R}^2 \times S^1. \quad (1)$$

The camera pose is obtained from the robot base pose ${}^{\mathcal{W}}\mathbf{T}_{\mathcal{R}}$ and known extrinsics ${}^{\mathcal{R}}\mathbf{T}_{\mathcal{C}}$ as ${}^{\mathcal{W}}\mathbf{T}_{\mathcal{C}} = {}^{\mathcal{W}}\mathbf{T}_{\mathcal{R}} {}^{\mathcal{R}}\mathbf{T}_{\mathcal{C}}$.

C. Object state and measurements

The object state consists of planar position and velocity,

$$\mathbf{x}^o(t) := [\mathbf{p}^o(t), \dot{\mathbf{p}}^o(t)]^\top \in \mathbb{R}^4. \quad (2)$$

At each planning time t_k , a separate sensing source provides a noisy measurement of the object's planar position,

$$\tilde{\mathbf{p}}_k^o = \mathbf{p}_k^o + \mathbf{n}_k, \quad \mathbf{n}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}), \quad (3)$$

where $\mathbf{R} = \sigma^2 \mathbf{I} \in \mathbb{R}^{2 \times 2}$ is the planar position measurement-noise covariance. Velocity is not directly observed and is inferred as a latent component of the state via fixed-lag GP smoothing (Sec. IV-A). The object heading is assumed to align with the object's tangential direction of motion.

IV. METHODOLOGY

We use an online receding-horizon planning loop that combines object-state prediction with coverage-driven NBV selection (Fig. 2). At each planning time t_k , the planner incorporates the latest depth and object-position measurements, propagates the object state one step ahead to $t_{k+1} = t_k + \Delta t$ (Sec. IV-A), and samples candidate viewpoints around the predicted object location and filters them by reachability (Sec. IV-B). Each reachable viewpoint is then evaluated by Monte Carlo estimation of its PB-NBV score under the predictive object-state belief (Sec. IV-C and Sec. IV-D). The viewpoint with the highest expected score is executed, and the loop repeats as new measurements arrive.

A. Object State Estimation and Prediction

At each planning time t_k , a fixed-lag sparse GP smoother estimates a Gaussian belief over the object state $\mathbf{x}_k^o$ and predicts this belief one step ahead to t_{k+1} . Because only planar position is measured, velocity is treated as a latent state. The resulting predictive belief is used for candidate viewpoint generation and evaluation by the online NBV planner.

1) *Fixed-lag GP-based smoother*: Following sparse GP trajectory estimation [20, 21], the object trajectory is modeled using a white-noise acceleration constant-velocity prior,

$$\ddot{\mathbf{p}}^o(t) = \mathbf{w}(t), \quad \mathbf{w}(t) \sim \mathcal{GP}(\mathbf{0}, \mathbf{Q}_C \delta(t - t')), \quad (4)$$

where $\mathbf{Q}_C = q_c \mathbf{I} \in \mathbb{R}^{2 \times 2}$ is the continuous-time acceleration-noise power spectral density, and $\delta(\cdot)$ denotes the Dirac delta.

At time t_k , the position-only measurement model is

$$p(\tilde{\mathbf{p}}_k^o | \mathbf{x}_k^o) = \mathcal{N}(\tilde{\mathbf{p}}_k^o; \mathbf{H}\mathbf{x}_k^o, \mathbf{R}), \quad \mathbf{H} = [\mathbf{I} \ \mathbf{0}] \in \mathbb{R}^{2 \times 4}.$$

A fixed-lag Maximum A Posteriori smoother is then applied to the most recent L measurements $\mathcal{D}_k = \{\tilde{\mathbf{p}}_{k-L+1}^o, \dots, \tilde{\mathbf{p}}_k^o\}$.

Let $\boldsymbol{\mu}_k^o$ and $\boldsymbol{\Sigma}_k^o$ denote the marginal mean and covariance of the current object state $\mathbf{x}_k^o$ obtained from the fixed-lag posterior; the current belief is

$$p(\mathbf{x}_k^o | \mathcal{D}_k) = \mathcal{N}(\mathbf{x}_k^o; \boldsymbol{\mu}_k^o, \boldsymbol{\Sigma}_k^o). \quad (5)$$

2) *One-step GP predictor*: For a time increment Δt , the motion prior in Eq. 4 yields the closed-form transition

$$\Phi(\Delta t) = \begin{bmatrix} \mathbf{I} & \Delta t \mathbf{I} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}, \quad \mathbf{Q}(\Delta t) = \begin{bmatrix} \frac{1}{3} \Delta t^3 \mathbf{Q}_C & \frac{1}{2} \Delta t^2 \mathbf{Q}_C \\ \frac{1}{2} \Delta t^2 \mathbf{Q}_C & \Delta t \mathbf{Q}_C \end{bmatrix}.$$

Given the current belief $p(\mathbf{x}_k^o | \mathcal{D}_k)$ from Eq. 5, the one-step predictive mean and covariance are

$$\boldsymbol{\mu}_{k+1|k}^o = \Phi(\Delta t) \boldsymbol{\mu}_k^o, \quad (6)$$

$$\boldsymbol{\Sigma}_{k+1|k}^o = \Phi(\Delta t) \boldsymbol{\Sigma}_k^o \Phi(\Delta t)^\top + \mathbf{Q}(\Delta t). \quad (7)$$

Candidate viewpoint generation and evaluation at planning step k are conditioned on the predictive Gaussian

$$p(\mathbf{x}_{k+1}^o | \mathcal{D}_k) = \mathcal{N}(\mathbf{x}_{k+1}^o; \boldsymbol{\mu}_{k+1|k}^o, \boldsymbol{\Sigma}_{k+1|k}^o). \quad (8)$$

This predictive belief contains both planar position and velocity components, together with their covariance.

3) *Object nominal heading*: Following the tangential-heading assumption in Sec. III, the object's nominal heading is derived from the predicted mean velocity:

$$\hat{\theta}_{k+1}^o = \text{atan2}((\boldsymbol{\mu}_{\mathbf{p}, k+1|k}^o)_y, (\boldsymbol{\mu}_{\mathbf{p}, k+1|k}^o)_x), \quad (9)$$

where $\boldsymbol{\mu}_{\mathbf{p}, k+1|k}^o$ denotes the velocity block of the predicted mean in (6). This expression assumes $\|\boldsymbol{\mu}_{\mathbf{p}, k+1|k}^o\| > 0$; when the speed is below a small threshold, the previous nominal heading is retained. The nominal heading $\hat{\theta}_{k+1}^o$ is used only to define the predicted scoring frame $\tilde{\mathcal{O}}_{k+1}$ for candidate generation (Sec. IV-B); uncertainty in the object state is handled explicitly during viewpoint evaluation (Sec. IV-D).

B. Prediction-Conditioned Candidate Viewpoint Generation

Using the predictive belief in Eq. 8, we sample candidate viewpoints on an uncertainty-adaptive ellipse that preserves a desired stand-off distance while expanding along directions of larger positional uncertainty.

