# Information-Theoretic Interactive Sensing and Inference for Autonomous Systems

Christopher Robbiano, *Member, IEEE,* Mahmood R. Azimi-Sadjadi, *Life Member, IEEE,* and Edwin K. P. Chong, *Fellow, IEEE*

***Abstract*—This paper addresses an autonomous exploration problem in which a mobile sensor, placed in a previously unseen search area, utilizes an information-theoretic navigation cost function to dynamically select the next sensing action, i.e., location from which to take a measurement, to efficiently detect and classify objects of interest within the area. The information-theoretic cost function proposed in this paper consist of two *information gain* terms, one for detection and localization of objects and the other for sequential classification of the detected objects. We present a novel closed-form representation for the cost function, derived from the definition of mutual information. We evaluate three different policies for choosing the next sensing action: lawn mower, greedy, and non-greedy. For these three policies, we compare the results from our information-theoretic cost functions to the results of other information-theoretic inspired cost functions. Our simulation results show that search efficiency is greater using the proposed cost functions compared to those of the other methods.**

***Index Terms*—Autonomous Navigation, Occupancy Grids, Sequential Classification, Information Gain, Mutual Information, Sonar**

## I. Introduction

In this paper, we consider the problem of autonomous exploration for the purpose of interactive sensing and inference in previously unseen search areas. At each time step, the autonomous platform performs a sensing action in the form of selecting and moving to the next position to collect a measurement that is used to update the detection, localization, and classification estimates. In this exploration problem, often referred to as the *active perception* problem [1], no pre-planned platform path is assumed as there is no *a priori* information about objects in the search area, and all initial sensing actions are regarded as providing the same amount of information. Additionally, the motion of the platform is restricted by some dynamical model, hence precluding arbitrary sequential sensing locations.

In active perception problems, the efficiency with which the search area is surveyed is typically the most important criterion as it directly relates to operational costs per sensing action, a need to minimize surveillance time during an information gathering sortie, as well as other time- and cost-sensitive objectives. That is, one is concerned with achieving high efficiency through minimizing the number of sensing actions, while maximizing the detection and classification performance.

To achieve such goals, information-theoretic measures have typically been used [2]–[6] for choosing optimal sensing actions in autonomous navigation and exploration problems. In the case of parameter estimation using measurements that are corrupted by Gaussian noise, maximizing the Shannon entropy of the error distribution is equivalent to minimizing the determinant of the parameter estimate covariances [2]. This provides a rule for selecting the sensing action that maximizes the predicted variance of the measurement produced after a sensing action is performed.

In [3], a subclass of the active perception problem is addressed, where an autonomous underwater vehicle (AUV) is used to inspect the hull of a large ship and estimate its surface shape. Gaussian process function approximation is exploited to approximate a mutual information-based cost function. In this particular multi-hypothesis testing problem, *a priori* information is available that allows the entire set of sensing actions, and their outcomes, to be observed prior to visiting all sensing locations.

Information-theoretic cost functions, specifically utilizing *information gain*, have been previously developed [4]–[6], and used successfully, in the context of navigation using information from the occupancy grid estimation process [7], [8].

In [4], the positions of the AUV and potential targets are estimated for a given sensing action using an extended Kalman filter (EKF), and the mutual information is directly calculated following an update to the occupancy grid. In addition to the occupancy grid based information gain cost, the authors in [4] suggest formulating an additional information-theoretic cost function from the outputs of the simultaneous localization and mapping (SLAM) problem using an EKF to estimate the positions of the AUV and objects in the search area. Specifically, the cost function they choose is related to the determinant of the AUV and object position error covariance matrices, similar to [2]. A convex combination of the two normalized cost functions is used in the sensing action selection, providing the ability to trade-off performance in localization (through SLAM) and detection (through occupancy grids) of objects.

In [5], the mutual information is directly calculated after each sensing action and subsequent measurement is taken *a priori*, and used to train the Gaussian process regression for estimating the mutual information for future sensing actions. Bayesian optimization is then used in conjunction with the

C. Robbiano, M.R. Azimi-Sadjadi, E.K.P. Chong, are with the Electrical and Computer Engineering department at Colorado State University, Fort Collins, CO 80524 USA (email: {chris.robbiano, edwin.chong, azimi}@colostate.edu).

This work was supported by the Office of Naval Research (ONR) under contract N00014-18-1-2805.

Manuscript received March ??, 2020; revised Month ??, 2020.

Gaussian process upper confidence bound to estimate the information gain for each point in an occupancy grid.

The formulation for explicitly calculating the predicted mutual information in [6] is developed using an occupancy grid framework under the assumption of statistical independence of measurements. Measurements are also assumed to be conditionally independent of the occupancy state of obscured grid cells, i.e., grid cells behind occupied cells in the perceptual range of the sensor, given an occupied grid cell.

In this paper, we propose a new approach to the problem of active perception using two information-theoretic cost functions. The first cost function is associated with object detection and localization, and measures the mutual information between the occupancy state variable for a single grid cell and a binary measurement random variable. Solving for its closed-form representation relies on the measurement model and posterior occupancy distributions produced through the occupancy grid estimation process presented in [9]. The second cost function associated with the classification of detected objects measures the mutual information between a class state variable for a single grid cell and random variable that is the parameter to the class state variable distribution. In this formulation, we choose to model the class state variable as a Categorical random variable, and its distribution parameter as a Dirichlet random variable [10]. The motivation for choosing this modeling scheme stems from the need to perform sequential updating of the class state variable distribution, akin to occupancy grid estimation process. This sequential updating process has a closed form due to the conjugacy between the Dirichlet and Categorical distributions, and allows for fast tracking of the class state distribution as new measurement are drawn. A *one-step* classification process is used in our formulation, producing class labels used to sequentially update the class state variable distribution. We call the entire process of using a one-step classifier for producing class labels and the Dirichlet-Categorical model for tracking the classification state the DCM. Similar to [4], a convex combination weighting of two normalized cost functions for sensing action selection is also utilized here.

A series of experiments are conducted to illustrate the utility of the proposed sequential state updating in conjunction with the proposed cost functions. Three sensing action selection policies are compared—lawn mower, greedy, and non-greedy. The lawn mower policy does not use a cost function in the choice of the next sensing action. For the two policies that use a cost function for selecting the next sensing action, greedy and non-greedy, two different methods for estimating the information gain are used: (a) the convex combination of our two theoretical derived proposed cost functions, and (b) a Gaussian process regression (GPR) for approximating the proposed cost function from training data [3], [5]. The GPR provides a robust data driven estimate of the proposed information gain based cost function. The performance of all three policies is evaluated for their ability to explore the interrogation area, and detect and classify targets, while performing only a limited number of actions. The results show that policies using our theoretical derived cost function outperform all policies using the data driven approximation.

The main contributions of this paper are as follows. The development of the DCM and one-step classifier provides a novel application for sequential classification and tracking of the classification state of each grid cell. The derivation of our proposed cost functions and their convex combination provides the theoretical estimates of the information gained from each sensing action, hence providing a framework for optimal navigation informed by detection and classification state estimates.

This paper is organized as follows. A review of the occupancy grid estimation process is presented in Section III. The sequential classification process, including a description of the one-step classifier, is presented in Section IV. The derivations of the information-theoretic cost functions for detection and classification are presented in Section V. The different navigation policies are described in Section VI. Simulation results on synthetic sonar data, and a comparison with other methods for choosing sensing actions [4], [5] are presented in Section VII. Finally, concluding remarks are made in Section VIII.

## II. System Overview

The specific active perception problem considered here involves undersea mine hunting in littoral zones using an AUV equipped with a side-looking sonar system, though the proposed formulations are not restricted only to this sensor configuration. The AUV explores previously unseen areas and simultaneously performs detection and classification of undersea objects. We divide this active perception problem into four segments:

(1) Generating a map of the scene through an occupancy grid estimation process, which produces a set of marginal posterior probabilities that any one point in an area is occupied [9].
(2) Classifying occupied regions through a sequential classification process, which produces a set of marginal posterior probabilities of class membership for each occupied region.
(3) Computing the (i) mutual information between the occupancy state of a grid cell and a random variable modeling a measurement on that grid cell and (ii) mutual information between the class state of a grid cell and a random variable modeling the class distribution parameter.
(4) Exploiting the information gain to select the sensing action that produces the best opportunity to detect, localize, and classify an object.

