

ADAPTIVE MULTI-ASPECT TARGET CLASSIFICATION AND DETECTION WITH HIDDEN MARKOV MODELS

Shihao Ji, Xuejun Liao and Lawrence Carin

Department of Electrical and Computer Engineering
Duke University

ABSTRACT

We consider target classification and detection based on back-scattered observations measured from a sequence of target-sensor orientations. The multi-aspect scattered waves from a given target are modeled with a hidden Markov model (HMM). The targets are assumed concealed and the absolute target-sensor orientation is assumed unknown; therefore, it is only possible to control the angular displacements (change in orientation) between consecutive measurements. The performance of the HMM classifiers/detectors is influenced by the choice of the angular displacements, the optimization of which motivates adaptive search strategies developed in this paper, based on entropy-driven optimality criteria. The search proceeds in a sequential fashion. Based on the previous observations and their associated angular displacements, one determines the optimal next displacement to perform an associated observation. The search strategies are detailed and example results presented on adaptive classification and detection of underwater targets.

1. INTRODUCTION

There are many sensing scenarios for which the target is stationary and the sensor position may be moved sequentially, to observe the unknown and concealed target from a sequence of target-sensor orientations (also called aspect angles throughout the paper). For example, an unmanned underwater or airborne vehicle (UUV or UAV) may observe a target from a sequence of aspect angles. Such multi-aspect sensing is widely observed in nature, for example in bats [1]. The motivation behind multi-aspect sensing is that the scattered waves produced by a real-world target are usually a strong function of the target-sensor orientation (aspect). While this aspect-dependence complicates the classification/detection task, it may also be utilized to enhance the performance of the classifier/detector. For example, consider two targets A and B producing a set of scattered waves E_A and E_B , respectively, over the 4π range of aspect angles (i.e., including the angles of azimuth and elevation). If A and B have many features in common, the intersection $E_A \cap E_B$ may be large and any classifier attempting to differentiate A and B by using the scattered wave at a single aspect angle will produce high error rates. However, if multiple scattered waveforms are observed at a proper *sequence* of aspect angles, the error rate may be reduced as the ambiguity is resolved by the *sequential* information - A and B are less likely to generate the same *sequence* of observations.

Multi-aspect scattering from a target can often be modeled well with a hidden Markov model (HMM) [7]. Specifically, the target-sensor orientations, denoted by θ , are divided into a set of states. Each state is defined by a contiguous subset of θ for which the associated scattered waves are approximately stationary. When performing scattering measurements at a sequence of θ , one is implicitly sampling scattered waves from a sequence of target states. The sequence of sampled states is often modeled well as a Markov process [2,3,4]. In practice the underlying sampled states are unobserved, or "hidden", and only the associated scattered waves are observed. This therefore yields a *hidden* Markov model.

For the task of classification, each target is characterized by a distinct HMM. After performing a sequence of observations $\mathbf{O}_L = \{O_1, \dots, O_L\}$ at L target-sensor orientations, one computes the likelihoods $P(\mathbf{O}_L|T_k)$, $k=1, \dots, K$, and applies the maximum likelihood (ML) criterion to make a decision, i.e., assigns \mathbf{O}_L to T_i if $P(\mathbf{O}_L|T_i) > P(\mathbf{O}_L|T_k)$, $\forall T_k \neq T_i$. For detection, one must distinguish between targets T and clutter C . Due to the diversity of C , usually no HMM is assumed for C and one has a single HMM built for T . An observation sequence \mathbf{O}_L is declared as having been generated by T if $P(\mathbf{O}_L|T) > \text{threshold}$, and it is otherwise deemed generated by C .

In general, the performances of the classifiers or detectors are dependent on the choice of observation sequences. For example, if targets A and B are similar around $\theta = \theta^*$, then a sequence \mathbf{O}_L observed in the vicinity of θ^* can rarely be distinguished as coming from A or B . One should change the aspect angle and use those observations in regions of θ where A and B exhibit significant disparities.

