

A POMDP for Multi-view Target Classification with an Autonomous Underwater Vehicle

Vincent Myers
Defence R&D Canada
Halifax, Canada

Email: vincent.myers@drdc-rddc.gc.ca

David P. Williams
NATO Undersea Research Centre
La Spezia, Italy

Email: williams@nurc.nato.int

Abstract—A partially observable Markov decision process (POMDP) is proposed to perform multi-view classification of underwater objects. The model allows one to adaptively determine which additional views of an object would be most beneficial for reducing classification uncertainty. Acquiring additional views is made possible by employing a sonar-equipped autonomous underwater vehicle (AUV) for data collection. The POMDP model is validated using real synthetic aperture sonar (SAS) data collected at sea, with promising results. The approach provides an elegant way to fully exploit multi-view information in a methodical manner.

Index Terms—POMDP, Automatic Target Recognition, Synthetic Aperture Sonar

I. INTRODUCTION

The automatic detection and classification of underwater objects on the seabed using high-frequency imaging sonars is a challenging task owing largely to the difficulty of inferring the three-dimensional shape of an object from a two-dimensional projection [1]. Traditionally, data would be collected for this goal via the use of an unmanned system, such as an autonomous underwater vehicle (AUV), that follows pre-planned survey routes designed by a human operator. Increasingly, focus is shifting away from this purely *deliberative* model for AUV control to instead incorporate *reactive* capabilities. In this new paradigm, an AUV is permitted to adapt its route so that the sensor data it collects can be exploited in the remainder of its mission. This adaptivity allows more valuable data to be collected, which should in turn improve classification performance.

Most AUVs used for target classification tasks are equipped with a side-looking sonar. This sensor transmits high-frequency acoustic signals in a direction orthogonal to the vehicle's travel direction in order to create a strip-map acoustic image of the seafloor. This high-resolution image data is then exploited to classify objects on the seafloor.

Some recent work [2], [3] has examined the use of multiple views of a given object to improve underwater classification performance beyond that achievable with a single view. However, these studies assumed that there was no freedom in choosing the views to be collected. The objective of this work is to develop a technique for performing multi-view target classification in which the views of an object that the AUV obtains can be determined adaptively. By incorporating this

knowledge into the data collection process, the most valuable data for classification purposes can be collected.

This goal is achieved by modeling the multi-view underwater mine classification problem as a partially observable Markov decision process (POMDP) [4]. The solution to a POMDP, called a *policy*, provides a principled way for a robot (*e.g.*, AUV) to react to collected sensor data. For the underwater mine classification problem, the POMDP policy will dictate the manner in which the AUV route is adapted and determine the number and particular aspects of views needed to confidently classify an object.

POMDPs have been applied to several problems in many fields, including robot navigation and target identification (see *e.g.*, [5]). The idea of performing aspect-dependent target classification by a robot has been considered for land mine identification in [6].

The remainder of this paper is organized as follows. The relevant theory of POMDPs is reviewed in Section 2. The proposed POMDP model tailored to the underwater mine classification problem with an AUV is described in Section 3. Experimental results of the proposed approach on real synthetic aperture sonar (SAS) data collected at sea are shown in Section 4. Concluding remarks are made in Section 5.

II. POMDP BACKGROUND

A POMDP [4] is a model that governs the actions of an agent (*e.g.*, robot) in a specified manner. A POMDP is composed of the following parts: a finite set of states \mathcal{S} ; a finite set of actions \mathcal{A} , and a transition function $T : \mathcal{S} \times \mathcal{A} \mapsto \Pi(\mathcal{S})$ that maps states and actions to probabilities over states, such that $T(s, a, s')$ is defined as the probability of transitioning from state s to state s' as a result of taking action a ; a set of observations Ω and an observation function $O : \mathcal{S} \times \mathcal{A} \mapsto \Pi(\Omega)$ that maps states and actions to probabilities over observations, such that $O(s', a, o)$ specifies the probability of obtaining observation o when action a was taken that resulted in state s' ; and a reward function $R : \mathcal{S} \times \mathcal{A} \mapsto \mathbb{R}$, where $R(s, a)$ is the reward for taking action a when in state s .