An illustration of the proposed active perception problem is given in Figure 1. The search area is discritized into grid cells where the knowledge about object locations in the environment is captured through an occupancy grid estimation process [9]. The knowledge about the class of each detected and localized object is then provided in the form of a Dirichlet-Categorical model (DCM) [10]. The occupancy grid estimation process and the DCM produce a set of distributions over occupancy state and class state of grid cells, respectively, from which the uncertainty in the distributions can be measured through Shannon entropy [11]. The state distribution of each grid cell is updated using the measurement collected after each sensing action, and the reduction in the uncertainty of each distribution

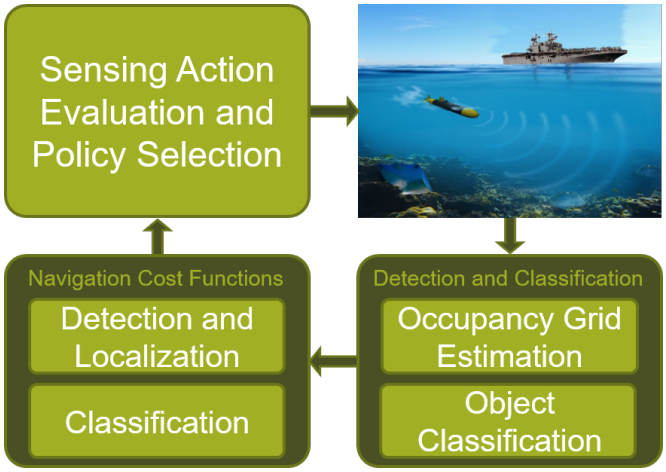



Fig. 1. The proposed active perception problem: An AUV takes measurements to perform occupancy grid estimation and sequential classification. The outputs from each of these processes are then used in evaluating the navigation cost function for sensing action and policy selection.

after each sensing action can be measured through *information gain* or *mutual information* [11]. The information gain is expressed as the difference between the entropy of a prior distribution for a state variable and the entropy of the posterior distribution for that state variable after a measurement has been observed. It is natural to seek sensing actions that will produce a measurement maximizing the reduction in uncertainty in both occupancy state and class state for all grid cells, and thus we choose to utilize information gain as our information-theoretic cost function for choosing sensing actions. These components of the proposed system are described in the next sections.

In the remainder of this paper, we shall use lowercase italic $x$ for scalars, lowercase bold italic symbols $\boldsymbol{x}$, and uppercase bold italic symbols $\boldsymbol{X}$, for vectors and matrices, respectively.

## III. Occupancy Grid Estimation

Occupancy grid estimation is a popular process for generating an occupancy map of an area given a set of measurements taken from that area [7]–[9]. The map is partitioned into a set of $B$ disjoint *grid cells* $\{g_i\}_{i=1}^{B}$, all with the same shape and size. To each grid cell a binary occupancy state indicator variable $b_i \in \{0,1\}$ is attached with $b_i = 1$ indicating that a grid cell $g_i$ is *occupied* (e.g., by a scatterer of radiation), and $b_i = 0$ indicating that $g_i$ is *empty*. We call the set $\boldsymbol{b} = \{b_i\}_{i=1}^{B}$ the set of *cellular occupancies*, commonly referred to as a *map*. The map $\boldsymbol{b}$ can be any one of $2^B$ possible unique maps from the set of all possible maps $\mathbb{B}$.

Now, given the measurement matrix $\boldsymbol{J}_S = \left[\boldsymbol{j}_1, \cdots, \boldsymbol{j}_s, \cdots, \boldsymbol{j}_S\right]$, consisting of a collection of measurement vectors $\boldsymbol{j}_s = \left[j_{s,1}, \cdots, j_{s,K}\right] \in \mathcal{J}^K = \{0,1\}^K$ for $s \in \{1, \cdots, S\}$ with $K$ elements that are the thresholded detection statistics taken at time $s$, the estimation problem produces the set of marginal posterior probabilities of occupancy grids (OGs) arranged as a vector

$$\boldsymbol{p} = \{p_{b|\boldsymbol{J}}(b_r = 1|\boldsymbol{J}_S)\}_{r=1}^{B}. \tag{1}$$

Using the following occupancy grid formulation presented in [9], these marginal posterior probabilities at time step $S$ can be expressed as

$$\begin{aligned} p_{b|\boldsymbol{J}}(b_r = 1|\boldsymbol{J}_S) &\propto \sum_{\boldsymbol{b}\in\mathbb{B}(r,1)} p_{\boldsymbol{j}|\boldsymbol{b}}(\boldsymbol{j}_S|\boldsymbol{b}) p_{\boldsymbol{b}|\boldsymbol{J}}(\boldsymbol{b}|\boldsymbol{J}_{S-1}) \\ &= \sum_{\boldsymbol{b}\in\mathbb{B}(r,1)} \prod_k \prod_i \Big[ \left(p_{ki}^{00}(1-b_i) + p_{ki}^{01} b_i\right)(1 - j_{S,k}) \\ &\quad + \left(1 - \left(p_{ki}^{00}(1-b_i) + p_{ki}^{01} b_i\right)\right) j_{S,k} \Big] \\ &\quad \times p_{\boldsymbol{b}|\boldsymbol{J}}(\boldsymbol{b}|\boldsymbol{J}_{S-1}), \end{aligned} \tag{2}$$

where $\mathbb{B}(r,1)$ is the set of all maps with the $r$th occupancy state pinned to occupied. Given the map $\boldsymbol{b}$, for any arbitrary time $s \in [1, S]$, the *sensor model* $p(\boldsymbol{j}_s|\boldsymbol{b})$ can be written as

$$p_{\boldsymbol{j}|\boldsymbol{b}}(\boldsymbol{j}_s|\boldsymbol{b}) = \prod_k p_{j|\boldsymbol{b}}(j_{s,k}|\boldsymbol{b}). \tag{3}$$

To express the terms under the product in (3) the BAC model was adopted in [9]. Figure 2 illustrates the interaction between all of the grid cell occupancy states and a single measurement where each $j_{s,k}$ is a Boolean function of *virtual occupancies* $\tilde{b}_i$ (outputs of the BACs); specifically $j_{s,k} = \sum_{i=1}^{B} \tilde{b}_i$. Then, we can write,

$$\begin{aligned} p_{j|\boldsymbol{b}}(j_{s,k} = 0|\boldsymbol{b}) &= \prod_i p_{\tilde{b}|b}(\tilde{b}_i = 0|b_i) \\ &= \prod_i p_{ki}^{00}(1-b_i) + p_{ki}^{01} b_i, \end{aligned} \tag{4}$$

The quantity $p_{ki}^{00}$ is the probability that the occupancy state of grid cell $g_i$ is transmitted through the BAC and correctly received as measurement $j_{s,k} = 0$ when $b_i = 0$ (probability of true non-detection), and $p_{ki}^{01}$ is the probability that the occupancy state of grid cell $g_i$ is transmitted through the BAC and incorrectly received as measurement $j_{s,k} = 0$ when $b_i = 1$ (probability of missed detection). As $j_{s,k}$ is a binary random variable, $p_{j|\boldsymbol{b}}(j_{s,k} = 1|\boldsymbol{b}) = 1 - p_{j|\boldsymbol{b}}(j_{s,k} = 0|\boldsymbol{b})$. The last term in (2) $p_{\boldsymbol{b}|\boldsymbol{J}}(\boldsymbol{b}|\boldsymbol{J}_{S-1})$ is the posterior probability of the map $\boldsymbol{b}$ at the previous time step $S-1$ calculated as

$$\begin{aligned} p_{\boldsymbol{b}|\boldsymbol{J}}(\boldsymbol{b}|\boldsymbol{J}_{S-1}) \propto \prod_k \prod_i \Big[ &\left(p_{ki}^{00}(1-b_i) + p_{ki}^{01} b_i\right)(1 - j_{S-1,k}) \\ &+ \left(1 - \left(p_{ki}^{00}(1-b_i) + p_{ki}^{01} b_i\right)\right) j_{S-1,k} \Big] \\ &\times p_{\boldsymbol{b}|\boldsymbol{J}}(\boldsymbol{b}|\boldsymbol{J}_{S-2}). \end{aligned} \tag{5}$$

One method for choosing the BAC transition probabilities $p_{ki}^{00}$ and $p_{ki}^{01}$ is to allow $p_{ki}^{00} = (1-p_{\text{fa}})/(1+\text{dist}(b_i, j_{s,k}))^{\alpha}$ and $p_{ki}^{01} = (1-p_{\text{d}})/(1+\text{dist}(b_i, j_{s,k}))^{\alpha}$, where $p_{\text{d}}$ and $p_{\text{fa}}$ are the probability of detection and false alarm, respectively, of the physical sonar system, $\text{dist}(b_i, j_{s,k})$ represents the Euclidean distance between the location of grid cell $g_i$ and that at which sample $j_{s,k}$ was taken, and $\alpha \geq 1$ [9]. This particular modeling is used to emulate degraded detection performance due to attenuation in the sonar return signal strength as a function of distance.

This formulation of occupancy grid estimation is used over other estimation techniques as it is able to account for the correlation between occupancy states of neighboring grid cells, and was developed with the measurement type (binary detection statistics) that are used in the implementation of our system.

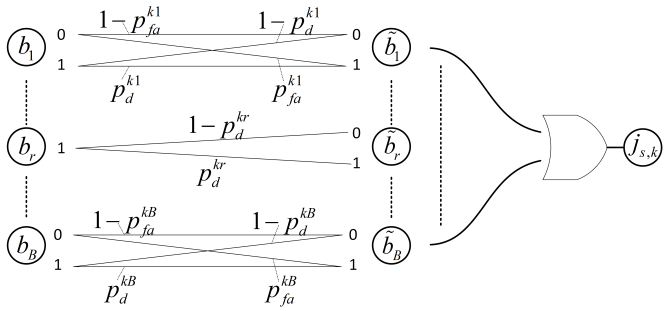



Fig. 2. Modeling of interaction between occupancy states and a single measurement $j_{s,k}$, conditioned on $b_r = 1$.

