2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

Sensor Control for Search and Identification of Markov Objects

Darin C. Hitchings and David A. Castañón

Abstract— **In this paper, we discuss stochastic control approaches to sensor control problems for the purposes of locating and classifying objects that can enter and leave areas of interest, and there are many objects to interrogate. Noisy sensors with limited energy can choose to interrogate areas to find and identify objects while they are present in the scenario, and can use different modes to either search or identify objects. The goal is to identify objects appearing in the scenario as soon as they are present. Although the resulting stochastic control problem is a partially observed Markov decision problem with combinatorially large action and state spaces, we develop an approximate stochastic control formulation based on relaxing constraints concerning the utilization of sensor energy, and obtain an efficient algorithm for generating near-optimal sensor control decisions. The resulting algorithm is illustrated in a simple scenario with a single sensor observing multiple areas of interest.**

I. INTRODUCTION

Advances in embedded computing have introduced a new generation of sensors that have the capability of adapting their sensing dynamically in response to collected information. Unmanned aerial vehicles (UAVs) have multiple sensors which can change their fields of view and resolution dynamically. However, there is a need for sensor control algorithms that exploit processed information obtained from sensor collections and selects measurement actions in order to improve the performance of the sensor system. Such control algorithms have numerous applications in surveillance problems, as well as fault identification and diagnosis.

An early example of sensor control can be found in search theory, where sensors moved and allocated search effort over time and space to locate objects [1], [2]. Although much of search theory focuses on the design of open-loop sensor control, there are interesting extensions to problems requiring adaptive feedback control based on noisy measurements [3]. Other early examples included Wald's theory of sequential hypothesis testing with costly observations [4], [5], as well as work on sequential nonlinear regression [6], [7] that used Cramer-Rao bounds for adaptive selection of measurements. Most of these control approaches involve one-step lookahead optimization criteria. Alternative approaches to adaptive Other approaches to adaptive control of sensing using single-stage optimization have been proposed using information theoretic objectives and performance bounds [8]–[10].

Feedback control approaches to sensor control based on optimization over time have been explored in different contexts. Athans [11] controlling the error covariance in linear estimators by choosing among potential linear measurements using maximum principle techniques. Multi-armed bandit formulations have been used to control individual sensors in applications related to target tracking [12], [13]. Such approaches are restricted to single-sensor control in order to obtain solutions using Gittins indices [14], [15]. Approximate dynamic programming (DP) techniques have also been proposed using approximations to the optimal cost-to-go based on information theoretic measures evaluated using Monte Carlo techniques [16], [17]. An overview of these techniques is available in [18].

The above approaches for dynamic feedback control are limited in application to problems with a small number of sensor-action choices and simple constraints because the algorithms must enumerate and evaluate the various control actions. For problems with many actions, [19] integrates combinatorial optimization techniques with stochastic dynamic programming to obtain stochastic control algorithms. Subsequent work in [20] derived a formulation for sensor control using partially observed Markov decision processes (POMDPs) and obtained a computable lower bound to the achievable performance of feedback strategies for complex multi-sensor management problems involving classification of stationary objects. The lower bound was obtained by a convex relaxation of the original combinatorial POMDP. The results in [20] were extended in [21] to obtain sensor control algorithms with performance close to the lower bound, using a receding horizon control approach.

In this paper, we extend the results of [20] and [21] to the problem of adaptive sensor control in the presence of Markovian objects, where the underlying state of the object may change with time. We impose a structural condition on the Markovian objects, in that the state at each location evolves independently of states at other locations, which allows extension of our previous approaches to this class of problems. We pose the sensor control problem as resource-constrained adaptive control problem using a POMDP framework, which is computationally intractable. As in [20], we develop a convex relaxation that provides a lower bound on the performance objective. Our results present a simpler derivation of the bound result and provide approaches for computing optimal solutions to the lower bound problem using combinations of integer programming and stochastic dynamic programming. The resulting algorithms are used to develop approximate control strategies that satisfy the required constraints, and achieve performance close to the lower bound, as in [21]. We illustrate the performance of the algorithms in a simple example.

