Sequential Inspection Using Loitering*

Meir Pachter[1], Swaroop Darbha[2], and P. Chandler[3]

[1] Professor, Dept. of Electrical and Computer Engineering
[2] Visiting Professor, AFRL/VACA, & Associate Professor, Department of Mechanical Engineering, Texas A&M University at College Station
[3] Air Force Research Laboratory, AFRL/VACA, Department of Electrical and Computer Engineering, Air Force Institute of Technology, Wright-Patterson Air Force Base, OH 45433

Abstract. A set of objects of interest is to be sequentially inspected by a Micro Aerial Vehicle (MAV) equipped with a camera. Upon arriving at an object of interest, an image of the object is sent to a human operator, who, upon inspecting the image, sends his feedback to the MAV. The feedback from the operator may consist of the pose angle of the object and whether he has seen any distinguishing features of the object. Upon receiving the feedback, the MAV uses this information to decide whether it should perform a secondary inspection of the object of interest or continue to the next object. A secondary inspection has a reward (or value or information gain) that is dependent on the operator's feedback. There is an associated cost of reinspection and it depends on the delay of the operator's feedback. It seems reasonable to let the MAV loiter for a while near the most recently inspected object of interest so that it expends a small amount of endurance from the reserve after receiving the feedback from the operator. The objective is to increase the information and hence, the total expected reward about the set of objects of interest. Since the endurance of the MAVs is limited, the loiter time near each object of interest must be carefully determined. This paper addresses the determination of the optimal loiter time through the use of Stochastic dynamic programming. Numerical results are presented that show the optimal loiter time is a function of the maximum expected operator delay.

1 Introduction

The following inspection scenario is considered. A set of n objects of interest O_i, $i = 1, ..., n$, is sequentially visited by an MAV equipped with a camera. Upon arriving at an object of interest, an image of the object is sent to a human operator for classification. The operator, upon inspecting the image, sends his feedback, e.g., the object's pose angle and whether he has seen a distinguishing feature in the object's image, to the MAV. When the operator's feedback is received, the MAV must make a decision whether it should revisit the object for a secondary

* The views expressed in this article are those of the authors and do not reflect the official policy or position of the United States Air Force, Department of Defence, or U.S. Government.

M.J. Hirsch et al. (Eds.): Adv. in Cooper. Ctrl. & Optimization, LNCIS 369, pp. 281–291, 2007.
springerlink.com

inspection. The information gain (reward) associated with a secondary inspection is dependent on the feedback from the operator. Each MAV has a finite endurance reserve and a revisit of an object requires expenditure from the reserve. This expenditure is a function of the operator's delay and the action that the MAV decides to take. The operator functions as a sensor/classifier in the inspection loop and the MAV decides on the course of action. The objective of the decision making is to maximize the total expected reward given the constraints on the endurance.

The MAV makes decisions *sequentially* based on the information available to it - operator's delay and feedback about the object, the number of objects left to be visited by the MAV, and the current reserve. While the operator's delay, τ_i, associated with the i^{th} object is a random variable (whose probability density function (p. d. f) $f(\tau_i)$ is known), we emphasize that the realized value of this random variable may only be known at the time of decision making. We do not allow for the possibility of the MAV revisiting an object more than once or revisiting an object after it has decided to go to the next object in the sequence. At the time of making a decision, the actions that we allow the MAVs to take are the following: loiter around the object; move onto the next object; or revisit the object. Associated with the i^{th} object, there is a continuous decision variable, u_i, which indicates the maximum allowable loitering time and a binary decision variable, v_i, which indicates whether the object should be revisited.

The motivation for the introduction of loitering is as follows: If the MAV were to move away from the object after the first visit, then the time (and hence, expenditure of the reserve) for a revisit is at least twice the operator's delay; by allowing the MAV to loiter near the object, the time to get back is shortened. If the operator's feedback is received by the MAV before the maximum allowable loiter time, only then does the MAV take a decision about revisiting the object; otherwise, it will go to the next object in the sequence. Since the objective is to increase the information about the set of objects of interest and since the MAV's endurance reserve c_1 is limited, the maximum allowable loiter time associated with the objects must be carefully determined.

