Using TDA to Analyze the Swimming Movement of Fish

Hannah Proctor and Khush Agrawal, Advised by Dr. Michael Jackson,
Grove City College Department of Mathematics, January 2024

Introduction

Motivation

Does the tail beat of fish exhibit a resonance allowing the fish to find a tail beat frequency that achieves a higher efficiency? Is topological data analysis (TDA) a viable approach to answering this question?

Abstract

In the children's movie Finding Nemo, Dory imparts the inspiring words of wisdom – "Just keep swimming!" For the past several years, Dr. Erik Anderson, a professor of Mechanical Engineering at Grove City College, has put her catch phrase to the test. His research centers around the question above, attempting to identify whether or not a fish's tail naturally beats with a resonance that makes for better swimming. In doing so, he has collected swimming data for many different species at various speeds. Using python, we processed a selection of this data and created animations that mirrored the original fish movement. We then generated persistence diagrams correlating to three differently sized fish of the same species. Ultimately, we aimed to analyze differences between these diagrams and to identify patterns that lie within the data to inform Dr. Anderson's research.

Data Collection and Processing

About the Data

Dr. Anderson's data is collected using a specially crafted flow tank at the Woods Hole Oceanographic Institute. It allows him to control the speed of the water and operate a camera that repeatedly captures the bottom-view of a fish swimming (see figure 1.1). The pictures are then stripped into data that keeps track of the position and outline of the fish in each frame. In this project, our data selection was taken from three differently sized fish of the same species: scup 43, scup 44, and scup 45 (see figures 2.1-3).

Figure 1.1: Original Image

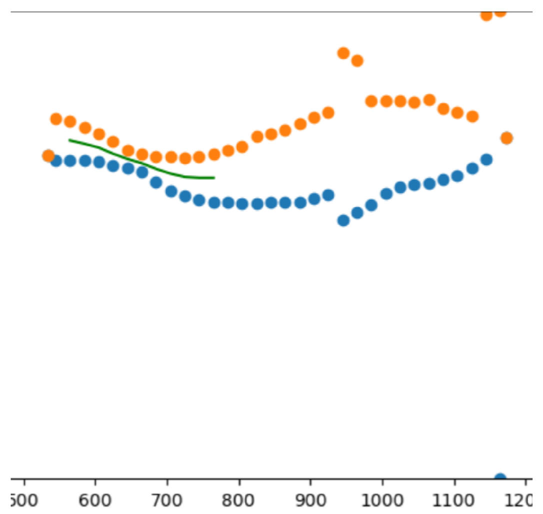
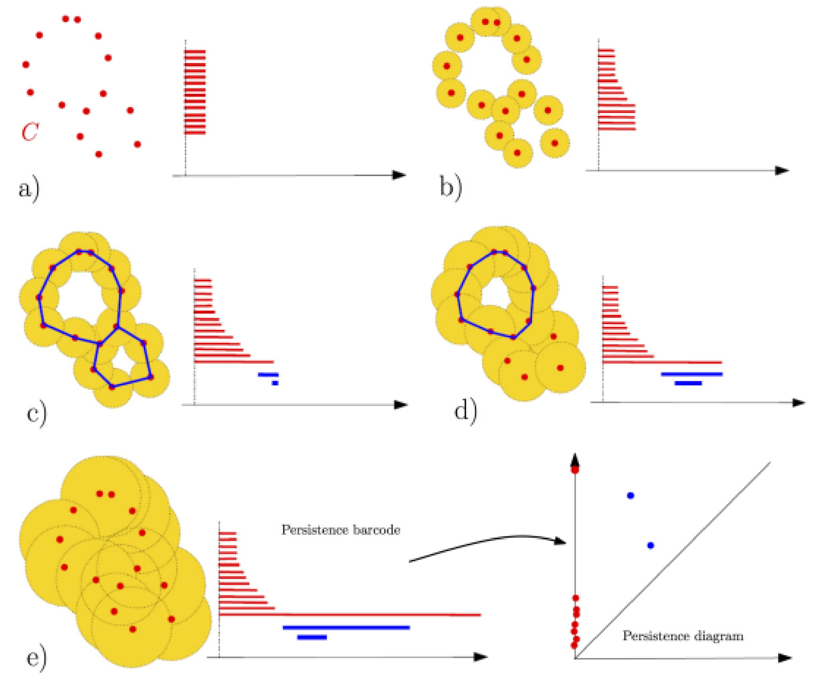


