Uncertainty-Aware Informative Path Planning for UAV: An Interpretable Bayesian-DRL Approach

Jeonghyeon Pak, Song Ma, *Member, IEEE*, Yuanchang Liu, *Member, IEEE*, and Hyoung Il Son, *Senior Member, IEEE*

***Abstract*—Efficient unmanned aerial vehicle (UAV)-based crop monitoring requires disease-relevant observations within a limited flight budget. This paper presents a task-oriented informative path planning (IPP) framework for motion-efficient disease source localization in partially observable agricultural environments. A voxelwise Bayesian belief represents the latent disease field, while a twin delayed deep deterministic policy gradient (TD3) policy generates continuous sensing actions from the UAV pose, local crop descriptors, row-alignment information, and spectral disease cues. The reward combines posterior covariance-trace reduction, local disease evidence, and translational motion cost, and the accumulated belief supports source estimation. The method was compared with coverage path planning, three adaptive IPP methods, and privileged oracle References under a common protocol. Across three fixed-hotspot layouts and one random-hotspot condition, each evaluated over 20 trials, it achieved 90-100% localization success and mean localization errors of 3.80-4.73 m. Motion cost was reduced by 41.9-65.5% relative to coverage planning and by 20.9-65.8% relative to the adaptive baseline with the highest average localization success. Sensitivity and ablation studies examined voxel resolution, prior initialization, and framework components. Field-derived LiDAR and spectral maps further assessed trajectory-level applicability to measured agricultural data.**

***Note to Practitioners*—This work addresses UAV-based crop monitoring tasks in which disease-related evidence must be acquired under a limited flight budget. Conventional coverage planning systematically surveys the field but may require unnecessary traversal when the operational objective is to identify a disease-relevant region rather than to reconstruct the entire field. The proposed method combines a Bayesian spatial belief with a continuous TD3 policy that selects sensing motions according to domain-aware agricultural features and translational motion cost. In simulation, the method maintained high disease source localization success while providing a favorable overall tradeoff between localization reliability and traversal cost relative to CPP and the evaluated adaptive IPP methods. The field-data evaluation applies the planners to maps constructed from actual LiDAR and RGN measurements and is intended to assess trajectory-level applicability rather than biologically verified disease diagnosis. Physical UAV deployment will additionally require sensor calibration, onboard computation, localization under canopy-induced disturbances, and independently confirmed disease ground truth.**

***Index Terms*—Aerial monitoring, informative path planning, precision agriculture, remote sensing, source localization.**

Received 13 May 2026; revised 2 September 2026 and 2 October 2026; accepted 7 October 2026. Date of publication 26 October 2026; date of current version 29 October 2026. This article was recommended for publication by Associate Editor G. Silano and Editor J. Gu upon evaluation of the reviewers' comments. This work was supported in part by Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry (IPET) through the Agriculture and Food Convergence Technologies Program for Research Manpower Development, funded by Ministry of Agriculture, Food and Rural Affairs (MAFRA) under Project RS-2024-00397026; and in part by the National Research Foundation of Korea (NRF) grant funded by the Ministry of Science and ICT (MSIT) under Grant NRF-2023R1A2C1003701. *(Corresponding author: Hyoung Il Son.)*

Jeonghyeon Pak is with the Department of Convergence Biosystems Engineering, Chonnam National University, Gwangju 61186, Republic of Korea, and also with the Interdisciplinary Program in IT-Bio Convergence System, Chonnam National University, Gwangju 61186, Republic of Korea (e-mail: jhpak@jnu.ac.kr).

Song Ma and Yuanchang Liu are with the Department of Mechanical Engineering, University College London, WC1E 7JE London, U.K. (e-mail: yuanchang.liu@ucl.ac.uk).

Hyoung Il Son is with the Department of Convergence Biosystems Engineering, Chonnam National University, Gwangju 61186, Republic of Korea, also with the Interdisciplinary Program in IT-Bio Convergence System, Chonnam National University, Gwangju 61186, Republic of Korea, and also with the Institute of AI Green-Bio, Chonnam National University, Gwangju 61186, Republic of Korea (e-mail: hison@jnu.ac.kr).

Digital Object Identifier 10.1109/TASE.2026.3746465

I. Introduction

PRECISION agriculture aims to enhance productivity and sustainability by enabling data-driven crop management based on fine-grained environmental observations [1]. Accurate monitoring of crop health and disease distribution is necessary for timely interventions and efficient resource utilization. Recent advances in sensing technology, including remote sensing and artificial intelligence, have enabled large-scale data collection for agricultural monitoring.

In agricultural monitoring, a critical task is collecting comprehensive, high-quality environmental information. Unmanned aerial vehicles (UAVs) have emerged as a promising platform for monitoring due to their flexibility and ability to cover large areas rapidly [2], [3]. However, agricultural environments are complex and only partially observable, as they are characterized by dense crop canopies, unstructured terrain, and dynamic environmental conditions (Fig. 1). These factors make the efficient collection of informative data using conventional exploration strategies challenging.

Most existing UAV-based exploration approaches employ coverage path planning (CPP) or predefined trajectories to cover the environment uniformly. Although effective for geometric coverage, these methods do not consider the informativeness of observations. Consequently, they often waste resources on low-value regions, failing to prioritize areas that offer critical information (e.g., disease hotspots). Resource-efficient decision-making has also been investigated in heterogeneous multirobot systems through deadline-constrained task assignment and precedence-constrained vehicle routing [4], [5]. However, these studies primarily focus on task allocation and path coordination rather than the active selection of informative sensing actions [6], [7]. This gap is particularly

1558-3783 © 2026 IEEE. All rights reserved, including rights for text and data mining, and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: CHONNAM NATIONAL UNIVERSITY. Downloaded on October 30,2026 at 07:49:35 UTC from IEEE Xplore. Restrictions apply.

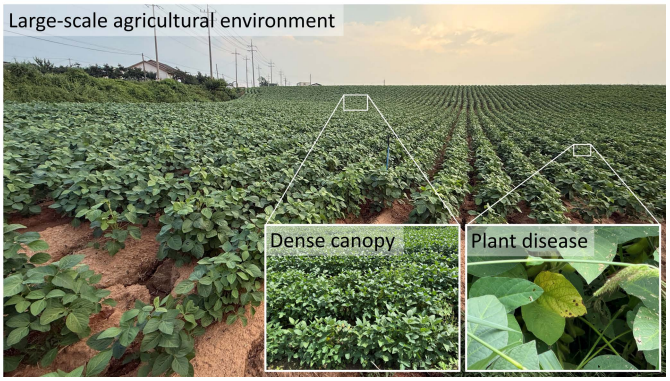


Fig. 1. Large-scale agricultural environments with heterogeneous disease distributions and complex, dense crop canopies.

relevant in agricultural environments, where spatially heterogeneous and clustered phenomena require information-driven exploration strategies.

To address these limitations, this study considers informative path planning (IPP) as a trajectory-level mechanism for acquiring the spatial evidence required for disease source localization. The mission-level objective is to identify the disease source location with limited traversal under a finite mission budget, whereas posterior uncertainty reduction serves as the information-theoretic criterion used to evaluate locally acquired observations. This distinction is important because exhaustive reconstruction of the entire disease field is not required once sufficient evidence has been accumulated for source localization.

The proposed framework reformulates the trajectory-level IPP objective as a sequential decision-making problem and uses TD3 to generate continuous sensing actions. A voxelwise Bayesian belief represents the latent disease field, while local crop geometry, crop-row orientation, and disease-related spectral observations constitute the policy state. The policy is trained using a motion-normalized uncertainty-reduction reward, while the accumulated Bayesian belief is used to estimate the disease source location. Accordingly, the formulation separates three functional roles: disease source localization defines the mission-level endpoint, posterior uncertainty reduction provides the information-acquisition criterion during policy learning, and locally observable agricultural features constitute the actor state for continuous action selection. The main contributions of this study are summarized as follows:

1) We formulate UAV-based crop monitoring as IPP, distinguishing disease source localization as the mission-level objective from posterior uncertainty reduction as the trajectory-level sensing criterion.
2) We develop a DRL-based local policy that approximates IPP by maximizing information gain under motion cost constraints.
3) We incorporate row-structural directional guidance into the local state representation to provide structured progression cues and reduce dependence on purely myopic local observations.
4) We integrate a voxelwise Bayesian belief that represents the latent disease field in an interpretable spatial form and supports final disease source localization.
5) We demonstrate in simulation that the proposed framework achieves reliable disease source localization with reduced traversal relative to the evaluated baselines, and further assess its trajectory-level applicability using field-derived LiDAR and RGN measurements.

The remainder of this paper is organized as follows. Section II reviews the related work, and Section III describes the problem. Next, Section IV details the proposed informative monitoring system, and Section V discusses the experimental setup. Then, Section VI provides the experimental results, and Section VII discusses the study limitations. Finally, Section VIII presents the conclusion.

II. Related Work

A. Aerial Agricultural Monitoring

UAVs have emerged as a primary platform for large-scale agricultural monitoring, enabling high-resolution, temporally consistent observations over extensive agricultural fields [8], [9]. Recent studies have continued to improve sensing reliability via better spectral calibration and domain-robust disease recognition. For example, recent studies have demonstrated that deep learning-based computer vision methods that apply RGB, multispectral, and hyperspectral imaging enable accurate, online plant disease detection without manual feature engineering [10]. These studies underscore the increasing maturity of UAV-based sensing in agriculture.

However, these studies have focused on sensing robustness, calibration, and recognition performance after data collection, rather than the active selection of informative viewpoints during data acquisition. These approaches improve inference from UAV data but do not address where the UAV should next sense, given spatial heterogeneity, occlusion, and a limited flight budget. This limitation is crucial in agricultural disease monitoring, where symptoms are commonly clustered and only partially visible in dense crop canopies. These observations motivate a perception- and information-aware exploration framework.

B. Classical Informative Path Planning Approaches

IPP seeks to maximize information gain under sensing and motion constraints [11]. Early approaches formulate IPP using information-theoretic criteria (e.g., mutual information and entropy reduction) and are often modeled with Gaussian processes (GP) [12], [13]. These problems are generally NP-hard but exhibit submodular properties, enabling efficient approximation algorithms with theoretical guarantees [14]. Nonmyopic, active-learning, sparse-optimization, and scalable mapping formulations have subsequently been developed to improve planning quality and computational tractability [15], [16], [17], [18]. These methods are primarily evaluated according to map uncertainty, reconstruction error, or information accumulated within a prescribed mission budget.

A related line of research uses informative sensing for active source identification, source seeking, or localization of high-interest regions. One study [19] formulates model-based active source identification by selecting sensing actions according to information associated with unknown source

Authorized licensed use limited to: CHONNAM NATIONAL UNIVERSITY. Downloaded on October 30,2026 at 07:49:35 UTC from IEEE Xplore. Restrictions apply.

parameters. Another study [20] investigates information-based source seeking to improve source-estimation convergence, while a related approach [21] considers information-guided maximum seek-and-sample in partially observable continuous environments. These formulations couple the sensing objective directly to source-related quantities or to the location of an unknown field maximum.

A more closely related formulation [22] uses a Gaussian-process field representation and optimizes a modified mutual-information criterion that also favors measurements in regions with high field magnitude. Source parameters are then estimated from the acquired measurements. This method is model-based rather than policy-learning-based, but it provides a relevant example of combining field-level information acquisition with source-directed exploitation. Unlike approaches that define sensing utility directly over source-related quantities, the present study maintains and refines a spatial disease belief from which the disease source is subsequently inferred. Accordingly, disease source localization is defined as the mission-level objective, whereas motion-normalized posterior uncertainty reduction serves as the trajectory-level sensing criterion.

C. Learning-Based Informative Path Planning Approaches

Learning-based adaptive IPP transfers part of the computational burden from online trajectory search to offline policy training. Rückin et al. [23] combine Monte Carlo tree search with an offline-trained policy-value network for adaptive information gathering, with performance evaluated primarily through map uncertainty and reconstruction quality. Vashisth et al. [24] construct dynamic local graphs in initially unknown three-dimensional environments and train a policy that balances environmental exploration with the exploitation of discovered targets of interest. More recent work has investigated policy transfer across different map representations and incremental sampling-based replanning [25], [26].

These approaches span a continuum from map-centric uncertainty reduction to target-aware exploration-exploitation. [23] is primarily directed toward informative map reconstruction, [24] incorporates target discovery into adaptive exploration, and [22] combines field estimation with source-directed sensing. The proposed method is positioned at the task-oriented end of this continuum: disease source localization defines mission completion, while motion-normalized posterior uncertainty reduction guides the acquisition of source-relevant evidence. The comparative evaluation therefore focuses on localization reliability and the traversal required to achieve it, rather than claiming uniformly superior global field reconstruction.

