IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

Efficient 3D Reconstruction in Noisy Agricultural Environments: A Bayesian Optimization Perspective for View Planning

Athanasios Bacharis[1], Konstantinos D. Polyzos[2], Henry J. Nelson[1], Georgios B. Giannakis[3], and Nikolaos Papanikolopoulos[1]

***Abstract*—3D reconstruction is a fundamental task in robotics that gained attention due to its major impact in a wide variety of practical settings, including agriculture, underwater, and urban environments. While this task can be carried out using a large number of arbitrarily taken 2D images, their processing may become laborious, time-consuming, and in some instances may not provide the necessary information about the object of interest. An efficient alternative is the so-termed view planning (VP), which aims to optimally place a certain number of cameras in positions that maximize the visual information. Nonetheless, in most real-world settings, existing environmental noise can significantly affect the performance of 3D reconstruction. To that end, this work advocates a novel geometric-based reconstruction quality function for VP, that accounts for the existing noise of the environment, without requiring its closed-form expression. With no analytic expression of the objective function, this work puts forth an adaptive Bayesian optimization algorithm for accurate 3D reconstruction in the presence of noise. Numerical tests on simulated and real noisy agricultural environments showcase the merits of the proposed VP approach for efficient 3D reconstruction with even a small number of available cameras.**

***Index Terms*—Agricultural Automation, Robotics and Automation in Agriculture and Forestry, Computer Vision for Automation**

I. Introduction

ACQUIRING visual information is a key component in robotics for scene understanding, planning, and decision-making. In particular, the acquisition of 3D information, known as 3D scene reconstruction, has gained popularity in different robotic settings, including agricultural [1]–[4], underwater [5], and urban [6], [7] environments. In the agricultural domain, informative 3D reconstructions have a major environmental and financial impact, including limited water usage for sustainable farming practices and increased revenue from large crop cultivation, just to list a few [8].

Obtaining an informative 3D scene reconstruction can be carried out using a large number of arbitrarily taken (2D) images. Yet this process may be inefficient since processing a large number of images to obtain a 3D reconstruction might be time-consuming. This discourages time-critical or resource-constrained applications such as surveillance and monitoring in agricultural fields where there is a need for real-time coverage of multiple areas of interest. Moreover, having multiple random images may lead to erroneous 3D reconstruction. To that end, an alternative is to model 3D reconstruction as an optimization problem, termed view planning (VP) [9], [10], that aims to decide the optimal one-shot placement of the available cameras in the 3D space. This will provide the optimal location and orientation of each available camera to get an image of the environment. In that way, this one-shot VP setting can obtain the necessary information that will result in an efficient and effective reconstruction of an area of interest. VP can be carried out offline, given a fixed number of cameras and a static model of the environment. The use of a VP approach can significantly improve the quality of the 3D reconstruction (typically expressed as point clouds) compared to a set of arbitrarily placed cameras, which usually results in low-density point clouds.

Nonetheless, obtaining the optimal camera placement can become a challenging optimization problem when the size of the environment becomes large; see e.g., urban environments. To cope with this challenge, existing methods have formulated VP as a discrete optimization problem to reduce the space of feasible solutions [1], [11]. A class of methods employed to solve the VP problem is based on search algorithms; e.g., [1], [6]. Recently, neural networks have been utilized in the VP setting to get an estimate of the function that maps any camera placement to its corresponding reconstruction quality [12].

Albeit interesting, all these approaches do not account for any type of environmental noise, which can prove to be important when modeling the environment and can significantly affect the VP performance in practical settings. Furthermore, existing methods typically entail a sufficiently large number of cameras which implies an increased cost of collecting images that is worth considering in challenging environments such as agricultural ones. In Fig. 1, there is an example of how the noise affects the geometry of the environment in VP. However, obtaining an analytic expression of the noise may be a challenging task, particularly when the number of noise realizations is not sufficient. With no analytic expression of the noise, the expression of the reward function that incorpo-

Manuscript received: April, 30, 2025; Revised August, 8, 2025; Accepted September, 28, 2025.

This paper was recommended for publication by Editor Hyungpil Moon upon evaluation of the Associate Editor and Reviewers' comments. This work was supported by the Minnesota Robotics Institute (MnRI) and NSF grants 2220292, 2212318, 2312547, 2126052, 2128593, 2332173, 1439728, 1531330, 1939033. USDA/NIFA has also supported this work through the grants 2020-67021-30755 and 2023-67021-39829. The work of Athanasios Bacharis was also supported by the University of Minnesota Doctoral Dissertation Fellowship. The work of Konstantinos D. Polyzos was carried out while he was at the University of Minnesota. His work was also supported by the Onassis Foundation Scholarship.

[1] Department of Computer Science and Engineering, University of Minnesota Twin Cities. Emails: {bacha035, nels8279, papan001}@umn.edu.

[2] Department of Electrical and Computer Engineering, University of California San Diego. Email: kpolyzos@ucsd.edu.

[3] Department of Electrical and Computer Engineering, University of Minnesota Twin Cities. Email: georgios@umn.edu.

Digital Object Identifier (DOI): see top of this page.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

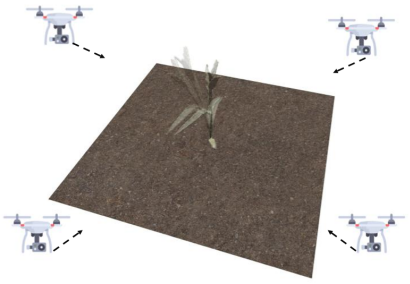


Fig. 1: Example of a noisy environment with four drones aiming to carry out the 3D reconstruction task.

rates the noise also becomes *unknown*, and thus conventional optimization techniques including gradient-based solvers, may not be applicable.

This motivates the use of the Bayesian optimization (BO) paradigm that aims at optimizing a black-box (unknown) function by leveraging a (typically small) number of function evaluations to form a probabilistic *surrogate model* that will guide the selection of new query points [13]. Relying on the surrogate model, the so-termed *acquisition function* (AF), typically expressed in closed-form, is used to sequentially select the next query points in the optimization process. BO has been adopted in a range of practical settings including hyperparameter tuning [14], drug discovery [15] and robotic-based tasks [16]–[18] to name a few. Nonetheless, none of the existing works has considered adopting BO in the VP problem under *noisy* environments.

This paper presents a novel approach for modeling and optimizing the reward function of the VP problem in order to effectively and efficiently reconstruct an area of interest in the presence of noise. Relying on an initial point cloud using collected data from the noisy environment, an innovative adaptive Bayesian optimization approach is proposed to judiciously select only a few yet informative viewpoints under new noise conditions, thereby enhancing efficiency and reducing monitoring costs in agricultural settings. The contributions of this work can be summarized in the following directions:

- To the best of our knowledge, this work is the first to incorporate existing noise in the reward function of a VP problem for 3D reconstruction.
- To optimize the reward function with *unknown* closed-form expression, the present work is the first to introduce a well-motivated adaptive BO method for the VP problem.
- Compared to most existing approaches that require discretization of the search space in order to be tractable, this work focuses on a *continuous* optimization problem that considers the entire search space.
- Promising experimental results on three different corn-based agricultural environments demonstrate the effectiveness of the proposed approach for accurate 3D reconstruction with a *limited* number of available cameras.

