Single Beacon Cooperative Path Planning Using Cross-Entropy Method

Tan Yew Teck
Department of Electrical and Computer Engineering
National University of Singapore
Singapore, 119222.

Mandar Chitre
ARL, Tropical Marine Science Institute and
Department of Electrical and Computer Engineering
National University of Singapore
Singapore 119227.

Abstract—**This paper focuses on path planning problem for a single beacon vehicle supporting a team of autonomous underwater vehicles (AUVs) performing surveying missions. Underwater navigation is a challenging problem due to the absence of GPS signal. The positioning error grows with time even though AUVs nowadays are equipped with onboard navigational sensors like compass for dead reckoning. One way to minimize this error is to have a moving beacon vehicle equipped with high accuracy navigational sensors to transmit its position acoustically at strategic locations to other AUVs. When it is received, the AUVs can fuse this data with the range measured from the travel time of acoustic transmission to better estimate their own positions and minimize the error. In this work, we address the beacon vehicle's path planning problem which takes into account the position errors being accumulated by the supported survey AUVs. The resultant path will position the beacon AUV at the strategic locations during the acoustic signal transmission. We formulate the problem within a Markov Decision Process (MDP) framework where the path planning policy is being learned through Cross-Entropy (CE) method. We show that the resultant planned path using the policy learned is able to keep the position error of the survey AUVs bounded throughout the simulated runs.**

I. Introduction

Recent advancement in the development of Autonomous Underwater Vehicles (AUV) and underwater communication have made acoustic ranging a viable option to be used for underwater cooperative positioning and/or localization. The idea of AUV cooperative localization is to have an AUV with good quality positioning information, termed beacon AUV, to transmit its position and range information to other AUVs, termed survey AUVs, within its communication range during navigation. Generally, the beacon vehicle is equipped with high accuracy sensors that is able to estimate its position with minimum errors. The range information between the AUVs can then be fused with the data obtained from onboard sensors like compass, Doppler Velocity Log (DVL) and Inertial Navigation System (INS) in the survey AUVs to reduce the positioning error during underwater navigation [1].

Since GPS signal is not available underwater, AUVs rely on the on-board sensors such as compass, DVL and INS for their position estimation. However, dead reckoning upon these sensors suffers from unbounded error growth due to the integration involved. Although this problem can be avoided by having the AUV surface and obtain a GPS fix, or deploying fixed beacons around the mission area, it may put the AUV and the beacons' safety in jeopardy especially around busy shipping channels. Besides that, in an AUV team operation which has the advantages of simultaneous monitoring and surveying, it is not cost effective to have every AUV carry expensive DVL or INS that can provide accurate position estimate. With the development of underwater acoustic modem which is capable of measuring the time of travel of acoustic signals among the AUVs, having a single beacon AUV that is equipped with accurate position estimate to support other AUVs within its acoustic range seems an attractive option.

However, in order for the single beacon range-only cooperative localization to perform well, the relative motion between the beacon and the survey AUVs should vary as close to 90 deg as possible for every consecutive range information transmission. Although several researchers [1], [2], [3], and [4] acknowledge this operation requirement, the problem of determining the optimal maneuver for the beacon AUV to obtain the required relative motion only received little attention. For instance, [3] proposed having the Communication and Navigation Aid (CNA), as they termed the beacon AUV, to maneuver in zigzagging pattern or encircling the AUV so that the CNA's path fully observable, while in the work [2] the authors even concluded 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 Side-scan sonar. More recently, the author in [5] attempted to solve the optimal beacon AUV's maneuver by developing a path planning algorithm that takes into account and minimizes the positioning errors being accumulated by other AUVs. The simulation results showed that given the required maneuvering speed as well as consistent range information, the beacon AUV is able to bound the positioning errors accumulated by other AUVs within a few meters. Extending the work mentioned in [5], we cast the beacon AUV's path planning problem within the framework of Markov Decision Process (MDP) and learn the policy using the Cross-Entropy method. This approach allows the beacon AUV to "learn" how to position itself at the locations where the acoustic range signal will help to minimize the survey AUVs' position error.

