

**IDENTIFYING HUMPBACK WHALE FLUKES BY
SEQUENCE MATCHING OF TRAILING EDGE
CURVATURE**

By

Zachary Jablons

A Thesis Submitted to the Graduate
Faculty of Rensselaer Polytechnic Institute
in Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE
Major Subject: COMPUTER SCIENCE

Examining Committee:

Dr. Charles Stewart, Thesis Adviser

Dr. Barbara Cutler, Member

Dr. Bülent Yener, Member

Rensselaer Polytechnic Institute
Troy, New York

April 2016
(For Graduation May 2016)

© Copyright 2016
by
Zachary Jablons
All Rights Reserved

CONTENTS

LIST OF TABLES	v
LIST OF FIGURES	vi
ACKNOWLEDGMENTS	vii
ABSTRACT	vii
1. Introduction	1
1.1 Humpback Whales	1
1.2 Current Identification Methods	1
1.2.1 Based on Trailing Edge	1
1.2.2 Based on general Fluke appearance	1
1.3 Hotspotter	1
1.4 The Dataset	1
2. Background	2
2.1 Seam Carving	2
2.2 Convolutional Networks	2
2.2.1 Facial Keypoint Prediction	2
2.2.2 Fully Convolutional Networks	2
2.2.3 Semantic Segmentation	2
2.3 Curvature Measures	2
3. Methods	3
3.1 Trailing Edge Extraction	3
3.1.1 Gradient Information	3
3.1.2 Keypoints	3
3.1.3 Trailing Edge Scoring	3
3.2 Trailing Edge Matching	3
3.2.1 Curvature Measurement	3
3.2.2 Sequence Matching	3
3.3 Alternative Approaches	3
3.3.1 Embedding via Convolutional Networks	3

3.3.1.1	Raw Images	3
3.3.1.2	Trailing Edges	3
3.3.2	Aligning Trailing Edges	3
3.3.2.1	Keypoint Alignment	3
3.3.2.2	Dynamic Time Warping Alignment	3
3.3.3	Histogram Matching	3
4.	Results	4
4.1	Main method	4
4.1.1	Characterization of Success cases	4
4.1.2	Characterization of Failure cases	4
4.1.3	Variations	4
4.2	In Combination with Hotspotter	4
4.2.1	Failure cases	4
5.	Discussion	5
5.1	Issues with current method	5
5.2	Future work	5
5.3	Conclusion	5
	REFERENCES	6
	APPENDIX	

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

This project could not have been completed so quickly and thoroughly without the invaluable help from the rest of the IBEIS team. I would like to thank Jon Crall for helping to write the experimentation framework which has saved a lot of time and effort. I would also like to thank Jason Parham and Hendrik Weidemann for the great discussions and advice, as well as the handy \LaTeX template. Additionally, without the development and annotation efforts of Andrew Batbouta, a lot of models could not have been so successfully trained. Importantly, I would like to thank the Wildbook team (Jason Holmberg, Jon van Oast) for providing the dataset and corresponding annotations with which a lot of models were trained and experiments run. Of course, this work could never have been undertaken without the help, advice, and direction of my advisor, Charles Stewart.

ABSTRACT

Humpback whales (*Megaptera novaeangliae*) are an important part of our ocean's ecosystems [citation needed], and have historically been at risk for extinction [citation needed]. While they are currently rated as 'Least Concern' [citation needed], tracking their migration patterns is important for helping the (currently small) population grow [citation needed]. In order to discern these migration patterns, conservationists need to be able to track individual humpback whales [citation needed]. One of the easiest (and cheapest) ways to do this is to watch for their tails as they breach the water surface [citation needed], giving a clear view of what is known as a Humpback 'fluke'. These are often patterned and scarred in unique ways, allowing conservationists to identify individuals [citation needed]. However, until recently, most automated identification methods still rely on significant manual effort to describe and identify the fluke, severely limiting the amount of humpbacks that can be tracked [citation needed?].

This thesis lays out a method that automates the identification of Humpback flukes directly from still images thereof, using the 'trailing edge' of the fluke. Using this method, we achieve a fairly high top-1 ranking accuracy on a large dataset (consisting of about 400 identified individuals). We also show that this method significantly helps the accuracy of a pure appearance based method, Hotspotter [citation needed], giving 89% top-1 accuracy.

To our knowledge, this is the first method that can achieve this level of accuracy on Humpback fluke identification without any manual effort at test-time.

TODO: Put in citations

CHAPTER 1

Introduction

1.1 Humpback Whales

1.2 Current Identification Methods

Photo-identification of Humpback whale flukes has been attempted since the early 90s [1]

1.2.1 Based on Trailing Edge

1.2.2 Based on general Fluke appearance

1.3 Hotspotter

1.4 The Dataset

CHAPTER 2

Background

2.1 Seam Carving

2.2 Convolutional Networks

2.2.1 Facial Keypoint Prediction

2.2.2 Fully Convolutional Networks

2.2.3 Semantic Segmentation

2.3 Curvature Measures

CHAPTER 3

Methods

3.1 Trailing Edge Extraction

3.1.1 Gradient Information

3.1.2 Keypoints

3.1.3 Trailing Edge Scoring

3.2 Trailing Edge Matching

3.2.1 Curvature Measurement

3.2.2 Sequence Matching

3.3 Alternative Approaches

3.3.1 Embedding via Convolutional Networks

3.3.1.1 Raw Images

3.3.1.2 Trailing Edges

3.3.2 Aligning Trailing Edges

3.3.2.1 Keypoint Alignment

3.3.2.2 Dynamic Time Warping Alignment

3.3.3 Histogram Matching

CHAPTER 4

Results

4.1 Main method

4.1.1 Characterization of Success cases

4.1.2 Characterization of Failure cases

4.1.3 Variations

4.2 In Combination with Hotspotter

4.2.1 Failure cases

CHAPTER 5

Discussion

5.1 Issues with current method

5.2 Future work

5.3 Conclusion

REFERENCES

- [1] S. A. Mizroch, J. A. Beard, and M. Lynde. Computer assisted photo-identification of humpback whales. *Report of the International Whaling Commission*, 12:63–70, 1990.

APPENDIX