III. Problem Description

A. Partial Observability in Field-Scale Sensing

This study considers UAV-based informative exploration for monitoring plant diseases in an unknown agricultural environment. The principal mathematical symbols used in the problem formulation and the proposed framework are summarized in Table I. Disease severity varies spatially across individual plants within the structured crop-row layout. Due to the limited sensing range and occlusions, the UAV acquires only partial observations at each time step, preventing a full reconstruction of the environment.

TABLE I
Principal Mathematical Notation

Symbol	Definition
$\mathcal{F}$	Continuous agricultural monitoring domain
$H(x)$	Latent disease field at location x
$\mathcal{V}$	Discretized voxel set
p_t	UAV pose at time t
o_t	UAV observation at time t
s_t	Policy state
a_t	Continuous UAV action
q_t	Local crop observation descriptors
ξ_t	Domain-aware feature vector
$\hat{\psi}_t$	Estimated dominant crop-row orientation
τ_t	Row-alignment feature
h_t	Local disease-hotspot feature
μ_t	Posterior mean of the latent disease field
P_t^+	Posterior covariance
$b_t(v_i)$	Disease probability at voxel v_i
$\mathcal{U}_t$	Aggregate posterior uncertainty
O_t	Observation history up to time t
φ	UAV trajectory
$C(\varphi)$	Cumulative motion cost
$\hat{\zeta}$	Estimated disease source location
T	Maximum mission horizon
B	Maximum motion budget

The spatial domain is defined as $\mathcal{F} \subset \mathbb{R}^3$, and a latent disease field is defined as follows:

$$H : \mathcal{F} \to \mathbb{R}, \tag{1}$$

where $H(x)$ denotes the disease severity at location $x \in \mathcal{F}$. At time step t, the UAV with pose p_t acquires an observation:

$$o_t = h(p_t) + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2), \tag{2}$$

where $h(p_t)$ represents the observation obtained by fusing four-dimensional (4D) LiDAR and red-green-NIR (RGN) camera data. This observation provides cropwise spectral information to estimate disease-related features (e.g., vegetation indices).

$$\mathcal{O}_t = \{o_0, o_1, \ldots, o_t\}. \tag{3}$$

This sequence represents all measurements available to the UAV for decision-making under partial observability.

B. Task Objective and Information Criterion

The task objective is to obtain a reliable disease source estimate while limiting the traversal required during monitoring. Because the unknown source location cannot be used directly for online action selection, this task objective is supported by a trajectory-level informative sensing criterion. Specifically, posterior uncertainty reduction is balanced against traversal cost under a finite motion budget. Accordingly, the following resource-constrained IPP objective formalizes the trajectory-level tradeoff between information acquisition and traversal cost:

$$\varphi^\star = \arg\max_{\varphi \in \Phi} \frac{\mathcal{I}(\varphi)}{\mathcal{C}(\varphi)} \quad \text{s.t.} \quad \mathcal{C}(\varphi) \le B, \tag{4}$$

Authorized licensed use limited to: CHONNAM NATIONAL UNIVERSITY. Downloaded on October 30,2026 at 07:49:35 UTC from IEEE Xplore. Restrictions apply.

where $\varphi = (p_0, p_1, \ldots, p_T)$ denotes the UAV trajectory. The mission horizon T quantifies the total number of decision steps during a single exploration episode. In addition, Φ represents the feasible trajectory set, $\mathcal{I}(\varphi)$ indicates the cumulative information gain, and $\mathcal{C}(\varphi)$ denotes the total traversal cost.

The cumulative information gain is defined as follows:

$$\mathcal{I}(\varphi) = \sum_{t=0}^{T-1} \left[\mathrm{tr}(P_t^+) - \mathrm{tr}(P_{t+1}^+)\right], \tag{5}$$

where P_t^+ denotes the posterior covariance at time t. The trace operator offers a scalar measure of uncertainty. It should be noted that $\mathcal{I}(\varphi)$ is evaluated from the maintained posterior belief, whereas the incremental reduction at each step is generated only by measurements acquired within the current sensing footprint. Therefore, the objective does not assume global observation of the field; rather, aggregate belief uncertainty is progressively reduced through spatially localized sensing actions. The traversal cost is defined as follows:

$$\mathcal{C}(\varphi) = \sum_{t=0}^{T-1} c(a_t), \tag{6}$$

where $c(a_t)$ represents the incremental traversal cost associated with action a_t, penalizing large displacements and encouraging motion-efficient trajectories. Direct optimization of (4) is computationally intractable in continuous control spaces. This motivates a sequential decision-making formulation.

C. Online Decision-Making in Continuous Control Space

Solving the trajectory optimization problem in (4) is computationally intractable for continuous state and action spaces. The UAV operates in an online setting in which each action must be selected without prior knowledge of future observations.

Formally, the UAV must select an action a_t at each time step based on the current state s_t, without solving the global optimization problem:

$$a_t = \pi(s_t), \tag{7}$$

where $\pi : \mathcal{S} \to \mathcal{A}$ indicates the policy mapping from states to actions. Accordingly, the following sequential learning objective is adopted:

$$\pi^\star = \arg\max_{\pi} \mathbb{E}\left[\sum_{t=0}^{T-1} \gamma^t r(s_t, a_t, s_{t+1})\right], \tag{8}$$

where the cumulative stepwise reward provides a sequential learning objective motivated by the trajectory-level information-motion tradeoff in (4).

Algorithm 1 Informative Exploration Algorithm for Monitoring System

Input: Initial UAV pose p_0, horizon T, policy π_ϕ
Output: Replay buffer $\mathcal{D}$ and trajectory φ
for each training episode **do**
  Initialize $\varphi \leftarrow p_0$
  for $t = 0$ to $T - 1$ **do**
    Acquire partial observations
    $o_t \leftarrow h(p_t) + \varepsilon_t$
    Domain-aware feature extraction
    Estimate the row direction $\hat{\psi}_t$
    $\tau_t \leftarrow \cos(\psi_t - \hat{\psi}_t)$
    Compute normalized difference vegetation index for crops $k \in \mathcal{N}_t$
    $h_t \leftarrow \frac{1}{|\mathcal{N}_t|} \sum_{k \in \mathcal{N}_t} (1 - N_t^{(k)})$
    State construction
    $\xi_t \leftarrow (\tau_t, h_t)$
    $q_t \leftarrow \{(dx_t^{(k)}, dy_t^{(k)}, \ell_t^{(k)})\}_{k=1}^{K}$
    $s_t \leftarrow \{p_t, q_t, \xi_t\}$
    Select an informative action using the policy
    $a_t \leftarrow \pi_\phi(s_t) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma)$
    Environment transition
    $p_{t+1} \leftarrow f(p_t, a_t)$
    Update the Bayesian belief
    $K_t \leftarrow P_t^+ \mathbf{H}_t^\top (\mathbf{H}_t P_t^+ \mathbf{H}_t^\top + R_t)^{-1}$
    $\boldsymbol{\mu}_{t+1} \leftarrow \boldsymbol{\mu}_t + K_t(o_t - \mathbf{H}_t \boldsymbol{\mu}_t)$
    $P_{t+1}^+ \leftarrow (I - K_t \mathbf{H}_t) P_t^+$
    Reward
    $c_t \leftarrow \|p_{t+1} - p_t\|_2$
    $\widetilde{c}_t \leftarrow \max(c_t, \varepsilon_c)$
    $\mathcal{I}_t \leftarrow \mathrm{tr}(P_t^+) - \mathrm{tr}(P_{t+1}^+)$
    $\bar{d}_t \leftarrow \frac{1}{K} \sum_{k=1}^{K} \mathbb{1}_{d_t^{(k)}}$
    $r_t = \mathrm{clip}\left(\frac{\mathcal{I}_t(1+\lambda \bar{d}_t)}{\widetilde{c}_t}, r_{\min}, r_{\max}\right)$
    Construct the next state s_{t+1}
    Store transition (s_t, a_t, r_t, s_{t+1}) in $\mathcal{D}$
    $p_t \leftarrow p_{t+1}$
    Create the probability map
    for each $v_i \in \mathcal{V}_t$ **do**
      $b_{t+1}(v_i) \leftarrow \sigma(\mu_{t+1}(v_i))$
    end for
    $\hat{\zeta}_t \leftarrow \arg\max_{v_i \in \mathcal{V}} b_{t+1}(v_i)$
    $\varphi \leftarrow (\varphi, p_{t+1})$
  end for
end for
return: $\mathcal{D}, \varphi$

IV. Informative Monitoring System

A. System Framework

This paper proposes a DRL-based approximation framework that addresses the intractable trajectory-level optimization in (4) through a sequential decision-making formulation. The proposed informative monitoring system (Algorithm 1) is designed to support task-oriented informative exploration and disease source localization in agricultural environments. As illustrated in Fig. 2, the overall architecture comprises two closely coupled components: (i) an informative exploration training module and (ii) a feedback-loop execution module. Within this architecture, the Bayesian estimator and the TD3 actor play distinct roles: posterior uncertainty provides the information signal used during policy learning, whereas the actor generates continuous sensing actions from the compact local state.

Authorized licensed use limited to: CHONNAM NATIONAL UNIVERSITY. Downloaded on October 30,2026 at 07:49:35 UTC from IEEE Xplore. Restrictions apply.

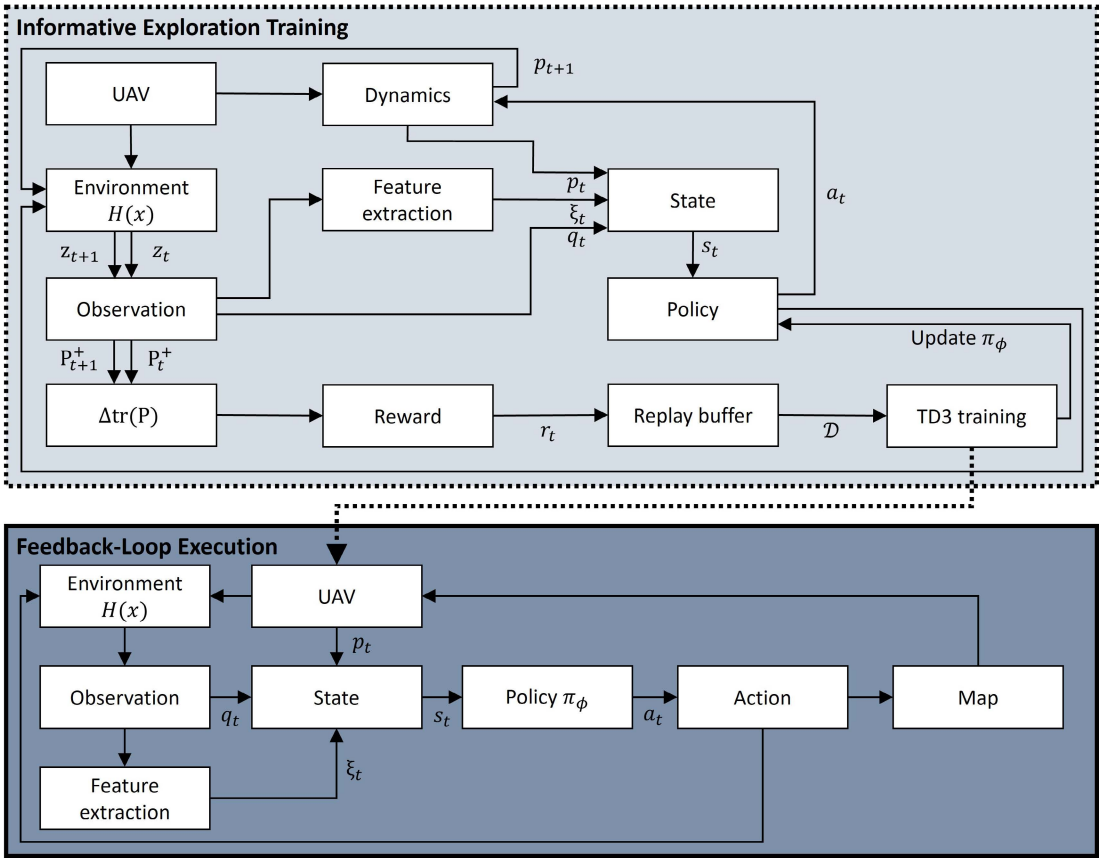


Fig. 2. Proposed system framework for informative exploration in agricultural environments. The UAV acquires partial observations and updates a probabilistic belief map of the underlying spatial distribution (e.g., disease intensity). A learning-based policy is trained with a task-oriented stepwise reward based on posterior uncertainty reduction, disease-related evidence, and motion cost, and generates continuous sensing actions from the local state.