In this paper, we first introduce the problem in detail, followed by the derivation of the path planning algorithm within MDP framework. After that, we apply the Cross-

Entropy method to learn the beacon AUV's path planning policy. We show that the beacon AUV's path generated by the learned policy is able to keep the positioning errors of other AUVs bounded throughout the simulated runs.

II. Problem Statement and Formulation

A. Position Error Estimation

Let $t \in \{0, ..., T\}$ represent the time step and we assume that the beacon vehicle is able to transmit a beacon signal every nt seconds. This transmission enables all the survey AUVs within acoustic range of the beacon vehicle to estimate their range from the beacon vehicle using the propagation delay of the signal. Let M be the number of survey AUVs supported by the beacon AUV and are indexed by $j \in \{1...M\}$. Let x_t^j represent the position of AUV j and x_t^{B} represent the position of beacon vehicle at time step t. At every time t, the range between the beacon AUV and each of the survey AUVs is estimated by $\hat{R}_t^j$, where its error is modeled by a zero-mean Gaussian random variable with variance σ^2:

$$\hat{R}_t^j = \mathcal{N}(|x_t^j - x_t^{\mathrm{B}}|, \sigma^2) \tag{1}$$

For the survey AUV j at time step t, we model the error in position estimation as a 2-dimensional 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. After a beacon signal transmission at time $t+1$, the error is minimum along the line joining the beacon and the survey vehicle:

$$\theta_{t+1}^j = \angle(x_{t+1}^j - x_{t+1}^{\mathrm{B}}) \tag{2}$$

$$\epsilon_{t+1}^j = \sigma \tag{3}$$

The error will grow linearly in the tangential direction if we assume velocity estimates are used for dead reckoning [5] and is given by:

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

where $\gamma_t^j = \theta_{t+1}^j - \theta_t^j$ and α is a constant of proportionality. Since x_t^j are usually pre-determined before mission, we want to plan $\{x_t^{\mathrm{B}} \dots x_T^{\mathrm{B}}\}$, the beacon AUV's path such that sum square of both (3) and (4) are minimized with respect to all the survey AUVs throughout the mission:

$$\{x_t^{\mathrm{B}} \dots x_T^{\mathrm{B}}\} = \arg\min_x \sum_{j,t} [(\epsilon_t^j)^2 + (\bar{\epsilon}_t^j)^2] \tag{5}$$

We assume that the starting position of both beacon and survey AUVs are known, and with accuracy of ϵ_0 in all directions:

$$\epsilon_0^j = \bar{\epsilon}_0^j = \epsilon_0 \tag{6}$$

$$\theta_0^j = 0 \tag{7}$$

III. Markov Decision Process formulation

In this section, we show the formulation of the beacon AUV's path planning problem within the MDP framework. Generally, MDP is defined by four main components: the state and action sets, the state transition probability matrix and the rewards/cost function. From section II-A, $\hat{R}_t^j$ is the estimated distance between beacon and survey AUV, ϕ_t^{B} represent the beacon AUV's current bearing at time t and ϕ_{t+1}^j be the survey AUV's bearing at time $t+1$ respectively, our state set is defined as a tuple: $z_t = \{\theta_t^j, \hat{R}_t^j, \phi_t^{\mathrm{B}},\ \phi_{t+1}^j\}$. 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 (5) for every time step t. This means having γ_t^j in (4) to be as close as possible to 90 deg. Thus, the ability of beacon AUV 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 the $\hat{R}_t^j$ and θ_t^j can be obtained from the acoustic ranging and communication between the AUVs while ϕ_{t+1}^j is usually pre-planned before the mission. The action a_t is the turning angle from the beacon AUV's current bearing (ϕ_t^{B}), $|a_t| \leq \dot{\phi}_{\max}^{\mathrm{B}}$ where $\dot{\phi}_{\max}^{\mathrm{B}}$ is the maximum turning angle per unit time t seconds achievable by the beacon AUV.

At every time t, after a_t is selected, the corresponding 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 (5). In MDP, a 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, z_t into N_z states and require that for all z rows in $\mathbf{P}_{za}$, the sum of each z^{th} row is equals to 1. We then define 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}.