Darin Hitchings is with Livevol, Inc, in San Francisco, CA, USA
David Castañón is Professor with the Department of Electrical Engineering, Boston University, Boston, MA 02215, USA dac@bu.edu

This work was supported by AFOSR grants FA9550-07-1-0361 and by ODDR&E MURI Grants FA9550-06-1-0324 and FA9550-07-1-0528

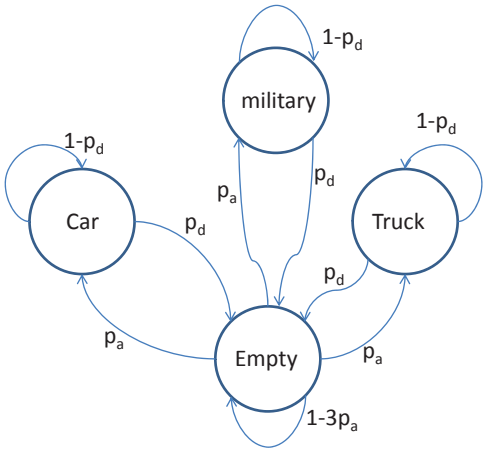

Fig. 1. An example HMM that can be used for each of the N locations. p_a is an arrival probability and p_d is a departure probability for the Markov chain.

The rest of this paper is organized as follows: Section II describes the formulation of the stochastic sensor control problem. Section III describes an approximate stochastic control problem that provides a lower bound to the original problem, and develops an algorithm for computing optimal strategies for this approximate control problem. Section IV discusses a computational example that illustrates the control algorithm. Section V summarizes our results and discusses areas for future work.

II. Problem Statement

Assume that there are a finite number of locations $1, \ldots, N$ where objects of different types may enter and leave. We assume there is a set of S sensors, each of which has multiple sensor modes, and that the sensor can distribute its sensing energy at each time over multiple locations by choosing its observation mode for each location. The sensor control problem is determining, for each sensor, which locations and which modes to use for collecting information at each discrete time step.

The contents of location i at time t is represented by state $x_i(t) \in \{0, 1, \ldots, D\}$, where $x_i(t) = 0$ if location i is unoccupied, and otherwise $x_i(t) = k > 0$ indicates location i contains an object of type k at time t. The initial knowledge about the contents of location i is represented by a discrete a priori probability distribution $\pi_i(0) \in \Re^{D+1}$ over the possible states for the i^{th} location for $i = 1, \ldots, N$ where $D \geq 2$. Assume that the random variables $x_i(t)$ for $i = 1, \ldots, N$ are mutually independent for each time t. The state of each of the N locations evolves dynamically as a Markov chain, such as Fig. 1. The transition probabilities are specified as a stochastic matrix $\{p_{jk}\}$ with stationary transition probabilities $p_{jk} = P(x_i(t+1) = j | x_i(t) = k)$. We assume the Markov transition probabilities are the same for each location for simplicity; extensions to different transition chains per location or time varying transition probabilities are straightforward.

There are $s = 1, \ldots, S$ sensors, each of which has $m = 1, \ldots, M_s$ possible modes of observation. Let there be a series of T discrete decision stages with $t = 1, \ldots, T$ for sensors to make measurements. Each sensor s has a limited set of locations that it can observe at each stage, denoted by $O_s(t) \subseteq \{1, \ldots, N\}$. At each stage, each sensor can choose to employ one of its sensor modes to collect noisy measurements concerning the states $x_i(t)$ of the sensed locations in its Field of View (FOV) (location i is in the FOV of sensor s if $i \in O_s(t)$).

To define the control objective, at each time, one must make an estimate of the content of each location after measurements are collected and processed. Thus, the control actions at each stage consist of two types: first, each sensor selects locations and modes to observe; information is collected and processed, and then the system makes a tentative classification of the current content of each location.

A sensor action by sensor s at stage t is the set of pairs:

$$u_s(t) = \{(i_s(t), m_s(t)) \mid i_s(t) \in O_s(t), m_s(t) \in M_s\} \quad (1)$$

where each pair consists of a location to observe $i_s(t)$, and a sensor mode (independent for each location) used to observe this location, $m_s(t)$, where the mode is restricted to the set of feasible modes given the resource levels for each sensor. We assume that no two sensors observe the same location at the same time in order to minimize the complexity of the associated action and observation spaces. Let $u_{i,s}(t)$ refer to the sensor action taken on location i with sensor s at stage t if any, or let $u_{i,s}(t) = \emptyset$ otherwise.

Sensor measurements at time t of location i with mode m are denoted as $y_{i,s,m}(t) \in \{1, \ldots, L_s\}$. Measured values $y_{i,s,m}(t)$ are assumed to be conditionally independent of other values $y_{j,\sigma,n}(\tau)$ given the underlying values of the states at the measured locations in $u_s(t), u_\sigma(\tau)$ whenever $i \neq j$, or $\tau \neq t$ or $\sigma \neq s$. Denote the conditional probability of the measurement as $P(y_{i,s,m}(t) | x_i(t), i, s, m)$. Thus, the underlying Markov states at each location are observed through noisy measurements, and the optimal Bayesian inference can be performed using Hidden Markov Model estimation. We assume that the conditional probability of measurements given $x_i(t)$ is time-invariant.

