PASP: Policy Based Approach for Sensor Planning

Sankalp Arora[1] and Sebastian Scherer[1]

***Abstract*— Capabilities of mobile autonomous systems is often limited by the sensory constraints. Range sensors moving in a fixed pattern are commonly used as sensing modalities on mobile robots. The performance of these sensors can be augmented by actively controlling their configuration for minimizing the expected cost of the mission. The related information gain problem in NP hard. Current methodologies are either computationally too expensive to run online or make simplifying assumptions that fail in complex environments.**

We present a method to create and learn a policy that maps features calculated online to sensory actions. The policy developed in this work actively controls a nodding lidar to keep the vehicle safe at high velocities and focuses the sensor bandwidth on gaining information relevant for the mission once safety is ensured. It is validated and evaluated on an autonomous full-scale helicopter (Boeing Unmanned Little Bird) equipped with an actively controlled nodding laser. It is able to keep the vehicle safe at its maximum operating velocity, $56\ m/s$, and reduce the landing zone evaluation time by 500% as compared to passive nodding. The structure of the policy and efficient learning algorithm should generalize to provide a solution for actively controlling a sensor for keeping a mobile robots safe while exploring regions of interest to the robot.

I. Introduction

Navigation through partially known environments, localization, mapping and manipulation of objects etc. are all tasks for which the robot is expected to detect objects, free space or features in the environment. The rate and accuracy of detection of information of interest dictate the performance of the robotic system. Sensors to detect such features of the environment are often limited by their properties like field of view (FOV), resolution, sampling rate and signal to noise ratios. These sensors are either fixed [1], [2], [3] with respect to the vehicle or in some cases move in constant pre-computed patterns [4], [5]. The capabilities of such sensors and the robotic system can be augmented by actively controlling the sensor configuration to minimize the cost of completing the task assigned to the robot.

The problem of active perception is a well studied one and finds its roots in sequential hypothesis testing [6]. Active perception and adaptive sampling problems have since been extended to account for mobile sensing within the framework of Bayesian reasoning [7]. Later works have resulted in various of solutions that incorporate information theoretic measures for problems like object recognition, mapping, and scene reconstruction [8]. Gradient based methods for next best view and belief space planning have improved [9], [10]. While such algorithms have shown to be useful in their respective applications, they typically rely on restrictive assumptions on the representations, objective functions and do not have guarantees on global optimality. Finite-horizon model predictive control methods [11], [12] provide improvement over myopic techniques, but they do not have performance guarantees beyond the horizon depth and the run times to operate online in large state-spaces. POMDP based solvers [13],suffer from the same curse of dimensionality. The recursive greedy algorithm [14], Branch and bound [15] use the budget to restrict the search space. But require computation exponential in the size of the problem instance due to the large blowup in the search space with increasing budget.

Fig. 1: Application Scenario: Top Left – Test Vehicle, Boeing Unmanned Littlebird. Top Right – Near Earth Autonomy M3 sensor suite, equipped with a actively controllable high range laser. Bottom – Example mission scenario, the vehicle is suppose to autonomously navigate to a pre-defined landing zone at high speeds, evaluate it and decide to whether to land or not while ensuring safety.

We overcome the curse of dimensionality by searching for a policy to actively control the sensor configuration. The policy function is learnt such that it maximizes the gain of information that is important for the completion of the mission of the robot while ensuring its safety as it navigates through unknown environments. We implement the policy on a full-scale autonomous helicopter (Boeing Unmanned Little Bird) equipped with an actively controlled nodding lidar, using occupancy grid map as a world representation. The policy optimizes the use of sensor bandwidth to enable safe, high speed navigation and fast detection of landing zones. The main contributions of the paper are as follows:

[1]The Robotics Institute, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA asankalp, basti@cmu.edu

- Computational complexity analysis for calculating expected information gain for a range sensor and an occupancy grid map representation .
- Policy function and feature design to allow for online active sensor control that allows the vehicle to stay safe at high speeds.
- Results from online evaluation on an autonomous full sized helicopter that show guaranteed safe autonomous navigation at speeds of 56m/s and straight in approaches to landing zones, eliminating the need to hover.

The task of enabling safe, high speed flight enforces tight temporal constraints on the sensor controller motivating the reactive, online approach. In the next section we formally setup the problem, section III presents with the approach overview. Section IV, section V cover the implementation details including construction and learning of policy. Results of evaluation of the performance of the algorithm are discussed in section VII.

II. Problem Setup

In this section we formulate the problem as that of minimization of expected cost of traversal from initial to goal state, through optimization of sensor trajectory. This problem is then shown to be the same as maximizing the gain of information that minimizes the cost of traversal. This problem formulation motivates and guides our approach.

Let the robot's state space be $\mathcal{X} \subset \mathbb{R}^n$. Let $\sigma : [0, T] \rightarrow \mathcal{X}$ be the state space trajectory and $C : \mathcal{X} \rightarrow \mathbb{R}$ be the cost function, where T is the time horizon. The boundary values are $\sigma(0) = \sigma_0$ and $\sigma(T) = \sigma_f$. Let the dynamics constraints on the robot be given by $h(\sigma(t), \dot{\sigma}(t), \ddot{\sigma}(t)) \leq 0$. The cost of the trajectory in a fully deterministic environment is $\int_0^T C(\sigma(t))\,\mathrm{dt}$.

