Journal of Robotics and Control (JRC)
Volume 5, Issue 5, 2024
ISSN: 2715-5072, DOI: 10.18196/jrc.v5i5.22298

The Role of Occasional Assessment of Sensor Performance for Improved Subsea Search Efficiency

Harun Yetkin [1*]
Dept. of Mechatronics Engieering, Bartin University, Bartin, Turkiye
Email: hyetkin@bartin.edu.tr
*Corresponding Author

Abstract—**This study addresses the subsea search performance of an autonomous underwater vehicle equipped with a search sensor and an environment characterization sensor. The performance of the search sensor is assumed to be dependent on characteristics of the local environment, and thus sensor performance in some locations can be different than in other locations. For the case that the agent is able to occasionally characterize the environment, and therefore estimate the performance of its search sensor, we describe a method for selecting when and where to characterize the environment and when and where to search in order to maximize overall search effectiveness. Our work accounts for false positives, false negatives and uncertainty in the performance of the search sensor that varies geographically. We show that effort applied to characterizing the environment, and therefore the performance of the search sensor, can improve search performance. We derive a utility function that is used to compute the best path and when to switch between the tasks of search and environmental characterization. The objective of the subsea search mission is to maximize the probability of attaining a desired level of risk reduction, and we terminate the search mission as soon as it is found that the desired risk reduction cannot be attained. To the best of our knowledge, this is the first study that addresses the problem of attaining a desired level of risk and stopping the mission when the desired risk is found to be unachievable. Through numerical illustrations, we show realistic scenarios where the findings of this study can be useful to improve search effectiveness and attain the desired level of risk where the standard exhaustive search techniques will fail to achieve.**

Keywords*—*Search Theory, Path Planning, Subsea Search, Attained Bayes Risk

I. Introduction

This paper presents an approach to subsea search planning that is based on minimizing the expected value of a decision-theoretic loss function. We address the case where an autonomous underwater vehicle (AUV) searches for an unknown number of stationary objects distributed in a bounded search domain and the search mission is subject to a time or distance constraint. The goal of the search mission is to estimate the number of objects at a location. We assume the local environmental conditions affect the performance of the search sensor such that some locations in the search domain are more informative than others. We consider the case where AUV is equipped with a search sensor and a method of characterizing the performance of the search sensor in the environment. This case arises in search applications where there are two distinct sensors within a system. While in a majority of these applications the sensors can operate simultaneously, there are specific cases where only one sensor can be active due to the physical limitations of the on-board sensors or the computational limitations associated with signal processing. In the context of subsea search (see: [1] for an application survey), marine systems have employed both acoustic sidescan sensors (search sensor) and acoustic sub-bottom profilers (environmental characterization sensor). Simultaneous data collection is possible in this scenario because the frequency bands of the sensors do not overlap (high frequency and low to mid frequencies, respectively). In a prior work [2]–[4], we offer path planning strategies for such cases. However, in some scenarios, simultaneous data collection may not be possible. For example, in order to detect buried objects, some sensing approaches operate at low to mid frequency bands [5]. Similarly, operating in low to mid frequencies has shown increased imaging performance using synthetic aperture sonar processing [6]. In both of these approaches the search sensor would overlap with typical sub-bottom profiling sensors resulting in decreased performance due to mutual interference. To address this, measurements may be taken in serial in order to avoid decreased sensor performance. We note that the approach to compute the optimal paths for such cases is fundamentally different than the approach reported in [2]–[4]. To this end, this paper specifically considers the cases where a search agents is equipped with a search sensor and an environment characterization sensor, but these sensors cannot operate simultaneously. Our main contributions in this study are three-fold:

1) we derive a decision-theoretic cost function that minimizes the Bayes risk of incorrectly estimating the number of objects,
2) we present a mathematically rigorous method to determine when to search and when to characterize the environment,
3) to the best of our knowledge, this is the first study that addresses the problem of attaining a desired level of risk and stopping the mission when the desired risk is found to be unachievable.

The question of when to search a location or to characterize the environment at a location can be important in practice whenever the environment affects the performance of the search sensor and the total search effort is limited. Our approach to search and environmental characterization informs applications such as search and rescue and mine-hunting.

Search theory is concerned with finding an optimal allocation of available search effort to locate a lost or hidden target, such that a reward specified as a measure of search effectiveness is maximized. A large number of studies in the literature address various aspects of the search theory and its applications. Examples include [7]–[25]. While the effect of missed detections (i.e., failing to detect an object that is present) on search effectiveness is often addressed in the literature, the effect of false alarms (i.e., detection of an object that is not present) is mostly ignored. Exceptions include [9], [15], [26]–[29]. In these studies, the environment is assumed to be homogeneous, and thus, the effect of the environment on search effectiveness is not accounted for in the devised search plan. However, it is well known that the environment is an important factor in real-world search problems [30]–[35]. We refer the reader to [36], [37] for a comprehensive survey of the existing literature on the search theory. We note that our cost function to assess the value of searching a location accounts for false alarms (i.e., detection of an object that is not present), missed detections (i.e., failing to detect an object that is present) and uncertainty in the environment.

In this study, we assume switching between search and characterization has no cost. We seek search strategies for which the risk of estimating the incorrect number of objects is below a desired value. We say that search mission is successful if a desired probability of attaining the desired risk is achieved. We terminate the mission when the time/distance limit is met, or we find out that the mission goal cannot be accomplished. When the search agent can either acquire environment data to reduce the uncertainty in the environment or engage in search to estimate the number of objects in the environment, but not both activities simultaneously, the search problem poses an exploration-exploitation trade-off. Exploration implies increasing our knowledge about the environment, and exploitation implies using the current knowledge of the environment to more efficiently conduct the search mission. The exploration-exploitation trade-off has been extensively addressed in the search literature (see, for instance, [38]–[44] for a list of recent studies). However, the approach proposed in these studies do not apply to the specific search problem that we address here.

Contrary to the existing literature, in this work, we consider that the search agent explores the search environment when it performs environment characterization, and exploits its current knowledge on the environment when it performs search. When the search agent searches a location, it returns with an uncertain reward drawn from a known distribution. On the other hand, when the agent characterizes the environment at a location, the expected payoff is zero, but it learns more about the distribution that the reward of a search visit to that location is drawn from. In the literature, the exploration-exploitation trade-off is sometimes modeled through the well-known multi-armed bandit problem where the decision-maker is faced with the decision of either exploiting the current knowledge to improve the current utilization or exploring the other alternatives to increase the future utilization, (see for example [45]–[55]). However, we note that the problem we describe in this paper is fundamentally different and thus it cannot be modeled through a multi-armed bandit problem.

The exploration-exploitation trade-off is also a well-studied problem in other disciplines. In [56], the authors present a framework in monetary problems to help the decision-maker determine whether buying the information to reduce the uncertainty in the outcome outweighs the cost of buying the information. The behavioural study in [57] considers the case where the decision-maker chooses between buying the information that may decrease the expected loss and buying the information that may increase the expected gain. However, meeting a certain goal, such as attaining a desired level of risk, is not the objective of these studies. In [58], the author presents a method to allocate the funds in an investment portfolio where the objective is to collect a certain amount of reward. However, the results are applicable only for a small class of rewards; when the total sums of the rewards are normally distributed or when each reward obeys a Poisson distribution. Thus, new insights into the exploration-exploitation trade-off, as those attaining a target reward with a non-specific rewards distribution developed in our work, have a broad impact in applications as diverse as finance and health-care.