In our case, this would translate into the probability of choosing a particular turning angle from the beacon AUV's current bearing (termed as desired heading in the rest of the paper) at time $t+1$, given the beacon AUV's current bearing, survey AUV's next heading as well as distance and relative angle between the AUVs. As a result, one can easily see that we can solve the cost minimization problem as the beacon AUV's path planning policy.

IV. Policy Learning using Cross-Entropy Method

A. Cross Entropy Method

In this section, we briefly introduce the Cross-Entropy (CE) Method and its application in learning the policy in MDP. The CE method was initially introduced for estimating the probability of rare events in complex stochastic networks [6]. 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 problem, once defined, can be tackled efficiently by adaptive algorithms. We refer the interested readers to [6] and [7] for detailed development and formulation of the CE method.

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 III. Let η^* denotes the minimum of C on χ, $\eta \in \mathbb{R}^+$:

$$\eta^* = \min_{x\in\chi} C(x) \tag{8}$$

The deterministic problem shown in (8) can be randomized by defining a family of probability density functions (pdf's) $\{f(x,p), p \in V\}$ on the set χ, with p being the pdf's (discrete) parameter. By ASP, we can associate with (8) the following estimation problem:

$$l(\eta) = \mathbb{P}_u(C(\mathbf{x}) \leq \eta) = \mathbb{E}_u I_{\{C(\mathbf{x})\leq\eta\}} \tag{9}$$

where $\mathbb{P}_u$ is the probability measure under which the random vector $\mathbf{x}$ has pdf $f(x,u)$, for some $u \in V$, and $I\{\cdot\}$ is the indicator function. The association comes from the fact that the probability $\mathbb{P}_u(C(\mathbf{x}) \leq \eta)$ will be very small (rare event) when η is close to η^*. By 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^*). With p_0 initialized to u, the tuple $(\hat{\eta}_n, \hat{p}_n)$ can be updated iteratively by:

1) Let η_n be the (1-ρ)-quantile of $C(\mathbf{x})$ under p_{n-1}. An estimate of η_n, denoted $\hat{\eta}_n$, can be obtained by generating a set of N sample random vectors from $f(x, p_{n-1})$ and assigning $\hat{\eta}_n$ as the (1-ρ)-quantile of $C(\mathbf{x}_k)$ where $C(\mathbf{x}_k) \in \{C(\mathbf{x}_1) \leq ... \leq C(\mathbf{x}_N)\}$.
2) With fixed $\hat{\eta}_n$ and p_{n-1}, the estimate of p_n, denoted $\hat{p_n}$, can be derived from [6]:

$$\hat{p_n} = \arg\max_p \frac{1}{N}\sum_{k=1}^{N} I_{\{C(\mathbf{x}_k)\leq\hat{\eta}_n\}} \ln f(\mathbf{x}, \mathbf{P}_{za}) \tag{10}$$

where the $\mathbf{x}_k$ are generated from $f(x, p_{n-1})$.

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

1) Generation of 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) Selecting 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 AUV's Path Planning Policy Learning using CE method

In order to apply CE method for learning the path planning policy, we must specify the two important phases stated before, which in our case are: (a) how to generate the sample beacon path, and (b) 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 τ steps, we can generate a set of beacon paths with the same path length via Markov process with the policy matrix $\mathbf{P}_{za}$. Let N be the total number of paths generated in the set, each beacon path, $\mathbf{x}_k$, $k \in \{0 \dots N\}$, consists of sequence of state-action pair, $x_k = (z_0, a_0, ..., z_\tau, a_\tau)$. The cost of each resultant beacon AUV's path can be estimated through through (3) and (4) as shown in section II-A.