In terms of constraints, assume each sensor has a quantity R_s of resources available for measurements during *each time*. Associated with the use of mode m by sensor s on location i at time t is a resource cost $r_s(u_{i,s}(t))$ to use this mode, representing power or some other type of resource required to operate the sensor, represented as

$$\sum_{i \in O_s(t)} r_s(u_{i,s}(t)) \leq R_s \;\; \forall\, s \in [1 \ldots S]; \;\; \forall\, t \in [1 \ldots T] \quad (2)$$

This constraint applies for each realization of observations and decisions. In contrast with [20], the available resources per sensor are limited at each time rather than across all times; in addition, at each time a tentative decision must be made as to the contents of each location, rather than a single decision at the end of a time horizon.

Let $I(t)$ denote the sequence of past sensing actions and measurement outcomes up to and including time $t-1$:

$$I(t) = \{(u_{i,s}(\tau), y_{i,s,m}(\tau))|\ i \in O_s(\tau);\ s = 1, \ldots, S; \\ \tau = 1, \ldots, t-1\}$$

Under the assumption of conditional independence of measurements and independence of the Markov chains governing each location, the joint probability $\pi(t) = P(x_1(t) = k_1, x_2(t) = k_2, \ldots, x_N(t) = k_N|I(t))$ can be factored as the product of marginal conditional probabilities $\pi_i(t) = p(x_i(t)|I_i(t))$ at each location, where $I_i(t)$ denotes the sequence of past sensing actions and measurement outcomes of location i up to and including time $t-1$. This structure will be exploited in our algorithms later, as it allows us to represent a complex sufficient statistic in terms of the marginal statistics. This independence implies that objects that leave one location do not go to a nearby location. A measurement of location i with the sensor-mode combination $u_{i,s}(t) = (i, m)$ at stage t that generates observation value $y_{i,s,m}(t)$ updates the marginal conditional probabilities as:

$$\pi_i(t+1) = \frac{\text{diag}\{P(y_{i,s,m}(t)|x_i(t) = j, i, s, m)\}\pi_i(t)}{\underline{1}^T \text{diag}\{P(y_{i,s,m}(t)|x_i(t) = j, i, s, m)\}\pi_i(t)} \quad (3)$$

where $\underline{1}$ is the $D+1$ dimensional vector of all ones. (3) captures the relevant information dynamics that are controlled by the choice of sensor actions.

Given the information $I(t)$ at stage t, an estimate $v_i(t)$ of the state $x_i(t)$ of each location i is made. The Bayes' cost of selecting estimate $v_i(t)$ when the true state is $x_i(t)$ is denoted as $c(x_i(t), v_i(t)) \in \Re$ with $c(x_i(t), v_i(t)) \geq 0$. The objective of this problem is to estimate the state of each location at each time with minimum error:

$$J = \min_{\gamma \in \Gamma} \mathop{\mathbf{E}}_{\gamma} \left[\sum_{i=1}^{N} \sum_{t=1}^{T} c(x_i(t), v_i(t)) \right] \quad (4)$$

subject to (2). The minimization is done over the finite space of admissible, adaptive feedback strategies $\gamma \in \Gamma$, corresponding of time-varying maps from information sets $I(t)$ to sensor actions $u_s(t)$ and from information sets $I(t+1)$ to tentative classification decisions $v_i(t)$. Note that determining an optimal classification decision strategy is straightforward, and corresponds to minimizing the Bayes' risk of the classification decision at each stage given the available information. Thus, the hard part of determining an optimal strategy is determining the sensing strategy.

The above formulation is a Partially Observed Markov Decision Problem (POMDP) that extends the formulation of [20] from static location contents to Markovian contents. As in [20], this POMDP has combinatorially large state and action spaces, and is intractable for all but the simplest of problems. Our approach will be to modify this problem to obtain a lower bound, as in [20], and to develop algorithms that can obtain solutions to this lower bound and can be used to generate feasible control strategies for the original problem.

