A Stochastic Controller for Tracking the Periodic Movement of Marine Animals

Kevin Smith[1], Jerry Hsiung[2],
Christopher G. Lowe[3] and Christopher M. Clark[4]

Abstract— **This electronic document is a live template. The various components of your paper [title, text, heads, etc.] are already defined on the style sheet, as illustrated by the portions given in this document.**

I. Introduction

II. Background

III. Modeling

Previous work in fish tracking by an autonomous underwater vehicle (AUV) has been limited to tracking a single individual. However, to monitor entire populations by AUV tracking, it is not feasible to pair a large number of AUVs with the members of a representative sample of the population. Instead, a smaller number of AUVs may track the population by statistical means under a stochastic control.

Stochastic tracking consists of three steps. First, the problem is simplified by discretizing the space of possible fish locations into a finite graph, so that the fish's location is represented by a node and the fish's movement is represented by an edge. Second, data in the robot's memory of past fish locations is used to generate a *transition model*, which predicts future distributions of fish population density over the graph. Finally, a Markov process to is constructed to control the AUVs' positions on the graph, minimizing the difference between the population densities of the robots and fish at each node and time.

A. Discretizing the Space

We will represent the fish's location by Cartesian coordinates in the plane of the surface of the water, with the origin set to some latitude / longitude coordinate O. Latitude / longitude coordinates of the fish measured during the track are mapped to this plane. Let $r[t]$ denote the x-coordinate (distance east of O) and y-coordinate (distance north of O) of the fish at time $t > 0$.

Since the fish's location cannot be determined at regular intervals, the times for which $r[t]$ is defined also fail to occur at regular intervals. Such an uneven sampling is inconvenient for modeling the fish's location. Interpolation is used to generate an approximation $R[t]$ of the fish's location, defined for any $t = kT_{\text{sample}}$, where k is a nonnegative integer. For convenience, we may use the notation $R[k] = R[kT_{\text{sample}}]$ to index the fish's location by integer k.

This work simplifies the tracking problem by discretizing the plane of all possible fish coordinates into N discrete locations, known as *nodes*. These nodes may be chosen by any process, such as dividing the region into a grid or clustering historical data to calculate N centroids. While the nodes used in this work were stationary, this is not a requirement. The set of nodes will be denoted by V, and the physical location of node i at time t will be represented by the function $V_i(t)$.

The path of a robot is confined to certain *edges* between nodes. An edge is an ordered pair (i, j), with $i, j \in V$, and for each edge a physical path $e(i, j)$ exists that the robot is to follow when moving from node i to node j. There need not exist an edge between any pair of nodes, as in some cases, it may not make sense for the robot to move between such nodes, particularly if finding an obstacle-free path between the nodes would be difficult. If E is the set of all edges, then the directed graph (V, E) is the *location graph*.

B. Transition Model

Since many species of fish show periodic patterns of movement, analysis of previous fish locations may predict future distributions of the fish population over the location graph. These predictions take the form of a *transition model*, which consists of a time-inhomogeneous transition matrix $T[k]$ and a population density vector $\mathbf{p}[k]$. If $N = |V|$, then $T[k]$ is an $N \times N$ left stochastic matrix in which the j,i-th entry indicates the probability of the fish at node i transitioning to node j at time k. The population density $\mathbf{p}[k]$ is a stochastic column vector in which the i-th entry is proportional to some measure of the density of the fish population at node i and time k. The transition model can either be constructed by fitting a transition matrix to previous fish data and calculating a corresponding population density vector, or a population density vector can be fit to previous data and a corresponding transition matrix be extrapolated.