Let $C(x_k)$ represent the total cost of path $\mathbf{x}_k$, and N be the total number of paths generated for policy learning at every iteration, we sort the paths' cost in increasing order and evaluate the (1-ρ)-quantile η. Once the η is selected, the policy matrix can be updated by solving (10) to obtain the formula (see [6], [7]):

$$p_{za} = \frac{\sum_{k=1}^{N} I_{\{C(\mathbf{x}_k)\leq\eta\}} I_{\{\mathbf{x}_k\in\chi_{za}\}}}{\sum_{k=1}^{N} I_{\{C(\mathbf{x}_k)\leq\eta\}} I_{\{\mathbf{x}_k\in\chi_{z}\}}} \tag{11}$$

where $C(\mathbf{x}_k) \leq \eta$ means 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 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 convergence to an acceptable limit. Detailed steps are shown in Algorithm 1.

```
Algorithm 1: Policy Learning
Require: $\mathbf{P}_{za}$ uniformly initialized with $(1/ \mid N_a \mid)$
  - Let $\eta_0$ = 0, set $n = 0$
  repeat
    - Set $n = n + 1$

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

      repeat
        - Generate a set of random surveying paths with path length of $\tau$ steps.
        - 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 $i$ = $\tau$.
        - Output the total cost $(C(x))$ of the trajectory $(z_0, a_0, ..., z_\tau, a_\tau)$.
      until N trajectories
      - Sort the $N$ scores in descending order, take $\eta_n$ as the (1 -$\rho$)-percentile of the score set.
      - Update the parameter matrix $\mathbf{P}_{za}$ according to equation (11).
    end for
  until $|\eta_n - \eta_{n-1}| \leq \psi$
```

Instead of updating the policy matrix $\mathbf{P}_{za}$ directly with equation (11), we apply a simple smoothing filter:

$$\hat{p}_{za,n} = \mu\tilde{p}_{za,n} + (1-\mu)\hat{p}_{za,n-1} \tag{12}$$

where $\tilde{p}_{za,n}$ is the solution of (11) and μ is the smoothing parameter with 0.7< μ <1. The filter serves two purposes: (i) smoothing the policy matrix update, (ii) avoiding the $\hat{p}_{za,n}$ from becoming zero especially during the initial stage of learning process. This is crucial as to avoid the learning algorithm from falling into a local minima and converge to a wrong solution.

V. Policy Learning and Simulation Setup

A. Policy Learning Setup

The learning algorithm shown in section IV-B was used with the the setup shown in Table I.

TABLE I
PARAMETER FOR POLICY LEARNING

Parameter	Value
τ	20s
σ	1 m
$\dot{\phi}^{\mathrm{B}}_{max}$	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

In our approach, we do not need to discretize our map into grid map since we are only concerned with the relative angle between the AUVs. However, we do discretize the angle between the AUVs and the AUVs' bearing into 36 states each representing an angle section of 10 deg spanning from 0 $\sim$ 360 deg. The AUVs are allowed to navigate between 100m and 1000m within each other, the distance is discretized into 3 states with first 2 zone having 300m each while the last zone spanning 400m. Any distance closer than 100m or more than 1000m apart will be given a heavy penalty that will contribute to the accumulated error. This is necessary to prevent the AUVs from colliding if they are too close together while keeping the AUVs within the communication range (which in our case, assumed to be 1000m). The maximum turning angle of the AUV is 40 deg and is discretized into 8 action states. The State and Action Space formulated for the policy learning are summarized in Table II.

TABLE II
STATE AND ACTION SPACE DISCRETIZATION

State Space, N_z	Number of States
Beacon AUV's current bearing	36
Surverying 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 AUV's desired turning angle	8

B. Simulation Setup

1) Supporting Single Survey AUV: The policy matrix learned using the CE method is used for simulation to determine its performance in keeping the accumulated position errors of supported survey AUVs bounded. In the first simulation scenario, a survey AUV was given a lawn-mower mission surveying an area of 500m by 700m as shown in Fig. 1. The survey AUVs' paths are pre-planned and are shared with the beacon vehicle. With this information, the beacon vehicle plans its path iteratively using the policy matrix until the survey AUVs' missions have completed. During the simulation, all the vehicles are assumed to be moving at the speed of 1.5 m/s.