II. Related Work

A. View Planning

VP aims at placing cameras in an optimal way that maximizes the gain of information about the environment, to successfully carry out the 3D reconstruction task [9], [10]. In order to reduce the complexity of the search space, which may be intractable in large environments, several works have modeled the problem in a discrete manner [1], [9], [10], [12]. However, this discrete formulation is known to be NP-complete, as shown in [10], leading to the use of sampling-based search methods [1] or approximate algorithms that require two or more optimization phases [11].

Besides an affordable complexity, a critical component in the VP optimization problem is the proper selection of the reward function. Given the information about the environment, two different types of reward functions have been mainly explored, namely occupancy maps and geometric evaluations. Although offering a simple representation of the environment, occupancy maps may reduce the quality of the 3D reconstruction [12], [19], [20]. On the other hand, geometric functions typically achieve more dense point clouds in the resulting reconstruction by imposing more constraints in the view selection process [1], [3], [4]. In our work, we will focus on a geometric function since we aim to improve the reliability of the 3D reconstruction.

Finally, an aspect of great interest in real-world settings is the existing noise of the environment. Although the information of the *static* environment is given, in several applications there are many parameters, such as the wind, that may alter the representation of the environment, hence affecting the outcome of the 3D reconstruction. Therefore, it is of paramount importance to incorporate these types of disturbances in the selection of the views to achieve the best reconstruction.

Despite the effectiveness of the aforementioned VP approaches in noise-free environments, none of them has considered the effects of the noise in the VP process, or the merits of a continuous formulation over a discrete one where any position in the 3D space can be explored.

B. Bayesian Optimization

Bayesian optimization (BO) provides a principled framework to optimize a *black-box* or *costly* to evaluate function, by judiciously exploiting a number of function values to build a surrogate model, that allows for the sequential acquisition of new points. The key components of BO are (i) the selection of a probabilistic surrogate model and (ii) the design of the AF to select new points to query on-the-fly. Gaussian processes (GPs) have been extensively adopted as a *nonparametric* Bayesian surrogate model in a range of BO applications due to their well-documented merits in learning an unknown function along with its probability density function (pdf) in a sample-efficient manner [13], [21]. Although interesting and effective in several practical settings, their performance hinges on the proper selection of the kernel function to evaluate the pairwise similarity of different input points, whose selection is a non-trivial task. To cope with this challenge, existing works adaptively learn the kernel as new data arrive on-the-fly; see

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

e.g., [22]–[26]. Capitalizing on GP-based surrogate models, several AFs have been designed including Thompson sampling [27], expected improvement (EI) [28] or upper confidence bound [29] to name a few. These AFs are expressed in closed-form, they are easy to optimize and have well-documented merits in balancing *exploration* and *exploitation* of the search space.

III. Problem Formulation

The objective of the VP problem is to find an optimal camera setup that will maximize the information that can be extracted from a given environment to assist the 3D reconstruction task. Specifically, given a number of cameras N, and a point cloud $\{\mathbf{p}_i\}_{i=1}^{\mathcal{P}} \subseteq \mathbb{R}^3$, sampled from the given representation of the environment, the goal is to find

$$\mathbf{z}^* = \underset{\mathbf{z}\in\mathcal{Z}}{\arg\max}\ r(\mathbf{z};\mathbf{p}_1,\ldots,\mathbf{p}_{\mathcal{P}}) \tag{1}$$

where $\mathbf{z} := [\mathbf{x}_1,\ldots,\mathbf{x}_N]$, with $\{\mathbf{x}_i\}_{i=1}^{N} \subseteq \mathbb{R}^6$ denoting each camera feature vector and $\mathcal{Z}$ is the set of feasible solutions. In particular, $\mathbf{x}_i = [\mathbf{c}_i, \mathbf{v}_i]$ with $\mathbf{c}_i \in \mathbb{R}^3$ and $\mathbf{v}_i \in \mathbb{R}^3$ containing the information of the location, and the orientation of the i^{th} camera. The reward function $r(\cdot)$ represents the quality of the 3D reconstruction. Note that each camera $i \in \{1,\ldots,N\}$ at location $\mathbf{c}_i$ with orientation $\mathbf{v}_i$ renders a *single* image, and thus VP relies on a total of N images to provide *time-efficient* solutions for 3D reconstruction, that are particularly appealing in time-critical applications.

Following the discrete geometrical formulation of [1] for the $r(\cdot)$ function, an extended version of the $r(\cdot)$ in the continuous space is expressed as

$$r(\mathbf{z};\mathbf{p}_1,\ldots,\mathbf{p}_{\mathcal{P}}) = \frac{1}{\mathcal{PS}} \sum_{\substack{\mathbf{p}\in\mathcal{P} \\ (\mathbf{c}_i,\mathbf{v}_i,\mathbf{c}_j,\mathbf{v}_j)\in\mathbf{z}}} f(\mathbf{c}_i,\mathbf{c}_j;\mathbf{p})\, g(\mathbf{c}_i,\mathbf{v}_i,\mathbf{c}_j,\mathbf{v}_j;\mathbf{p}) \tag{2}$$

where f indicates the reconstruction quality per point $\mathbf{p}$ from the camera location pair $(\mathbf{c}_i,\mathbf{c}_j)$, g represents the geometric constraints for each $(\mathbf{c}_i,\mathbf{v}_i),(\mathbf{c}_j,\mathbf{v}_j)$ camera pair, and $\mathcal{S}$ is the total number of (i,j) pairs.

The geometric reconstruction quality, defined by f, relies on the angle between the vectors $(\mathbf{c}_i-\mathbf{p}),(\mathbf{c}_j-\mathbf{p})\ \forall(i,j)$. When the angle value increases, the reconstruction quality error of $\mathbf{p}$ becomes smaller. This can be expressed by the sine of the angle between these two vectors as

$$f(\mathbf{c}_i,\mathbf{c}_j;\mathbf{p}) = \frac{\big\|(\mathbf{c}_i-\mathbf{p})\times(\mathbf{c}_j-\mathbf{p})\big\|}{\|\mathbf{c}_i-\mathbf{p}\|\big\|\mathbf{c}_j-\mathbf{p}\big\|} \tag{3}$$