[1]Kevin Smith is with the Department of Physics, Harvey Mudd College, Claremont, CA 91711 `ksmith@g.hmc.edu`

[2]Jerry Hsiung is with the Department of Computer Science, Harvey Mudd College, Dayton, Claremont, CA 91711 `shiung@g.hmc.edu`

[3]Christopher G. Lowe is with the Department of Biological Sciences, California State University Long Beach, Long Beach, CA 90840 `clowe@csulb.edu`

[4]Christopher M. Clark is with the Department of Engineering, Harvey Mudd College, Dayton, Claremont, CA 91711 `clark@hmc.edu`

C. Transition Matrix Approach

A transition model may be constructed by generating a transition matrix from previous data. Once the fish's location $R[k]$ is discretized to the N nodes of the graph, we may represent the fish's movement between each time step as $d_i[k]$, defined recursively as follows:

$$d_i[k] = \begin{cases} R[k+1] & R[k] = i \\ d_i[k-1] & R[k] \neq i, k > 0 \\ i & R[k] \neq i, k = 0 \end{cases}$$

Informally, $d_i[k]$ represents the node that the fish will occupy at time step $k+1$ if it's location at time step k is node i. Of course, we only observe the fish's transition from one source node per time step. If $R[k] = i$ and $R[k+1] = j$, then we may easily choose $d_i[k] = j$; however, we are left to guess where the fish *would* have gone from some other node $x \neq i$ to choose the value of $d_x[k]$. The above definition makes the guess that, if $R[n+1] = y$ for the largest $n < k$ at which $R[n] = x$, then $d_x[k] = y$. An alternative definition of $d_i[k]$ may make the guess that $d_i[k] = R[k+1]$ for any i, though such a definition may overemphasize the current location of the fish, at the expense of losing information about the fish's path.

From $d_i[k]$, the transition matrix is calculated as follows:

$$T[k]_{i,j} = \sum_{m=0}^{L} f(m|k, \sigma/T_{\text{sample}}) \langle d_j[m] = i \rangle,$$

where L is the largest value such that $d_i[L]$ is defined, $f(x|\mu,\sigma)$ is a discrete normal distribution (defined for integers 0 to $L-1$ with a total sum of 1), and $\langle P \rangle$ is 1 if the proposition P is true and 0 otherwise. The population density vector can then be predicted as

$$\mathbf{p}[k] = \left(\prod_{m=0}^{k-1} T[m] \right) \mathbf{p}[0],$$

where $\mathbf{p}[0]$ can either be an estimate of the population density at $k = 0$ or the vector $\mathbf{1}/N$.

D. Population Density Approach

Alternatively, a transition model may be constructed by generating a population density vector from previous data. Let $\vec{R}[k] = (x_k, y_k)$ be a vector denoting the fish's location at time step k, and let $\vec{V}_i = (x_i, y_i)$ denote the location of node i. From these quantities we may define an *attenuation vector* $\mathbf{a}[k]$ as follows:

$$\mathbf{a}_i[k] = e^{-||\vec{R}[k] - \vec{V}_i[k]||^2/(2\sigma_{\text{atten}}^2)}$$

where σ_{atten} is a free parameter. Informally, $\mathbf{a}[k]$ describes the weight of a particular node in the fish's location. If the fish's location at some k precisely equals the location of node i, then $\mathbf{a}_i[k] = 1$; but as the fish moves away from i, then $\mathbf{a}_i[k] \to 0$. The parameter σ_{atten} sets the "attenuation" of the a node's assigned weight as the fish moves away from it.

We then define the population density vector as a weighted sum of the attenuation vector:

$$\mathbf{p}[k] = \sum_{m}^{L} f(m|k, \sigma/T_{\text{sample}}) \mathbf{a}[m].$$

Ideally, we would like to construct a transition matrix $T[k]$ such that $\mathbf{p}[k+1] = T[k]\mathbf{p}[k]$. In some cases, however, such a construction may not be possible. Instead, we find a $T[k]$ that minimizes $||\mathbf{p}[k+1] - T[k]\mathbf{p}[k]||^2$, subject to the constraint that $T[k]$ is a stochastic matrix.

IV. Controller

The final step is to construct a Markov process that will minimize the difference in robot population density and fish population density at each time. Again, there are two approaches: tracking the fish transition matrix, or attempting to match population density directly. In the former case, the robot simply uses the transition matrix generated by the transition model. In the latter case, a feedback control is used to generate a transition matrix at each time step that will match the robot population density to fish population density.

