Automatic Optical Detection and Classification of Marine Animals around MHK Converters using Machine Vision

Steven L. Brunton
University of Washington
sbrunton@uw.edu 609.921.6415
February 2017

Project Overview

Automatic optical detection and classification of marine animals around MHK converters using machine vision

The Challenge: As part of environmental review and monitoring, MHK developers are often required to perform studies to examine and monitor potential impact of projects on presence, behavior, and abundance of species in prospective sites.

Continuous monitoring of the marine environment at MHK sites is essential to quantify and manage environmental risk uncertainties, including interaction of animals with converters, noise levels, and changes to marine animal distribution and habitat.

However, the deluge of optical data makes expert review time-consuming and expensive, leading to a so-called data mortgage.

The goal of this project is to develop a software pipeline to leverage machine learning for the automatic detection and classification of marine animals to improve MHK site monitoring and alleviate the growing data mortgage.

Partners: Brian Polagye [UW]: Provided MHK data and expertise
Jenq-Neng Hwang [UW]: Machine learning, fish recognition
Sharon Kramer [H.T. Harvey]: Environmental consulting

Mitigate Environmental Risk Uncertainty at MHK Sites

Technology Maturity

- Test and demonstrate prototypes
- Develop cost effective approaches for installation, grid integration, operations and maintenance
- Conduct R&D for Innovative MHK components
- Develop tools to optimize device and array performance and reliability
- Develop and apply quantitative metrics to advance MHK technologies

Deployment Barriers

- Identify potential improvements to regulatory processes and requirements
- **Support research focused on retiring or mitigating environmental risks and reducing costs**
- Build awareness of MHK technologies
- Ensure MHK interests are considered in coastal and marine planning processes
- Evaluate deployment infrastructure needs and possible approaches to bridge gaps

Market Development

- Support project demonstrations to reduce risk and build investor confidence
- Assess and communicate potential MHK market opportunities, including off-grid and non-electric
- Inform incentives and policy measures
- Develop, maintain and communicate our national strategy
- Support development of standards
- Expand MHK technical and research community

Crosscutting Approaches

- Enable access to testing facilities that help accelerate the pace of technology development
- Improve resource characterization to optimize technologies, reduce deployment risks and identify promising markets
- Exchange of data information and expertise

Mitigate Environmental Risk Uncertainty at MHK Sites

Deployment Barriers

- Identify potential improvements to regulatory processes and requirements
- **Support research focused on retiring or mitigating environmental risks and reducing costs**
- Build awareness of MHK technologies
- Ensure MHK interests are considered in coastal and marine planning processes
- Evaluate deployment infrastructure needs and possible approaches to bridge gaps

The Impact

- The target of this project is to 1) develop a modular software package to automatically detect and flag events for storage and eventual human review, and 2) to train and test various classifiers on MHK image data to assess the effectiveness of automatic classification. The goal for event detection and classification accuracy rates is at least a 50% improvement over a random guess across the possible categories. For event detection, this translates to 75% accuracy.
- This project may significantly reduce the burden of data collection and manual expert review, providing a valuable tool to assess and retire environmental risk uncertainty around the effect of MHK sites on marine animals.
- This project has culminated in the development of an open-source software framework to scrub and classify image data from MHK sites.

Technical Approach

There are two key components to this project:

1. Build an **open-source software framework** to process and classify MHK image data. A major goal is to be modular and flexible to encourage future development.
2. Develop and test various **data scrubbing and machine learning** algorithms to effectively detect and classify images. This will flag important data to be stored and reduce the data mortgage.

Key Issue: Reduce data mortgage by detecting/classifying images so that only important images are kept for future human review.

Classification near MHK site is particularly difficult since environment is unstructured; low-lighting and occlusions also challenging.

Expert team leveraging 1) software engineering, 2) environmental consulting, 3) machine learning, and 4) fish recognition.