Since the states s are not directly observable, a belief state b is introduced as a probability distribution over the states \mathcal{S} , where $b(s)$ specifies the probability that the agent is in state s . Belief is propagated using a *state estimator* that employs Bayes' rule to update the belief when starting from state s ,

performing action a , and obtaining observation o :

$$b^{ao} = \frac{\Pr(s'|a, o, b) O(s', a, o) \sum_{s \in \mathcal{S}} T(s, a, s') b(s)}{\Pr(o|a, b)}, \quad (1)$$

where

$$\Pr(o|a, b) = \sum_{s' \in \mathcal{S}} O(s', a, o) \sum_{s \in \mathcal{S}} T(s, a, s') b(s), \quad (2)$$

is a normalization constant ensuring that $\sum_{s \in \mathcal{S}} b(s) = 1$.

The solution to a POMDP is called a *policy*, π , which determines the action to take for any belief state b . The policy maximizes the expected reward over t steps:

$$E \left[\sum_{n=0}^{t-1} \gamma^n R(s_n, a_n) \right], \quad (3)$$

where $0 < \gamma < 1$ is a discount factor that regulates the tradeoff between future and immediate rewards, and t defines the horizon of the POMDP. The POMDP policy is learned via the value function, which is given by [7]

$$V_n^*(b) = \max_{a \in \mathcal{A}} \left[\rho(b, a) + \gamma \sum_{o \in \Omega} \Pr(o|a, b) V_{n-1}^*(b^{ao}) \right], \quad (4)$$

where $\rho(b, a) = \sum_{s \in \mathcal{S}} b(s) R(s, a)$ is the reward function conditioned on the state estimator, and $\Pr(o|a, b)$ is from (2). The optimal n -horizon policy, π_n^* is the one which is defined by V_n^* . Equation (4) is formulated for the finite horizon case. In many instances, however, it is not desirable or possible to specify the horizon n in advance. In the case of $n \rightarrow \infty$ one obtains an infinite horizon POMDP. It can be shown that

$$\lim_{n \rightarrow \infty} \|V_n^* - V^*\| = 0 \quad (5)$$

holds and it is possible to approximate the infinite horizon value function arbitrarily close by using a sufficiently long horizon. The policy defined by V^* is denoted π^* .

Computing the value function and policy for a POMDP is a challenging task because the belief space is an infinite continuous space over \mathcal{S} . The method used to solve the POMDP in this study is called Focused Real-Time Dynamic Programming (FRTDP) [8], specific details of which are given in [9].

The iterative decision cycle for a POMDP is summarized as follows: With the current belief state b , perform the action a given by the policy $\pi(b)$. Next, obtain an observation o . Then, update the belief state using the state estimator in (1).

III. MULTI-VIEW TARGET CLASSIFICATION WITH A POMDP

In this section, we define the various POMDP components that were defined above to in terms of the problem of performing multi-view target classification of underwater objects when data in the form of sonar imagery is collected by an AUV.

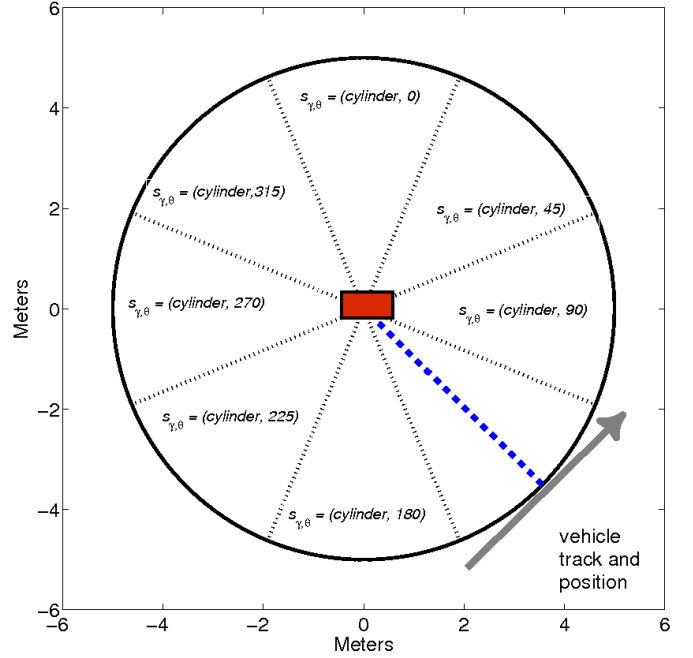


Fig. 1. A sketch of the problem setup with the cylinder target. The vehicle track is shown by the gray arrow. The target aspects are discretized into distinct sectors, each one as a separate state and labeled according to the sector's center aspect, in degrees.