The paper is organized as follows. In Section 2, a stochastic optimal control problem which models the class of decision scenarios at hand is formulated. In Section 3, the method of Dynamic Programming is brought to bear on the sequential decision problem and numerical results corroborating the methodology presented for the decision problems considered in this paper are provided. In Section 4, a generalization of the present formulation is explored. Conclusions are drawn in Section 5.

We use the following notation throughout the paper:

i	Index of stage in a Stochastic Dynamic Program (SDP)
p	*a priori* probability
r_i	Running reward for the i^{th} stage
u_i	Decision variable at the i^{th} stage
τ_i	Delay in communicating the first observation
$f(\tau_i)$	Probability density function (p.d.f) of the operator delay τ_i
τ_0	Fixed communication delay

2 Stochastic Optimal Control

The Dynamic Program (DP) has n stages and one state variable, c_i - the endurance reserve on arrival to O_i. The operator's delay at O_i, $0 \leq \tau_i \leq \tau_{max}$ is a random variable whose realization may *not* be known at the time of making a decision. The p.d.f. of τ_i is $f(\tau_i)$ and is assumed known. The decision/control variable is u_i, the *maximal* loiter or waiting time at O_i.
The nonlinear dynamics are driven by the control variable u and by the random variable τ:

$$c_{i+1} = c_i - \min\ (u_i,\ \tau_i)\ \ i = 1, ..., n \tag{1}$$

The initial reserve, c_1, is known. The control variable is constrained according to

$$0 \leq u_i \leq \min(c_i, \tau_{\max}) \tag{2}$$

and the random variable τ_i is characterized by its p.d.f. $f(\tau)$.

The running payoff is

$$r_i(u_i, \tau_i) = \begin{cases} 1 & \text{if } \tau_i < u_i \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

The payoff function

$$J(u_1, ..., u_n; c_1) = \mathcal{E}_{\tau_1,...,\tau_n}\ (\ \sum_{i=1}^{n} r_i(u_i, \tau_i)\ ) \tag{4}$$

The optimal strategy is a state feedback control law $u_i^*(c_i),\ \ i = 1, ..., n$.

3 Dynamic Programming Recursion

The stochastic optimal control problem (1)-(4) is solved using the method of Dynamic Programming (DP). We emphasize that the realization of the random variable may not be known at the time of making a decision.
We shall require the following definition:

$$p(u) \equiv \int_u^{\tau_{max}} f(\tau) d\tau$$

The term $p(u_i)$ is the probability that the MAV comes out empty handed, that is, the MAV leaves the vicinity of O_i after waiting for a time u_i without receiving the operator's feedback.

The value function $V_i(c_i)$ is the maximal expected reward at the time of making a decision concerning O_i, given the endurance reserve of the MAV at O_i is c_i.

The DP recursion is established as follows: For each $u_i \in [0, \min\{c_i, \tau_{\max}\}]$,

$$\begin{aligned}
V_i(c_i|u_i,\tau_i) &= r_i(u_i,\tau_i) + V_{i+1}(c_i - \min(u_i,\tau_i)) \\
\Rightarrow V_i(c_i|u_i) &= \mathcal{E}_{\tau_i}[r_i(u_i,\tau_i) + V_{i+1}(c_i - \min(u_i,\tau_i))] \\
&= [0 + V_{i+1}(c_i - u_i)]p(u_i) + \int_0^{u_i} [1 + V_{i+1}(c_i - \tau)]f(\tau)d\tau \\
V_i(c_i) &= \max_{0 \le u_i \le \min\ (c_i, \tau_{\max})} \mathcal{E}_{\tau_i}\ (\ r_i(u_i,\tau_i) + V_{i+1}(c_{i+1}(c_i,u_i,\tau_i))\ ) \\
&= \max_{0 \le u_i \le \min\ (c_i, \tau_{\max})} \{[0 + V_{i+1}(c_i - u_i)]p(u_i) + \\
&\quad \int_0^{u_i} [1 + V_{i+1}(c_i - \tau_i)]f(\tau_i)d\tau_i\} \\
&= \max_{0 \le u_i \le \min\ (c_i, \tau_{\max})} [p(u_i)V_{i+1}(c_i - u_i) + 1 - p(u_i) + \\
&\quad \int_0^{u_i} V_{i+1}(c_i - \tau_i)f(\tau_i)d\tau_i]
\end{aligned}$$