Operating in partially known, unstructured environments the robot has to decide on its next action based on its current belief. Let the high dimensional state space of the world be $\mathcal{W} \subset \mathbb{R}^m$, where world includes the uncertainty of the robot about its environment and it's pose. Let, belief be a probability distribution over the state of the world, $b : \mathcal{W} \rightarrow \{0, 1\}$. Let's assume the belief of the robot at the start of the mission is b_0. The belief of the robot changes as observations are made using a sensor. Let the sensor's state space be $\mathcal{S} \subset \mathbb{R}^s$. Let $\sigma_s : [0, T] \rightarrow \mathcal{S}$ be the sensor trajectory.

The cost function changes as the belief of the robot evolves. To highlight this fact we represent the cost functional as $C_b : \mathcal{X} \times \mathcal{W} \rightarrow \mathbb{R}$. Due to the stochasticity of the belief of the robot, it can only reason about expected cost of its policy, $\mathbb{E}_{p(b|\sigma,\sigma_s,b_o)}\, C_b(.)$. The distribution of belief trajectories $p(b|\sigma, \sigma_s, b_o)$ at any given time is dependent on both state and sensor trajectory. Let the dynamics constraint on the motion of the sensor be given by $h_s(\sigma_s(t), \dot{\sigma}_s(t), \ddot{\sigma}_s(t)) \leq 0$. The full optimization problem can then be defined as follows.

$$\underset{\sigma(t),\sigma_s(t)}{\arg\min} \underset{p(b|\sigma,\sigma_s,b_o)}{\mathbb{E}} \int_0^T C_b(\sigma(t), \sigma_s(t), b(t))\,\mathrm{dt} \qquad (1)$$

$$h(\sigma(t), \dot{\sigma}(t), \ddot{\sigma}(t)) \leq 0$$
$$h_s(\sigma_s(t), \dot{\sigma_s}(t), \ddot{\sigma_s}(t)) \leq 0$$

In this paper we study the problem of optimizing the sensor trajectory given a fixed vehicle trajectory $\sigma(.)$. The problem is then reduced to.

$$\underset{\sigma_s(t)}{\arg\min} \underset{p(b|\sigma,\sigma_s,b_o)}{\mathbb{E}} \int_0^T C_b(\sigma(t), \sigma_s(t), b(t))\,\mathrm{dt} \qquad (2)$$

The constraints on the sensor actuation specified in eq. 1 also apply to eq. 2. Notice that, in this formulation the sensor motion can only result in gaining of information about the environment. Therefore, the minimization of the cost function with the $\sigma_s(t)$ as the only variable, results in sensor gaining information that is important to minimize the required cost function. We call this information contextually important information, as it is important to sense this information to reduce the cost function. Let $\mathbb{IG}(x, \sigma(t), \sigma_s(t), b(t))$ be the information gain at $x \in (W)$ and time t. Let $M(x, \sigma(t), b(t)) \in \{0, 1\}$ return 1 if the information is contextually important and 0 otherwise. The optimization problem then reduces to eq: 3. Proof in [16].

$$\underset{\sigma_s(t)}{\arg\max} \underset{p(b|\sigma,\sigma_s,b_o)}{\mathbb{E}} \int_0^T \int_{\forall x \in \mathbb{X}} M(x, \sigma(t), b(t)) \mathbb{IG}(x, \sigma(t), \sigma_s(t), b(t))\mathrm{dxdt} \qquad (3)$$

In our application, the robot is a full scale autonomous helicopter, the sensor is a nodding lidar, with an actively controlled pitch axis. The trajectory $\sigma_s(t)$ of the nodding sensor relative to the helicopter can be completely defined by its pitch angle profile in time $\rho(t)$, laser configuration ρ_{conf}, which consists laser's fast axis resolution ρ_{frr} and its pulse rate, ρ_{prr}. The pose of the robot is given by a high accuracy GPS/INS system, the world representation is an occupancy grid map [17]. The information gain evaluation for an occupancy grid map is presented in detail in [16]. $M(x, \sigma(t), b(t))$, the contextual importance function is discussed in detail in section IV. In the next section we present an overview of our approach to solve eq. 3 for a mobile autonomous robot.

III. Approach Overview

Information gain based path planning is an NP-hard problem [14]. Moreover, the calculation of information gain itself is computationally expensive. In this work, we use an occupancy grid map based representation. For a laser sensory action with R rays, interacting with N cells of an occupancy grid map, the computation complexity of calculating information gain is $O(3^R N)$ [16]. As a result, the computational cost involved in calculating information gain for a large number of rays over a large range, makes it non-trivial to conduct gradient based optimizations.

To overcome the computational complexity issues and solve the problem online, we propose learning a reactive policy function that maps from features to action, (see

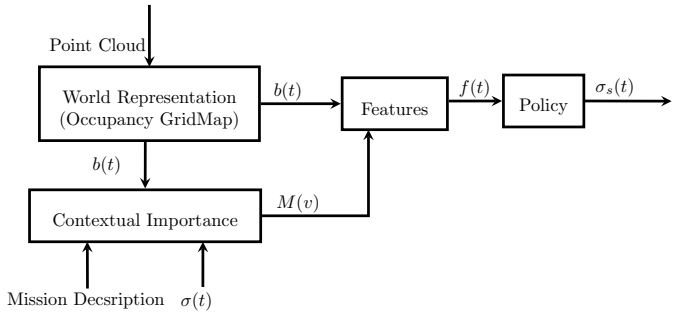

Fig. 2: Approach Overview: The sensor is actively controlled through a policy that takes in the features that are extracted from robot's belief. The features take expected information gain and contextual importance function into account.

