Next-Best-View-based Task and Motion Planning for Autonomous Photography & Inspection

Shijie Gao[1], Lauren Bramblett[2] and Nicola Bezzo[1,2]

Abstract— **Autonomous mobile robots (AMRs) equipped with high-quality cameras have revolutionized the field of inspections by providing efficient and cost-effective means of conducting surveys. While autonomous inspection is gaining traction in various applications, acquiring the best inspection information autonomously using effective task and motion planning (TAMP) remains a complex challenge. In situations where objects may block a robot's view, it is necessary to use reasoning to adjust the task and seek the optimal points for collecting data. Many researchers assume that inspection information can be continuously collected and stored either locally or via a cloud-based solution; however, these approaches are limited to storage capacity and network restrictions and require labor-intensive post-processing. To address this challenge, we present an autonomous Next-Best-View (NBV) framework that maximizes the inspection information while reducing the number of pictures needed during operations. The framework consists of a formalized evaluation metric using ray-tracing and Gaussian process interpolation to estimate information reward based on the current understanding of the partially-known environment. Particle swarm optimization (PSO) is used to sample candidate views in the environment and identify the NBV point. The effectiveness of the proposed framework is validated in simulations and experiments.**

Note—**Simulations and experiments videos can be accessed at `https://www.bezzorobotics.com/sg-lb-tamp23`.**

I. Introduction

Autonomous mobile robots (AMRs) have been particularly useful for inspection purposes because of their agility and mobility. However, most existing autonomous inspection frameworks are not fully automated and collect all available information during operations, relegating data processing and decision-making to humans. With this in mind, we note two gaps in autonomous inspection and photography: i) existing approaches overlook the complexity associated with continuously and repetitively acquiring information due to the absence of an explicit definition of a "good" picture, and ii) the ability for an autonomous vehicle to perform prolonged tasks while minimizing post-processing efforts, decreasing data storage, and ensuring safety will make autonomous inspection more useful and accessible to real-world operations.

In this work, our goal is to bridge these gaps. We assume that the robot knows the target's $\mathbf{T}$ locations and dimensions while the obstacles, $\mathcal{O}$, can either be known or unknown in the environment – represented by an occupancy map $\mathcal{M}$. To find the best point for inspection, the vehicle should update the best view position as obstacles are detected that compromise inspection photo quality. The goal is to finish the task by taking as few pictures as possible while preserving as much information of $\mathbf{T}$ as possible. The problems are formally defined as follows:

Problem 1. ***Inspection Task Planning:*** *Given the target's inspection locations and dimensions, derive an optimal task plan that minimizes energy consumption for the robot.*

Problem 2. ***Metrics of the Best View for Target Inspection:*** *Given the location and dimensions of the target $\mathbf{T}$ and a set of known obstacles represented by an occupancy map $\mathcal{M}$, find an evaluation metric $G(\mathbf{T}, \mathcal{M}, \mathbf{P})$ that scores the sampled viewpoints $\mathbf{P}$ for the vehicle to visit at runtime.*

Problem 3. ***Max-Info Motion Planning:*** *Given a partially-known environment and metrics for picture evaluation, dynamically adjust the task and plan a path that minimizes the time to the inspecting task while ensuring the robot's safety. The TAMP framework should accommodate changes in the environment, as new obstacles may appear.*

II. Methodology

We propose a Next-Best-View-based (NBV) TAMP framework[1] to provide a comprehensive strategy to dynamically plan the inspection task to achieve the fewest number of viewpoints that covers the entire inspected target. The vehicle is initialized at a starting point in a partially-known environment with the target information known. The high-level task plan is derived by solving a traveling salesman problem (TSP). During the mission, the vehicle updates its map using its equipped depth sensor. After the update, candidate views are generated and evaluated via our photo quality scoring metrics. Particles migrate toward locations that offer the best viewpoints using PSO. Particles are reassessed at the new viewpoints until the termination criterion is met. The best viewpoint is then passed to the vehicle's planner and controller to drive it toward the location. As the vehicle moves, the map is updated and the viewpoint is reevaluated based on any new information. Once the vehicle reaches the best viewpoint, a photo is captured and the target interest is updated based on the estimated information gained from the captured photo using Gaussian process interpolation.