During training, the UAV interacts with the environment to learn an information-driven exploration policy. At each time step, the UAV collects partial LiDAR-RGN observations and updates the belief about the latent disease field, as described in Section III-A. The dominant crop-row orientation is estimated from local LiDAR measurements and, although inferred locally, represents the principal progression direction of the surrounding field, thereby providing row-structural directional guidance. The acquired observations are processed using a feature extraction module to generate domain-aware features and local descriptors. These features are combined with the UAV pose to construct a compact state representation, serving as the input to the policy. The policy is trained via DRL to maximize the discounted cumulative stepwise reward defined in Section IV-D. This reward serves as a task-oriented sequential surrogate motivated by the information-motion objective in (4). State transitions are stored in a replay buffer and used for iterative TD3 updates.

After training, the learned policy is deployed in a closed-loop execution framework. At each time step, the UAV acquires new observations and constructs the state using the same feature extraction process. The policy maps the current state to an action, which is executed to update the pose of the UAV. This process forms a feedback loop in which new observations update the local state features used for action selection while simultaneously refining the Bayesian belief used for uncertainty estimation and source localization. The probability map is updated based on the Bayesian belief, supplying a spatial representation of disease likelihood. This representation is progressively refined during exploration, enabling disease source estimation from the accumulated Bayesian belief.

B. State Space

At each discrete time step t, with a given state $s \in S$, the agent selects actions $a \in A$ according to its policy $\pi : S \rightarrow A$, receiving a reward r and the new state of the environment s'. In actor-critic methods, the policy (the actor) can be updated using the deterministic policy gradient algorithm [27]. The problem is formulated as a Markov decision process, where the state at time step t is defined as follows:

$$s_t = \{p_t, q_t, \xi_t\}, \quad s_t \in \mathcal{S}. \tag{9}$$

The posterior mean $\boldsymbol{\mu}_t$ and covariance $\mathbf{P}_t^+$ are not included explicitly in the actor state $\mathbf{s}_t$. Posterior uncertainty affects policy learning through the uncertainty-reduction reward, while the Bayesian belief is updated in parallel during execution for uncertainty estimation and final disease source localization.

The UAV state is defined as follows:

$$p_t = (x_t, y_t, z_t, \psi_t), \tag{10}$$

which represents the UAV position and heading. The local observation feature is defined as follows:

$$q_t = \left\{ (dx_t^{(k)},\ dy_t^{(k)},\ \ell_t^{(k)}) \right\}_{k=1}^{K}, \tag{11}$$

Authorized licensed use limited to: CHONNAM NATIONAL UNIVERSITY. Downloaded on October 30,2026 at 07:49:35 UTC from IEEE Xplore. Restrictions apply.

where $dx_t^{(k)}$ and $dy_t^{(k)}$ denote relative positions, and $\ell_t^{(k)}$ represents the disease level of nearby crops. K denotes the number of neighboring crops included in the local observation. The domain knowledge vector is defined as follows:

$$\xi_t = (\tau_t, h_t), \tag{12}$$

where τ_t represents the row alignment, and h_t denotes the disease hotspot feature. The row-alignment feature τ_t represents the dominant crop-row direction, whereas the relative crop descriptors q_t encode the local spatial arrangement and disease levels of surrounding plants. The hotspot feature h_t complements this structural representation by summarizing local disease-related spectral evidence. Together, these features provide the policy with crop-structural and task-relevant guidance without requiring a predefined row map.

The row alignment feature is defined as follows:

$$\tau_t = \cos(\psi_t - \hat{\psi}_t), \tag{13}$$

where $\hat{\psi}_t$ is estimated from local LiDAR measurements. The normalized difference vegetation index (NDVI) for crop k is computed as follows:

$$N_t^{(k)} = \frac{\alpha_{\mathrm{NIR}}^{(k)} - \alpha_{\mathrm{R}}^{(k)}}{\alpha_{\mathrm{NIR}}^{(k)} + \alpha_{\mathrm{R}}^{(k)} + \epsilon}, \tag{14}$$

where $\alpha_{\mathrm{NIR}}^{(k)}$ and $\alpha_R^{(k)}$ denote the NIR- and red-band reflectances of crop k, respectively, and $\epsilon > 0$ prevents division by zero. The hotspot feature is defined as follows:

$$h_t = \frac{1}{|\mathcal{N}_t|} \sum_{k \in \mathcal{N}_t} \left(1 - N_t^{(k)}\right), \tag{15}$$

where $\mathcal{N}_t$ represents the set of crops in the sensing region.

C. Action Space

The UAV operates in a continuous control space, and the action at time step t is computed as follows:

$$a_t = (\Delta x_t, \Delta y_t, \Delta z_t, \Delta \psi_t), \tag{16}$$

where Δx_t, Δy_t, and Δz_t denote the incremental movements along each axis. This continuous formulation enables fine-grained control and adaptive trajectory generation. The corresponding motion cost is defined as follows:

$$c(a_t) = \|p_{t+1} - p_t\|_2. \tag{17}$$

D. Reward Function

Directly optimizing the trajectory-level objective in (4) is computationally intractable in continuous domains. We therefore define the following task-oriented stepwise reward for TD3 training:

$$r_t = \mathrm{clip}\left(\frac{\Delta \mathrm{tr}(P)\left(1 + \lambda \bar{d}_t\right)}{\widetilde{c}_t}, r_{\min}, r_{\max}\right), \tag{18}$$

where $\Delta \mathrm{tr}(P)$ denotes the reduction in posterior covariance trace at time step t, λ controls the weighting of local disease-related evidence, and $r_{\min}$ and $r_{\max}$ are the lower and upper reward bounds, respectively. Accordingly, (18) serves as a sequential surrogate motivated by the trajectory-level information-motion objective in (4), rather than as an exact mathematical decomposition of that objective. The local disease density is defined as

$$\bar{d}_t = \frac{1}{K} \sum_{k=1}^{K} \mathbb{1}_{d_t^{(k)}}, \tag{19}$$

where the term $\sum_{k=1}^{K} \mathbb{1}_{d_t^{(k)}}$ aggregates disease indicators in the local neighborhood. The effective motion cost is lower-bounded as

$$\widetilde{c}_t = \max\left(c(a_t), \varepsilon_c\right), \tag{20}$$

where $\varepsilon_c > 0$ is a small positive constant. The lower bound prevents excessively large reward values for near-zero displacement, while reward clipping limits extreme reward magnitudes and critic targets during TD3 training. Under the constant-speed setting, incremental displacement is proportional to travel time and is therefore used as a surrogate for traversal effort rather than as a direct energy model.

Algorithm 2 TD3-Based Training Algorithm for Informative Exploration Policy

Input: Replay buffer $\mathcal{D}$, actor π_ϕ, critics $Q_{\theta_1}, Q_{\theta_2}$
Initialize: Target networks $\pi_{\phi'}, Q_{\theta_1'}, Q_{\theta_2'}$
for each training iteration **do**
  Sample minibatch $(s_t, a_t, r_t, s_{t+1}) \sim \mathcal{D}$
  Compute the target action:

$$a'_{t+1} \leftarrow \mathrm{clip}\left(\pi_{\phi'}(s_{t+1}) + \varepsilon_t\right)$$

  Compute the target value:

$$y_t \leftarrow r_t + \gamma \min_{i=1,2} Q_{\theta_i'}(s_{t+1}, a'_{t+1})$$

  Update the critic network by minimizing:

$$L(\theta_i) \leftarrow \mathbb{E}\left[(Q_{\theta_i}(s_t, a_t) - y_t)^2\right], \quad i \in \{1, 2\}$$

  Update the actor network by minimizing:

$$L(\phi) \leftarrow -\mathbb{E}[Q_{\theta_1}(s_t, \pi_\phi(s_t))]$$

  Soft-update the actor and critic target networks
end for

E. Exploration Policy

The control policy is a deterministic mapping $a_t = \pi_\phi(s_t)$, where π_ϕ is parameterized by ϕ (Algorithm 2). The policy is trained to maximize the expected cumulative reward:

$$\pi_\phi^\star = \arg\max_{\pi_\phi} \mathbb{E}\left[\sum_{t=0}^{T-1} \gamma^t r_t\right], \tag{21}$$

where γ denotes the discount factor. The actor is therefore trained to maximize the discounted cumulative stepwise reward defined in (18).

The critic networks are trained by minimizing the Bellman error:

$$L(\theta_i) = \mathbb{E}\left[(Q_{\theta_i}(s_t, a_t) - y_t)^2\right], \tag{22}$$

Authorized licensed use limited to: CHONNAM NATIONAL UNIVERSITY. Downloaded on October 30,2026 at 07:49:35 UTC from IEEE Xplore. Restrictions apply.

where the target value is

$$y_t = r_t + \gamma \min_{i=1,2} Q_{\theta_i'}(s_{t+1}, a'_{t+1}). \tag{23}$$

The actor is updated using the deterministic policy gradient:

$$L(\phi) = -\mathbb{E}[Q_{\theta_1}(s_t, \pi_\phi(s_t))]. \tag{24}$$

Moreover, the learned policy is executed in a closed-loop manner:

$$s_{t+1} = f(s_t, a_t), \tag{25}$$

where the next state is obtained from new observations $o_t = h(p_t)$ and updated domain knowledge. This structure allows continuous adaptation to the environment during exploration.

F. Probability Map

For a spatially consistent representation of the disease likelihood, the proposed framework maintains a voxelwise probability map coupled with the Bayesian belief about the latent disease field. Unlike the state representation, where the cropwise NDVI is employed as a local spectral feature, the probability map applies the same NDVI quantity as an observation for a voxelwise belief update. This representation links belief estimation to the uncertainty-based training signal and the final source-estimation process.

Let the discretized spatial domain be defined as the following voxel set:

$$\mathcal{V} = \{v_1, v_2, \ldots, v_N\}, \quad v_i \in \mathbb{R}^3. \tag{26}$$

Each voxel v_i is associated with a latent disease score, whose belief is characterized by a Gaussian posterior distribution with mean $\mu_t(v_i)$ and covariance P_t^+. The initial belief is represented by a Gaussian prior characterized by the mean $\boldsymbol{\mu}_0$ and covariance P_0^+, which denote the initial disease estimate and its associated uncertainty, respectively. Using 4D LiDAR-based measurements, individual plants are segmented as geometric entities $\mathcal{P}_j$. For each reconstructed plant, the cropwise NDVI $N_t^{(j)}$ defined in (14) is applied as a plant-level phenotyping observation. Each reconstructed plant $\mathcal{P}_j$ is associated with a subset of voxels:

$$\mathcal{V}_j \subset \mathcal{V}. \tag{27}$$

For each voxel $v_i \in \mathcal{V}_j$, the observation is computed as follows:

$$o_t(v_i) = 1 - N_t^{(j)}. \tag{28}$$

The posterior belief is updated using the following observation model:

$$\boldsymbol{\mu}_{t+1} = \boldsymbol{\mu}_t + K_t\big(o_t - \mathbf{H}_t\boldsymbol{\mu}_t\big), \tag{29}$$

$$P_{t+1}^+ = (I - K_t\mathbf{H}_t)P_t^+, \tag{30}$$

where $\mathbf{H}_t$ denotes the observation model that maps the latent voxel field to the currently observed measurement space. The observation uncertainty is represented by the covariance R_t, and the gain is calculated as

$$K_t = P_t^+\mathbf{H}_t^\top \left(\mathbf{H}_t P_t^+ \mathbf{H}_t^\top + R_t\right)^{-1}. \tag{31}$$

The gain K_t determines the relative contribution of the current observation and the prior belief according to their respective uncertainties.

The voxelwise probability map is calculated as a normalized transformation of the posterior mean:

$$b_t(v_i) = \sigma(\mu_t(v_i)), \tag{32}$$

where $\sigma(\cdot)$ represents a monotonic squashing function mapping the latent disease score to $[0, 1]$. The uncertainty of the belief is quantified using the trace of the covariance, as follows:

$$\mathcal{U}_t = \mathrm{tr}(P_t^+), \tag{33}$$

where a larger trace indicates greater aggregate epistemic uncertainty over the voxelized disease field. The information gain at time step t is computed as follows:

$$\Delta\mathrm{tr}(P) = \mathrm{tr}(P_t^+) - \mathrm{tr}(P_{t+1}^+), \tag{34}$$

which is consistent with the reward in (18). Although $\mathcal{U}_t$ represents aggregate uncertainty over the maintained belief, its stepwise reduction is induced by measurements acquired within the current sensing region. Thus, new information enters the Bayesian belief through spatially localized observations rather than through global observation of the entire field.

