

Cooperative Path Planning for Range-Only Localization Using a Single Moving Beacon

Yew Teck Tan, Rui Gao, and Mandar Chitre

Abstract—Underwater navigation that relies solely on dead reckoning (DR) suffers from unbounded position error growth. A common approach for alleviating the problem is to have the underwater vehicle surface occasionally for a Global Positioning System (GPS) fix, at the risk of jeopardizing the vehicle's safety and consuming precious mission time. Other alternatives include deploying a long-baseline (LBL) acoustic positioning system in the mission area; this involves substantial deployment effort. The idea of having active mobile beacons as navigational aids has recently gained interest. We explore the use of a single-beacon vehicle for range-only localization to support other autonomous underwater vehicles (AUVs). Specifically, we focus on cooperative path-planning algorithms for the beacon vehicle using dynamic programming and Markov decision process formulations. These formulations take into account and minimize the positioning errors being accumulated by the supported AUV. This approach avoids the use of LBL acoustic positioning systems as well as allows the supported AUV to remain submerged for a longer period of time with small position error. Simulation results and field trials data demonstrate that the beacon vehicle is able to help keep the position error of the supported AUV small via acoustic range measurements.

Index Terms—Autonomous underwater vehicles (AUVs), dynamic programming, Markov decision processes, cross-entropy method, positioning, navigation.

I. INTRODUCTION

UNDERWATER navigation is a challenging problem and has received considerable attention in recent years [1]. As the Global Positioning System (GPS) signals are unavailable, underwater, autonomous underwater vehicles (AUVs) rely on the proprioceptive sensors such as compass, Doppler velocity log (DVL), and Inertial Navigation System (INS) to estimate their position. However, dead reckoning (DR) based on these sensors suffers from an unbounded navigation error growth over time. With inexpensive sensors, this error growth is rapid, while expensive high-quality sensors result in slower (but still unbounded) error growth. Although this problem can be avoided by having the AUV surface and obtain a GPS fix, precious mission time is wasted for making the round trip to the water sur-

face. In addition, the act of surfacing poses safety concerns in busy shipping channels, and may also be undesirable in missions where the AUV is required to be at a specific depth.

Besides using the AUV's proprioceptive sensors, AUVs may make use of fixed beacons for underwater navigation. For example, an underwater long-baseline (LBL) acoustic positioning system functions by measuring distance of the AUV with respect to a framework of baseline beacons to estimate its position. In order for this to work well, the LBL beacons have to be deployed around the AUV's mission area and retrieved upon the completion of the mission. Other underwater acoustic positioning systems such as the ultrashort baseline (USBL) systems offer simpler deployment at the cost of lower positioning accuracy. In the literature, Matos *et al.* [2] have developed a low-cost LBL navigation system for the AUVs, while Rigby *et al.* [3] combined data from a DVL and an USBL system to provide superior 3-D position estimates to the AUV. Another recent solution uses a GPS intelligent buoy (GIB) system which consists of four surface buoys equipped with differential GPS (DGPS) receivers and submerged hydrophones for tracking the position of the submerged AUV [4]. Although these systems act as good navigational aids for AUVs, they suffer from a few drawbacks. First, deploying and retrieving these positioning systems require considerable operational effort. Second, they are generally expensive and operate only within a few square kilometers, making them inflexible and impractical for longer range missions.

Recent advancements in the development of AUVs and underwater communications have made intervehicle acoustic ranging a viable option for underwater cooperative positioning and localization. The idea of AUV cooperative positioning is to have a vehicle with good quality positioning information (beacon vehicle), to transmit its position and range information acoustically to supported AUVs (survey AUVs) within its communication range during navigation. Generally, the beacon vehicle is equipped with high accuracy sensors that are able to estimate its position with minimum errors. In some cases, the beacon vehicle may operate at the surface and have access to GPS for position estimation. The range information between the vehicles can then be fused with the data obtained from proprioceptive sensors in the survey AUVs to reduce the positioning error during underwater navigation [5], [6].

The idea of cooperative positioning with a few vehicles that know their position well and other AUVs with poor navigational sensors is not new. The vehicles with accurate position estimates are referred to by some authors as master vehicles [7], and by others as communication and navigation aids (CNAs) [6], [8]. Although multiple-beacon vehicles can provide higher accuracy

Manuscript received October 31, 2011; revised January 25, 2013; accepted December 15, 2013.

Guest Editor: M. Seto.

The authors are with the Department of Electrical and Computer Engineering and the Acoustic Research Laboratory, Tropical Marine Science Institute, National University of Singapore, Singapore 119227, Singapore (e-mail: william@arl.nus.edu.sg).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JOE.2013.2296361

navigation, our research focuses on single-beacon cooperative positioning due to its operational advantages and lower inter-vehicle communication requirement. This approach has been explored by several researchers [7]–[12], and their work includes observability analysis, algorithms for position determination based on range measurements, and some experimental results. Although all of these authors acknowledge that the relative motion of the vehicles is key to having single-beacon range-only navigation perform well, the problem of determining the optimal path of the beacon vehicle given the desired path of the survey AUVs has received little attention. For example, the work in [7] assumes a circular path for the beacon vehicle, while Fallon *et al.* [8] use a zigzag path during experiments. To maximize the transect period of a survey AUV for cable or pipeline surveyings, Hartsfiel [10] suggested that the leading beacon vehicle would likely have to maneuver off course from its pre-planned path to achieve sufficient relative change of motion to fix the survey AUV's position. More recently, Webster *et al.* [12] also adopted a similar approach and maneuvered the beacon vehicle above the survey site in a diamond shape while keeping station at each apex to increase observability.

The rest of this paper is organized as follows. First, we introduce the concept of multivehicle cooperative positioning using acoustic ranging. This is followed by the cooperative positioning problem formulation. We develop two different cooperative path-planning algorithms for the beacon vehicle which take into account and aim to minimize the localization errors being accumulated by the supported survey AUV. We evaluate the performance of the algorithms through simulation and using field trial data to show that the path followed by the beacon vehicle enables the survey AUV to get acoustic ranging information and keep its position errors small over time. Symbols used in the paper are listed in Table I.

II. COOPERATIVE POSITIONING THROUGH ACOUSTIC RANGING

Cooperative positioning missions typically consist of a beacon vehicle that acts as a navigational aid for the survey AUVs which are deployed for monitoring or surveying missions. By having a beacon vehicle supporting a team of survey AUVs, we can avoid having to equip every single AUV with expensive navigational sensors. This not only reduces the space required to house all the electronics in the vehicles, but also prolongs precious mission time due to lower power consumption.

Cooperative AUVs need to communicate to cooperate. Hence, they are usually fitted with underwater acoustic modems that may also be used to measure range between two vehicles using the travel time of the acoustic signals (Fig. 1). The measurements are typically performed under the assumption of known sound-speed profile. If time synchronization is available, a one-way propagation delay can be measured and used to compute the range between the AUVs. This is known as one-way-travel-time (OWTT) ranging [12]. In the absence of time synchronization, two-way-travel-time (TWTT) ranging has to be used to compute the range. In either case, the position of the beacon vehicle and the estimated range between the vehicles are communicated to the survey AUVs periodically.

TABLE I
SYMBOLS USED IN THE PAPER

Symbols	Description
τ	beacon transmission period
Γ	total mission time
M	total number of survey AUVs
$\mathbf{x}_t^j$	position of the survey AUV j at time t
R_t^j	range between beacon vehicle and survey AUV j at time t
σ^2	variance of range error
ϕ_t^j	heading of the survey AUV at time t
θ_t^j	direction of minimum error for survey AUV j at time t
ϵ_t^j	error along the direction θ_t^j at time t
$\bar{\epsilon}_t^j$	error tangential to the direction θ_t^j at time t
σ_B^2	variance of the beacon vehicle's error
$\mathbf{x}_t^B$	position of the beacon vehicle at time t
ϕ_t^B	heading of the beacon vehicle at time t
s^B	navigational speed of the beacon vehicle
α	constant of proportionality, position error growth factor
δ_t^B	turning angle made by the beacon vehicle at time t
$\dot{\phi}_{\max}^B$	maximum turning rate of the beacon vehicle
$D_{\min,\max}$	minimum and maximum distance between the vehicles
η	optimization criteria for the MDP formulation
ψ	stopping criteria for the MDP formulation
ρ	quantile of beacon paths to be evaluated in the MDP formulation
μ	smoothing parameter for the MDP formulation
ς	tidal current estimation
l	time step between prediction
χ_l^j	state vector of AUV j at time step l
$\mathbf{F}_{l+1}$	state transition matrix from time step l to $l+1$
$\mathbf{u}_{l+1}$	control input
$\mathbf{B}_{l+1}$	control-input matrix
$\mathbf{P}_l$	estimate error covariance
$\mathbf{w}_{l+1}$	process noise vector
$\mathbf{Q}_{l+1}$	covariance of process noise $\mathbf{w}_{l+1}$
$\Upsilon(\phi)$	rotation matrix by the rotation angle ϕ
$\mathbf{z}_{l+1}^j$	measurement of AUV j at time step $l+1$
$h(\chi_l^j)$	measurement function
$\mathbf{C}_{l+1}$	transformation matrix for the tidal current ς

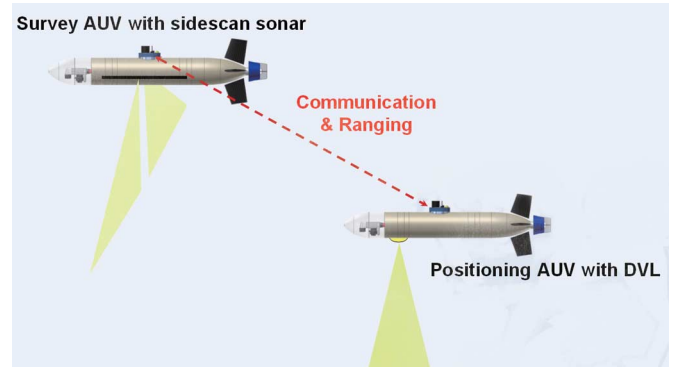


Fig. 1. Two AUVs for cooperative positioning.

Although our focus in this paper is on the problem of a single-beacon vehicle supporting a single-survey AUV, we provide a general mathematical formulation where the beacon vehicle may support multiple-survey AUVs. The approximate paths to be followed by the survey AUVs are preplanned. The beacon

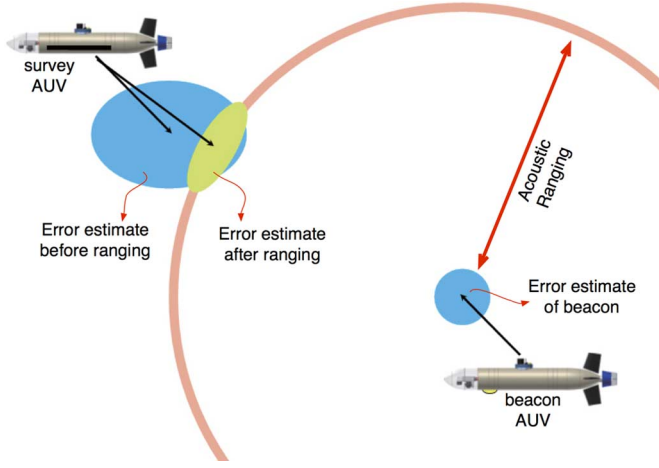


Fig. 2. Illustration of error estimates by range measurements. The error ellipse of the survey AUV (larger blue ellipse next to survey AUV) was reduced (yellow ellipse) by acoustic ranging with beacon vehicle. The error estimate of the beacon vehicle is assumed constant (circle next to beacon vehicle).

vehicle's path is planned in real time through a series of sequential decisions made by the onboard command and control system, using information about the survey AUVs' desired path and reported positions during mission execution. The decisions are made with an optimization criterion that minimizes the error of the survey AUVs' positions, avoids collision between the vehicles, and attempts to keep the vehicles within communication range.