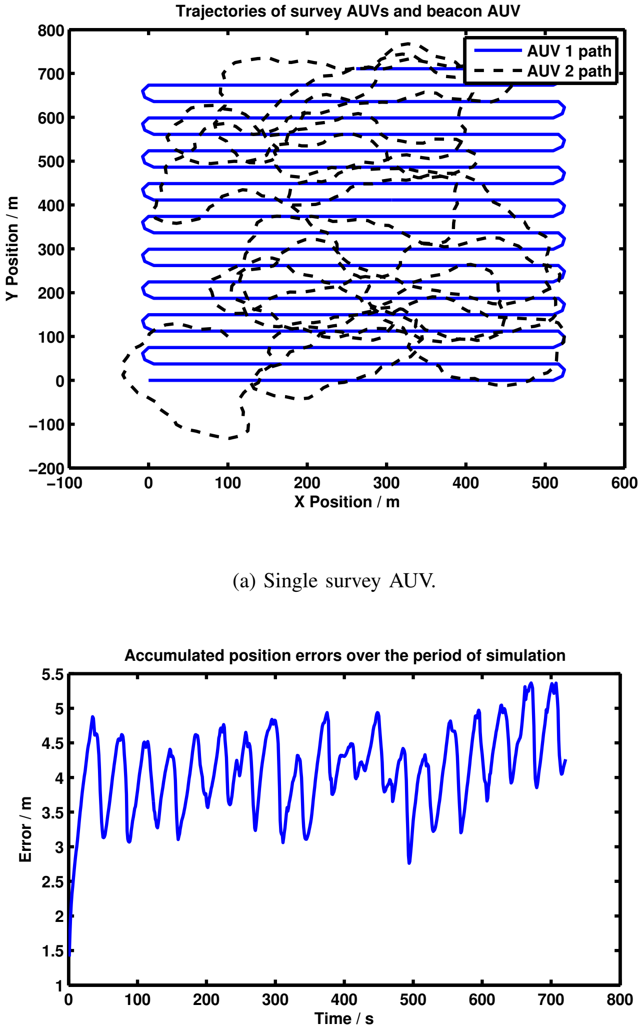
2) Supporting multiple Survey AUVs: In the second simulation scenario, we looked into having a single beacon AUV to support multiple survey AUVs. Two AUVs are put into a surveying mission where they were required to navigate in a lawn-mower pattern adjacent to each other as shown in Fig. 2 with an area of around 400m by 700m. At every time step t during the simulation, the beacon AUV generates one desired heading with respect to each of the survey AUVs using the same policy matrix, $\{A^{\mathrm{B}}_{j=1}, A^{\mathrm{B}}_{j=2}\}$. Since choosing one of the desired headings may reduce the accumulated error of one survey AUV while increasing the other, care has to be taken while making the final decision.

Three different methods were studied to explore the best strategy for the beacon AUV in deciding the desired heading during the course of supporting the survey AUVs:

1) *S-1*: A simple strategy is to choose the beacon AUV's desired heading that will favor the survey AUV whose current accumulated error is the highest. Since the position errors for each of the survey AUVs can be estimated after every range information transmission, the beacon AUV can simply decide on the desired heading according to the survey AUV with higher accumulated error.
2) *S-2*: One of the factors that affects the beacon AUV's capability in achieving the maximum relative angle with respect to the survey AUVs is to maintain close distance with all the vehicles it is supporting. However, this is impossible for the case of multiple survey AUVs where during the surveying mission, survey AUVs may navigate far apart from each other. In this strategy, we have the beacon AUV choose the desired heading that will navigate it in the proximity of the centroid location among the survey AUVs.
3) *S-3*: Instead of generating one desired heading with respect to each of the survey AUV, the sum of square of both the relevant rows of the policy matrix was used as the beacon AUV's action space. This results in only a single desired heading being produced at every step for the beacon AUV.

VI. Simulation Results

The result of the simulation for beacon AUV supporting single survey AUV is shown in Fig. 1 along with the accumulated position error over the period of the simulation. Without the supporting beacon AUV, the position error of the survey AUV is expected to grow linearly without bound. However, it can be seen in Fig. 1(b) that the error is bounded at around 5m throughout the simulation. We also observed that the beacon AUV position itself within the mission area during its course of supporting the survey AUV. The beacon AUV seems to have “learned” that by keeping a close distance to the survey AUV, the chance for it to achieve maximum change in relative angle with respect to the survey AUV is higher.