In some search missions, when searching the area further will not improve the search results, it might be important to terminate the mission to free up the search agent to perform other tasks. There are a few studies in which the search mission is stopped when there is adequate belief on the presence or absence of a target (see, for example, [59] and [60]). However, in these studies, the search continues until adequate information is acquired. In this paper, we terminate the search mission when we find out that adequate information on the number of objects cannot be acquired. This strategy improves search efficiency in cases where the goal of the search mission can be achieved under only certain environmental conditions. In other words, the search mission can be terminated early if the environment is such that the search sensor cannot perform well enough to meet the goals of the mission.

The remainder of this paper is organized as follows. In Section II, we formulate the search problem and describe the value of performing a search visit and an environment characterization visit to a location. In Section III, we discuss the problem of determining when to search and when to characterize the environment, and we derive the corresponding cost function. Section IV provides the numerical results.

II. Problem Formulation

A. Preliminaries

We are given a bounded search grid $\mathcal{G} \subset \mathbb{R}^2$ that consists of K disjoint cells. We associate with each cell random variables X and E that represent the number of targets and the environmental conditions in the cell, respectively. Since environmental conditions predict search sensor performance, we presume that knowledge of environmental conditions is equivalent to knowledge of search sensor performance. We presume X_i is independent of X_j and E_i is independent of E_j when $i \neq j$. The searcher's objective is to estimate $X_1, \dots, X_K$. We assume that the environment in each cell is from a finite set of possible environments $\{w_1, w_2, \dots, w_m\}$. We presume that the actual environmental condition in each cell is not known, but that a probability distribution is known for each cell. The environment probability distribution for the ith cell is expressed $P\big(E_i\big) = [p_1(i), p_2(i), \dots, p_m(i)]$ where $p_j(i) = P(E_i = w_j)$ is the probability that the environment is w_j.

When the search vehicle visits a location, it activates either the search sensor to detect the number of objects in the location, or the environmental characterization sensor to observe the environmental conditions at that location. When the search sensor is activated, the vehicle acquires a noisy observation $z \in Z$ of the number of objects. We use Bayesian update law to update our belief on the number of objects, $P(X)$ after acquiring the measurement z. We model the likelihood of observing z objects when x is the true number of objects and w is the true environment

$$P\big(z \mid x, w\big) = \sum_{k=0}^{\min(x,z)} \binom{x}{k} D^k (1-D)^{x-k} (1-\alpha)\alpha^{z-k} \quad (1)$$

where $0 \leq \alpha < 1$ denote the probability of one or more false alarms and $0 < D \leq 1$ the probability of detection. Both α and D are assumed to vary as functions of the environment type w. We refer the reader to [61] for more details on the observation model in (1).

After acquiring the measurement z, we apply Bayes' rule to update our belief on the number of objects.

$$P\big(x \mid z, w\big) \;\propto\; P\big(z \mid x, w\big)\; P\big(x \mid w\big) \quad (2)$$

where $P\big(z \mid x, w\big)$ is the observation model in (1). We assume the number of objects in a cell is statistically independent of the environmental conditions in that cell.

Similarly, when the environmental characterization sensor is activated, the vehicle acquires a noisy observation $Y = y$ of the environmental conditions at that location. We again apply Bayes' rule to update our belief on the environmental conditions.

$$P\big(w \mid y\big) \;\propto\; P\big(y \mid w\big)\; P\big(w\big) \quad (3)$$

where $P\big(y \mid w\big)$ is the likelihood of observing the environment $y \in Y$ given true environment w. We assume that the probability of observing a particular environment is known before the mission starts and does not change. Insight on the sensor model for environment characterization arises from research on subsea bottom-type characterization, such as in [62].

We first show how to formally assess the value of searching the number of objects at a location, and then we show to assess the value of characterizing the environment at that location.

B. Value of Searching a Location

The overall goal of the search mission is to estimate the number of objects in each cell. In our decision-theoretic framework, the value of a search measurement is associated with how likely it is to contribute to a good estimate on the number of objects with the acquired measurement.

After the vehicle searches a location, we compute an estimate, denoted $\delta(z)$, of the number of objects x at that location, based on the observation z. When $\delta(z)$ is greater than x, we overestimate the number of objects, i.e. we declare more than the actual number of objects are present. When $\delta(z)$ is less than x, we underestimate the number of objects, i.e. we fail to declare some of the objects that are present. Both overestimation and underestimation may degrade the utility of the search results. Within our decision-theoretic framework, we define a linear loss function to penalize deviations from the true number of objects. Given the measured data z, we define the loss corresponding to the estimate $\delta(z)$ when x is the true number of objects

$$L\big(x, \delta(z)\big) = c_i \big| x - \delta(z) \big| \quad \text{for } i \in \{1,\ 2\} \quad (4)$$

where $c_1 > 0$ and $c_2 > 0$ are relative costs of overestimating $(\delta(z) > x)$ and underestimating $(\delta(z) < x)$ the number of objects. For some applications, such as mine-hunting and search and rescue, overestimating the number of objects is preferred to underestimation. In mine-hunting missions, overestimating the number of mines, due to false positives, may lead to wasted follow-on effort in neutralizing mines that are not present or avoiding future maneuvering in a location due to the threat of mines, when there are actually no mines present. However, underestimating the number of mines may have disastrous consequences. Thus, we may assign the relative costs such that $c_1 > c_2$.

The posterior expected loss of computing the estimate $\delta(z)$ when the environment is w is

$$E\Big[L\big(x, \delta(z)\big) \mid w\Big] = \sum_x P\big(x \mid z, w\big) L\big(x, \delta(z)\big) \quad (5)$$

where the expectation is taken over the parameter space X with respect to the posterior distribution $P(x \mid z, w)$. The estimator

$\delta^\star$ is called Bayes estimator when it minimizes the expected loss in (5).

$$\delta^\star = \arg\min_{\delta(z)} \; E\Big[L\big(x, \delta(z)\big) \mid w\Big] \tag{6}$$

Expected loss in (5) is called Bayes' risk when $\delta(z)$ is the Bayes estimator in (6).

For notational convenience and clarity of the presentation, we define two types of risk associated with the value of searching a location: the *conditional current risk* and the *conditional anticipated risk*. Loosely speaking, the conditional current risk is the risk of incorrectly estimating the number of objects with the information at hand, and the conditional anticipated risk is the the risk we expect to attain with the additional information after searching the location, both conditioned on the environment w. The conditional current risk for the ith cell with the previously acquired measurement z is denoted

$$\rho(i \mid z, w) = \mathbb{E}\Big[L\big(x, \delta^\star(z)\big) \mid w, z\Big] \tag{7}$$

and the conditional anticipated risk for visiting the ith cell k times is denoted by

$$r(i, k \mid w) = \sum_{\mathbf{z}_k} P\big(\mathbf{z}_k \mid w\big)\mathbb{E}\Big[L\big(x, \delta^\star(\mathbf{z}_k)\big) \mid \mathbf{z}_k, w\Big] \tag{8}$$

where $\mathbf{z}_k = \{z_1, z_2, ..., z_k\}$. Note that both the conditional current risk in (7) and the conditional anticipated risk in (8) are conditional on the environment w.