Due to the limited sensing range, the voxelwise probability map is updated only within the current sensing region $\mathcal{V}_t \subset \mathcal{V}$:

$$b_{t+1}(v_i) = \begin{cases} \sigma\big(\mu_{t+1}(v_i)\big), & v_i \in \mathcal{V}_t, \\ b_t(v_i), & \text{otherwise.} \end{cases} \tag{35}$$

At mission termination, the disease source is estimated as

$$\hat{\zeta} = \arg\max_{v_i \in \mathcal{V}} b_T(v_i), \tag{36}$$

where $\hat{\zeta}$ denotes the estimated disease source location obtained from the accumulated Bayesian belief. The corresponding ground-truth source location ζ is used only for evaluating localization performance.

V. Experimental Setup

A. Environments

The simulation experiments were conducted in a Unity-based agricultural environment covering a 50×50 m area. The UAV operated at a fixed altitude of 3 m. The environment contained row-structured soybean plants and spatially heterogeneous disease distributions generated using the dispersion model described in Section V-C. The simulated sensing suite modeled a Unitree L2 4D LiDAR and a MAPIR Survey3 RGN camera to provide crop-geometry and spectral observations, respectively.

The simulation evaluation comprised three fixed-hotspot layouts and one random-hotspot condition. The fixed layouts enabled repeatable comparisons under different source arrangements, whereas the random-hotspot condition assessed planner behavior under varying hotspot configurations. Each planning method was evaluated over 20 independent trials per condition.

A separate trajectory-level evaluation was conducted using georeferenced LiDAR and RGN measurements acquired with an F660 UAV platform (Fig. 3) in three representative

Authorized licensed use limited to: CHONNAM NATIONAL UNIVERSITY. Downloaded on October 30,2026 at 07:49:35 UTC from IEEE Xplore. Restrictions apply.

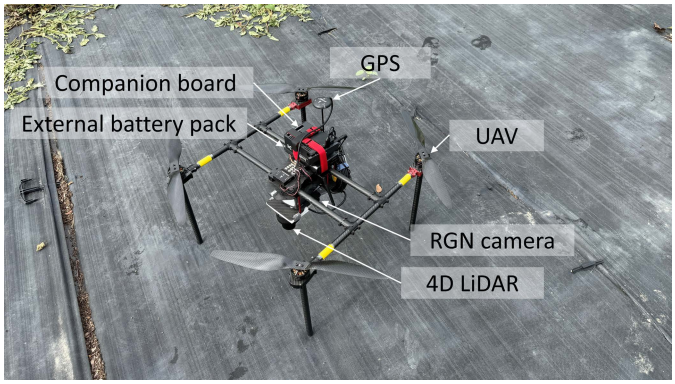


Fig. 3. Hardware setup of the UAV-based informative monitoring system.

soybean-field environments (Fig. 4). Each corresponding field-derived map covered a 50×50 m area, and the measurements were acquired at a nominal altitude of 3 m. The sensor data were collected during field flights, whereas the planners were evaluated offline using the resulting georeferenced maps. At each planning step, only the measurements within the current sensing region were made available to the planner; the complete map was not provided as an observation. Each planner was evaluated over 20 trials on each field-derived sensor map. The full NDVI map was not exposed to the planner and was used only after each trial to define an offline spectral proxy reference and visualize the resulting trajectory. Accordingly, this experiment assesses trajectory-level applicability to measured agricultural data rather than biological validation of disease source localization or physical UAV flight performance.

B. Compared Methods

The comparative evaluation includes CPP, Brouwer et al. [22], Rückin et al. [23], Vashisth et al. [24], and the proposed method. CPP follows a fixed-altitude lawnmower trajectory composed of parallel sweep lines. Brouwer et al. select sensing motions using an exploration-exploitation tradeoff that balances field exploration and source-directed behavior. Rückin et al. use tree search guided by a learned policy and value representation for informative action selection. Vashisth et al. construct local graph representations online and perform learned adaptive replanning. These adapted comparisons are not intended to reproduce the original benchmark results of the respective methods. Instead, they evaluate their defining planning mechanisms under a common downstream disease source localization task. The three literature-based IPP methods were directly implemented in the same Unity environment. Interface-level adaptations were made only to apply the common sensing model, feasible motion space, mission budget, and disease source evaluation procedure, while retaining the defining planning mechanism of each method.

An oracle reference is additionally reported for interpreting motion efficiency. For the single-hotspot condition, the oracle follows the shortest path from the initial UAV position to the ground-truth hotspot center. For the multi-hotspot conditions, the oracle follows the exact minimum-length open traveling-salesperson path through all ground-truth hotspot centers without returning to the initial position. Because the oracle has privileged access to the ground-truth hotspot locations, it is used only as a reference and is not considered an executable planner.

All executable planners were evaluated using the same field configuration, initial UAV pose, sensing footprint, observation data, translational and rotational action limits, mission horizon, and source-estimation procedure. Importantly, ground-truth hotspot locations were not used for online action selection. An evaluation episode was terminated when the predefined online hotspot-detection condition was satisfied using only information available to the planner, or when the maximum mission horizon T was reached. After termination, the estimated source locations were compared offline with the ground-truth hotspot locations to calculate the task-level localization metrics described in Section V-E.

C. Wind-Driven Disease Dispersion Model

To generate realistic disease distributions, this work employs a wind-driven dispersion model based on the Monin-Obukhov similarity theory [28]. The vertical wind profile is modeled as follows:

$$u(z) = \frac{u^*}{\kappa} \ln\left(\frac{z-d}{z_0}\right), \tag{37}$$

where u^* denotes the friction velocity representing the shear stress between the ground and atmosphere, κ represents the von Kármán constant (typically $\approx$ 0.4), z indicates the height above ground level, d measures the zero-plane displacement height, and z_0 denotes the surface roughness length.

Disease propagation is modeled using a source term that captures the interaction between airflow and pathogen transport:

$$S_k = \rho C_d a(z)(\beta_p |u|^3 - \beta_d |u| k), \tag{38}$$

where ρ denotes the air density, C_d is the drag coefficient, $a(z)$ represents the vegetation density, k denotes the turbulent kinetic energy, and β_p and β_d are the turbulence-production and dissipation coefficients, respectively.

D. Model Parameters

The exploration policy was trained using the twin delayed deep deterministic policy gradient (TD3) algorithm. Table II summarizes the critical training hyperparameters. The localization tolerance used to calculate SR was set to $\delta = 5$ m. The weighting parameter λ controls the additional emphasis assigned to local disease density in the motion-normalized reward and was set to $\lambda = 0.3$.

E. Performance Metrics

The framework's performance was evaluated using quantitative metrics derived from the formulations in Sections III and IV. The primary evaluation focuses on localization success, localization error, and motion cost. Uncertainty reduction, coverage, and redundancy are additionally reported as diagnostic measures of exploration behavior.

Authorized licensed use limited to: CHONNAM NATIONAL UNIVERSITY. Downloaded on October 30,2026 at 07:49:35 UTC from IEEE Xplore. Restrictions apply.



(a) (b) (c)

Fig. 4. Representative soybean-field environments in which the georeferenced LiDAR and RGN measurements were acquired: (a) Environment 1, (b) Environment 2, and (c) Environment 3.

TABLE II
TD3 TRAINING AND IMPLEMENTATION PARAMETERS

Parameter	Value
Grid resolution	50×50 cells
Voxel size	1 m
Maximum training episode length	700 steps
Maximum evaluation episode length	1,000 steps
Total training timesteps	500,000
Learning rate	1×10^{-4}
Batch size	256
Replay buffer size	200,000
Action noise	Gaussian ($\mu = 0$, $\sigma = 0.1$)
Discount factor γ	0.99
Actor hidden layers	2
Critic hidden layers	2
Actor hidden-layer units	400, 300
Critic hidden-layer units	400, 300 (each Q-network)
Hidden activation	ReLU
Actor output activation	Tanh
Optimizer	Adam
Target update coefficient τ	0.005
Policy update delay	2
Target policy noise	$\mathcal{N}(0, 0.2^2)$; clipped to $[-0.5, 0.5]$
Disease-belief prior mean $\boldsymbol{\mu}_0$	0.5 per cell
Initial disease-belief covariance $\mathbf{P}_0^+$	0.2 per cell
Motion-cost lower bound ε_c	10^{-3}
Reward clipping range $(r_{\min}, r_{\max})$	$(-5, 5)$

Steps: The number of policy decisions executed before satisfying an online termination condition or reaching the maximum mission horizon.

Uncertainty Reduction (*UR*): The cumulative information acquired during an episode is quantified from the reduction in the posterior covariance trace. This metric represents the aggregate change in the maintained disease belief resulting from sequential, spatially localized observations. A larger value indicates greater overall posterior uncertainty reduction over the observed trajectory.

Coverage Area (*CA*): The extent of exploration was quantified using the ratio of visited spatial units to the total number of units. This metric is reported as the spatial coverage ratio, indicating how broadly each method explores the environment.

Localization Error and Success Rate (*LE*, *SR*): *LE* is averaged over all evaluated trials, including unsuccessful trials, whereas *SR* represents the proportion of trials satisfying the corresponding localization criterion. Disease source localization performance was evaluated using the Euclidean distance between the estimated source $\hat{\zeta}$ and the corresponding reference source ζ. For the single-hotspot condition, a trial was counted as successful when the hotspot-detection condition was satisfied and the estimated source was located within the localization tolerance δ of the reference hotspot. For the multi-hotspot conditions, a trial was counted as successful only when the detection and localization conditions were satisfied for every reference hotspot during the same mission. *SR* is the percentage of evaluated trials satisfying the corresponding success criterion. Each individual simulation condition was evaluated over 20 independent trials.

Redundancy (*RED*): To measure the exploration inefficiency, this study evaluates the ratio of revisited spatial units to the total number of visited units. This variable is reported as the redundancy ratio, capturing unnecessary repeated observations during exploration.

Motion Cost (*MC*): The total traversal effort was measured as the cumulative motion cost, reflecting the overall resource consumption during exploration.

For the field-derived maps, localization-related quantities are reported as proxy metrics. The reference location was determined offline from the minimum-NDVI region of the complete map. *LE* and *SR* therefore measure the agreement between a trajectory-derived estimate and this offline spectral reference; they do not represent biologically verified disease source accuracy.

VI. EXPERIMENTAL RESULTS

The evaluation addresses three questions: (1) whether the proposed method achieves reliable disease source localization under the prescribed mission budget, (2) how much traversal is required to achieve this task relative to the compared planners, and (3) how the resulting uncertainty reduction, coverage, and redundancy patterns differ across planning strategies. The latter quantities are treated as diagnostic measures of exploration behavior rather than as the primary task-success criteria.

A. Sensitivity Analysis

1) Voxel Resolution: Table III summarizes the results obtained using voxel sizes of 1, 2, and 5 m under the same training budget and evaluation protocol. For the proposed method, the 1 m configuration achieved a localization success rate of 95%, a localization error of 4.33±0.48 m, and a motion cost of 112.24±31.47. Increasing the voxel size to 2 m reduced the success rate to 80% and increased the localization error to 7.58 ± 2.12 m. The 5 m configuration also achieved an

Authorized licensed use limited to: CHONNAM NATIONAL UNIVERSITY. Downloaded on October 30,2026 at 07:49:35 UTC from IEEE Xplore. Restrictions apply.