Hence, the DP recursion is

$$V_i(c_i) = 1 + \max_{0 \le u_i \le \min\ (c_i, \tau_{\max})} [p(u_i)V_{i+1}(c_i - u_i) - p(u_i) + \int_0^{u_i} V_{i+1}(c_i - \tau_i)f(\tau)d\tau] \quad (5)$$

Assuming that the value function, V_{i+1}, is known, the above recursion allows one to compute the value function V_i. The optimal control,

$$u_i^*(c_i) = \arg \max_{0 \le \min c_i, \tau_{\max}} [p(u_i)V_{i+1}(c_i - u_i) - p(u_i) + \int_0^{u_i} V_{i+1}(c_i - \tau_i)f(\tau)d\tau]. \quad (6)$$

For one to begin the recursion, the value function V_n must be specified. This is presented in the next subsection.

3.1 The Boundary Condition

Obviously, at time n

$$u_n^*(c_n) = c_n$$

Hence, if $c_n > \tau_{\max}$, $r_n = 1$ and therefore $V_n(c_n) = 1$. If, however, $c_n \le \tau_{\max}$, then

$$\begin{aligned}
V_n(c_n) &= \mathcal{E}_{\tau_n}\ (\ r_n(c_n,\tau_n)\ ) \\
&= 0 \cdot p(c_n) + 1 \cdot (1 - p(c_n)) \\
&= 1 - p(c_n)
\end{aligned}$$

Thus, the boundary condition is

$$V_n(c_n) = \begin{cases} 1 & if \quad c_n > \tau_{max} \\ 1 - p(c_n) & if\ 0 \le c_n \le \tau_{max} \end{cases} \quad (7)$$

3.2 Computing Suboptimal Value Functions

In the examination of sensitivity of the value function to a perturbation in the probability distribution of the operator's feedback delay, one is interested in the computation of suboptimal value functions. Let $U_i(c)$ be the sub-optimal value function for a given strategy. Let $f_p(\tau)$ denote the perturbed distribution and $\tilde{\tau}_{max}$ denote the maximum value of the corresponding delay. Let $\tilde{p}(c)$ denote the integral

$$\int_c^{\tilde{\tau}_{\max}} f_p(\tau)d\tau.$$

The function $U_i(c)$ may be computed recursively as follows:

$$U_n(c) = \begin{cases} 1 - \tilde{p}(c) & c \in [0, \tilde{\tau}_{max}], \\ 1 & c \geq \tilde{\tau}_{max} \end{cases}$$

and

$$U_i(c|u_i^*(c)) = 1 + [\tilde{p}(u_i^*)U_{i+1}(c_i - u_i^*) - \tilde{p}(u_i^*) + \int_0^{u_i^*} U_{i+1}(c_i - \tau_i)f_p(\tau)d\tau].$$

3.3 Numerical Implementation

For the purposes of implementation, we discretize the cumulative density function, $f_c(\tau)(:= \int_0^\tau f(\eta)d\eta)$ and deal with the corresponding discrete probability density function. We specifically assume the following: the maximum reserve $c_{\max}$ and the maximum number of objects of interest ($N_{\max}$) are known a priori. Further, we assume the reserve to be an integer multiple of a fixed increment of reserve, Δ, i.e., $c = k\Delta$ for some $k \geq 0$, and that the fixed delay, τ_0 and the delay, τ are also integral multiples of a fixed increment of reserve, i.e., $\tau = l\Delta$ for some integral $l, l_0 \geq 0$. Since the delay can only take discrete values (which are integral multiples of Δ), one may approximate the continuous p.d.f by a discrete p.d.f. for $f(\tau)$ as: $P(\tau = l\Delta) = p_l$ and hence,

$$f(\tau) = \sum_{j=1}^{\infty} p_j \delta(\tau - j\Delta),$$

where $\delta(k) = 1$ if $k = 0$ and is 0 otherwise.

