Adaptive Multi-Aspect Target Classification and Detection with Hidden Markov Models

Shihao Ji, Xuejun Liao, *Senior Member, IEEE,* and Lawrence Carin, *Fellow, IEEE*

***Abstract*— Target detection and classification are considered based on backscattered signals observed from a sequence of target-sensor orientations, with the measurements performed as a function of orientation (angle) at a fixed range. The theory of optimal experiments is applied to adaptively optimize the sequence of target-sensor orientations considered. This is motivated by the fact that if fewer, better-chosen measurements are used then targets can be recognized more accurately with less time and expense. Specifically, based on the previous sequence of observations $\boldsymbol{o}_t = (o_1, \cdots, o_t)$, the technique determines what change in *relative* target-sensor orientation $\Delta\theta_{t+1}$ is optimal for performing measurement $t+1$, to yield observation o_{t+1}. The target is assumed distant or hidden, and therefore the *absolute* target-sensor orientation is unknown. We detail the adaptive-sensing algorithm, employing a hidden Markov model (HMM) representation of the multi-aspect scattered fields, and example classification and detection results are presented for underwater targets using acoustic scattering data.**

***Index Terms*— Classification, detection, HMM, entropy, optimal experiments.**

I. Introduction

There are many sensing scenarios for which the target is stationary and the sensor position may be moved sequentially, allowing observation of the unknown and concealed target from a sequence of target-sensor orientations. For example, an unmanned underwater or airborne vehicle (UUV or UAV) may observe a target from a sequence of orientations. Such multi-aspect sensing is also widely observed in nature, for example in bats [1]. Multi-aspect sensing exploits the fact that many targets generate scattered fields that are a strong function of the target-sensor orientation. This strong aspect dependence complicates the sensing task, but it may also enhance classification and detection performance. For example, one often finds that two targets T_1 and T_2 scatter waves in a nearly identical manner, when observed from distinct target-sensor orientations θ_1 and θ_2, respectively. The *absolute* target-sensor orientation θ_1 or θ_2 is typically unknown, since the target is distant or concealed. The classification ambiguity may be resolved if the targets are observed at a sequence of orientations (for which only the *relative* target-sensor orientations are known), since two targets are less likely to generate the same *sequence* of observations.

It has been demonstrated that multi-aspect target scattering is often well characterized via a hidden Markov model (HMM) [2]–[5]. Specifically, the target-sensor orientations, denoted by angle θ, are divided into a set of states. Each state is defined as a contiguous subset of angles for which the associated scattered fields are approximately stationary. When performing a sequence of scattering measurements, from a sequence of target-sensor orientations, one implicitly samples scattered waveforms from a sequence of target states. It has been demonstrated that the sequence of sampled states is often modeled well as a Markov process [2]–[5]. In practice the underlying sampled states are unobserved, or "hidden", and only the associated scattered fields are observed. This therefore yields the aforementioned *hidden* Markov model.

For the detection phase, it is assumed that one has available an HMM for the target of interest (or HMMs for multiple targets of interest). The sequence of observations $\boldsymbol{o}_t = (o_1, \cdots, o_t)$ is associated with a target if the HMM likelihood is greater than a threshold; by varying the threshold the receiver operating characteristic (ROC) is generated. When performing classification, it is assumed that the object under interrogation is one of a finite number of known targets, each described by a distinct HMM. After performing a sequence of observations $\boldsymbol{o}_t$, from t target-sensor orientations, classification is effected by determining which HMM was most likely to have generated $\boldsymbol{o}_t$. Assuming that $p(\boldsymbol{o}_t|T_k)$ represents the likelihood of observing $\boldsymbol{o}_t$ for target T_k, the sequence $\boldsymbol{o}_t$ is associated with target T_i if $p(\boldsymbol{o_t}|T_i) > p(\boldsymbol{o}_t|T_k), \forall T_k \neq T_i$.

Assume that the sensor has observed the sequence $\boldsymbol{o}_t$, with *relative* angular change between consecutive measurements $\Delta\boldsymbol{\theta}_t = (\Delta\theta_2, \cdots, \Delta\theta_t)$. This paper addresses the question of determining what change in angular position $\Delta\theta_{t+1}$ should be considered for collection of observation o_{t+1}, with the goal of optimizing detection and classification. We solve this problem by exploiting the theory of optimal experiments. Several decades ago Fedorov wrote a book on this theory [6], and more recently this topic has attracted the attention of the machine-learning community, in the context of adaptive algorithm training [7]. As discussed further below, for the classification problem the choice of $\Delta\theta_{t+1}$ is similar to that reported in [6], [7], in which we employ a measure of entropy. The key distinction is that in most previous optimal-experiment studies it has been assumed that the measurements are statistically independent [6], [7]. This assumption is required to achieve a simple and usable form of the adaptive-sensing cost function. We demonstrate here that the HMM assumption yields a cost function that may also be evaluated easily, constituting a sensing process that accounts for statistical dependence between the sequence of observations. We also extend this concept to the problem of target detection.