TABLE III
SENSITIVITY TO VOXEL RESOLUTION

Voxel size (m)	Method	*Steps* (↓)	*UR* (%) (↑)	*CA* (%) (↑)	*RED* (%) (↓)	*MC* (m) (↓)	*LE* (m) (↓)	*SR* (%) (↑)
1	Oracle	149.9 ± 21.7	8.50 ± 1.51	16.28 ± 2.59	95.03 ± 0.11	61.02 ± 8.07	0.00 ± 0.00	100
	CPP	640.0 ± 28.4	38.88 ± 10.08	60.99 ± 9.53	95.57 ± 0.31	228.34 ± 34.45	$\mathbf{4.08 \pm 2.58}$	90
	Vashisth et al. [24]	510.4 ± 115.7	26.88 ± 4.13	45.55 ± 7.35	95.27 ± 0.58	191.59 ± 33.22	5.96 ± 1.32	80
	Rückin et al. [23]	999.0 ± 0.0	0.00 ± 0.00	9.49 ± 3.37	99.47 ± 0.08	157.98 ± 102.07	38.13 ± 22.03	0
	Brouwer et al. [22]	343.5 ± 118.1	16.75 ± 2.65	30.79 ± 4.91	94.56 ± 0.59	$\mathbf{134.28 \pm 43.54}$	7.47 ± 1.10	60
	Proposed	$\mathbf{248.9 \pm 76.1}$	6.34 ± 1.25	16.97 ± 1.65	96.19 ± 0.52	$\mathbf{112.24 \pm 31.47}$	$\mathbf{4.33 \pm 0.48}$	**95**
2	Oracle	143.7 ± 15.5	7.60 ± 0.42	14.87 ± 0.99	95.27 ± 0.25	58.82 ± 5.97	0.00 ± 0.00	100
	CPP	560.1 ± 155.8	29.46 ± 11.13	51.51 ± 14.65	95.39 ± 0.03	212.86 ± 57.72	$\mathbf{3.13 \pm 0.38}$	**100**
	Vashisth et al. [24]	656.3 ± 94.3	36.06 ± 3.64	58.10 ± 4.60	95.36 ± 0.28	273.22 ± 38.91	$\mathbf{4.35 \pm 0.36}$	80
	Rückin et al. [23]	999.0 ± 0.0	0.00 ± 0.00	5.11 ± 0.00	99.63 ± 0.00	266.10 ± 3.87	43.27 ± 14.20	0
	Brouwer et al. [22]	$\mathbf{406.6 \pm 143.8}$	18.27 ± 4.04	34.19 ± 7.51	95.30 ± 0.47	$\mathbf{157.82 \pm 52.85}$	5.01 ± 0.60	70
	Proposed	$\mathbf{407.0 \pm 115.1}$	3.62 ± 1.16	16.72 ± 1.17	97.18 ± 0.45	$\mathbf{135.55 \pm 33.68}$	7.58 ± 2.12	**80**
5	Oracle	216.6 ± 110.8	7.94 ± 2.83	15.96 ± 2.42	95.56 ± 0.72	85.90 ± 41.22	0.00 ± 0.00	100
	CPP	575.2 ± 167.6	31.52 ± 13.10	51.81 ± 14.48	95.79 ± 0.24	219.93 ± 63.03	$\mathbf{3.94 \pm 0.69}$	**80**
	Vashisth et al. [24]	736.0 ± 169.8	41.05 ± 8.06	63.91 ± 10.22	95.48 ± 0.32	308.48 ± 70.48	5.84 ± 0.64	50
	Rückin et al. [23]	999.0 ± 0.0	0.00 ± 0.00	7.29 ± 0.22	99.67 ± 0.01	220.52 ± 16.30	39.89 ± 23.77	0
	Brouwer et al. [22]	508.7 ± 247.8	20.13 ± 5.53	38.43 ± 11.57	95.53 ± 0.77	196.45 ± 92.45	$\mathbf{4.55 \pm 0.31}$	60
	Proposed	$\mathbf{418.3 \pm 74.5}$	4.00 ± 2.04	16.79 ± 1.19	97.26 ± 0.64	$\mathbf{112.78 \pm 1.55}$	9.51 ± 3.14	**80**

(↑) and (↓) indicate whether higher or lower values are better, respectively.
Results are aggregated over the three fixed-hotspot scenarios, each evaluated over 20 independent trials, and are reported as the mean ± standard deviation of the three scenario-level means.

TABLE IV
SENSITIVITY TO BAYESIAN PRIOR INITIALIZATION

Prior condition	Prior mean	Prior variance	Successful trials	Localization step	Delay from flat	*MC* at localization
Flat	0.5	0.20	20/20	150.5 [148.8, 152.0]	Reference	69.155 [68.410, 69.485]
Pessimistic	0.1	0.20	20/20	159.5 [159.0, 161.2]	+9.5 [8.0, 13.0]	72.959 [72.830, 73.547]
Optimistic	0.9	0.20	0/20	N/R	N/R	-
Overconfident	0.9	0.02	0/20	N/R	N/R	-

Values for localization step, delay from the flat prior, and *MC* at localization are reported as median [25th percentile, 75th percentile] over 20 paired evaluation seeds.
N/R: localization criterion not reached within the 1,000 step evaluation horizon.

80% success rate but produced a larger localization error of 9.51 ± 3.14 m.

Motion cost did not vary monotonically with voxel size. However, the localization results consistently favored the 1 m representation. Coarser discretization assigns spatially distinct spectral observations to the same cell, reducing the spatial resolution available for source estimation. Based on its higher localization success and lower localization error, the 1 m voxel size was retained for the remaining experiments.

2) Bayesian Prior Initialization: The prior-initialization sensitivity was evaluated using spatially uniform priors. The same initial mean was assigned to every voxel, and the initial covariance was diagonal with an identical per-voxel variance. The localization step denotes the first step at which the localization criterion defined in Section V-E was satisfied. The delay from the flat prior was calculated using paired trials, and *MC* at localization denotes the cumulative motion cost at the corresponding localization step.

The influence of prior initialization was evaluated using four spatially uniform conditions. The flat prior used an initial mean of 0.5 and a per-voxel variance of 0.20. The pessimistic prior used a mean of 0.1 and a variance of 0.20, whereas the optimistic and overconfident priors used a mean of 0.9 with variances of 0.20 and 0.02, respectively. The same trained TD3 policy and 20 paired evaluation seeds were used under all four conditions.

As shown in Table IV, the flat and pessimistic priors satisfied the localization criterion in all 20 trials. The median localization step increased from 150.5 under the flat prior to 159.5 under the pessimistic prior, corresponding to a median delay of 9.5 steps. The associated median motion cost increased from 69.155 to 72.959. In contrast, neither the optimistic nor the overconfident prior satisfied the localization criterion within the 1,000-step horizon. These results indicate that the Bayesian estimate recovered consistently from flat and conservative initialization but remained sensitive to a uniformly optimistic initial mean. The flat prior was therefore retained for the main experiments, which assume that no historical disease map is available.

B. Quantitative Comparison in Simulation

Fig. 5 and Table V compare the oracle reference, CPP, Vashisth et al., Rückin et al., Brouwer et al., and the proposed method. Because the compared planners were originally designed with different information-acquisition objectives, the results are interpreted jointly in terms of localization success, localization error, motion cost, and exploration behavior.

Across the four simulation conditions, the proposed method achieved localization success rates of 100%, 90%, 95%, and 100%, with corresponding localization errors of 4.45, 4.73, 3.80, and 4.28 m. It was the only executable

Authorized licensed use limited to: CHONNAM NATIONAL UNIVERSITY. Downloaded on October 30,2026 at 07:49:35 UTC from IEEE Xplore. Restrictions apply.

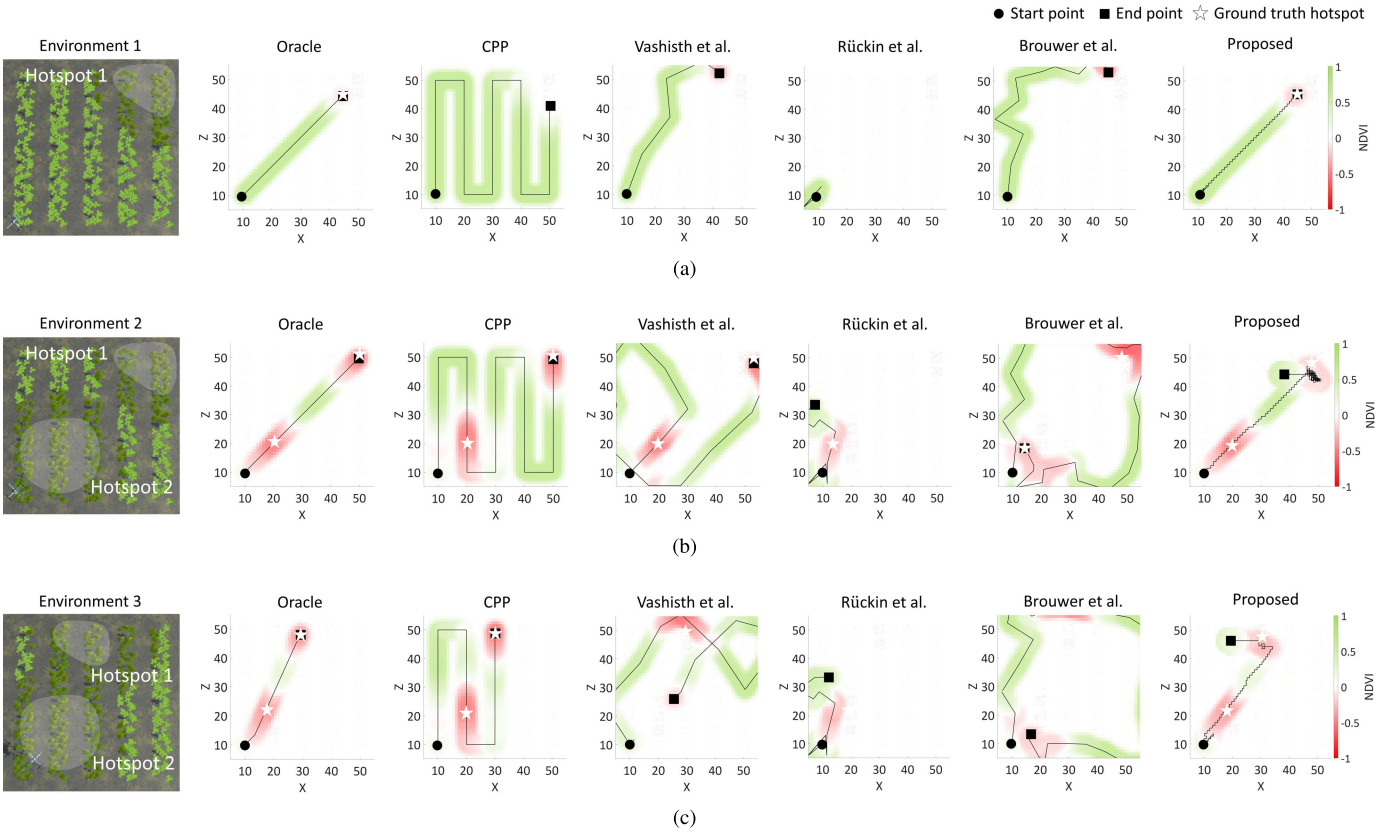


Fig. 5. Qualitative comparison of exploration trajectories in the three fixed-hotspot simulation layouts: (a) Environment 1, (b) Environment 2, and (c) Environment 3. The first column shows the simulated disease distribution, and the remaining columns show the oracle reference, CPP, Vashisth et al., Rückin et al., Brouwer et al., and the proposed method. The black circle and square denote the initial and terminal UAV positions, respectively; the white stars denote the ground-truth hotspot centers. The solid line represents the UAV trajectory, and the background color indicates the NDVI distribution.