The value function $V_n(k\Delta)$ may be readily computed from the discretization of the cumulative distribution function as follows:

$$V_n(k\Delta) = \sum_{l \leq k} p_l.$$

Clearly, if $l\Delta > \tau_{\max} = D_{\max}\Delta$, then $V_n(l\Delta) = 1$. The optimal decision, u_n^* is specified by the following equation:

$$u_n^*(k\Delta) = k\Delta.$$

The recursive equations for the value functions, $V_i(k\Delta),\ i < n$, can be expressed as follows:

$$V_i(k\Delta) = 1 + \min_{0\leq j\leq \min\{k, D_{max}\}} (\sum_{l>j} V_{i+1}((k-j)\Delta) - \sum_{l>j} p_l + \sum_{l\leq j} V_{i+1}(k-l)p_l).$$

The corresponding optimal decisions are:

$$u_i^*(k\Delta) = \Delta \arg \min_{0\leq j\leq \min\{k, D_{max}\}} (\sum_{l>j} V_{i+1}((k-j)\Delta) - \sum_{l>j} p_l + \sum_{l\leq j} V_{i+1}(k-l)p_l).$$

Real-time Implementation: Once the $V_i(c)$ and $u_i^*(c)$ is computed for each $c \in [0, C_{\max}]$ and $i = 1, \ldots, N_{\max}$, it is stored as two matrices in the MAV's on-board processor. From the knowledge of c and the number of objects to visit, one can compute the relevant optimal decision (waiting time) from the table. If the operator does not provide any feedback before the optimal waiting time, then the MAV moves onto the next object; if he does provide feedback, it will revisit the object.

3.4 Numerical Results

We have considered the following case: The maximum number, $N_{\max}$ of objects to visit is 20, the maximum reserve, $c_{\max} = 1000$ *units*, the maximum delay is 200 *units* and $\Delta = 1$ *unit*. The operator delay is initially assumed to be uniform. The corresponding value functions and optimal decisions are shown in Figures 1 and 2.

From Figure 1, we can see that the expected number of revisits increases monotonically with the reserve and the number of objects to visit. From Figure 2, for any fixed reserve (smaller than the maximum operator's feedback delay), the optimal wait time decreases with the number of the objects to visit. This is consistent with our intuition. In particular if the number of objects is arbitrarily large, then the optimal waiting time is the minimum delay of 1 unit and the maximum number of revisits that are possible is c if the initial reserve is c units.

In order to examine the sensitivity of the optimal strategy to the distribution of the operator's feedback delay, we perturbed the distribution. When we refer to a sub-optimal value function, we decide on a revisit based on the optimal waiting time obtained when the operator's delay is uniformly distributed between 1 and 200 seconds. An optimal value function will correspond to the optimal waiting time associated with the perturbed distribution. We considered four perturbed distributions and associated with each perturbed distribution there is a sub-optimal revisit function and a sub-optimal value function. The case when the operator's feedback delay is randomly (as opposed to uniformly) distributed between 1 and 200 units of reserve shows little difference with the optimal (uniform) revisit and value functions.