Fig. 2 shows that the position error estimate of the survey AUV is reduced in the radial direction of the ranging circle centered at the beacon vehicle each time a range estimate becomes available. The error in the tangential direction remains approximately unchanged. The cooperative positioning algorithm for the beacon vehicle uses the estimated error ellipse of the survey AUV's position to plan its own motion. If the beacon vehicle can maneuver in such a way that the next range measurement occurs along the direction of the major axis of the error ellipse, the position error of the survey AUV can be minimized. This is the key idea underlying the path planning for the beacon vehicle. Thus, in order for the single-beacon range-only cooperative positioning to perform best, the relative motion between the beacon and the survey vehicles should vary as close to 90° as possible for every consecutive range information transmission. We term this change in relative bearing between vehicles as the change of *relative aspect*. This observation agrees with the work in [13] which claims that "ranging from the same relative direction" is one of the factors that results in the reduction of performance of their approach in AUV navigation using both the acoustic ranging and the sidescan sonar.

III. PROBLEM FORMULATION

We assume that the clocks of the vehicles are synchronized (e.g., using an oven-controlled crystal oscillator [14]), and, therefore, the survey AUVs can estimate their range from the beacon AUV using OWTT ranging. If synchronization is unavailable, TWTT ranging may be used at the expense of a

larger number of communication packets. Since the beacon vehicle makes a navigation decision per beacon transmission period, we represent time using an index $t \in \{0 \dots \Gamma\}$. The elapsed time in seconds from the start of the mission to instant t is simply $t\tau$.

Although the underwater environment is 3-D, it is common that the depth of the beacon and survey vehicles is specified in a mission and may not be altered by our path-planning algorithm. Therefore, we represent the position of each vehicle using a 2-D position vector and the direction of travel of each vehicle by a yaw angle. Let $\mathbf{x}_t^B$ be the position and ϕ_t^B be the heading of the beacon vehicle B at time t . Let M be the number of survey AUVs supported by the beacon vehicle. We index the survey AUVs by $j \in \{1 \dots M\}$. Let $\mathbf{x}_t^j$ represent the position of survey AUV j at time t . At every time index t , we have estimates $\hat{R}_t^j$ of the 2-D range (easily estimated from the measured range by taking into account the difference in depths between the vehicles) between the beacon vehicle and each of the survey AUVs. We model the error in range estimation as a zero-mean Gaussian random variable with variance σ^2 :

$$\hat{R}_t^j \sim \mathcal{N}\left(\left|\mathbf{x}_t^j - \mathbf{x}_t^B\right|, \sigma^2\right). \quad (1)$$

We further model the error in position estimation of the survey AUVs as a 2-D zero-mean Gaussian random variable described by three parameters: the direction θ_t^j of minimum error, the error ϵ_t^j along direction θ_t^j , and the error $\bar{\epsilon}_t^j$ in the tangential direction. Assuming the error in range measurement is much smaller than prior error in survey AUVs' position estimate, the posterior error is minimum along the line joining the beacon and the survey vehicle (see the Appendix)

$$\theta_{t+1}^j = \angle\left(\mathbf{x}_{t+1}^B - \mathbf{x}_{t+1}^j\right) \quad (2)$$

$$\left(\epsilon_{t+1}^j\right)^2 = \sigma^2 + \sigma_B^2 + \alpha\tau \quad (3)$$

where σ_B^2 is the variance of zero-mean Gaussian random variable $\mathcal{N}(0, \sigma_B^2)$, describing the position error of the beacon vehicle and α is the constant of proportionality (determined by the accuracy of the velocity estimate of the survey AUV). The position error of the beacon vehicle is assumed to be isotropic and constant throughout the mission (other error models can easily be accommodated in the formulation). The error in ranging is independent of the error in position. When the distance between the beacon vehicle and the survey AUV is much larger than the positioning error of the survey AUV, the range measurement gives almost no information in the tangential direction, and, therefore, the position error grows in that direction. Assuming that the survey AUVs use velocity estimates (e.g., using DVL or thruster-induced speed) for DR, the position error variance in the tangential direction will grow linearly with time (see the Appendix)

$$\left(\bar{\epsilon}_{t+1}^j\right)^2 = \frac{\left(\epsilon_t^j \bar{\epsilon}_t^j\right)^2}{\left(\epsilon_t^j \cos \gamma_t^j\right)^2 + \left(\bar{\epsilon}_t^j \sin \gamma_t^j\right)^2} + \alpha\tau \quad (4)$$

where $\gamma_t^j = \theta_{t+1}^j - \theta_t^j$.

The navigation decision made by the beacon vehicle at each time step t is δ_t^B , the turning angle during the time interval until the next decision. If $\dot{\phi}_{\max}^B$ is the maximum turning rate

$$|\delta_t^B| \leq \dot{\phi}_{\max}^B \tau. \quad (5)$$

If s^B is the speed of the beacon vehicle, then the heading and position of the vehicle at time $t + 1$ is approximately given by

$$\phi_{t+1}^B = \phi_t^B + \delta_t^B \quad (6)$$

$$\mathbf{x}_{t+1}^B = \mathbf{x}_t^B + \tau s^B \begin{pmatrix} \cos \phi_{t+1}^B \\ \sin \phi_{t+1}^B \end{pmatrix}. \quad (7)$$

To ensure that the beacon and survey vehicles do not collide but are within communication range of each other, we require that

$$D_{\min} \leq \|\mathbf{x}_{t+1}^j - \mathbf{x}_{t+1}^B\| \leq D_{\max} \quad \forall j. \quad (8)$$

We assume that the position of each survey AUV is known at the start of the mission with an accuracy of ϵ_0 in all directions

$$\epsilon_0^j = \bar{\epsilon}_0^j = \epsilon_0 \quad (9)$$

$$\theta_0^j = 0 \text{ (arbitrary choice)}. \quad (10)$$

Given the desired paths $\{\mathbf{x}_t^j \forall t\}$ of the survey AUVs and the initial position $\mathbf{x}_0^B$ and heading ϕ_0^B of the beacon vehicle, we wish to plan a path for the beacon vehicle such that we minimize the sum-square estimated position error across all survey AUVs for the entire mission duration. The path is fully determined by the sequence of decisions $\{\delta_t^B\}$ made during the mission

$$\{\delta_t^B\} = \arg \min \sum_{j,t} \left[\left(\epsilon_t^j \right)^2 + \left(\bar{\epsilon}_t^j \right)^2 \right]. \quad (11)$$

IV. PATH PLANNING FOR THE BEACON VEHICLE

Although the zigzagging and encircling patterns proposed in [7] and [8] managed to reduce the positioning errors of the supported AUVs to some extent, the beacon vehicle did not explicitly take into account the change of relative aspect. Similar results were observed in [12] where the beacon vehicle maneuvered in a diamond-shaped pattern above the survey site to provide range measurements for the survey AUVs. Even though these approaches reduce the problem of unbounded position error growth due to DR, the position errors of the supported survey AUVs may still grow with time if the range measurements were obtained from similar relative bearings among the vehicles.

In this section, we propose two-beacon-vehicle path-planning algorithms that take into account the estimated positioning errors accumulated by the supported survey AUVs. The formulations not only keep track of the relative angles for position error estimation, but also maintain the distance constraints in (8). This, to the best of our knowledge, is the first attempt (other than our previous work in [15]) in maneuvering the beacon vehicle according to a well-defined optimization criterion with respect to the supported survey AUVs.

A. Method 1: Dynamic Programming Formulation

We rewrite the problem definition in the previous section as a dynamic program [15]. The state S_t of the system consists of the positions and heading of the beacon vehicle and the estimated position error of each survey AUV

$$S_t = \left\{ \mathbf{x}_t^B, \phi_t^B, \left(\epsilon_t^j, \bar{\epsilon}_t^j, \theta_t^j \right) \forall j \right\}. \quad (12)$$

The decision space at time t is represented by the set $\mathcal{A}(S_t)$ of turn angles δ_t^B that the beacon vehicle can adopt subject to constraints (5) and (8). Unlike typical dynamic programs, the decision space is not discrete, but continuous, and therefore the set $\mathcal{A}(S_t)$ is an infinite set. When a decision $a_t \in \mathcal{A}(S_t)$ is made, the state change is given by the state transition function $S_{t+1} = \mathcal{T}(S_t, a_t)$ defined by (2), (3), (4), (6), and (7). The decision incurs a cost $C(S_t, a_t)$ equal to the sum-square position error at time $t + 1$ across all survey AUVs

$$C(S_t, a_t) = \sum_j \left[\left(\epsilon_{t+1}^j \right)^2 + \left(\bar{\epsilon}_{t+1}^j \right)^2 \right]. \quad (13)$$

We wish to find an optimal policy (representing a route for the beacon vehicle, comprising a sequence of decisions $\{a_t \in \mathcal{A}(S_t) \forall t \in \{0 \dots \Gamma - 1\}\}$) that will minimize the total cost over the mission duration

$$\{a_t\} = \arg \min_{\{a_t\}} \sum_{t=0}^{\Gamma-1} C(S_t, a_t). \quad (14)$$

A dynamic programming problem naturally lends itself to a recursive formulation based on Bellman's principle of optimality [16]. Bellman's equation introduces a value function $V_t(S_t)$ that represents the cost of applying an optimal policy from a given state S_t until the end of the mission

$$\begin{aligned} V_t(S_t) &= \min_{\{a_{t'}\}} \sum_{t'=t}^{\Gamma-1} C(S_{t'}, a_{t'}) \\ &= \min_{a \in \mathcal{A}(S_t)} \left[C(S_t, a) + \min_{\{a_{t'}\}} \sum_{t'=t+1}^{\Gamma-1} C(S_{t'}, a_{t'}) \right] \\ &= \min_{a \in \mathcal{A}(S_t)} [C(S_t, a) + V_{t+1}(S_{t+1})] \end{aligned} \quad (15)$$

where $S_{t+1} = \mathcal{T}(S_t, a)$. The optimal decision at any point in the mission is then simply given by

$$a_t = \arg \min_{a \in \mathcal{A}(S_t)} [C(S_t, a) + V_{t+1}(S_{t+1})]. \quad (16)$$

1) *Decision Space Approximation*: Problems of the form described above are typically solved using standard methods in dynamic programming such as value iteration or policy iteration [16]. Unfortunately, these methods cannot be applied to our problem as our state space and decision space are continuous (and the corresponding sets of states and decisions are each infinite). Therefore, we resort to approximate methods of solving this problem.

The problem of a continuous decision space can be solved by discretizing the decision space yielding a finite set of decisions that can be made at each stage. The small dimensionality and the constraints placed on the decisions allow us to approximate the decision space by a set of discrete decisions distributed in the space. The decision space is then approximated by the finite

set $\tilde{A}(S_t)$ consisting of A , the number of discrete turn angles uniformly spaced in the range $-\dot{\phi}_{\max}^B \tau$ to $\dot{\phi}_{\max}^B \tau$ and satisfying constraint (8).