This paper addresses the problem of optimizing a sequence of aspect angles $\boldsymbol{\theta}_L = \{\theta_1, \dots, \theta_L\}$ such that the corresponding sequence of observations $\mathbf{O}_L = \{O_1, \dots, O_L\}$ measured at $\boldsymbol{\theta}_L$ are optimal in the sense that the identity of \mathbf{O}_L is most easily determined. The targets are assumed concealed, therefore the absolute values of $\boldsymbol{\theta}_L$ are not known and one can only control the angular displacements between consecutive observations, i.e., $\Delta\boldsymbol{\theta}_L = \{\Delta\theta_2, \dots, \Delta\theta_L\}$ with $\Delta\theta_i = \theta_i - \theta_{i-1}$ if one considers only azimuthal angles. Joint optimization of all elements in $\Delta\boldsymbol{\theta}_L$ is usually difficult as the search space grows exponentially with L . Alternatively, we take a sequential approach here. We assume to have observed the sequence \mathbf{O}_L , associated with angular displacements $\Delta\boldsymbol{\theta}_L$, and ask what should be the next angular displacement $\Delta\theta_{L+1}$ to move the sensor and measure observation O_{L+1} such that the identity of \mathbf{O}_{L+1} is most readily determined.

This work can be considered an extension of the research in [5,6] to the case for which the statistical-independence assumption on the observations at different aspect angles is replaced by a hidden Markov assumption. The independence

assumption is usually favored for its ability to achieve a simple and usable form of the adaptive-search objective function. We demonstrate here that the hidden Markov assumption yields an objective function that may also be evaluated easily, constituting a search process that accounts for statistical dependence between the sequence of observations.

2. GEOMETRIC HMM WITH AN ANGLE-DEPENDENT STATE-TRANSITION MATRIX

The geometric HMM was developed in [2,3] to model multi-aspect acoustic scattering from an underwater elastic target, and it has been extended to radar scattering [4]. A geometric HMM utilizes the fact that a state corresponds to a contiguous range of aspect angles θ , allowing one to establish the HMM parameters based on geometrical considerations. Assume that the i -th state of target T_k , denoted by $S_i^{(k)}$, resides in the angular region $\theta_{i,1}^{(k)} \leq \theta \leq \theta_{i,2}^{(k)}$, with the angular extent $\hat{\theta}_i^{(k)} = \theta_{i,2}^{(k)} - \theta_{i,1}^{(k)}$. Based on simple geometrical considerations, assuming all target orientations are equally probable, the probability of T_k being in $S_i^{(k)}$ on the first observation is [2]

$$\pi_i^{(k)} = \hat{\theta}_i^{(k)} / \sum_{j=1}^{M^{(k)}} \hat{\theta}_j^{(k)} \quad (1)$$

assuming $M^{(k)}$ states for T_k . In (1) we assume the scattered waves are only a function of azimuth over 2π . This can be extended to the case where dependence on both azimuth and elevation angles is considered.

For each state $S_m^{(k)}$, $m=1, \dots, M^{(k)}$, of target T_k , we define the probability of making an observation O . Let $P(O | S_m^{(k)}, T_k)$ be the probability of observing O in $S_m^{(k)}$ of T_k . There are numerous ways to define $P(O | S_m^{(k)}, T_k)$, for example in terms of Gaussian mixtures [7]. For simplicity we here employ vector quantization, with which O is mapped to a code in a pre-defined codebook. If the codebook is composed of N codes $\mathbf{C} = \{c_1, \dots, c_N\}$, each observed O is mapped to one member of \mathbf{C} , and the state-dependent observation probabilities are quantized in terms of $P(O = c_n | S_m^{(k)}, T_k)$, with $\sum_{n=1}^N P(O = c_n | S_m^{(k)}, T_k) = 1$.