## IV. Sequential Classification Using Dirichlet-Categorical Models

In this section, we present a new sequential classification method that allows tracking of the class state for each grid cell, formulated with the need of an information-theoretic classification cost function in mind. We desire a method that produces a set of distributions as its output, and the sequential updating of these distributions to be performed quickly, without the need to iterate an algorithm to convergence. As such, we chose the Dirichlet-Categorical model (DCM) [10], [12] for representing the class state variable and its associated distribution random variable.

The idea behind this sequential updating process is to take a measurement from the sensor at time $s$ and use it to update the class membership probabilities for the grid cell. The measurement is first converted into a crude estimate of the class label $l_s$, and that label is merged with all previous labels to update the posterior predictive class distribution for the grid cell.

Let $c$ be the class state variable for grid cell $g_i$. At each time step $s$, a *one-step* classifier is employed to assign a class label $l_s \in [1, L]$ to the most recent measurement from grid cell $g_i$. The one-step classifier in this system can be any commonly used classifier such as support vector machine (SVM), relevance vector machine (RVM) [13] or a deep neural network (DNN) [14]. The collection of sequential class labels $l_s$ for grid cell $g_i$ up to the current sensing time $S$, are formed into a set $\mathcal{L} = \left[l_1, \cdots, l_S\right]$. Now, the goal here is to generate the posterior predictive distribution of the class state variable $c$, $p_{c|\mathcal{L}}(c|\mathcal{L})$, given all the past and present labels in $\mathcal{L}$.

To begin, we model $c$ as a Categorical random variable taking on $L$ possible, non-orderable, values. A random variable $c$ is Categorically distributed if $p_{c|\boldsymbol{\lambda}}(c = l|\boldsymbol{\lambda}) = \lambda_l = P(c = l)$ for $l = 1, \cdots, L$, where $\boldsymbol{\lambda} = \left[\lambda_1, \cdots, \lambda_L\right]$ and $\sum_{l=1}^{L} \lambda_l = 1$, and can be expressed as $c|\boldsymbol{\lambda} \sim \text{Cat}(\boldsymbol{\lambda})$ [10]. The probability mass function of the Categorical distribution can be written as

$$p_{c|\boldsymbol{\lambda}}(c|\boldsymbol{\lambda}) = \prod_{l=1}^{L} \lambda_l^{\delta_{cl}}, \qquad \delta_{cl} = \begin{cases} 1 & c = l \\ 0 & \text{otherwise} \end{cases}. \tag{6}$$

The Categorical distribution parameter $\boldsymbol{\lambda}$ is modeled as a Dirichlet distributed random variable with distribution parameter $\boldsymbol{\alpha}$, $\boldsymbol{\lambda} \sim \text{Dir}(\boldsymbol{\alpha})$. The probability density function of $\boldsymbol{\lambda}$ is defined as

$$p(\boldsymbol{\lambda}) = \frac{1}{B(\boldsymbol{\alpha})} \prod_{l=1}^{L} \lambda_l^{\alpha_l - 1}, \tag{7}$$

where $B(\boldsymbol{\alpha}) = \frac{\prod_{l=1}^{L} \Gamma(\alpha_l)}{\Gamma(\alpha_0)}$ is the multivariate beta function and $\alpha_0 = \sum_{l=1}^{L} \alpha_l$ [10]. The parameter vector $\boldsymbol{\alpha} = \left[\alpha_1, \cdots, \alpha_L\right]$ is non-random, with $\alpha_l > 0 \ \forall \ l$.

The Dirichlet distribution is the conjugate prior for the Categorical distribution, and thus the posterior distribution of $\boldsymbol{\lambda}|c$ is $\text{Dir}(\boldsymbol{\alpha}^\circ)$ where $\boldsymbol{\alpha}^\circ = \left[\alpha_1^\circ, \cdots, \alpha_L^\circ\right]$ and $\alpha_l^\circ = \alpha_l - 1 + \delta_{cl}$ [10], [12]. That is, we can write $\boldsymbol{\lambda}|c = \boldsymbol{\lambda}^\circ \sim \text{Dir}(\boldsymbol{\alpha}^\circ)$. This shows that after getting a new class label $l_s$, the updated estimate of the distribution parameter $\boldsymbol{\lambda}$ is now Dirichlet distributed with parameter $\boldsymbol{\alpha}^\circ$. We call this updated distribution parameter $\boldsymbol{\lambda}^\circ \sim \text{Dir}(\boldsymbol{\alpha}^\circ)$.

The DCM provides an efficient closed-form equation for calculating the posterior predictive distribution [12] of the class state variable $c$ given the label data in $\mathcal{L}$ using

$$\begin{aligned} p_{c|\mathcal{L}}(c|\mathcal{L}) &= \int p_{c|\boldsymbol{\lambda}}(c|\boldsymbol{\lambda}) p_{\boldsymbol{\lambda}|\mathcal{L}}(\boldsymbol{\lambda}|\mathcal{L}) d\boldsymbol{\lambda} \\ &= \int \lambda_c \frac{1}{B(\boldsymbol{\alpha})} \prod_{l=1}^{L} \lambda_l^{\alpha_l - 1 + \sum_{l' \in \mathcal{L}} \delta_{cl'}} d\boldsymbol{\lambda} \\ &= \frac{B(\boldsymbol{\alpha}')}{B(\boldsymbol{\alpha})} \int \text{Dir}(\boldsymbol{\alpha}') d\boldsymbol{\lambda} = \frac{B(\boldsymbol{\alpha}')}{B(\boldsymbol{\alpha})}, \end{aligned}$$

where $\alpha_l' = \alpha_l - 1 + \sum_{l' \in \mathcal{L}} \delta_{cl'}$. Thus, the posterior predictive distribution is also Categorically distributed as $c|\mathcal{L} \sim \text{Cat}(\boldsymbol{\lambda}')$ with $p_{c|\mathcal{L}}(c|\mathcal{L}) = \frac{B(\boldsymbol{\alpha}')}{B(\boldsymbol{\alpha})} = \lambda_c'$. Only one label $l_s$ is added at each time $s$, thus using the recursive property of the Gamma function, $\Gamma(n+1) = n\Gamma(n)$, we see that

$$\begin{aligned} \lambda_c' = \frac{B(\boldsymbol{\alpha}')}{B(\boldsymbol{\alpha})} &= \frac{\Gamma(\alpha_0)}{\prod_{l=1}^{L} \Gamma(\alpha_l)} \frac{\Gamma(\alpha_c + 1) \prod_{l=1 \neq c}^{L} \Gamma(\alpha_l)}{\Gamma(\alpha_0 + 1)} \\ &= \frac{\Gamma(\alpha_0)}{\prod_{l=1}^{L} \Gamma(\alpha_l)} \frac{\alpha_c' \prod_{l=1}^{L} \Gamma(\alpha_l)}{\alpha_0' \Gamma(\alpha_0)} = \frac{\alpha_c'}{\alpha_0'}. \end{aligned} \tag{8}$$

Similar to occupancy grids, which capture the posterior marginal probabilities of occupancy for all grid cells, the posterior predictive class distribution generated by the sequential classification process is captured in classification maps (CMs). We represent a CM as a set of distributions denoted as

$$\boldsymbol{q} = \{p_{c|\mathcal{L}}(c_r|\mathcal{L})\}_{r=1}^{B}, \tag{9}$$

where $c_r$ is the class state variable for the $r$th grid cell $g_r$. Figure 3 depicts the idea behind the proposed class state tracking process. The sequential class updating process takes each new class label $l_s$ as well as the previous distribution parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\alpha}$ to produce the new distribution parameters $\boldsymbol{\lambda}'$ and $\boldsymbol{\alpha}'$, hence allowing us to successively update the parameter estimates for every new measurement.

## V. Information-Theoretic Cost Functions

In this section, we develop a novel information-theoretic cost function that is used in evaluating the utility of sensing actions during navigation policy selection. This cost function, called the *total information gain* at time $s$ $IG_T(s)$, is a convex combination of two individual information-theoretic cost functions,

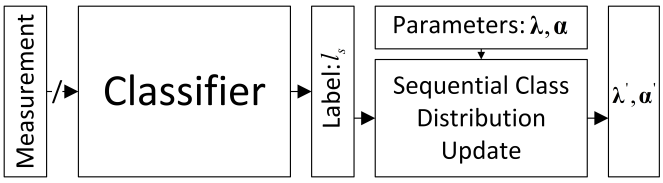



Fig. 3. Block diagram of sequential classification state tracking. A measurement is converted into a class label $l_s$ by means of a one-step classifier. The class label is used in conjunction with the previous distribution parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\alpha}$ to produce a new distribution parameter $\boldsymbol{\lambda}'$ and $\boldsymbol{\alpha}'$.

one for detection $IG_D(s)$ and one for classification $IG_C(s)$. We first present the detection cost function, derived with the sensor model and outputs from the occupancy grid estimation process. The classification cost function is then presented, derived with the sequential classification process using the DCM.

Letting $X$ and $Y$ be two random variables, and $H(\cdot)$ the Shannon entropy, then the information gain is defined as [11]

$$I(X;Y) = H(X) - H(X|Y), \tag{10}$$

which can be thought of as the reduction in uncertainty of the distribution on $X$ that the knowledge of random variable $Y$ would bring. In the active perception problem, $H(X)$ can be thought of as the prior distribution on $X$, while $H(X|Y)$ can be viewed of as the posterior distribution on $X$ after observing random variable $Y$.