In order for (3) to provide an accurate estimate of the reconstruction quality for each point $\mathbf{p}$, two conditions need to be satisfied. The first focuses on the number of those points that are able to be seen from each camera $(\mathbf{c}_i,\mathbf{v}_i)$, and the second on the ability of all the camera pairs $(\mathbf{c}_i,\mathbf{v}_i),(\mathbf{c}_j,\mathbf{v}_j)$ to gain 3D information of each common viewed point $\mathbf{p}$. Specifically, the function g has the form

$$g(\mathbf{c}_i,\mathbf{v}_i,\mathbf{c}_j,\mathbf{v}_j;\mathbf{p}) = \mathbb{1}(cond_1^i)\mathbb{1}(cond_1^j)\mathbb{1}(cond_2^{i,j}) \tag{4}$$

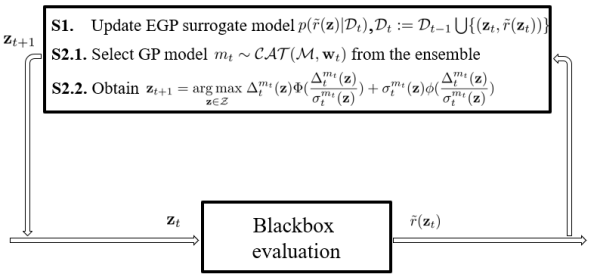


Fig. 2: EGP-VP in a nutshell.

where $\mathbb{1}(\cdot)$ is the indicator function, returning value 1 when the input condition is satisfied and 0 otherwise. The conditions $(cond_1^i, cond_2^{i,j})$ regarding each camera pair (i,j) are defined as

$$cond_1^i := \cos\big(\phi_1^i\big) - \cos\left(\frac{FoV}{2}\right) \geq 0 \tag{5}$$

$$cond_2^{i,j} := \cos\left(\phi_2^{i,j}\right) - \cos(\theta_{match}) \geq 0 \tag{6}$$

where $\cos\big(\phi_1^i\big) := \frac{(\mathbf{c_i}-\mathbf{p})\cdot\mathbf{v_i}}{\|\mathbf{c_i}-\mathbf{p}\|\|\mathbf{v_i}\|}$, $\cos\left(\phi_2^{i,j}\right) := \frac{(\mathbf{c}_i-\mathbf{p})\cdot(\mathbf{c}_j-\mathbf{p})}{\|\mathbf{c}_i-\mathbf{p}\|\|\mathbf{c}_j-\mathbf{p}\|}$, FoV is the camera field-of-view parameter and θ_{match} the maximum angle between the vectors $(\mathbf{c}_i-\mathbf{p})$, $(\mathbf{c}_j-\mathbf{p})$, so that to reconstruct the point $\mathbf{p}$.

The reward function in (1), although formulating the VP problem, does not account for real-world scenarios where there is noise in the environment. To alleviate this limitation, we assume that each point $\mathbf{p}$ under the noisy environment is modeled as $\tilde{\mathbf{p}} = h(\mathbf{p})$, where the function $h(\cdot)$ represents the effect of the noise, whose expression is considered unknown. For instance, the function $h(\cdot)$ may have a linear form of $h(\mathbf{p}) = \mathbf{p}+\mathbf{n}$, with $\mathbf{n}$ following a distribution $\mathscr{D}$ parameterized by $\boldsymbol{\omega}$; that is $\mathbf{n} \sim \mathscr{D}(\boldsymbol{\omega})$. Taking into account the function $h(\cdot)$, the equivalent noisy reward function becomes

$$\tilde{r}(\mathbf{z}) = r(\mathbf{z};\tilde{\mathbf{p}}_1,\ldots,\tilde{\mathbf{p}}_{\mathcal{P}}). \tag{7}$$

Finally, the VP optimization problem turns into

$$\mathbf{z}^* = \underset{\mathbf{z}\in\mathcal{Z}}{\arg\max}\, \tilde{r}(\mathbf{z}). \tag{8}$$

It is worth noticing that the analytic expression of the objective function in (8) becomes unknown when apriori information about the function $h(\cdot)$ is not provided. With no analytic expression at hand, conventional gradient-based solvers are not applicable for obtaining $\mathbf{z}^*$. This motivates well the BO paradigm that offers principled methods to effectively optimize black-box functions in a data-efficient manner, as outlined next.

IV. Optimization Method

Relying on a given budget of input-output data pairs $\mathcal{D}_t := \{(\mathbf{z}_\tau,\tilde{r}_\tau)\}_{\tau=1}^{t}$ with $\tilde{r}_\tau := \tilde{r}(\mathbf{z}_\tau)$, BO capitalizes on a probabilistic surrogate model $p(\tilde{r}(\mathbf{z})|\mathcal{D}_t)$ for $\tilde{r}$, that guides the acquisition of the next query input $\mathbf{z}_{t+1}$. Specifically, each iteration of the BO process follows a two-step approach: (i) update

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

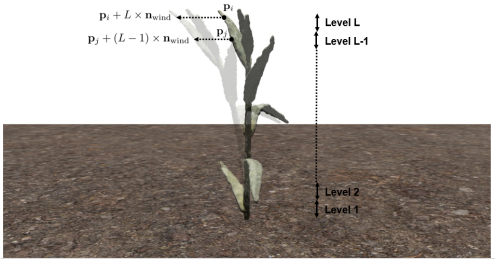


Fig. 3: Representation of the modeled wind noise in the 1-plant scenario. We consider L levels of the plant and for two randomly selected points $\mathbf{p}_i$ and $\mathbf{p}_j$ belonging to levels L and $L-1$ respectively, the corresponding noisy points are $\tilde{\mathbf{p}}_i = \mathbf{p}_i + L \times \mathbf{n}_{\text{wind}}$ and $\tilde{\mathbf{p}}_j = \mathbf{p}_j + (L-1) \times \mathbf{n}_{\text{wind}}$ where $\mathbf{n}_{\text{wind}} \sim \mathcal{N}(\mathbf{n}_{\text{wind}}; 0, 0.002 \times \mathbf{I}_2)$.

$p(\tilde{r}(\mathbf{z})|\mathcal{D}_t)$ using $\mathcal{D}_t$ and (ii) obtain $\mathbf{z}_{t+1} = \arg\max_{\mathbf{z}\in\mathcal{Z}} \alpha_{t+1}(\mathbf{z}|\mathcal{D}_t)$ using $p(\tilde{r}(\mathbf{z})|\mathcal{D}_t)$. The so-termed acquisition function (AF) $\alpha_{t+1}(\mathbf{z}|\mathcal{D}_t)$, often expressed in closed-form, is chosen so as to balance *exploration* and *exploitation* of the search space [13]. There are several choices for both $p(\tilde{r}(\mathbf{z})|\mathcal{D}_t)$ and $\alpha_{t+1}(\mathbf{z}|\mathcal{D}_t)$. In this work, we will focus on the Gaussian process (GP) based surrogate model that has been widely used in a gamut of BO applications; see e.g., [13], [22], and will use the expected improvement (EI) as AF.