The remainder of the paper is organized as follows. In Sec. II we review the HMM as applied to multi-aspect sensing, and

S. Ji, X. Liao and L. Carin are with the Department of Electrical and Computer Engineering, Duke University, Box 90291, Durham, NC 27708-0291.

demonstrate how it may be employed to yield a simple and effective sequential sensing strategy for detection and classification. Example sequential detection and classification results are presented in Sec. III, for multi-aspect acoustic sensing of underwater elastic targets. The results of the adaptive sensing process are compared to results using uniform (non-adaptive) angular sampling. Conclusions and future work are described in Sec. IV.

II. Optimal Multi-Aspect Detection and Classification

A. Geometric HMM with an angle-dependent state-transition matrix

The hidden Markov model (HMM) has been employed previously for the analysis of multi-aspect target identification [2]–[5]. In this context an HMM state corresponds to a contiguous range of target-sensor angles for which the associated scattered signal is approximately stationary as a function of orientation. Assume that state m of a given target resides within the angular region $\theta_{m,1} \leq \theta \leq \theta_{m,2}$, with the angular extent of state m denoted by $\hat{\theta}_m = \theta_{m,2} - \theta_{m,1}$. Based on simple geometrical considerations, and assuming that all target orientations are uniformly probable when performing the first measurement, the probability of being in state m on the first measurement is computed as [2]

$$\pi_m^{(k)} = \hat{\theta}_m^{(k)} / \sum_{i=1}^{M_k} \hat{\theta}_i^{(k)} \tag{1}$$

for M_k target states. In (1) we emphasize that the initial-state-probability vector is a function of which target is considered, here represented for target T_k. Note that we here assume that the measurements are only a function of a single angle, over 2π; this concept may be extended to motion in two angular dimensions, over a 4π solid angle.

For each state we define the likelihood of making an associated observation (or measurement) o. Let $p(o|s_m, T_k)$ represent the likelihood of observing o in state m (denoted s_m) of target T_k. There are numerous ways to define $p(o|s_m, t_k)$, for example in terms of a Gaussian mixture [8]. For simplicity we here employ vector quantization [9], with which o is mapped to a code in a pre-defined codebook. If the codebook is composed of N codes $\boldsymbol{C} = \{\tilde{o}_1, \tilde{o}_2, \cdots, \tilde{o}_N\}$, each observed o is mapped to one member of $\boldsymbol{C}$, and the state-dependent observation probabilities are now characterized in terms of $p(\tilde{o}_n|s_m, T_k)$, with $\sum_{n=1}^{N} p(\tilde{o}_n|s_m, T_k) = 1$. The probabilities $p(\tilde{o}_n|s_m, T_k)$ for N codes and M_k states defines an $N \times M_k$ matrix $\boldsymbol{B}^{(k)}$.

What remains is defining the Markovian likelihood of transitioning from state s_m to state s_l, in a given target T_k. In previous studies [2]–[5] the angular sample rate $\Delta\theta$ between consecutive measurements was assumed fixed, and the state-transition matrix was constant. We now extend this concept to the case in which $\Delta\theta$ varies from one measurement to the next, allowing adaptive sensing. Specifically, the state-transition matrix for target T_k, denoted $\boldsymbol{A}^{(k)}$, is a function of $\Delta\theta$.

For target T_k, let $d_{i,j}^{(k)} \in (0, 2\pi)$ represent the shortest angular distance to travel from the center of state i (at angle $\theta_{ic}^{(k)}$) to the center of state j (at angle $\theta_{jc}^{(k)}$) in a prescribed direction, i.e. clockwise or counter-clockwise. Further, assume that $\Delta\theta \geq 0$ represents the change in the relative angular position on consecutive measurements, performed in the same angular direction as used to define $d_{i,j}^{(k)}$. Then the (i,j)th element of $\boldsymbol{A}^{(k)}(\Delta\theta)$, denoting the likelihood of transitioning from state i to state j of target T_k when moving the angular distance $\Delta\theta$, is defined as

$$a_{i,j}^{(k)}(\Delta\theta) = \frac{w_j^{(k)}(d_{i,j}^{(k)} - \Delta\theta)}{\sum_j w_j^{(k)}(d_{i,j}^{(k)} - \Delta\theta)} \tag{2}$$

where $w_j^{(k)}(\theta)$ is a function defining state j of target k. One possible choice for $w_j^{(k)}(\theta)$, and that used here, is

$$w_j^{(k)}(\theta) = \frac{1}{\sqrt{2\pi(\sigma_j^{(k)})^2}} \exp\left[-\frac{\theta^2}{2(\sigma_j^{(k)})^2}\right] \tag{3}$$

To simplify the above analysis we have assumed that the sensor always moves in a fixed direction (clockwise or counter-clockwise). However, in practice the actual direction of sensor motion is dictated by which path is shortest, e.g. it is easier to move 5° counter clockwise than clockwise 355°.