A. Task Planning

To address *Problem* 1, we solve a TSP within a full connected graph. The graph is constructed in the way that task locations are served as nodes. The graph's edges are directional, and the associated costs are calculated by estimating energy consumption during the traversal between nodes. The estimates are given by the energy consumption model designed in our previous work [2]. Fig. 1 shows an example of task planning result for an inspection mission.

Shijie Gao, Lauren Bramblett, and Nicola Bezzo are with the Departments of Electrical & Computer Engineering[1] and Systems & Information Engineering[2], University of Virginia, Charlottesville, VA 22904, USA. Email: {sg9dn, qbr5kx, nb6be}@virginia.edu

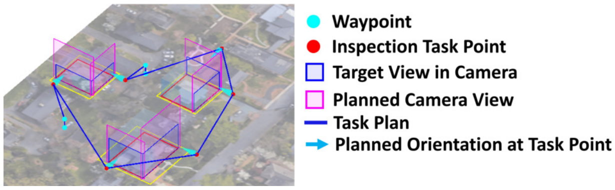

Fig. 1. Given the inspection tasks, a high-level task plan is derived by solving an energy-efficient TSP problem.

B. Motion and Path Planning

Given the task plan, motion and path planning is achieved via an NBV approach. The main goal of NBV [3] is to select the most informative viewpoint, addressing *Problem* 2. We propose an inspection photo evaluation metric based on the perspective distortion, the scale of a target within the viewing frame, and the estimated information gained from a viewpoint. Consider a target in continuous space $\mathbf{T} \in \mathbb{R}^{m \times n_d}$ where n_d is the dimension of the target defined by m coordinates. Viewpoints are evaluated as follows:

$$G(\mathbf{T}, \mathcal{M}, \mathbf{P}) = \gamma_d \cdot \gamma_s \cdot \int_{\mathbf{T}} g(\tau) d\tau, \qquad \tau \in \mathbf{T} \tag{1}$$

$$\gamma_d = \prod_{i=1}^{n_d - 1} U_d \left(\frac{|\ell_{i2} - \ell_{i1}|}{\ell_{i\mathbf{T}}} \right) \tag{2}$$

$$\gamma_s = \prod_{i=1}^{n_d - 1} U_s \left(\frac{(\phi_{i\max} - \phi_{i\min})}{\Phi_i}, \beta \right) \tag{3}$$

where $g(\tau)$ represents the importance of the target portion. $g(\tau) = 1$ denotes that the portion has not been inspected, while $g(\tau) = 0$ indicates that the region is obscured or has already been observed. γ_d and γ_s are discount factors that reflect the amount of distortion and the percentage of the target that occupies the frame. $U_d(\cdot)$ and $U_s(\cdot)$ represent the utility functions associated with the perspective distortion and the scale of the target, penalizing unbalanced target captures and undesired scale, respectively. β is a user-defined parameter that specifies the desired percentage of the target within the frame. A visual illustration of the metric using a two-dimensional example is presented in Fig. 2.

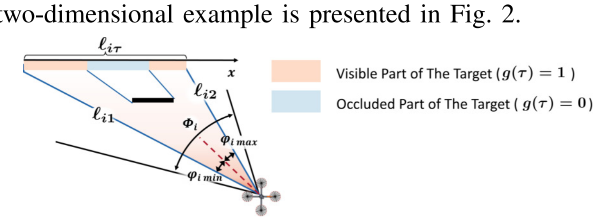

Fig. 2. Pictorial depiction of the proposed evaluation metric.