III. Lower Bound Formulation

Instead of solving the problem outlined in the previous section, we focus on a different version where the hard resource constraint in (2) are replaced with an expected-resource-use constraint for each of the S sensors, as:

$$\sum_{i \in O_s(t)} \mathbf{E}[r_s(u_{i,s}(t))] \leq R_s \ \ \forall\, s \in [1 \ldots S];\ \ \forall\, t \in [1 \ldots T] \quad (5)$$

This problem is a lower bound on the original problem (4) with sample path constraints (2) because every strategy that satisfies the original sample path constrained problem is feasible for the relaxed problem. However, it is a POMDP with combinatorial action and state spaces.

Define nonnegative multipliers $\lambda_s(t) \geq 0\ \forall s, t$. Define an augmented objective in Lagrangian form as

$$J_\lambda = \min_{\gamma \in \Gamma} \mathop{\mathbf{E}}_{\gamma} \left[\sum_{i=1}^{N} \sum_{t=1}^{T} c(x_i(t), v_i(t)) - \sum_{s=1}^{S} \sum_{t=1}^{T} \lambda_s(t) \left(R_s - \sum_{i \in O_s(t)} r_s(u_{i,s}(t)) \right) \right] \quad (6)$$

The following result is typical of weak duality results in nonlinear programming

Theorem 3.1 (Lower Bound on Performance): The solution of the unconstrained decision problem with augmented objective function (6) is a lower bound on the achievable performance of the decision problem (4) with sample path constraints (2).

Proof: Every admissible strategy $\gamma \in \Gamma$ that satisfies (2) also satisfies (5). Thus, for admissible, adaptive feedback strategies that satisfy (5), the second term in (6) is non-positive, and thus has a value less than or equal to the value in (4). Thus, the unconstrained minimization over strategies in (6) must yield a lower bound on the original optimization (4) with constraints (2). ■

Unfortunately, the optimization in (6) is still over a joint set of strategies across all locations. To further simplify the optimization, we show, similar to the work in [20], [22], that one can choose strategies for optimizing (6) where the actions at location i depend only on the information collected at location i, as established below.

Theorem 3.2: Under the assumption of independent Markov states across locations, and multiplier trajectories $\lambda_s(t)\ \forall\ s, t$, an optimal solution to the optimization problem in (6) can be achieved with local adaptive feedback strategies s γ_i that select sensor actions $u_{i,s}(t)$ for each location i based only on local information $I_i(t)$.

Proof: The following inequality follows from minimizing the sum of terms versus summing the minimum per term:

$$\min_{\gamma \in \Gamma} \mathop{\mathbf{E}}_{\gamma} \left[\sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ c(x_i(t), v_i(t)) + \sum_{s=1}^{S} \lambda_s(t) r_s(u_{i,s}(t)) \right\} \right] \geq \sum_{i=1}^{N} \min_{\gamma \in \Gamma} \mathop{\mathbf{E}}_{\gamma} \left[\sum_{t=1}^{T} \left\{ c(x_i(t), v_i(t)) + \sum_{s=1}^{S} \lambda_s(t) r_s(u_{i,s}(t)) \right\} \right] \quad (7)$$

Consider the minimization problem for each location i in the right hand side above:

$$\min_{\gamma\in\Gamma}\mathbf{E}\left[\sum_{t=1}^{T}\left\{c(x_i(t),v_i(t))+\sum_{s=1}^{S}\lambda_s(t)r_s(u_{i,s}(t))\right\}\right]$$

We can solve this problem via stochastic dynamic programming. We break the decision problem at each stage t into two stages: first, we select $u_{i,s}(t)$ and collect information on object i. Then, we select $v_i(t)$, the tentative classification. At the final stage, consider the selection of $v_i(T)$ as a function of the complete information-state $I(T+1)$, which includes the measurements and sensing actions collected at stage T collected over the entire set of locations, as:

$$\begin{aligned}J_i^*(I(T+1),T)&=\min_{v_i(T)}\mathbf{E}\left[c(x_i(T),v_i(T))|I(T+1)\right]\\&=\min_{v_i(T)}\mathbf{E}\left[c(x_i(T),v_i(T))|I_i(T+1)\right]\\&\equiv J_i^*(I_i(T+1),T)\end{aligned}$$

because of the independence of $x_i(T)$ from other $x_j(T)$ and conditional independence of the observations of location i from those of other locations. These independence assumptions imply $p(x_i(T)|I(T+1))=p(x_i(T)|I_i(T+1))$. Thus, the optimal decision, $v_i(T)$, and the optimal cost-to-go will be a function only of $I_i(T+1)$ and not all of $I(T+1)$.