Considering (2) and (3) in greater detail, we note that the likelihood of transitioning from state i to state j, $a_{i,j}^{(k)}$, is maximized when $\Delta\theta = d_{i,j}^{(k)}$, corresponding to transitioning an angular distance commensurate with the distance between the centers of these two states. Assume now that measurement t is performed in state i, and that the next measurement at time $t+1$ is performed at an angular displacement $\Delta\theta \neq d_{i,j}^{(k)}$. As $|\Delta\theta - d_{i,j}^{(k)}|$ increases, the likelihood of transitioning from state i to state j diminishes, as defined in (2) and (3). The rate of which this likelihood diminishes is dictated by the angular extent of state j relative to $|\Delta\theta - d_{i,j}^{(k)}|$, since for simplicity it is assumed that the measurement at time t was performed in the center of state i. The size of the scalar parameter $\sigma_j^{(k)}$ is therefore dictated by the angular extent of state j in target k (see Sec. IIIA).

Assuming equally probable occurrence of the K targets, the posterior probability of target T_k after observing the sequence $\boldsymbol{o}_t = (o_1, \cdots, o_t)$ for the relative angle displacements $\Delta\boldsymbol{\theta}_t = (\Delta\theta_2, \cdots, \Delta\theta_t)$ is

$$p(T_k|\boldsymbol{o}_t, \Delta\boldsymbol{\theta}_t) = \frac{p(\boldsymbol{o}_t|\Delta\boldsymbol{\theta}_t, T_k)}{\sum_{k=1}^{K} p(\boldsymbol{o}_t|\Delta\boldsymbol{\theta}_t, T_k)} \tag{4}$$

where we have assumed that $p(T_k) = p(T_k|\Delta\boldsymbol{\theta}_c)$. Note that the probability of which state is observed on the first measurement is controlled by the initial-state probabilities in (1); the probability of which states are observed on subsequent measurements is controlled by the angular displacements $\Delta\boldsymbol{\theta}_t$. Computation of $p(\boldsymbol{o}_t|\Delta\boldsymbol{\theta}_t, T_k)$ is performed using the classical forward-backward algorithm applied widely in HMMs [8]. Specifically, the forward operator is defined as

$$\alpha_t^{(k)}(i) \equiv p(\boldsymbol{o}_t, q_t = s_i^{(k)}|\Delta\boldsymbol{\theta}_t, T_k) \tag{5}$$

where q_t denotes the state observed on observation t and $s_i^{(k)}$ denotes the ith state of target T_k. Assuming that target T_k is composed of $M^{(k)}$ states, we now have

$$p(\boldsymbol{o}_t|\Delta\boldsymbol{\theta}_t, T_k) = \sum_{i=1}^{M^{(k)}} \alpha_t^{(k)}(i) \tag{6}$$

The expression $\alpha_{t+1}^{(k)}(j)$ may be computed recursively as [8]

$$\alpha_{t+1}^{(k)}(j) = p(o_{t+1}|q_{t+1} = s_j^{(k)}, T_k) \sum_{i=1}^{M^{(k)}} \alpha_t^{(k)}(i) a_{i,j}^{(k)}(\Delta\theta_{t+1}) \tag{7}$$

B. Sequential optimization of sensor angular motion: classification

Assume a target has been detected and that it is known to be one of K candidate targets, denoted by T_k, $k = 1, \cdots, K$. In the classification stage the task is to declare the target as being one of the K candidates, based on a sequence of observations $\boldsymbol{o}_t$ measured from the target under investigation. We assume occurrence of each of the K targets is equally probable.

After performing t measurements we have the probabilities $p(T_k|\boldsymbol{o}_t, \Delta\boldsymbol{\theta}_t)$ for all targets T_k, and we now ask which angular displacement $\Delta\theta_{t+1}$ should be considered for measurement $t+1$, to optimally enhance our ability to distinguish which of the K targets is being interrogated. In this context the uncertainty in the target identity may be quantified by the entropy associated with $p(T_k|\boldsymbol{o}_t, \Delta\boldsymbol{\theta}_t)$ for all targets T_k, for all targets T_k. Specifically, we compute the entropy [10]

$$H_t(\boldsymbol{o}_t, \Delta\boldsymbol{\theta}_t) = -\sum_{k=1}^{K} p(T_k|\boldsymbol{o}_t, \Delta\boldsymbol{\theta}_t) \times \log p(T_k|\boldsymbol{o}_t, \Delta\boldsymbol{\theta}_t) \tag{8}$$

To minimize the uncertainty in target classification, the objective is to minimize the entropy after performing measurement $t+1$, and therefore the angular motion $\Delta\theta_{t+1}$ is selected such that $H_t(\boldsymbol{o}_t, \Delta\boldsymbol{\theta}_t)$ is minimized. However, $\Delta\theta_{t+1}$ clearly must be selected before observing o_{t+1}. Consequently, using the HMM, we compute the *expected* decrease in entropy upon performing measurement $\Delta\theta_{t+1}$, with the expectation performed over o_{t+1}. The expected entropy after performing measurement $t+1$ and utilizing angular displacement $\Delta\theta_{t+1}$ is expressed as