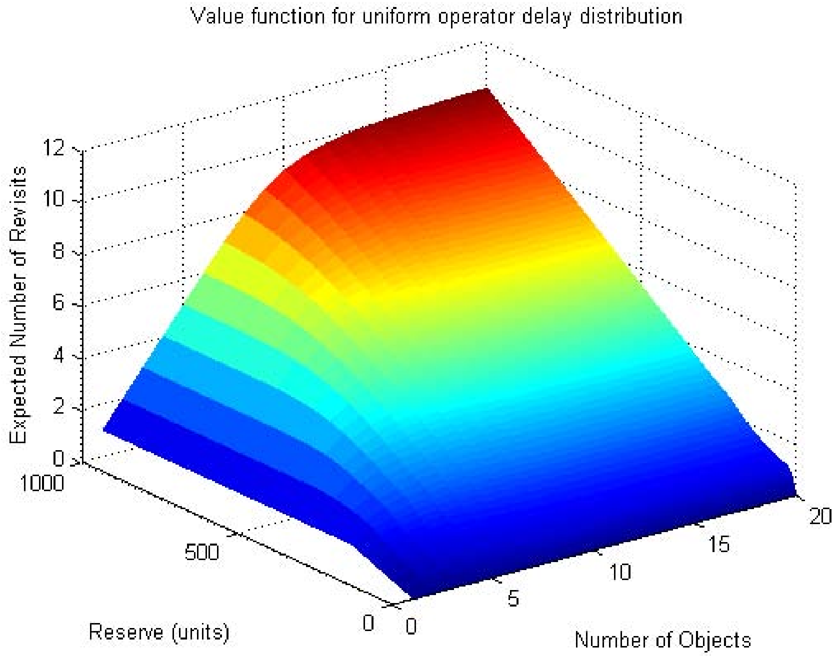

Fig. 1. Value function for uniform operator delay

Three other cases with varying degrees of randomness in operator delay were tested and similar results were found. The trends remained the same with minor variations in the expected number of revisits and the optimal waiting time. While the overall performance (expected number of revisits) is sensitive to the maximum operator delay assumed in the derivation of the optimal decisions, it is not sensitive to the exact distribution of the delay. In particular, if the maximum delay was underestimated, the performance deteriorated (i.e., the expected number of revisits was significantly smaller than the case corresponding to the knowledge of the exact value of the maximum delay) and if the maximum delay was overestimated, the degradation in the performance was insignificant. Essentially, it is better to overestimate the maximum value of the operator's feedback delay as opposed to underestimating it.

4 Generalization: Including an Endurance Cost When Revisiting an Object

Suppose there are L different ways to revisit an object. Suppose there is an overhead cost of τ_{ij} units of reserve when the i^{th} object is revisited in the j^{th} way. Let v_{ij} be a binary variable that takes a value 1 if the i^{th} object is revisited in the j^{th} way and is 0 otherwise. Then, the governing constraints may be expressed as:

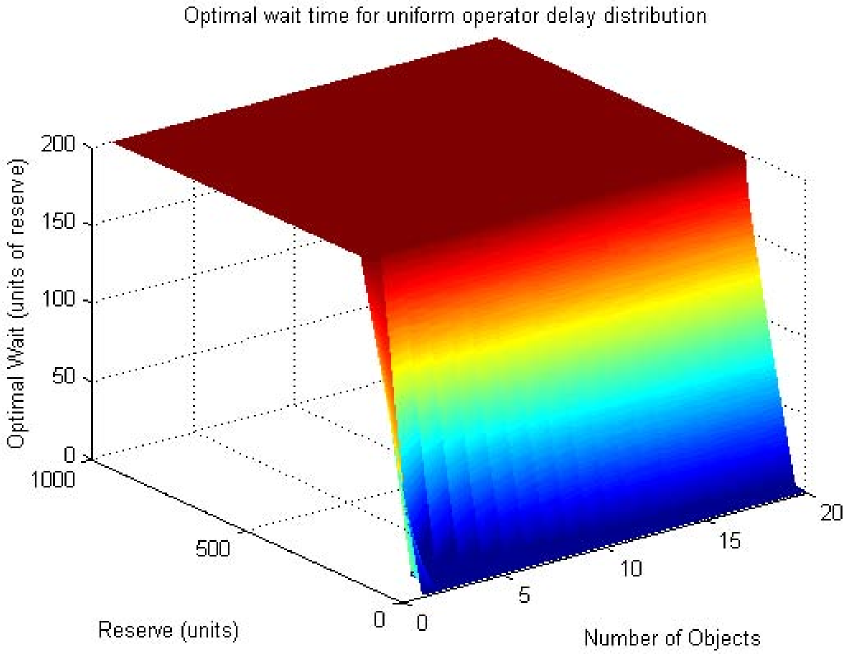

Fig. 2. Optimal wait for uniform operator delay

$$c_{i+1} = c_i - \min(u_i, \tau_i) - \sum_{j=1}^{L} \tau_{ij} v_{ij} \tag{8}$$

$$0 \leq c_i - \min(u_i, \tau_i) - \sum_{j=1}^{L} \tau_{ij} v_{ij}, \tag{9}$$

$$\sum_{j=1}^{L} v_{ij} = \begin{cases} 1 & \textit{if } u_i \geq \tau_i \;\; \textit{and} \;\; \tau_i \leq c_i - \min_j \tau_{ij} \\ 0 & \textit{otherwise.} \end{cases} \tag{10}$$

The last constraint indicates that a MAV will always revisit an object in exactly one way if the operator's feedback delay is smaller than the waiting time associated with the object provided it has sufficient reserve and will not revisit otherwise.

