Using High Performance Computing to Explore Large Complex Bioacoustic Soundscapes: Case Study for Right Whale Acoustics

Peter Dugan\textsuperscript{a}, Mohammad Pourhomayoun\textsuperscript{a}, Yu Shiu\textsuperscript{a}, Rosemary Paradis\textsuperscript{b} Aaron Rice\textsuperscript{a} and Christopher Clark\textsuperscript{a}

\textsuperscript{a}Bioacoustics Research Program (BRP), Cornell University, Ithaca, NY, USA 14850
\textsuperscript{b}Independent Consultant, Binghamton, NY, USA, 13850

Abstract

This paper describes ongoing work being done at Cornell University to investigate the development of a complex system designed for extracting information from large acoustic datasets. The system, called DeLMA, is based on integrating advanced machine learning with high performance computing (HPC). The goal of this work is to provide the capability to accurately detect and classify whale sounds in large acoustic datasets collected using underwater sensors. The case study for this work is focused on detecting the acoustic communication signals of the North Atlantic Right Whale, \textit{Eubalaena glacialis}, and uses data collected in the Stellwagen Bank National Marine Sanctuary (SBNMS), USA. A summary of the work done for developing a complex detection-classification system and brief description of several algorithms that are used for classifying whale sounds will be covered. A brief discussion on how standard detection algorithms can be incorporated, with no special modifications, into the HPC system for analysis will be mentioned, and two new right whale detection methods are presented, based on continuous region analysis (CRA) and histogram of oriented gradients (HOG). This paper presents a first-hand look at applying the DeLMA system and these algorithms on a large dataset containing over 60,000 channel-hours of acoustic data from the SBNMS. Results from these new detection methods are compared against Baseline algorithms. With the development of the DeLMA system, sound archives can now be explored using a powerful distributed processing architecture. This advancement will allow for rapid execution and visualization of the data using seasonal graphs called diel plots, which show the distribution of detections on a time-of-day vs. time-of-year plane. Diel plots of Baseline, CRA and HOG algorithm results reveal various large-scale features of the seasonality of whale calling behavior. Results are summarized and the authors discuss future areas for study, especially those relate to handling other big passive acoustic data projects.

1. Introduction

In the past several decades, scientists have been monitoring the ocean environments using a variety of sensor modalities. Passive acoustic monitoring is one of the primary methods used to investigate and understand animal behavioural patterns [1]. The acoustic modality is particularly appropriate for marine mammals because all species studied to date are known to produce sounds for foraging, navigating and/or communicating. Furthermore, acoustic

\textsuperscript{1}Work made possible by National Oceanic Partnership Program (NOPP), Office of Naval Research (ONR) N000141210585, National Fish and Wildlife Foundation (NFWF) -0309.07.28515.
monitoring methods are not subject to visual sighting limitations imposed by weather, daylight and ocean environmental conditions [2]. The main challenge, as with many fields, is the ability to process and analyse the vast amount of data collected. In many cases, analysis relies on primitive, inefficient and time-consuming processes, such as human inspection. This is often slow and inefficient, and only allows small percentages of the data to be inspected. In this paper, we apply algorithms based on image processing and machine-learning to detect North Atlantic Right Whales (NARW) contact calls, referred to as up-calls, in the presence of environmental noise. We will also introduce a new high performance computing (HPC) technology, applied within a system called DeLMA, which provides an efficient processing mechanism for the application of automatic detection and classification algorithms.

1.1 Right Whale Auto Detection Algorithms

The NARW is one of the world’s most highly endangered whales [6, 7]. Therefore, there is an urgent need to develop efficient techniques to detect its presence so as to determine their seasonal occurrences and protect them from possible harm [4]. Right whales produce up-calls for long-range communication. Up-calls are frequency-modulated upsweeps in the 50-250 Hz frequency band [3], and detection of up-calls has been shown to be the most effective mechanisms for determining whale presence in critical habitats [4, 5]. Monitoring, usually involves real-time information with the goal of minimizing ship strikes or the potential harm from seismic exploration [8]. Research typically uses archival data to understand the ocean environment and various ecological aspects, such as seasonal migration behaviour and food resources. In the past decade, researchers have been working to design effective automated algorithms for identifying marine mammal vocalizations, including NARW up-calls [1, 2, 5, 9-17]. Techniques from the field of computer vision and image processing have proven to be quite useful for building automatic algorithms. Recent developments in high performance computing have provided new ways to utilize large pools of CPU, memory and data storage. These technologies allow researchers the ability to access large amounts of data at high rates [18]. However, right whale detection algorithms have not been exercised using HPC technology. The work herein focuses on using HPC within a system called DeLMA to demonstrate how standard algorithms can be applied to a high performance computing system such that no special modifications are necessary to utilize a parallel-distributed processing environment.