Assume inductively that for stages $\tau>t$, the optimal cost-to-go $J_i^*(I(\tau+1),\tau)\equiv J_i^*(I_i(\tau+1),\tau)$ depends only on the information collected at location i, and the strategy for the optimal decision $v_i(\tau)$ and measurements $u_{i,s}(\tau+1)$ for $s=1,\ldots,S$ depends only on $I_i(\tau+1)$ and not all of $I(\tau+1)$. Consider the minimization over the choice of $v_i(t),u_{i,s}(t+1)$, $s=1,\ldots,S$. Under γ, these are functions of the full information-state $I(t+1)$. Bellman's equation becomes:

$$\begin{aligned}J_i^*(I(t+1),t)=&\min_{v_i(t),u_{i,s}(t+1)}\mathbf{E}\left[c(x_i(t),v_i(t))+\right.\\&\sum_{s=1}^{S}\lambda_s(t+1)r_s(u_{i,s}(t+1))+\\&\left.\mathbf{E}\Big[J_i^*(I_i(t+2),t+1)|I(t+1),\{u_{i,s}(t+1)\}\Big]|I(t+1)\Big]\end{aligned}$$

Although all of the conditioning is in terms of the information set $I(t+1)$, the dependence will be only on $I_i(t+1)$ because of the independence assumptions, which imply that $p(x_i(t)|I(t+1))=p(x_i(t)|I_i(t+1))$ and:

$$\begin{aligned}&\mathbf{E}\Big[J_i^*(I_i(t+2),t+1)\Big|I(t+1),\{u_{i,s}(t+1)\}\Big]=\\&\mathbf{E}\Big[J_i^*(I_i(t+2),t+1)\Big|I_i(t+1),\{u_{i,s}(t+1)\}\Big]\end{aligned}$$

so the minimizing strategies for $v_i(t)$, $u_{i,s}(t+1)$ will only depend on $I_i(t+1)$. Let Γ_i^L denote the set of local feedback strategies γ_i that select decisions $u_{i,s}(t),v_i(t-1)$ depending on $I_i(t)$ only. By induction through stochastic dynamic programming, we have shown:

$$\begin{aligned}&\min_{\gamma\in\Gamma}\mathbf{E}_{\gamma}\left[\sum_{t=1}^{T}\left\{c(x_i(t),v_i(t))+\sum_{s=1}^{S}\lambda_s(t)r_s(u_{i,s}(t))\right\}\right]=\\&\min_{\gamma_i\in\Gamma_i^L}\mathbf{E}_{\gamma_i}\left[\sum_{t=1}^{T}\left\{c(x_i(t),v_i(t))+\sum_{s=1}^{S}\lambda_s(t)r_s(u_{i,s}(t))\right\}\right]\end{aligned}$$

Carrying the induction to the initial time yields

$$\begin{aligned}&\min_{\gamma\in\Gamma}\mathbf{E}_{\gamma}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{c(x_i(t),v_i(t))+\sum_{s=1}^{S}\lambda_s(t)r_s(u_{i,s}(t))\right\}\right]\geq\\&\sum_{i=1}^{N}\min_{\gamma_i\in\Gamma_i^L}\mathbf{E}_{\gamma_i}\left[\sum_{t=1}^{T}\left\{c(x_i(t),v_i(t))+\sum_{s=1}^{S}\lambda_s(t)r_s(u_{i,s}(t))\right\}\right]\end{aligned}$$

To complete the proof, note that feedback strategies of the form $\gamma=(\gamma_1,\gamma_2,\ldots,\gamma_N)$ are admissible strategies for the optimization problem on the left. Hence, the optimal local strategies, γ_i, achieve equality in the above equation, establishing the theorem. ■

The above theorem establishes that the optimization problem (5) can be decoupled into the sum of N local cost functions as follows:

$$\begin{aligned}J_\lambda=&\min_{\gamma\in\Gamma}\mathbf{E}_{\gamma}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}c(x_i(t),v_i(t))+\right.\\&\left.\sum_{s=1}^{S}\sum_{t=1}^{T}\sum_{i\in O_s(t)}\lambda_s(t)r_s(u_{i,s}(t))-\sum_{s=1}^{S}\sum_{t=1}^{T}\lambda_s(t)R_s\right]\\=&\sum_{i=1}^{N}\left\{\min_{\gamma_i\in\Gamma_L}\mathbf{E}_{\gamma_i}\left[\sum_{t=1}^{T}\left(c(x_i(t),v_i(t))\right.\right.\right.\\&\left.\left.\left.+\sum_{s=1}^{S}\lambda_s(t)r_s(u_{i,s}(t))\right)\right]\right\}-\sum_{t=1}^{T}\sum_{s=1}^{S}\lambda_s(t)R_s\end{aligned}\quad(8)$$