Implementation note: For some states S , the decision set $\tilde{A}(S)$ may be an empty set ($\emptyset$). This happens when all potential decisions fail to satisfy constraint (8). To allow the algorithm to recover from such states, if $\tilde{A}(S_t) = \emptyset$, we allow $\tilde{A}(S)$ to contain a single decision that violates the constraint by the least amount. This ensures that the algorithm can continue and potentially recover to a state S' where $\tilde{A}(S') \neq \emptyset$.

2) *Value Function Approximations:* The approach of discretizing the space, unfortunately, cannot be used for the state space. The state space has a high dimensionality, and some of the dimensions represent a physically large geographical area. Discretizing it to a reasonable approximation would require a very large set of discrete states, leading to computationally unfeasible solutions.

Even though we could evaluate $V_t(S_t)$ on a continuous state space by using (15), the computational load of this approach is $\mathcal{O}(A^T)$ and grows very rapidly with the length of the mission. For modest length missions and small values of A , the computational load is already prohibitively large. Therefore, we seek an approximation $\tilde{V}_t(S_t)$ to the value function $V_t(S_t)$ to solve our optimization problem. The approximation can then be used to replace the value function in (16) to give

$$a_t = \arg \min_{a \in \tilde{A}(S_t)} [C(S_t, a) + \tilde{V}_{t+1}(S_{t+1})]. \quad (17)$$

a) *Greedy Strategy:* A trivial approximation for the value function is

$$\tilde{V}_t(S_t) = 0. \quad (18)$$

Since the value function represents the future costs to be incurred from the next state to the end of the mission, setting it to zero is equivalent to ignoring these future costs and making the current decision based purely on the cost incurred by the decision. Hence, we call this a *greedy strategy*.

b) *L-Level Look-Ahead Strategy:* A better approximation for the value function $V_t(S_t)$ is given by evaluating only L terms of the summation in (15)

$$\tilde{V}_t(S_t) = \min_{\{a_{t'}\}} \sum_{t'=t}^{t+L-1} C(S_{t'}, a_{t'}), \quad L \in \mathbb{Z}^+. \quad (19)$$

For example, a 1-level look ahead computes the accumulative cost incurred by the decisions made only for the current and the next time steps. The value of L can be increased for better value function approximation, with the cost of increased computational complexity. Simulation studies [15] showed the best performance-to-computation tradeoff is offered by setting $A = 3$ and 4-level look ahead. For brevity, we denote an L -level look-ahead strategy as LA- L .

B. Method 2: Markov Decision Process Formulation

In this section, we present the formulation of the beacon vehicle's path-planning problem within the Markov decision process (MDP) framework [17]. Generally, an MDP is defined by four main components: the state and action sets, the state

transition probability matrix, and the reward/cost function. From (1), $\hat{R}_t^j$ is the estimated distance between the beacon vehicle and the survey AUV, ϕ_t^B represents the beacon vehicle's current bearing at time t , and ϕ_{t+1}^j is the survey AUV's bearing at time $t+1$, respectively; our state set is defined as a tuple: $z_t = \{\theta_t^j, \hat{R}_t^j, \phi_t^B, \phi_{t+1}^j\}$. Since we assume that ϵ_{t+1}^j in (3) is a constant, we need to minimize $\bar{\epsilon}_{t+1}^j$ in (4) to obtain (11) for every time step t . This means having γ_t^j in (4) to be as close as possible to 90° . Thus, the ability of beacon vehicle B to achieve this with respect to survey AUV j will depend on its knowledge of the components in the state space as well as the actions that it can take. Both $\hat{R}_t^j$ and θ_t^j can be obtained from the acoustic range measurements and communication between the AUVs while ϕ_{t+1}^j is usually preplanned before the mission.

The action a_t is the turning angle from the beacon vehicle's current bearing (ϕ_t^B), $|a_t| \leq \dot{\phi}_{\max}^B \tau$. At every time t , after a_t is selected, the corresponding $\mathbf{x}_{t+1}^B$ can be calculated and the accumulated sum-square error can be estimated through (3) and (4). We model this accumulated error as the cost function C , and we are interested in minimizing this cost over the entire mission path, which is equivalent to solving (11). An MDP policy is the state-action mapping that determines the probability distribution of action a_t , when the process is in the state z_t at time step t . We discretize a_t into N_a action states and z_t into N_z states, and define a policy matrix $\mathbf{P}_{za} = (p_{za})$ with $z \in \{1 \dots N_z\}$ and $a \in \{1 \dots N_a\}$, such that for each state z , we choose action a with probability p_{za} . This requires that for all z rows in $\mathbf{P}_{za}$, the sum of each z th row is equal to 1. In the case of cooperative path planning, this translates into the probability of choosing a particular turning angle from the beacon vehicle's current bearing (termed as desired heading in the rest of the paper) at time $t+1$, given the beacon vehicle's current bearing, survey AUV's next heading as well as distance and relative angle between the AUVs. As a result, the cost minimization problem reduces to determining the beacon vehicle's path-planning policy.

1) Policy Learning Using the Cross-Entropy Method:

a) *Cross-Entropy Method:* We now briefly introduce the cross-entropy (CE) method and its application in learning the MDP policy. The CE method was initially introduced for estimating the probability of rare events in complex stochastic networks [18]. Later, it was modified to solve the combinatorial optimization problem (COP). The main idea behind the CE method in solving COP is the association of an estimation problem with the optimization problem, which is called associated stochastic problem (ASP). This ASP, once defined, can be tackled efficiently by iterative estimation procedure shown in Algorithm 1. In what follows, we present the simplified version of the CE method and refer the interested readers to [18] and [19] for its detailed development and formulation.

Suppose we wish to minimize some cost function C on space χ , where χ is the action space defined in the MDP shown in Section IV-B. Let η^* denote the minimum of C on χ , $\eta \in \mathbb{R}^+$

$$\eta^* = \min_{\mathbf{x} \in \chi} C(\mathbf{x}). \quad (20)$$

We define a collection of indicator functions $\{I_{\{C(\mathbf{x}) \leq \eta\}}\}$ on χ for various thresholds or levels η . Let $\{f(\cdot; \mathbf{P}_{za}), \mathbf{P}_{za} \in V\}$ be a family of (discrete) probability density function (pdfs) on

χ , parameterized by a real-valued parameter $\mathbf{P}_{za}$. For a certain $p \in V$, we can associate with (20) the following estimation problem:

$$l(\eta) = \mathbb{P}_p(C(\mathbf{x}) \leq \eta) = \sum_{\mathbf{x}} I_{\{C(\mathbf{x}) \leq \eta\}} f(\mathbf{x}; p) = \mathbb{E}_p I_{\{C(\mathbf{x}) \leq \eta\}} \quad (21)$$

where $\mathbb{P}_p$ is the probability measure under which the random vector $\mathbf{x}$ has pdf $f(\mathbf{x}, p)$. The association comes from the fact that the probability $\mathbb{P}_p(C(\mathbf{x}) \leq \eta)$ will be very small (rare event) when η is close to η^* . By the CE method, this rare event can be estimated by iteratively generating and updating a sequence of tuple $\{(\hat{\eta}_n, \hat{p}_n)\}$ such that it will converge to a small region of the optimal tuple (η^*, p^*) . Let ψ be the stopping criterion, then the tuple $(\hat{\eta}_n, \hat{p}_n)$ can be updated iteratively by Algorithm 1.

Algorithm 1: Iterative Estimation

- Let $\eta_0 = 0$ and $p_0 = 1/|\chi|$, set $n = 0$

repeat

- Set $n = n + 1$
- Let η_n be the $(1 - \rho)$ -quantile of $C(\mathbf{x})$ under p_{n-1}
- Generate a set of N random vector from $f(x, p_{n-1})$, denoted as $\mathbf{x}_k$ for $k \in \{1 \dots N\}$
- Estimate η_n , denoted as $\hat{\eta}_n$, by assigning it as the $(1 - \rho)$ -quantile of $C(\mathbf{x}_k)$ where $C(\mathbf{x}_k) \in \{C(\mathbf{x}_1) \leq \dots \leq C(\mathbf{x}_N)\}$
- Estimate p_n , denoted as $\hat{p}_n$, with fixed $\hat{\eta}_n$ and p_{n-1} . The estimation can be derived from [18] as

$$\hat{p}_n = \arg \max_p \frac{1}{N} \sum_{k=1}^N I_{\{C(\mathbf{x}_k) \leq \hat{\eta}_n\}} \ln f(\mathbf{x}, \mathbf{P}_{za}). \quad (22)$$

until $|\hat{\eta}_n - \hat{\eta}_{n-1}| \leq \psi$

To sum up, the CE method generally consists of two important phases.

- 1) Generate sample data $\mathbf{x}$, according to a specified random mechanism (pdf parameterized by the vector p). Score and rank the resultant sample data according to the cost function $C(\mathbf{x})$.
- 2) Select the η and updating the parameters of the pdfs on the basis of the data, to produce a “better” sample in the next iteration.

b) Beacon Vehicle's Path-Planning Policy Learning: To apply the CE method for learning the path-planning policy, we must specify the two important phases stated before, which in our case are: 1) how to generate the sample beacon path, and 2) how to update the policy matrix at each iteration.

Since we have formulated the path-planning problem within the MDP framework, for a given survey AUV's path with arbitrary path length of t_{LA} steps, we can generate a set of beacon paths with the same path length via the Markov process with the policy matrix $\mathbf{P}_{za}$. Let N be the total number of paths generated in the set, then each beacon vehicle's path $\mathbf{x}_k$, $k \in \{0 \dots N\}$, consists of a sequence of state-action

pairs $\mathbf{x}_k = (z_0, a_0, \dots, z_{t_{LA}}, a_{t_{LA}})$. The cost of each resultant beacon vehicle's path can be estimated through (3) and (4), as shown in Section III.

Let $C(\mathbf{x}_k)$ represent the total cost of path $\mathbf{x}_k$ generated for policy learning at every iteration, then we sort the paths' cost in increasing order and evaluate the $(1 - \rho)$ -quantile η . Once the η is selected, the policy matrix can be updated by solving (22) to obtain the formula (see [18] and [19])

$$p_{za} = \frac{\sum_{k=1}^N I_{\{C(\mathbf{x}_k) \leq \eta\}} I_{\{\mathbf{x}_k \in \chi_{za}\}}}{\sum_{k=1}^N I_{\{C(\mathbf{x}_k) \leq \eta\}} I_{\{\mathbf{x}_k \in \chi_z\}}} \quad (23)$$

where $C(\mathbf{x}_k) \leq \eta$ means that the total cost of path $\mathbf{x}_k$ is less than the selection score, the event $\{\mathbf{x}_k \in \chi_z\}$ means that the trajectory $\mathbf{x}_k$ contains a visit to state z , while the event $\{\mathbf{x}_k \in \chi_{za}\}$ means that the trajectory corresponding to path $\mathbf{x}_k$ contains a visit to state z , in which action a was taken. The learning process is repeated until η converges within the stopping criterion. Detailed steps are shown in Algorithm 2.