1.2 Time Series, High Performance Computing: The DeLMA System

Processing the data poses many challenges including highly variable ambient noise conditions and a host of biological and anthropogenic sound sources [19-21]. The DeLMA system is designed to offer a framework that utilizes high performance computing. The system architecture is shown in Figure 1.

Figure 1. The DeLMA system concept of operations for high performance computing is designed for processing large acoustic datasets using detection classification algorithms.

Both sounds and machine learning algorithms are distributed using a map-reduce architecture, in which various compute nodes \( n_j \) are allocated CPU and memory. Each computer is represented as \( j \), distributed nodes, or \( d\text{nodes} \) noted as \( J \); processors on each unique computer are represented as \( i \), max processors, or \( p\text{nodes} \), noted as \( I \). Sounds
are automatically mapped across \( n_d \) such that equal parts of work exist between compute nodes and processors, \( J \) and \( I \) respectively. At the output, the system collects distributed data from the distributed resources, reducing the output to a single thread. The simplest processing model is serial \((j=J=1, i=I=1)\), which is limited in computation speed, unlike parallel or distributed models. Distributed processing is defined by \( j > 1 \), such that each node becomes a virtual single computer, referred to as a single process, multiple data [22]. Coordination happens between the nodes \( \{j = 1, 2, \ldots, J\} \) using a scheduling manager to coordinate tasks, where the \( j^{th} \) node serves as the head node. The DeLMA software can operate using serial, parallel or distributed models where execution in distributed mode can occur at the head node, or between a client computer. This means that one server cluster can exist in a remote location and clients can issue commands to process data remotely. For performance reasons, sound data and processing nodes are located on the same local network, reducing latency and timeout conditions.

1.3 Algorithm Performance: Large Datasets and the Stellwagen Right Whale Study

True system performance is measured using continuous data. Realistically, however, access to complete human-validated truth data is not practical, the data are difficult to analyse and in many cases only fractions of days or a handful of days have been fully labelled for a given dataset. The SBNMS is located in the coastal waters east of Boston, USA. It is one of the many sites in which scientists have been studying underwater sound and ocean ecology for several decades [30-33]. Work funded by the National Oceanic Partnership Program (NOPP) from 2007-2010 recorded a large dataset of continuous multi-channel sounds using an array of Marine Autonomous Recording Units (MARU’s). As a result of audio recordings from the MARU, Morano et. al. [23] analysed selected portions of data for up-calls during a three year period by hand, to document right whale daily and seasonal occurrence patterns. This study serves as one of the most complete records for when animals are resident in the SBNMS area. Important results from this study are shown in Figure 2. As shown, more calls happen in the late winter, early spring period, February through April, and significantly less between October and December. This reflects a well-known migration pattern that the majority of right whales undergo, migrating southward to waters off Florida for calving during the winter months and northward to waters off New England during the spring and summer.

Methods

2.1 Training, Validation and Testing Data,

Cornell University and Marinexplore in collaboration with Kaggle (www.kaggle.com) provided the international machine learning community with a large dataset consisting of right whale up-calls. Two training and validation data sets were hosted through the data science websites [24, 25]; the two sets consisted of 30,000 and 47,841 training samples, and 54,503 and 25,468 validation samples respectively. Each sample was a short sound segment, sampled at 2 kHz, which contained either signal or noise. Noise samples consisted of random acoustic events containing sounds from ships, weather or other marine mammals. All together training and validation sets consisted of 88 hours of sounds. The split of data for [24] consisted of roughly 50% right whale up-calls to 50% environment noise. The second source [25] contained equal proportions to a true day of environmental conditions, with 7,849 unique right whale up-calls, or roughly 10% right whale up-calls to 90% noise. These data were extracted from NOPP datasets taken from continuous recordings collected during the peak of right whale calling activity for that year which occurred during 28-31 March 2009 2009. A third data set was furnished by [26] which consisted of 4 days of fully labeled, continuous data.