The states of the POMDP correspond to target-aspect pairs of different object types. Let Γ be the set of possible object types, and let Θ be the set of possible aspects. The set of states \mathcal{S} contains every possible target-aspect combination, resulting in the cardinality $|\mathcal{S}| = |\Gamma| \times |\Theta| + 1$. Define $s = (k, \ell)$ to be the state associated with the k th object γ_k , $k = [1, \dots, |\Gamma|]$ and ℓ th aspect θ_ℓ , $\ell = [1, \dots, |\Theta|]$. The additional state is a zero-reward, fully absorbing end state denoted \bar{s} , which is explained below.

In this work, we assume $|\Gamma| = 4$ object types: a cylinder, a truncated cone, a wedge-shaped object, and an irregularly-shaped non-target object. In addition, views are discretized into $|\Theta| = 8$ aspects. The geometry of the side-looking sonar setup and an example of the aspect divisions are shown in Figure 1.

The set of actions \mathcal{A} is composed of several actions which fall into two broad categories. The first type of action — movement — directs the AUV to move and obtain another view of the object from a specific aspect. The second type of action — declaration — instead classifies the object being interrogated. Therefore, additional data is collected only when performing movement actions. The possible view aspects are divided into $|\Theta|$ sectors, resulting in $|\Theta|-1$ movements actions, and $|\Gamma|$ declaration actions, one for each possible object type. This formulation is derived from the *tiger* problem presented in [10].

The reward function, R , of the POMDP is defined to treat misclassifications equally, rather than heavily penalizing certain types (such as mistakenly declaring targets as non-targets). The reward for a correct classification is $+1$, the

reward for an incorrect classification is -1 , and the reward for obtaining an additional view is 0.

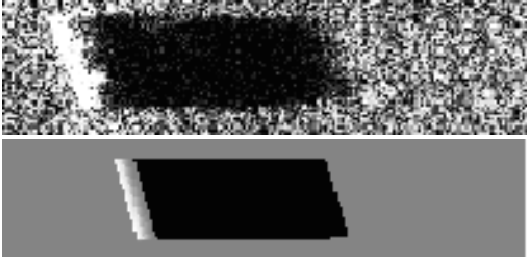


Fig. 2. A SAS image of a large cylinder (top) and its corresponding ideal template (bottom).

Because modern navigation systems on AUVs are generally reliable and accurate, the transition function, T , of the POMDP is trivial. Specifically, an action to obtain a view at a desired aspect will result in that aspect being obtained with probability one.

The use of the end state \bar{s} is a commonly used technique in Markov Decision Processes. In this case, following a declaration action, the system will move into state \bar{s} with probability one, and any action will cause the system to remain in this state with probability one. The reward for being in this state is zero. This simply indicates to the agent that once an object has been classified, it is no longer possible to perform any action that would result in additional rewards.

The observation o that is obtained from performing a movement (and data collection) action is defined to be the *index-pair* associated with the target-aspect pair that achieves the maximum correlation between (i) the newly acquired view and (ii) a template corresponding to that target-aspect pair. This observation is described in more detail below.

Let $\mathcal{S}^\dagger \subset \mathcal{S}$ be the subset of \mathcal{S} which belong to objects for which it is possible to generate a three dimensional representation which can be used as input to a computer model which will simulate the sonar imaging process. This set will include known target types whose size and shape information can be obtained by various means; it will generally *not* include clutter objects, such as rocks. For all the target-aspect pairs in \mathcal{S}^\dagger , an image template representing the ideal signature for the given target at the specified aspect is constructed using a basic ray-tracing model. An example template for a cylindrical target is shown in Figure 2.

Let $M^{(i,j)}$ denote the modeled template for target type $\gamma_i \in \Gamma$ at aspect $\theta_j \in \Theta$. Let I denote the newly acquired sonar image (*i.e.*, view) under consideration. Define $g(M, I)$ as a function which computes the similarity between the image I and the template M . This function is described in detail in [11]. The observation o is defined to be the *index-pair* associated with the target-aspect pair that both achieves the

maximum similarity and is also above a set threshold τ :

$$o = \begin{cases} \arg \max_{\substack{i = [1, \dots, |\Gamma|] \\ j = [1, \dots, |\Theta|]}} g(M^{(i,j)}, I) & \text{if } g(M^{(i,j)}, I) > \tau \\ (0, 0) & \text{otherwise.} \end{cases} \quad (6)$$

The observation $o = (i, j)$ indicates that the maximum similarity corresponded to the i -th target type, $\gamma_i \in \Gamma$, at the j -th aspect, $\theta_j \in \Theta$. The observation $o = (0, 0)$ is assigned to the case for which none of the $|\mathcal{S}^\dagger|$ target-aspect pairs achieve a maximum similarity greater than τ .