In the context of our problem, both the detection and classification cost functions are closed-forms of the information gain, or mutual information, between a state variable $X$ (occupancy state or class state) and a random variable $Y$ that capture information about the latest measurements. As such, they are both predictors of the information that is gained by taking a measurement from a grid cell given the current state of the grid cell. The detection and classification cost functions can be viewed as predicting the theoretical mutual information between the current state and the updated state after performing a sensing action, and aide in the selection of the next sensing action that provides the most information.

In the sequel, we will denote the set of grid cells observed at time $s$ by $\mathcal{G}$.

### A. Detection Cost Function

We define the detection information gain at time $s$, $IG_D(s)$, as the sum of the mutual information between the occupancy state and the measurement random variable for all observed grid cells in $\mathcal{G}$. This is stated mathematically as

$$IG_D(s) \triangleq \sum_{g_i \in \mathcal{G}} I(b_i; j_{s,k}), \tag{11}$$

where $b_i$ is the occupancy state variable for grid cell $g_i$ and $j_{s,k}$ is the random variable representing a measurement at time $s$ and range $k$. The information gain for grid cell $g_i$ is given by,

$$\begin{aligned} I(b_i;j_{s,k}) &= H(b_i) - H(b_i|j_{s,k}) \\ &= -\mathbb{E}_{\mathcal{B}} \log p_b(b_i) - \mathbb{E}_{\mathcal{J}} \log p_{b|j}(b_i|j_{s,k}) \\ &= -\sum_{b\in\mathcal{B}} p_b(b_i) \log p_b(b_i) \\ &\quad - \sum_{b\in\mathcal{B}} \sum_{j\in\mathcal{J}} p_{j|b}(j_{s,k}|b_i) p_b(b_i) \log \frac{p_{j|b}(j_{s,k}|b_i) p_b(b_i)}{p_j(j_{s,k})}. \end{aligned} \tag{12}$$

where $\mathcal{J} = \{0, 1\}$ is the set of possible values that realizations of $j_{s,k}$ can take, and $\mathcal{B} = \{0, 1\}$ is the set of possible values that realizations of $b_i$ can take.

Using the occupancy grid estimation model in (4), the interaction between the occupancy state variable $b_i$ and the measurement random variable $j_{s,k}$ can be represented by,

$$\begin{aligned} p_{j|b}(j_{s,k}|b_i) = &\left[(1-p_{\text{fa}})(1-b_i) + (1-p_{\text{d}})b_i\right](1-j_{s,k}) \\ &+ \left[p_{\text{fa}}(1-b_i) + p_{\text{d}} b_i\right] j_{s,k}, \end{aligned} \tag{13}$$

where $p_{\text{d}}$ and $p_{\text{fa}}$ are the probabilities of detection and false alarm, respectively, for the physical detector that produces the measurement vectors $\boldsymbol{j}_s$.

We treat the output of the occupancy grid estimation $p_{b|\boldsymbol{J}}(b_i|\boldsymbol{J}_{s-1})$, i.e. the *posterior estimate* from the previous time step, as the *prior* $p_b(b_i)$ for time $s$ [15]. To simplify notation, we let $p_i = p_b(b_i = 1)$. Now, using this together with (13), and invoking the total probability, the marginal probability mass function for $j_{s,k}$ is,

$$\begin{aligned} p_j(j_{s,k}) &= \sum_{\beta\in\mathcal{B}} p_{j|b}(j_{s,k}|b_i = \beta) p(b_i = \beta) \\ &= p_{j|b}(j_{s,k}|b_i = 1) p_i + p_{j|b}(j_{s,k}|b_i = 0)(1-p_i) \\ &= \left[(1-p_{\text{d}})(1-j_{s,k}) + p_{\text{d}} j_{s,k}\right] p_i \\ &\quad + \left[(1-p_{\text{fa}})(1-j_{s,k}) + p_{\text{fa}} j_{s,k}\right](1-p_i) \\ &= \left[(1-p_{\text{fa}})(1-p_i) + (1-p_{\text{d}})p_i\right](1-j_{s,k}) \\ &\quad + \left[p_{\text{fa}}(1-p_i) + p_{\text{d}} p_i\right] j_{s,k}, \end{aligned} \tag{14}$$

Using this result, the prior and conditional entropy in (12), become

$$H(b_i) = -\left[p_i \log p_i + (1-p_i)\log(1-p_i)\right]. \tag{15}$$

and

$$\begin{aligned} &H(b_i|j_{s,k}) \\ &= -\Big[(1-p_{\text{fa}})\big(1-p_i\big) \log \frac{(1-p_{\text{fa}})\big(1-p_i\big)}{(1-p_{\text{d}})p_i + (1-p_{\text{fa}})\big(1-p_i\big)} \\ &\quad + p_{\text{fa}}\big(1-p_i\big) \log \frac{p_{\text{fa}}\big(1-p_i\big)}{p_{\text{d}} p_i + p_{\text{fa}}\big(1-p_i\big)} \\ &\quad + (1-p_{\text{d}})p_i \log \frac{(1-p_{\text{d}})p_i}{(1-p_{\text{d}})p_i + (1-p_{\text{fa}})\big(1-p_i\big)} \\ &\quad + p_{\text{d}} p_i \log \frac{p_{\text{d}} p_i}{p_{\text{d}} p_i + p_{\text{fa}}\big(1-p_i\big)}\Big] \\ &= (1-p_{\text{fa}})\big(1-p_i\big) \log\big[1 + (1-p_{\text{d}})p_i\big] \\ &\quad + p_{\text{fa}}\big(1-p_i\big) \log\big[1 + p_{\text{d}} p_i\big] \\ &\quad + (1-p_{\text{d}})p_i \log\big[1 + (1-p_{\text{fa}})\big(1-p_i\big)\big] \\ &\quad + p_{\text{d}} p_i \log\big[1 + p_{\text{fa}}\big(1-p_i\big)\big], \end{aligned} \tag{16}$$

respectively.

Plugging (13), (14), (16), and (15) into (12) gives a closed-form expression for the detection information gain as

$$\begin{aligned} I(b_i; j_{s,k}) = p_i\Big[&(1-p_{\mathrm{d}})\log\big[1+(1-p_{\mathrm{fa}})(1-p_i)\big] \\ &+ p_{\mathrm{d}}\log\big[1+p_{\mathrm{fa}}(1-p_i)\big] - \log p_i\Big] \\ + \big(1-p_i\big)\Big[&(1-p_{\mathrm{fa}})\log\big[1+(1-p_{\mathrm{d}})p_i\big] \\ &+ p_{\mathrm{fa}}\log\big[1+p_{\mathrm{d}}p_i\big] - \log\big[1-p_i\big]\Big]. \end{aligned} \quad (17)$$

### B. Classification Cost Function

We define the classification information gain at time $s$ $IG_C(s)$ as the sum of the mutual information between the class state variable, $c_i$, and the Dirichlet distributed parameter vector, $\boldsymbol{\lambda}$, for all observed grids $g_i \in \mathcal{G}$. This is stated mathematically as

$$IG_C(s) \triangleq \sum_{g_i \in \mathcal{G}} I(\boldsymbol{\lambda}; c_i). \quad (18)$$

The distribution parameter vector $\boldsymbol{\lambda}$ for the Categorical distribution on $c_i$ in essence captures information about the latest measurements.

For a *mixed-pair* of discrete scalar random variable $X$ and continuous random vector $\boldsymbol{Y}$, assuming they satisfy the sufficient conditions to be a *good mixed-pair* [16], their mutual information is,

$$\begin{aligned} I(\boldsymbol{Y}; X) &= h(\boldsymbol{Y}) - h(\boldsymbol{Y}|X) \\ &= -\int p_{\boldsymbol{y}}(\boldsymbol{y}) \log p_{\boldsymbol{y}}(\boldsymbol{y}) d\boldsymbol{y} \\ &\quad + \sum_{x \in \mathcal{X}} \int p_{\boldsymbol{y},x}(\boldsymbol{y}, x) \log p_{\boldsymbol{y}|x}(\boldsymbol{y}|x) d\boldsymbol{y}, \end{aligned} \quad (19)$$

where $h(\cdot)$ is the differential entropy [11].

Applying (19) to (18), the mutual information $I(\boldsymbol{\lambda}; c_i)$ can be evaluated as

$$\begin{aligned} I(\boldsymbol{\lambda}; c_i) &= h(\boldsymbol{\lambda}) - h(\boldsymbol{\lambda}|c_i) \\ &= h(\boldsymbol{\lambda}) + \sum_{c_i=1}^{L} \int_{\Delta^L} p_{\boldsymbol{\lambda},c_i}(\boldsymbol{\lambda}, c_i) \log p_{\boldsymbol{\lambda}|c_i}(\boldsymbol{\lambda}|c_i) d\boldsymbol{\lambda}. \end{aligned} \quad (20)$$

The entropy of a Dirichlet distributed random vector is well-known [17] and can be written as

$$h(\boldsymbol{\lambda}) = \log B(\boldsymbol{\alpha}) + (\alpha_0 - L)\psi(\alpha_0) - \sum_{l=1}^{L}(\alpha_l - 1)\psi(\alpha_l), \quad (21)$$

where $\psi(x) = \frac{d}{dx}\log\Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}$ is the digamma function.