A. BO with single GP and EI

Belonging to the family of nonparametric Bayesian approaches, GPs offer a principled framework to learn an unknown nonlinear function with well-quantified uncertainty, which can readily guide the acquisition of new query input points. Learning with GPs begins with the assumption that a GP prior is postulated over the function $\tilde{r}(\cdot)$ as $\tilde{r} \sim \mathcal{GP}(0, \kappa(\mathbf{z}, \mathbf{z}'))$, with $\kappa(\mathbf{z}, \mathbf{z}')$ representing a positive-definite kernel function that measures the pairwise similarity between $\mathbf{z}$ and $\mathbf{z}'$. This means that the random vector $\tilde{\mathbf{r}}_t := [\tilde{r}(\mathbf{z}_1), \ldots, \tilde{r}(\mathbf{z}_t)]^\top$ comprising the function values at inputs $\mathbf{Z}_t := [\mathbf{z}_1, \ldots, \mathbf{z}_t]^\top$ is Gaussian distributed as $\tilde{\mathbf{r}}_t \sim \mathcal{N}(\tilde{\mathbf{r}}_t; \mathbf{0}_t, \mathbf{K}_t)(\forall t)$, where $\mathbf{K}_t$ is the $t \times t$ kernel (covariance) matrix whose (i,j)th element is $[\mathbf{K}_t]_{i,j} = \text{cov}(\tilde{r}(\mathbf{z}_i), \tilde{r}(\mathbf{z}_j)) := \kappa(\mathbf{z}_i, \mathbf{z}_j)$ [21].

Let $\mathbf{y}_t := [y_1, \ldots, y_t]^\top$ denote the observed output vector where $y_\tau = \tilde{r}(\mathbf{z}_\tau) + \varepsilon_\tau$ $(\forall \tau)$ and $\varepsilon_\tau \sim \mathcal{N}(\varepsilon_\tau; 0, \sigma_n^2)$ representing the GP model noise, also known as observation noise, that is assumed to be white Gaussian uncorrelated across τ. Note that the GP model noise ε_τ is different from the environmental noise $h(\mathbf{p})$ described in Sec. III. Then it can be shown that the function posterior pdf of $\tilde{r}(\mathbf{z})$ at any input $\mathbf{z}$ is Gaussian distributed as

$$p(\tilde{r}(\mathbf{z})|\mathcal{D}_t) = \mathcal{N}(\tilde{r}(\mathbf{z}); \mu_t(\mathbf{z}), \sigma_t^2(\mathbf{z})) \tag{9}$$

with mean and variance given in closed form as [21]

$$\mu_t(\mathbf{z}) = \mathbf{k}_t^\top(\mathbf{z})(\mathbf{K}_t + \sigma_n^2 \mathbf{I}_t)^{-1} \mathbf{y}_t \tag{10a}$$

$$\sigma_t^2(\mathbf{z}) = \kappa(\mathbf{z}, \mathbf{z}) - \mathbf{k}_t^\top(\mathbf{z})(\mathbf{K}_t + \sigma_n^2 \mathbf{I}_t)^{-1} \mathbf{k}_t(\mathbf{z}) \tag{10b}$$

where $\mathbf{k}_t(\mathbf{z}) := [\kappa(\mathbf{z}_1, \mathbf{z}), \ldots, \kappa(\mathbf{z}_t, \mathbf{z})]^\top$. Note that the mean in (10a) provides a point estimate of the function $\tilde{r}$ evaluated at $\mathbf{z}$ and the variance in (10b) quantifies the associated uncertainty.

Leveraging the Gaussian posterior pdf $p(\tilde{r}(\mathbf{z})|\mathcal{D}_t)$, the next input point $\mathbf{z}_{t+1}$ can be queried using the EI AF as [28]

$$\mathbf{z}_{t+1} = \arg\max_{\mathbf{z}\in\mathcal{Z}} \Delta_t(\mathbf{z}) \Phi\left(\frac{\Delta_t(\mathbf{z})}{\sigma_t(\mathbf{z})}\right) + \sigma_t(\mathbf{z}) \phi\left(\frac{\Delta_t(\mathbf{z})}{\sigma_t(\mathbf{z})}\right) \tag{11}$$

where $\Delta_t(\mathbf{z}) := \mu_t(\mathbf{z}) - \hat{\tilde{r}}_t^{\max}$ and $\Phi(\cdot), \phi(\cdot)$ represent the Gaussian cumulative density function and Gaussian pdf respectively. The EI AF is one of the most prevalent acquisition criteria in BO due to its well-documented merits in balancing well exploration and exploitation of the search space [28]. This is particularly appealing in the VP problem under noisy environments introduced in the present work, since the $\tilde{r}(\cdot)$ function is unknown with no prior information about the noise, and the goal is to find the function optimum relying solely on some available function values.

Although effective in several practical settings, the performance of the single GP-EI method hinges on the *pre-selected* kernel function, whose proper selection is a nontrivial and challenging task without any additional prior knowledge.

B. BO with ensemble GPs and adaptive EI

To cope with this challenge, an ensemble of M GP models (EGPs) is advocated to model $\tilde{r}(\cdot)$ which is *agnostic* to the appropriate kernel at first, but will iteratively *learn* the latter as new input-output data will become available. Each GP model $m \in \mathcal{M} := \{1, \ldots, M\}$ relies on a distinct kernel function $\kappa^m(\cdot,\cdot)$ implying that for each model m, a unique GP prior is placed on $\tilde{r}$ as $\tilde{r}|m \sim \mathcal{GP}(0, \kappa^m(\mathbf{z}, \mathbf{z}'))$. Note that the kernel set $\mathcal{K} := \{\kappa_m\}_{m=1}^M$ comprises kernels of different types and with different hyperparameters. Combining the M GP priors in the ensemble with the initial weights $\{w_0^m\}_{m=1}^M$ forms the Gaussian mixture (GM) ensemble prior $\tilde{r} \sim \sum_{m=1}^M w_0^m \mathcal{GP}(0, \kappa^m(\mathbf{z}, \mathbf{z}'))$ with $\sum_{m=1}^M w_0^m = 1$. Each weight $w_0^m := \Pr(i = m)$ represents the probability that captures the influence of GP model m in the ensemble.

Capitalizing on the ensemble GM prior and the available budget of input-output pairs $\mathcal{D}_t$ at slot t, one can obtain the ensemble posterior pdf via the sum-product rule as

$$p(\tilde{r}(\mathbf{z})|\mathcal{D}_t) = \sum_{m=1}^M w_t^m p(\tilde{r}(\mathbf{z})|i=m, \mathcal{D}_t) \tag{12}$$

which is a GM posterior pdf used as a surrogate model for $\tilde{r}$. For each GP model m, the corresponding weight $w_t^m := \Pr(i = m|\mathcal{D}_t)$ is obtained via Bayes' rule as

$$w_t^m \propto p(\mathcal{D}_t|i=m)\Pr(i=m) = p(\mathcal{D}_t|i=m) w_0^m \tag{13}$$

where $p(\mathcal{D}_t|i=m)$ is the marginal likelihood of $\mathcal{D}_t$ at slot t and i is a latent variable representing the influence of model m in the ensemble. The latter in the GP regression setting is Gaussian distributed as $p(\mathcal{D}_t|i=m) = \mathcal{N}(\mathbf{y}_t; \mathbf{0}_t, \mathbf{K}_t^m + (\sigma_n^m)^2 \mathbf{I}_t)$ where $\mathbf{K}_t^m$ and $(\sigma_n^m)^2$ are the kernel matrix and noise variance of the m^{th} GP model respectively [21].