Algorithm 2: Policy Learning Through Iterative Estimation

Require: $\mathbf{P}_{za}$ uniformly initialized with $(1/|N_a|)$

- Let $\eta_0 = 0$, set $n = 0$

repeat

- Set $n = n + 1$

for all z^s in $\mathbf{P}_{za}$ **do**

- Generate a random surveying path with the first path segment satisfies z^s and path length of t_{LA} steps.

repeat

- Start from the initial state $z_0 = z^s$, set $i = 0$.
- Generate an action a_i according to the z_i th row of $\mathbf{P}_{za}$, calculate the cost $C_i = c(z_i, a_i)$, and generate a new state z_{i+1} . Set $i = i + 1$. Repeat till $i = t_{LA}$.
- Output the total cost ($C(\mathbf{x}_k)$) of the trajectory $(z_0, a_0, \dots, z_{t_{LA}}, a_{t_{LA}})$.

until N trajectories

- Sort the N scores in descending order, and take η_n as the $(1 - \rho)$ -percentile of the score set.

- Update the parameter matrix $\mathbf{P}_{za}$ according to (23).

end for

until $|\eta_n - \eta_{n-1}| \leq \psi$

Instead of updating the policy matrix $\mathbf{P}_{za}$ directly with (23), we apply a simple smoothing filter

$$\hat{p}_{za,n} = \mu \tilde{p}_{za,n} + (1 - \mu) \hat{p}_{za,n-1} \quad (24)$$

where $\tilde{p}_{za,n}$ is the solution of (23) and μ is the smoothing parameter with $0.7 < \mu < 1$. The filter serves two purposes: i) smoothing the policy matrix update, and ii) preventing $\hat{p}_{za,n}$ from becoming zero especially during the initial stage of the learning process. This is crucial as to prevent the learning algorithm from finding a local minimum and converging to an incorrect solution.

2) *Application to Cooperative Path Planning:* Once the policy learning is completed, the path planning for the beacon vehicle supporting a single-survey AUV reduces to a policy

matrix lookup. At every planning step, the beacon vehicle determines its current state and decides on its next heading using the corresponding action row's probability distribution. This process is repeated until the survey AUV's mission is completed.

C. Comparison of the Computational Load

Γ is the length of the mission and A is the size of decision space. The computational load for the DP method in generating an optimal route using the greedy strategy to support a single-survey AUV is $\mathcal{O}(\Gamma A)$ and only increases linearly with the length of the mission. However, its computational load increases exponentially to $\mathcal{O}(\Gamma A^{L+1})$ if L -level of look-ahead strategy is employed, which can be significantly higher than that of the greedy strategy (equivalent to 0-level look-ahead strategy).

On the other hand, the computational load incurred by the MDP method is only $\mathcal{O}(\Gamma)$. Although the computational load is significantly lower than the DP method, the process of the decision making in the MDP method is heuristic, and generates suboptimal routes. Furthermore, the policy learning step can be time consuming even though it can be performed offline.

We performed a bench test using the STARFISH [20] AUV's single-board computer [LIPPERT Cool LiteRunner-ECO, 1.6-GHz central processing unit (CPU) clock speed, 1-GB DDR2 RAM] to determine the amount of time taken by both methods in producing comparable results. The DP method with $L = 4$ and $A = 8$ took about 480 ms to compute the first desired heading, while the MDP method trained with the same size of decision space took about 600 ns for each of its policy lookups.

V. POSITION ESTIMATION OF THE SURVEY AUV

The model in Section III is used by the beacon vehicle to keep track of the position errors accumulated by the supported survey AUVs. For position estimation, the survey AUVs must take into account the range measurements from the beacon vehicle as well as the information from its navigational sensors. In this section, we formulate a state-space-based position estimation for the survey AUV using the extended Kalman filter (EKF) [21]. The formulation also estimates the ocean current, which can be utilized by the onboard control system for navigation.

A. Dynamic System Model

Let the time step between each state prediction be l . The elapsed time between steps $l + 1$ and l is Δl . This is usually equal to or less than the range update interval τ . The state vector of the survey AUV j at time l is $\mathbf{x}_l^j = [s_e, s_n, v_e, v_n, c_e, c_n]^T$, where the AUV's 2-D position $\mathbf{x}_l^j = [s_e, s_n]^T$, AUV's velocity $[v_e, v_n]^T$, and tidal current's velocity $[c_e, c_n]^T$ are expressed in east and north directions in the navigation frame. Let $\hat{\mathbf{x}}$ be the estimate of the state vector. The state estimate $\hat{\mathbf{x}}_{l+1|l}$ denotes the prediction of the state vector from time l to $l + 1$ before any measurement update. It is expressed as

$$\hat{\mathbf{x}}_{l+1|l}^j = \mathbf{F}_{l+1} \hat{\mathbf{x}}_l^j + \mathbf{B}_{l+1} \mathbf{u}_{l+1} \quad (25)$$

where $\mathbf{u}_{l+1} = [u_e, u_n]^T$ is the control input that determines the AUV's motion. The control input (derived from commanded heading and thrust) is the commanded velocity at which AUV should move in the east and north directions for the next step, given the current position and heading estimates, and the pre-planned path. The state transition matrix is

$$\mathbf{F}_{l+1} = \begin{bmatrix} 1 & 0 & \Delta l & 0 & \Delta l & 0 \\ 0 & 1 & 0 & \Delta l & 0 & \Delta l \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (26)$$

and the control-input matrix is

$$\mathbf{B}_{l+1} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}. \quad (27)$$

Correspondingly, the predicted estimate covariance is

$$\mathbf{P}_{l+1|l} = \mathbf{F}_{l+1} \mathbf{P}_l \mathbf{F}_{l+1}^T + \mathbf{Q}_{l+1} \quad (28)$$

where $\mathbf{P}_{l+1|l}$ is the error covariance of the state estimate $\hat{\mathbf{x}}_{l+1|l}$ and $\mathbf{Q}_{l+1}$ is the covariance of the zero-mean independently distributed process noise $\mathbf{w}_{l+1}$ in the state propagation

$$\mathbf{x}_{l+1}^j = \mathbf{F}_{l+1} \mathbf{x}_l^j + \mathbf{B}_{l+1} \mathbf{u}_{l+1} + \mathbf{w}_{l+1}. \quad (29)$$

The off-diagonal elements of $\mathbf{Q}$ are zero, and the diagonal elements are defined by σ_w^2 .

B. Measurement Model

There are three possible measurements that can be obtained onboard the survey AUV. When equipped with the DVL, the velocity measured is the surge and sway speed in the AUV's body frame $[\nu_h, \nu_s]^T$. In the horizontal plane, the compass measures the heading as ϕ (to be positive clockwise from north). During the mission, the acoustic range measurement R is measured from a beacon with position $\mathbf{x}^B = [b_e, b_n]$.

The body-frame velocity measured by the DVL can be expressed as the multiplication between the velocity in navigation frame and a rotational matrix $\Upsilon(\phi)$

$$\begin{bmatrix} \nu_s \\ \nu_h \end{bmatrix} = \Upsilon(\phi) \begin{bmatrix} v_e + c_e \\ v_n + c_n \end{bmatrix} = \begin{bmatrix} (v_e + c_e) \cos \phi - (v_n + c_n) \sin \phi \\ (v_e + c_e) \sin \phi + (v_n + c_n) \cos \phi \end{bmatrix}. \quad (30)$$

This can be easily extended to 3-D space by measuring the 6 degrees of freedom if desired.

When all the measurements are available, the observation matrix for the survey AUV j is

$$\begin{aligned} \mathbf{z}_{l+1}^j &= \begin{bmatrix} \nu_h \\ \nu_s \\ \phi \\ R \end{bmatrix}_{l+1} \\ &= h(\mathbf{x}_{l+1}^j) + \mathbf{v}_{l+1} \end{aligned} \quad (31)$$

where the measurement noise $\mathbf{v}$ is assumed to be independent Gaussian $\sim \mathcal{N}(\mathbf{0}, \mathbf{V})$. The diagonal elements of the covariance matrix $\mathbf{V}$ are σ_v^2 . The measurement function is then

$$h(\mathbf{x}_{l+1}^j) = \begin{bmatrix} (v_e + c_e) \cos \varphi + (v_e + c_e) \sin \varphi \\ (v_e + c_e) \cos \varphi - (v_e + c_e) \sin \varphi \\ \varphi \\ \|\mathbf{x}^j - \mathbf{x}^B\| \end{bmatrix}_{l+1} \quad (32)$$

where we define $\varphi = \arctan(v_e/v_n)$ and the distance between the survey AUV j and the beacon vehicle

$$\|\mathbf{x}^j - \mathbf{x}^B\| = \sqrt{(\hat{s}_e - b_e)^2 + (\hat{s}_n - b_n)^2}.$$

Once the prediction and measurement models are formulated, the EKF's update step follows the standard procedures, as shown in [21].

VI. RESULTS AND DISCUSSION

We implemented the cooperative path-planning algorithms and evaluated their performance in a simulated environment. Various simulation studies were conducted to investigate the localization performance of the survey AUV supported by a single-beacon vehicle. For comparison purposes, the simulations were conducted with four different types of ranging aids, each transmitted from a single beacon. The ranging aids used were: single fixed beacon, circularly moving beacon, and cooperative beacon (beacon vehicle) where its paths were planned using the DP method and the MDP method, respectively. The fixed beacon remains stationary throughout the mission, while the circularly moving beacon maneuvers in a circular pattern around the center of the survey site. Finally, we also conducted a simulation where the survey AUV relied solely on DR for navigation to further illustrate the rate of position error growth without range measurement.

This section consists of two parts. In the first part, the AUV's dynamic model, sensor measurements, tidal conditions, and vehicle control are all simulated to emulate the real conditions. We evaluate the performance of a single-beacon vehicle in supporting a single-survey AUV. In the second part, the sensor data from a survey AUV during a field experiment are used in conjunction with simulated ranging data to estimate the performance of the proposed algorithms. The GPS measurements of the survey AUV were used as the ground truth to show the advantages of having a cooperative beacon vehicle.

A. Simulation

1) *Simulation Setup*: The acoustic range updates are assumed to occur at a fixed interval $\tau = 10$ s for the first part of the simulation study. However, the updates may be sporadic in reality due to the communication packet loss, and this may affect the computation of the position estimate and the error estimate of the survey AUV. From the measured acoustic ranging statistics during the field trial, the range updates occur between 5 and 20 s with some exceptions due to packet loss. In the second part of the simulation study, we simulate range updates received by the survey AUV to occur at any time uniformly distributed between 5 and 20 s, with a packet loss probability of

TABLE II
PARAMETER FOR POLICY LEARNING

Parameter	Value
t_{LA}	20
σ	1 m
$\dot{\phi}_{\max}^B$	0.07 rad/s
ϵ_0	1 m
α	0.1 m ² /s
N	200
ρ	0.1
μ	0.9
ψ	0.1

TABLE III
STATE-SPACE AND ACTION-SPACE DISCRETIZATION

State Space, N_z	Number of States
Beacon vehicle's current bearing	36
Surveying AUV's next bearing	36
Relatives angle between AUVs	36
Distance between AUVs	3
Total :	139968

Action Space, N_a	Number of Action States
Beacon vehicle's desired turning angle	8

0.46, to match the statistics collected from a recent field trial. The difference in terms of the range update frequency allows us to investigate the robustness of the resultant path-planning algorithms in handling the uncertainty associated with the acoustic communication.

In total, 100 runs were conducted for each simulation scenario. Throughout the simulation studies, we assume that the survey AUV is not equipped with a DVL. The only available measurements for position estimation are the compass and the acoustic ranges. We observed that performance of the survey AUV's localization is affected both favorably and unfavorably depending on the starting position of the beacon vehicle (Section VI-A4b). Thus, the fixed beacon and starting position of the circularly moving beacon are positioned randomly for each run to average out the effect. Four levels of look-ahead (LA-4) strategy and three discrete turn angles ($A = 3$) were used for the DP method, while the policy table trained with the CE method was used in the MDP method. Detailed policy training setup for the MDP method is shown in Section VI-A2.