Next, this observation feature is used to specify observation probabilities $P(o = (i, j) | s = (k, \ell))$. We first begin by fixing the probability that, given a non-clutter object, the image matches *none* of the templates to a degree which exceeds τ . In this case:

$$P(o = (0, 0) | s = (k, \ell), (\gamma_k, \theta_\ell) \in \mathcal{S}^\dagger) = q. \quad (7)$$

The remaining balance of probability $p = 1 - q$ is apportioned in the following way. In order to calculate $P(o = (i, j) | s = (k, \ell))$, we begin by estimating the potential for the similarity function g to return a match. Define α as the ratio between the number of non-zero pixels in the template $M^{(i,j)}$ and the template $M^{(k,\ell)}$. The value of $\xi_{(k,\ell)}^{(i,j)} = |\alpha - 1|$ is presumed to be inversely proportional to the expected output of $g(M^{(i,j)}, I^{(k,\ell)})$. A perfect match ($\xi_{(k,\ell)}^{(i,j)} = 0$) is expected when $i = k$ and $j = \ell$. The larger the value of $\xi_{(k,\ell)}^{(i,j)}$, the less likely it is that g will produce a peak for the template $M^{(i,j)}$ when presented with an image $I^{(k,\ell)}$. This effect is modeled using the exponential density function, where

$$f(\xi) = \lambda e^{-\lambda \xi}, \quad (8)$$

with parameter λ . The final probability for all states in \mathcal{S}^\dagger is:

$$P(o = (i, j) | s = (k, \ell)) = \frac{f\left(\xi_{(k,\ell)}^{(i,j)}\right)}{\sum_{i,j} f\left(\xi_{(k,\ell)}^{(i,j)}\right)} \times p. \quad (9)$$

Therefore, the probability p that given the sonar image of a target $I^{(k,\ell)}$, at least *one* of the templates will exceed the threshold τ is assigned according to (9), and is estimated using the templates themselves.

Finally, for the objects in the set $\{\mathcal{S} - \mathcal{S}^\dagger\}$, corresponding to the non-target class representing all possible types of unmodeled clutter, a uniform distribution is placed over the number of possible observations. With this choice, the problem of explicitly specifying the non-target class is neatly circumvented.

IV. EXPERIMENTAL RESULTS

The proposed POMDP model for underwater target classification was evaluated using SAS data collected by the MUSCLE AUV [12] during trials in the Baltic Sea. The data consisted of multi-view imagery (at several aspects) of a cylinder, a truncated cone, a wedge-shaped object, and two (non-target) rocks. The adaptive sensing nature of the POMDP

TABLE I
CLASSIFICATION CONFUSION MATRICES FOR SINGLE VIEW, TWO ORTHOGONAL VIEWS, AND THE PROPOSED POMDP. (CLASSES: CYLINDER (C), WEDGE (W), TRUNCATED CONE (TC), AND NON-TARGET (N))

	1 VIEW				2 VIEWS				POMDP			
	W	TC	C	N	W	TC	C	N	W	TC	C	N
W	0.90	0.10	0	0	0.89	0	0	0.11	0.95	0	0	0.05
TC	0	0.89	0	0.11	0	1.0	0	0	0	0.96	0	0.04
C	0	0	1.0	0	0	0	1.0	0	0	0	1.0	0
N	0	0.69	0	0.31	0	0.31	0	0.69	0	0.06	0	0.94