1) *Uncertainty-adaptive ellipse discretization*: We define a predicted object-centric scoring frame $\tilde{\mathcal{O}}_{k+1}$ for viewpoint generation and evaluation at the future time t_{k+1} . Let

$$\mathbf{c}_{k+1} := \boldsymbol{\mu}_{\mathbf{p}, k+1|k}^o \quad (10)$$

denote the predicted mean position extracted from the position block of (6). The planar origin of $\tilde{\mathcal{O}}_{k+1}$ is located at $\mathbf{c}_{k+1}$, and its x -axis is aligned with the nominal predicted heading derived in Eq. 9. This defines the planar rotation

$$\mathcal{W}_{\mathbf{R}_{\tilde{\mathcal{O}}, k+1}} := \begin{bmatrix} \cos(\hat{\theta}_{k+1}^o) & -\sin(\hat{\theta}_{k+1}^o) \\ \sin(\hat{\theta}_{k+1}^o) & \cos(\hat{\theta}_{k+1}^o) \end{bmatrix} \in SO(2).$$

Inspired by uncertainty-aware observation strategies [36], we generate candidate viewpoints on a planar ellipse centered at $\mathbf{c}_{k+1}$ (10), with semi-axes adapted to the predicted positional uncertainty. The idea is to enlarge the candidate search radius along directions where the future object position is more uncertain, while preserving a nominal stand-off distance when uncertainty is low. Let $\boldsymbol{\Sigma}_{\mathbf{p}, k+1|k}^o \in \mathbb{R}^{2 \times 2}$ denote the position block of the predicted covariance in (7). The positional covariance expressed in the coordinates of $\tilde{\mathcal{O}}_{k+1}$ is

$$\tilde{\boldsymbol{\Sigma}}_{\mathbf{p}, k+1|k}^o := (\mathcal{W}_{\mathbf{R}_{\tilde{\mathcal{O}}, k+1}})^\top \boldsymbol{\Sigma}_{\mathbf{p}, k+1|k}^o \mathcal{W}_{\mathbf{R}_{\tilde{\mathcal{O}}, k+1}}.$$

We use the diagonal of $\tilde{\boldsymbol{\Sigma}}_{\mathbf{p}, k+1|k}^o$ as heading-aligned marginal variances and define forward and lateral standard deviations as the square roots of the diagonal entries,

$$\sigma_{\parallel, k+1} := \sqrt{(\tilde{\boldsymbol{\Sigma}}_{\mathbf{p}, k+1|k}^o)_{11}}, \quad \sigma_{\perp, k+1} := \sqrt{(\tilde{\boldsymbol{\Sigma}}_{\mathbf{p}, k+1|k}^o)_{22}}.$$

The ellipse semi-axes are given by

$$a_{k+1} := r_{\text{off}} + \kappa \sigma_{\parallel, k+1}, \quad b_{k+1} := r_{\text{off}} + \kappa \sigma_{\perp, k+1},$$

where $r_{\text{off}} > 0$ is a desired stand-off distance chosen from the camera sensing range, and $\kappa > 0$ is an uncertainty-inflation factor that trades off coverage of likely object locations against reachability. Uncertainty is reflected through anisotropic scaling of the viewpoint generation ellipse axes, while the ellipse orientation is governed by the nominal motion direction (Eq. 9) for stable and interpretable viewpoint placement.

The ellipse is discretized using I uniformly spaced azimuth angles $\varphi_i = 2\pi(i-1)/I$, for $i \in \{1, \dots, I\}$. Each index i defines one candidate viewpoint on the ellipse.

2) *Candidate viewpoint positions and yaw*: For each azimuth index i , the planar offset in the predicted object-centric scoring frame $\tilde{\mathcal{O}}_{k+1}$ is defined as

$$\mathbf{q}_{k+1, i} := \begin{bmatrix} a_{k+1} \cos \varphi_i \\ b_{k+1} \sin \varphi_i \end{bmatrix} \in \mathbb{R}^2.$$

The corresponding planar candidate position in $\mathcal{W}$ is then

$$\mathbf{p}_{k+1, i}^{\text{cand}} := \mathbf{c}_{k+1} + \mathcal{W}_{\mathbf{R}_{\tilde{\mathcal{O}}, k+1}} \mathbf{q}_{k+1, i} \in \mathbb{R}^2, \quad (11)$$

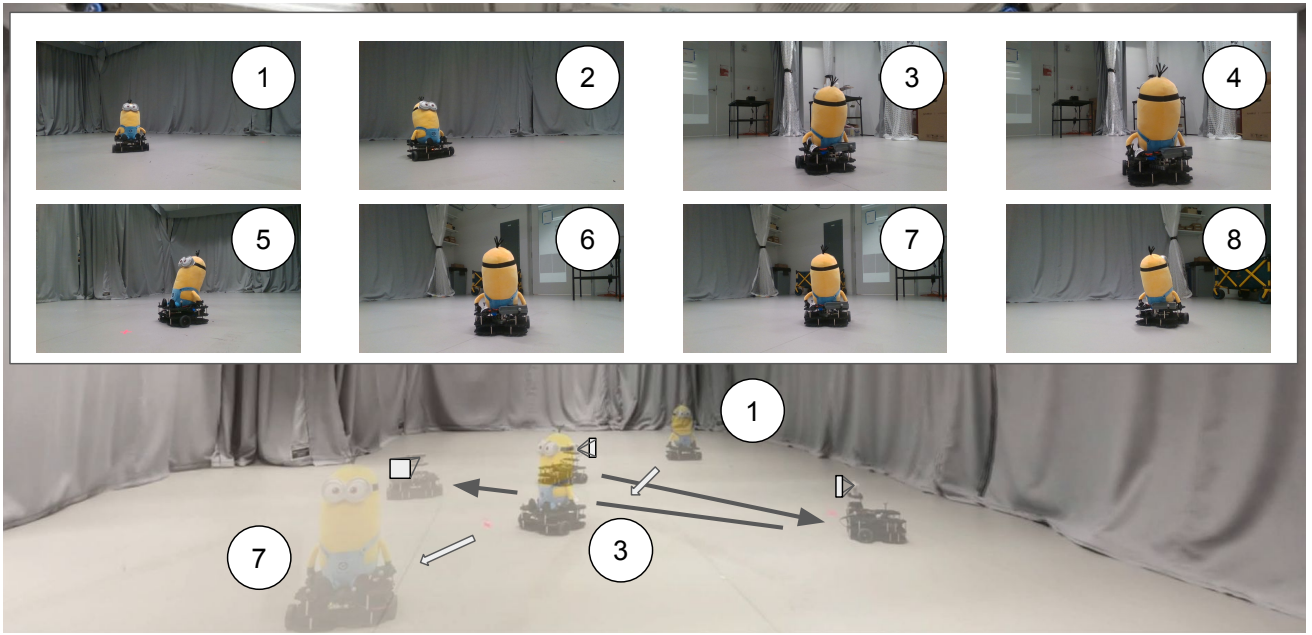


Fig. 3. **Real-world replanning sequence from one representative trial.** **Top:** RGB-camera views acquired over eight successive planning iterations for visual illustration. **Bottom:** corresponding robot and object configurations overlaid in a common frame; numbered markers indicate selected iterations. The sequence shows that the robot does not simply follow the moving object, but changes perspectives to observe new object surfaces as reconstruction progresses.