To evaluate the conditional entropy term $h(\boldsymbol{\lambda}|c)$ in (20), we use the fact that the Dirichlet distribution is the conjugate prior of the Categorical distribution. Thus, we can write

$$\begin{aligned} h(\boldsymbol{\lambda}|c) &= -\sum_{c=1}^{L} \int_{\Delta^L} p_{\boldsymbol{\lambda},c}(\boldsymbol{\lambda}, c) \log p_{\boldsymbol{\lambda}|c}(\boldsymbol{\lambda}|c) d\boldsymbol{\lambda} \\ &= -\sum_{c=1}^{L} \int_{\Delta^L} p_{c|\boldsymbol{\lambda}}(c|\boldsymbol{\lambda}) p_{\boldsymbol{\lambda}}(\boldsymbol{\lambda}) \log p_{\boldsymbol{\lambda}|c}(\boldsymbol{\lambda}|c) d\boldsymbol{\lambda} \\ &= -\sum_{c=1}^{L} \frac{B(\boldsymbol{\alpha}')}{B(\boldsymbol{\alpha})} \int_{\Delta^L} \frac{1}{B(\boldsymbol{\alpha}')} \prod_{l=1}^{L} \lambda_l^{\alpha'-1} \log \frac{1}{B(\boldsymbol{\alpha}')} \prod_{l=1}^{L} \lambda_l^{\alpha'-1} d\boldsymbol{\lambda} \\ &= -\sum_{c=1}^{L} \frac{\alpha'_c}{\alpha'_0} E_{\boldsymbol{\lambda}'}\left[\log \frac{1}{B(\boldsymbol{\alpha}')} \prod_{l=1}^{L} \lambda_l^{\alpha'-1}\right] = \sum_{c=1}^{L} \frac{\alpha'_c}{\alpha'_0} h(\boldsymbol{\lambda}') \quad (22) \\ &= \sum_{c=1}^{L} \frac{\alpha'_c}{\alpha'_0}\left[\log B(\boldsymbol{\alpha}') + (\alpha'_0 - L)\psi(\alpha'_0) - \sum_{l=1}^{L}(\alpha'_l - 1)\psi(\alpha'_l)\right] \end{aligned}$$

Combining (18), (21), and (22) provides the information gain for classification as

$$\begin{aligned} IG_C(s) &= \sum_{g_i \in \mathcal{G}} I(\boldsymbol{\lambda}; c_i) \quad (23) \\ &= \sum_{g_i \in \mathcal{G}}\Big[\log B(\boldsymbol{\alpha}) + (\alpha_0 - L)\psi(\alpha_0) - \sum_{l=1}^{L}(\alpha_l - 1)\psi(\alpha_l) \\ &\quad - \sum_{c_i=1}^{L} \frac{\alpha'_c}{\alpha'_0}\Big(\log B(\boldsymbol{\alpha}') + (\alpha'_0 - L)\psi(\alpha'_0) - \sum_{l=1}^{L}(\alpha'_l - 1)\psi(\alpha'_l)\Big)\Big]. \end{aligned}$$

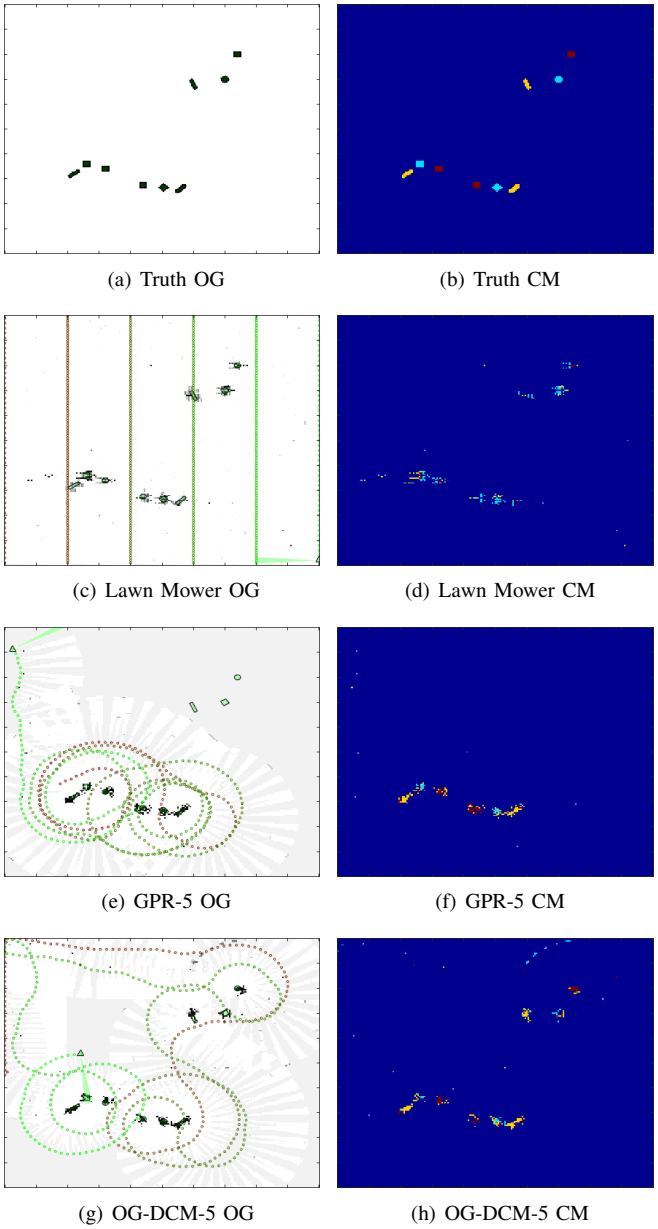
### C. Total Information Gain

As previously stated, the total information gain is defined as the convex combination of the detection and classification information metrics. To ensure that one information metric does not dominate the other at all times $s$ due to scaling, we normalize them by their respective maximal values,

$$IG_T(s) = w_D \frac{IG_D(s)}{IG_{D_{\max}}} + w_C \frac{IG_C(s)}{IG_{C_{\max}}}, \quad w_D + w_C = 1. \quad (24)$$

The values for $IG_{D_{\max}}$ and $IG_{C_{\max}}$ are calculated as the sum of the maximum information gain available from a grid cell over all grid cells $g_i \in \mathcal{G}$ for the detection and classification states, respectively. The information gain for an individual grid cell is maximized when the prior distribution is uniform and the posterior distribution is a delta, implying that the maximum value is equal to the entropy of a uniform random variable. Hence, $IG_{D_{\max}} = |\mathcal{G}|\log(2)$ and $IG_{C_{\max}} = |\mathcal{G}|\log(L)$ where $|\mathcal{G}|$ is the cardinality of the set $\mathcal{G}$. The values for $IG_{D_{\max}}$ and $IG_{C_{\max}}$ are considered constant across all times $s$, as the distribution of the state variables could be uniform at any time $s$ (whether or not they have been fluctuating at all times $< s$) and then completely determined at time $s+1$ thus achieving the maximal information gain.

This convex weighting allows for different strategies to be employed by the system. For example, at the beginning of a sortie, there will likely be insufficient information to consider classification for choosing sensing actions. In this case, we can assign a higher weight (e.g., $w_D = 0.9$) for the detection while choosing a lower weight (e.g., $w_C = 1 - w_D = 0.1$) for the classification. In contrast, once most of the grid cells are observed, there may be little to no information left in performing target detection and localization, in which case we can choose the sensing actions primarily based on the classification criterion by choosing $w_C = 0.9$ and $w_D = 1 - w_C = 0.1$.



(a) Truth OG (b) Truth CM

(c) Lawn Mower OG (d) Lawn Mower CM

(e) GPR-5 OG (f) GPR-5 CM

(g) OG-DCM-5 OG (h) OG-DCM-5 CM

Fig. 4. Occupancy grids (OG) on left and classification maps (CM) on right. OG and CM shown for: the underlying truth, lawn mower, GPR-5, and OG-DCM-5. Cylinders are colored yellow, cubes light blue, and spheres red. The deep blue color indicates no target. Each figure shows the same $50 \times 50$ meter area.

## VI. Trajectory Planning

In this section, we discuss three types of trajectory-planning policies for performing interactive sensing and navigation. Recall that the action at each time step is the selection of the next location for the sensor platform. For all policies, let $\boldsymbol{a}$ denote an action, and $\mathcal{A}_s$ be the set of all feasible actions the sensor can take under vehicle dynamical constraints at time step $s$. The three policies choose sensing locations according to: a pre-determined lawn mower path; and maximizing a cost function for choosing $\boldsymbol{a} \in \mathcal{A}_s$ in a greedy, and non-greedy manner.

In many traditional underwater target detection and classification operations, a predetermined lawn mower path, as shown in Figure 4(c), is used. The sequence of prespecified actions $\boldsymbol{a}$ guarantees that each region in the search area is observed though at the cost of potentially poor detection and classification accuracy due to inadequate viewing angles.