Fig. 2). The policy learnt, is designed to maximize the information gain in the next time step. For the application considered in the paper, the policy learnt keeps the robot safe at high speeds and enables quick landing zone evaluation, allowing for low mission times. The data flow in the block diagram (Fig. 2) is explained in the rest of this section to provide an intuitive understanding of the suggested approach and its application.

The point cloud generated by the laser is used by the perception representation to form the robot's belief about the world, in our case an occupancy grid map is used as the world representation. The belief of the robot, along with the mission description and robot trajectory are used to infer the contextual importance function. The contextual importance function presents the locations from which it is important for the robot to gain information to complete its task safely. The construction of this function is further described in section IV. The contextual importance function along with the belief of the robot are used to calculate features. The policy maps these features to sensory actions. The features are the only input to the policy. Therefore, they should capture the saliencies of the interaction between sensory actions, the representation and the environment. The construction of one such set of features is described in section V-A.

The policy takes the features as input and provides the trajectory for the sensor, $\sigma_s(.)$. For the application considered in the paper, the policy is designed for high range laser, nodding in pitch axis with respect to the vehicle. The construction of the policy is described in section V. The parameters of the constructed policy function need to be learnt to maximize the contextually important information gain. The large partially known state space in which the robot operates makes the learning of the correct policy parameters intractable. We reduce the scenarios for which the policy is to be learnt to overcome this problem. The procedure to learn the policy parameters is presented in section VI. In the next section we describe the construction of contextual importance function for our application.

IV. Contextual Importance Function

Finding the contextual importance function is as hard as solving for the original optimization problem [16]. We approximate this function by defining $M(x, \sigma(t), b(t)) = 1$, for all x, that might be important to the vehicle given the belief $b(t)$ and the robot state σ. In this work, we consider the structure of the contextual importance function to have two components - safety $M_{safe}(.)$ and landing zone evaluation $M_{lz}(.)$.

A. Safe Navigation

We use the emergency maneuver library [18] to enforce the safety constraint. It suggests using an offline computed emergency maneuver library to ensure vehicle safety. If there exists a trajectory in the emergency maneuver library that starts from the current state of the vehicle and stays within the known obstacle free region for infinite time, the vehicle can be considered safe. Therefore, it is important to sense the volume occupied by emergency maneuver trajectories corresponding to the planned trajectory.
Given a vehicle state $\sigma(t)$ at time t, let there be a function $\kappa(x, \sigma(t)) \in \{0, 1\}$ which is one for the points inside the volume occupied by emergency maneuver library for state $\sigma(t)$ if it is unsafe and zero otherwise. The the contextual importance function for safety is then given as.

$$M_{safe}(x, \sigma(t), b(t)) = \kappa(x, \sigma(t)) \tag{4}$$

B. Landing Zone

The mission objective of the helicopter is to navigate safely from start to goal and then land. Therefore, it is important to evaluate landing zones before the helicopter commits to a landing. The landing zone is represented as a $2D$ grid, with each cell storing the probability of that cell being a safe landing site. Let, the landing zone grid be represented by $\mathbb{LZ}_{\mathbb{G}}$. Each cell inside the landing zone is allocated contextual importance of one.

$$M_{lz}(x, \sigma(t), b(t)) = 1 \; \forall x \in \mathbb{LZ}_{\mathbb{G}} \tag{5}$$

The combined contextual importance function of safety and landing zone evaluation is given as

$$M(.) = \max(M_{safe}(.), M_{lz}(.)) \tag{6}$$

Fig. 3, illustrates the contextual importance function for a helicopter navigating towards a landing zone to land.

V. Policy Parametrization

We have discussed that the problem of directly optimizing the cost function itself is NP hard and the cost function evaluation is computationally expensive, making local optimization intractable. Therefore, we develop a mapping from features to actions that can be used online.

We learn a policy, $\pi : \mathbb{R}^f \rightarrow \mathbb{R}^a$, where $\mathbb{R}^f$ is the feature space and $\mathbb{R}^a$ is the action space. The policy is developed such that it can guarantee the safety of the vehicle at high speeds if the environment allows so. We restrict the action space to constant velocity nods to make the search computationally feasible. We learn a policy offline that maximizes

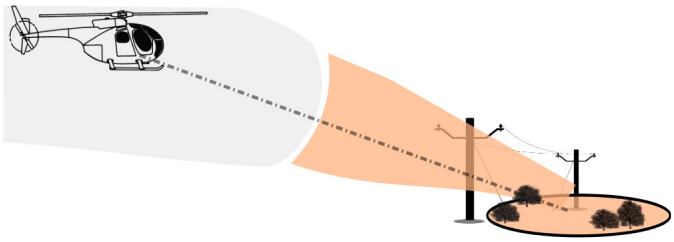

Fig. 3: Contextual Importance Function – The figure illustrates the region with contextual importance function greater than one in orange, and regions which were contextually important in grey. As the robot navigates, the regions known in the past are now irrelevant, as there is no more information to be gained that might affect its future actions. The volume around the trajectory, bounded by the emergency maneuver library for the future as yet unsafe states and volume inside the landing zone is contextually important.

the information gain in the worst case scenario. The worst case scenario and the process of learning the policy function is presented in section VI.