$$\begin{aligned}\tilde{H}_{t+1}(\Delta\theta_{t+1}) &= E_{\tilde{o}_{t+1}|\boldsymbol{o}_t, \Delta\boldsymbol{\theta}_t, \Delta\theta_{t+1}} H_{t+1}(\boldsymbol{o}_t, \tilde{o}_{t+1}, \Delta\boldsymbol{\theta}_t, \Delta\theta_{t+1}) \\ &= \sum_{n=1}^{N} \Bigg[H_{t+1}(\boldsymbol{o}_t, \tilde{o}_{t+1} = n, \Delta\boldsymbol{\theta}_t, \Delta\theta_{t+1}) \\ &\qquad \times \frac{\sum_{k=1}^{K} p(\boldsymbol{o}_t, \tilde{o}_{t+1} = n|\Delta\boldsymbol{\theta}_t, \Delta\theta_{t+1}, T_k)}{\sum_{k=1}^{K} p(\boldsymbol{o}_t|\Delta\boldsymbol{\theta}_t, T_k)} \Bigg]\end{aligned} \tag{9}$$

where N is the size of codebook, and we explicitly note that the observations are quantized, with each quantized observation a member of the codebook $\boldsymbol{C}$. The likelihood $p(\boldsymbol{o}_t, \tilde{o}_{t+1}|\Delta\boldsymbol{\theta}_t, \Delta\theta_{t+1}, T_k)$ is computed efficiently by exploiting the recursive properties of the HMM, as demonstrated by (5)-(7). The detailed derivation of (9) is provided in the Appendix.

Based on the above discussion, the angular displacement $\Delta\theta_{t+1}$ that maximizes the *expected* reduction in target-classification uncertainty is expressed as

$$\Delta\theta_{t+1} = \arg\max_{\Delta\theta} \Big[H_t(\boldsymbol{o}_t, \Delta\boldsymbol{\theta}_t) - \tilde{H}_{t+1}(\Delta\theta) \Big] \tag{10}$$

The operations implied by (9) and (10) are performed sequentially, to determine the angular position of all measurements after the first.

C. Sequential optimization of sensor angular motion: detection

When performing detection one usually cannot assume *a priori* knowledge of the clutter or false targets that may be encountered (the targets of interest typically constitute a finite set, while the potential false targets are of infinite number). For simplicity we assume interest in detecting a single target, represented by an associated HMM, with this readily extended to the case of multiple targets of interest. No knowledge is assumed with regard to the false targets.

Let $\boldsymbol{o}_t = (o_1, o_2, \cdots, o_t)$ represent the sequence of t observations performed during detection, with associated relative angular displacements $\Delta\boldsymbol{\theta}_t = (\Delta\theta_2, \cdots, \Delta\theta_t)$. Assuming that target T is the target of interest, the log-likelihood that $\boldsymbol{o}_t$ is associated with T is $\log p(\boldsymbol{o}_t|\Delta\boldsymbol{\theta}_t, T)$. We select the angular displacement for measurement $t+1$, $\Delta\theta_{t+1}$, such that the *expected* log-likelihood of target T is maximized. Specifically, we have

$$\Delta\theta_{t+1} = \arg\max_{\Delta\theta} E_{o_{t+1}|\boldsymbol{o}_t, \Delta\boldsymbol{\theta}_t, \Delta\theta, T} \log p(\boldsymbol{o}_t, o_{t+1}|\Delta\boldsymbol{\theta}_t, \Delta\theta, T) \tag{11}$$

which corresponds to minimizing the entropy or uncertainty in the likelihood of target T. Note that, as in the case of classification, (11) may be computed efficiently via the recursive formulae (5)-(7). In the case of a discrete HMM, for which the observations are vector quantized, (11) is

$$\begin{aligned}\Delta\theta_{t+1} = \arg\max_{\Delta\theta} \sum_{n=1}^{N} & p(\tilde{o}_{t+1} = n|\tilde{\boldsymbol{o}}_t, \Delta\boldsymbol{\theta}_t, \Delta\theta, T) \\ & \times \log p(\tilde{\boldsymbol{o}}_t, \tilde{o}_{t+1} = n|\Delta\boldsymbol{\theta}_t, \Delta\theta, T)\end{aligned} \tag{12}$$

The maximization in (11)-(12) is based on the assumption that the object under interrogation is the target of interest, T. If this assumption is true, the resulting optimal angular displacements will produce a sequence that maximizes (on average) the HMM log likelihood, with the maximization performed in a "greedy" manner, one measurement at a time. On the other hand, if the object is not T, then the selected sequence of angles will in general not maximize the associated likelihood, since in this case there is likely a mismatch between the HMM and the scattering characteristics of the object under interrogation. In this manner we implicitly distinguish between T and false targets, using no *a priori* knowledge of the latter.

III. Example Detection and Classification Results

A. HMM design

We demonstrate the sensing strategy outlined in Sec. II by considering acoustic scattering data from underwater elastic

targets. Details on the data and targets may be found in [3]. The targets are rotationally symmetric, and therefore the scattering data is collected over 360° in a plane bisecting the target axis of rotation. The data are sampled in 1° increments, in the far zone of the target (at radial distance large with respect to the target dimensions).