The greedy policy for action selection chooses the action as the one that maximizes the one-step *reward*, or one-step navigation cost function, i.e.

$$\boldsymbol{a}^*_{s+1} = \text{argmax}_{\boldsymbol{a} \in \mathcal{A}_s} R_s(\boldsymbol{b}_s, \boldsymbol{c}_s, \boldsymbol{a}), \tag{25}$$

where $R_s(\boldsymbol{b}_s, \boldsymbol{c}_s, \boldsymbol{a}) = IG_T(s)$ is the immediate reward for performing action $\boldsymbol{a}$ with the current state variable distributions at time $s$, where $\boldsymbol{b}_s$ and $\boldsymbol{c}_s$ are the occupancy state and the class state of the system at time $s$, respectively.

The non-greedy policy for action selection chooses the next sensing action that maximizes the one-step reward, while also considering the reward from future actions along a finite horizon of $T$ future actions. This policy, in essence, chooses the next sensing action that it believes will generate the largest cumulative reward given the current information available. This type of problem is typically cast as a partially observable Markov decision problem (POMDP), which admits a number of solution methods [18]–[20]. In this paper, the *rollout policy* method [19] is used to solve the problem given our choice of a heuristic navigation cost function $IG_T(s)$.

Mathematically speaking, this policy can be described as choosing the sensing action that maximizes the one-step reward plus the *expected reward-to-go* associated with a prespecified policy, typically a heuristic rule. The decision rule for action selection used by the rollout policy is defined as

$$\boldsymbol{a}^*_{s+1} = \text{argmax}_{\boldsymbol{a} \in \mathcal{A}_s} R_s(\boldsymbol{b}_s, \boldsymbol{c}_s, \boldsymbol{a}) + E_{s+1}, \tag{26}$$

where $R_s(\boldsymbol{b}_s, \boldsymbol{c}_s, \boldsymbol{a})$, $\boldsymbol{b}_s$, and $\boldsymbol{c}_s$ are the same as the greedy case, $E_{s+1} = \mathbb{E}\left[\sum_{i=s+1}^{T} R_i(\boldsymbol{b}_i, \boldsymbol{c}_i, \boldsymbol{a}_i) | \boldsymbol{b}_s, \boldsymbol{c}_s\right]$ is the expected reward-to-go, $R_i(\boldsymbol{b}_i, \boldsymbol{c}_i, \boldsymbol{a}_i) = IG_T(i)$ is the reward for performing action $\boldsymbol{a}_i$ with the current state variable distributions at time $i \in [s+1, T]$, where $\boldsymbol{b}_i$ and $\boldsymbol{c}_i$ are the occupancy state and the class state of the system at time $i$, respectively, and $T$ is the length of the finite horizon.

## VII. Experimental Results

In this section, we present the simulation results from autonomous navigation experiments utilizing the action selection policies described in Section VI which use the total information gain as the reward function. The experiments conducted in this section expose each action selection policy's ability to simultaneously perform detection, localization, and classification of targets, while exploring new areas. For the remainder of this section, we call our proposed method for calculating $IG_T(s)$ the occupancy grid Dirichlet Categorical model (OG-DCM).

### A. Benchmark Method

As shown in the previous sections, the theoretical representation for calculating the information gained from each measurement utilizes the sequential detection and classification state estimates. As such, there is no appropriate non-deterministic benchmark method other than data driven function approximation methods. Although, many function approximation methods such as neural networks and deep learning

architectures exist, the Gaussian progress regression (GPR) has been shown to be successful in approximating the mutual information for the active perception problems [3], [5]. Thus, here we benchmark OG-DCM to that of a GPR for estimating the information gain expected from both the detection and classification components, $IG_D(s)$ and $IG_C(s)$, respectively. The ground truth outputs used to train the GPR were the detection and classification information gains, calculated as the difference between the entropy of prior and posterior state distributions for detection and classification, respectively [11]. The GPR estimates mean and associated covariance of an input data vector using an appropriate kernel function and pairs of training data and their corresponding outputs. In our problem, the current position and previous detection and classification mutual information compose the input vector. The output, estimated by the GPR, is a composite vector of the detection and classification mutual information calculated for the current sensing action. We trained the Gaussian process with a Matérn kernel function [5] and utilized K-D trees [3] to find the nearest $100$ neighbors for forming the covariance matrix on a per-cell basis.

The total information gain is approximated by the Gaussian process regression following the Gaussian process upper bound confidence algorithm [5], [21], and is given by

$$IG_{GPR}(s) = w_D \frac{\mu_D(\boldsymbol{x}) + \beta\sigma_D(\boldsymbol{x})}{IG_{D_{\max}}} + w_C \frac{\mu_C(\boldsymbol{x}) + \beta\sigma_C(\boldsymbol{x})}{IG_{C_{\max}}},$$

where $\beta$ is the tradeoff parameter between exploration and exploitation, $\mu_\cdot(\boldsymbol{x})$ and $\sigma_\cdot(\boldsymbol{x})$ are the predicted mean and variance for the detection and classification components, respectively, derived from the GPR.

### B. Experimental Data & Description

In our active perception problem, a sonar platform is used to mimic the behavior of an autonomous underwater vehicle (AUV) that is searching littoral zones for mine-like underwater targets. The system is equipped with multiple (11) hydrophones arranged in a uniform linear array (ULA), all pointing slightly downwards from horizontal (positive depression angle). The transmitted waveform was a linearly frequency modulated (LFM) chirp with center frequency $f_c$ = 80 kHz, bandwidth $BW$ = 20 kHz, and sampling frequency $f_s$ = 60 kHz. The ULA has a 7° beamwidth with an interrogation range up to tens of meters. The sensor is attached to a platform, which is 10 meters above the seafloor.

The experiments use simulated side-looking sonar (SLS) that directs acoustic radiation to the starboard side of the AUV. The sonar data is generated by a cutting-edge, physics-based sonar simulator namely the Personal Computer Shallow Water Acoustic Toolset (PC SWAT) [22]. PC SWAT models scattering from the target by a combination of the Kirchhoff approximation and the geometric theory of diffraction. Propagation of sound into a marine sediment with ripples is described by an application of Snell's law and second order perturbation theory in terms of Bragg scattering [22]. PC SWAT has been used to produce simulations providing *exemplar* template measurements that closely match real data generated by sonar systems [23].

The stave data generated by PC SWAT was fed through an adaptive coherence estimator (ACE) detector [24]–[26] to produce a single beamformed measurement (detection statistics) vector. This beamformed vector is thresholded at a predetermined value to provide the measurement vector $\boldsymbol{j}_s$. The threshold is chosen to yield a desired $p_{\text{fa}}$ and hence $p_{\text{d}}$.

A total of 500 pings (actions) with 1 meter ping separation were simulated for each experiment. Nine targets, in three clusters of three different targets (i.e. $L$ = 3), are proud on the seafloor within the $50 \times 50$ meter search field. Medium sandy bottom was used in all the experiments to simulate bottom clutter. Each cluster contains a 2 meter long cylindrical target with a radius of $0.25$ meters, a 1 meter$^3$ cubic target, and a partially hollow sphere with 1 meter radius. The spatial orientation of the three clusters can be seen in the classification map (CM) of Figure 4(b), where the cylinders are color-coded yellow, the cubes light blue, and the spheres red. The deep blue color indicates no target.

### C. Evaluation Metrics

To evaluate the performance of each policy, three different metrics are used, as detailed below. For all metrics, with the exception of the percentage of grid cells observed, we evaluate the performance for detection (occupancy grids) and classification (classification maps) separately to better illustrate the strengths and weaknesses of each. In the following, $\boldsymbol{t}$ represents the true set of distributions, either true occupancy $\boldsymbol{b}$ or true classification $\boldsymbol{c}$, and $\boldsymbol{e}$ represents the estimated set of distributions, either occupancy grid $\boldsymbol{p}$ or classification map $\boldsymbol{q}$. Only like pairs are compared, i.e., $\boldsymbol{b}$ and $\boldsymbol{p}$ or $\boldsymbol{c}$ and $\boldsymbol{q}$. The true distributions are formed from delta functions (e.g., $p_{b|\boldsymbol{J}}(b_r|\boldsymbol{J}_S) = [1, 0]$ if a cell is occupied, and $p_{c|\mathcal{L}}(c_r|\mathcal{L}) = [0, 0, 1, 0]$ if a cell is of class 3).

1) Similarity between the true distribution $\boldsymbol{t}$ to that of the estimated distribution $\boldsymbol{e}$:

$$\rho = \frac{\langle \boldsymbol{t}, \boldsymbol{e} \rangle_F}{||\boldsymbol{t}||_F^2 ||\boldsymbol{e}||_F^2},$$

where $||\cdot||_F$ is the Frobenius norm, and $\langle \cdot, \cdot \rangle_F$ is the Frobenius inner product. For calculating this metric, we form $\boldsymbol{t}$ and $\boldsymbol{e}$ into matrices by making each row the vectorized form of the distribution for each grid cell. Clearly, $0 \le \rho \le 1$, and $\rho = 1$ when $\boldsymbol{t} = \boldsymbol{e}$

2) Sum of the Jensen-Shannon distance (SJSD) $D_{\text{JS}}(t_i||e_i)$ [27] over all $i = 1, \cdots, B$ grid cells:

$$\begin{aligned}\text{SJSD} &= \sum_r D_{\text{JS}}(t_i||e_i) \\ &= \sum_i \frac{1}{2} D_{\text{KL}}(t_i||m_i) + \frac{1}{2} D_{\text{KL}}(e_i||m_i) \\ &= -\frac{1}{2} \sum_i \Big[ \sum_{x \in \mathcal{X}} t_i(x) \log\Big(\frac{m_i(x)}{t_i(x)}\Big) + e_i(x) \log\Big(\frac{m_i(x)}{e_i(x)}\Big) \Big],\end{aligned}$$

where $m_i(x) = \frac{1}{2}\big(t_i(x) + e_i(x)\big)$, $t_i(x)$ and $e_i(x)$ are the distributions for grid cell $g_i$ evaluated at point $x$, and $D_{\text{KL}}(\cdot, \cdot)$ is the Kullback-Leibler (KL) divergence [27]. The Jensen-Shannon distance is used in favor of the KL divergence as it is symmetric, positive, and always finite. The maximum value of SJSD is $\log(2) \times B$, with smaller

values indicating that $\boldsymbol{t}$ and $\boldsymbol{e}$ are similar and $\boldsymbol{t} = \boldsymbol{e}$ when SJSD is 0.