Solution of the decoupled problems in (8) for any trajectories of $\lambda_s(t)$ provides a lower bound to the performance of our original problem. In particular, we have

$$J^*\geq\sup_{\lambda_1,\ldots,\lambda_S\geq 0}J_{\lambda_1,\ldots,\lambda_S}\quad(9)$$

The problem in (9) is a dual optimization problem in terms of optimizing the multipliers associated with the constraints (5). Because of strong duality, this is the dual of the linear programming optimization problem that optimizes over mixtures $q(\gamma_i)$ of local feedback strategies (see [21], [23] for details)

$$\min_{q\in Q(\Gamma^L)}\sum_{\gamma\in\Gamma^L}q(\gamma)\sum_{i=1}^{N}\mathbf{E}_{\gamma_i}\left[\sum_{t=1}^{T}c(x_i(t),v_i(t))\right]\quad(10)$$

$$\sum_{\gamma\in\Gamma^L}q(\gamma)\sum_{i=1}^{N}\mathbf{E}_{\gamma_i}\left[\sum_{i\in O_s(t)}r_s(u_{i,s}(t))\right]\leq R_s\ \ \forall\, s,t\quad(11)$$

$$\sum_{\gamma\in\Gamma^L}q(\gamma)=1\quad(12)$$

where we have one constraint for each of the S sensor resource pools *for each time* t and an additional simplex constraint in (12), which ensures that $q \in Q(\Gamma^L)$ forms a valid probability distribution.

Note that the above linear program is an optimization problem over a very large space, mixtures of local feedback strategies. However, the number of constraints is equal to the number of sensors times the number of decision times, which means that the optimal mixture will be sparsely supported. This suggests the use of Column Generation [24] to obtain optimal mixtures, which operates as follows:

- Initalize the problem with a set of local strategies, Γ^{L_rest}.
- Solve the above linear programming (10)-(12) *restricted* to mixtures of strategies in Γ^{L_rest}, and compute the optimal dual variables $\lambda_s(t)$ for the constraints (11).
- Using these dual variables, sove the resulting decoupled POMDP problems for each location arising from the decomposition of Theorem 3.2 to obtain a new local strategy γ.
- If the new composite local strategy γ is not included in Γ^{L_rest}, add the strategy to Γ^{L_rest} and repeat. Otherwise, stop and the optimal mixture is provided by the solution of the linear program.

In consequence, we have *randomized strategies* that mix not only in terms of which sensor is utilized, but when and where a sensor is utilized. These strategies can be used to generate sensor control actions by various mechanisms such as sampling or truncation, as explored in [21].

IV. Simulation Example

Although the preceding implementation provides algorithms that handle arbitrary Markov processes at each location, we focus on demonstrating the algorithm for a *Markov Birth Process*, which is easier to implement with existing POMDP solvers. In this example, there were 100 locations, each of which could be empty, or have objects of three types, so the possible states of location i were $x_i \in \{0, 1, 2, 3\}$ where type 1 represents cars, type 2 trucks, and type 3 military vehicles. Sensors can have four observation modes: a search mode, a low resolution mode, a high resolution mode and a wait mode where no measurement is taken. The search mode primarily detects the presence of objects; the low resolution mode can identify cars, but confuses the other two types, whereas the high resolution mode can separate the three types. Observations are modeled as having three possible values. The search mode consumes 0.25 units of resources, whereas the low-resolution mode consumes 1 unit and the high resolution mode 5 units, uniformly for each sensor and location. Table I shows the conditional probability functions for the different sensing modes.

Initially, each location has a state with prior probability distributions described as $\pi_i(0) = [0.1\ 0.6\ 0.2\ 0.1]'$ $\forall\, i \in [1, \ldots, 10]$, $\pi_i(0) = [0.80\ 0.12\ 0.06\ 0.02]^T$ $\forall\, i \in [11, \ldots, 100]$. Thus, the first ten locations are likely to contain an object initially, and the subsequent 90 locations are likely to start as empty locations. The Markov chain model has transitions from the empty state 0 to the other states 1, 2, 3, such that

$$P(x_i(t+1) = 1 | x_i(t) = 0) = 0.06;$$
$$P(x_i(t+1) = 2 | x_i(t) = 0) = 0.03;$$
$$P(x_i(t+1) = 3 | x_i(t) = 0) = 0.01;$$

States 1, 2, 3 are absorbing states.