To design an HMM for target T_k we require the initial-state probabilities $\pi_m^{(k)}$, the state-transition matrix $\boldsymbol{A}^{(k)}$, and the state-dependent-probability matrix $\boldsymbol{B}^{(k)}$. Based on the discussion in Sec. IIA, once the state decomposition of target T_k is performed, these HMM parameters are computed directly. With regard to defining the state-transition matrix for non-uniform sampling, assume $\hat{\theta}_j^{(k)}$ represents the angular extent of state j for target T_k. The standard deviation $\sigma_j^{(k)}$ in (3) is defined here as $\hat{\theta}_j^{(k)}/2$.

The state decomposition is dependent on the target scattering physics, not on the angular sampling rate, and therefore we may apply HMM-training tools developed previously using uniform sampling [2], [3]. When training to learn that target state decomposition, we consider angular sampling at 5° between consecutive measurements, and the Baum-Welch algorithm [2], [3], [8] is applied.

B. Classification results

To demonstrate the utility of the adaptive-sensing procedure, we consider the following example. It is assumed that a target has been detected, and that it is one of the $K = 5$ known targets for which HMMs have been trained (the five targets are described in [3]). The objective is to choose $\Delta\theta_t$ to optimize classification, in the manner discussed in Sec. II. Assume that, for adaptive sensing, the average total angular extent spanned after t adaptive measurements is $\langle\theta_{total}\rangle$. For comparison purposes, we compare the results of adaptive sensing to sensing with uniform angular sampling between consecutive measurements, using an angular sampling rate $\langle\theta_{total}\rangle/(t-1)$. One would not know $\langle\theta_{total}\rangle$ without the adaptive algorithm, but this comparison allows us to address whether improved performance is accrued because of the average total angular aperture *vis-à-vis* the specific adaptively-determined angular sampling positions considered.

The identification results using the adaptively sampled sequences and uniformly sampled sequences are presented in the form of confusion matrices, in Tables 1-3, for a total of $t = 5$ observations. Table 1 shows results for $\Delta\theta = 5^\circ$ sampling, corresponding to the uniform sampling rate used while training to define the target states (as well as the sampling used in previous studies [2], [3]), and Table 2 shows results for uniform sampling corresponding to $\Delta\theta = \langle\theta_{total}\rangle/4 = 48^\circ$. Finally, Table 3 shows results for the adaptive-sensing algorithm. The confusion-matrix results are averaged across all possible initial angles of observation. As indicated by examining Tables 1 and 2, the increased angular displacement on consecutive measurements (with uniform angular sampling) significantly enhances classification performance. This is expected, since Table 2 has the potential of exploiting more fully the angular diversity in the scattering physics from the five targets.

TABLE I

CONFUSION MATRIX FOR THE FIVE TARGETS IN [3], USING 5° UNIFORM ANGULAR SAMPLING.

	T_1	T_2	T_3	T_4	T_5
T_1	**0.7667**	0.0139	0.0333	0.0528	0.0133
T_2	0.0694	**0.7278**	0.1111	0	0.0917
T_3	0.0194	0.1861	**0.7083**	0.0111	0.0750
T_4	0.0417	0	0.0222	**0.9083**	0.0278
T_5	0.0389	0.0972	0.0500	0.0389	**0.7750**

TABLE II

CONFUSION MATRIX AS IN TABLE 1, USING 48° UNIFORM ANGULAR SAMPLING.

	T_1	T_2	T_3	T_4	T_5
T_1	**0.9417**	0.0278	0.0222	0.0056	0.0028
T_2	0.0722	**0.7917**	0.1194	0	0.0167
T_3	0.0333	0.2333	**0.7167**	0	0.0167
T_4	0.0028	0	0	**0.9556**	0.0417
T_5	0.0056	0.0083	0.0111	0.0361	**0.9389**

TABLE III

CONFUSION MATRIX AS IN TABLE 1, USING ADAPTIVE ANGULAR SAMPLING (WITH AN AVERAGE ANGULAR SAMPLE RATE OF 48°).

	T_1	T_2	T_3	T_4	T_5
T_1	**0.9750**	0.0139	0.0028	0	0.0083
T_2	0.0333	**0.8806**	0.0833	0	0.0028
T_3	0.0083	0.1333	**0.8444**	0	0.0139
T_4	0.0056	0	0	**0.9639**	0.0306
T_5	0.0028	0.0028	0.0056	0.0083	**0.9806**

It is important to note that for the targets considered [2], [3] the angular extent of a given state $\hat{\theta}_m$ satisfies $5^\circ < \hat{\theta}_m < 35^\circ$. Hence, five measurements with 5° angular sampling *may* constitute transitions between two or more states. This explains the relatively good results in Table 1. However, since the total angular aperture in this case is only 20°, often only a single state is interrogated, undermining performance. The improved results in Table 2 are reflective of guaranteed state transitions, with uniform angular sampling at $\Delta\theta = 48^\circ$. We note that when uniform $\Delta\theta = 5^\circ$ angular sampling is used for a complete 360° rotation about the target, perfect classification is achieved. While this is encouraging, when sensing underwater targets such as mines, such high-density sampling is very expensive in time and energy, and therefore undesirable (often a very large number of targets must be interrogated), motivating the adaptive algorithms. In the adaptive algorithm, as discussed in the context of (2) and (3), for mathematical simplicity we have assumed that the sensor moves in one direction. In practice the adaptive direction of motion is dictated by the closest angular distance required for movement to a desired target state (with adaptive motion generally both in a clockwise and counter-clockwise direction), and therefore the actual total angular distance traveled in the adaptive measurement will typically be smaller than that considered for the uniform-sampled results in Table 2. The adaptive procedure conserves both time and energy.