TABLE V

SIMULATION PERFORMANCE UNDER FIXED- AND RANDOM-HOTSPOT CONDITIONS

Scenario	Method	*Steps* (↓)	*UR* (%) (↑)	*CA* (%) (↑)	*RED* (%) (↓)	*MC* (m) (↓)	*LE* (m) (↓)	*SR* (%) (↑)
Fixed hotspot 1	Oracle	150.4 ± 4.9	8.68 ± 0.22	16.55 ± 0.11	94.95 ± 0.18	61.16 ± 3.64	0.00 ± 0.00	100
	CPP	644.8 ± 4.9	$\mathbf{43.71 \pm 0.15}$	$\mathbf{65.39 \pm 0.17}$	95.40 ± 0.03	244.38 ± 3.47	4.73 ± 0.05	**100**
	Vashisth et al. [24]	251.7 ± 142.1	16.12 ± 7.25	27.80 ± 11.75	94.49 ± 0.46	106.76 ± 60.63	5.01 ± 1.17	85
	Rückin et al. [23]	999.0 ± 0.0	0.00 ± 0.00	5.92 ± 0.23	99.56 ± 0.04	54.72 ± 9.66	60.89 ± 2.06	0
	Brouwer et al. [22]	240.6 ± 15.3	13.80 ± 1.68	25.67 ± 2.02	94.85 ± 0.22	94.93 ± 6.35	6.36 ± 1.73	40
	Proposed	$\mathbf{171.2 \pm 8.0}$	8.43 ± 0.12	16.46 ± 0.26	95.59 ± 0.18	$\mathbf{84.43 \pm 5.52}$	$\mathbf{4.45 \pm 0.48}$	**100**
Fixed hotspot 2	Oracle	171.3 ± 5.9	9.92 ± 0.35	18.72 ± 0.51	94.99 ± 0.13	69.02 ± 3.37	0.00 ± 0.00	100
	CPP	665.7 ± 8.7	$\mathbf{45.64 \pm 0.60}$	$\mathbf{67.53 \pm 0.66}$	95.39 ± 0.03	251.84 ± 4.29	$\mathbf{1.24 \pm 0.54}$	**100**
	Vashisth et al. [24]	538.6 ± 313.5	30.78 ± 14.49	49.51 ± 21.09	95.05 ± 0.80	226.13 ± 129.33	7.10 ± 7.14	80
	Rückin et al. [23]	999.0 ± 0.0	0.00 ± 0.00	9.93 ± 2.80	99.43 ± 0.14	160.40 ± 116.72	36.58 ± 12.01	0
	Brouwer et al. [22]	$\mathbf{317.6 \pm 176.5}$	17.55 ± 7.82	31.24 ± 12.40	$\mathbf{94.94 \pm 0.54}$	$\mathbf{126.86 \pm 68.97}$	7.49 ± 11.90	**90**
	Proposed	$\mathbf{323.2 \pm 258.8}$	9.33 ± 0.69	18.82 ± 1.90	96.46 ± 1.08	$\mathbf{146.40 \pm 105.82}$	$\mathbf{4.73 \pm 5.68}$	**90**
Fixed hotspot 3	Oracle	128.0 ± 4.9	6.91 ± 0.30	13.56 ± 0.38	95.15 ± 0.17	52.88 ± 3.16	0.00 ± 0.00	100
	CPP	609.5 ± 293.4	$\mathbf{27.29 \pm 5.97}$	$\mathbf{50.06 \pm 13.35}$	95.93 ± 0.75	188.79 ± 48.83	6.27 ± 6.78	65
	Vashisth et al. [24]	383.2 ± 288.4	22.56 ± 14.60	37.07 ± 21.61	94.83 ± 0.71	159.86 ± 119.41	$\mathbf{4.52 \pm 2.78}$	**95**
	Rückin et al. [23]	999.0 ± 0.0	0.00 ± 0.00	12.62 ± 2.24	99.41 ± 0.11	258.81 ± 75.65	16.91 ± 1.87	0
	Brouwer et al. [22]	472.4 ± 397.2	18.91 ± 10.50	35.46 ± 20.79	$\mathbf{93.88 \pm 5.86}$	181.06 ± 144.58	8.55 ± 10.71	60
	Proposed	$\mathbf{252.2 \pm 193.2}$	7.20 ± 0.31	15.64 ± 1.36	96.51 ± 0.96	$\mathbf{105.90 \pm 62.39}$	$\mathbf{3.80 \pm 2.69}$	**95**
Random hotspot	Oracle	80.4 ± 42.5	4.38 ± 1.88	8.74 ± 3.84	85.22 ± 29.21	34.40 ± 17.75	0.00 ± 0.00	100
	CPP	280.3 ± 208.6	$\mathbf{16.75 \pm 13.35}$	28.67 ± 20.61	$\mathbf{85.59 \pm 29.29}$	108.90 ± 78.88	$\mathbf{4.64 \pm 0.80}$	85
	Vashisth et al. [24]	662.7 ± 397.9	8.76 ± 5.62	$\mathbf{29.15 \pm 13.82}$	87.70 ± 30.01	142.76 ± 72.94	21.01 ± 21.21	40
	Rückin et al. [23]	849.6 ± 364.9	0.23 ± 0.58	5.23 ± 1.21	88.74 ± 30.62	$\mathbf{50.72 \pm 22.37}$	35.13 ± 16.47	10
	Brouwer et al. [22]	557.7 ± 439.1	3.27 ± 2.44	19.85 ± 11.73	87.00 ± 29.90	96.52 ± 65.82	19.21 ± 19.44	50
	Proposed	$\mathbf{105.0 \pm 58.0}$	4.75 ± 2.06	9.62 ± 4.45	$\mathbf{85.84 \pm 29.43}$	$\mathbf{48.84 \pm 26.60}$	$\mathbf{4.28 \pm 0.54}$	**100**

(↑) and (↓) indicate whether higher or lower values are better, respectively.
Values are reported as mean ± standard deviation over 20 independent trials.

planner that maintained at least 90% localization success in every condition. Averaged across the four conditions, the proposed method achieved a localization success rate of 96.25%, compared with 87.5% for CPP, 75.0% for Vashisth et al., 60.0% for Brouwer et al., and 2.5% for Rückin et al.

The proposed method also achieved the lowest average motion cost among the executable planners. Its motion costs

Authorized licensed use limited to: CHONNAM NATIONAL UNIVERSITY. Downloaded on October 30,2026 at 07:49:35 UTC from IEEE Xplore. Restrictions apply.

TABLE VI
ABLATION STUDY ON CRUCIAL COMPONENTS OF THE PROPOSED FRAMEWORK

Scenario	Method	Steps (↓)	UR (%) (↑)	CA (%) (↑)	RED (%) (↓)	MC (m) (↓)	LE (m) (↓)	SR (%) (↑)
Fixed hotspot 1	w/o Bayesian	859.85 ± 286.88	0.00 ± 0.00	$\mathbf{19.83 \pm 1.98}$	98.23 ± 0.85	214.70 ± 77.44	70.00 ± 0.00	0
	w/o Domain	999.00 ± 0.00	4.34 ± 0.38	11.91 ± 0.61	99.45 ± 0.03	84.92 ± 15.41	70.00 ± 0.00	0
	w/o Reward	999.00 ± 0.00	2.02 ± 0.30	4.25 ± 0.44	99.37 ± 0.46	106.67 ± 110.56	68.95 ± 3.24	0
	Proposed	$\mathbf{171.20 \pm 8.02}$	$\mathbf{8.43 \pm 0.12}$	16.46 ± 0.26	$\mathbf{95.59 \pm 0.18}$	$\mathbf{84.43 \pm 5.52}$	$\mathbf{4.45 \pm 0.48}$	**100**
Fixed hotspot 2	w/o Bayesian	999.00 ± 0.00	0.00 ± 0.00	$\mathbf{20.92 \pm 5.07}$	98.04 ± 0.47	240.17 ± 62.92	48.79 ± 0.00	0
	w/o Domain	999.00 ± 0.00	4.15 ± 0.43	11.70 ± 0.67	99.46 ± 0.03	$\mathbf{75.87 \pm 13.50}$	48.79 ± 0.00	0
	w/o Reward	999.00 ± 0.00	2.54 ± 0.37	5.47 ± 1.29	99.70 ± 0.11	85.45 ± 45.04	27.80 ± 2.11	0
	Proposed	$\mathbf{323.15 \pm 258.84}$	$\mathbf{9.33 \pm 0.69}$	18.82 ± 1.90	$\mathbf{96.46 \pm 1.08}$	146.40 ± 105.82	$\mathbf{4.73 \pm 5.68}$	**90**
Fixed hotspot 3	w/o Bayesian	999.00 ± 0.00	0.00 ± 0.00	$\mathbf{18.14 \pm 2.04}$	98.13 ± 0.56	198.88 ± 65.96	42.60 ± 0.00	0
	w/o Domain	999.00 ± 0.00	4.26 ± 0.44	11.85 ± 0.66	99.45 ± 0.03	$\mathbf{80.31 \pm 13.76}$	42.60 ± 0.00	0
	w/o Reward	999.00 ± 0.00	3.36 ± 2.17	6.44 ± 2.23	99.67 ± 0.12	99.38 ± 66.66	20.78 ± 3.05	0
	Proposed	$\mathbf{252.2 \pm 193.2}$	$\mathbf{7.20 \pm 0.31}$	15.64 ± 1.36	$\mathbf{96.51 \pm 0.96}$	105.90 ± 62.39	$\mathbf{3.80 \pm 2.69}$	**95**
Random hotspot	w/o Bayesian	341.50 ± 348.21	0.00 ± 0.00	$\mathbf{11.89 \pm 5.27}$	97.51 ± 1.18	94.13 ± 73.03	42.18 ± 15.53	0
	w/o Domain	999.00 ± 0.00	4.14 ± 0.44	11.59 ± 0.63	99.47 ± 0.03	77.25 ± 14.88	42.18 ± 15.53	0
	w/o Reward	453.10 ± 423.65	3.07 ± 1.65	6.65 ± 3.28	88.45 ± 30.28	57.58 ± 34.09	22.12 ± 25.23	65
	Proposed	$\mathbf{105.00 \pm 58.04}$	$\mathbf{4.75 \pm 2.06}$	9.62 ± 4.45	$\mathbf{85.84 \pm 29.43}$	$\mathbf{48.84 \pm 26.60}$	$\mathbf{4.28 \pm 0.54}$	**100**

(↑) and (↓) indicate whether higher or lower values are better, respectively.

of 84.43, 146.40, 105.90, and 48.84 represent reductions of 65.5%, 41.9%, 43.9%, and 55.2%, respectively, relative to CPP. Compared with Vashisth et al., the proposed method matched or exceeded its localization success in every condition while reducing motion cost by 20.9-65.8%. The corresponding oracle-normalized motion-cost ratios were 1.38, 2.12, 2.00, and 1.42, indicating the additional traversal required to infer source locations that were unknown to the executable planners.

CPP achieved the largest uncertainty reduction and coverage because it systematically observed a broad portion of the field. It achieved 100% localization success in the first two fixed layouts but required motion costs of 244.38 and 251.84. Its success rate decreased to 65% in the third fixed layout and 85% in the random-hotspot condition. Vashisth et al. achieved success rates of 40-95% but generally required longer and more variable trajectories than the proposed method. Brouwer et al. achieved 90% success in the second fixed layout, where its mean motion cost was lower than that of the proposed method, but its success rates were limited to 40-60% in the remaining conditions. Rückin et al. frequently reached the maximum mission horizon without successful localization. Its relatively low motion cost in some conditions therefore reflects limited trajectory progression rather than efficient task completion.

The trajectories in Fig. 5 are consistent with the quantitative results. CPP follows systematic sweep paths, whereas Brouwer et al. and Vashisth et al. produce longer or more variable trajectories depending on the source arrangement. Rückin et al. remain concentrated near the initial region in several trials. The low localization success of Rückin et al. in the present evaluation should not be interpreted as a contradiction of its original informative-mapping performance. The method was originally designed for a different information-acquisition objective, and under the common disease source localization interface its policy-value search did not consistently produce source-directed progression in the evaluated layouts. The proposed method generates comparatively compact trajectories directed toward disease-relevant regions while maintaining consistent localization performance.

C. Ablation Study

Table VI evaluates the effects of removing the Bayesian belief component, domain-aware features, and the proposed information-motion reward. The reported terminal localization errors for unsuccessful trials indicate the final estimate at mission termination and should be interpreted together with SR.

Removing the Bayesian component resulted in zero localization success in all four conditions. In this variant, posterior belief updates are disabled, so the terminal source estimate is no longer informed by the accumulated Bayesian belief used in the complete framework. Consequently, the zero-success result is expected under the current task definition and should not be interpreted as isolated evidence that posterior uncertainty directly improves the actor's online path-selection decisions.

Removing the domain-aware features also resulted in zero success across all conditions. These variants generally reached the maximum mission horizon while exploring only a limited portion of the environment. The result indicates that row alignment, local crop geometry, and disease-related spectral cues provide information that cannot be replaced by the UAV pose alone.

Removing the information-motion reward resulted in zero success in all three fixed-hotspot layouts and 65% success in the random-hotspot condition. Thus, partial exploration can emerge without the complete reward in some randomized trials, but it does not provide consistent localization across the evaluated source arrangements. In comparison, the full framework achieved 90-100% success in all conditions. These results are therefore interpreted as system-level component-removal analyses rather than as isolated causal assessments of each component.

D. Quantitative Comparison in Field Evaluation

The planning methods were additionally evaluated using georeferenced LiDAR and RGN measurements acquired in three agricultural environments. At each decision step, a planner accessed only the measurements within its current sensing region. The complete NDVI map was used only after

Authorized licensed use limited to: CHONNAM NATIONAL UNIVERSITY. Downloaded on October 30,2026 at 07:49:35 UTC from IEEE Xplore. Restrictions apply.