Figure 1.2: Processed Image

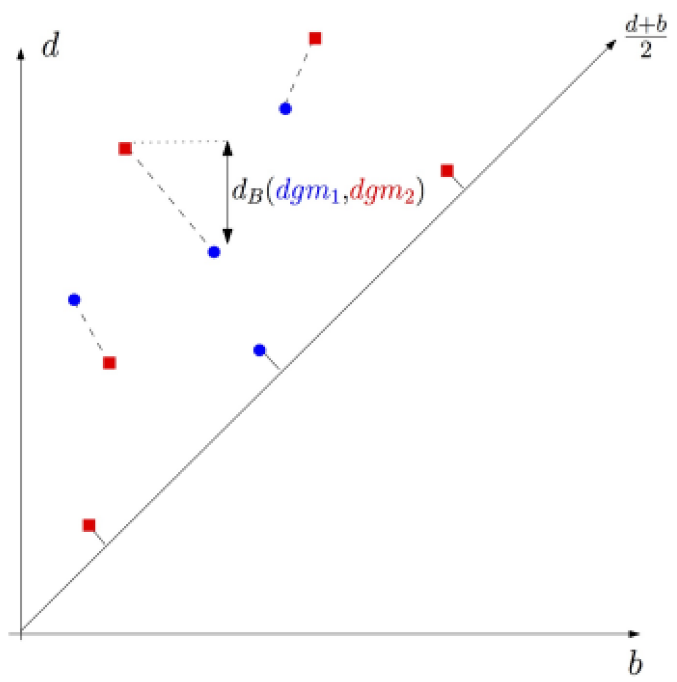
Utilizing Python

For each fish, there were about 60-80 csv files representing various water speeds. So, to process and analyze the data, we turned to python. First we used the pandas library to read in the points from each file. To eliminate as many outliers as possible, we automated a process to identify 'good data' frames. Good data frames included frames where the fish was completely in view and the tail portion was clear without missing or misplaced points. From there, we isolated the tail movement into a numpy array using a basic points averaging system. Plotting both the original and the average array with matplotlib.pyplot generates the image above (figure 1.2). We then formulated a metric for the curvature of the tail, and applied the sliding window technique on the resulting array, this was inspired by Thomas et. al.(4). Finally, we created persistence diagrams through the ripser and persim plot_diagrams modules.

Finding Nemo Meets TDA



Persis. Diagram (1)



Bottleneck Distance (1)

Definitions

Persistence diagrams are used to analyze the structure of data using the process depicted on the left (also see [1] or [2]). To compare persistence diagrams, we used **bottleneck distance**:

$$b_d(dgm1, dgm2) = \inf_{\substack{\text{matching} \\ m}} \max_{(p,q)\in m} ||p - q||_\infty.$$

Graphically, this measure is shown to the right.

Implementation Details

During implementation we considered three things: size, speed, and movement. Notice that scup 43 is the smallest of the three fish, followed by 44 and 45, respectively. We looked at 9.5 hz and 14 hz specifically (where a higher hertz refers to a faster flow speed while the unit is based on tank operation details). To ensure similarities between the data, we chose one 100-frame window for each fish where there movement was the least sporadic. Using this framework, we got persistence diagrams like those shown in Table 1.