The value of acquiring a search measurement at a location is the reduction in uncertainty associated with not knowing the true number of objects at that location due to the acquired measurement. Thus, we define the benefit of searching a location for k times given the environment at the location as the difference between the conditional current risk and the conditional anticipated risk, and we call this the *attained risk reduction*.

$$B\big(i, k \mid w\big) = \rho(i \mid w) - r(i, k \mid w) \tag{9}$$

C. Value of Characterizing the Environment at a Location

Due to the stochasticity in the environment, we may not deterministically know the attained risk reduction after a mission. We instead represent our belief on the attained risk reduction through a probability distribution. This probability distribution maps any possible attained risk reduction that follows from (9) to a probability of it being the true value of the attained risk reduction after a mission. We note that characterizing the environmental conditions at a location does not change the attained risk reduction, but it modifies the probability distribution on it. That is, we expect to reduce the uncertainty in the attained risk reduction due to environmental uncertainty rather than to directly increase the attained risk reduction. Thus, the value of acquiring an environment measurement at a location is associated with reducing the environmental uncertainty since it may allow us to better anticipate the attained risk reduction after a mission. We may also choose to characterize the environment when doing so will lead us to better determine when to terminate the search mission. Note that when the environment at a location is deterministically known, the value of performing environmental characterization at that location is zero.

III. Achieving a Target Level of Risk Reduction

In this section we address the question of when to search and when to characterize the environment. We assume that the sensing agent is equipped with an environment characterization sensor and a search sensor, and that both sensors cannot operate simultaneously. Therefore, when the sensing agent visits a location, it activates either the search sensor to observe the number of objects, or the characterization sensor to observe the environmental conditions at the location. Such cases can arise in subsea search applications where both classes of sensors are sonar systems that cannot be operated simultaneously. A modest modification of our results also informs search planning for the case that search and environmental characterization are conducted using entirely different assets that are not operated simultaneously. In this case, the question is when to recover one asset and deploy another in order to maximize overall risk reduction. The objective of the search mission is to reduce the risk of incorrectly estimating the number of objects below a desired level of risk $\bar{\beta}$. We note that this objective can also be interpreted as attaining a desired risk reduction

$$\beta = \sum_{i \in \mathcal{G}} \rho(i) - \bar{\beta} \tag{10}$$

which can be computed more efficiently. Thus, our goal is to determine when to search and when to characterize the environment in order to maximize the probability of attaining the desired level of risk reduction. We assume that switching between the search sensor and the environmental characterization sensor has zero cost and can happen any time during the mission. However, our approach is also applicable if there is a cost associated with switching between the sensors or there is a constraint on when it can happen.

A. Probability Distribution on Risk Reduction

When the attained risk reduction after a mission is greater than the desired level of risk reduction, we consider that the mission is successfully accomplished. Let y denote the environment measurement acquired at a location and let β represent the desired risk reduction. Because the attained risk reduction in (9) is conditioned on the environment w, the attained risk reduction is $B\big(i, k \mid w\big)$ with probability $P(w \mid y)$.

The *probability of success* is the probability of attaining the desired risk reduction and is denoted by $\mathcal{P}$. Given the attained risk reduction conditioned on each environment $w_1, w_2, \ldots, w_m$, we compute the probability of success by

$$\mathcal{P} = \sum_{j:\ B(i,k|w_j)\geq\beta} P(w_j \mid y) \tag{11}$$

We consider the N-length path $\gamma = \{q_1, q_2, \ldots, q_N\}$ and the set γ_d of distinct cells in γ. Let $\mathcal{A}_s$ be the action space, and let $a \in \mathcal{A}_s$ be an action the vehicle takes when it visits a cell. Since the vehicle can activate either the search sensor or the environment characterization sensor, the action space is $\mathcal{A}_S = \{\text{search}, \text{characterize}\}$. We denote the sequence of actions taken along the path γ by $\mathbf{a}_\gamma$. Let m_i be the multiplicity of search measurements acquired at q_ith cell. Then the attained risk reduction when traversing γ is the sum of the attained risk reduction for each cell in the path,

$$B\big(\gamma, \mathbf{a}_\gamma \mid e_\gamma\big) = \sum_{q_i\in\gamma_d} B\big(q_i, m_{q_i} \mid e_{q_i}\big) \tag{12}$$

where $e_\gamma = \{e_{q_1}, \ldots, e_{q_N}\}$, and $e_i \in \{w_1, \ldots, w_m\}$ is the assumed environment in the ith cell. The notation $(\gamma, \mathbf{a}_\gamma)$ indicates that the attained risk reduction is associated with the path γ and the sequence of actions $\mathbf{a}_\gamma$ taken over the path.

Let y_{q_i} be the environment measurements acquired at q_ith cell, and $\mathbf{y}_\gamma$ be the set of environment measurements acquired along the path γ. Then, the risk reduction in (12) is attained with probability

$$P\big(e_\gamma \mid \mathbf{y}_\gamma\big) = \prod_{q_i\in\gamma_d} P\big(e_{q_i} \mid y_{q_i}\big) \tag{13}$$

Since the true environment in each cell may not be known, we compute (12) and (13) for each possible set of true environments e_γ. This yields the probability distribution on attained risk reduction conditioned on the set of environment observations $\mathbf{y}_\gamma$. Then the probability of success for traversing γ, taking actions $\mathbf{a}_\gamma$ and observing $\mathbf{y}_\gamma$ is

$$\mathcal{P}_{\gamma,\mathbf{a}_\gamma} = \sum_{e_\gamma:\ B(\gamma,\mathbf{a}_\gamma|e_\gamma)\ \geq\ \beta} P\big(e_\gamma \mid \mathbf{y}_\gamma\big) \tag{14}$$

B. Gain of Selecting a Sequence of Actions

The optimization problem we address yields a desired path *and* a desired sequence of actions. That is, we seek to determine when and where to search and when and where to characterize the environment. Thus, we are interested in finding the best available path and the best sequence of actions along this path so that the probability of accomplishing the search mission is maximized.

We denote the *desired probability of success* by $\mathcal{B}$. It is the minimum acceptable probability of attaining the desired risk reduction. Thus, the mission is successful if the probability of success $\mathcal{P}$ in (14) is greater than or equal to desired probability of success $\mathcal{B}$. Selecting a path and a set of actions along the path yields a probability distribution on the attained risk reduction conditioned on the environment measurements $\mathbf{y}_\gamma$ acquired along the path. Let Π_B denote the probability distribution in (13) on the attained risk reduction in (12). We consider that after Π_B is computed, a decision upon whether the mission is successfully accomplished or not has to be made. Let $d : \Pi_B \to \mathcal{A}_D$ be the decision rule that maps the distribution on attained risk reduction to an action in the action space $\mathcal{A}_D = \{a_0, a_1\}$. The action $a_0 \in \mathcal{A}_D$ represents the decision that the mission will be successful, and the action $a_1 \in \mathcal{A}_D$ represents the decision that the mission will not be successful.

Candidate actions are assessed by evaluating a gain function that results from the utility of forming the decision d. Given decision d when R is the true attained risk reduction, we define the corresponding loss

$$U\big(R, d\big) = \begin{cases} l_1 & \text{if } \ d = a_0 \text{ and } R > \beta \\ -l_2 N_d & \text{if } \ d = a_1 \\ -l_3 & \text{if } \ d = a_0 \text{ and } R \leq \beta \end{cases} \tag{15}$$

where $l_1 > 0, l_2 > 0$, $l_3 > 0$, and $l_2 << l_1$, $l_2 << l_3$.