The entries of the robot transition matrix often exhibit periodic behavior. In these cases, performing a Fourier transform of the transition probabilities, filtering out some number of the largest Fourier coefficients, and reconstructing the transition probabilities from these coefficients can fit a sum of sinusoids to entries of the transition matrix.

A. Feedback Control for Density Tracking

In tracking population density, the objective is to minimize the difference between fish population density $\mathbf{p}[k]$ and robot population density $\rho[k]$. A feedback control may be used to this end. After defining an error vector $\mathbf{e}[k] = \mathbf{p}[k] - \rho[k]$, This control generates a transition matrix $T[k]$ at each time k as follows:

$$T_{j,i}[k] = \begin{cases} K\mathbf{e}_j[k]/(\rho_j[k] + \varepsilon) & i \neq j \\ 1 + K'\mathbf{e}_j[k]/(\rho_j[k] + \varepsilon) & i = j \end{cases}$$

The columns are subsequently normalized so that $T[k]$ is a stochastic matrix.

B. Proof of Feedback Control Convergence

C. Edge Removal

In some cases, it may be necessary to ensure that the robot has zero probability of traversing a particular edge, if such a path would be physically difficult to navigate. In these cases, linear programming may be recursively used to calculate a transition matrix $T[k]$ so that the error

$$\left|\left| \left(\prod_{m=0}^{k-1} T[m] \right) T[k] \rho[0] - \mathbf{p}[k] \right|\right|^2,$$

(where $\rho[0]$ is the initial population density of the robots) is minimized subject to the constraint that any entries of T corresponding to a disallowed transition are set to zero.

V. Results and Discussion

A. Simulation

Both the feedback control and the periodic control were tested against a historic dataset in a MATLAB simulation. The transition matrix approach was used to generate a transition model from one week of location data of a single fish off the shore of Catalina Island. The fish's behavior during this time was highly periodic, as shown in Figure 1. Using this transition model, 100 fish were simulated, and a varying number of AUVs were also simulated using either the feedback control or the periodic control to track this population.

To evaluate the control, we used three error metrics:

1) **Closest AUV Distance** The Closest AUV Distance error tracked the distance from each simulated fish to the closest AUV, averaged across all of the fish and all time steps of the simulation.
2) **Fraction In Range** The Fraction In Range metric tracked the fraction of the simulated fish within some distance of an AUV. Since acoustic tracking is only feasible if the AUV is within a certain range of the tag, this error is useful in determining how many fish would have been lost during a track.
3) **Population Density Error** Population Density Error is a measure between 0 and 1 quantifying the difference between the distribution of simulated fish across the nodes and the distribution of AUVs. For a fish density $\mathbf{p}_{\text{fish}}$ and AUV density $\mathbf{p}_{\text{AUV}}$, this error is calculated as

$$e = \sqrt{\frac{1}{2}(\mathbf{p}_{\text{fish}} - \mathbf{p}_{\text{AUV}}) \cdot (\mathbf{p}_{\text{fish}} - \mathbf{p}_{\text{AUV}})}$$

For this analysis, $\mathbf{p}_{\text{AUV}}$ is measured from the AUVs' destination nodes; i.e., from the nodes the AUVs are moving towards rather than the nodes to which they are closest at the time.

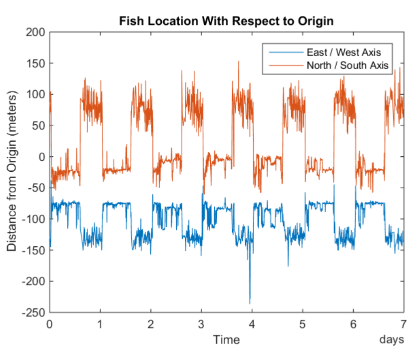

Fig. 1. Simulations used a transition model generated from a historical week-long track of a single fish. The above graph shows the fish's coordinates along the East-West and North-South axes from a constant origin.