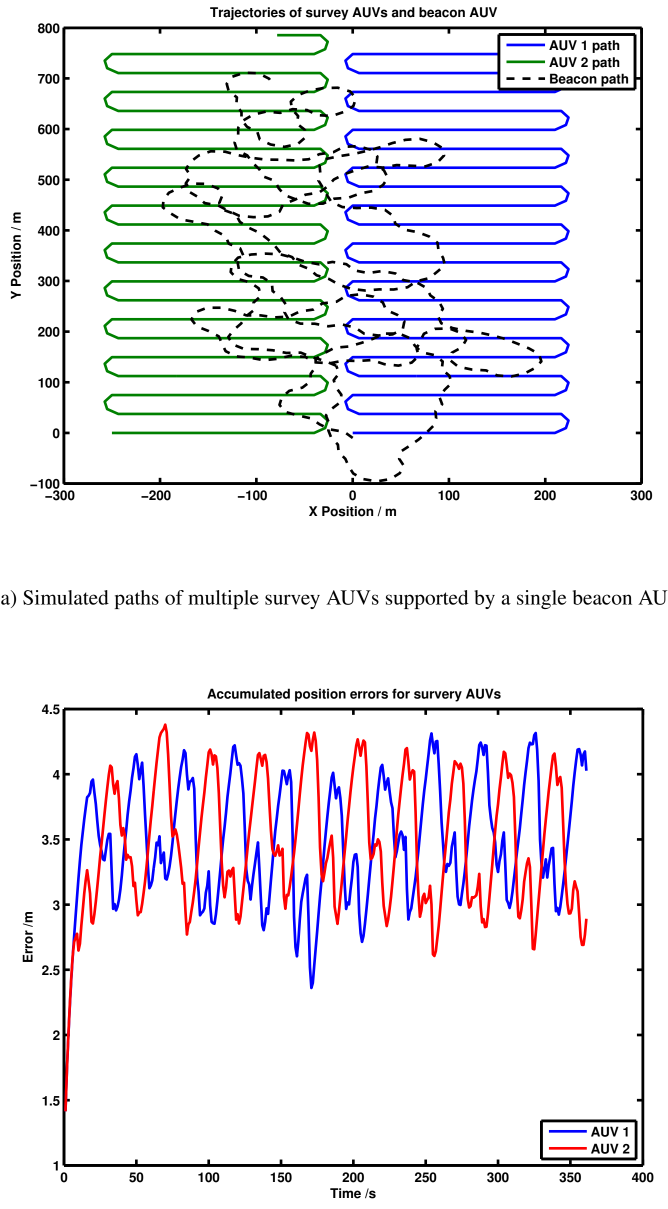


(a) Single survey AUV.

(b) Accumulated Position Error for single survey AUV.

Fig. 1. Simulation results showing a single beacon vehicle supporting single survey AUVs.

In the case of beacon AUV supporting multiple survey AUVs, the results from 10 simulated runs using the different strategies are summarized in Table III. We are interested in comparing the Root-Mean-Square (RMS) error as well as the Maximum (Max) error accumulated by both the survey AUVs. From the results, *S-2* and *S-3* performed slightly better compared to *S-1*. This is because choosing the desired heading that favors the survey AUV (AUV 1) with the highest accumulated error may lead the beacon AUV to move away from the other survey AUV (AUV 2) and reduce the relative angle that it can achieve with AUV 2. Although *S-3* managed to keep the RMS error as low as *S-2*, it has higher Max error on both the survey AUVs. However, *S-3* is preferable since it simplifies the decision making process of the beacon AUV. Overall, the accumulated error for both the survey AUVs are bounded throughout the simulated runs. A sample result using *S-3* is also shown in Fig. 2.



(a) Simulated paths of multiple survey AUVs supported by a single beacon AUV.

(b) Accumulated Position Error for multiple survey AUVs.