We want to optimize the use of sensor bandwidth and avoid wasting sensor's time looking at regions that are already known, inaccessible or unimportant to the mission. Therefore the features have to cover visibility, information gain and contextual importance function. In the next subsection we cover the feature design and then present the policy function.

A. Features

To calculate the relevant field of view for the sensor, we calculate and store the contextually important expected information gain along a grid of pitch (θ) and yaw (ϕ) directions, with rays originating from laser's center. This forms a $2D$ map $\zeta_m(\theta, \phi) \rightarrow \mathbb{R}$, mapping view direction to contextual information gain. We generate this map by tracing rays through 3D occupancy grid representation and calculating contextually weighed expected information gain along the rays. The algorithm for calculating the expected information gain along the rays is presented in [19]. Average range at which information is gained r_e is also used as a feature for the policy. r_e can be calculated while calculating ζ_m. The algorithms to calculate ζ_m and r_e are presented in detail in [16]. The policy function can then be represented as eq. 7 below.

$$[\dot{\rho}, \rho_{conf}, \rho_{max}, \rho_{min}] = \pi(\zeta_m, r_e) \tag{7}$$

The FOV to scan ρ_{max}, ρ_{min}, is inferred using ζ_m. The ρ_{max} is given by maximum pitch in ζ_m for which the information gain is greater than 0. The ρ_{min} is given by minimum pitch in ζ_m for which the information gain is greater than 0. The velocity of the nod is stored as lookup table against r_e at which the information to be sensed. We now look at how we learn the mapping (lookup table) between range and velocity that maximizes the information gain at that range.

VI. Policy Search

We want to find a nodding speed at which to scan the volume to maximize the gain of contextually important information. The standard method to find the optimal policy is to run the system either in simulation or in field and search the parameter space for values that minimize the average cost over a range of scenarios. But given we are designing a policy to keep the vehicle safe in all conditions, we search for the optimal policy for the worst case scenario. In the worst case scenario, the time available to the sensor for scanning is minimal and the volume to be covered is maximum.

The following subsections develop the worst case scenario for a guaranteed safe autonomous mobile robot navigating through unknown environments. This scenario is used to learn the optimal nodding action for a given query range. The optimal action maximizes contextually important information gain. Where, information gain is a function of reduction in uncertainty about the presence of obstacle of interest.

In the case of a rotorcraft the smallest obstacles of interest are wires. This section describes how information gain is related to probability of detection of obstacles of interest. We further describe algorithms to efficiently calculate probability of detection given sensory actions, enabling search of policy parameters $(\dot{\rho})$ in reasonable times.

A. Policy Search Scenario

Proposition 1 (Worst Case Scenario for Safety)**.** *Given the average range at which contextually important information is requested is r. The sensor would have to strive to gain maximum information to ensure vehicle safety while the vehicle follows the planned trajectory if it is moving at the maximum speed that allows the vehicle to stay safe at range r. Assuming the vehicle is flying unknown environment and the environment allows for safe following of the nominal trajectory.*

The proof of the proposition [16] relies on the fact that the volume occupied by emergency maneuvers is monotonically increases with vehicle's speed. We optimize the nodding speed for this worst case scenario. Assuming the world representation takes t_r seconds to refresh. We search for a nodding speed that maximizes the gain of information in the relevant volume V_r in t_r seconds, where V_r is given by equation 8.

$$V_r = max(S_V(r)) \tag{8}$$

where $S_V(r)$, returns a vector of bounding volumes of emergency maneuver libraries that lie completely within a sphere of radius r of the vehicle. In the next subsection we develop the definition of information gain for safety, and show how to optimize for it efficiently.

B. Information gain as probability of detection of worst case obstacle

Let us assume, the smallest obstacle against which the autonomous system has to guarantee safety is given by o, the probability that an object is present at x is given by $p_x(o)$.

For brevity we use $p(o)$ instead of $p_x(o)$. Let the probability of detection of obstacle given there is exists an obstacle at x be given by $p(d|o)$ and probability of detecting an obstacle given there is no obstacle at x is given by $p(d|o')$. We express the the expected information gain given $p_x(o)$ as a function of $p(d|o)$ and $p(d|o')$

.

$$\begin{aligned}\mathbb{E}\,\mathbb{IG}(o) = p(o)(p(d|o)[H(p(o)) - H(p(o|d))+ \\ p(d'|o)(H(p(o)) - H(p(o|d')))]+ \\ p(o')(p(d|o')(H(p(o)) - H(p(o|d)))+ \\ p(d'|o')(H(p(o)) - H(p(o|d'))))\end{aligned} \tag{9}$$

where, $H(.)$ is the entropy [20].

Proposition 2 (Monotonicity of Information Gain)**.** *Assuming there are no false positives,* $p(d|o') = 0, p(d'|o') = 1$. *For a given* $p(o)$, $\mathbb{E}\,\mathbb{IG}$ *can be completely expressed as a function of* $p(d|o)$ *and monotonically increases in* $p(d|o)$.