We note that the utility function (15) does not yield a decision. Rather, it is used to evaluate the effect of selecting among the actions search and environmental characterization. For the decision $d = a_0$, corresponds to the mission being successfully accomplished, there is a positive utility if true attained risk reduction is greater than the desired risk reduction, and there is a negative utility (cost) if true attained risk reduction is less than the desired risk reduction. The negative utility represents the severe consequences of incorrectly estimating that the mission is successfully accomplished. For the decision $d = a_1$, corresponds to a mission not being successful, the associated cost is proportional to the traversed path length until the decision is formed $(N_d \leq N)$. Incurring a cost in such cases promotes early termination of the mission when the mission cannot be accomplished under the present environmental conditions.

Given a path γ, let $\mathbf{A}_\gamma$ represent the set of possible sequence of actions we can take along the path, and let $B\big(\gamma, \mathbf{a}_\gamma \mid e_\gamma\big)$ be the attained risk reduction conditioned on the set of environments e_γ. Then, for each sequence of actions $\mathbf{a}_\gamma \in \mathbf{A}_\gamma$, the gain of taking these actions is

$$G\big(\mathbf{a}_\gamma\big) = \max_d \sum_{\mathbf{y}_\gamma} \sum_{e_\gamma} P\big(e_\gamma, \mathbf{y}_\gamma\big)\ U\big(B\big(\gamma, \mathbf{a}_\gamma \mid e_\gamma\big), d\big) \tag{16}$$

Let $\Gamma(t)$ denote the finite collection of N-length paths available to the vehicle at time step t. Then, the optimal path and the best sequence of actions are

$$\big(\gamma^\star(t), \mathbf{a}^\star(t)\big) = \arg\max_{\gamma\in\Gamma(t)} \Big(\arg\max_{\mathbf{a}_\gamma\in\mathbf{A}_\gamma} G\big(\mathbf{a}_\gamma\big)\Big) \tag{17}$$

C. Reducing Computational Complexity of the Solution

Computing the optimal path and the optimal set of actions in (17) is equivalent to determining when and where to search and when and where to characterize the environment. However, the maximization of (17) is computationally prohibitive when the search space is large, making the proposed approach infeasible in real-time applications. Thus, we briefly comment on computational issues.

Branch and bound methods are commonly applied in optimization problems [63]–[72]. However, for the specific problem considered in this paper, there are two drawbacks of applying the branch-and-bound approach. First, computing a meaningful upper bound on the probability of success for a given node can be computationally very challenging. Second, an additional visit to a cell modifies the probability distribution on the attained risk reduction of that cell (to be specific, an additional search visit modifies the range of the probability distribution, and an additional characterization visit modifies the shape of the probability distribution). Thus, in order to update the probability distribution on the attained risk reduction of a path when an additional visit is made to a cell in this path, it is necessary to keep track of how the previous visits to that cell affected the probability distribution up to that visit. Thus there are significant memory requirements and corresponding computational requirements.

Instead, we apply a simple trick to reduce the computational effort. We quantize attained risk reduction of a path into N_0 possible values and normalize attain risk reduction to 1. This allows us to compactly represent the probability distribution on the attained risk reduction, and we no longer need to store the risk values but only the probability distribution on them. With N_0 risk reduction values, which is independent of the path length N and the number of possible environments m, we represent the probability distribution on the attained risk reduction with only N_0 values. This not only reduces the required memory storage, but also significantly speeds up the corresponding computations since the number of addition and multiplication operations are drastically reduced. For example, when a new cell is added to a path of length N, we perform mN_0 addition and mN_0 multiplication operations instead of m^{N+1} addition and m^{N+1} multiplication operations. Using this method results in computing the probability of success on a desired risk reduction of $\beta - \frac{1}{N_0}$ which is negligibly different than β when N_0 is large. Alternatively, one can modify the desired risk reduction as $\beta + \frac{1}{N_0}$ to account for the effect of quantization.

In addition, we pre-compute all probability distributions or the attained risk reduction at each cell corresponding to all possible environment measurements and number of search visits. Then given the path, the actions along the path and the environment measurements (if any), we compute the attained risk reduction and the corresponding probability distribution on it by using the pre-computed values for each cell along the path.

While these methods reduce the computational time from several hours to several minutes, the proposed approach can still be computationally infeasible for large-scale problems. In on-going work not reported herein, we are pursuing ways to efficiently reduce the search space by pruning the paths that are guaranteed to be not the optimal.

IV. Numerical Results

In this section, we present simulation results that illustrate the efficacy of the proposed strategy for attaining the desired risk reduction. We consider that the search agent is equipped with a search sensor and an environment characterization sensor, but that these sensors cannot operate simultaneously. Simulations are conducted over a 6-by-6 cell search area. For each cell, we assume a stochastic description of the environment in that cell is available to the vehicle.

When the vehicle visits a location, it either activates the search sensor to observe the number of objects or the environment characterization sensor to observe the environmental conditions at that location. The performance of the search sensor is dependent on the environmental conditions. The particular sensor model that we use for the numerical illustrations is (1). We assume there are three types of environment, $\{w_1, w_2, w_3\}$, in the search area. For each environment, the probability of detection D, and the probability of at least one false alarm α are shown in Table I. Sensor performance increases with increasing probability of detection and decreases with increasing probability of false alarm.

TABLE I. Environment Types

Environment	Probability of Detection	Probability of False Alarm
w_1	0.65	0.4
w_2	0.8	0.3
w_3	0.95	0.05

Thus, environment w_1 is the least and environment w_3 is the most informative. To illustrate the relationship between sensor performance, the probability of false alarm, and the probability of detection, the attained risk reduction corresponding to searching a location that possess one of each of the three environment types is shown in Table II when the relative costs of overestimating and underestimating the number of objects are $c_1 = 3$, $c_2 = 1$. We also show the attained risk reduction for searching that location a second time. Larger attained risk reduction implies better sensor performance. When the vehicle characterizes the environment at a location, it acquires an environment measurement with respect to the characterization sensor model and the true environment at the location. We consider that the sensor model for environment characterization is

$$a_{ij} = P\big(Y = w_i \mid E = w_j\big) \quad \text{for all} \quad i, j \in \{1, 2, 3\} \qquad (18)$$

where a_{ii} is the probability of observing the true environment w_i. For the numerical illustrations, we use the characterization sensor model with $a_{11} = 0.9$, $a_{22} = 0.92$, $a_{33} = 0.94$. That is, for example, there is 0.9 probability of acquiring environment measurement w_1 when w_1 is the true environment at the location. The noisy environment observations are due to nonzero probabilities of observing environment w_i when true environment is w_j, denoted by a_{ij} for $i \neq j$. We assume the probability of acquiring incorrect environment measurement is the same for all possible environments other than the true environment. For example, when the true environment at a location is w_1, since $a_{11} = 0.9$, the probability of acquiring environment measurement w_2 and probability of acquiring environment measurement w_3 are $a_{21} = a_{31} = 0.5$.

Although types of possible environments are known, the specific environment at any location is uncertain. Fig. 1 shows the search area and the corresponding probability distribution for each cell. The number in each cell is simply a label for each cell. For each distribution $\Pi = [p_1, p_2, p_3]$, p_j is the probability that the environment in the cell is w_j. For example, Fig. 1 indicates that there is a 0.95 probability that the environment in cell 11 is w_1, and a 0.05 probability that the environment is w_3. Note that the lighter cells are more likely to be belong to environment w_3, and the darker cells are more likely to belong to environment w_1. The greatest uncertainty about the environment occurs in cells 3, 9 and 15.