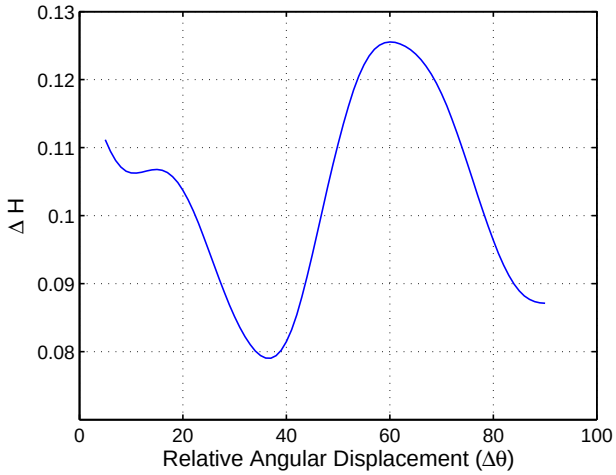


Fig. 1. An example of the objective function defined in (10).

In addition, note that for uniform sampling in 5° increments 72 measurements are required to cover 360°, for a complete rotation. For 72 adaptively determined angles of orientation we also do perfectly; in fact, we do perfectly after 10 adaptively determined observations. In these examples we chose a relatively smaller number of observations (for which uniform sampling does not achieve perfect performance) to demonstrate the relative utility of the adaptive algorithm *vis-à-vis* uniform sampling.

From Table 2 we note that Targets T_2 and T_3 are classified least effectively, with this addressed with the adaptive algorithm summarized in Table 3. Note that, on average, Tables 2 and 3 consider the same total angular aperture, so the performance improvement is attributed to the specific (non-uniform) angles observed adaptively for Table 3. The classification performance for T_2 and T_3 improves 9% and 13% respectively when sensing adaptively (from Table 2 to Table 3). The average correct classification rates across all targets are respectively 77%, 87% and 93% for Tables 1-3, again underscoring the utility of adaptive sensing.

To examine the adaptive sampling procedure in greater detail, Fig. 1 shows an example of the objective function $H_t - \tilde{H}_{t+1}(\Delta\theta)$ as defined in (10). Results are shown for 90^circ variation of the sensor position, since this represents the only non-redundant sensor positions as a consequence of target symmetry [3]. In this example we consider target T_1 as defined in [3], and this figure is representative of the form of $H_t - \tilde{H}_{t+1}(\Delta\theta)$ for all examples considered. This figure demonstrates that different selections of $\Delta\theta_{t+1}$ do indeed lead to different decreases in the *expected* entropy, and it also indicates by the smoothness of this function that the maximization in (10) is implemented easily. This has been our observation for all examples considered on this data set.

In Figs. 2(a) and 2(b) we demonstrate the probability that the sequence $\boldsymbol{o}_t = (o_1, \cdots, o_t)$ is associated with each of the five HMMs, considering an example $\boldsymbol{o}_t$ from target T_1 [3]. The initial angle of observation is the same in both cases, but in Fig. 2(a) uniform angular sampling is used, while in Fig. 2(b) the sampling is determined adaptively. The uniform sampling rate in Fig. 2(a) corresponds to $\Delta\theta = \langle\theta_{total}\rangle/4 = 48^\circ$. It is interesting to note in Figs. 2(a) and 2(b) that after a single measurement the HMMs deem target T_2 to be most likely, with the actual target (T_1) second most likely. As indicated in Fig. 2(a), with uniform sampling the HMM alternates between deeming T_1 and T_2 most likely, while after the second measurement the adaptive algorithm has strongly indicated (via the HMM probabilities) that the target under interrogation is (correctly) $T1$. After $t = 4$ observations the adaptive algorithm correctly identifies the target with almost complete certainty, i.e. the probability of $T1$ is almost unity after four observations. The phenomenon observed in Figs. 2(a) and 2(b) does not happen all the time, based on numerical studies, but it does happen often enough to yield the confusion-matrix improvements indicated in Tables 2, 3.

For the examples considered in Figs. 2(a) and 2(b), in Fig. 2(c) we plot the associated entropies of the posterior target distribution as a function of number of observations t. As demonstrated in Fig. 2(c), the adaptive sampling method selects the observation angle such that each observation yields a significant decrease in the uncertainty (entropy) of the target under interrogation, while the uniform-sampling method clearly does not address this criterion.