Proposition 2 is proved in [16], and Fig. 4 illustrates it. Hence, to find the effectiveness of an action we evaluate

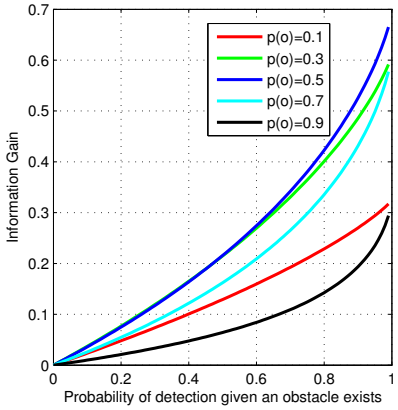

Fig. 4: Information Gain – The expected information gain given a prior ($p(o)$) is monotonic in probability of detecting the event w if the event occurred, $p(d|o)$. The assumption being, there are no false positives, $p(d|o') = 0$. This implies for maximizing information gain, one may maximize $p(d|o)$

$p(d|o)$ for a given $p(o)$ and calculate the information gain in the mission relevant region. The action that provides the maximum information gain is selected as policy for given feature inputs. In the next section we present an efficient algorithm to calculate $p(d|o)$, that allow the creation of policy lookup table tractably.

C. Efficient calculation of probability of detection

We describe a method that exploits the symmetry and the structure of the problem to calculate information gain for a range sensor efficiently. The naive method of calculating $p(d|o)$ is through sampling configurations of the object from $p(o)$ and computing ray intersections with the object given the sensor nodding motion. This algorithm has the complexity of $O(RM)$, where R is the total number of rays generated by the laser and M is the number of samples drawn from $p(o)$. Let's assume an objects exists in a $6D$ space, and for each dimension we sample N particles. The computational complexity is then $O(RN^6)$. We introduce an algorithm that reduces this computational complexity to $O(R^3N^4)$ by reasoning about ray-object intersections in the object's cspace in spherical coordinates. We further exploit the symmetry and structure of the problem of sensing for safety to reduce the complexity to $O(R_l^3N)$, where R_l is a small fraction of the actual number of rays.

The algorithm for fast calculation $P(d|o)$ is presented in Alg. 1. The input to the algorithm is sensor trajectory, σ_s whose information gain is to be calculated, an object's partial pose $x_{partial}$, only stating the distance of the object from the sensor and its relative orientation. We assume that the size of the object is small as compared to the query range, which results in negligible change in projection of the object to $[\theta, \phi]$ plane of the spherical coordinate frame, centered at the laser. The algorithm returns a function $p(d|\theta, \phi, x_{partial})$ that provides the probability of detection of object if its in $x_{partial}$ configuration at any $[\theta, \phi]$. Lets, assume the object is detected if n points hits are detected on the object. Each laser beam is represented as polygon in $[\theta, \phi]$ space of spherical coordinates It is assumed that if the ray and the object intersect, the hit will be reported with a probability of p_h.

The $Intersect(.)$ function (Alg. 2), calculates the intersection of set of unique polygons amongst themselves and stores it as Y_{int} and also returns the list of members of its input set that intersected to form a polygon $\forall i \in Y_{int}$ as $\lambda^i_{int} \in \Lambda_{int}$. The output of Alg. 2 are used by $Probability(.)$ function (Alg. 3), to calculate the probability of detection given an object at $x_{partial}$ $\forall[\theta, \phi]$. It is obvious that given a fixed

Algorithm 1: $EfficientCalculationofP(d|o)$

Input : $\sigma_s, x_{partial}$

Output: $p(d|\theta, \phi, x_{partial})$

$S_o \leftarrow ProjSpherical(s_{partial})$
$S_L \leftarrow GeneratePolygons(\sigma_s)$
$S_{L,o} = S_L \bigoplus S_o$
// The minkowski sum projects polygons in
// S_L to the space of S_o.
$[Y_{int}, \Lambda_{int}] \leftarrow Intersect(S_{L,o})$
$p(d|\theta, \phi, s_{partial}) \leftarrow DetectionProbability(Y_{int}, \Lambda_{int})$
Return: $p(d|\theta, \phi, s_{partial})$

angular motion profile for a range sensor the number of rays intersecting an object decreases with increasing distance. This implies that the probability of detection of the object decreases with increasing distance [16]. We maximize the information gain at the range at which the information is requested, maximizing the minimum information gain. This results on maximizing the information gain on a 2d spherical manifold $[\theta, \phi]$. So the configuration space of the object restricted to a 2D plane is x, y, θ. Assuming, we use N particles for each dimension to approximate $p(o)$,

Algorithm 2: $Intersect(.)$

Input : $S_{L,o}$

Output : $Y_{int}^{s_{partial}}, \Lambda_{int}$

Initialize: $Y_{int}^{s_{partial}} = S_{L,o}, \lambda_{int}^{i} = i \forall i \in Y_{int}^{s_{partial}}, Y_{prev} = S_{L,o}$

while $|Y_{prev}| > 1$ **do**
  $Y_{new} =$
  for $j \in Y_{prev}$ **do**
    for $k \in Y_{prev} - [1:j]$ **do**
      $Y_{new} = Y_{new} \cup (j \cap k)$
      $\lambda_{int}^{end+1} = \lambda_{int}^{j} \cup \lambda_{int}^{k}$
    end
  end
  $Y_{int}^{s_{partial}} = Remove(Y_{int}^{s_{partial}}, Y_{new})$
  $Y_{int}^{s_{partial}} = Y_{int}^{s_{partial}} \cup Y_{new}$
  $Y_{pref} = Y_{new}$
end
Return : Y_{int}, Λ_{int}