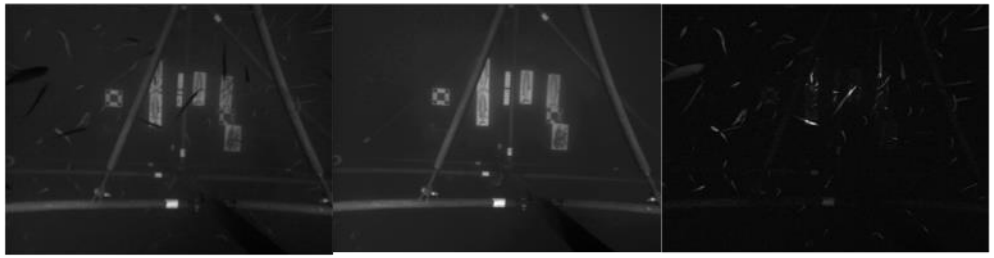
Unique data set: 60GB, 14,000 HD images from Sunset Bay

Technical Approach

There are two key components to this project:

1. **Open-source software framework:**
   a) Version control (multiple teams can develop and branch)
   b) Documentation and unit tests (changes easily understood and verified)
   c) Modular (better algorithms easily implemented, flexible protocols)
   d) Graphics processing unit (GPU) accelerated computations (real-time)
2. **Data scrubbing and machine learning:**
   a) Background subtraction and lighting correction
   b) Feature extraction (in consultation with marine experts)
   c) Hierarchical data labeling for flexible detection/classification protocols
   d) Detection and classification algorithms developed/tested

Original Filtered Occlusion

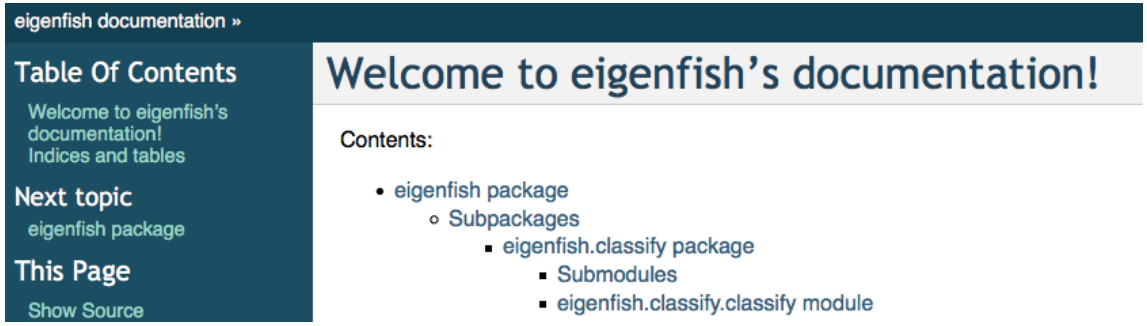


Original Background Fish

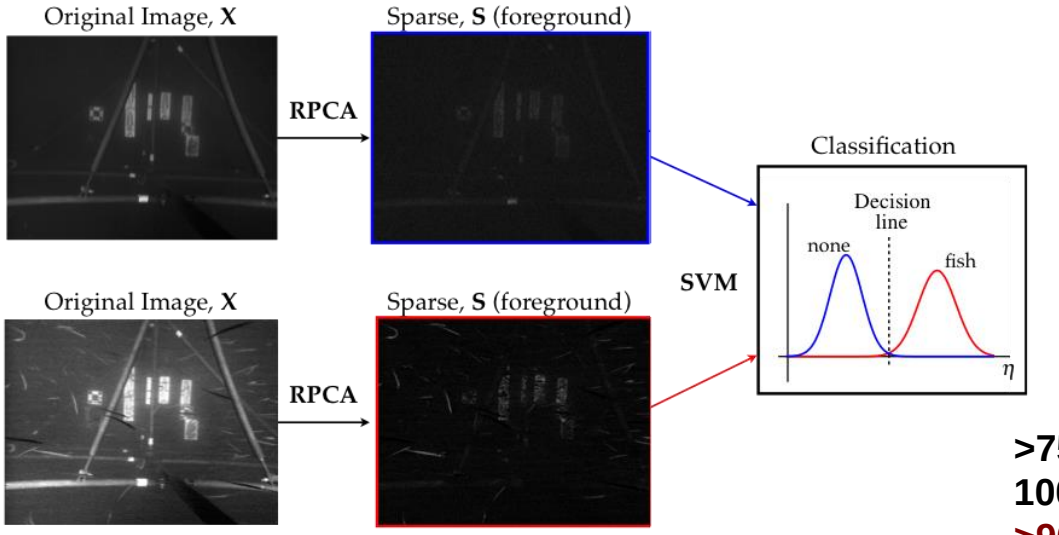
Robust Principal Components Analysis (RPCA) for Background Subtraction (on GPU)