What remains is defining the Markovian probability of T_k transitioning from state $S_i^{(k)}$ to state $S_j^{(k)}$. In previous studies [2,3,4] the angular displacement $\Delta\theta$ between consecutive observations 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 observation to the next, to allow adaptive sensing. In particular, the state-transition matrix for target T_k , denoted $\mathbf{A}^{(k)}$, is a function of $\Delta\theta$. The (i,j) -th element of $\mathbf{A}^{(k)}(\Delta\theta)$, which denotes the probability of T_k transiting from state $S_i^{(k)}$ to $S_j^{(k)}$, given the angular displacement $\Delta\theta$, is defined as

$$a_{ij}^{(k)}(\Delta\theta) = \frac{w_j^{(k)}((d_{i,j}^{(k)} - \Delta\theta) \bmod 2\pi)}{\sum_{j=1}^{M^{(k)}} w_j^{(k)}((d_{i,j}^{(k)} - \Delta\theta) \bmod 2\pi)} \quad (2)$$

where mod represent modulus, $d_{i,j}^{(k)}$ is the angular distance traveling clockwise from the center of $S_i^{(k)}$ to the center of $S_j^{(k)}$, the sign of $\Delta\theta$ is defined as positive for clockwise angular displacement and negative for counterclockwise angular

displacement, and $w_j^{(k)}(\theta)$ is a bell-shaped function integrating to one. One possible choice of $w_j^{(k)}(\theta)$, and that used here, is

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

Assuming equally probable occurrence of the K targets in consideration, the posterior probability of target T_k after the sequence $\mathbf{O}_L = \{O_1, \dots, O_L\}$ is observed is found to be

$$P(T_k | \mathbf{O}_L, \Delta\theta_L) = \frac{P(\mathbf{O}_L | \Delta\theta_L, T_k)}{\sum_{k=1}^K P(\mathbf{O}_L | \Delta\theta_L, T_k)} \quad (4)$$

where $\Delta\theta_L = \{\Delta\theta_2, \dots, \Delta\theta_L\}$ are the angular displacements associated with \mathbf{O}_L with $\Delta\theta_i$ the displacement from O_{i-1} to O_i , and $P(T_k) = P(T_k | \Delta\theta_L)$ is used. Computation of $P(\mathbf{O}_L | \Delta\theta_L, T_k)$ can be effected as [7]

$$P(\mathbf{O}_L | \Delta\theta_L, T_k) = \sum_{i=1}^{M^{(k)}} \alpha_L^{(k)}(i) \quad (5)$$

where

$$\alpha_L^{(k)}(i) = P(\mathbf{O}_L, q_L = S_i^{(k)} | \Delta\theta_L, T_k) \quad (6)$$

is the forward variable and q_L the state variable for observation O_L . The forward variable can be computed efficiently in a recursive fashion as [7]

$$\alpha_{L+1}^{(k)}(j) = \left[\sum_{i=1}^{M^{(k)}} \alpha_L^{(k)}(i) a_{ij}^{(k)}(\Delta\theta_{L+1}) \right] P(O_{L+1} | q_{L+1} = S_j^{(k)}, T_k) \quad (7)$$

3. ADAPTIVE SEARCH OF OPTIMAL SENSOR ANGULAR DISPLACEMENTS

3.1. Adaptive Search in the Case of Classification

Assume that the K possible targets under consideration occur with equal probability. After a sequence of observations \mathbf{O}_{L+1} is made with associated angular displacements $\Delta\theta_{L+1}$, one has the posterior probabilities $P(T_k | \mathbf{O}_{L+1}, \Delta\theta_{L+1})$, $k=1, \dots, K$. Consider $P(T_k | \mathbf{O}_{L+1}, \Delta\theta_{L+1})$ as a distribution in k , from which one can use the maximum posterior probability (MAP) criterion to determine the identity of the interrogated target. To minimize uncertainty, one minimizes the entropy of $P(T_k | \mathbf{O}_{L+1}, \Delta\theta_{L+1})$ in k , i.e.,

$$H(T_k | \Delta\theta_{L+1}, \mathbf{O}_{L+1}) = - \sum_{k=1}^K P(T_k | \Delta\theta_{L+1}, \mathbf{O}_{L+1}) \log P(T_k | \Delta\theta_{L+1}, \mathbf{O}_{L+1}) \quad (8)$$