2) *Policy Learning Setup for the MDP Method*: The learning algorithm shown in Section II was used with the setup shown in Table II.

In this approach, we do not need to discretize our map into a grid map since we are only concerned with the relative angle between the vehicles. However, we do discretize the angle between the vehicles and the vehicles' bearing into 36 states, each representing an angle section of 10° spanning from 0° to 360°. The vehicles are allowed to navigate between 100 and 1000 m from each other, and the distance is discretized into three states with first two zones having 300 m each, while the last zone spans 400 m. Any distance shorter than 100 m or longer than 1000 m will be given a heavy penalty that will contribute to the accumulated error. This is necessary to prevent the vehicles from colliding if they are too close to each other, while keeping the vehicles within the communication range (which in our case is

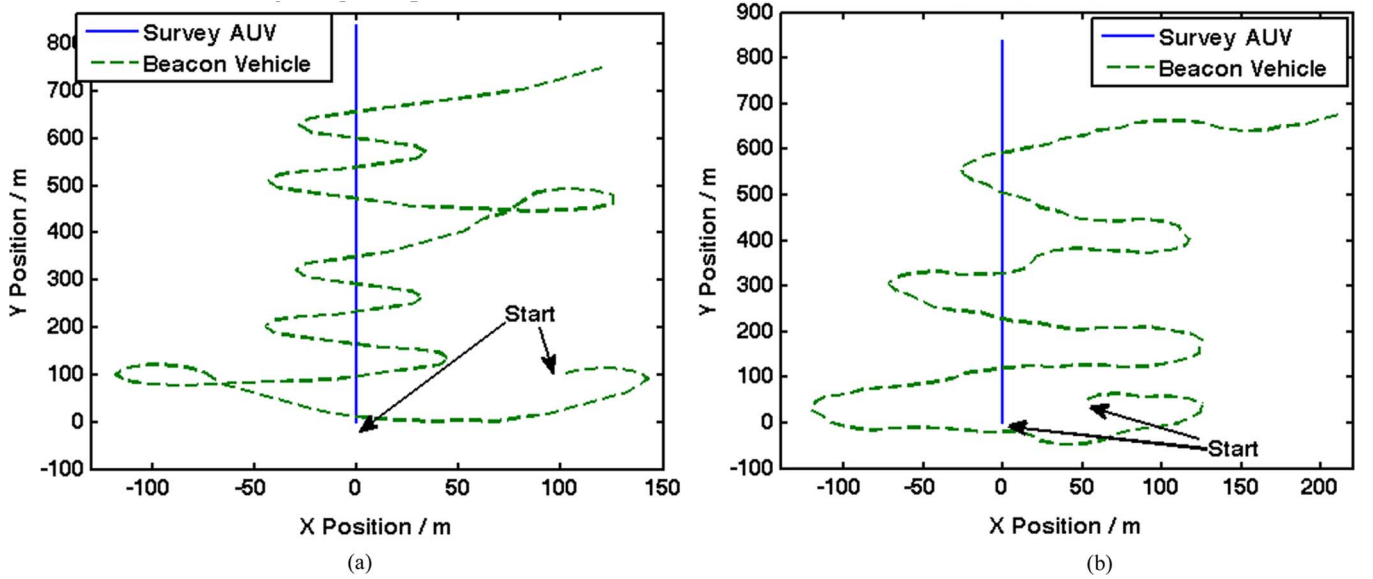


Fig. 3. Beacon vehicle's paths planned by (a) DP and (b) MDP methods in supporting the survey AUV moving in a straight line.

assumed to be 1000 m). The maximum turning angle achievable by the vehicle is 40° per beacon transmission period (assuming $\tau = 10$ s) and is discretized into eight action states. The number of beacon vehicle's paths N generated during the policy learning was determined through experimentation. The policy learning may fail to converge if the value of N is set too low while setting it too high will significantly slow down the learning process. The state space and the action space formulated for the policy learning are summarized in Table III.

3) *Survey AUV's State Estimation*: The state propagation of the survey AUV is simulated as

$$\mathbf{x}_{l+1}^j = \mathbf{F}'_{l+1} \mathbf{x}_l^j + \mathbf{B}_{l+1} \mathbf{u}_{l+1} + \mathbf{C}_{l+1} \boldsymbol{\varsigma}_{l+1} + \mathbf{w}_l. \quad (33)$$

The difference between this propagation model from that of (29) is the propagation matrix $\mathbf{F}'$ and the tidal current $\boldsymbol{\varsigma}$. The propagation $\mathbf{F}'$ is the same as $\mathbf{F}$ shown in (29), except that the last two diagonal elements in $\mathbf{F}'$ are 0. This is to simulate the ocean current $[c_e, c_n]^\top$ as a combination of the tidal current and the nontidal current. The tidal current $\boldsymbol{\varsigma}$ is assumed constant throughout the mission and is randomly selected between 0.1 and 0.2 m/s, while the nontidal current is simulated as the process noise $\mathbf{w}$. The transformation matrix for the tidal current is

$$\mathbf{C}_{l+1} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad (34)$$

At the beginning of the simulation, the ocean current velocity in the state vector is assumed to be zero as we do not have an estimate of the ocean current. The EKF updates from acoustic ranging will help estimate the actual ocean current. In practice, if a tidal current estimate is available from tidal forecasts, then it should be used to seed the state vector.

TABLE IV
PARAMETERS OF PATH PLANNING USING THE MDP AND DP METHODS

Parameters	value
$\phi_{\max}^B$	0.07 rad/s
σ	0.2 m
$[D_{\min}, D_{\max}]$	[100, 1000] m
α	0.1 m ² /s
ϵ_0	1 m

Let the thrusting speed of the beacon vehicle and the survey AUV j be s_B and s_S^j , both at 1.5 m/s. The process noise for the AUV's position (in meters) and velocity and ocean current velocity (both in meters per second) has standard deviation $\boldsymbol{\sigma}_w = [0.1, 0.1, 0.1, 0.1, 0.005, 0.005]^\top$. The measurement noise $\mathbf{v}$ has a standard deviation $\boldsymbol{\sigma}_v = [0.01, \sigma]^\top$ where 0.01 is for the heading measurement in degrees and σ is for the ranging accuracy defined in Table IV.

A simple control model is simulated for the navigation of the survey AUVs to follow the preplanned mission path. Based on the estimated position $\hat{x}_l^j$ and heading $\hat{\phi}_l$ at time l , the survey AUV calculates its next thrusting velocity $\mathbf{u}_{l+1}$ to navigate through its preplanned path. The navigation is subject to AUV's dynamic constraints such as the maximum turning rate and speed.

4) Beacon Vehicle Supporting Single-Survey AUV:

a) *Survey AUV With Straight Path*: In the first simulation scenario, a survey AUV was given a simple straight path (blue solid line), as shown in Fig. 3. The survey AUV paths are preplanned and are shared with the beacon vehicle. Starting from the same initial position, the beacon vehicle plans its path using the DP method [Fig. 3(a)] and the MDP method [Fig. 3(b)]. The objective of the simple simulation is to provide better intuition behind both DP and MDP methods in cooperative positioning algorithms. The simulation results show that, given a straight survey path, the beacon vehicle maneuvers back and forth across the survey AUV's track to maximize the change of relative aspect when the acoustic range information is exchanged. Also,

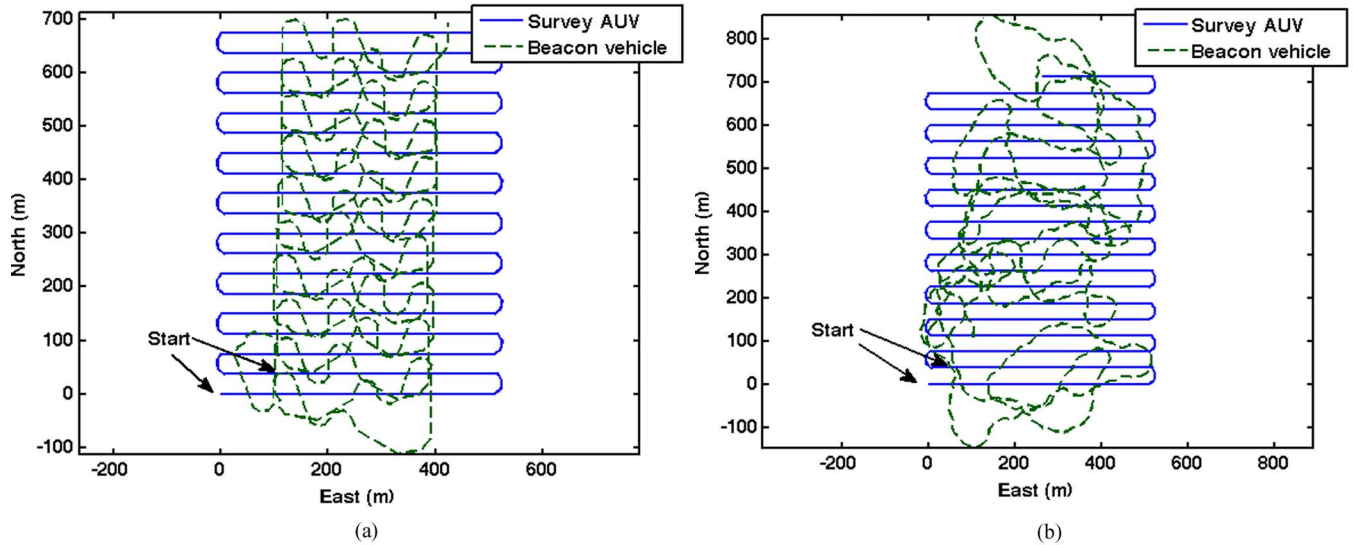


Fig. 4. Beacon vehicle path planned by the (a) DP and (b) MDP methods supporting single-survey AUVs moving in a lawn-mowing pattern.

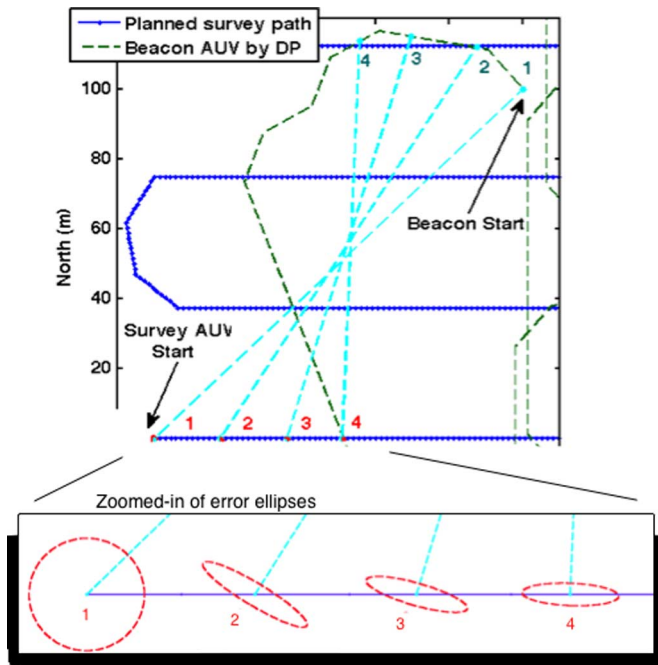


Fig. 5. Evolution of error ellipse estimated by the survey AUV for the first four acoustic range updates provided by the beacon vehicle using the DP method.

the resultant paths maneuver the beacon vehicle in the direction of the survey AUV to keep them within the communication range.

b) Survey AUV With Lawn-Mowing Path: In the second simulation scenario, a survey AUV was given a lawn-mower mission surveying an area of $500 \times 700 \text{ m}^2$. Fig. 4 shows the resultant paths of the beacon vehicle supporting a single-survey AUV using the DP and MDP methods. To further illustrate the process of cooperative path planning, we provide a zoomed-in capture of the first four steps planned by the DP method [Fig. 4(a)] together with the corresponding position error ellipses estimated by the supported survey AUV in Fig. 5.