Algorithm 3: $Probability(.)$

Input : $Y_{int}^{s_{partial}}, \Lambda_{int}^{s_{partial}}, \phi, \theta$

Output: $p(d|\phi, \theta, s_{partial})$

$i \leftarrow ReturnPolygonContaining(\theta, \phi, Y_{int}^{s_{partial}})$
$p(d|\phi, \theta, s_{partial}) = \sum_{k=n}^{|\lambda_{int}^{i}|} \binom{|\lambda_{int}^{i}|}{k} p_h^k (1-p_h)^{|\lambda_{int}^{i}|-k}$
Return: $p(d|\phi, \theta, s_{partial})$

the computational cost of evaluating an action is given by $O(R^3N)$.

We restricted the action space to constant velocity nods, hence the laser points lie at a constant distance form each other in pitch and yaw space, forming a grid. This symmetry can be exploited by finding $p(d|o)$ for every configuration in a single grid cell as $p(d|o)$ is the same for every cell. Also, the intersection of the object need only be checked with the points in l^2 radius of the grid cell. Let these number of points in l^2 radius be R_l. The complexity of evaluating an action is $O(R_l^3 N)$. For further details about reduction in computational complexity please refer to [16].

Alg. 1 is evaluated exhaustively for the complete nodding velocity space at finite discrete intervals, to create the lookup table of range vs nodding velocity. The slowest nodding speed is restricted such that one complete nod occurs in maximum $t_r = 1.4s$. The action providing maximum information gain is stored as the nodding time for that range. Fig 5 presents plots for expected probability of detection of a wire, given a uniform distribution. This gives a more intuitive evaluation of the nodding trajectories.

The policy for scanning for landing zone evaluation may also be calculated by the method presented above, but for our application continuously scanning the region of landing zone to get a uniform point distribution on it sufficed as an adequate policy. The sampling rate of the laser is fixed for safety scans to 29000 points per second, as changing the sampling rate means switching off the laser for 7 seconds and reducing the maximum range of the laser from $1400m$ to $650m$, which rendered the vehicle unsafe. Once the safety of the entire mission is ensured, the sampling rate of the sensor is increased for faster landing zone evaluation.

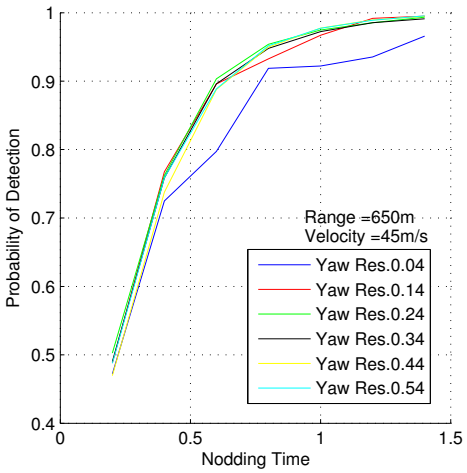

Fig. 5: Expected Probability of Detection: The expected probability of detection of a wire if it exists, $\mathbb{E}_{p(o)}\, p(d|o)$ for a given action is an intuitive indicator of how good an action is. The plot shows $\mathbb{E}_{p(o)}\, p(d|o)$ versus varying nodding time period for the worst case scenario for the sensor at vehicle velocity, $v = 45\ m/s$. Each sensor velocity corresponding to different nodding time periods is evaluated till $t_r = 1.4seconds$. The evaluation shows scans with slower scanning speed are better, this is also intuitively the correct behavior as slower nodding speeds means more uniform point distribution in the $[\theta, \phi]$ manifold.

VII. Results

We test our approach on an autonomous full-sized helicopter which uses the Near Earth Autonomy, Mark 3 active nodding range sensor. We tested our approach in Mesa, AZ, Manassas, VA, Pittsburgh, PA and Quantico, VA for more than 30 autonomous missions in varying conditions. Fig. 6 shows a typical collection of missions. All the missions need the vehicle to navigate from start to the landing zone at high speeds, evaluate the landing zone and land without hovering if the landing zone is safe. The sensor had to clear enough region to keep the vehicle safe, even at the speeds of 56 m/s and sample a $100m$ diameter landing zone with enough number of points to evaluate it for landing the vehicle on it. On an average the time available to the laser to sample the landing zone (LZ) was below 20seconds. Such missions were impossible to conduct with a passive scanning approach. Passive scans designed to ensure safety and evaluate landing zone are only able to keep the vehicle safe up to a maximum

speed of 18 m/s and evaluate the complete landing zone in 112 seconds. Fig. 7 shows a detailed analysis of the sensor

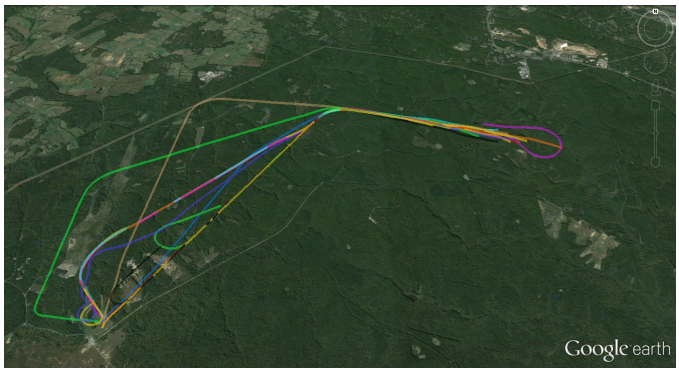

Fig. 6: Mission Definition: The helicopter navigates from loiter point to landing zone in less than 210 seconds. It has to navigate the environment while being provably safe and touch down at the LZ without hovering over it to look for potential sites.