	Search			Low-res			Hi-res		
	y1	y2	y3	y1	y2	y3	y1	y2	y3
empty	0.92	0.04	0.04	0.95	0.03	0.02	0.95	0.03	0.02
car	0.08	0.46	0.46	0.05	0.85	0.10	0.02	0.95	0.03
truck	0.08	0.46	0.46	0.05	0.10	0.85	0.02	0.90	0.08
military	0.08	0.46	0.46	0.05	0.10	0.85	0.02	0.03	0.95

TABLE I

Observation likelihoods for different sensor modes with the observation symbols y1, y2 and y3.

We consider a problem with a single sensor that has a total of 100 units of resource, with 20 units to be used at each of f5 mesurement times, with a horizon corresponds to $T = 5$. After measurements are collected at each time, a tentative classification decision must be made, corresponding to $v_i(t) \in \{0, 1\}$, where 1 corresponds to a military vehicle and 0 corresponds to otherwise. The inremental cost of each decision is given as:

$$c(x_i, v_i) = \begin{cases} 0 & v_i = 0, x_i \in \{0, 1, 2\} \\ 0 & v_i = 1, x_i = 3 \\ 1 & \text{otherwise} \end{cases}$$

For this problem, we solve the associated POMDP problems using the Point Based Value Iteration Method, a fast POMDP technique [25], and we compute the optimal mixed strategies using column generation. Since we have a single sensor and 5 decision times, there are 5 dual variables to be determined in the algorithm, and the optimal mixed strategies are mixtures of 6 pure strategies. In order to obtain feasible decisions, we sample the mixed strategies according to their mixture probabilities in order to determine which strategy should be used for each location. To guarantee feasibility at each period, resources are assigned incrementally across locations, so that if there are not enough resources to implement the desired action for a location, the most informative feasible action is implemented instead.

As a reference set of strategies, we compare the performance of our strategies with a myopic based on entropy reduction. This algorithm selects actions as follows: for each location and potential action, one computes an index consisting of the expected reduction in sample entropy, or discrimination gain [8], per unit resource assigned. Actions are then assigned for each object in order of decreasing index until all resources for each interval are exhausted.

Table II compares compares the performance of our algorithm, termed Relaxed DP, with that of the alternative algorithm, Discrimination Gain, and the lower bound computed in our relaxation. The performance of the algorithms was computed averaging over 100 independent realizations of our experiments, while the bound was computed from

	Average Cost
Bound	64.45
Relaxed DP	70.81
Discrimination Gain	77.23

TABLE II
AVERAGE COST OBTAINED BY DIFFERENT ALGORITHMS FOR 100 MONTE CARLO SIMULATIONS.

a single run. The results illustrate that our Relaxed DP algorithm achieves performance closer to the lower bound than the alternative Discrimination Gain algorithm, as the strategies selected coordinated sensing activities across time, by recognizing the need to reserve resources to classify accurately future arrivals. Note that neither algorithm achieves performance close to the lower bound, suggesting that alternative approaches should be explored for determining feasible strategies from the optimal mixed strategies, such as the receding horizon approaches discussed in [21].

V. CONCLUSION

The problem of optimal control of observation processes is a complex, partially observed stochastic control problem that requires feedback from information states generated by the acquired information. When the problem involves observation of multiple locations with multiple sensors, the resulting combinatorial control problem is intractable, and requires approximations to obtain computable control strategies. In our previous work [20], [21], we have developed an approximate control strategy based on finding the optimal solution to a lower bound to the optimal cost, obtained through relaxation of sample path constraints and using mixed strategies to obtain a convex optimization problem that can be solved through combinations of decoupled stochastic dynamic programs for each location, coordinated by a master problem to select optimal prices of sensor resources.

The previous results applied only to locations where the unknown state was constant over time. In this paper, we extend these results to locations where the state can change according to a Markov chain, representing the possibility that objects arrive and depart. We obtain a similar lower bound to the previous results available in the literature, and develop a hierarchical algorithm that combines dynamic programming with column generation to obtain near-optimal feedback strategies.

The results of this paper provide a foundation for determining achievable performance of adaptive sensor control schemes. However, real-time implementation of these controllers is still computationally intensive, as it requires iterative solutions involving small POMDP problems. We are investigating alternative approaches for control of information acquisition that use other approximations involving information theory bounds, and hope to report on these results in future publications.