With the posterior pdf $p(\tilde{r}(\mathbf{z})|\mathcal{D}_t)$ and the weights $\{w_t^m\}_{m=1}^M$ at hand, the next input query point $\mathbf{z}_{t+1}$ can be obtained by first selecting a certain GP model in the ensemble as $m_t \sim \mathcal{CAT}(\mathcal{M}, \mathbf{w}_t)$, where $\mathcal{CAT}(\mathcal{M}, \mathbf{w}_t)$ is a categorical distribution that draws a value from $\mathcal{M}$ with probabilities

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

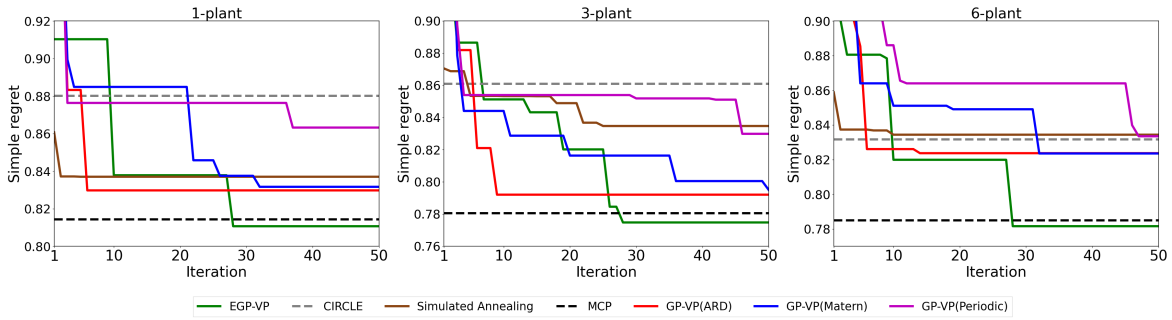
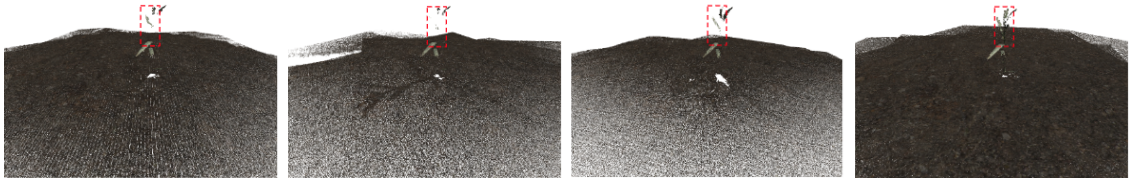
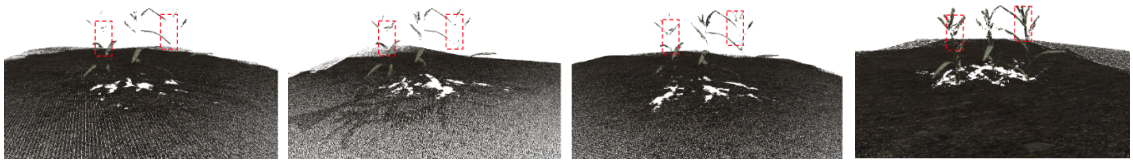


Fig. 4: Simple regret for (a) 1-plant, (b) 3-plant, and (c) 6-plant scenarios.



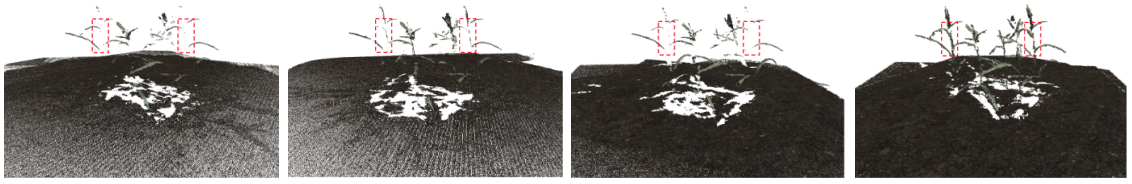
(a) Circle **(b)** Simulated Annealing **(c)** MCP **(d)** EPG-VP

Fig. 5: Reconstruction results from all competing methods shown on the 1-plant case from the same viewpoint to demonstrate differences.



(a) Circle **(b)** Simulated Annealing **(c)** MCP **(d)** EPG-VP

Fig. 6: Reconstruction results from all competing methods shown on the 3-plant case from the same viewpoint to demonstrate differences.



(a) Circle **(b)** Simulated Annealing **(c)** MCP **(d)** EPG-VP

Fig. 7: Reconstruction results from all competing methods shown on the 6-plant case from the same viewpoint to demonstrate differences.

$\mathbf{w}_t := [w_t^1, \ldots, w_t^M]^\top$. Upon selecting the GP model m_t, $\mathbf{z}_{t+1}$ is obtained using the following adaptive ensemble EI-based acquisition criterion

$$\mathbf{z}_{t+1} = \arg\max_{\mathbf{z}\in\mathcal{Z}} \Delta_t^{m_t}(\mathbf{z})\Phi\left(\frac{\Delta_t^{m_t}(\mathbf{z})}{\sigma_t^{m_t}(\mathbf{z})}\right) + \sigma_t^{m_t}(\mathbf{z})\phi\left(\frac{\Delta_t^{m_t}(\mathbf{z})}{\sigma_t^{m_t}(\mathbf{z})}\right) \quad (14)$$

with $\Delta_t^{m_t}(\mathbf{x}) := \mu_t^{m_t}(\mathbf{x}) - \hat{\tilde{r}}_t^{\max}$ and $\mu_t^{m_t}(\mathbf{x})$, $\sigma_t^{m_t}(\mathbf{x})$ denoting the posterior mean and variance of the m^{th} GP model respectively (c.f. (10a) and (10b)). Note that unlike the single GP-EI criterion in (11), the advocated AF in (14) prudently adapts to the m^{th} GP model at each slot t as new input-output data are processed in an online fashion, balancing well the exploration and exploitation of the search space. Also note that upon acquiring the output value $\tilde{r}(\mathbf{z}_{t+1})$, the weights and the posterior pdf of each GP model in the ensemble are updated using (9) and (13) respectively. Fig. 2 illustrates the steps at each iteration of the BO process of the our EGP method with adaptive EI AF, used for the VP problem of interest (henceforth abbreviated as 'EGP-VP').

Remark 1. In this paper, we consider noise on the point cloud $\{\mathbf{p}_i\}_{i=1}^{\mathcal{P}}$ representing the environment. Incorporating other types of noise including sensor noise or noise in the cameras location, belongs to our future research agenda.