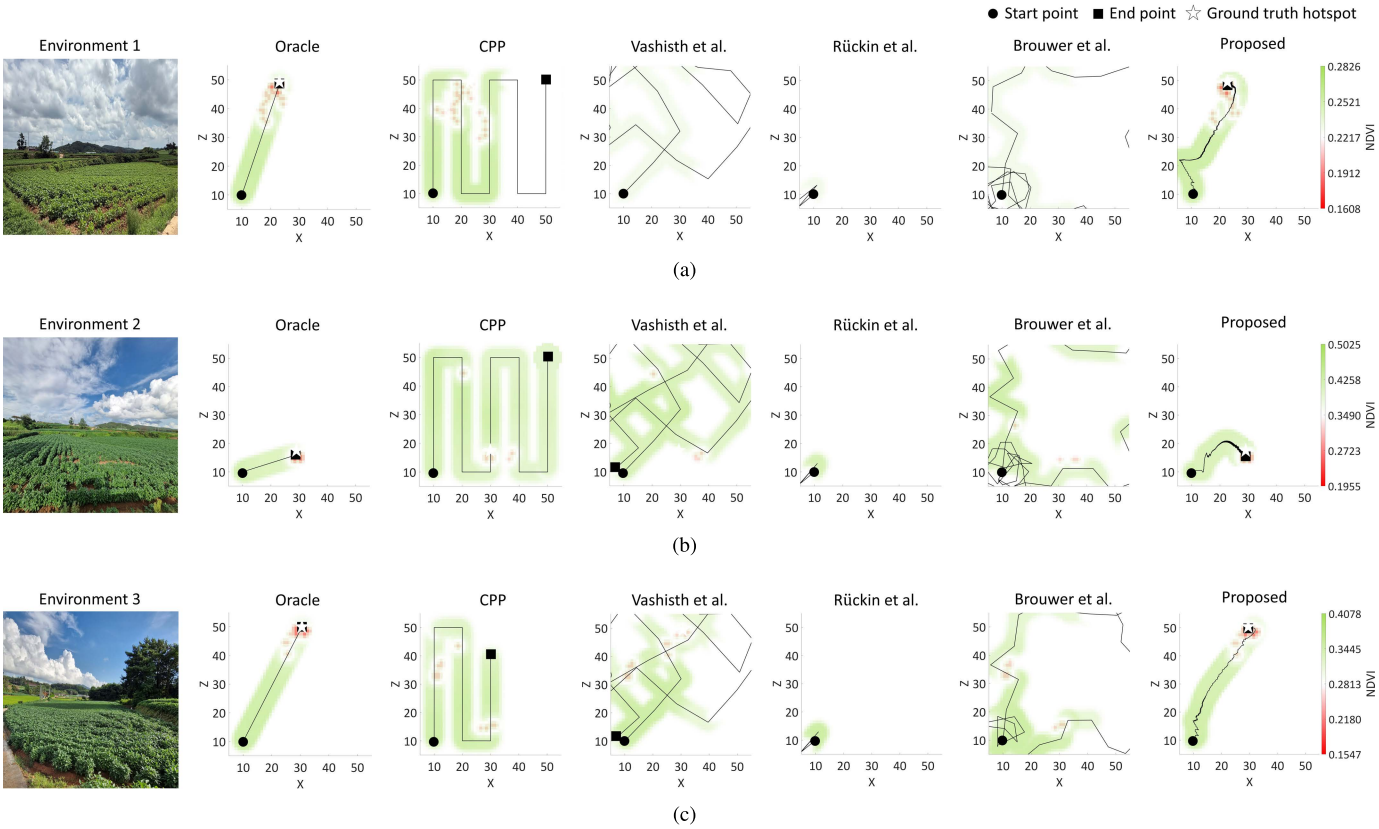


Fig. 6. Trajectory-level comparison on three field-derived agricultural sensor maps: (a) Environment 1, (b) Environment 2, and (c) Environment 3. The first column shows the corresponding field image, and the remaining columns show the privileged reference path, CPP, Vashisth et al., Rückin et al., Brouwer et al., and the proposed method. The black circle and square denote the initial and terminal UAV positions, respectively. The white star denotes the offline spectral proxy reference obtained from the complete NDVI map and does not represent biologically verified disease ground truth. The background color indicates the field-derived NDVI distribution.

TABLE VII

TRAJECTORY-LEVEL EVALUATION ON FIELD-DERIVED SENSOR MAPS

Environment	Method	Steps (↓)	CA (%) (↑)	RED (%) (↓)	MC (m) (↓)	LE (m) (↓)	SR (%) (↑)
1	Oracle	443.6 ± 5.4	13.61 ± 0.00	98.61 ± 0.02	45.4 ± 1.0	0.003 ± 0.000	100
	CPP	999.0 ± 0.0	67.78 ± 0.00	96.93 ± 0.00	240.3 ± 0.0	3.301 ± 0.000	0
	Vashisth et al. [24]	955.2 ± 196.1	73.85 ± 4.47	96.25 ± 0.20	384.9 ± 8.9	3.132 ± 1.666	10
	Rückin et al. [23]	999.0 ± 0.0	5.72 ± 0.00	99.53 ± 0.00	46.1 ± 3.4	37.608 ± 0.000	0
	Brouwer et al. [22]	999.0 ± 0.0	57.90 ± 2.77	97.14 ± 0.12	415.4 ± 4.8	6.467 ± 1.466	0
	Proposed	$\mathbf{939.0 \pm 0.0}$	16.51 ± 0.26	99.24 ± 0.01	$\mathbf{204.5 \pm 2.1}$	$\mathbf{0.777 \pm 0.018}$	**80**
2	Oracle	226.7 ± 6.6	7.61 ± 0.00	98.49 ± 0.05	24.9 ± 1.3	0.009 ± 0.000	100
	CPP	999.0 ± 0.0	67.78 ± 0.00	96.93 ± 0.00	240.3 ± 0.0	1.390 ± 0.000	0
	Vashisth et al. [24]	938.8 ± 185.5	77.80 ± 3.89	96.08 ± 0.24	402.2 ± 21.3	3.549 ± 2.431	20
	Rückin et al. [23]	999.0 ± 0.0	5.88 ± 0.22	99.52 ± 0.01	46.6 ± 2.9	17.711 ± 0.088	0
	Brouwer et al. [22]	891.3 ± 221.2	58.75 ± 2.12	97.10 ± 0.08	411.6 ± 9.1	3.718 ± 3.449	30
	Proposed	$\mathbf{392.9 \pm 145.5}$	8.70 ± 0.36	98.93 ± 0.17	$\mathbf{82.1 \pm 63.3}$	$\mathbf{0.522 \pm 0.186}$	**75**
3	Oracle	479.1 ± 4.8	14.67 ± 0.00	98.61 ± 0.01	48.2 ± 0.9	0.019 ± 0.000	100
	CPP	$\mathbf{380.0 \pm 0.0}$	40.61 ± 0.00	95.15 ± 0.00	$\mathbf{140.3 \pm 0.0}$	$\mathbf{0.225 \pm 0.000}$	**100**
	Vashisth et al. [24]	919.9 ± 243.6	71.84 ± 4.12	96.38 ± 0.22	370.8 ± 14.1	3.300 ± 2.469	15
	Rückin et al. [23]	999.0 ± 0.0	5.90 ± 0.26	99.52 ± 0.02	46.3 ± 4.8	41.695 ± 0.162	0
	Brouwer et al. [22]	999.0 ± 0.0	58.52 ± 2.02	97.09 ± 0.08	338.2 ± 10.1	4.121 ± 2.142	10
	Proposed	803.4 ± 14.0	16.40 ± 0.10	99.06 ± 0.01	$\mathbf{153.9 \pm 17.9}$	$\mathbf{0.470 \pm 0.012}$	**80**

(↑) and (↓) indicate whether higher or lower values are better, respectively.

each trial to define an offline spectral reference and to visualize the resulting trajectory.

As reported in Fig. 6 and Table VII, the proposed method achieved reference-based success rates of 80%, 75%, and 80% in Environments 1-3, respectively. The corresponding distances to the offline spectral reference were 0.777, 0.522, and 0.470 m. In Environments 1 and 2, the proposed method reduced motion cost relative to CPP by 14.9% and 65.8%, respectively. In Environment 3, CPP achieved a lower motion cost and a 100% reference-based success rate, whereas the proposed method achieved 80% success with a motion cost approximately 9.7% higher than that of CPP.

Authorized licensed use limited to: CHONNAM NATIONAL UNIVERSITY. Downloaded on October 30,2026 at 07:49:35 UTC from IEEE Xplore. Restrictions apply.

The adaptive IPP baselines exhibited greater trajectory variability and lower reference-based success across the three field-derived maps. Vashisth et al. achieved reference-based success rates of 10%, 20%, and 15%, whereas Brouwer et al. achieved 0%, 30%, and 10%. Rückin et al. did not satisfy the reference-based success criterion in any of the three environments. The proposed method therefore provided the most consistent agreement with the offline spectral reference among the adaptive planners, although it did not outperform CPP in every individual environment.

These results show that the state construction and action-selection procedures can generate task-directed trajectories when simulated observations are replaced by field-derived measurements. Because the reference was obtained from the complete NDVI map rather than from independently verified disease ground truth, the reported distances and success rates evaluate trajectory-level agreement with a spectral reference. They do not establish biological disease source localization accuracy, physical-flight robustness, or onboard computational performance.

VII. Discussion

A. Exploration Performance

The results distinguish broad field reconstruction from task-oriented disease source localization. CPP achieved the largest aggregate uncertainty reduction because it systematically observed a large portion of the field. The proposed method did not maximize global belief reduction; instead, it maintained localization success rates of 90-100% while reducing motion cost by 41.9-65.5% relative to CPP. Its principal advantage is therefore the reduction in traversal required for reliable source localization rather than uniformly superior reconstruction of the complete disease field.

Among the adaptive planners, the proposed method achieved the highest average localization success and the lowest average motion cost. It matched or exceeded the success rate of Vashisth et al. in every condition and reduced the corresponding motion cost by 20.9-65.8%. Nevertheless, it did not dominate every individual metric: Brouwer et al. required less motion in the second fixed layout, and CPP achieved a lower localization error in some conditions. The oracle-normalized ratios of 1.38-2.12 indicate the additional traversal required when the source locations are initially unknown. The proposed method also did not consistently minimize redundancy, suggesting that repeated local observations may contribute to belief refinement around disease-relevant regions but can also introduce unnecessary traversal. Revisit-aware regularization or short-horizon path refinement may reduce this behavior.

B. Model Simplification

The voxel-resolution analysis showed that the 1 m representation provided the highest localization success and the lowest localization error for the proposed method. The 2 and 5 m configurations both achieved 80% success, but their localization errors increased to 7.58 and 9.51 m, respectively. The prior-initialization analysis further showed that the flat and pessimistic priors satisfied the localization criterion in all 20 trials, although the pessimistic prior introduced a median delay of 9.5 steps. In contrast, neither the optimistic nor the overconfident prior satisfied the criterion within the evaluation horizon. These results identify spatial discretization and uniformly optimistic initialization as important factors affecting source estimation.

The current framework uses a linear-Gaussian belief and a stepwise reward to obtain a computationally tractable approximation of trajectory-level planning. The Gaussian representation may not fully capture multimodal disease distributions, and the linear observation model may not describe nonlinear effects caused by canopy occlusion, illumination variation, or nonuniform reflectance. In addition, maximizing the discounted cumulative reward encourages motion-efficient sensing but does not guarantee global trajectory optimality. More expressive belief models, prior tempering, adaptive covariance inflation, and local trajectory refinement may be investigated to address these limitations.

C. Field Deployment Considerations

The field-derived sensor-map experiments replaced simulated observations with georeferenced LiDAR and RGN measurements. The proposed method achieved reference-based success rates of 80%, 75%, and 80%, with final estimates located 0.777, 0.522, and 0.470 m from the corresponding offline spectral references. It reduced motion cost relative to CPP in two environments, whereas CPP achieved lower motion cost and higher reference-based success in the third environment. These results show that the proposed state representation and action-selection procedure can generate task-directed trajectories from measured agricultural data, but they do not establish superiority under every field configuration.

Several challenges must be addressed for field deployment. The simulated environment assumes a controlled disease distribution, whereas real agricultural fields display more complex, stochastic spatial dynamics influenced by environmental factors (e.g., wind, humidity, and plant growth conditions). Physical deployment must additionally account for UAV localization error, vehicle dynamics, time-varying illumination, sensor synchronization, wind disturbances, canopy occlusion, and onboard computation time. Future work should therefore evaluate the complete sensing, belief-update, and planning pipeline on a physical UAV using calibrated sensors and biologically confirmed reference locations. Subsequent extensions may consider cross-field policy adaptation, robust prior handling, and resource-aware cooperative UAV-UGV monitoring.