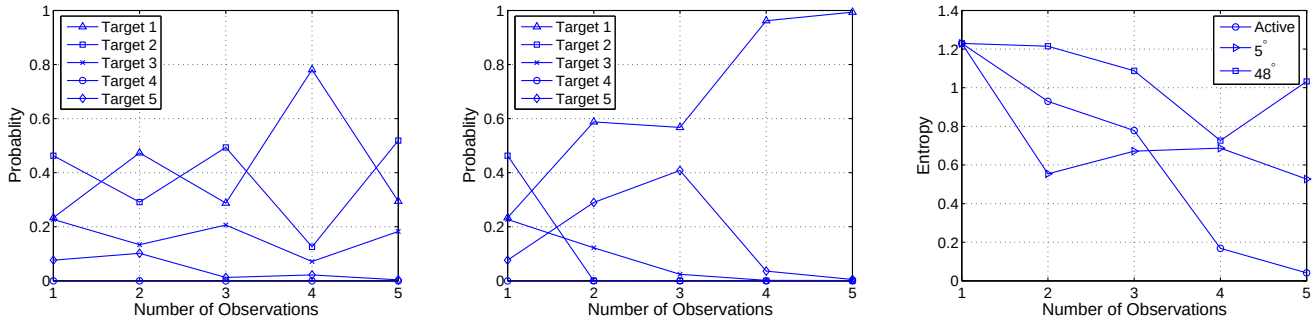
C. Detection results

In the detection example we assume that an HMM is available for target T_5 [3], this representing the "target of interest". Target T_5 is a cylindrical shell, while the six false targets are two rocks, a log, a 55-gallon drum, a plastic container and a small missile-like object (see [11] for details on the false targets, or clutter). These false targets were not considered when training the HMM for T_5. The detection results in Fig. 3 are presented for a total of $t = 5$ observations, considering all possible initial angles of observation for the targets and false targets. Results are shown for adaptive sampling as well as for uniform sampling; the uniformly-sampled results are shown for 5° and 22° increments, the latter representing the average sample rate for the adaptive algorithm. The improved detection performance of the adaptive algorithm is evident from Fig. 3.

IV. Conclusions

Hidden Markov models have been used for multi-aspect target detection and identification, with the objective of optimizing the angular displacement considered on consecutive observations. The method considered here represents an extension of previous research on optimal sequential experiments [6], [7]. Specifically, through use of the HMM and its attractive computational properties, we have removed the assumption that the measurements are statistically independent. The ideas developed here are applicable to multi-aspect sensing, as well as other applications for which HMMs are applied sequentially to process data.

The theory has been demonstrated using measured acoustic-scattering data from five underwater elastic targets [3] and six "false targets" or clutter items [11]. It has been demonstrated that the adaptive sensing procedure yields significant improvements in detection and classification performance.



(a) uniform sampling (b) active selection (c) entropy

Fig. 2. (a) Probability that the object under interrogation is targets $T_1 - T_5$, as computed by the respective HMMs. The true target identity is T_1. The angular sample rate in this example is $48°$; (b) using adaptive selection of the change in angular position; (c) Entropy defined by the probabilities $p(T_k|\boldsymbol{o}_t, \Delta\boldsymbol{\theta}_t)$, for $T_1 - T_5$, as a function of the number of observations t. Results are shown for uniform angular sampling at $5°$ and $48°$, as well as for adaptive sensing.

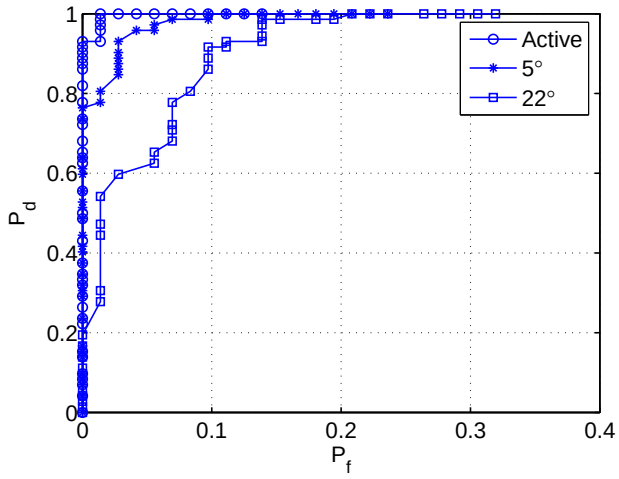


Fig. 3. Receiver operating characteristic (ROC) for distinguishing target T_5 [3] from six false targets [11]. Results are shown for adaptive selection of the change in angular position as well as for uniform angular sampling at $5°$ and $22°$. The latter represents the average angular sample rate of the adaptive algorithm.

There are many issues that deserve consideration in future studies. For example, in this study the sensor was restricted to move along an angular arc, and therefore the sensor range to the target was assumed fixed. For the examples considered here, for which there was no additive noise, this is reasonable, since there is no increase in the signature signal-to-noise ratio (SNR) as one moves closer to the target. In future research, for which clutter or additive noise is considered, the optimal sensor motion may be to move closer to the target (enhancing the SNR), rather than to move along an angular arc. The concept of expected entropy may still be used in this scenario, but the underlying HMM representation must be augmented to account for radial sensor motion.