with an assigned inward-facing yaw that points toward the predicted object center $\mathbf{c}_{k+1}$ (10):

$$\psi_{k+1,i}^{\text{cand}} := \text{atan2}\left((\mathbf{c}_{k+1} - \mathbf{p}_{k+1,i}^{\text{cand}})_y, (\mathbf{c}_{k+1} - \mathbf{p}_{k+1,i}^{\text{cand}})_x\right). \quad (12)$$

Embedding the planar candidate position and yaw as a robot-base pose with zero roll and pitch gives

$$\mathcal{W}\mathbf{T}_{\mathcal{R},k+1,i}^{\text{cand}} := \begin{pmatrix} R_z(\psi_{k+1,i}^{\text{cand}}) & \begin{bmatrix} \mathbf{p}_{k+1,i}^{\text{cand}} \\ 0 \\ 1 \end{bmatrix} \\ \mathbf{0}^\top & 1 \end{pmatrix}.$$

Applying the known camera extrinsics ${}^{\mathcal{R}}\mathbf{T}_{\mathcal{C}}$ yields the corresponding candidate camera pose

$$\mathcal{W}\mathbf{T}_{\mathcal{C},k+1,i}^{\text{cand}} := \mathcal{W}\mathbf{T}_{\mathcal{R},k+1,i}^{\text{cand}} {}^{\mathcal{R}}\mathbf{T}_{\mathcal{C}} \in SE(3).$$

Collecting all camera poses, we obtain the candidate set

$$\mathbb{T}_{k+1}^{\text{cand}} := \{\mathcal{W}\mathbf{T}_{\mathcal{C},k+1,i}^{\text{cand}} \mid i \in \{1, \dots, I\}\}.$$

3) *Reachability filtering*: Not all candidate viewpoints are attainable within a single planning step. We therefore filter candidates according to the current robot state $\mathbf{x}_k^r$ and bounded unicycle motion limits. A candidate with planar position $\mathbf{p}_{k+1,i}^{\text{cand}}$ (11) and yaw $\psi_{k+1,i}^{\text{cand}}$ (12) is feasible if

$$\|\mathbf{p}_{k+1,i}^{\text{cand}} - \mathbf{p}_k^r\| \leq v_{\max} \Delta t, \quad |\psi_{k+1,i}^{\text{cand}} - \theta_k^r| \leq \omega_{\max} \Delta t. \quad (13)$$

The candidates satisfying (13) form the feasible set $\mathbb{T}_{k+1}^{\text{feas}} \subseteq \mathbb{T}_{k+1}^{\text{cand}}$, which is passed to the NBV evaluation in Sec. IV-D.

C. Coverage-Driven NBV Evaluation

The objective is to improve surface coverage of an initially unknown object during online replanning. We therefore use an efficient coverage-driven NBV proxy that scores viewpoints by the amount of new surface they are expected to reveal from an object-centric bounded volumetric representation.

1) *Object-centric bounded voxel proxy*: As the robot moves, point clouds are registered to $\mathcal{O}_k$ and incrementally fused into a bounded object-centric voxel grid. This bounded-domain representation is widely used in active 3D reconstruction to keep mapping and viewpoint evaluation tractable [6, 18, 37]. Within this voxel proxy, voxels are labeled as *Occupied* (observed surface), *Free* (empty space), and *Unknown* (unobserved). *Frontier* voxels are extracted at the boundary between the observed region, consisting of Occupied and Free voxels, and the Unknown region, indicating where additional measurements are likely to add new information [38–40].

2) *Efficient voxel map proxy*: Given a fixed relative pose ${}^{\mathcal{O}}\mathbf{T}_{\mathcal{C}}$, we evaluate a deterministic coverage-driven viewpoint score $S({}^{\mathcal{O}}\mathbf{T}_{\mathcal{C}})$. In our implementation, $S(\cdot)$ is instantiated using the projection-based PB-NBV score [6] computed from the current object-centric voxel proxy. Our framework is agnostic to the choice of deterministic coverage-driven NBV score; PB-NBV is used as an efficient per-view scoring surrogate.

We consider two voxel sets from the bounded voxel grid: Frontier voxels and Occupied voxels. Following PB-NBV [6], each voxel set is clustered using Gaussian mixture models, and each cluster is approximated by a minimum-volume enclosing ellipsoid. This yields two ellipsoid collections,

$${}^{\mathcal{O}}\mathcal{E}_F = \{{}^{\mathcal{O}}E_{F,u}\}_{u=1}^{K_F}, \quad {}^{\mathcal{O}}\mathcal{E}_O = \{{}^{\mathcal{O}}E_{O,v}\}_{v=1}^{K_O}$$

where K_F and K_O are the numbers of Frontier and Occupied clusters, respectively. These collections summarize where additional surface is likely to be revealed and where surface has already been observed.

3) *Projection-based pixel-area score*: For a candidate relative pose ${}^{\mathcal{O}}\mathbf{T}_{\mathcal{C}}$, each ellipsoid is transformed to the camera frame using ${}^{\mathcal{C}}\mathbf{T}_{\mathcal{O}} = ({}^{\mathcal{O}}\mathbf{T}_{\mathcal{C}})^{-1}$. A point $\mathbf{x}_{\mathcal{O}}$ on ellipsoid E

is mapped to the camera frame as $\mathbf{x}_C = {}^C\mathbf{R}_O \mathbf{x}_O + {}^C\mathbf{t}_O$ and projected into the image using the calibrated intrinsics $\mathbf{K}$:

$$\tilde{\mathbf{u}} \sim \mathbf{K}({}^C\mathbf{R}_O \mathbf{x}_O + {}^C\mathbf{t}_O), \quad (\mathbf{x}_O - \mathbf{m}_E)^\top \mathbf{A}_E (\mathbf{x}_O - \mathbf{m}_E) \leq 1, \quad (14)$$

where $\tilde{\mathbf{u}} = (u, v, 1)^\top$ denotes homogeneous image coordinates, and $\mathbf{m}_E$ and $\mathbf{A}_E$ denote the ellipsoid center and shape matrix in $\mathcal{O}$. The projected silhouette region Ω_E is obtained by projecting the ellipsoid into the image and rasterizing the filled image region [41]. Large projected Frontier regions correspond to viewpoints expected to observe more currently unknown surface, whereas large projected Occupied regions indicate views dominated by already observed surface.