Remark 2. The present work focuses on settings where the number of available cameras N is small to show that the proposed approach can offer 3D reconstruction of good quality, even with a limited budget of available cameras. If more cameras are available, they can be readily exploited by the advocated method to improve the reconstruction quality.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

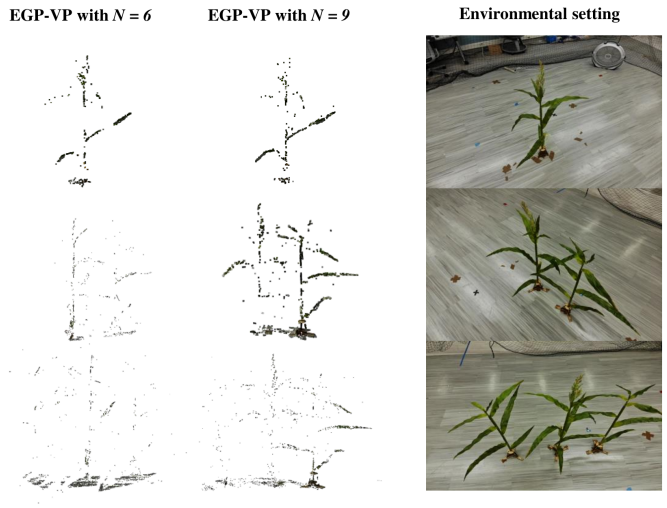


Fig. 8: Reconstructed corn plants using EGP-VP with $N = 6$ cameras (left) and $N = 9$ cameras (middle), shown alongside the real-world agricultural setting (right) with 1-, 2-, and 3- corn plants.

Note that when N becomes large, the dimensionality of the input vector $\mathbf{z}$ also becomes large, motivating the extension of our approach to EGP-based high-dimensional BO techniques that belong to our future research agenda.

V. Experiments

A. Implementation Details

Simulated realistic environments. In our experimental setup and evaluation, we used the PyRender software [30] for the photo-realistic scene creation and the image rendering process. We considered three different environmental scenarios with 1-plant, 3-plant (in row formation), and 6-plant (in rectangle formation). The data used to create the simulated scenes were real-world corn-plant reconstructions as presented in [31]. Each plant was given a random rotation around its vertical axis and a scale within 10% of the original size to get a realistic representation of the environment. In our setup, we consider *additive* and *linear* noise emulating the wind which is the most common source in agricultural environments that creates disturbances in the crops. To emulate this noise, we considered L levels of the plants and for each level $l \in \{1,\dots,L\}$ we have applied a noise value $l \times \mathbf{n}_{\text{wind}}$ where $\mathbf{n}_{\text{wind}} \sim \mathcal{N}(\mathbf{n}_{\text{wind}}; 0, 0.002 \times \mathbf{I}_2)$ as shown in Fig. 3. The intuition is that the effect of the wind is more evident in the upper parts of the plants in real-world settings. In order to get solutions (i.e., locations and orientations of cameras) that can handle different behaviors of the wind, we considered five distinct realizations of $\mathbf{n}_{\text{wind}}$ to create five different noisy environments during the optimization process, and three new unseen noise realizations for testing. It is worth noticing that although the present work emphasizes on the additive and linear wind noise for demonstration purposes, our formulation in Sec. III can account for other types of noise as well. All the experiments where conducted using an Intel Core i7-5930K CPU and a Titan X GPU. The run time for our method was 6.31, 8.89, and 12.20 minutes for the 1-, 3-, and 6-plant cases accordingly.

Real-world experiment. In addition to realistic simulated environments, we further show the benefits of our EGP-VP approach in a real-world agricultural setting consisting of 1-, 2-, and 3-plant models. We used EGP-VP to find the optimal cameras configuration to guide a commercial UAV equipped with a 12 MP camera, in collecting informative images of the sought agricultural environment. Note that each iteration of the EGP-VP approach involves the placement of N cameras, which in our setting corresponds to N distinct positions from which the UAV captures images during a single iteration. To account for environmental noise perturbations, we used a fan to emulate wind disturbances at various locations. We considered two distinct noise realizations to collect images and build a noisy point cloud representation of the environment $\{\tilde{\mathbf{p}}_i\}_{i=1}^{\mathcal{P}}$. Upon identifying the optimal camera configuration for the real-world environment using the proposed EGP-VP method, we then collected the images at the corresponding locations under a new third noise realization used for testing. A representative demonstration of our experimental setting is presented in the supplementary video material.

The 3D reconstruction process was conducted with COLMAP [32], [33] using the default settings in both simulated and real environments. For the RGB camera parameters, we set the $FoV = \frac{\pi}{2}$, $\theta_{match} = \frac{\pi}{4}$, and the rendering resolution at 2000×1500 for the simulated and 4056×3040 for the real-world environments respectively. To demonstrate the benefits of reconstructing 3D areas of interest in noisy environments with a small camera budget, the number of cameras N in our experiments was set to $N = 6$ for all cases. This mathematically means that each input vector $\mathbf{z}$ in the optimization method has dimensions 36×1 since each camera is represented as a 6×1 vector (c.f. Sec. III).

To optimize the noisy black-box reward function $\tilde{r}(\cdot)$, the adaptive EGP-VP framework described in Sec. IV-B was employed, with the ensemble consisting of (i) a GP model with a periodic kernel, (ii) a GP with an RBF kernel with automatic relevance determination (ARD), and (iii) a GP with a Matern kernel with parameter $\nu = 2.5$. For a given camera location and orientation $\mathbf{z}$ (input), the corresponding $\tilde{r}(\mathbf{z})$ (output) in all cases is the average reward function value of the five distinct noisy environments created by the five distinct realizations of $\mathbf{n}_{\text{wind}}$. In our experimental setting, 50 randomly selected input-output evaluation pairs $\{\mathbf{z}_\tau, \tilde{r}(\mathbf{z}_\tau)\}_{\tau=1}^{50}$ were used to train the initial parameters of the EGP-VP model and 50 iterations were considered for the BO process. In the training process, the kernel parameters of each GP model in the ensemble were obtained maximizing the marginal log-likelihood via *GPytorch* [34] and *Botorch* [35] packages, using the initial 50 training data. Note that the kernel parameters of each GP model were also updated at every iteration of the BO process, where newly acquired input-output data become available online.

To demonstrate the benefits of the EGP model in the BO process, we have initially compared the advocated EGP-VP approach with the single GP-based counterparts abbreviated as 'GP-VP'. Then, we compared our proposed EGP-VP method with three baselines; namely (i) a standard circular formation where we experimented with different values of the radius and the altitude and considered the best-performing one, (ii)

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

TABLE I: Quantitative results of the reconstructions based on point cloud completion using the average CD and MAE metrics (lower is better) along with the corresponding standard deviations. The results shown consider 3 testing (unseen) realizations of noise.