Suppose the vehicle visits cell 3 twice. It can either observe the number of objects in the cell twice or observe both the number of objects and the environmental conditions in the cell once. Based on the results in Table II, since the prior environment distribution for cell 3 assigns equal probabilities to environments w_1 and w_3, the attained risk reduction for searching the cell twice is either 0.337 or 0.924 with equal probabilities. Now, suppose the vehicle searches the cell once, then characterizes the environment in the cell and observes $Y = w_3$. Then, the attained risk reduction is either 0.824 with a probability of 0.95 or 0.196 with a probability of 0.05. Thus, the benefit of searching the cell twice is the increase in the attained risk reduction conditioned on the environment, and the benefit of characterizing the environment in the cell is the reduction in the uncertainty of the attained risk.

The objective is to find the best path and the best sequence of actions to attain a desired level of risk reduction. When the best path is computed, the vehicle visits the first cell of the best path and executes the corresponding action. After acquiring data, either on the number of objects or on the environmental conditions, we update the corresponding distribution for that cell and re-plan the best path and best sequence of actions for the remaining mission length using the new information. Suppose the vehicle searches the ith cell and acquires the measurement z. Then, with a probability of $P(E_i = w)$, the *achieved* risk reduction in cell i is

$$\rho(i) - \rho(i \mid z, w) \tag{19}$$

which yields a probability distribution on the achieved risk reduction. When re-planning, the attained risk reduction of a path in (12) and its probability distribution in (13) are modified accordingly to account for the probability distribution on achieved risk reduction in cell i. We will now show the numerical results for simplistic scenarios.

We assume the vehicle starts the mission in cell 1. The vehicle's motion is constrained such that it can only move in four directions - up, down, left or right - as long as it remains in the search area. The mission is terminated either when the maximum mission length is met or when the desired level of risk reduction cannot be attained under present environment conditions.

For numerical illustrations, we consider three different mission objectives to show how the objective of a mission, a desired probability of attaining a desired risk reduction, affects the best path and the best set of actions. For all illustrations, the mission length is 20, the relative costs in (15) are $l_1 = l_3 = 1, l_2 = 0.01$, and the risk is quantized into $N_0 = 100$ values.

We first consider $\beta = 13$ and $\mathcal{B} = 0.85$. For a mission length of 20, this implies that at least 0.85 probability of attaining, on average, a 0.65 risk reduction per cell is required. Note that the attained risk reduction for a single search visit is given as 0.824 in Table II even when the environment is the most informative environment w_3.

TABLE II. Attained Risk with Deterministic Environments

Environment	Attained Risk for Single Search Pass	Attained Risk for Two Search Passes
w_1	0.196	0.337
w_2	0.356	0.559
w_3	0.824	0.924

The best path and the best set of actions corresponding to this case are shown in Fig. 2a. The blue solid line represents the best path for the vehicle, the circles represent search actions, and the asterisks represent the characterization actions. We see that cells 3, 9 and 15 are first characterized and then searched while the other cells in the path are searched once. There is a large uncertainty in the environment for cells 3, 9 and 15. The attained risk reduction for these cells might be either very small so that the desired risk reduction cannot be attained with the desired probability, or it might be large enough to meet the mission objective. Thus, accomplishing the mission is conditioned on the environment observations acquired from cells 3, 9 and 15. The mission objective can be satisfied only if the environment w_3 is observed in each of these cells. If a different environment is observed in any of these cells, the mission cannot be accomplished and the vehicle terminates the mission. Note that the cells 3, 9 and 15 are characterized as early as possible to promote early termination of the mission if the mission cannot be accomplished under

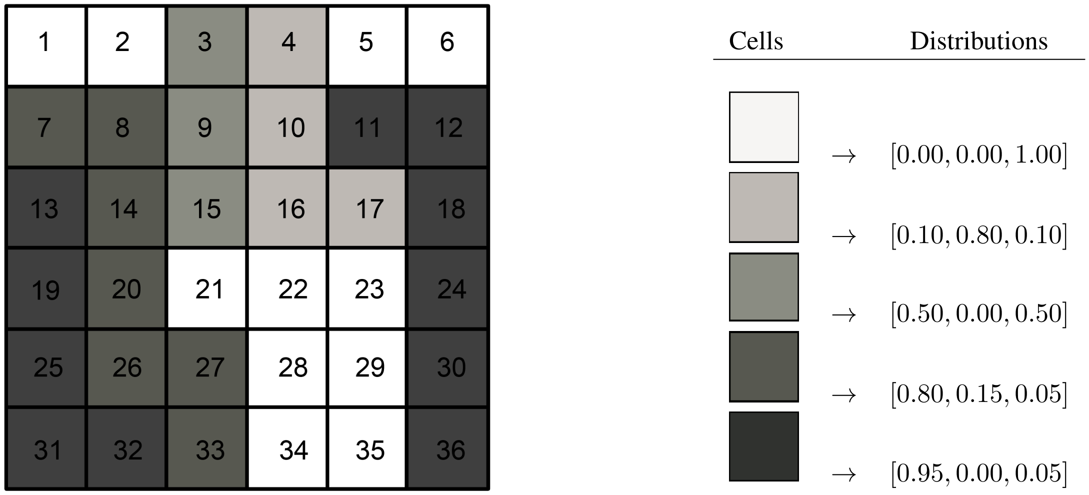
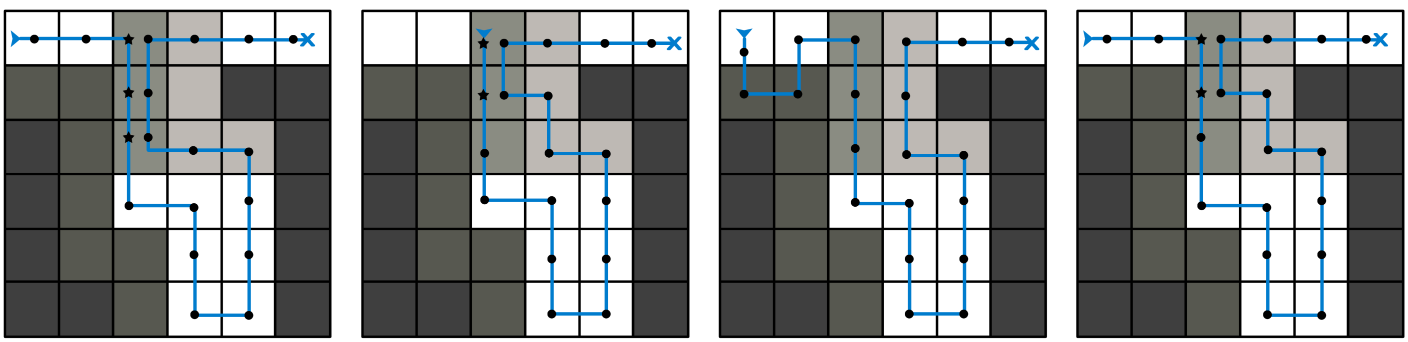


Fig. 1. Search area and cell-wise environment distributions



(a) (b) (c) (d)

Fig. 2. Best path and best sequence of actions when (a) $\beta = 13$ and $\mathcal{B} = 0.85$, (b) $\beta = 13$ and $\mathcal{B} = 0.85$ and $z_1 = z_2 = 2$ are observed, (c) $\beta = 11$ and $\mathcal{B} = 0.85$, and (d) $\beta = 23$ and $\mathcal{B} = 0.65$

present environmental conditions. This is due to the cost $l_2 N_d$ in (15), which is proportional to the length of the traversed path before terminating the mission.