In plain words, the problem may be stated as follows: The MAV can revisit objects after it receives a feedback from the operator in one of L ways. The MAV does not revisit if the operator does not provide his feedback within the optimal waiting time associated with each object in the specified sequence. A revisit of the i^{th} object in the j^{th} way fetches a reward or payoff of β_{ij}. At the time of receiving the feedback, if it happens to be smaller than the optimal waiting time, the MAV must decide which of the L following ways it should revisit so as to maximize the total expected payoff.

One can think of a different waiting time for each way of revisiting the object; however, for simplicity, we do not consider such a scheme in this paper.

Mathematically, the payoff for visiting the i^{th} object in the k^{th} way is feasible only if the operator's feedback delay is smaller than the waiting time set for that object and there is sufficient reserve to revisit the object of interest. Hence,

$$r_i(u_i, \tau_i | v_{ik} = 1) = \begin{cases} \beta_{ik} & \text{if } \tau_i \leq \min\{c_i - \tau_{ik} v_{ik}, u_i, \tau_{max}\} \\ 0 & \text{otherwise.} \end{cases}$$

We shall assume, henceforth, the following: $\beta_{i1} > \beta_{i2} > \ldots > \beta_{iL}$ and correspondingly, $\tau_{i1} > \tau_{i2} > \ldots > \tau_{iL}$. This assumption implies that to get a higher payoff, one must pay a higher overhead cost (reserve).

Let

$$J = \mathcal{E}_{\tau_1,\ldots,\tau_n}[\sum_{i=1}^{n} \sum_{j=1}^{L} \underbrace{r_i(u_i, \tau_i | v_{ij} = 1) v_{ij}}_{r_i(u_i, \tau_i, v_{ij})}].$$

The term J indicates the total expected payoff for any given set of decisions, $u_i, v_{ij}, j = 1, \ldots, L,\ i = 1, \ldots, n$. The objective of the optimization is to maximize the expected payoff, J, over the possible set of decisions, $u_i,\ i = 1, \ldots, n$ and $v_{ij}, j = 1, \ldots, L,\ i = 1, \ldots, n$.

Let

$$V_i(c_i) := \max_{u_k, v_{kj},\ i \leq k \leq n,\ 1 \leq j \leq L} \mathcal{E}_{\tau_i,\ldots,\tau_n}[\sum_{k=i}^{n} \sum_{j=1}^{L} r_k(u_k, \tau_k, v_{ik})]. \tag{11}$$

One can then use DP to get the following recursion:
$V_i(c_i | u_i, \tau_i) =$

$$\begin{cases} V_{i+1}(c_i - u_i) & \text{if } \tau_i > \min\{u_i, c_i - \tau_{iL}\}, \\ \max_{1 \leq j \leq L}\{\beta_{ij} + V_{i+1}(c_i - \tau_{ij} - \tau_i) \colon c_i - \tau_i - \tau_{ij} \geq 0\} & \text{otherwise.} \end{cases}$$

Let $\mathcal{D}_{ij} := \{\tau : j = \arg\max_{1 \leq k \leq L}\{\beta_{ij} + V_{i+1}(c_i - \tau_{ik} - \tau_i) :\ c_i - \tau_i - \tau_{ik} \geq 0\}\}$. Therefore,

$$\begin{aligned} V_i(c_i | u_i) &= \mathcal{E}_\tau(V_i(c_i | u_i, \tau_i)) \\ &= \int_{\min(u_i, c_i - \tau_{iL})}^{\tau_{max}} V_{i+1}(c_i - u_i) f(\tau) d\tau + \\ &\quad \sum_{j=1}^{L} \int_{\tau \in \mathcal{D}_{ij}} (\beta_{ij} + V_{i+1}(c_i - \tau_{ij} - \tau_i)) f(\tau) d\tau. \end{aligned}$$