We note that evaluating numerous potential viewpoints can be expensive at runtime. To mitigate this issue, we use ray-tracing [4] to examine the voxels in $\mathcal{M}$ to validate the visibility of the target. However, using an unlimited number of rays is impractical in real world. Thus, we use Gaussian process interpolation [5] to estimate the likelihood of the target visibility across non-sampled target portions in between ray termination coordinates.

To solve *Problem* 3, the viewpoint evaluation is used in conjunction with a particle swarm optimization [6] (PSO) which samples the environment to find the best inspection point. PSO is employed not only because of the highly nonlinear nature of the problem, which includes expansive flat regions posing challenges to gradient descent methods, but also due to its adaptability to runtime considerations. In our inspection application, a candidate viewpoint is represented as a point in an N-dimensional solution space. A swarm of particles is used to represent a number of potential viewpoints. The information gain G is calculated for each viewpoint and the particles migrate toward the particle with the highest evaluation score. The viewpoint associated with the highest-rated particle is the desired inspection point. The robot can then take advantage of algorithms such as A* and RRT* to find a collision-free path.

III. Results

Fig. 3 shows results of our approach in simulation in which a quadrotor detects a T-shape obstacle and autonomously decides to focus on the right part of the target first and subsequently moves to the left side to capture the rest of the target. Fig. 4 shows experiment results for an inspection task with an a priori unknown slash-shaped obstacle. The robot locates the best vantage point at which the obstacle appears as a thin line.

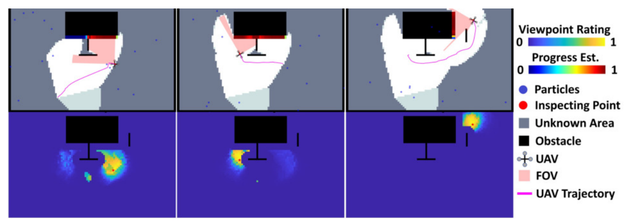

Fig. 3. Simulation result of a quadrotor inspecting the front and side of a building. Heatmaps in the second row show the ground truth information gain at all viewpoints with the proposed metric.

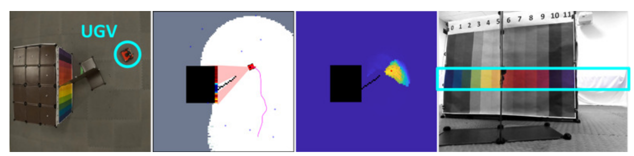

Fig. 4. An example of inspecting an object with a slash-shaped obstacle. The robot preserves almost the entire target with minimal distortion.

IV. Conclusion

In this work, we have presented a novel TAMP framework for an autonomous vehicle to take a quality image of a target in a partially-known environment. The extensive results show the validity, applicability, and generality of the method.

References

[1] S. Gao, L. Bramblett, and N. Bezzo, "Take your best shot: Sampling-based next-best-view planning for autonomous photography & inspection," *IEEE Robotics and Automation Letters*, 2023, under review.

[2] S. Gao, C. Di Franco, D. Carter, D. Quinn, and N. Bezzo, "Exploiting ground and ceiling effects on autonomous uav motion planning," in *2019 International Conference on Unmanned Aircraft Systems (ICUAS)*. IEEE, 2019, pp. 768–777.

[3] C. Connolly, "The determination of next best views," in *Proceedings. 1985 IEEE international conference on robotics and automation*, vol. 2. IEEE, 1985, pp. 432–435.

[4] J. I. Vasquez-Gomez, L. E. Sucar, and R. Murrieta-Cid, "Hierarchical ray tracing for fast volumetric next-best-view planning," in *2013 International conference on computer and robot vision*. IEEE, 2013, pp. 181–187.

[5] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*. MIT press Cambridge, MA, 2006, vol. 2, no. 3.

[6] L. Kang, R.-S. Chen, N. Xiong, Y.-C. Chen, Y.-X. Hu, and C.-M. Chen, "Selecting hyper-parameters of gaussian process regression based on non-inertial particle swarm optimization in internet of things," *IEEE Access*, vol. 7, pp. 59 504–59 513, 2019.