is minimized [8]. One then searches for the $\Delta\theta_{L+1}$ that minimizes (8). Joint search of the L unknowns in $\Delta\theta_{L+1}$ is difficult, as the search space grows exponentially with L . Alternatively, one can use a sequential search strategy, which searches the next angular displacement $\Delta\theta_{L+1}$ based on the previously determined displacements $\Delta\theta_L = \{\Delta\theta_2, \dots, \Delta\theta_L\}$. Clearly, the choice of $\Delta\theta_{L+1}$ must be made before O_{L+1} is actually observed. One can remove the dependence of (8) on O_{L+1} in the search of $\Delta\theta_{L+1}$, by taking conditional expectation of (8) with respect to O_{L+1} given \mathbf{O}_L , $\Delta\theta_L$, and $\Delta\theta_{L+1} = \Delta\theta$, and minimizing the expected entropy. In the case of quantized observations, the expected entropy is

$$\begin{aligned}
g_{L+1}(\Delta\theta) &\stackrel{Def.}{=} E_{O_{L+1}|\mathbf{O}_L, \Delta\theta_L, \Delta\theta} H(T_k | \Delta\theta_L, \Delta\theta, \mathbf{O}_L, O_{L+1}) \\
&= \sum_{n=1}^N [H(T_k | \Delta\theta_L, \Delta\theta, \mathbf{O}_L, O_{L+1} = c_n) \\
&\quad \times \frac{\sum_{k=1}^K P(\mathbf{O}_L, O_{L+1} = c_n | \Delta\theta_L, \Delta\theta, T_k)}{\sum_{k=1}^K P(\mathbf{O}_L | \Delta\theta_L, T_k)}]
\end{aligned} \tag{9}$$

which can be efficiently evaluated in a recursive manner using (4)-(7). The detailed derivation of (9) is omitted here for brevity. The optimal angular displacement $\Delta\theta_{L+1}$ maximizes the reduction in the expected entropy

$$\Delta\theta_{L+1} = \arg \max_{\Delta\theta} [H(T_k | \Delta\theta_L, \mathbf{O}_L) - g_{L+1}(\Delta\theta)] \tag{10}$$

Note the first term in the objective function of (10) is independent of $\Delta\theta$ and O_{L+1} and can be treated as a constant in the maximization. The operations in (9) and (10) are performed sequentially, to determine the angular positions of all observations except the first.

3.2. Adaptive Search in the Case of Detection

When performing detection one usually cannot assume *a priori* knowledge of the possible clutter, as their number is usually infinite, unlike the targets, which are of a finite number. Therefore for detection a single HMM is built representing the targets T and no HMM is built for the clutter set C .

Assume a sequence of observations \mathbf{O}_L has been made with associated angular displacements $\Delta\theta_L$. The next angular displacement $\Delta\theta_{L+1}$ for observation O_{L+1} is determined by maximizing the *expected* logarithmic likelihood of \mathbf{O}_{L+1} having been produced by T , i.e.,

$$\Delta\theta_{L+1} = \arg \max_{\Delta\theta} E_{O_{L+1}|\mathbf{O}_L, \Delta\theta_L, \Delta\theta, T} [\log P(\mathbf{O}_L, O_{L+1} | \Delta\theta_L, \Delta\theta, T)] \tag{11}$$

where the expectation is taken with respect to O_{L+1} conditional on \mathbf{O}_L , $\Delta\theta_L$, $\Delta\theta_{L+1} = \Delta\theta$, and T . By using $\log P(\mathbf{O}_L, O_{L+1} | \Delta\theta_L, \Delta\theta, T) = \log P(O_{L+1} | \mathbf{O}_L, \Delta\theta_L, \Delta\theta, T) + \log P(\mathbf{O}_L | \Delta\theta_L, \Delta\theta, T)$ and the fact that \mathbf{O}_L is independent of O_{L+1} , one finds (11) reduces to