Table 1	**1-Plant**		**3-Plant**		**6-Plant**	
Method	CD ($\times$ 100)	MAE (cm)	CD ($\times$ 100)	MAE (cm)	CD ($\times$ 100)	MAE (cm)
Circle	2.257 ± 0.787	1.09 ± 0.337	2.934 ± 0.159	1.084 ± 0.321	2.726 ± 0.0356	1.101 ± 0.34
SA	2.278 ± 0.229	1.124 ± 0.37	2.151 ± 0.004	1.121 ± 0.375	2.095 ± 0.187	1.115 ± 0.376
MCP	1.533 ± 0.094	$\mathbf{1.07 \pm 0.312}$	1.725 ± 0.058	$\mathbf{1.054 \pm 0.294}$	2.797 ± 0.015	1.018 ± 0.312
EGP-VP (Ours)	$\mathbf{1.347 \pm 0.052}$	$\mathbf{1.07 \pm 0.408}$	$\mathbf{1.452 \pm 0.048}$	1.068 ± 0.403	$\mathbf{1.749 \pm 0.141}$	$\mathbf{1.017 \pm 0.286}$

the discrete formulation counterpart described in [1] that uses Simulated Annealing (SA) [36] for the optimization problem, and (iii) the discrete formulation of the maximum coverage problem (MCP) [37] integrated into the optimization problem described in (8), while also accounting for the existing noise in the environment.

B. Evaluation

Initially, the performance of the proposed approach in optimizing the reward function $\tilde{r}(\cdot)$ is evaluated using the simple regret (SR) metric, which at iteration t is expressed as

$$SR(t) := \tilde{r}(\mathbf{z}^*) - \max_{\tau \in \{1,\ldots,t\}} \tilde{r}(\mathbf{z}_\tau). \tag{15}$$

Fig. 4 illustrates the SR performance of all competing approaches for the (a) 1-plant, (b) 3-plant, and (c) 6-plant cases, where the optimal value $\tilde{r}(\mathbf{z}^*) = 1$ in (15) represents the maximum reward function value. It is evident that our EGP-VP approach not only enjoys smaller SR compared to the single GP counterparts, but also consistently outperforms the baselines in all cases. This showcases its merits in balancing well the exploration and exploitation of the search space, and adopting a continuous formulation over a discrete one as in the SA approach [1].

Upon obtaining the camera placement vector $\mathbf{z}^{\text{EGP-VP}}$ that maximizes the $\tilde{r}(\cdot)$ function while considering five distinct noise realizations in the optimization process, the reconstruction quality was quantitatively assessed using the Chamfer Distance (CD) [38], that is a widely used metric for point cloud completion [39], [40], and the mean absolute error (MAE) of the depth images computed by the reconstruction process. Table I reports the average CD and the average MAE along with the corresponding standard deviation of all approaches considering three new test distinct noise $\mathbf{n}_{\text{wind}}$ realizations, to assess the generalization of the $\mathbf{z}^{\text{EGP-VP}}$ solution in new unseen environmental noise perturbations. One important notice here is that compared to the MAE criterion, the CD metric takes into account only the crop information of the environment and not that of the ground, with the latter being less important for phenotype analysis in agricultural applications. It can be clearly seen that our advocated EGP-VP method enjoys the lowest average CD and MAE compared to the baselines, which translates to improved qualitative 3D reconstructions for all three environmental cases. This corroborates the effectiveness of the proposed EGP-VP approach in selecting appropriate camera placements that provide sufficient information to accurately reconstruct areas of interest, even with a limited budget of available cameras. In Figs. 5, 6, and 7, we demonstrate the reconstructed point clouds for all methods in every scenario we considered. It is evident in Figs. 5, 6, and 7 that our EGP-VP method offers the best reconstruction of the corn plants qualitatively, as can be observed by the annotated red boxes, that demonstrate substantial point cloud completeness compared to the baselines. It is also worth noticing that our approach can reconstruct better the crop stems compared to the baselines. This enables the phenotype analysis in agricultural applications [2] since stems constitute an essential component of the structure of the plants.

Having established the effectiveness of the proposed EGP-VP framework in the realistic simulated environments, we then evaluate its performance on three real-world agricultural setting considering 1-,2-, and 3-plants. Fig. 8 presents the three environments along with the reconstructions by EGP-VP using $N = 6$ and $N = 9$ cameras, clearly illustrating that EGP-VP successfully identifies an appropriate camera configuration for capturing informative images, enabling accurate reconstruction of the real and noisy 1-, 2-plant agricultural environments and satisfactory reconstruction in the 3-plant environment. In the latter case the reconstruction quality can be further improved increasing the number of cameras N. In addition, the running time for obtaining the noisy point cloud of the environment $\{\tilde{\mathbf{p}}_i\}_{i=1}^{\mathcal{P}}$ in the 1-plant environment is $t_{\text{init_pc}} = 1189.9$s, and the VP process by our EGP-VP method is $t_{\text{VP}} = 404.69$s. It is evident that $t_{\text{VP}} << t_{\text{init\_pc}}$, demonstrating that EGP-VP can identify a few yet informative viewpoints with reduced running time, thereby enabling enhanced efficiency and reduced monitoring costs.

VI. Conclusions

This paper focuses on the VP problem of optimally placing a given set of cameras in a 3D space, to obtain sufficient visual information for accurate and efficient reconstruction of any 3D area of interest. Unlike existing approaches in the literature, this work is the first to incorporate the existing noise of the environment in the VP problem without knowledge of the analytic noise expression and considers a small number of cameras during the optimization for efficiency. To optimize the so-termed reward function that gives an estimate of the reconstruction quality and whose closed-form expression is unknown due to the embodied noise, an adaptive Bayesian optimization technique is advocated. This provides assistance in effectively optimizing the black-box reward function in a sample-efficient manner. While most existing VP approaches entail discretizing the search space, the proposed approach depends on a continuous optimization problem where any position belonging to the entire space can be explored. Testing on noisy agricultural settings demonstrates the efficacy of the novel approach in accurately reconstructing 3D areas with only a small budget of available cameras.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2026, Vienna, Austria. Cite as RA-L paper.

Future directions include additional experiments with noise in the input of the optimization and the images, along with theoretical and robustness analysis of the proposed method. Additional environmental scenarios can also be considered, with applications to forestry and fire monitoring or urban surveillance. Beyond using a fixed number of cameras, a future extension is to adapt our method to dynamically estimate the number of viewpoints based on scene complexity, similarly as in e.g. [12]. Additionally, while the EGP-VP-derived camera configuration and extracted images have so far been used only with COLMAP, the next step is to apply them to alternative approaches, including Neural Radiance Field-based ones.

REFERENCES

[1] A. Bacharis, H. J. Nelson, and N. Papanikolopoulos, "View planning using discrete optimization for 3d reconstruction of row crops," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 9195–9201.