Although both paths generated by the algorithms are different, they generally maneuver the beacon vehicle around the survey site to maximize the change of relative aspects while

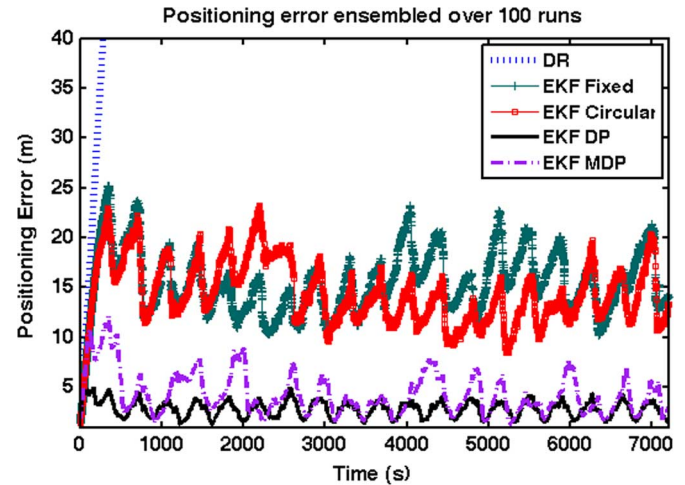


Fig. 6. Positioning error accumulated by the survey AUV supported by different types of beacons.

maintaining the intervehicle distance constraints. Fig. 6 shows the root-mean-square (RMS) positioning errors of the survey AUV compared with its actual position over 100 simulated runs. Without ranging, the positioning error of the survey AUV using the DR grows sharply and unbounded. This is mainly due to the drift caused by the simulated tidal current. However, with the help of ranging, the positioning error growth can be reduced and the ocean current estimation can be improved.

It should be noted that the positioning error estimated by the EKF with a fixed beacon also grows unbounded, but at a rate slower than DR. As the survey AUV moves farther away from the fixed beacon, the achievable change of relative aspect between the two vehicles decreases. Although the positioning error in the direction between the beacon and the survey vehicles is bounded by the range updates, the tangential position error grows. The total positioning error, therefore, increases with time. On the other hand, the positioning error estimated by the EKF with circularly moving beacon shows slightly better performance. We observe that the performance of each run varies significantly with the initial position of the beacon vehicle. This is further illustrated later in this section.

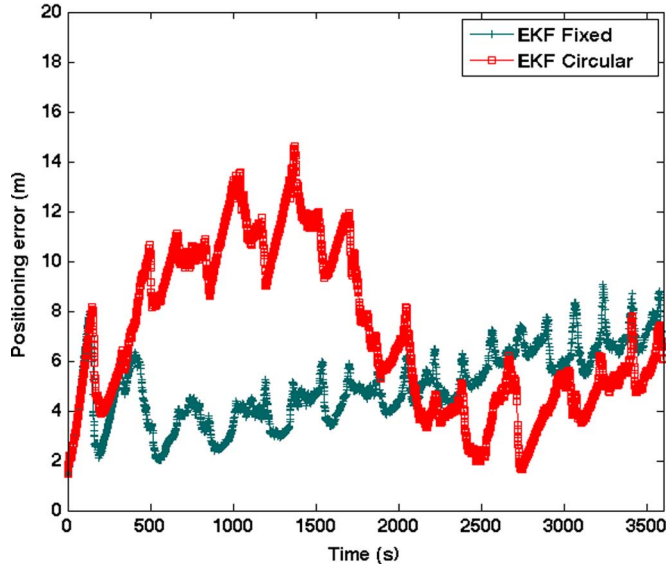


Fig. 7. Nonrandom starting position (at $[0, -10]$ m): Positioning error of multiple-survey AUVs supported by different types of beacon.

The positioning error of the survey AUV is lower (as compared to the case of fixed and circularly moving beacons) when the DP method or the MDP method is used (Fig. 6). The oscillating pattern of the positioning errors calculated by the EKF DP and the EKF MDP are due to the lawn-mower pattern of the survey AUV. Whenever the survey AUV is moving in a straight line, it is more difficult for the beacon vehicle to achieve maximum change of relative aspect ($\pm 90^\circ$) at every range update, and, hence, the positioning error grows slowly. Although the DP method performs slightly better than the MDP method, the latter method requires much lower computational power as the decision making of the beacon vehicle is just a policy matrix lookup, while the computationally intensive policy matrix generation is an offline onetime task.

One important observation is that performance of both DP and MDP methods is fairly independent of the starting position of the beacon vehicle, provided that it is within the communication range of the survey AUV. The same is not true for the case of the fixed and circularly moving beacons. We observed that their performance relies heavily on the initial location of the beacon, even though it is not clearly seen in Fig. 6. To illustrate this, we performed the same cooperative mission shown in Fig. 4 with both fixed and circularly moving beacons started at the initial location $([0, -10])$ for all the 100 runs. The results (in Fig. 7) clearly show that the positioning error using the fixed beacon grows rapidly over time, while the positioning error with the circularly moving beacon fluctuates. Since the exchange of the acoustic ranging is solely periodic without considering the survey AUV's position error accumulation, the performance of the fixed and circularly moving beacons improves when the periodic ranging occurs in the direction of a maximum error and drops otherwise.

B. Performance Estimation With Field Data

Next, we present performance estimates based on the survey AUV's data obtained during field experiments with a simulated beacon vehicle and range measurements. The field experiment

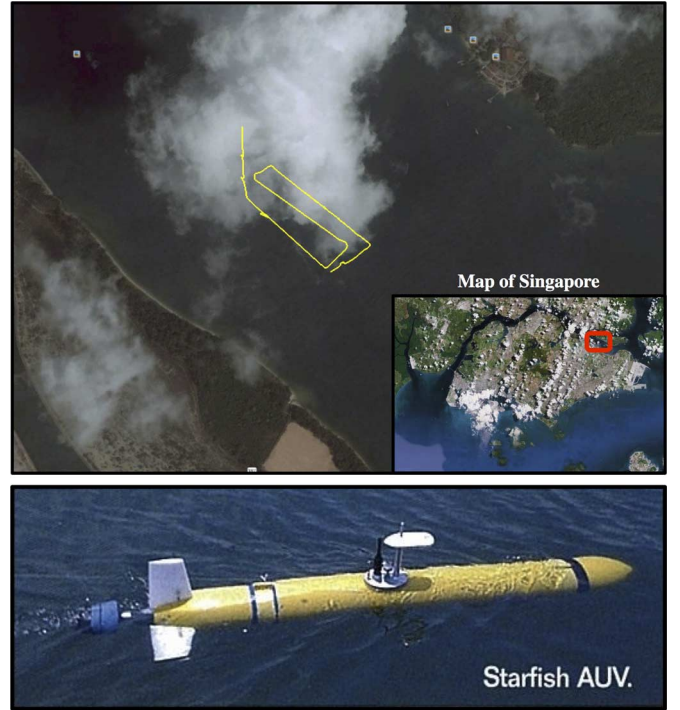


Fig. 8. Top: Field trial near the Serangoon Island, Singapore. Bottom: The STARFISH AUV [20].

provided us with valuable navigational data collected from the survey AUV's proprioceptive sensors that could not otherwise be reproduced in the simulation environment. In addition, the environmental uncertainties due to tidal current also allowed us to test the robustness of the algorithms in handling unexpected natural events.

1) *Experimental Setup*: On July 9, 2011, a field trial was conducted near Serangoon Island, Singapore, using the STARFISH AUV (Fig. 8). The STARFISH AUV [20] was deployed to perform a simple surface surveying mission with GPS position estimates as ground truth. The navigational data were collected and used as the preplanned path for the simulated beacon vehicle. During the simulation, the position of the survey AUV was estimated (assuming no GPS) using only compass measurements and simulated acoustic range updates. Only the first few GPS updates were used to initialize the position of the survey AUV. Simulated acoustic range updates, as described in Section VI-A1, were used; these were available only when a GPS fix was available, as it was required as ground truth.

2) *Experimental Results and Discussion*: Fig. 9(a) shows the real path of the survey AUV and the resultant beacon paths generated with different beacon types, while Fig. 9(b) shows the accumulated positioning errors of the supported survey AUV.

Throughout the mission, in total, 77 simulated acoustic range updates were received by the survey AUV (simulated with packet lost probability of 0.46). It can be seen that the DR method without range measurement produces the worst position estimation for the survey AUV. Since the GPS updates were only available at the beginning, the position estimation using the DR method started to drift uncontrollably throughout the rest of the mission. However, the error growth rate is different at different mission legs depending on the prevalent

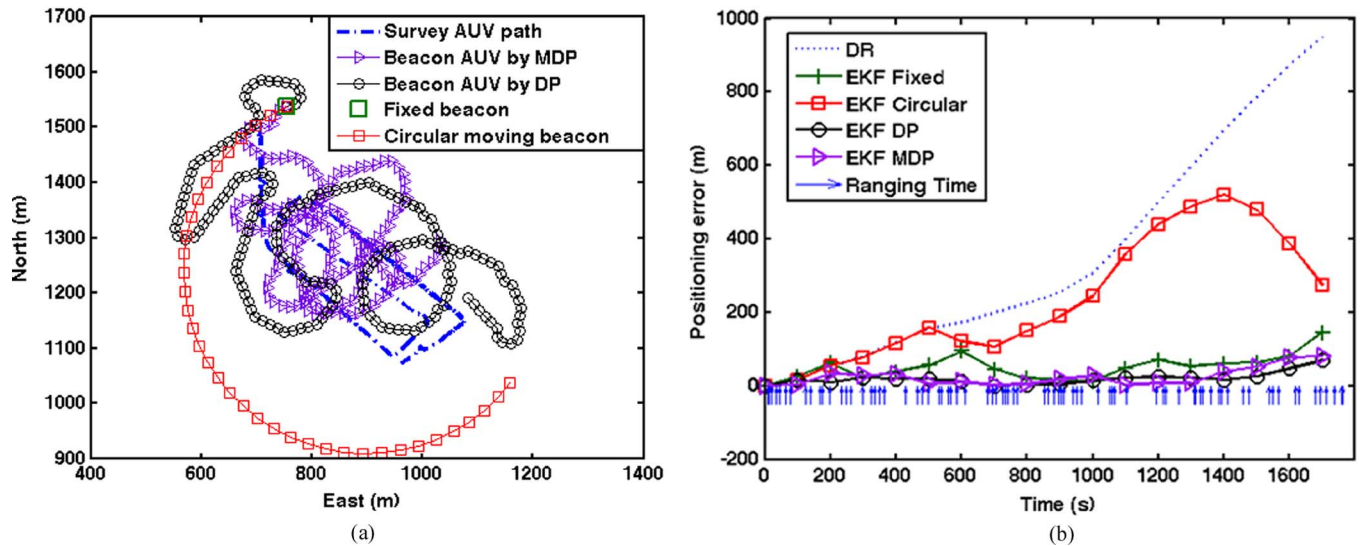


Fig. 9. Performance estimate using the field data collected on July 9, 2011, near the Serangoon Island, Singapore. The beacon vehicle starts at an offset of [50, 50] m from the survey AUV. (a) Planned paths by varies types of beacons overlaying the preplanned path of the survey AUV. (b) Positioning errors of the survey AUV supported by different types of beacons. The vertical lines (blue) at the bottom of the plot show the time when there is an acoustic range update.