To approximate visibility ordering along the camera optical axis, the ellipsoids $\mathcal{O}_{\mathcal{F}} \cup \mathcal{O}_{\mathcal{O}}$ are sorted by increasing depth. Let $r(E)$ denote the resulting depth rank of ellipsoid E , with $r(E) = 1$ corresponding to the closest ellipsoid. A depth-ordering weight is assigned as

$$w_E = \alpha^{r(E)-1}, \quad \alpha \in (0, 1], \quad (15)$$

so that farther ellipsoids contribute less to the score. Let $\Omega_{F,u}$ and $\Omega_{O,v}$ denote the projected image regions of Frontier and Occupied ellipsoids, with corresponding weights $w_{F,u}$ and $w_{O,v}$. The deterministic PB-NBV score [6] is then given by

$$S({}^O\mathbf{T}_C) = \sum_{u=1}^{K_F} w_{F,u} |\Omega_{F,u}| - \sum_{v=1}^{K_O} w_{O,v} |\Omega_{O,v}|, \quad (16)$$

where $|\Omega|$ denotes the projected pixel area. The score rewards projected Frontier area, which approximates newly observable surface, and penalizes projected Occupied area, which corresponds to already observed surface.

D. Motion-Uncertainty-Aware NBV Planning

We select the NBV under uncertainty in the object state at execution time t_{k+1} . The key idea is to evaluate candidate viewpoints under the distribution of future camera-object configurations induced by the predictive object-state belief, rather than at a single predicted pose.

1) *Camera-object configuration under belief sampling*: We draw N i.i.d. samples $\{\mathbf{x}_{k+1}^{o,(n)}\}_{n=1}^N$ from the predictive belief in Eq. 8. From each sample, the planar position $\mathbf{p}_{k+1}^{o,(n)}$ and velocity $\dot{\mathbf{p}}_{k+1}^{o,(n)}$ are extracted. For samples with nonzero speed, the corresponding heading is

$$\theta_{k+1}^{o,(n)} = \text{atan2}\left((\dot{\mathbf{p}}_{k+1}^{o,(n)})_y, (\dot{\mathbf{p}}_{k+1}^{o,(n)})_x\right).$$

When the sampled speed is below a small threshold, the nominal heading from Eq. 9 is retained. Each sample then induces a world-frame pose of the predicted object frame,

$$\mathcal{W}\mathbf{T}_{\tilde{\mathcal{O}},k+1}^{(n)} := \begin{pmatrix} R_z(\theta_{k+1}^{o,(n)}) & \begin{bmatrix} \mathbf{p}_{k+1}^{o,(n)} \\ 0 \end{bmatrix} \\ \mathbf{0}^\top & 1 \end{pmatrix} \in SE(3).$$

For a fixed candidate camera pose $\mathcal{W}\mathbf{T}_{C,k+1}^{\text{cand}}$, the sample-induced relative transform is

$$\tilde{\mathcal{O}}_{k+1}^{(n)} \mathbf{T}_{C,k+1}^{\text{cand},(n)} := (\mathcal{W}\mathbf{T}_{\tilde{\mathcal{O}},k+1}^{(n)})^{-1} \mathcal{W}\mathbf{T}_{C,k+1}^{\text{cand}}.$$

With known robot pose, object-state uncertainty appears as variability in this relative camera-object configuration.

2) *Expected NBV score*: For each feasible candidate $\mathcal{W}\mathbf{T}_{C,k+1}^{\text{cand}} \in \mathbb{T}_{k+1}^{\text{feas}}$, the uncertainty-aware score is defined as the predictive expectation of the deterministic NBV score in Eq. 16. For an object state $\mathbf{x}_{k+1}^o$, let $\mathcal{W}\mathbf{T}_{\tilde{\mathcal{O}}}(\mathbf{x}_{k+1}^o)$ denote the predicted object-frame pose constructed from its position and velocity using the tangential-heading rule above. For a fixed candidate pose, the induced relative transform is

$$\tilde{\mathcal{O}} \mathbf{T}_C^{\text{rel}}(\mathbf{x}_{k+1}^o) := (\mathcal{W}\mathbf{T}_{\tilde{\mathcal{O}}}(\mathbf{x}_{k+1}^o))^{-1} \mathcal{W}\mathbf{T}_{C,k+1}^{\text{cand}}.$$

The motion-uncertainty-aware NBV score is then

$$J_{k+1}(\mathcal{W}\mathbf{T}_{C,k+1}^{\text{cand}}) := \mathbb{E}_{\mathbf{x}_{k+1}^o} \left[S(\tilde{\mathcal{O}} \mathbf{T}_C^{\text{rel}}(\mathbf{x}_{k+1}^o)) \right]. \quad (17)$$

This expectation admits no closed form because $S(\cdot)$ (Eq. 16) depends on nonlinear camera projection, discretization, and depth-ordering operations (Eq. 14 and Eq. 15), which make the integrand in Eq. 17 non-smooth in the object state. By the law of large numbers, belief-space samples from $p(\mathbf{x}_{k+1}^o | \mathcal{D}_k)$ provide a consistent Monte Carlo estimate of the predictive expectation. We therefore approximate (17) by Monte Carlo sampling:

$$J_{k+1}(\mathcal{W}\mathbf{T}_{C,k+1}^{\text{cand}}) \approx \frac{1}{N} \sum_{n=1}^N S(\tilde{\mathcal{O}} \mathbf{T}_C^{\text{rel}}(\mathbf{x}_{k+1}^{o,(n)})). \quad (18)$$

3) *Motion-Uncertainty-Aware NBV Selection*: The next viewpoint is selected by maximizing the expected motion-uncertainty-aware NBV score over the feasible candidate set:

$$\mathcal{W}\mathbf{T}_{C,k+1}^* := \arg \max_{\mathcal{W}\mathbf{T}_{C,k+1}^{\text{cand}} \in \mathbb{T}_{k+1}^{\text{feas}}} J_{k+1}(\mathcal{W}\mathbf{T}_{C,k+1}^{\text{cand}}). \quad (19)$$

This policy favors viewpoints with high expected coverage-driven score under future object-state uncertainty.

Remark

The method uses object-state prediction twice: (i) to shape candidate geometry according to the object-state uncertainty, and (ii) to evaluate viewpoints by expected observation quality over the predictive belief.

V. EXPERIMENTS

A. Experimental Goals and Metrics

The experiments evaluate whether motion-uncertainty-aware NBV improves coverage-driven reconstruction of a moving object. They are organized around three questions: (i) whether one-step predictive belief-space evaluation improves reconstruction over scoring viewpoints at the current estimated object state; (ii) whether coverage-driven viewpoint selection provides benefits beyond prediction-only tracking; and (iii) how uncertainty-aware candidate generation and Monte Carlo belief-space evaluation contribute to performance.