3) Percentage of the grid cells in the map that are observed during the experiment. This measure provides the efficiency with which the AUV is able to explore new areas while estimating the occupancy grid.

The true distributions are illustrated in Figure 4(a) and 4(b). In Figure 4(a), occupied grid cells are black while empty grid cells are white. In Figure 4(b), each grid cell is color-coded according to the non-zero component of its distribution (i.e., cylindrical class grid cells are yellow, etc.).

### D. Results and Discussion

In the experiments conducted here, OG-DCM utilized the modified version of the matched subspace classifier (MSC) [28] as the one-step classifier in Figure 3. The modified MSC (MMSC) was used owing to the proven success in classifying underwater objects [29]. Additionally, the MMSC has many desirable properties including the ability to use any learned subspace dictionaries, and incremental updating of the dictionary matrices when operating in new measurements [30].

At each time step, a sensing action in the form of selecting and moving to the next position to collect a measurement is taken. The measurement is collected and used to update the occupancy grid and classification map. The next location from which to take a measurement was chosen according to (25) and (26) for the greedy and non-greedy policies, respectively. Each reward, $R(\boldsymbol{b}_s, \boldsymbol{c}_s, \boldsymbol{a}) = IG_T(s)$ for OG-DCM and $R(\boldsymbol{b}_s, \boldsymbol{c}_s, \boldsymbol{a}) = IG_{GPR}(s)$ for GPR, was calculated by first generating the ping from the new location with PC SWAT, then performing the occupancy grid estimation and classification map estimation, and finally calculating the reward for that action. The non-greedy policy was evaluated to time step $s+T$ for different finite horizon lengths of $T = 0$ (greedy), $T = 5$ and $T = 10$ for both GPR and OG-DCM methods. That is, each decision involves considering $T$ time steps into the future. We denote the finite horizon policies as GPR-$T$ and OG-DCM-$T$, i.e., GPR-0 and OG-DCM-0 are for $T = 0$, etc. Thus, including the lawn mower policy, a total of 7 different experiments were conducted. Each experiment uses a different action selection policy under different configuration, i.e., different estimation methods for the navigation cost function.

For each type of non-deterministic policy, i.e., greedy, and non-greedy, 20 different trials were conducted where the starting locations and headings of the AUV were randomly chosen in each trial. The lawn mower policy was only executed once. Table I gives the mean values of SJSD, $\rho$, and the percentage of grid cells observed after 500 sensing actions. Bold values in each column of the table represent the best performance for the metric associated with that column.

As seen from these results the OG-DCM non-greedy policy with a $T = 10$ step finite horizon outperformed all other non-deterministic policies in all metrics. The lawn mower policy outperformed OG-DCM-10 in the number of observed grid cells and the SJSD for detection. These two metrics can be thought of as measuring the same information, as an increase in the number of grid cells seen implies a relatively good estimate of occupancy for any reliable detector. Note that

TABLE I
SJSD AND $\rho$ FOR DETECTION (DET.) AND CLASSIFICATION (CLASS.), AND % OF GRID CELLS SEEN FOR DIFFERENT NAVIGATION POLICIES AFTER 500 SENSING ACTIONS. **BOLD** VALUES INDICATE BEST PERFORMANCE PER METRIC.

| Policy | % Seen | SJSD Det. | SJSD Class. | $\rho$ Det. | $\rho$ Class. |
|---|---|---|---|---|---|
| Lawn Mower | **0.97** | **157.5** | 69.2 | **0.52** | 0.58 |
| GPR-0 | 0.10 | 764 | 66.5 | 0.22 | 0.61 |
| GPR-5 | 0.45 | 542 | 68.6 | 0.26 | 0.58 |
| GPR-10 | 0.41 | 558.1 | 70.83 | 0.15 | 0.57 |
| OG-DCM-0 | 0.35 | 604.7 | 62.7 | 0.38 | 0.64 |
| OG-DCM-5 | 0.62 | 442.1 | 57.7 | 0.47 | 0.67 |
| OG-DCM-10 | 0.62 | 427.1 | **48.7** | **0.52** | **0.75** |

the lawn mower policy is designed to observe as many grid cells as possible, which is why it achieves an almost perfect score in the percent-seen metric, while matching the best non-deterministic policy in detection performance $\rho$. The accuracy of the occupancy grid, as measured by $\rho$ for detection, is much higher for the percentage of grid cells observed when comparing the OG-DCM to the lawn mower policy, indicating that it extracts more information per grid cell it observes. In other words, because there is a tradeoff between $\rho$ and the percentage of cells seen, showing that OG-DCM-10 outperforms the lawnmower scheme entails comparing *both*. The tradeoff clearly favors OG-DCM-10.

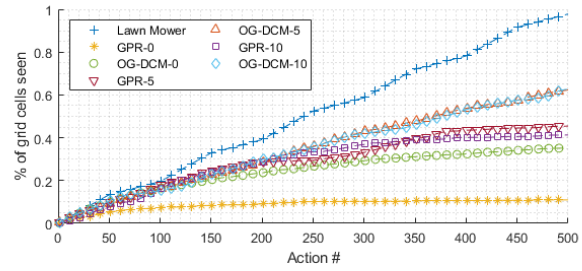
The policies that used OG-DCM generally outperformed those using GPR. In fact, there are only a few cases where even the greedy policy using OG-DCM performed worse than the best non-greedy policy using GPR. This shows that the proposed OG-DCM method provides better estimation of the information gain for future sensing actions than that of GPR.

The OGs and CMs illustrated in Figures 4(c)-4(h) show one realization of the results for the lawn mower, and non-greedy policy experiments. Realizations for finite horizons of $T = 0$ and $T = 10$ have been omitted due to space limitation. The policies that use GPR wound up falling into local minima, i.e., over-observing a particular region without extracting any new detection or classification information, and thus reduced the overall efficiency of the system. Again, as can be observed the OG-DCM method generally provided much better classification results when compared with the GPR results.

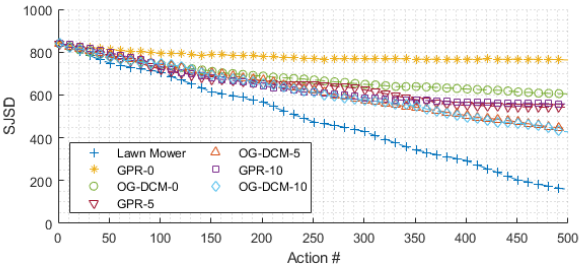
Finally, to compare the temporal evolution of these metrics during the navigation the mean value of each metric is plotted in Figures 5(b)-5(a) against the action number for each case. A major take away, illustrated in Figure 5(c), 5(d), and 5(e), is that OG-DCM-10 not only outperforms the other policies, but does so in a relatively small number of sensing actions. From these results, one can conclude that as more sensing actions are taken, and more grid cells are observed, the OG-DCM-10 outperforms all other policies in all metrics.

## VIII. CONCLUSION

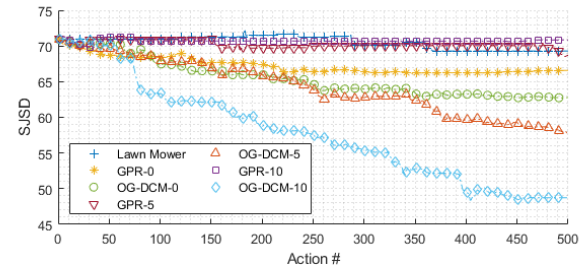
An autonomous navigation system using a novel information-theoretic cost function is proposed based on the outputs of two state tracking algorithms, namely the ones for object detection and classification. This navigation cost function provides a way to compare multiple locations from which a sensor can take measurements and declare which of those locations



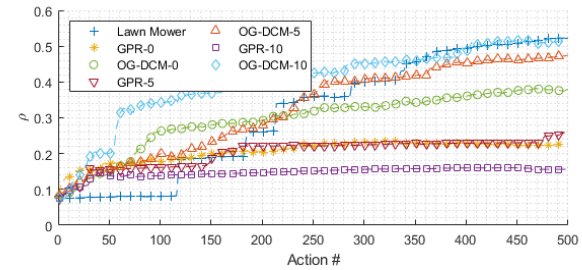

(a) % of grid cells seen



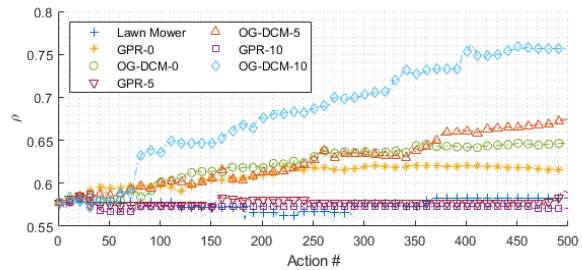

(b) SJDS for detection




(c) SJDS for classification




(d) $\rho$ for detection




(e) $\rho$ for classification

Fig. 5. Performance plots for each of the navigation policies used in experiments.

provide maximal information gain for updating state variable estimates. The performance of three navigation policies, using the proposed cost function, were evaluated and compared to the ground truth. The experimental results show that the use of the proposed information-theoretic cost function along with the non-greedy policy produces the most accurate occupancy grid estimates and target classification while observing more of the map in the same duration of time.