[2] H. J. Nelson, C. E. Smith, A. Bacharis, and N. P. Papanikolopoulos, "Robust plant localization and phenotyping in dense 3d point clouds for precision agriculture," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 9615–9621.

[3] C. Peng and V. Isler, "View selection with geometric uncertainty modeling," *arXiv preprint arXiv:1704.00085*, 2017.

[4] P. Roy and V. Isler, "Active view planning for counting apples in orchards," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 6027–6032.

[5] E. Vidal, N. Palomeras, K. Istenič, N. Gracias, and M. Carreras, "Multi-sensor online 3d view planning for autonomous underwater exploration," *Journal of Field Robotics*, vol. 37, no. 6, pp. 1123–1147, 2020.

[6] W. Jing, J. Polden, P. Y. Tao, W. Lin, and K. Shimada, "View planning for 3d shape reconstruction of buildings with unmanned aerial vehicles," in *2016 14th International Conference on Control, Automation, Robotics and Vision (ICARCV)*. IEEE, 2016, pp. 1–6.

[7] N. Smith, N. Moehrle, M. Goesele, and W. Heidrich, "Aerial path planning for urban scene reconstruction: A continuous optimization method and benchmark," *ACM Transactions on Graphics*, vol. 37, no. 6, dec 2018.

[8] D. Zermas, V. Morellas, D. Mulla, and N. Papanikolopoulos, "Extracting phenotypic characteristics of corn crops through the use of reconstructed 3d models," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 8247–8254.

[9] K. A. Tarabanis, P. K. Allen, and R. Y. Tsai, "A survey of sensor planning in computer vision," *IEEE Transactions on Robotics and Automation*, vol. 11, no. 1, pp. 86–104, 1995.

[10] G. H. Tarbox and S. N. Gottschlich, "Planning for complete sensor coverage in inspection," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 84–111, 1995.

[11] C. Peng and V. Isler, "Adaptive view planning for aerial 3d reconstruction," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 2981–2987.

[12] S. Pan, H. Hu, and H. Wei, "Scvp: Learning one-shot view planning via set covering for unknown object reconstruction," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 1463–1470, 2022.

[13] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. De Freitas, "Taking the human out of the loop: A review of Bayesian optimization," *Proceedings of the IEEE*, vol. 104, no. 1, pp. 148–175, 2015.

[14] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," *Neural Information Processing Systems*, vol. 25, 2012.

[15] K. Korovina, S. Xu, K. Kandasamy, W. Neiswanger, B. Poczos, J. Schneider, and E. Xing, "Chembo: Bayesian optimization of small organic molecules with synthesizable recommendations," *International Conference on Artificial Intelligence and Statistics*, pp. 3393–3403, 2020.

[16] Z. Wang and S. Jegelka, "Max-value entropy search for efficient Bayesian optimization," *International Conference on Machine Learning*, pp. 3627–3635, 2017.

[17] A. Cully, J. Clune, D. Tarapore, and J.-B. Mouret, "Robots that can adapt like animals," *Nature*, vol. 521, no. 7553, pp. 503–507, 2015.

[18] R. Marchant and F. Ramos, "Bayesian optimisation for informative continuous path planning," in *International Conference on Robotics and Automation*, 2014, pp. 6136–6143.

[19] 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*. IEEE, 2014, pp. 4227–4233.

[20] S. Pan, H. Hu, H. Wei, N. Dengler, T. Zaenker, and M. Bennewitz, "One-shot view planning for fast and complete unknown object reconstruction," *arXiv preprint arXiv:2304.00910*, 2023.

[21] C. E. Rasmussen and C. K. Williams, *Gaussian processes for machine learning*. MIT press Cambridge, MA, 2006.

[22] Q. Lu, K. D. Polyzos, B. Li, and G. B. Giannakis, "Surrogate modeling for bayesian optimization beyond a single gaussian process," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 9, pp. 11 283–11 296, 2023.

[23] K. D. Polyzos, Q. Lu, and G. B. Giannakis, "Bayesian optimization with ensemble learning models and adaptive expected improvement," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023.

[24] D. Nguyen, S. Gupta, S. Rana, A. Shilton, and S. Venkatesh, "Bayesian optimization for categorical and category-specific continuous inputs," *AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, pp. 5256–5263, 2020.

[25] S. Gopakumar, S. Gupta, S. Rana, V. Nguyen, and S. Venkatesh, "Algorithmic assurance: An active approach to algorithmic testing using Bayesian optimisation," *Neural Information Processing Systems*, pp. 5470–5478, 2018.

[26] K. D. Polyzos, Q. Lu, and G. B. Giannakis, "Weighted ensembles for adaptive active learning," *IEEE Transactions on Signal Processing*, 2024.

[27] W. R. Thompson, "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples," *Biometrika*, vol. 25, no. 3/4, pp. 285–294, 1933.

[28] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *Journal of Global optimization*, vol. 13, no. 4, pp. 455–492, 1998.

[29] N. Srinivas, A. Krause, S. Kakade, and M. Seeger, "Gaussian process optimization in the bandit setting: No regret and experimental design," in *International Conference on Machine Learning*, 2010.

[30] M. Matl, "Pyrender," https://github.com/mmatl/pyrender, 2019.

[31] H. J. Nelson and N. Papanikolopoulos, "Learning continuous object representations from point cloud data," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 2446–2451.

[32] J. L. Schönberger and J.-M. Frahm, "Structure-from-motion revisited," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[33] J. L. Schönberger, E. Zheng, M. Pollefeys, and J.-M. Frahm, "Pixelwise view selection for unstructured multi-view stereo," in *European Conference on Computer Vision (ECCV)*, 2016.

[34] J. R. Gardner, G. Pleiss, D. Bindel, K. Q. Weinberger, and A. G. Wilson, "Gpytorch: Blackbox matrix-matrix gaussian process inference with gpu acceleration," in *Advances in Neural Information Processing Systems*, 2018.

[35] M. Balandat, B. Karrer, D. R. Jiang, S. Daulton, B. Letham, A. G. Wilson, and E. Bakshy, "BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization," in *Advances in Neural Information Processing Systems 33*, 2020.

[36] S. Kirkpatrick, C. D. Gelatt Jr, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[37] M. S. Daskin, "A maximum expected covering location model: formulation, properties and heuristic solution," *Transportation science*, vol. 17, no. 1, pp. 48–70, 1983.

[38] F. Williams, "Point cloud utils," 2022, https://www.github.com/fwilliams/point-cloud-utils.

[39] Y. Guo, H. Wang, Q. Hu, H. Liu, L. Liu, and M. Bennamoun, "Deep learning for 3d point clouds: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 12, pp. 4338–4364, 2020.

[40] T. Wu, L. Pan, J. Zhang, T. Wang, Z. Liu, and D. Lin, "Balanced chamfer distance as a comprehensive metric for point cloud completion," *Advances in Neural Information Processing Systems*, vol. 34, pp. 29 088–29 100, 2021.