$$\begin{aligned}
&\Delta\theta_{L+1} \\
&= \arg \max_{\Delta\theta} E_{O_{L+1}|\mathbf{O}_L, \Delta\theta_L, \Delta\theta, T} [\log P(O_{L+1} | \mathbf{O}_L, \Delta\theta_L, \Delta\theta, T)] \tag{12} \\
&= \arg \max_{\Delta\theta} [-H(O_{L+1} | \mathbf{O}_L, \Delta\theta_L, \Delta\theta, T)]
\end{aligned}$$

where H is the entropy. This shows $\Delta\theta_{L+1}$ is equivalently a minimizer of the entropy, similar to the case in Sec. 3.1. For quantized observations (11) can be expressed as

$$\Delta\theta_{L+1} = \arg \max_{\Delta\theta} \sum_{n=1}^N [P(\mathbf{O}_L, O_{L+1} = c_n | \Delta\theta_L, \Delta\theta, T) \times \log P(\mathbf{O}_L, O_{L+1} = c_n | \Delta\theta_L, \Delta\theta, T)] \tag{13}$$

The objective function in (13) can be recursively computed via (5)-(7).

The maximization in (11)-(13) is based on the assumption that the object being interrogated is T . If this assumption is true, the resulting optimal angular displacements will produce a sequence that maximizes, on average, the HMM logarithmic likelihood, with the maximization performed in a “greedy” manner, one observation at a time. 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 the targets and clutter, using no *a priori* knowledge of the latter.

4. RESULTS

We consider acoustic scattering data from five underwater elastic targets T_1, T_2, \dots, T_5 [2,3]. For classification, we first build a distinct HMM for each of the five targets, assuming a uniform angular displacement of 5° between consecutive observations. During this training phase the target state decomposition is performed using the Baum-Welch algorithm [7]. Then the state-transition-probability matrix is augmented to handle non-uniform angular displacements, using (2)-(3) with the variances in (3) determined from the state decomposition.

Specifically, if $\hat{\theta}_j^{(k)}$ represents the angular extent of state $S_j^{(k)}$ of target T_k , then $\sigma_j^{(k)}$ in (3) is defined as $\hat{\theta}_j^{(k)}/2$.

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. We compare three methods. In Method 1 we adaptively determine the angular displacements $\Delta\theta_L$ for $L=5$ via the search strategies in Sec. 3. In Method 2 the angular displacements are constant and span the same angular extent (aperture) as the adaptive displacements for the same sequence length. In Method 3 the displacements are constant and equal to 5° .

The confusion matrices of classifying the 5 targets using Methods 1, 2, and 3 are presented in Tables 1-3, respectively. These results are averaged across all possible initial angles for the length-five ($L=5$) sequences. The (i,k) -th element of the confusion matrix quantifies the probability that a sequence of T_i is declared as coming from T_k . Tables 1-3 demonstrate that Method 1, the adaptive search method, consistently outperforms the other two.

Fig. 1 shows an example of the objective function as defined in (10). In this example we consider target T_1 , and this figure shows a representative shape of the objective function for all examples considered. This figure demonstrates that different selections of $\Delta\theta_{L+1}$ 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.

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 wood log, a 55-gallon drum, a plastic container and a small missile-like object (see [9] 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. 2 are presented for a total of $L=5$ observations, considering all possible initial angles of observation for the targets and false targets. The results of Methods 1, 2, and 3 are presented in the form of the receiver operating characteristic (ROC) curve. It is seen that Method 1, the adaptive method, achieves significantly improved results over the other two methods, which both use sequences of uniform angular displacements. The uniformly-sampled results are shown for 5° and 22° increments, the latter representing the average sample rate for the adaptive algorithm.

5. CONCLUSIONS

Hidden Markov models have been used for multi-aspect target identification and detection, with the objective of optimizing the angular displacement between consecutive observations. The method considered here represents an extension of the work on optimal sequential experiments [5-6]. Specifically, through use of the HMM we have removed the assumption that the sequence of 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 effectiveness of the presented methods has been demonstrated using measured acoustic-scattering data from five underwater elastic targets. The results showed that by using adaptive search procedure the performance of the HMM classifiers and detectors are significantly improved.