Suppose the vehicle searches cell 1 and cell 2 and acquires *good* search measurements that yield a greater risk reduction than expected to be attained. For example, in our simulations, when the deterministic environment is w_3, acquiring search measurement $z = 2$ yields an achieved risk reduction of 0.88 while the attained risk reduction prior to acquiring the measurement is 0.824. After acquiring search measurements, the vehicle computes the achieved risk reduction in (19) for both cells. When re-planning a new path with the remaining mission length, a smaller risk reduction is required to be attained since a greater risk reduction than expected is achieved in cells 1 and 2. The corresponding path and the set of actions for this specific case are shown in Fig. 2b. Note that the resulting path is different than the path in Fig. 2a. Instead of characterizing the environment in cell 15, the vehicle searches cell 10 so that accomplishing the mission is now conditional on observing environment w_3 in only cells 3 and 9, but not in cell 15. It is evident that observing environment w_3 in cells 3 and 9 occurs with greater probability compared to observing environment w_3 in cells 3, 9 and 15.

In our second illustration, we consider a lower risk reduction $\beta = 11$ that should be easier to achieve and the same desired probability of success $\mathcal{B} = 0.85$. By lowering the desired risk reduction, we expect that the vehicle chooses search actions more often compared to achieving a higher risk reduction. That is, lowering the desired risk reduction increases the probability of attaining it with fewer number of characterization actions so that accomplishing the mission is conditioned on fewer environment measurements. The corresponding path and set of actions are shown in Fig. 2c. We see that the vehicle performs only search actions. Environmental characterization actions are not required since the probability of success can be met even if some environments correspond to poor sensor performance.

Hence, accomplishing the mission is not conditioned on specific environment measurements.

We finally consider the case when $\beta = 13$ but the required probability of success is lowered to $\mathcal{B} = 0.65$. The effect of lowering the desired probability of success is similar to the effect of lowering the desired risk reduction. However, these two cases, lowering the desired probability of success and lowering the desired risk reduction, can yield different paths and different set of actions depending on the search area characteristics. We again expect that the vehicle conducts fewer number of environment characterizations compared to a higher value of desired probability of success. Fig. 2d shows the resulting path and the set of actions. Compared to Fig. 2a, the environment in cell 15 is not characterized so that accomplishing the mission is conditioned on acquiring environment measurement w_3 only in cells 3 and 9.

We note that the corresponding path and set of actions for the first illustration result in a positive gain in (16) when the desired risk reduction or the desired probability of success is lowered as in the later illustrations. However, since the gain of taking a path and a set of actions in (16) depends on the probability of acquiring a particular set of environment measurements along the path $P(\mathbf{y}_\gamma)$, the path in Fig. 2a is suboptimal and therefore not preferred.

V. Conclusions

In this paper, we derive a utility function for subsea search missions that yields when and where to search and when and where to characterize the environment so that the probability of attaining a desired level of risk reduction is maximized. The benefit of search is expressed through a linear error function. We show that if environmental characterization of a location is beneficial for follow-on search, then environmental characterization should be conducted as soon as possible during mission so that in case the mission goals cannot be met under the present environmental conditions, the sensing agent will be freed up sooner. Our results highlight the importance of addressing the adaptive assessment of the local environment in subsea search missions in order to improve overall search effectiveness. Future work will focus on improving the scalability of our approach by efficiently reducing the search space and pruning suboptimal paths.

References

[1] T. R. Clem, “Sensor technologies for hunting buried sea mines,” *OCEANS '02 MTS/IEEE*, vol. 1, pp. 452-460, 2002, doi: 10.1109/OCEANS.2002.1193312.

[2] H. Yetkin, C. Lutz and D. Stilwell, “Acquiring environmental information yields better anticipated search performance,” *OCEANS 2016 MTS/IEEE Monterey*, pp. 1-6, 2016, doi: 10.1109/OCEANS.2016.7761175.

[3] J. McMahon, H. Yetkin, A. Wolek, Z. J. Waters and D. J. Stilwell, “Towards real-time search planning in subsea environments,” *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 87-94, 2017, doi: 10.1109/IROS.2017.8202142.

[4] H. Yetkin, C. Lutz, and D. J. Stilwell, “A decision-theoretic approach to acquire environmental information for improved subsea search performance,” *Ocean Engineering*, vol. 204, p. 107280, 2020, doi: 10.1016/j.oceaneng.2020.107280.

[5] J. A. Bucaro *et al.,* “Acoustic identification of buried underwater unexploded ordnance using a numerically trained classifier (l),” *The Journal of the Acoustical Society of America*, vol. 132, no. 6, pp. 3614–3617, 2012, doi: 10.1121/1.4763997.

[6] M. P. Hayes and P. T. Gough, “Broad-band synthetic aperture sonar,” in *IEEE Journal of Oceanic Engineering*, vol. 17, no. 1, pp. 80-94, 1992, doi: 10.1109/48.126957.

[7] B. O. Koopman, “The theory of search: III. The optimum distribution of searching effort,” *Operations Research*, vol. 5, no. 5, pp. 613–738, 1957, doi: 10.1287/opre.5.5.613.

[8] A. Charnes and W. W. Cooper, “The theory of search: optimum distribution of search effort,” *Management Science*, vol. 5, no. 1, pp. 44–50, 1958, doi: 10.1287/mnsc.5.1.44.

[9] L. D. Stone, J. A. Stanshine, and C. A. Persinger, “Optimal search in the presence of Poisson-distributed false targets,” *SIAM Journal on Applied Mathematics*, vol. 23, no. 1, pp. 6–27, 1972, doi: 10.1137/0123002.

[10] H. R. Richardson, *Search theory*, Center for Naval Analyses, 1986.

[11] J. B. Kadane, “Discrete search and the Neyman-Pearson lemma,” *Journal of Mathematical Analysis and Applications*, vol. 22, no. 1, pp. 156–171, 1968, doi: 10.1016/0022-247X(68)90167-4.

[12] J. B. Kadane, “Optimal whereabouts search,” *Operations Research*, vol. 19, no. 4, pp. 845–1117, 1971, doi: 10.1287/opre.19.4.894.

[13] M. C. Chew Jr, “Optimal stopping in a discrete search problem,” *Operations Research*, vol. 21, no. 3, pp. 661–865, 1973, doi: 10.1287/opre.21.3.741.

[14] G. Kimeldorf and F. H. Smith, “Binomial searching for a random number of multinomially hidden objects,” *Management Science*, vol. 25, no. 11, pp. 1045–1174, 1979, doi: 10.1287/mnsc.25.11.1115.

[15] M. Kress, K. Y. Lin, and R. Szechtman, “Optimal discrete search with imperfect specificity,” *Mathematical Methods of Operations Research*, vol. 68, no. 3, pp. 539–549, 2008, doi: 10.1007/s00186-007-0197-2.

[16] T. H. Chung and J. W. Burdick, “Analysis of Search Decision Making Using Probabilistic Search Strategies,” in *IEEE Transactions on Robotics*, vol. 28, no. 1, pp. 132-144, 2012, doi: 10.1109/TRO.2011.2170333.

[17] T. Cheng, B. Kriheli, E. Levner, and C. Ng, “Scheduling an autonomous robot searching for hidden targets,” *Annals of Operations Research*, vol. 298, no. 1, pp. 95–109, 2021, doi: 10.1007/s10479-019-03141-1.