We use a voxelized surface-completeness metric inspired by standard reconstruction completeness measures [42]. For both simulation and real-world experiments, the ground-truth object

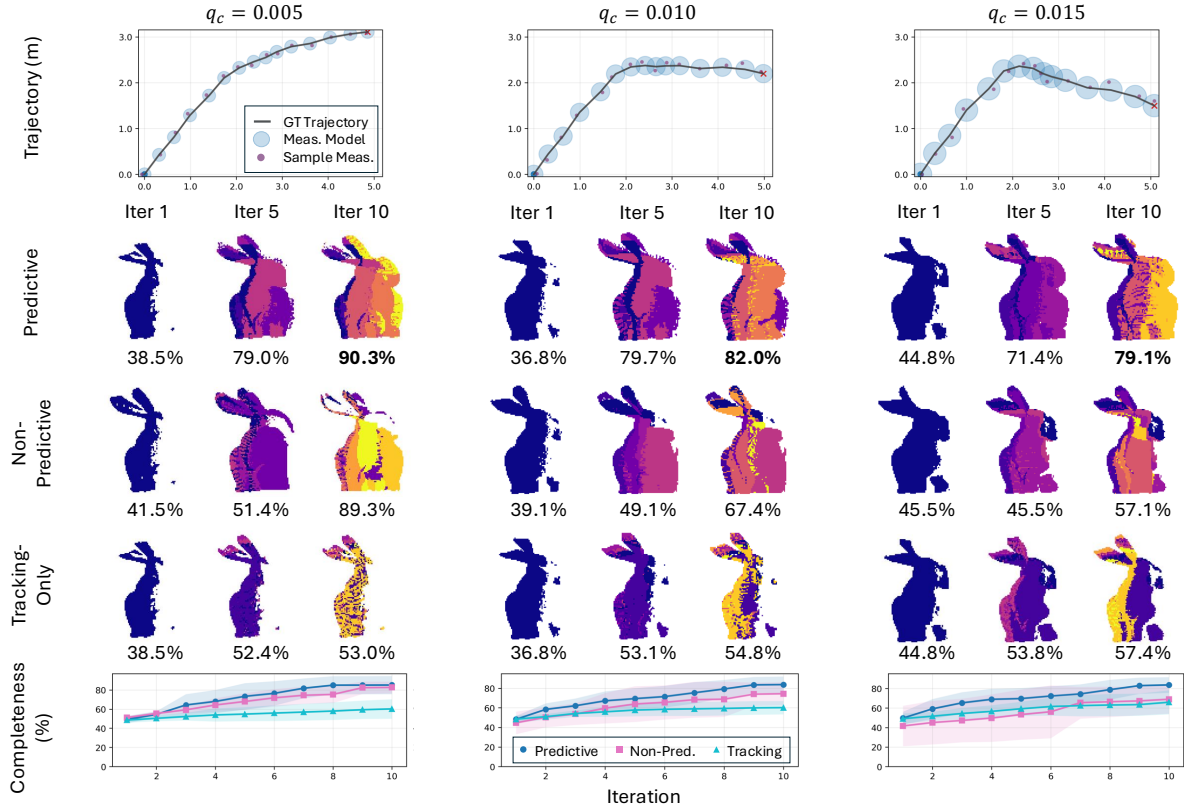


Fig. 4. **Simulation reconstruction under varying process noise.** The process-noise level increases across columns, $q_c \in \{0.005, 0.010, 0.015\}$, with measurement noise fixed at 0.10 m. **Top:** representative ground-truth (GT) object trajectories, position-uncertainty regions, and noisy position measurements. **Middle:** representative reconstructions at iterations 1, 5, and 10 for the three viewpoint selection methods; numbers report reconstruction completeness, and colors indicate the surface revealed by each iteration. **Bottom:** completeness over planning iterations, reported as mean $\pm$ std over 10 random seeds.

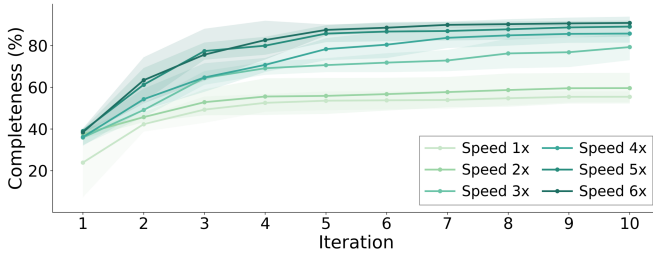


Fig. 5. **Effect of robot-object relative speed in simulation.** The speed factor scales the robot’s maximum per-step travel distance relative to the object’s displacement. Curves report reconstruction completeness as mean $\pm$ std over 10 random-seeded trajectories with fixed process and measurement noise.

point cloud is constructed by merging depth observations from robot-accessible viewpoints uniformly spaced around the object at 15° angular increments. At each planning iteration k , the accumulated reconstructed points are transformed into the object-centric evaluation frame and voxelized to form $V_{1:k}$. The ground-truth object point cloud is voxelized at the same resolution to form V_{gt} . Completeness is defined as

$$\text{Comp}(k) = 100 \cdot \frac{|V_{1:k} \cap V_{gt}|}{|V_{gt}|}.$$

This metric reports the percentage of ground-truth object voxels observed by the reconstruction.

Unless otherwise stated, all compared methods use the same object-position measurements, fixed-lag GP smoother window size, candidate discretization, reachability constraints, and voxel-map update procedure. They differ only in how

candidate viewpoints are generated, scored, or selected.

We compare the proposed method with two baselines representing the paradigms our framework bridges:

- 1) **Predictive NBV (ours)** selects viewpoints by maximizing the expected coverage-driven score under the one-step predictive object-state belief.
- 2) **Non-predictive NBV** evaluates viewpoints under the current object-state belief, without propagating the belief to execution time.
- 3) **Tracking-only** uses the predicted object state but selects the feasible candidate closest to the predicted object location, without coverage-driven NBV scoring.

The comparison with Non-predictive NBV isolates the effect of accounting for object motion during the decision-to-execution interval, while the comparison with Tracking-only isolates the effect of coverage-driven viewpoint selection beyond following the predicted target.

B. Simulation Experiments

The simulations evaluate how reconstruction performance changes with object process noise, robot reachability limits, and the main components of the proposed planner. Experiments are run online in Gazebo on Ubuntu 24.04 with ROS 2 Jazzy, using a workstation equipped with a 12th Gen Intel Core i9-12900K. All planning is performed on CPU, with subsecond computation per planning iteration. A mobile robot with a depth camera replans viewpoints to reconstruct a rigid object undergoing planar motion.

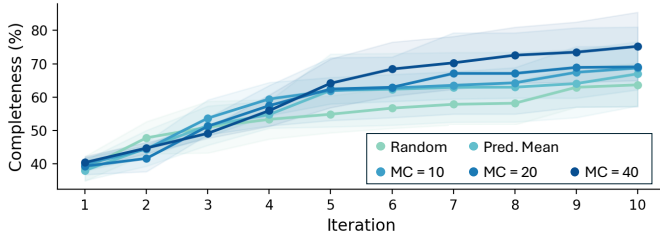


Fig. 6. **Ablation of viewpoint evaluation under the predictive belief.** *Random* selects uniformly from the feasible candidate set, *Pred. Mean* scores only at the predicted object-state mean, and MC uses $N \in \{10, 20, 40\}$ samples from the one-step predictive belief. Curves report mean $\pm$ std over 10 seeds using the same trajectories and measurement realizations.