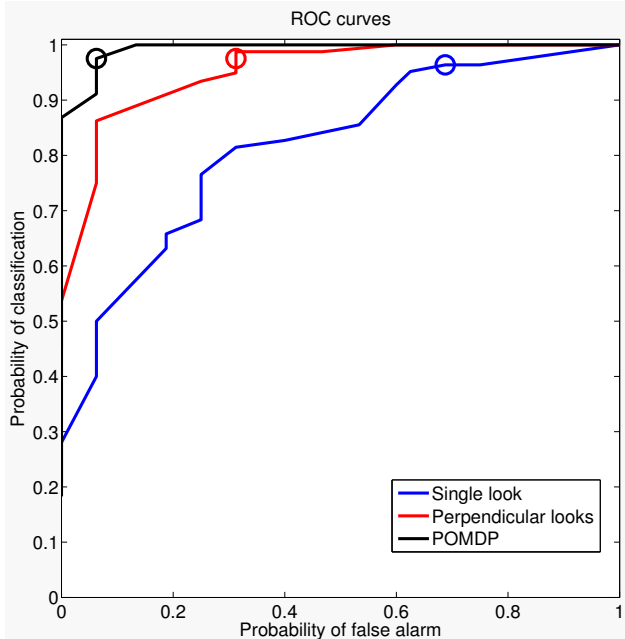


Fig. 3. Receiver-Operating Characteristic curves for the three methods considered. The points on the curves that correspond to the most accurate confusion matrices for each method are also shown; these matrices are displayed in Table 1.

was simulated for these experiments. The performance of the proposed approach is compared to the performance when a single view is used to classify the target, and when two orthogonal views are used. For these two views, the target which corresponds to the best matching template (that is, the highest value of g) is chosen as the class of the object. If neither of these views exceeds τ , then the image is classified as a non-target.

Receiver-Operating Characteristic (ROC) curves, shown in Fig. 3 were calculated for all three methods by varying the threshold τ in (6) over the interval $[0.5, 1.2]$. Each point in the curve is computed from a randomly drawn bootstrap sample of size 100. The other parameter values used for the POMDP were $\lambda = 10$ in (8) and $p = 0.9$ in (9). The discount factor γ in the value function (4) was 0.75. The FRTDP program was instructed to continue to try and improve the policy until the regret bound, defined as the difference between the expected reward of the optimal policy and the current policy [13], of less than 0.02.

In order to create a ROC curve which plots the probability of detection as a function of false alarm, all of the targets are

grouped together as a single class for the purposes of accounting; therefore, a cone classified as a cylinder is still a correct classification of a target as a target. A more detailed analysis of the performance can be performed via the use of confusion matrices, some of which are shown in Table 1. Each operating point on the curves shown in Fig. 3 represents a specific threshold τ for which one can compute additional performance measures, such as confusion matrices. The points which were chosen correspond to the value of τ which achieved the greatest classification accuracy; these are shown for all three methods in Fig. 3. For the methods of the POMDP method and perpendicular views this meant a threshold $\tau = 1.03$ and for the single view a threshold of $\tau = 0.98$. The performance in terms of both target classification and clutter rejection is superior for the POMDP method. The average number of views selected by the POMDP before choosing a declaration action (*i.e.*, classifying an object) was 2.65, broken down as follows: 1.89 views for cylinders, 3.11 views for wedges, 2.85 views for truncated cones, and 3.5 views for rocks. This breakdown can be attributed to the relative complexity of each object shape.

Since the aspects that will be obtained from an object are not known in advance, it is possible to show the path that the vehicle would (adaptively) determine when classifying a target. For these examples, a simple greedy algorithm calculates the route [14] and waypoints required to obtain an image of a target at a given aspect. For the examples shown in Fig 4, a turn radius of 10 meters was used, with a lead-in length, often required with imaging techniques which require vehicle stability, of 100 meters and a lead-out length of 20 meters.

V. DISCUSSION AND FUTURE WORK

A POMDP model for performing multi-view classification of underwater objects with a side-scan equipped AUV has been presented. The promise of the proposed model was demonstrated on a set of real SAS data collected at sea. The approach allows one to adaptively determine which additional views of an object would be most beneficial for reducing classification uncertainty.

One point of discussion is that of the Markov property of the model, which maintains that the future states are only dependent on the present state. This creates a situation where an object can be viewed from the same view aspect twice. In this case, one would expect (in general) that no additional information would be gained from reacquiring the same view multiple times. It may be worth investigating ways to avoid