In the study presented here only the angular motion was considered for the sensor. In future research, additive noise may be considered and the sensor's radial motion may also be optimized to enhance the signal to noise ratio. In addition other measures of entropy may be considered, such as Renyi entropy [8], rather than the Shannon entropy used here.

6. REFERENCES

- [1] Simmons, J. A., Saillant P. A., and Dear, S. P., "Through a Bat's ear", *IEEE Spectrum*, March 1992.
- [2] Runkle, P., Bharadwaj, P., and Carin, L., "Hidden Markov model multi-aspect target classification", *IEEE Trans. Signal Proc.*, vol. 47, pp. 2035-2040, July 1999.
- [3] Runkle, P., Carin, L., Couchman, L., Yoder, T. J., and Bucaro, J. A., "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] Liao, X., Runkle, P., and Carin, L., "Identification of ground targets from sequential HRR radar signatures", *IEEE Transactions on Aerospace and Electronic System*, Vol. 38, No. 4, Oct 2002, pp. 1230-1242.
- [5] Fedorov, V. V., *Theory of Optimal Experiments*, Academic Press, 1972.
- [6] MacKay, D. J. C., "Information-based objective functions for active data selection", *Neural Computation*, 4, 1992, pp. 590-604.
- [7] Rabiner, L. R., "A tutorial on hidden Markov models and selected applications in speech recognition", *Proc. IEEE*, Vol. 77, No. 2, pp. 257-286, 1989.
- [8] Cover, T. M. and Thomas, J. A., *Elements of Information Theory*, John Wiley & Sons, New York, 1991.
- [9] Dasgupta, N., Runkle, P., Carin, L., Couchman, L., Yoder, T., Bucaro, J., and Dobeck, G. J., "Class-based target identification with multispect scattering data", *IEEE J. Oceanic Eng.*, vol. 28, pp. 271-282, April 2003.

Table 1. Confusion matrix of Method 1

	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

Table 2. Confusion matrix of Method 2

	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 3. Confusion matrix of Method 3

	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

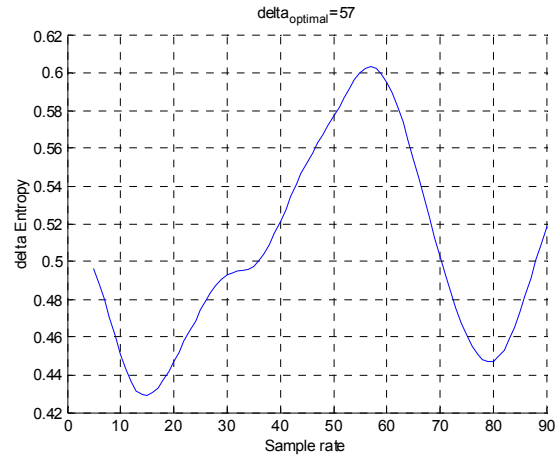


Figure 1. An example of the objective function defined in equation (10).

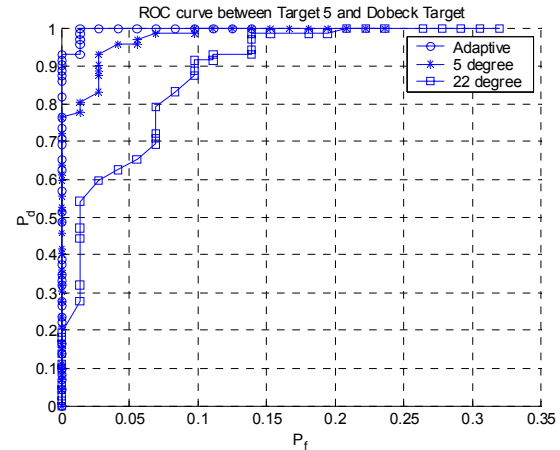


Figure 2. Receiver operating characteristic (ROC) for distinguishing target T₅ [3] from six false targets [9]. Results are shown for adaptive search of angular displacements as well as for uniform angular displacements of 5° and 22°. The latter represents the average angular sampling rate of the adaptive algorithm.