Baseline comparison under process noise. Figure 4 compares the three viewpoint selection methods as object process noise increases. Our proposed Predictive NBV achieves the highest final completeness across all process-noise settings, with the largest gains at higher process noise. The representative reconstructions explain this trend: Predictive NBV continues to reveal new surface regions in later iterations, while Tracking-only often stays near the predicted target without improving surface coverage. The improvement over Non-predictive NBV shows that evaluating viewpoints under the current belief becomes unreliable when object motion changes the execution-time camera-object configuration. The improvement over Tracking-only shows that prediction must be coupled with coverage-driven scoring to reveal previously unseen surfaces. Overall, propagating the belief to execution time produces viewpoint choices that remain reconstruction-informative under plausible future object motion.

Effect of robot reachability. Figure 5 evaluates how reachability limits affect reconstruction by varying the sensing robot’s maximum per-step travel distance relative to the object’s displacement, which is approximately 0.5 m per planning step. Higher relative speed improves both coverage growth and final completeness because more informative predicted viewpoints become reachable within one planning step. This highlights a practical limit of moving object reconstruction: prediction can identify useful future viewpoints, but reconstruction improves only when the robot can physically reach them before the object state changes again.

Ablation of belief-space evaluation. Figure 6 isolates the effect of viewpoint evaluation under the predictive belief. Random feasible selection performs worst, showing that the candidate set alone is not sufficient. Scoring only at the predicted mean object state improves performance, but Monte Carlo expected scoring improves completeness further by accounting for nonlinear changes in viewpoint utility across plausible execution-time object states. The improvement from $N = 10$ to $N = 40$ indicates that better approximating the predictive expectation leads to more reliable viewpoint choices. Average scoring time remains subsecond, increasing modestly from 427 ms for predicted-mean scoring to 479 ms for MC scoring with $N = 40$.

Ablation of candidate geometry. Table I isolates the effect of uncertainty-aware candidate viewpoint generation.

TABLE I

Ablation of candidate viewpoint generation geometry. Final completeness at iteration 10 for an isotropic ring and the proposed uncertainty-adaptive ellipse. The ring radius is determined from the maximum eigenvalue of the predicted object-position covariance. Results are reported as mean $\pm$ std over 10 seeds under three different scoring methods.

Candidate Shape	Random	Predicted Mean	MC = 40
Ring (isotropic)	61.1% $\pm$ 8.2%	65.9% $\pm$ 9.1%	69.4% $\pm$ 10.7%
Ellipse (adaptive)	64.8% $\pm$ 6.4%	67.1% $\pm$ 9.5%	70.8% $\pm$ 8.7%

The adaptive ellipse improves final completeness across all three scoring methods, indicating that directional uncertainty helps place candidate viewpoints in regions that remain useful under plausible object motion. The gains are smaller than those from Monte Carlo belief-space evaluation, indicating that candidate generation and expected scoring are complementary, with belief-space evaluation contributing the larger gain.

C. Real-world Experiments

The real-world experiments test whether the simulation trends transfer to a physical robot with real depth sensing and injected object-position measurement noise. A mobile robot is equipped with an Intel RealSense D435i depth camera, as shown in Fig. 1. The moving object follows a predefined S-curve trajectory while the robot replans viewpoints for 10 iterations. A motion-capture system provides ground-truth object poses for evaluation. Noisy planar position measurements are generated by adding zero-mean Gaussian noise to the measured object position, with $\sigma \in \{0.10, 0.15, 0.20\}$ m. For each setting, four trials are conducted using the same three viewpoint selection methods as in Sec. V-B. Figure 3 shows one real-world trial. The robot changes its relative viewpoint over successive planning iterations rather than simply following the object. This behavior is important for reconstruction: as the object moves, the planner selects views that expose new object surfaces instead of only maintaining proximity.

Baseline comparison under measurement noise. Figure 7 compares the three viewpoint selection methods as injected object-position measurement noise increases. Predictive NBV achieves the highest final completeness across all noise settings. The trend matches simulation: Predictive NBV continues to improve over later iterations, while Non-predictive NBV and Tracking-only often plateau earlier. This indicates that views selected from the current belief or by proximity to the predicted target can become less informative after the object moves. Because the methods share the same smoothing module, reachability constraints, and reconstruction update, the performance gap reflects the viewpoint-selection strategy: whether prediction is propagated to execution time and whether candidate views are scored for coverage. The results show that motion-uncertainty-aware NBV transfers to the physical robot and improves reconstruction when noisy partial observations induce uncertainty over the object state.

VI. LIMITATIONS

Our framework has the following limitations. First, viewpoint evaluation and voxel integration rely on accurate camera