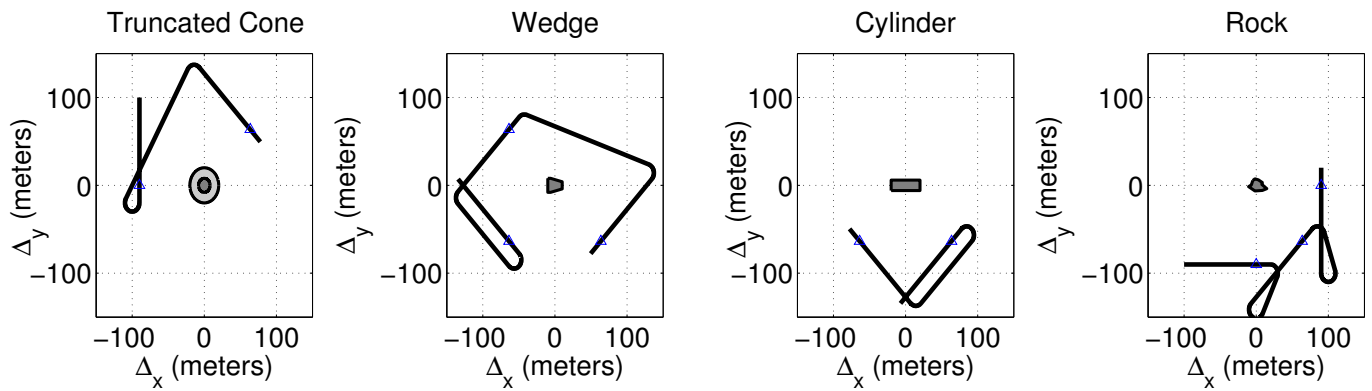


Fig. 4. Example AUV routes selected by the POMDP in order to classify different object types.

this situation by modifying the state space, for instance by introducing a dimension that is based on which views have already been obtained to reflect this condition. However, it would result in an even larger state space, creating a significantly more difficult problem to solve than the one presented here.

Future work will focus on developing a more sophisticated observation model in the POMDP, partially with the goal of eliminating the parameters p and λ which are currently set in an admittedly *ad hoc* manner. In addition, the method of using the ratio of non-zero pixels in the templates (α) to estimate a degree of similarity between target types could also be improved. While the number of pixels can indicate that objects are of similar sizes, a metric which is derived from the template matching method g itself may be of greater use, since shape and intensity information will be captured as well.

REFERENCES

- [1] S. Reed, Y. Petillot, and J. Bell, "Model-based approach to the detection and classification of mines in sidescan sonar," *Applied Optics*, vol. 43, no. 2, pp. 237–246, 2004.
- [2] B. Zerr, J. Fawcett, and D. Hopkin, "Adaptive algorithm for sea mine classification," in *Underwater Acoustic Measurements (UAM) Conference Proceedings*, 2009.
- [3] D. Williams and J. Groen, "Multi-view target classification in synthetic aperture sonar imagery," in *Underwater Acoustic Measurements (UAM) Conference Proceedings*, 2009, pp. 699–704.
- [4] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," *Artificial Intelligence*, 1998.
- [5] A. Cassandra, "A survey of POMDP applications," *AAAI fall symposium*, 1998.
- [6] L. He, S. Ji, and L. Carin, "Application of partially observable Markov decision processes to robot navigation in a minefield," in *ICAPS*, 2006.
- [7] A. Cassandra, M. Littman, and L. Kaelbling, "A simple, fast, exact methods for partially observable Markov decision processes," in *Proceedings of Uncertainty in Artificial Intelligence*, 1997.
- [8] T. Smith, "ZMDP software for POMDP and MDP planning," <http://www.cs.cmu.edu/~trey/zmdp/>, Accessed May 2009.
- [9] —, "Probabilistic planning for robotic exploration," Ph.D. dissertation, Carnegie Mellon Institute, 2007.
- [10] A. Cassandra, L. Kaelbling, and M. Littman, "Acting optimally in partially observable stochastic domains," in *AAAI Conference Proceedings*, 1994, pp. 1023–1028.
- [11] V. Myers and J. A. Fawcett, "A template matching procedure for automatic target recognition in synthetic aperture sonar imagery," *IEEE Signal Processing Letters*, 2010.
- [12] A. Bellettini and M. Pinto, "Design and experimental results of a 300 kHz synthetic aperture sonar optimized for shallow-water operations," *IEEE Journal of Oceanic Engineering*, vol. 34, pp. 285–293, 2008.
- [13] T. Smith and R. Simmons, "Heuristic search value iteration for POMDPs," in *Proc. Int. Conf. on Uncertainty in Artificial Intelligence (UAI)*, 2004.
- [14] L. Dubins, "On curves of minimal length with a constraint on average curvature, and with prescribed initial and terminal positions and tangents," *American Journal of Mathematics*, vol. 79, pp. 497–516, 1957.