[18] C. Wang and C. Chen, “A heuristic lowest unknown-degree target search strategy under non-structured environment for multi-agent systems,” *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 24, no. 7, pp. 934–943, 2020, doi: 10.20965/jaciii.2020.p0934.

[19] M. Dunbabin and L. Marques, “Robots for Environmental Monitoring: Significant Advancements and Applications,” in *IEEE Robotics & Automation Magazine*, vol. 19, no. 1, pp. 24-39, 2012, doi: 10.1109/MRA.2011.2181683.

[20] R. Pyla *et al.*, “Design and development of swarm AGV's alliance for search and rescue operations,” *Journal of Robotics and Control (JRC)*, vol. 4, no. 6, pp. 791–807, 2023, doi: 10.18196/jrc.v4i6.18392.

[21] T. Furukawa, F. Bourgault, B. Lavis and H. F. Durrant-Whyte, “Recursive Bayesian search-and-tracking using coordinated uavs for lost targets,” *Proceedings 2006 IEEE International Conference on Robotics and Automation*, pp. 2521-2526, 2006, doi: 10.1109/ROBOT.2006.1642081.

[22] B. Doroodgar, Y. Liu and G. Nejat, “A Learning-Based Semi-Autonomous Controller for Robotic Exploration of Unknown Disaster Scenes While Searching for Victims,” in *IEEE Transactions on Cybernetics*, vol. 44, no. 12, pp. 2719-2732, 2014, doi: 10.1109/TCYB.2014.2314294.

[23] S. Y. Ku, G. Nejat, and B. Benhabib, “Wilderness search for lost persons using a multimodal aerial-terrestrial robot team,” *Robotics*, vol. 11, no. 3, p. 64, 2022, doi: 10.3390/robotics11030064.

[24] S. C. Mohamed, A. Fung and G. Nejat, “A Multirobot Person Search System for Finding Multiple Dynamic Users in Human-Centered Environments,” in *IEEE Transactions on Cybernetics*, vol. 53, no. 1, pp. 628-640, 2023, doi: 10.1109/TCYB.2022.3166481.

[25] B. AlKhlidi, A. T. Abdulsadda, and A. Al Bakri, “Optimal robotic path planning using intelligents search algorithms,” *Journal of Robotics and Control (JRC)*, vol. 2, no. 6, pp. 519–526, 2021, doi: 10.18196/jrc.26132.

[26] S. M. Pollock, *Sequential search and detection*, Massachusetts Institute of Technology, Dept. of Physics, 1964.

[27] J. M. Dobbie, "Some search problems with false contacts," *Operations Research*, vol. 21, no. 4, pp. 867–1016, 1973, doi: 10.1287/opre.21.4.907.

[28] T. H. Chung and J. W. Burdick, "Analysis of Search Decision Making Using Probabilistic Search Strategies," in *IEEE Transactions on Robotics*, vol. 28, no. 1, pp. 132-144, 2012, doi: 10.1109/TRO.2011.2170333.

[29] B. Kriheli, E. Levner, and A. Spivak, "Optimal search for hidden targets by unmanned aerial vehicles under imperfect inspections," *American Journal of Operations Research*, vol. 6, no. 2, pp. 153–166, 2016, doi: 10.4236/ajor.2016.62018.

[30] J. De Guenin, "Optimum distribution of effort: an extension of the Koopman basic theory," *Operations Research*, vol. 9, no. 1, pp. 1–144, 1961, doi: 10.1287/opre.9.1.1.

[31] P. A. Elmore, W. E. Avera, M. M. Harris and K. M. Duvieilh, "Environmental Measurements Derived from Tactical Mine-Hunting Sonar Data," *OCEANS 2007 - Europe*, pp. 1-5, 2017, doi: 10.1109/OCEANSE.2007.4302463.

[32] A. Zare and J. T. Cobb, "Sand ripple characterization using an extended synthetic aperture sonar model and MCMC sampling methods," *2013 OCEANS - San Diego*, pp. 1-7, 2013, doi: 10.23919/OCEANS.2013.6741000.

[33] K. Takahashi, J. Igel and H. Preetz, "Clutter Modeling for Ground-Penetrating Radar Measurements in Heterogeneous Soils," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 4, no. 4, pp. 739-747, 2011, doi: 10.1109/JSTARS.2011.2106481.

[34] K. Takahashi, H. Preetz, and J. Igel, "Soil properties and performance of landmine detection by metal detector and ground-penetrating radar—soil characterisation and its verification by a field test," *Journal of Applied Geophysics*, vol. 73, no. 4, pp. 368–377, 2011, doi: 10.1016/j.jappgeo.2011.02.008.

[35] P. D. Gader, M. Mystkowski and Yunxin Zhao, "Landmine detection with ground penetrating radar using hidden Markov models," in *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 6, pp. 1231-1244, 2001, doi: 10.1109/36.927446.

[36] S. J. Benkoski, M. G. Monticino, and J. R. Weisinger, "A survey of the search theory literature," *Naval Research Logistics (NRL)*, vol. 38, no. 4, pp. 469–494, 1991, doi: 10.1002/1520-6750.

[37] T. H. Chung, G. A. Hollinger, and V. Isler, "Search and pursuit-evasion in mobile robotics," *Autonomous robots*, vol. 31, no. 4, pp. 299–316, 2011, doi: 10.1007/s10514-011-9241-4.

[38] P. Ghassemi and S. Chowdhury, "Decentralized informative path planning with balanced exploration-exploitation for swarm robotic search," in *Proceedings of the ASME 2019 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, vol. 1, pp. 1–11, doi: 10.1115/DETC2019-97887.

[39] H. L. Kwa, J. Leong Kit, and R. Bouffanais, "Balancing collective exploration and exploitation in multi-agent and multi-robot systems: A review," *Frontiers in Robotics and AI*, vol. 8, p. 771520, 2022, doi: 10.3389/frobt.2021.771520.

[40] H. Bai *et al*., "A study of robotic search strategy for multi-radiation sources in unknown environments," *Robotics and Autonomous Systems*, vol. 169, p. 104529, 2023, doi: 10.1016/j.robot.2023.104529.

[41] A. Munir and R. Parasuraman, "Exploration–exploitation tradeoff in the adaptive information sampling of unknown spatial fields with mobile robots," *Sensors*, vol. 23, no. 23, p. 9600, 2023, doi: 10.3390/s23239600.

[42] M. Park, S. An, J. Seo and H. Oh, "Autonomous Source Search for UAVs Using Gaussian Mixture Model-Based Infotaxis: Algorithm and Flight Experiments," in *IEEE Transactions on Aerospace and Electronic Systems*, vol. 57, no. 6, pp. 4238-4254, 2021, doi: 10.1109/TAES.2021.3098132.

[43] C. Tholen, T. A. El-Mihoub, L. Nolle, and O. Zielinski, "Artificial intelligence search strategies for autonomous underwater vehicles applied for submarine groundwater discharge site investigation," *Journal of Marine Science and Engineering*, vol. 10, no. 1, 2021, doi: 10.3390/jmse10010007.

[44] C. Rhodes, C. Liu and W. -H. Chen, "Autonomous Source Term Estimation in Unknown Environments: From a Dual Control Concept to UAV Deployment," in *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 2274-2281, 2022, doi: 10.1109/LRA.2022.3143890.