We also note that the adaptive procedure developed here is myopic or greedy, in the sense that we select the optimal next angular sensor position without consideration of its impact on future measurements. In a non-myopic approach one would look two or more measurements ahead. The procedure developed here may be extended to the non-myopic case, by considering more than just one measurement ahead when performing the expectations in (9) and (12). For example, if two measurements ahead are considered, the expectation is performed over o_{t+1} and o_{t+2}, with entropy minimization performed over $\Delta\theta_{t+1}$ and $\Delta\theta_{t+2}$. Issues to be explored include the tradeoff between computational complexity and algorithm performance.

Finally, in this study we have placed no restrictions on the angular distance traveled on consecutive measurements. In practice one may wish to penalize long travel distances, with the goal of conserving energy. To address such an extension the expected-entropy cost functions must be augmented with a penalization term for large values of $\Delta\theta$. This is a simple algorithmic extension of significant practical importance.

APPENDIX

For the case of a discrete HMM, assume the codebook is of size N. We have

$$\tilde{H}_{t+1}(\Delta\theta_{t+1}) = E_{\tilde{o}_{t+1}|\boldsymbol{o}_t,\Delta\boldsymbol{\theta}_t,\Delta\theta_{t+1}} H_{t+1}(\boldsymbol{o}_t, \tilde{o}_{t+1}, \Delta\boldsymbol{\theta}_t, \Delta\theta_{t+1})$$
$$= \sum_{n=1}^{N} p(\tilde{o}_{t+1}=n|\tilde{\boldsymbol{o}}_t, \Delta\boldsymbol{\theta}_t, \Delta\theta_{t+1}) \times H_{t+1}(\boldsymbol{o}_t, \tilde{o}_{t+1}=n, \Delta\boldsymbol{\theta}_t, \Delta\theta_{t+1}) \quad (13)$$

where

$$\begin{aligned} p(\tilde{o}_{t+1}=n|\tilde{\boldsymbol{o}}_t, \Delta\boldsymbol{\theta}_t, \Delta\theta_{t+1}) &= \sum_{k=1}^{K} p(\tilde{o}_{t+1}=n|\tilde{\boldsymbol{o}}_t, \Delta\boldsymbol{\theta}_t, \Delta\theta_{t+1}, T_k) p(T_k|\tilde{\boldsymbol{o}}_t, \Delta\boldsymbol{\theta}_t) \\ &= \sum_{k=1}^{K} \frac{p(\tilde{\boldsymbol{o}}_t, \tilde{o}_{t+1}=n|\Delta\boldsymbol{\theta}_t, \Delta\theta_{t+1}, T_k)}{p(\tilde{\boldsymbol{o}}_t|\Delta\boldsymbol{\theta}_t, T_k)} p(T_k|\tilde{\boldsymbol{o}}_t, \Delta\boldsymbol{\theta}_t) \\ &= \frac{\sum_{k=1}^{K} p(\tilde{\boldsymbol{o}}_t, \tilde{o}_{t+1}=n|\Delta\boldsymbol{\theta}_t, \Delta\theta_{t+1}, T_k)}{\sum_{k=1}^{K} p(\tilde{\boldsymbol{o}}_t|\Delta\boldsymbol{\theta}_t, T_k)} \end{aligned} \quad (14)$$

Substituting (A4) into (A1), we obtain (9).

REFERENCES

[1] J. A. Simmons, P. A. Saillant, and S. P. Dear, "Through a bat's ear," *IEEE Spectrum*, vol. 23, no. 3, pp. 46–48, mar. 1992.

[2] P. R. Runkle, P. K. Bharadwaj, L. Couchman, and L. Carin, "Hidden Markov models for multiaspect target classification," *IEEE Trans. Signal Proc.*, vol. 47, pp. 2035–2040, Jul. 1999.

[3] P. Runkle, L. Carin, L. Couchman, T. J. Yoder, and J. A. Bucaro, "Multi-aspect target identification with wave-based matching pursuits and continuous hidden Markov models," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, pp. 1371–1378, Dec. 1999.

[4] X. Liao, P. Runkle, and L. Carin, "Identification of ground targets from sequential HRR radar signatures," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, no. 4, pp. 1230–1242, Oct. 2002.

[5] P. D. Gader, M. Mystkowski, and Z. Yunxin, "Landmine detection with ground penetrating radar using hidden Markov models," *IEEE Trans. Geoscience and Remote Sensing*, vol. 39, pp. 1231–1244, Jun. 2001.

[6] V. V. Fedorov, *Theory of optimal experiments*. Academic Press, 1972.

[7] D. MacKay, "Information-based objective functions for active data selection," *Neural Computation*, vol. 4, no. 4, pp. 590–604, 1992.

[8] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[9] R. M. Gray, "Vector quantization," *IEEE ASSP Magazine*, pp. 4–29, Apr. 1984.

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY: Wiley, 1991.

[11] N. Dasgupta, P. Runkle, L. Carin, L. Couchman, T. Yoder, J. Bucaro, and G. Dobeck, "Class-based target identification with multiaspect scattering data," *IEEE J. Oceanic Eng.*, vol. 28, pp. 271–282, Apr. 2003.