REFERENCES

[1] B. Koopman, *Search and Screening: General Principles with Historical Applications*. Pergamon, New York NY, 1980.

[2] S. J. Benkoski, M. G. Monticino, and J. R. Weisinger, "A survey of the search theory literature," *Naval Research Logistics*, vol. 38, no. 4, pp. 469–494, 1991.

[3] D. A. Castañón, "Optimal search strategies in dynamic hypothesis testing," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 25, no. 7, pp. 1130–1138, Jul 1995.

[4] A. Wald, "On the efficient design of statistical investigations," *The Annals of Mathematical Statistics*, vol. 14, pp. 134–140, 1943.

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

[6] H. Chernovv, *Sequential Analysis and Optimal Design*. SIAM, Philadelphia, PA, 1972.

[7] V. V. Fedorov, *Theory of Optimal Experiments*. Academic Press, New York, 1972.

[8] K. Kastella, "Discrimination gain to optimize detection and classification," *IEEE Transactions on Systems, Man and Cybernetics, Part A*, vol. 27, no. 1, pp. 112–116, Jan. 1997.

[9] C. Kreucher, K. Kastella, and I. Alfred O. Hero, "Sensor management using an active sensing approach," *Signal Processing*, vol. 85, no. 3, pp. 607–624, 2005.

[10] K. L. Jenkins and D. A. Castañón, "Receding horizon stochastic control algorithms for sensor management," in *Proc. Conference on Decision and Control*, Atlanta, GA, Dec 2010.

[11] M. Athans, "On the determination of optimal costly measurement strategies for linear stochastic systems," *Automatica*, vol. 8, no. 4, pp. 397–412, 1972.

[12] V. Krishnamurthy and R. Evans, "Hidden Markov model multiarm bandits: a methodology for beam scheduling in multitarget tracking," *IEEE Transactions on Signal Processing*, vol. 49, no. 12, pp. 2893–2908, Dec 2001.

[13] R. Washburn, M. Schneider, and J. Fox, "Stochastic dynamic programming based approaches to sensor resource management," in *Proceedings of the 5th International Conference Information Fusion, 2002*, vol. 1, 2002, pp. 608–615.

[14] J. C. Gittins, "Bandit processes and dynamic allocation indices," *Journal of the Royal Statistical Society. Series B*, vol. 41, no. 2, pp. 148–177, 1979.

[15] W. Macready and I. Wolpert, D.H., "Bandit problems and the exploration/exploitation tradeoff," *IEEE Transactions on Evolutionary Computation*, vol. 2, no. 1, pp. 2–22, Apr. 1998.

[16] C. Kreucher and A. Hero, "Monte carlo methods for sensor management in target tracking," in *IEEE Nonlinear Statistical Signal Processing Workshop*, 2006.

[17] E. Chong, C. Kreucher, and A. Hero, "Monte-carlo-based partially observable Markov decision process approximations for adaptive sensing," *International Workshop on Discrete Event Systems*, pp. 173–180, May 2008.

[18] D. A. Castañón and L. Carin, "Stochastic control theory for sensor management," in *Foundations and Applications of Sensor Management*, A. Hero, D. Castañón, D. Cochran, and K. Kastella, Eds. New York, NY: Springer Verlag, 2008.

[19] D. A. Castañón, "Approximate dynamic programming for sensor management," in *Proceedings 36th IEEE Conference on Decision and Control*, 1997, pp. 1202–1207.

[20] ——, "Stochastic control bounds on sensor network performance," in *Proceedings 44th IEEE Conference on Decision and Control*, Dec. 2005, pp. 4939–4944.

[21] D. C. Hitchings and D. A. Castañón, "Receding horizon stochastic control algorithms for sensor management," in *Proc. American Control Conference*, Baltimore, MD, June 2010.

[22] D. A. Castañón, "A lower bound on adaptive sensor management performance for classification," 2005, preprint (2005).

[23] D. C. Hitchings, "Near-optimal multi-platform search and exploitation with humans in the loop," Ph.D. dissertation, Boston University, May 2010.

[24] P. C. Gilmore and R. E. Gomory, "A linear programming approach to the cutting-stock problem," *Operations Research*, vol. 9, no. 6, pp. 849–859, 1961.

[25] J. Pineau, G. Gordon, and S. Thrun, "Point-based value iteration: An anytime algorithm for POMDPs," in *International Joint Conference on Artificial Intelligence (IJCAI)*, Aug. 2003, pp. 1025–1032.