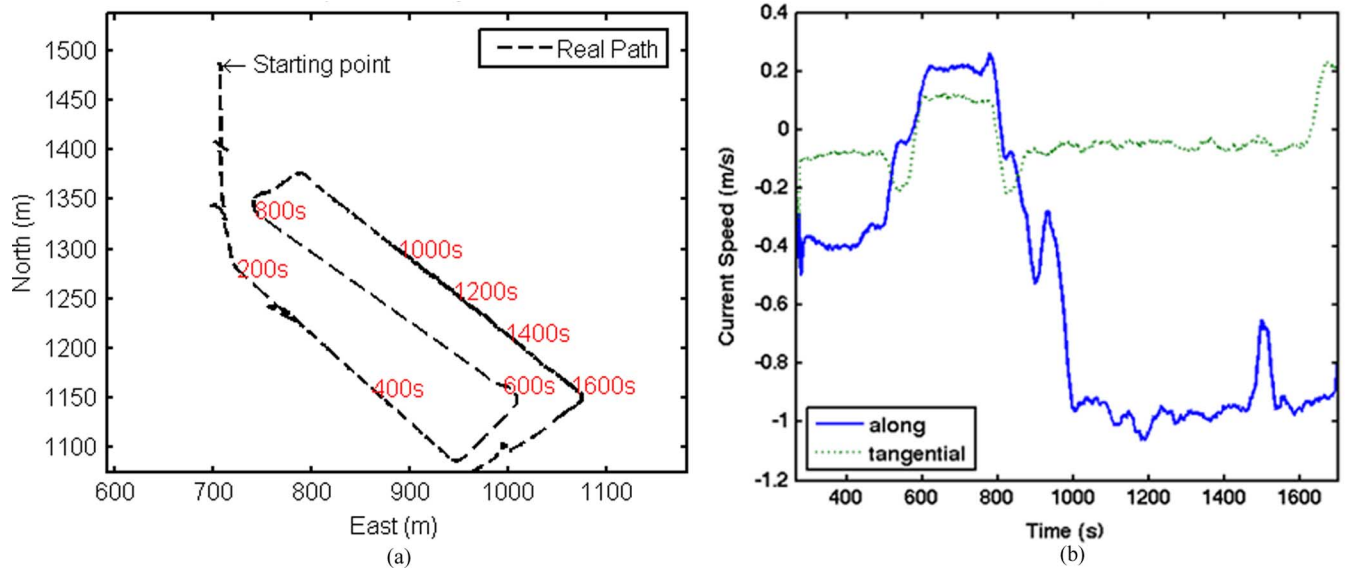


Fig. 10. Real path of the survey AUV during the field trial on the July 9, 2011, and the estimates of the ocean current in the AUV's body frame. The AUV encountered a strong ocean current stream from 1000th second onwards. (a) Survey AUV's executed path with time stamps of every 200 s. (b) Ocean current estimated in the AUV's body frame.

tidal current. The tidal current during the mission varied significantly in different mission legs. Fig. 10(a) shows the real trajectory of the survey AUV with time noted at every 200 s. The survey AUV was commanded to thrust at a constant level of 70% throughout the whole mission, which gives about 1.5-m/s relative speed with respect to the water. Since the observed displacement is not 300 m for every 200 s, it is clear that there was some tidal current slowing down or speeding up the AUV along its heading direction, in addition to the local nontidal current variations along the channel, as reported in [22] and estimated in Fig. 10(b): in the first leg (200–400 s), there was a mild current stream (about 0.5 kn) against the AUV's direction; in the second leg (600–800 s), the effect of ocean current was along the AUV's heading direction, thus it increased its effective speed; in the last leg (1000–1600 s), there was a strong current stream (up to 2 kn) slowing down

the AUV and caused it to move only about 100 m for every 200-s interval.

In Fig. 9(b), the fixed and circularly moving beacons perform poorly in correcting the positioning error of the supported survey AUV since the changes of relative aspect between the vehicles were small during the acoustic range updates. This causes the tidal current estimation in the state vector to become worse. The poor tidal current estimation, in turn, results in poor estimation of future positions. This feedback cycle escalates the growth of positioning error in the survey AUV. However, when the ocean current is almost zero and the survey AUV is moving in favor of the fixed beacon's location (600–800 s), the positioning error can be significantly reduced by its range updates. This observation further supports the claim that the location of the fixed beacon is one of the important factors that determines the performance of the beacon.

TABLE V
POSITIONING ERRORS INCURRED BY VARIOUS METHODS

Methods	DR	Fixed beacon	Circular Moving Beacon	DP	MDP
Max Errors (m)	553.6	241.0	521.7	79.8	104.0
% o.d.t.*	35.57	15.49	33.52	5.13	6.69
Average Errors (m)	343.7	54.7	233.4	19.1	26.5
% o.d.t.*	22.09	3.51	15.00	1.23	1.71

* over the distance traveled of ~ 1.5 km

Both DP and MDP methods keep the positioning error of the survey AUV fairly small throughout the mission, even though they were under the effect of varying tidal currents. This demonstrates the robustness of both DP and MDP methods in handling the environmental uncertainties. For the entire survey path of around 1.5 km, the survey AUV position error with both fixed and circularly moving beacons reaches a maximum of around 16% and 34%, while the DP and MDP methods yielded a maximum error well below 7%. The average errors accumulated over the entire mission were around 3.5% and 15% for both fixed and circularly moving beacons, and 1.2% and 1.7% for the DP and MDP methods. Detailed position error estimates using various methods are shown in Table V.

VII. CONCLUSION AND FUTURE WORK

We proposed two different algorithms of cooperative path planning for a beacon vehicle that utilizes acoustic range measurements to support one or more survey AUVs in minimizing their accumulated positioning error during underwater navigation. The algorithms plan the beacon vehicle's path around the survey AUVs such that when range information is exchanged, the position errors of the supported survey AUVs can be kept small. Simulation studies were conducted to compare the performance of the algorithms with two simpler methods where the beacon is fixed to a location or moving in a circular motion around the mission area. The positioning error due to DR was used as a benchmark.

The simulation results showed that the proposed algorithms kept the positioning errors of the supported survey AUVs small throughout the mission runs. The algorithms were shown to be robust in handling varying range update rate as well as environmental uncertainties. While the performance of the DP method was slightly better than the MDP method, the latter algorithm required much lower computational power during runtime. Although the fixed and circularly moving beacons were able to correct the positioning errors of the survey AUVs to a lesser extent than the DP and MDP algorithms, their performance was highly dependent on the initial location of the beacon. We conclude that cooperative path planning using a single moving beacon has the potential for minimizing the positioning errors of the supported survey AUVs.

Future research in cooperative path planning will focus on online learning of the planning policy to support single- as well as multiple-survey AUVs. The methods and results reported herein will serve as a benchmark for our future investigation on the cooperative positioning missions among a team of AUVs.

APPENDIX

ERROR ESTIMATE COVARIANCE DUE TO RANGE UPDATES

Let the state vector $\mathbf{x}$ represent the vehicle's position in 2-D space, that is, $\mathbf{x}_l = [x, y]_l^\top$. At time $l + 1$, the beacon vehicle's position is $\mathbf{x}_{l+1}^B = [x^B, y^B]_{l+1}^\top$. Let β_{l+1} be the angle formed by the line joining the beacon vehicle and survey AUV $\beta_{l+1} = \angle(\mathbf{x}_{l+1}^B - \mathbf{x}_{l+1})$, then the observation matrix with respect to the true position is

$$\mathbf{H}_{l+1} = [\cos \beta_{l+1} \quad \sin \beta_{l+1}]. \quad (35)$$

Assuming that the position error of the survey AUV can be described as an error ellipse, the error estimate covariance $\mathbf{P}_l$ can be written as

$$\mathbf{\Upsilon}_l \mathbf{P}_l \mathbf{\Upsilon}_l^\top = \mathbf{P}_l$$

$$\mathbf{\Upsilon}_l \begin{bmatrix} \epsilon_l^2 & 0 \\ 0 & \bar{\epsilon}_l^2 \end{bmatrix} \mathbf{\Upsilon}_l^\top = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}_l \quad (36)$$

where rotation matrix $\mathbf{\Upsilon}_l$ is formed by the angle of the minor axis θ_l , counterclockwise from the x -axis. $\sigma_{x,l}^2 = \epsilon_l^2$ and $\sigma_{y,l}^2 = \bar{\epsilon}_l^2$ denote the length of the minor and major axis at time step l after propagation.

Let the measurement error be $\sim N(0, \sigma_z^2)$, which includes the ranging error $\sim N(0, \sigma^2)$ and the position error of the beacon vehicle $\sim N(0, \sigma_B^2)$. We have $\sigma_z^2 = \sigma^2 + \sigma_B^2$. The innovation covariance is then derived as

$$\begin{aligned} \mathbf{S}_{l+1} &= \mathbf{H}_{l+1} \mathbf{P}_l \mathbf{H}_{l+1}^\top + \sigma_z^2 \\ &= \mathbf{H}_{l+1} \mathbf{\Upsilon}_l \mathbf{P}_l \mathbf{\Upsilon}_l^\top \mathbf{H}_{l+1}^\top + \sigma_z^2 \\ &= \epsilon_l^2 \cos^2(\beta_{l+1} - \theta_l) + \bar{\epsilon}_l^2 \sin^2(\beta_{l+1} - \theta_l) + \sigma_z^2. \end{aligned} \quad (37)$$

The Kalman gain is

$$\mathbf{K}_{l+1} = \mathbf{P}_l \mathbf{H}_{l+1}^\top \mathbf{S}_{l+1}^{-1} \quad (38)$$

and the error estimate covariance is updated as

$$\mathbf{P}_{l+1} = \mathbf{P}_l - \mathbf{K}_{l+1} \mathbf{H}_{l+1} \mathbf{P}_l. \quad (39)$$

$\mathbf{P}_{l+1}$ is a 2×2 symmetric matrix with the components in the upper triangle as

$$\begin{aligned} \mathbf{P}_{11,l+1} \mathbf{S}_{l+1} &= \epsilon_l^2 \bar{\epsilon}_l^2 \sin^2 \beta_{l+1} + (\epsilon_l^2 \cos^2 \theta_l + \bar{\epsilon}_l^2 \sin^2 \theta_l) \sigma_z^2 \\ \mathbf{P}_{12,l+1} \mathbf{S}_{l+1} &= -\frac{1}{2} \epsilon_l^2 \bar{\epsilon}_l^2 \sin(2\beta_{l+1}) + \frac{1}{2} (\epsilon_l^2 - \bar{\epsilon}_l^2) \sigma_z^2 \sin^2(2\theta_l) \\ \mathbf{P}_{22,l+1} \mathbf{S}_{l+1} &= \epsilon_l^2 \bar{\epsilon}_l^2 \cos^2 \beta_{l+1} + (\epsilon_l^2 \sin^2 \theta_l + \bar{\epsilon}_l^2 \cos^2 \theta_l) \sigma_z^2. \end{aligned} \quad (40)$$