Accomplishments and Progress

1) Open-source software platform deployed on GitHub to encourage broad adoption and development by the MHK community.



1) Data processing and machine learning implemented and tested



	LDA (%)	QDA (%)	SVM (%)
Fish vs. No Fish	85.2	89.1	100
Something vs. Nothing	66	71.3	79.1
One Species vs. Two	90.8	90.8	83.1
Algae vs. Invertebrates vs Vertebrates	84.1	92.8	85.3
Uninteresting, Mildly Interesting, Very Interesting	83.7	91.7	85.1

(all based on expert labels)

>75% Detection of something vs. nothing☑

100% Detection of fish vs. no fish☑

>90% Classification of image as "interesting"☑

Project Plan & Schedule

- Project Start Date: October 1, 2014
- Project End Date: September 30, 2016
- No Cost Extension: June 30, 2017

- All milestones met on time.

- Go/No-Go #1 [M12, Q4, Sep. 30, 2015]: Software interface decided on, labeled data acquired and converted into common format, and RPCA algorithm used for image background subtraction. [Status: complete on date]

- Final Deliverable [M24, Q8, Sep. 30, 2016]: Final software delivered on GitHub repository, fully documented, with unit tests that pass. [Status: complete on date]

Project Budget

U.S. DEPARTMENT OF ENERGY | Energy Efficiency & Renewable Energy

Budget History					
FY2014		FY2015		FY2016	
DOE	Cost-share	DOE	Cost-share	DOE	Cost-share
22619	3981	111829	13909	76631	6772

- We requested a 9-month no-cost extension to continue writing up results in peer-reviewed journals and presenting at conferences.
- We have spent 94% of the budget to date.

Research Integration & Collaboration

Partners, Subcontractors, and Collaborators:
Sharon Kramer [H. T. Harvey & Associates] and her team provided expert consultation on which features in data are important for classification. Together, we developed a universal image labeling system and they then made an extensive labeled data set to train algorithms.

Communications and Technology Transfer: We have written 5 papers at various stages [published, under review, in preparation]:

1. "Data-Driven Methods in Fluid Dynamics: Sparse Classification from Experimental Data", Ch 17 in *Whither Turbulence and Big Data in the 20th Century,* Springer 2017.
2. "Compressed Dynamic Mode Decomposition for Real-Time Object Detection", *Accepted to Journal of Real-Time Image Processing*, 2016.
3. "Automated Fish Detection and Identification in Underwater Video: A Technology Roadmap"*, In preparation (w/ Shari Matzner),* 2016.
4. "Streaming GPU Dynamic Mode Decomposition", *In preparation*, 2016.
5. "Automatic optical detection and classification of marine animals around marine hydrokinetic converters using machine vision", *In preparation*, 2016.

Open Source Code Available at: https://github.com/sethdp/eigenfish
Outcome: Already incorporated into PNNL MHK effort [Harker-Klimes, Matzner].

Next Steps and Future Research

FY17/Current research: Project period is finished and all milestones/deliverables are complete. We have requested a no cost extension to continue disseminating these results via: 1) finishing publications, 2) presenting at conferences, and 3) interacting with other teams exploring MHK data.

Proposed future research: The software pipeline was designed to be a flexible platform for future expansion and development by users. In particular, the software is modular, so improved data processing and machine learning algorithms can easily be included. Future research would extend the methods from post-processing detection to real-time detection and higher resolution classification.