VIII. Conclusion

This paper presented a task-oriented IPP framework for motion-efficient UAV-based disease source localization. The framework combines a voxelwise Bayesian disease belief, crop-structural and disease-related state features, and a continuous TD3 policy trained using posterior uncertainty reduction normalized by translational motion cost. In this formulation, posterior uncertainty reduction serves as the trajectory-level sensing criterion, whereas disease source localization is the mission-level objective inferred from the accumulated belief.

Authorized licensed use limited to: CHONNAM NATIONAL UNIVERSITY. Downloaded on October 30,2026 at 07:49:35 UTC from IEEE Xplore. Restrictions apply.

Across three fixed-hotspot layouts and one random-hotspot condition, the proposed method achieved localization success rates of 90-100% and mean localization errors of 3.80-4.73 m. Relative to CPP, it reduced motion cost by 41.9-65.5%. Compared with Vashisth et al., which achieved the highest average localization success among the adaptive baselines, the proposed method matched or exceeded its success rate in every condition while reducing motion cost by 20.9-65.8%. Its motion cost was 1.38-2.12 times that of the corresponding oracle reference, indicating the additional traversal required to infer initially unknown source locations.

The 1 m voxel representation achieved the highest localization success and the lowest localization error among the evaluated resolutions. The flat and pessimistic priors both achieved successful localization in all 20 trials, although the pessimistic prior introduced a median delay of 9.5 steps. In contrast, the optimistic and overconfident priors did not satisfy the localization criterion within the 1,000-step horizon, identifying a limitation under globally optimistic initialization. The ablation results further characterized the roles of the Bayesian belief, domain-aware state representation, and information-motion reward within the complete localization pipeline.

The evaluation on three field-derived maps constructed from LiDAR and RGN measurements further examined the applicability of the planning framework to measured agricultural data. Because the field reference was obtained from the complete NDVI map rather than from independently confirmed disease ground truth, this evaluation represents trajectory-level applicability rather than biological disease validation or physical-flight verification. Future work will focus on physical UAV experiments with calibrated sensing, biologically confirmed references, and measured onboard computation time. Meta-learning-based policy adaptation [29], robust prior-handling mechanisms, and resource-aware cooperative UAV-UGV monitoring [30], [31] will also be investigated.

ACKNOWLEDGMENT

The authors declare that generative AI tools were used only for language editing and not for scientific content generation.

REFERENCES

[1] J. Seol, J. Pak, and H. Il Son, "Development of tributary mapping system using multiple unmanned aerial vehicles with LiDAR," *IEEE Access*, vol. 13, pp. 4081–4097, 2025.

[2] C. Ju and H. I. Son, "Multiple UAV systems for agricultural applications: Control, implementation, and evaluation," *Electronics*, vol. 7, no. 9, p. 162, Aug. 2018.

[3] J. Pak and H. I. Son, "3D LiDAR-based semantic SLAM for intelligent irrigation using UAV," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 18, pp. 7495–7508, 2025.

[4] X. Bai, C. Li, B. Zhang, Z. Wu, and S. S. Ge, "Efficient performance impact algorithms for multirobot task assignment with deadlines," *IEEE Trans. Ind. Electron.*, vol. 71, no. 11, pp. 14373–14382, Nov. 2024.

[5] X. Bai, B. Li, J. Li, Z. Wu, W. Zhang, and S. S. Ge, "Efficient routing for multitruck multidrone package delivery with precedence constraints," *IEEE Trans. Robot.*, vol. 41, pp. 6514–6528, 2025.

[6] Y. Jo and H. I. Son, "Field evaluation of a prioritized path-planning algorithm for heterogeneous agricultural tasks of multi-UGVs," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2024, pp. 11891–11897.

[7] C. Kim, Y. Jo, and H. Il Son, "A Voronoi-based isotemporal task allocation system for autonomous tractor fleet," *IEEE Access*, vol. 13, pp. 166012–166026, 2025.

[8] C. Costa, U. Schurr, F. Loreto, P. Menesatti, and S. Carpentier, "Plant phenotyping research trends, a science mapping approach," *Frontiers Plant Sci.*, vol. 9, p. 1933, Jan. 2019.

[9] J. Pak, B. Kim, C. Ju, and H. I. Son, "Unmanned aerial vehicle-based autonomous tracking system for invasive flying insects," *Comput. Electron. Agricult.*, vol. 227, Dec. 2024, Art. no. 109616.

[10] A. Upadhyay et al., "Deep learning and computer vision in plant disease detection: A comprehensive review of techniques, models, and trends in precision agriculture," *Artif. Intell. Rev.*, vol. 58, no. 3, p. 92, Jan. 2025.

[11] M. Popović, J. Ott, J. Rückin, and M. J. Kochenderfer, "Learning-based methods for adaptive informative path planning," *Robot. Auto. Syst.*, vol. 179, Sep. 2024, Art. no. 104727.

[12] A. Singh, A. Krause, C. Guestrin, and W. J. Kaiser, "Efficient informative sensing using multiple robots," *J. Artif. Intell. Res.*, vol. 34, pp. 707–755, Apr. 2009.

[13] R. Marchant and F. Ramos, "Bayesian optimisation for informative continuous path planning," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2014, pp. 6136–6143.

[14] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies," *J. Mach. Learn. Res.*, vol. 9, no. 2, pp. 235–284, 2008.

[15] S. Arora and S. Scherer, "Randomized algorithm for informative path planning with budget constraints," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2017, pp. 4997–5004.

[16] J. Rückin, F. Magistri, C. Stachniss, and M. Popović, "An informative path planning framework for active learning in UAV-based semantic mapping," *IEEE Trans. Robot.*, vol. 39, no. 6, pp. 4279–4296, Dec. 2023.

[17] S. Dutta, N. Wilde, and S. L. Smith, "Informative path planning for active regression with Gaussian processes via sparse optimization," *IEEE Trans. Robot.*, vol. 41, pp. 2184–2199, 2025.

[18] Y. Chen et al., "UAV-aided efficient informative path planning for autonomous 3D spectrum mapping," *IEEE Trans. Cognit. Commun. Netw.*, vol. 12, pp. 1664–1677, 2026.

[19] R. Khodayi-Mehr, W. Aquino, and M. M. Zavlanos, "Model-based active source identification in complex environments," *IEEE Trans. Robot.*, vol. 35, no. 3, pp. 633–652, Jun. 2019.

[20] T. Zhang, V. Qin, Y. Tang, and N. Li, "Distributed information-based source seeking," *IEEE Trans. Robot.*, vol. 39, no. 6, pp. 4749–4767, Dec. 2023.

[21] G. Flaspohler, V. Preston, A. P. M. Michel, Y. Girdhar, and N. Roy, "Information-guided robotic maximum seek-and-sample in partially observable continuous environments," *IEEE Robot. Autom. Lett.*, vol. 4, no. 4, pp. 3782–3789, Oct. 2019.

[22] Y. Brouwer, A. Vale, and R. Ventura, "Informative path planner with exploration–exploitation trade-off for radiological surveys in non-convex scenarios," *Robot. Auto. Syst.*, vol. 136, Feb. 2021, Art. no. 103691.

[23] J. Rückin, L. Jin, and M. Popović, "Adaptive informative path planning using deep reinforcement learning for UAV-based active sensing," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2022, pp. 4473–4479.

[24] A. Vashisth, J. Rückin, F. Magistri, C. Stachniss, and M. Popović, "Deep reinforcement learning with dynamic graphs for adaptive informative path planning," *IEEE Robot. Autom. Lett.*, vol. 9, no. 9, pp. 7747–7754, Sep. 2024.

[25] J. Rückin, D. Morilla-Cabello, C. Stachniss, E. Montijano, and M. Popović, "Towards map-agnostic policies for adaptive informative path planning," *IEEE Robot. Autom. Lett.*, vol. 10, no. 5, pp. 5114–5121, May 2025.

[26] B. Moon et al., "IA-TIGRIS: An incremental and adaptive sampling-based planner for online informative path planning," *IEEE Trans. Robot.*, vol. 42, pp. 1695–1713, 2026.

[27] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, "Deterministic policy gradient algorithms," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 387–395.

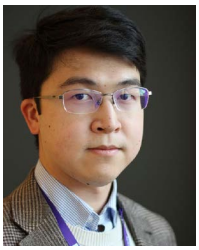
[28] M. Z. Jacobson, *Fundamentals of Atmospheric Modeling*. Cambridge, U.K.: Cambridge Univ. Press, 1999.

Authorized licensed use limited to: CHONNAM NATIONAL UNIVERSITY. Downloaded on October 30,2026 at 07:49:35 UTC from IEEE Xplore. Restrictions apply.

[29] M. L. Betalo, A. Mohamed, A. Sharafian, Z. Wu, J. Li, and X. Bai, "Meta-learning-enhanced task assignment and resource scheduling for UAV-assisted WSNs in 6G-enabled ITS," *IEEE Trans. Mobile Comput.*, vol. 25, no. 10, pp. 1–18, Oct. 2026.
[30] S. Ma, J. Ruan, Y. Du, R. Bucknall, and Y. Liu, "An end-to-end deep reinforcement learning based modular task allocation framework for autonomous mobile systems," *IEEE Trans. Autom. Sci. Eng.*, vol. 22, pp. 1519–1533, 2025.
[31] X. Bai, B. Li, I. Ullah, Z. Wu, S. Basheer, and A. K. Bashir, "Energy-efficient routing for IoT-enabled multi-truck multi-drone pickup and delivery systems," *Appl. Energy*, vol. 400, Dec. 2025, Art. no. 126546.



Jeonghyeon Pak received the B.S. degree from the Department of Rural and Biosystems Engineering, Chonnam National University, Republic of Korea, in 2022, and the M.S. degree from the Department of Convergence Biosystems Engineering, Chonnam National University, in 2024, where she is currently pursuing the Ph.D. degree. Her research interests include field robotics, multi-agent systems, SLAM, machine learning, and autonomous control of mobile robots.



Song Ma (Member, IEEE) received the M.Sc. degree in mechanical engineering from University College London in 2019, where he is currently pursuing the Ph.D. degree in robotics with the Department of Mechanical Engineering. His main research interests pertaining to the autonomous systems are in multi-agent planning and robotic exploration.



Yuanchang Liu (Member, IEEE) is currently a Professor of autonomy and automation and the Chair of Marine Research with the Department of Mechanical Engineering, University College London. He was also the Program Director of M.Sc. power systems engineering from 2023 to 2025. His research primarily focuses on automation and autonomy in marine, with a special emphasis on exploring technologies related to sensing, perception, and the guidance and control of intelligent and autonomous vehicles. He serves on the Editorial Board of several leading international journals, including *Research—Science Partner Journal*, *Journal of Marine Engineering and Technology*, and *Applied Ocean Research*.



Hyoung Il Son (Senior Member, IEEE) received the B.S. and M.S. degrees from the Department of Mechanical Engineering, Pusan National University, Republic of Korea, in 1998 and 2000, respectively, and the Ph.D. degree from the Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology (KAIST), Republic of Korea, in 2010. In 2015, he joined the Faculty of the Department of Convergence Biosystems Engineering, Chonnam National University, Gwangju, Republic of Korea, where he is currently a Full Professor. He also holds a joint appointment as a Professor with the Department of Robotics Engineering, Chonnam National University. Prior to joining Chonnam National University, he led the Telerobotics Group, Central Research Institute, Samsung Heavy Industries, Daejeon, Republic of Korea, as a Principal Researcher from 2012 to 2015. He has also held several positions in both academia and industry, including a Senior Researcher at LG Electronics, Pyungtaek, Republic of Korea, from 2003 to 2005, and Samsung Electronics, Cheonan, Republic of Korea, from 2005 to 2009, a Research Associate at the Institute of Industrial Science, The University of Tokyo, Tokyo, Japan, in 2010, and a Research Scientist at the Max Planck Institute for Biological Cybernetics, Tübingen, Germany, from 2010 to 2012. His research interests include field robotics, hybrid systems, systems and synthetic biology, and agricultural robotics.

Authorized licensed use limited to: CHONNAM NATIONAL UNIVERSITY. Downloaded on October 30,2026 at 07:49:35 UTC from IEEE Xplore. Restrictions apply.