Fig. 2. Simulation results showing a single beacon vehicle supporting multiple survey AUVs using *S-3*.

TABLE III
SIMULATION RESULTS FOR 2 AUV USING DIFFERENT DECISION STRATEGIES

	RMS Error (m)		Max Error (m)	
Strategy	AUV1	AUV2	AUV1	AUV2
S-1	4.6	4.0	8.8	6.2
S-2	3.5	3.6	4.8	4.9
S-3	3.5	3.6	5.0	5.2

VII. Discussion and Future Work

In this paper, we presented a path planning problem for a single beacon vehicle supporting survey AUVs in underwater localization via acoustic ranging. The path planning problem was formulated as a MDP problem and its policy was learned by using the CE method. The main purpose of this work was to minimize the position error accumulated by survey AUVs which use only velocity estimates for dead reckoning during surveying missions. Simulation studies showed that the accumulated position errors of survey AUVs could be kept bounded throughout the simulated runs as long as acoustic ranging information was available at particular relative angles between the survey and beacon AUVs. These results are comparable with those mentioned in [5] for the cases where the AUVs were assumed to be navigating at the speed of 1.5 m/s. However, once the policy matrix is trained, our approach has significant advantage in terms of computational complexity with respect to the number of survey AUVs being supported by a single beacon AUV. In [5], the computational load for generating an optimal path using L-level look-ahead strategy is $\mathcal{O}(TN_a^{L+1}M)$ where M is the number of survey AUVs being supported. Whereas in our case, the computational load is independent of number of action states and grows only linearly with the number of survey AUVs, $\mathcal{O}(TM)$.

Although simulation studies show the accumulated error was bounded by using a single beacon AUV supporting multiple survey AUVs, it may not be the case in the real environment as the error as well as the motion model of the AUVs may be different due to environmental noise like sea current. Besides that, the acoustic ranging information may not be available at the predictable interval mentioned in the paper due to the loss in acoustic pings, or data corruption. These potential issues will be taken into account when we implement and test the algorithm in field trials with the STARFISH AUV [8] in the future. Currently, the policy matrix trained using the CE method is based on the sole assumption that the AUVs are traveling at 1.5 m/s. In the next version, we would like to improve the robustness of the algorithm to handle beacon or survey AUVs traveling at different speeds.

References

[1] Gao Rui and M. Chitre, "Cooperative positioning using range-only measurements between two AUVs", in *OCEANS 2010 IEEE - Sydney*, may 2010, pp. 1 –6.

[2] M. F. Fallon, M. Kaess, H. Johannsson, and J. J. Leonard, "Efficient AUV navigation fusing acoustic ranging and side-scan sonar", in *IEEE International Conference on Robotics and Automation (ICRA),Shanghai, China*, May 2011.

[3] Maurice F Fallon, Georgios Papadopoulos, John J Leonard, and Nicholas M Patrikalakis, "Cooperative AUV Navigation using a Single Maneuvering Surface Craft", *The International Journal of Robotics Research*.

[4] A.S. Gadre and D.J. Stilwell, "Toward underwater navigation based on range measurements from a single location", in *Robotics and Automation, 2004. Proceedings. ICRA '04. 2004 IEEE International Conference on*, april-1 may 2004, vol. 5, pp. 4472 – 4477 Vol.5.

[5] M. Chitre, "Path planning for cooperative underwater range-only navigation using a single beacon", in *Autonomous and Intelligent Systems (AIS), 2010 International Conference on*, 2010, pp. 1 –6.

[6] Pieter tjerk De Boer, Dirk P. Kroese, Shie Mannor, and Reuven Y. Rubinstein, "A tutorial on the cross-entropy method", *Annals of Operations Research*, vol. 134, pp. 19–67.

[7] Shie Mannor, Reuven Rubinstein, and Yohai Gat, "The cross entropy method for fast policy search", in *In International Conference on Machine Learning*. 2003, pp. 512–519, Morgan Kaufmann.

[8] 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", in *3rd International Conference on Underwater System Technology: Theory and Applications 2010, (Cyberjaya, Malaysia)*, Nov 2011.