B. Periodic Control

The periodic control was simulated varying the number of Fourier components kept and the number of AUVs simulated. Since the week of historical data contained 5,040 samples, the resulting transition probability functions contained 2,520 frequency components revealed by an FFT. The periodic control was testing by keeping various numbers of the highest-magnitude frequency components, between 0 (resulting in the AUVs taking random transitions) and 2,520 (in which case the AUVs follow exactly the same transition probabilities as the fish). Also, varying numbers of AUVs, between 1 and 100, were tested.

Figure 2 shows the Closest AUV Distance error for each of the combinations of parameters. As would be expected, when more AUVs are present, fish are more likely to be close to an AUV. When 10 or more AUVs are used, i.e., when there are more AUVs than nodes, even random transitions result in an error below 12 meters.

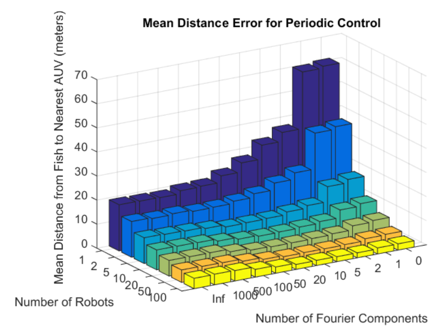

Fig. 2. The above plot shows the average distance from a fish to the closest AUV when using the periodic control with varying numbers of AUVs and Fourier components.

Figure 3 shows the Fraction In Range error for the periodic control. When 50 or more AUVs are present, regardless of the number of Fourier components used, more than 90% of the simulated fish remain within range of the AUV. This fraction remains above 80% when there are more AUVs than there are nodes, as long as 10 or more frequency components are kept. But when the number of AUVs falls below the number of nodes, the fraction of fish within range rapidly drop off to the best case of 35% in the single AUV trials.

Finally, Figure 4 shows the Population Density Error. With 50 or more frequency components and 10 or more robots, this error flattens out between 0.21 and 0.25. The error tends to increase with the number of AUVs, as more AUVs are available to fit the transition model's predicted population density more precisely. Additionally, error tends to decrease as more frequency components are used and the AUV's transition probabilities converge on those of the simulated fish.

C. Feedback Control

The feedback control was simulated with varying numbers of tagged fish, between 0 (resulting in random transitions)

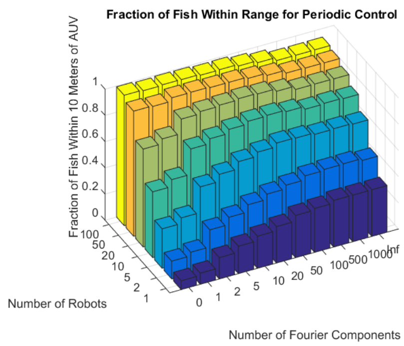

Fig. 3. The above plot shows the fraction of fish within 10 meters of an AUV when using the periodic control. Note that the independent axes are reversed with respect to the graphs in the other periodic control figures.

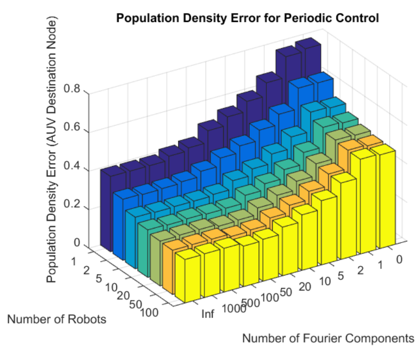

Fig. 4. The above plot shows the Population Density Error from periodic control simulations.

and 100; as well as the number of AUVs, also between 0 and 100. Figure 5 shows the Closest AUV Distance error for these parameters. As was the case with the periodic control, if there are significantly more AUVs than there are nodes, the control can perform well even with little information about the fish. Interestingly, after 5 fish are tagged, additional tagged fish have negligible effect on the error (and slightly increase error when only 1 or 2 AUVs are used). Since the simulated fish population in these tests follows the transition model of a single fish, the distribution of fish is unimodal; as such, a single tagged fish is usually sufficient to bring the AUVs near most of the fish.