The angle of the minor axis by $\mathbf{P}_{l+1}$ is θ_{l+1} . Thus

$$\tan(2\theta_{l+1}) = \frac{-\epsilon_l^2 \bar{\epsilon}_l^2 \sin(2\beta_{l+1}) + \sin(2\theta_l) (\epsilon_l^2 - \bar{\epsilon}_l^2) \sigma_z^2}{-\epsilon_l^2 \bar{\epsilon}_l^2 \cos(2\beta_{l+1}) + \cos(2\theta_l) (\epsilon_l^2 - \bar{\epsilon}_l^2) \sigma_z^2}. \quad (41)$$

With the assumption that $\sigma_z^2 \ll \epsilon_l^2 \leq \bar{\epsilon}_l^2$, we have $(\sigma_z^2 / \epsilon_l^2 \bar{\epsilon}_l^2) \rightarrow 0$

$$\lim_{\substack{\sigma_z^2 \rightarrow 0 \\ \epsilon_l^2 \bar{\epsilon}_l^2 \rightarrow 0}} \tan(2\theta_{l+1}) = \tan(2\beta_{l+1})$$

$$\lim_{\substack{\sigma_z^2 \rightarrow 0 \\ \epsilon_l^2 \bar{\epsilon}_l^2 \rightarrow 0}} \theta_{l+1} = \beta_{l+1}. \quad (42)$$

With $\gamma_l = \theta_{l+1} - \theta_l$, the minor and major axes of the error ellipse formed by $\mathbf{P}_{l+1}$ are

$$\lim_{\substack{\sigma_z^2 \rightarrow 0 \\ \epsilon_l^2 \bar{\epsilon}_l^2 \rightarrow 0}} \sigma_{x,l+1}^2 = \sigma_z^2$$

$$\lim_{\substack{\sigma_z^2 \rightarrow 0 \\ \epsilon_l^2 \bar{\epsilon}_l^2 \rightarrow 0}} \sigma_{y,l+1}^2 = \frac{\epsilon_l^2 \bar{\epsilon}_l^2}{\epsilon_l^2 \cos^2 \gamma_l + \bar{\epsilon}_l^2 \sin^2 \gamma_l}. \quad (43)$$

For a Kalman filter with an identity propagation matrix $\mathbf{F}$

$$\begin{aligned} \mathbf{P}_{l+1|l} &= \mathbf{F} \mathbf{P}_{l|l} \mathbf{F}^\top + \mathbf{Q} \\ &= \mathbf{P}_{l|l} + \mathbf{Q} \end{aligned} \quad (44)$$

where $\mathbf{Q} = \text{diag}(q, q)$, which includes the error growth $q = \alpha\tau$. The updated error estimate covariance forms an ellipse with the following.

- The direction of the minor axis (minimum error) θ_{l+1} is along the line joining the beacon and the survey AUV.
- The error in the minor axis has $\epsilon_{l+1}^2 = \sigma^2 + \sigma_B^2 + \alpha\tau$.
- The error in the major axis has

$$\bar{\epsilon}_{l+1}^2 = \frac{\epsilon_l^2 \bar{\epsilon}_l^2}{\epsilon_l^2 \cos^2 \gamma_l + \bar{\epsilon}_l^2 \sin^2 \gamma_l} + \alpha\tau$$

REFERENCES

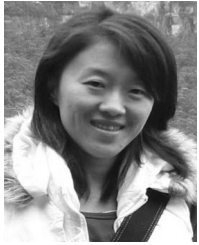
- [1] J. C. Kinsey, R. M. Eustice, and L. L. Whitcomb, "A survey of underwater vehicle navigation: Recent advances and new challenges," in *Proc. IFAC Conf. Manoeuvring Control Mar. Craft*, Lisbon, Portugal, Sep. 2006, invited paper.
- [2] A. Matos, N. Cruz, A. Martins, and F. L. Pereira, "Development and implementation of a low-cost LBL navigation system for an AUV," in *Proc. MTS/IEEE OCEANS Conf., Riding the Crest Into the 21st Century*, Sep. 1999, vol. 2, pp. 774–779.
- [3] P. Rigby, O. Pizarro, and S. Williams, "Towards geo-referenced AUV navigation through fusion of USBL and DVL measurements," in *Proc. OCEANS Conf.*, Sep. 2006, DOI: 10.1109/OCEANS.2006.306898.
- [4] A. Alcocer, P. Oliveira, and A. Pascoal, "Study and implementation of an EKF GIB-based underwater positioning system," *Control Eng. Practice*, vol. 15, no. 6, pp. 689–701, 2007.
- [5] G. Rui and M. Chitre, "Cooperative positioning using range-only measurements between two AUVs," in *Proc. IEEE OCEANS Conf.*, Sydney, Australia, May 2010, DOI: 10.1109/OCEANSSYD.2010.5603615.

- [6] A. Bahr, J. J. Leonard, and M. F. Fallon, "Cooperative localization for autonomous underwater vehicles," *Int. J. Robot. Res.*, vol. 28, no. 6, pp. 714–728, 2009.
- [7] J. C. Alleyne, "Position estimation from range only measurements," M.S. thesis, Dept. Mech. Eng., Naval Postgrad. Schl., Monterey, CA, USA, Sep. 2000.
- [8] M. F. Fallon, G. Papadopoulos, J. J. Leonard, and N. M. Patrikalakis, "Cooperative AUV navigation using a single maneuvering surface craft," *Int. J. Robot. Res.*, vol. 29, pp. 1461–1474, Oct. 2010.
- [9] T. L. Song, "Observability of target tracking with range-only measurements," *IEEE J. Ocean. Eng.*, vol. 24, no. 3, pp. 383–387, Jul. 1999.
- [10] J. Hartsfiel, "Single transponder range only navigation geometry (STRONG) applied to REMUS autonomous under water vehicles," M.S. thesis, Dept. Mech. Eng., Massachusetts Inst. Technol., Cambridge, MA, USA, 2005.
- [11] A. Gadre and D. Stilwell, "Toward underwater navigation based on range measurements from a single location," in *Proc. IEEE Int. Conf. Robot. Autom.*, New Orleans, LA, USA, May 2004, pp. 4472–4477.
- [12] S. E. Webster, R. M. Eustice, H. Singh, and L. L. Whitcomb, "Advances in single-beacon one-way-travel-time acoustic navigation for underwater vehicles," *Int. J. Robot. Res.*, vol. 31, no. 8, pp. 935–950, 2012.
- [13] M. F. Fallon, M. Kaess, H. Johannsson, and J. J. Leonard, "Efficient AUV navigation fusing acoustic ranging and side-scan sonar," in *Proc. IEEE Int. Conf. Robot. Autom.*, Shanghai, China, May 2011, pp. 2398–2405.
- [14] M. Chitre, I. Topor, and T. Koay, "The UNET-2 modem—An extensible tool for underwater networking research," in *Proc. OCEANS Conf.*, Yeosu, Korea, May 2012, DOI: 10.1109/OCEANS-Yeosu.2012.6263431.
- [15] M. Chitre, "Path planning for cooperative underwater range-only navigation using a single beacon," in *Proc. Int. Conf. Autonom. Intell. Syst.*, Jun. 2010, DOI: 10.1109/AIS.2010.5547044.
- [16] W. B. Powell, *Approximate Dynamic Programming: Solving the Curses of Dimensionality*, 1st ed. New York, NY, USA: Wiley, 2007, ch. 3, pp. 58–62.
- [17] Y. T. Tan and M. Chitre, "Single beacon cooperative path planning using cross-entropy method," in *Proc. IEEE/MTS OCEANS Conf.*, Sep. 2011.
- [18] P. T. De Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," *Ann. Oper. Res.*, vol. 134, pp. 19–67, 2005.
- [19] S. Mannor, R. Rubinstein, and Y. Gat, "The cross entropy method for fast policy search," in *Proc. Int. Conf. Mach. Learn.*, Aug. 2003, pp. 512–519.
- [20] T. B. Koay, Y. T. Tan, Y. H. Eng, R. Gao, M. Chitre, J. L. Chew, N. Chandhavarkar, R. Khan, T. Taher, and J. Koh, "STARFISH—A small team of autonomous robotics fish," *Indian J. Geo-Mar. Sci.*, vol. 40, pp. 157–167, Apr. 2011.
- [21] Eubank and L. Randy, *A Kalman Filter Primer*. Boca Raton, FL, USA: Chapman & Hall/CRC Press, 2006, ch. 8.
- [22] J. Wei, H. Zheng, H. Chen, B. H. Ooi, M. H. Dao, W. Cho, P. M. Rizzoli, P. Tkachik, and N. M. Patrikalakis, "Multi-layer model simulation and data assimilation in the Serangoon Harbor of Singapore," in *Proc. Int. Offshore (Ocean) Polar Eng. Conf.*, Jun. 2010 [Online]. Available: <http://hdl.handle.net/1721.1/64464>



Yew Teck Tan received the B.S. degree in computer science from the University Tunku Abdul Rahman (UTAR), Perak, Malaysia, in 2006, developing path planner for mobile robotic systems, and the M.Eng. degree from the National University of Singapore (NUS), Singapore, in 2009, where he is currently working toward the Ph.D. degree under the Singapore/MIT Alliance for Research and Technology (SMART) fellowship.

Currently, he is involved in developing a command and control system for the autonomous underwater vehicles at NUS. His research interests include intelligent and autonomous systems, robotics, multiagent multirobot systems, and biologically inspired computing.



Rui Gao received the B.S. degree in electrical and computer engineering and the M.Eng. degree for the image and signal processing for the classification of dolphin vocalization from the National University of Singapore (NUS), Singapore, in 2007 and 2011, respectively.

She is currently working at the Acoustic Research Laboratory (ARL), NUS, as a Research Engineer. Her research interests include information fusion, modeling, multi-AUV positioning, and navigation.



Mandar Chitre received the B.Eng. and M.Eng. degrees in electrical engineering from the National University of Singapore (NUS), Singapore, in 1997 and 2000, respectively, the M.Sc. degree in bioinformatics from the Nanyang Technological University (NTU), Singapore, in 2004, and the Ph.D. degree from NUS in 2006.

From 1997 to 1998, he worked with the ARL, NUS. From 1998 to 2002, he headed the technology division of a regional telecommunications solutions company. In 2003, he rejoined ARL, initially as

the Deputy Head (Research) and is now the Head of the laboratory. He also holds a joint appointment with the Department of Electrical and Computer Engineering at NUS as an Assistant Professor. His current research interests are underwater communications, autonomous underwater vehicles, model-based signal processing, and modeling of complex dynamic systems.

Dr. Chitre has served on the technical program committees of the IEEE OCEANS Conference, the International Conference on UnderWater Networks and Systems (WUWNet), the Defence Technology Asia Conference (DTA), the Offshore Technology Conference (OTC), and the IEEE International Conference on Communication Systems (ICCS), and has served as reviewer for numerous international journals. He was the Chairman of the Student Poster Committee for the 2006 IEEE OCEANS Conference in Singapore, and the Chairman for the 2013 IEEE Singapore AUV Challenge. He is currently the IEEE Ocean Engineering Society Technology Committee Co-Chair of Underwater Communication, Navigation, and Positioning.