Hence,

$$V_i(c_i) = \max_{u_i \in (0, \min(c_i, \tau_{max}))} V_i(c_i | u_i).$$

The boundary condition that completes the recursion and enables the determination of all the value functions. Clearly, the waiting time if there is only one object to visit is equal to the reserve—$u_n^*(c_n) = c_n$ and correspondingly:

$$V_n(c_n|\tau_n) = \begin{cases} \beta_{nj} & \tau_n \in (c_n - \tau_{n,j-1}, c_n - \tau_{nj}] \\ 0 & \tau_n > c_n - \tau_{nL}. \end{cases}$$

Let $\tau_{n,0} = c_n$ and

$$V_n(c_n) = \sum_{j=1}^{L} \int_{c_n - \tau_{n,j-1}}^{c_n - \tau_{n,j}} \beta_{nj} f(\tau) d\tau.$$

Once the value functions are computed, the computation of optimal waiting time is straight forward:

$$u_i^*(c_i) = argmax_{u_i \in (0, \min(c_i, \tau_{max}))} V_i(c_i|u_i), \; i = 1, \ldots, n-1,$$
$$u_n^*(c_n) = c_n.$$

The optimal decisions to revisit are as follows:

$$v_{ij}^*(c_i, \tau_i) = \begin{cases} 1 & \text{if } \tau_i \in \mathcal{D}_{ij}(c_i, u_i^*(c_i)), \\ 0 & \text{otherwise.} \end{cases}$$

5 Conclusion

We have observed in numerical simulations that the performance (i.e., the expected number of revisits) of the sequential inspection decision system is sensitive to the assumed value of the maximum operator delay, but not sensitive to the actual distribution of the delay. The structure of the strategy is reasonably simple for its actual real-time implementation on the MAVs: We store the optimal wait time as a function of the reserve and the number of objects to visit and based on the operator's delay, decide on the future course of action. If the operator's delay is smaller than the optimal wait time associated with the MAV's reserve and the number of objects to revisit, an appropriate action for revisiting the object is taken; otherwise, it is optimal for the MAV to continue to the next object in the sequence. The optimal loiter time comes from solving the stochastic dynamic programming problem.

References

1. Gross D, Rasmussen S, Chandler PR, Feitshans G. "Cooperative Operations in UrbaN TERrain (COUNTER)," *2006 SPIE Defense & Security Symposium*, Orlando, FL, 17-21 April 2006.
2. Derman, C., Lieberman G. J., and Ross, S. M.: "A Sequential Stochastic Assignment Problem", Management Science, Vol. 18, No. 7, Theory Series, pp. 349-355, March 1972.

3. Ross SM. *Introduction to Stochastic Dynamic Programming*, Academic Press, 1982.
4. Pachter M, Chandler PR, Darbha S. “Optimal Control of an ATR Module Equipped MAV–Human Operator Team,” to appear in Vol. 588 of *Lecture Notes in Economics and Mathematical Systems*, Springer 2007.
5. Girard AR, Pachter M, Chandler PR. “Decision Making Under Uncertainty and Human Operator Model for Small UAV Operations,” *Proceedings of the AIAA Guidance, Navigation and Control Conference*, Keystone, CO, August 2006.
6. Pachter, M., P. R. Chandler and Swaroop Darbha, “Optimal Sequential Inspections,” *Proceedings of the IEEE Conference on Decision and Control*, December 2006.
7. Pachter, M., P. R. Chandler and Swaroop Darbha, “Robust Control of Unmanned Aerial Vehicles in Uncertain Environments,” to appear in *International Journal of Robust and Nonlinear Control*, 2007.
8. R. J. Gallager: “Information Theory and Reliable Communication”, Wiley 1968, pp. 18.
9. Parzen, E. *Modern Probability Theory and its Applications*, Wiley 1960.