[45] J. C. Gittins, "Bandit processes and dynamic allocation indices," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 41, no. 2, pp. 148–164, 1979, doi: 10.1111/j.2517-6161.1979.tb01068.x.

[46] T. L. Lai and H. Robbins, "Asymptotically efficient adaptive allocation rules," *Advances in Applied Mathematics*, vol. 6, no. 1, pp. 4–22, 1985, doi: 10.1016/0196-8858(85)90002-8.

[47] O. Besbes, Y. Gur, and A. Z. Chen, "Optimal exploration-exploitation in a multi-armed-bandit problem with non-stationary rewards," *Stochastic Systems*, vol. 9, no. 4, pp. 319–337, 2019, doi: 10.2139/ssrn.2436629.

[48] I. Q. Lordeiro, D. B. Haddad and D. O. Cardoso, "Multi-Armed Bandits for Minesweeper: Profiting From Exploration–Exploitation Synergy," in *IEEE Transactions on Games*, vol. 14, no. 3, pp. 403-412, 2022, doi: 10.1109/TG.2021.3082909.

[49] M. Tajik, B. M. Tosarkani, A. Makui, and R. Ghousi, "A novel two-stage dynamic pricing model for logistics planning using an exploration–exploitation framework: A multi-armed bandit problem," *Expert Systems with Applications*, vol. 246, p. 123060, 2024, doi: 10.1016/j.eswa.2023.123060.

[50] X. Li, Y. Li, and X. Wu, "Empirical gittins index strategies with ε-explorations for multi-armed bandit problems," *Computational Statistics & Data Analysis*, vol. 180, p. 107610, 2023, doi: 10.1016/j.csda.2022.107610.

[51] S. Jamieson, J. P. How, and Y. Girdhar, "Finding the optimal exploration-exploitation trade-off online through Bayesian risk estimation and minimization," *Artificial Intelligence*, vol. 130, p. 104096, 2024, doi: 10.1016/j.artint.2024.104096.

[52] G. Elena, K. Milos, and I. Eugene, "Survey of multiarmed bandit algorithms applied to recommendation systems," *International Journal of Open Information Technologies*, vol. 9, no. 4, pp. 12–27, 2021.

[53] D. Padmanabhan, S. Bhat, K. Prabuchandran, S. Shevade, and Y. Narahari, "Dominant strategy truthful, deterministic multi-armed bandit mechanisms with logarithmic regret for sponsored search auctions," *Applied Intelligence*, vol. 52, no. 3, pp. 3209–3226, 2022, doi: 10.1007/s10489-021-02387-2.

[54] D. Marković, H. Stojić, S. Schwöbel, and S. J. Kiebel, "An empirical evaluation of active inference in multi-armed bandits," *Neural Networks*, vol. 144, pp. 229–246, 2021, doi: 10.1016/j.neunet.2021.08.018.

[55] I. Nasim, A. S. Ibrahim and S. Kim, "Learning-Based Beamforming for Multi-User Vehicular Communications: A Combinatorial Multi-Armed Bandit Approach," in *IEEE Access*, vol. 8, pp. 219891-219902, 2020, doi: 10.1109/ACCESS.2020.3043301.

[56] K. K. Damghani, M. Taghavifard, and R. T. Moghaddam, "Decision making under uncertain and risky situations," in *Enterprise Risk Management Symposium Monograph Society of Actuaries-Schaumburg, Illinois*, vol. 15, 2009.

[57] J. Bowen and Z.-l. Qiu, "Satisficing when buying information," *Organizational Behavior and Human Decision Processes*, vol. 51, no. 3, pp. 471–481, 1992, doi: 10.1016/0749-5978(92)90022-Y.

[58] M. I. Henig, "Risk criteria in a stochastic knapsack problem," *Operations Research*, vol. 38, no. 5, pp. 820–825, 1990, doi: 10.1287/opre.38.5.820.

[59] K. E. Wilson, R. Szechtman, and M. P. Atkinson, "A sequential perspective on searching for static targets," *European Journal of Operational Research*, vol. 215, no. 1, pp. 218–226, 2011, doi: 10.1016/j.ejor.2011.05.045.

[60] M. Kress, K. Y. Lin, and R. Szechtman, "Optimal discrete search with imperfect specificity," *Mathematical methods of operations research*, vol. 68, no. 3, pp. 539–549, 2008, doi: 10.1007/s00186-007-0197-2.

[61] H. Yetkin, C. Lutz and D. Stilwell, "Utility-based adaptive path planning for subsea search," *OCEANS 2015 - MTS/IEEE Washington*, pp. 1-6, 2015, doi: 10.23919/OCEANS.2015.7404367.

[62] S. Jaramillo and G. Pawlak, "AUV-based bed roughness mapping over a tropical reef," *Coral Reefs*, vol. 30, no. 1, pp. 11–23, 2011, doi: 10.1007/s00338-011-0731-9.

[63] J. Binney and G. S. Sukhatme, "Branch and bound for informative path planning," *2012 IEEE International Conference on Robotics and Automation*, pp. 2147-2154, 2012, doi: 10.1109/ICRA.2012.6224902.

[64] B. W. Wah and Chee Fen Yu, "Stochastic Modeling of Branch-and-Bound Algorithms with Best-First Search," in *IEEE Transactions on Software Engineering*, vol. SE-11, no. 9, pp. 922-934, 1985, doi: 10.1109/TSE.1985.232550.

[65] W. Zhang, "Depth-first branch-and-bound versus local search: A case study," in *National Conference on Artificial Intelligence*, pp. 930–935, 2000.

[66] T. Calogiuri, G. Ghiani, E. Guerriero, and R. Mansini, "A branch-and-bound algorithm for the time-dependent rural postman problem," *Computers & Operations Research*, vol. 102, pp. 150–157, 2019, doi: 10.1016/j.cor.2018.07.016.

[67] J. Ahn and H. -J. Kim, "A Branch and Bound Algorithm for Scheduling of Flexible Manufacturing Systems," in *IEEE Transactions on Automation Science and Engineering*, 2023, doi: 10.1109/TASE.2023.3296087.

[68] H. Wang, X. Zhao, S. Huang, Q. Li, and Y. Liu, "A branch-and-bound based globally optimal solution to 2d multi-robot relative pose estimation problems," *Automatica*, vol. 164, p. 111654, 2024, doi: 10.1016/j.automatica.2024.111654.

[69] J. G. Martin *et al*., "Multi-robot task allocation problem with multiple nonlinear criteria using branch and bound and genetic algorithms," *Intelligent Service Robotics*, vol. 14, pp. 707–727, 2021, doi: 10.1007/s11370-021-00393-4.

[70] L. Hong, Y. Wang, Y. Du, X. Chen, and Y. Zheng, "UAV search-and-rescue planning using an adaptive memetic algorithm," *Frontiers of Information Technology & Electronic Engineering*, vol. 22, no. 11, pp. 1477–1491, 2021, doi: 10.1631/FITEE.2000632.

[71] S. Saha, A. E. Vasegaard, I. Nielsen, A. Hapka, and H. Budzisz, "UAVs path planning under a bi-objective optimization framework for smart cities," *Electronics*, vol. 10, no. 10, p. 1193, 2021, doi: 10.3390/electronics10101193.

[72] B. Hermans, R. Leus, and J. Matuschke, "Exact and approximation algorithms for the expanding search problem," *INFORMS Journal on Computing*, vol. 34, no. 1, pp. 281–296, 2022, doi: 10.1287/ijoc.2020.1047.