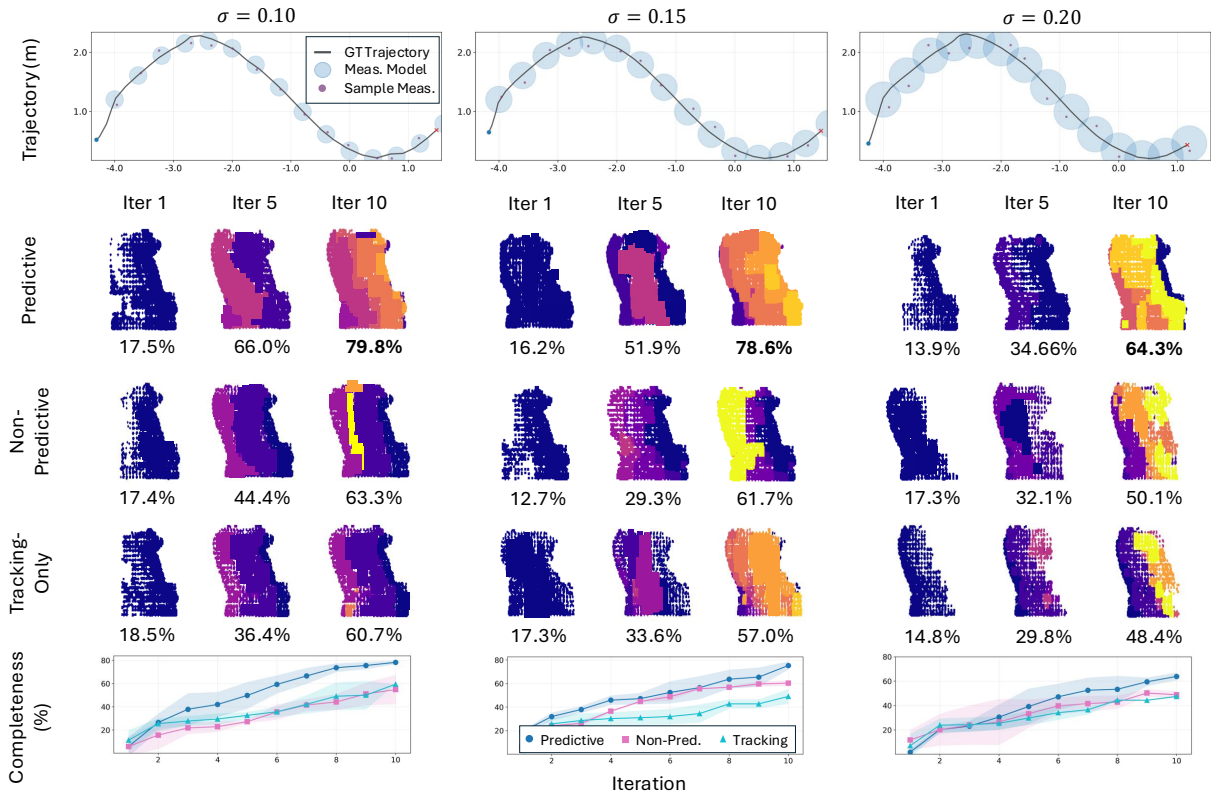


Fig. 7. **Real-world reconstruction under varying measurement noise.** The injected measurement noise increases across columns, $\sigma \in \{0.10, 0.15, 0.20\}$ m. **Top:** measured ground-truth (GT) object trajectory, position-uncertainty regions, and noisy position measurements. **Middle:** representative reconstructions at iterations 1, 5, and 10 for the three viewpoint selection methods; numbers report reconstruction completeness, and colors indicate the surface revealed by each iteration. **Bottom:** completeness over planning iterations, reported as mean $\pm$ std over four trials per setting.

extrinsics and robot-to-object transforms over the planning horizon. Localization drift, synchronization error, or calibration bias can distort the predicted camera-object configuration and degrade viewpoint scoring. Future work could incorporate robot pose uncertainty and online calibration checks into the NBV planning pipeline.

Second, the NBV planner relies on externally provided noisy object-position measurements at each planning time. This separates active reconstruction from the upstream problem of onboard target detection and tracking. Performance may degrade when target observations are intermittent, biased, or delayed. A natural extension is to integrate onboard target tracking with reconstruction, so that the robot jointly estimates the object state and selects viewpoints that support both tracking and surface coverage.

Third, NBV evaluation is performed on an object-centric bounded voxel proxy using a projection-based surrogate score. This introduces approximations from voxel discretization, frontier extraction, ellipsoid fitting, and simplified depth-ordering rather than explicit ray casting. These approximations may become brittle for thin structures, strong self-occlusion, or degraded depth measurements.

VII. CONCLUSIONS

This work presented a motion-uncertainty-aware next-best-view (NBV) planning framework for active reconstruction of a moving object. The framework addresses a key limitation of standard NBV planning for static objects: when the object

moves, the viewpoint selected during planning may no longer be informative at execution time. Instead of evaluating candidate views at a single predicted object pose, the proposed method propagates a predictive belief over the plausible object state and selects viewpoints by maximizing the expected coverage-driven score over that belief.

The resulting online planner integrates fixed-lag Gaussian Process prediction, uncertainty-adaptive candidate generation, reachability filtering, and Monte Carlo belief-space evaluation. Simulation and real-world experiments show that this formulation improves reconstruction completeness over both non-predictive NBV and prediction-only tracking baselines. The baseline comparisons show that motion prediction alone is not sufficient, and coverage-driven NBV without execution-time prediction is also insufficient under object motion uncertainty. Together, these results demonstrate that moving-object reconstruction benefits from explicitly coupling predictive state estimation with coverage-driven viewpoint selection.

VIII. ACKNOWLEDGMENT

This study was funded by Project CETI via grants from Dalio Philanthropies and Ocean X; Sea Grape Foundation; Virgin Unite and Rosamund Zander/Hansjorg Wyss through The Audacious Project: a collaborative funding initiative housed at TED. This work has been made possible in part by a gift from the Chan Zuckerberg Initiative Foundation to establish the Kempner Institute for the Study of Natural and Artificial Intelligence.

REFERENCES

- [1] S. Pan and H. Wei, "A global max-flow-based multi-resolution next-best-view method for reconstruction of 3d unknown objects," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 714–721, 2022.
- [2] D. Yan, J. Liu, F. Quan, H. Chen, and M. Fu, "Active implicit object reconstruction using uncertainty-guided next-best-view optimization," *IEEE Robotics and Automation Letters*, vol. 8, no. 10, pp. 6395–6402, 2023.
- [3] R. Border and J. D. Gammell, "The surface edge explorer (see): A measurement-direct approach to next best view planning," *Int. J. Rob. Res.*, vol. 43, no. 10, p. 1506–1532, Sep. 2024. [Online]. Available: <https://doi.org/10.1177/02783649241230098>
- [4] J. I. Vasquez-Gomez, L. E. Sucar, and R. Murrieta-Cid, "View planning for 3d object reconstruction with a mobile manipulator robot," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2014, pp. 4227–4233.
- [5] R. Monica and J. Aleotti, "Surfel-based next best view planning," *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3324–3331, 2018.
- [6] Z. Jia, Y. Li, Q. Hao, and S. Zhang, "Pb-nbv: Efficient projection-based next-best-view planning framework for reconstruction of unknown objects," *IEEE Robotics and Automation Letters*, vol. 10, no. 7, pp. 7444–7451, 2025.
- [7] B. Koger, A. Deshpande, J. T. Kerby, J. M. Graving, B. R. Costelloe, and I. Couzin, "Quantifying the movement, behaviour and environmental context of group-living animals using drones and computer vision." *The Journal of animal ecology*, 2023.
- [8] S. Gero, M. Milligan, C. Rinaldi, P. Francis, J. Gordon, C. Carlson, A. Steffen, P. Tyack, P. Evans, and H. Whitehead, "Behavior and social structure of the sperm whales of dominica, west indies," *Marine Mammal Science*, vol. 30, no. 3, pp. 905–922, 2014.
- [9] N. O. Hirtle, J. E. Stepanuk, E. I. Heywood, F. Christiansen, and L. H. Thorne, "Integrating 3d models with morphometric measurements to improve volumetric estimates in marine mammals," *Methods in Ecology and Evolution*, vol. 13, no. 11, pp. 2478–2490, 2022.
- [10] F. Christiansen, M. Sironi, M. J. Moore, M. Di Martino, M. Ricciardi, H. A. Warick, D. J. Irschick, R. Gutierrez, and M. M. Uhart, "Estimating body mass of free-living whales using aerial photogrammetry and 3d volumetrics," *Methods in Ecology and Evolution*, vol. 10, no. 12, pp. 2034–2044, 2019.
- [11] K. C. Bierlich, S. Karki, C. N. Bird, A. Fern, and L. G. Torres, "Automated body length and body condition measurements of whales from drone videos for rapid assessment of population health," *Marine Mammal Science*, vol. 40, no. 4, p. e13137, 2024.
- [12] C. Napoli, N. Hirtle, J. Stepanuk, F. Christiansen, E. I. Heywood, T. J. Grove, A. Stoller, F. Dodds, M. Glarou, M. H. Rasmussen *et al.*, "Drone-based photogrammetry reveals differences in humpback whale body condition and mass across north atlantic foraging grounds," *Frontiers in Marine Science*, vol. 11, p. 1336455, 2024.
- [13] S. Bhattacharya, N. Jadhav, H. Izhar, K. Li, K. George, R. Wood, and S. Gil, "Real-time remote tracking and autonomous planning for whale rendezvous using robots," *arXiv preprint arXiv:2512.05808*, 2025.
- [14] S. Kiciroglu, H. Rhodin, S. N. Sinha, M. Salzmann, and P. Fua, "Activemocap: Optimized viewpoint selection for active human motion capture," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 100–109.
- [15] Q. Wang, Y. Gao, J. Ji, C. Xu, and F. Gao, "Visibility-aware trajectory optimization with application to aerial tracking," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 5249–5256.
- [16] B. Penin, P. Giordano, and F. Chaumette, "Vision-based reactive planning for aggressive target tracking while avoiding collisions and occlusions," *IEEE Robotics and Automation Letters*, vol. 3, pp. 3725–3732, 2018.
- [17] R. Tallamraju, E. Price, R. Ludwig, K. Karlapalem, H. Bülthoff, M. J. Black, and A. Ahmad, "Active perception based formation control for multiple aerial vehicles," *IEEE Robotics and Automation Letters*, vol. 4, pp. 4491–4498, 2019.
- [18] J. Delmerico, S. Isler, R. Sabzevari, and D. Scaramuzza, "A comparison of volumetric information gain metrics for active 3d object reconstruction," *Autonomous Robots*, vol. 42, no. 2, pp. 197–208, 2018.
- [19] A. Gu'edon, P. Monasse, and V. Lepetit, "Scone: Surface coverage optimization in unknown environments by volumetric integration," *ArXiv*, vol. abs/2208.10449, 2022.
- [20] T. D. Barfoot, C. H. Tong, and S. Särkkä, "Batch continuous-time trajectory estimation as exactly sparse gaussian process regression." in *Robotics: Science and Systems*, vol. 10. Citeseer, 2014, pp. 1–10.
- [21] T. D. Barfoot, P. Furgale, and C. H. Tong, "Gaussian Process Gauss-Newton: Non-Parametric State Estimation," in *2013 International Conference on Computer and Robot Vision*. IEEE Computer Society, 2012, pp. 206–213.
- [22] C. H. Tong, P. Furgale, and T. Barfoot, "Gaussian process gauss-newton for non-parametric simultaneous localization and mapping," *The International Journal of Robotics Research*, vol. 32, pp. 507–525, 05 2013.
- [23] J. Santos, M. Oliveira, R. Arrais, and G. Veiga, "Autonomous scene exploration for robotics: A conditional random view-sampling and evaluation using a voxel-sorting mechanism for efficient ray casting," *Sensors (Basel, Switzerland)*, vol. 20, 2020.
- [24] A. Batinovic, T. Petrovic, A. Ivanovic, F. Petric, and S. Bogdan, "A multi-resolution frontier-based planner for autonomous 3d exploration," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 4528–4535, 2021.
- [25] F. Pratissoli, M. Mantovani, A. Prorok, and L. Sabatini,