angles during a mission in Mesa, AZ. The policy scans for safety in the beginning when the vehicle is navigating towards its destination. The focus of the sensor is moved to LZ evaluation once safety is ensured till the end of the mission. Notice the laser configuration is changed only after confirming the safety of the vehicle for the rest of the mission. This mitigates the problem of reduced sensor range at high sampling rates affecting vehicle safety. Fig. 8 shows

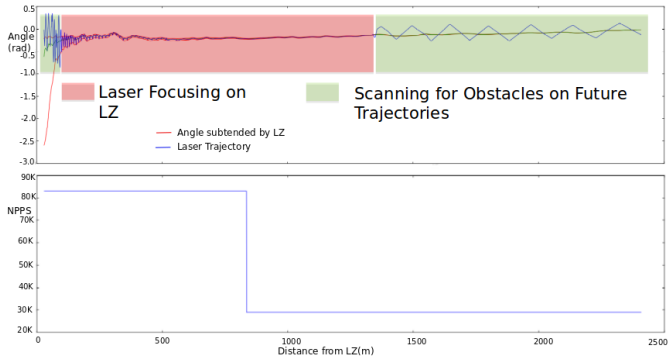

Fig. 7: Sensor Angles and Configuration: The top figure shows the sensor angles. In the initial part of the mission, the sensor plans to keeps the vehicle safe (oscillation of the blue line). It switches to focussing on the landing zone when safety for the remainder of the mission has been guaranteed. This focussing of the laser is shown by the narrow peak to peak of the nodding angles. At the very end, once enough points on the landing zone has been focussed, the sensor reverts back to ensuring that the vehicle can be safe should it desire to waveoff. The bottom figure shows the configuration switch. Since switching configuration reduces range, the policy triggers this event at the appropriate distance to landing zone, while making sure that the safety of the rest of the mission is ensured.

the performance of the policy in keeping the vehicle safe. The sensor always keeps the vehicle safe and only switches to scanning the landing zone when it has guaranteed safety for the entire trajectory leading to touchdown. Notice that the executed vehicle speed is always well below the safe speed of the vehicle. Fig. 9 shows performance of the system

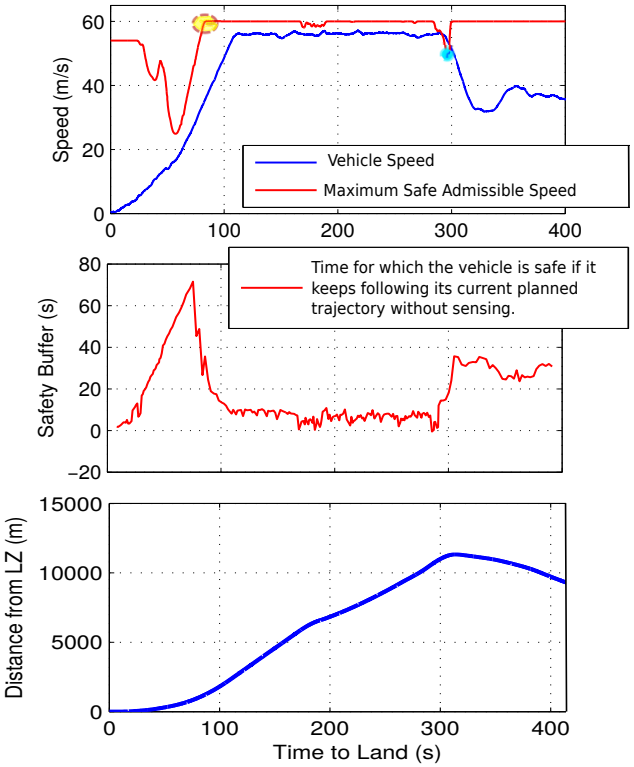

Fig. 8: Safety analysis for a single mission. In each figure, two events are marked - first (blue) is when a manual pilot makes an aggressive maneuver, second (yellow) is when the sensor switches to scanning the landing zone. The figure in the top shows that in autonomous mode the vehicle is flying slower than the safe speed limit. The middle figure shows the time after which the vehicle will be unsafe given its current belief. As the vehicle approaches to land, this time increases till its high enough for the sensor to switch to scanning the landing zone. The bottom figure shows the progress of the mission.

over 12 missions executed in Quantico, VA. The vehicle is always safe and evaluates the landing zone in time for all the missions, this demonstrates the reliability of the policy in the real world.