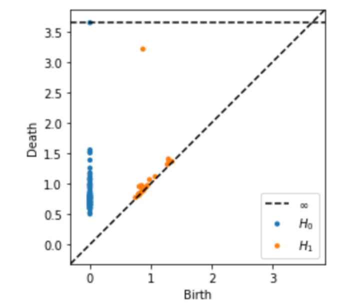
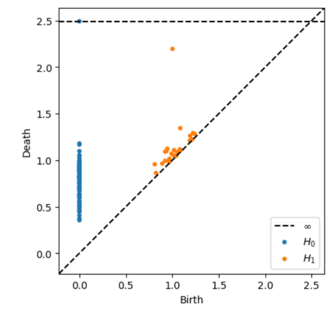
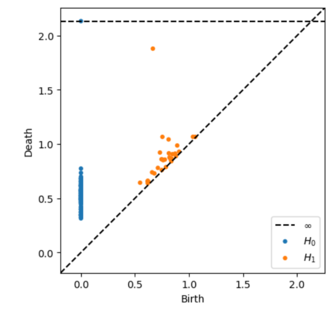
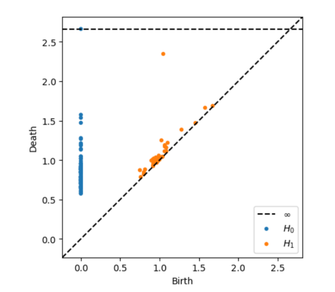
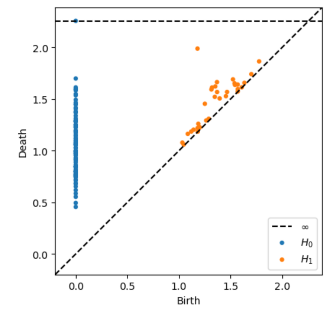
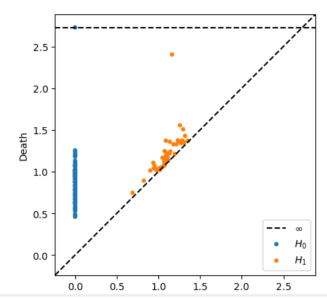
tank setting hz	Water speed cm / s	Scup 43	Scup 44	Scup 45
9.5	36.4			
14	54.0			

Table 1: Table of Persistence Diagrams

Analysis

As shown above, each persistence diagram has a single, prominent point away from the diagonal. This point results from the periodic nature of the fish motion, and indicates that our selection is actually good data. Notice that the birth number occurs earlier for the lower speed as opposed to the higher speed for each fish. The weight of this difference varies, but ultimately it signifies that our data is more spread out in higher speeds.

Conclusion

Final Remarks

The purpose of this project was twofold: (1) to analyze differences between a collection of persistence diagrams, and (2) to identify patterns that lie within the data to inform Dr. Anderson's research. We accomplished the first goal by analyzing persistence diagrams like those found in Table 1. We looked at size, speed, and movement to identify possible patterns and calculated the bottleneck distance between diagrams to measure difference. The next step is to calculate the bottleneck distance between the diagrams and to preform a more complete analysis with that data. This will provide a more concrete framework to discover correlations.

So, does the tail beat of fish exhibit a resonance allowing the fish to find a tail beat frequency that achieves a higher efficiency? We don't know, our second goal has yet to be completed. Our work only scratches the surface, but from what we've seen TDA is a viable approach to approaching the question.

Figure 2.1: Scup 43, 21.4 cm

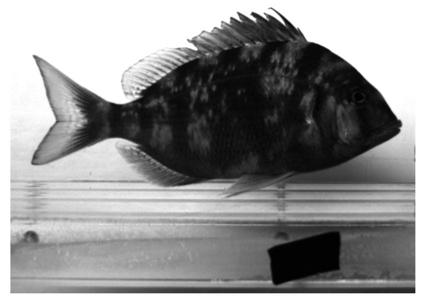


Figure 2.2: Scup 44, 19.6 cm

Figure 2.3: Scup 45, 28.6 cm

Further Study

We have shown that TDA is a viable option for investigating fish swimming motion. Including more water speeds in the analysis will allow us to look for trends as the speed varies. In addition, we can analyze fish movement within the good frames that we are using to see if the persistence diagrams are dependent on the movement of the fish. Once this initial work is completed with the three fish, we would like to expand the analysis by including more fish and additional species of fish.

References

(1) Tralie, Christopher J., and Perea, Jose A. "(Quasi)Periodicity Quantification in Video Data, Using Topology". CoRR abs/1704.08382 (2017).

(2) Chazal, Frédéric, and Bertrand, Michel. "An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists". Frontiers in Artificial Intelligence 4 (2021).

(3) Munch, Elizabeth. 2017. "A User's Guide to Topological Data Analysis". Journal of Learning Analytics 4 (2):47–61. https://doi.org/10.18608/jla.2017.42.6.

(4) Thomas, Ashleigh, Kathleen, Bates, Alex, Elchesen, Iryna, Hartsock, Hang, Lu, and Peter, Bubenik. "Topological Data Analysis of C. elegans Locomotion and Behavior". Frontiers in Artificial Intelligence 4 (2021).

Funding was provided by the Swezey Fund for Scientific Research and Instrumentation, Grove City College.

GROVE CITY COLLEGE