## REFERENCES

[1] R. Bajcsy, "Active perception," *Proceedings of the IEEE*, vol. 76, no. 8, pp. 966–1005, 1988.

[2] P. Whaite and F. P. Ferrie, "Autonomous exploration: Driven by uncertainty," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 3, pp. 193–205, 1997.

[3] G. A. Hollinger, B. Englot, F. S. Hover, U. Mitra, and G. S. Sukhatme, "Active planning for underwater inspection and the benefit of adaptivity," *The International Journal of Robotics Research*, vol. 32, no. 1, pp. 3–18, 2013.

[4] F. Bourgault, A. A. Makarenko, S. B. Williams, B. Grocholsky, and H. F. Durrant-Whyte, "Information based adaptive robotic exploration," in *IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, vol. 1. IEEE, 2002, pp. 540–545.

[5] S. Bai, J. Wang, K. Doherty, and B. Englot, "Inference-enabled information-theoretic exploration of continuous action spaces," in *Robotics Research*. Springer, 2018, pp. 419–433.

[6] B. J. Julian, S. Karaman, and D. Rus, "On mutual information-based control of range sensing robots for mapping applications," *Int. J. Robotics Research*, vol. 33, no. 10, pp. 1375–1392, 2014.

[7] A. Elfes, "Using occupancy grids for mobile robot perception and navigation," *Computer*, vol. 22, no. 6, pp. 46–57, 1989.

[8] S. Thrun, "Learning occupancy grid maps with forward sensor models," *Autonomous Robots*, vol. 15, no. 2, pp. 111–127, 2003.

[9] C. Robbiano, E. K. P. Chong, L. L. Scharf, M. R. Azimi-Sadjadi, and A. Pezeshki, "Bayesian learning of occupancy grids," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–12, 2020. [Online]. Available: https://doi.org/10.1109/TITS.2020.3019813

[10] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.

[11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 2012.

[12] S. Tu, "The dirichlet-multinomial and dirichlet-categorical models for bayesian inference," *Computer Science Division, UC Berkeley*, 2014.

[13] C. M. Bishop, *Pattern recognition and machine learning*. Springer, 2006.

[14] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[15] D. V. Lindley, "The philosophy of statistics," *Journal of the Royal Statistical Society. Series D (The Statistician)*, vol. 49, no. 3, pp. 293–337, 2000. [Online]. Available: http://www.jstor.org/stable/2681060

[16] C. Nair, B. Prabhakar, and D. Shah, "On entropy for mixtures of discrete and continuous variables," *arXiv preprint cs/0607075*, 2006, https://arxiv.org/abs/2007.05072v1.

[17] N. Ebrahimi, E. S. Soofi, and S. Zhao, "Information measures of dirichlet distribution with applications," *Applied Stochastic Models in Business and Industry*, vol. 27, no. 2, pp. 131–150, 2011.

[18] D. P. Bertsekas and D. A. Castanon, "Rollout algorithms for stochastic scheduling problems," *J. Heuristics*, vol. 5, no. 1, pp. 89–108, 1999.

[19] E. K. P. Chong, C. M. Kreucher, and A. O. Hero, "Partially observable markov decision process approximations for adaptive sensing," *Discrete Event Dynamic Systems*, vol. 19, no. 3, pp. 377–422, 2009.

[20] J. C. Goodson, B. W. Thomas, and J. W. Ohlmann, "A rollout algorithm framework for heuristic solutions to finite-horizon stochastic dynamic programs," *Eur. J. Operational Research*, vol. 258, no. 1, pp. 216–229, 2017.

[21] N. Srinivas, A. Krause, S. Kakade, and M. Seeger, "Gaussian process optimization in the bandit setting: no regret and experimental design," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, 2010, pp. 1015–1022.

[22] G. S. Sammelmann, "Propagation and scattering in very shallow water," in *MTS/IEEE Oceans 2001. An Ocean Odyssey. Conference Proceedings (IEEE Cat. No. 01CH37295)*, vol. 1. IEEE, 2001, pp. 337–344.

[23] R. Lim, "Data and processing tools for sonar classification of underwater uxo," *SERDP MR-2230*, 2015.

[24] L. L. Scharf and L. T. McWhorter, "Adaptive matched subspace detectors and adaptive coherence estimators," in *Conf. Rec. 13th Asilomar Conf. Signals, Systems and Computers*. IEEE, 1996, pp. 1114–1117.
[25] S. Kraut, L. L. Scharf, and L. T. McWhorter, "Adaptive subspace detectors," *IEEE Trans. Signal Processing*, vol. 49, no. 1, pp. 1–16, 2001.
[26] S. Kraut, L. L. Scharf, and R. W. Butler, "The adaptive coherence estimator: A uniformly most-powerful-invariant adaptive detection statistic," *IEEE Trans. Signal Processing*, vol. 53, no. 2, pp. 427–438, 2005.
[27] J. Lin, "Divergence measures based on the Shannon entropy," *IEEE Trans. Information theory*, vol. 37, no. 1, pp. 145–151, 1991.
[28] A.-B. Salberg, A. Hanssen, and L. L. Scharf, "Robust multidimensional matched subspace classifiers based on weighted least-squares," *IEEE transactions on signal processing*, vol. 55, no. 3, pp. 873–880, 2007.
[29] J. J. Hall, M. R. Azimi-Sadjadi, S. G. Kargl, Y. Zhao, and K. L. Williams, "Underwater unexploded ordnance (uxo) classification using a matched subspace classifier with adaptive dictionaries," *IEEE Journal of Oceanic Engineering*, vol. 44, no. 3, pp. 739–752, 2019.
[30] P. Pakrooh, L. L. Scharf, and M. R. Azimi-Sadjadi, "Underwater target classification using a pose-invariant matched manifold classifier," in *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2016, pp. 1–5.

**Edwin K. P. Chong** Edwin K. P. Chong (F'04) received the B.E. degree with First Class Honors from the University of Adelaide, South Australia, in 1987; and the M.A. and Ph.D. degrees in 1989 and 1991, respectively, both from Princeton University, where he held an IBM Fellowship. He joined the School of Electrical and Computer Engineering at Purdue University in 1991, where he was named a University Faculty Scholar in 1999. Since August 2001, he has been a Professor of Electrical and Computer Engineering and Professor of Mathematics at Colorado State University. He coauthored the best-selling book, *An Introduction to Optimization* (4th Edition, Wiley-Interscience, 2013).

Prof. Chong received the NSF CAREER Award in 1995 and the ASEE Frederick Emmons Terman Award in 1998. He was a co-recipient of the 2004 Best Paper Award for a paper in the journal *Computer Networks*. In 2010, he received the IEEE Control Systems Society Distinguished Member Award. He was the founding chairman of the IEEE Control Systems Society Technical Committee on Discrete Event Systems, and served as an IEEE Control Systems Society Distinguished Lecturer. He was a Senior Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL. He was the General Chair for the 2011 Joint 50th IEEE Conference on Decision and Control and European Control Conference. He has served as a member of the IEEE Control Systems Society Board of Governors and as Vice President for Financial Activities until 2014. He currently serves as President (2017).

**Christopher Robbiano** Christopher Robbiano received a BS in physics and a BS in electrical engineering, a MS in electrical engineering, and a PhD in electrical engineering from Colorado State University in 2011, 2017, and 2020, respectively. His recent research has been in the areas of detection, classification, and autonomy tasks for sonar applications. He currently works as a Senior Systems Engineer at Ball Aerospace. His current research interests include machine learning, autonomous systems, statistical signal processing, and applications of the aforementioned topics to infrared imaging systems.

**Mahmood R. Azimi-Sadjadi** Dr. Azimi-Sadjadi received his M.S. and Ph.D. degrees from the Imperial College of Science & Technology, University of London, England in 1978 and 1982, respectively, both in Electrical Engineering with specialization in Digital Signal/Image Processing.

He is currently a full professor at the Electrical and Computer Engineering Department at Colorado State University (CSU). He is also serving as the director of the Digital Signal/Image Laboratory at CSU. His main areas of interest include statistical signal and image processing, machine learning and adaptive systems, target detection, classification and tracking, sensor array processing, and distributed sensor networks.

Dr. Azimi-Sadjadi served as an Associate Editor of the IEEE Transactions on Signal Processing and the IEEE Transactions on Neural Networks. He is a Life Member of the IEEE.