- “Distributed coverage control for time-varying spatial processes,” *IEEE Transactions on Robotics*, vol. 41, pp. 1602–1617, 2025.
- [26] S. Sahl, E. Song, and D. Niu, “Robust cubature kalman filter for moving-target tracking with missing measurements,” *Sensors (Basel, Switzerland)*, vol. 24, 2024.
- [27] M. Mantovani, F. Pratissoli, and L. Sabattini, “Distributed coverage control for spatial processes estimation with noisy observations,” *IEEE Robotics and Automation Letters*, vol. 9, no. 5, pp. 4431–4438, 2024.
- [28] U. Lagerblad, H. Wentzel, and A. Kulachenko, “Study of a fixed-lag kalman smoother for input and state estimation in vibrating structures,” *Inverse Problems in Science and Engineering*, vol. 29, pp. 1260 – 1281, 2020.
- [29] A. Goudar, W. Zhao, and A. P. Schoellig, “Range-visual-inertial sensor fusion for micro aerial vehicle localization and navigation,” *IEEE Robotics and Automation Letters*, vol. 9, pp. 683–690, 2023.
- [30] X. Yan, V. Indelman, and B. Boots, “Incremental sparse gp regression for continuous-time trajectory estimation and mapping,” *Robotics and Autonomous Systems*, vol. 87, pp. 120–132, 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0921889016300434>
- [31] T. Li, H. Chen, S. Sun, and J. Corchado, “Joint smoothing and tracking based on continuous-time target trajectory function fitting,” *IEEE Transactions on Automation Science and Engineering*, vol. 16, pp. 1476–1483, 2019.
- [32] N. Saini, E. Bonetto, E. Price, A. Ahmad, and M. J. Black, “Airpose: Multi-view fusion network for aerial 3d human pose and shape estimation,” *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4805–4812, 2022.
- [33] C. Ho, A. Jong, H. Freeman, R. Rao, R. Bonatti, and S. Scherer, “3d human reconstruction in the wild with collaborative aerial cameras,” in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021, pp. 5263–5269.
- [34] H. Gao, P. Wu, Y. Su, K. Zhou, J. Ma, H. Liu, and C. Liu, “Probabilistic visibility-aware trajectory planning for target tracking in cluttered environments,” *2024 American Control Conference (ACC)*, pp. 594–600, 2023.
- [35] Z. Qinghui, L. Hao, L. Lei, L. Shenglin, and P. Wei, “3d reconstruction of the moving object with double-shooting based on phase shifting profilometry,” vol. 52, no. 12, pp. 20 220 891–1, 2023.
- [36] G. Liu, W. Jiang, B. Lei, V. Pandey, K. Daniilidis, and N. Motee, “Beyond uncertainty: Risk-aware active view acquisition for safe robot navigation and 3d scene understanding with fisherrf,” 2025. [Online]. Available: <https://arxiv.org/abs/2403.11396>
- [37] M. Lauri, J. Pajarinen, J. Peters, and S. Frintrop, “Multi-sensor next-best-view planning as matroid-constrained submodular maximization,” *IEEE Robotics and Automation Letters*, vol. 5, pp. 5323–5330, 2020.
- [38] B. Hepp, M. Nießner, and O. Hilliges, “Plan3d,” *ACM Transactions on Graphics (TOG)*, vol. 38, pp. 1 – 17, 2017.
- [39] D. Duberg and P. Jensfelt, “Ufomap: An efficient probabilistic 3d mapping framework that embraces the unknown,” *IEEE Robotics and Automation Letters*, vol. 5, pp. 6411–6418, 2020.
- [40] A. Dai, S. Papatheodorou, N. Funk, D. Tzoumanikas, and S. Leutenegger, “Fast frontier-based information-driven autonomous exploration with an mav,” *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 9570–9576, 2020.
- [41] P. J. Besl and N. D. McKay, “Method for registration of 3-d shapes,” in *Sensor fusion IV: control paradigms and data structures*, vol. 1611. Spie, 1992, pp. 586–606.
- [42] I. Petrovska and B. Jutzi, “Vision through obstacles - 3d geometric reconstruction and evaluation of neural radiance fields (nerfs),” *Remote. Sens.*, vol. 16, p. 1188, 2024.