VIII. Conclusion

The main contribution of this paper is to define an algorithm for actively controlling the sensor mounted on a vehicle to maximize the vehicle's performance. We reduce the problem of actively controlling the sensor to minimize trajectory cost to the problem maximizing contextually important information gain. We suggest a policy based approach to solve the NP hard problem of maximizing contextually

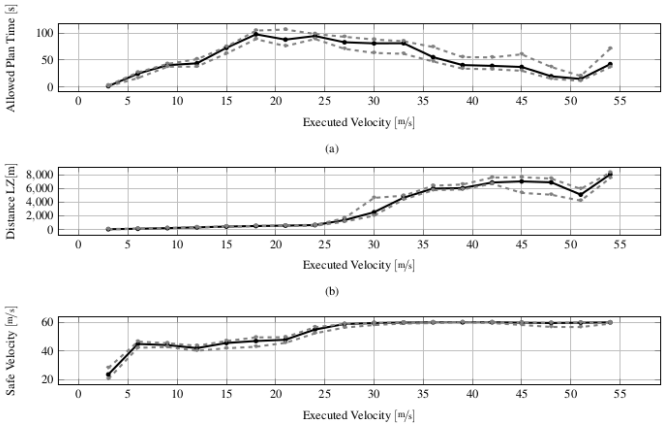

Fig. 9: Multiple Mission Safety Analysis: The safety analysis shown over several missions. The worst case performance of the system is still shown to be guaranteed safe.

important information gain. The method is evaluated through application on a full scale autonomous helicopter. We have shown clear benefits of actively controlling the sensor using our method as opposed to passive nodding. The method enabled the helicopter to perform autonomous high speed, guaranteed safe flights and land on landing sites without the need to hover.

In the future, we want to explore the possibility of inferring the contextual importance function, given a cost function. We are currently working towards applying the method on micro aerial vehicles and extending the policy search space to allow for richer sensory trajectories.

IX. Acknowledgement

This work would not have been possible without the dedicated efforts of the entire AACUS TALOS team and was supported by ONR under contract N00014-12-C-0671.

References

[1] A. S. Huang, A. Bachrach, P. Henry, M. Krainin, D. Maturana, D. Fox, and N. Roy, "Visual odometry and mapping for autonomous flight using an rgb-d camera," in *International Symposium on Robotics Research (ISRR)*, 2011, pp. 1–16.

[2] P. Furgale and T. D. Barfoot, "Visual teach and repeat for long-range rover autonomy," *Journal of Field Robotics*, vol. 27, no. 5, pp. 534–560, 2010.

[3] C. Bills, J. Chen, and A. Saxena, "Autonomous mav flight in indoor environments using single image perspective cues," in *Robotics and automation (ICRA), 2011 IEEE international conference on*. IEEE, 2011, pp. 5776–5783.

[4] S. Scherer, J. Rehder, S. Achar, H. Cover, A. Chambers, S. Nuske, and S. Singh, "River mapping from a flying robot: state estimation, river detection, and obstacle mapping," *Autonomous Robots*, vol. 33, no. 1-2, pp. 189–214, 2012.

[5] S. Scherer, L. Chamberlain, and S. Singh, "Autonomous landing at unprepared sites by a full-scale helicopter," *Robotics and Autonomous Systems*, vol. 60, no. 12, pp. 1545–1562, 2012.

[6] A. Wald, "Sequential tests of statistical hypotheses," *The Annals of Mathematical Statistics*, vol. 16, no. 2, pp. 117–186, 1945.

[7] A. Cameron and H. Durrant-Whyte, "A bayesian approach to optimal sensor placement," *The International Journal of Robotics Research*, vol. 9, no. 5, pp. 70–88, 1990.

[8] S. Chen, Y. Li, and N. M. Kwok, "Active vision in robotic systems: A survey of recent developments," *The International Journal of Robotics Research*, vol. 30, no. 11, pp. 1343–1377, 2011.

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

[10] J. Van Den Berg, S. Patil, and R. Alterovitz, "Motion planning under uncertainty using iterative local optimization in belief space," *The International Journal of Robotics Research*, vol. 31, no. 11, pp. 1263–1278, 2012.

[11] F. Bourgault, T. Furukawa, and H. F. Durrant-Whyte, "Coordinated decentralized search for a lost target in a bayesian world," in *Intelligent Robots and Systems, 2003.(IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, vol. 1. IEEE, 2003, pp. 48–53.

[12] A. Ryan and J. K. Hedrick, "Particle filter based information-theoretic active sensing," *Robotics and Autonomous Systems*, vol. 58, no. 5, pp. 574–584, 2010.

[13] V. Myers and D. Williams, "A pomdp for multi-view target classification with an autonomous underwater vehicle," in *OCEANS 2010*. IEEE, 2010, pp. 1–5.

[14] A. Singh, A. Krause, C. Guestrin, and W. J. Kaiser, "Efficient informative sensing using multiple robots," *Journal of Artificial Intelligence Research*, vol. 34, no. 2, p. 707, 2009.

[15] J. Binney, A. Krause, and G. S. Sukhatme, "Optimizing waypoints for monitoring spatiotemporal phenomena," *The International Journal of Robotics Research*, vol. 32, no. 8, pp. 873–888, 2013.

[16] S. Arora and S. Scherer, "Policy based approach for sensor planning," http://www.frc.ri.cmu.edu/~sankalp/publications/pasp.pdf.

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

[18] S. Arora, S. Choudhury, S. Scherer, and D. Althoff, "A principled approach to enable safe and high performance maneuvers for autonomous rotorcraft," in *American Helicopter Society 70th Annual Forum*. AHS, 2014.

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

[20] C. E. Shannon, "Prediction and entropy of printed english," *Bell system technical journal*, vol. 30, no. 1, pp. 50–64, 1951.