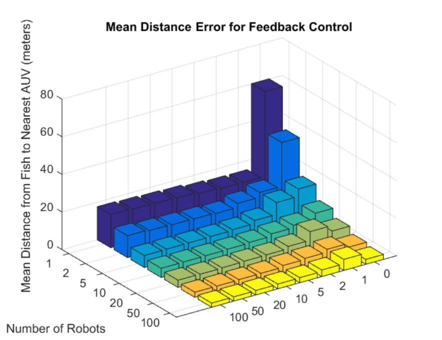

Fig. 5. The above plot shows the average distance from a fish to the closest AUV when using the feedback control with varying numbers of AUVs and tagged fish.

Figure 6 shows the Fraction In Range error for the feedback control. A striking feature of this plot is the dip in the fraction of fish in range when exactly one fish is tagged. When the number of AUVs significantly exceeds the number of nodes, the random transitions when zero fish are tagged result in a uniform distribution of AUVs across the nodes, resulting in a high probability that all fish will share a node with at least one AUV. But when the number of AUVs is comparable to or smaller than the number of nodes, this random distribution strategy fail to achieve a uniform distribution of AUVs. In this case, the information provided by tagged fish is necessary to best-distribute the limited number of AUVs.

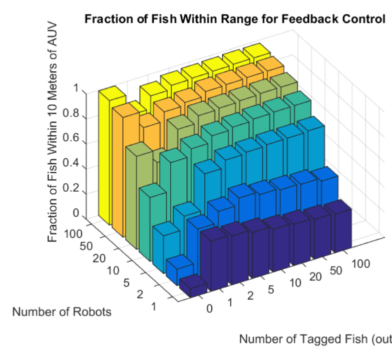

Fig. 6. The above plot shows the fraction of fish within 10 meters of an AUV when using the feedback control. Note that the independent axes are reversed with respect to the graphs in the other feedback control figures.

Figure 7 shows the Population Density Error. When there are more AUVs than nodes and at least 5 fish are tagged, the error flattens out between 0.10 and 0.13. Again, diminishing returns are evident for tagging more fish; after 1 or 2 fish are tagged, the benefit of additional tagged fish is minimal. In fact, the Population Density Error plot reveals another interesting trend when fewer AUVs are in use. With one AUV, tagging beyond one fish actually *increases* error, from 0.31 when one fish is tagged to 0.43 when all of the fish are tagged. The same effect is visible after tagging more than 5 fish when two AUVs are in use and after 20 fish when 5 AUVs are used. Increasing the number of AUVs seems to increase the number of tagged fish required to optimize the AUVs' population density.

D. Trends to Investigate

There are several questions from this data that I wish to investigate further.

1) What causes the Mean Distance Error for the periodic control to increase when more than 50 Fourier components are used? Why does the Population Density Error increase when more than 100 Fourier components are used? Why is this increase visible after the same number of Fourier components in both error metrics?
2) Can the spike in feedback control error when one fish is tagged and many AUVs are deployed be quantified?

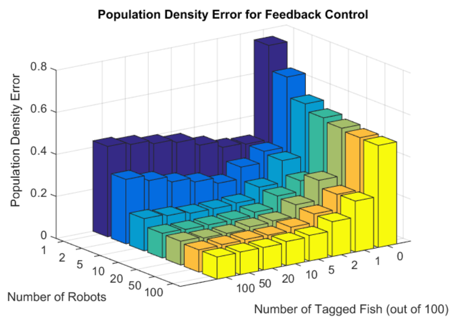

Fig. 7. The above plot shows the Population Density Error from feedback control simulations.

3) Feedback control error is increased when more fish are tagged and few AUVs are present, but decreased when more fish are tagged and many AUVs are present. Can this relationship be quantified? What theorems from statistics can be applied here?

E. Field Test

VI. Conclusion and Future Work

APPENDIX

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under Grant No. 1245813. This work was performed in part at the USC Wrigley Institute for Environmental Studies Wrigley Marine Science Center and at the Claremont Colleges Robert J. Bernard Biological Field Station.

References

[1] PLACE HOLDER