Testing data were established by using a complete 12 month long, continuous dataset from the SBNMS, from 7 September 2008 – 2 October 2009. Seasonal data were captured using an array of sensors taken across five deployments, totalling 60,912 hours of sound. Since a MARU is limited in recording duration, multiple deployments were conducted, during the fall and spring seasons. Typically ten MARUs were deployed in an array, otherwise one or two MARU units were deployed. During spring 2009 (see Figure 2) the largest number of calls occurred in February, March and April; estimating, 1300, 4800 and 3300 calls, respectively. In January and May an estimated 150 and 200 calls respectively. In fall 2008, right whale activity for October, November and December
accounted for 300, 500 and 400 calls, respectively. Estimated average call activity for the fall 2008 and spring 2009 is summarized in Table 1.

2.2 Algorithms and HPC System

Referring to Figure 1, the HPC system was setup to run four, independent dnodes \( (J = 4) \). For each separate distributed node twelve ppnodes \( (I = 12) \) were used totalling 48 unique processing units. NOPP data from September 2008 to October 2009 used several deployments. To ensure balanced processing, each \( J \) unit handled a separate sound archive. The process of detecting and classifying right whale contact calls is consistent within the framework for the HPC system. Sound is broken down into 5-minute intervals, called sound pages. The sound pages are assembled as part of the mapping processing in the DeLMA system. Each page is presented to a processing node, and each node contains the algorithms shown in Figure 1.

According to Figure 1, three algorithms were used for detection-classification; Baseline, histogram of oriented gradient (HOG) and continuous region analysis (CRA.) All three algorithms were distributed across the processing nodes \( n \), using a simple, common interface. The Baseline algorithm has served as the standard processing method for Cornell, and two different algorithms are used. The first Cornell Baseline is used for audio clips and is based on [11]. The second Cornell Baseline is used for continuous data and based on [17].

The two new right whale algorithms, HOG and CRA, are based on computer vision technology, [27] and [5], respectively. Both methods competed in two, recent international data machine learning competitions and finished in the top 10%. These approaches contain separate detection and classification stages, and both algorithms use region based detection that relies on converting signals to images using the Fourier spectrogram transform. This is followed by a power-law algorithm [12], which is used for “de-noising”. The maximally stable external regions (MSER) process is applied to detect the spectrogram regions of interest (ROI) that possess distinguishing, invariant and stable properties [28]. The CRA is one of two classifiers used to detect and classify up-calls, as shown in Figure 1. Input is received as detection events, from the MSER stages. A total of 22 features are extracted from the continuous regions and fed into an artificial neural net (ANN). More details on the CRA approach are available in [5]. The HOG detection-classifier is designed to capture edges and gradients from the spectrogram image. More than 200 features were extracted from the edges and gradients, and an Adaboost [29] was employed to distinguish between up-calls and noise.

3 Results

Both the CRA and HOG algorithms were applied to the datasets provided by though the Kaggle competitions [24, 25]. For clip data, performance was measured using standard receiver operator curve (ROC) and the area under the curve (AUC). AUC ranges from 0.0 to 1.0, where 1.0 is perfect. The HOG and CRA methods performed with AUC scores of 0.96 and 0.93, respectively. The Cornell Baseline using [11] was 0.72. HPC system, Figure 1, was used to exercise the NOPP datasets for 10 channels for the baseline algorithm, where output is plotted using a seasonal data graph called a diel plot, Figure 3. The HOG and CRA algorithms were also applied to process three channels of MARU data (channels 7, 8 and 10) for the same NOPP datasets. Output event information is shown in diel plots, Figure 4. From the results of the HOG and CRA algorithms, the fall time period has roughly 50 and 75 calls per month for HOG and CRA, respectively. In contrast, an estimated 300 calls per month were detected for the spring months with the HOG algorithm and slightly more for CRA, around 350 calls per month. Results are summarized in Table 2.
Figure 2. Results from Morano et. al. [23] show the number of right whale calls taken from independent sensors. The dotted box shows the time period used in this study, starting in September 2008 through 2009.

Figure 3. Diel plot [34] showing baseline results for automatic Right Whale Detector. Grey areas portions indicate missing data. Plot show a high degree of false positives. Seasonal patterns should exist as described by the Morano study.

Figure 4. Diel plots [34] showing time during the day (x-axis) and day of the deployment on the (y-axis). Results from the HOG algorithm (left), CRA algorithm (right).

<table>
<thead>
<tr>
<th>Season</th>
<th>Mean number of right whale calls per month</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall 2008</td>
<td>400</td>
</tr>
<tr>
<td>Spring 2009</td>
<td>1950</td>
</tr>
</tbody>
</table>

Table 1. Estimated number of calls from human operator [23].

<table>
<thead>
<tr>
<th>Season</th>
<th>Baseline Average Calls (Based on 10 Channels)</th>
<th>HOG Average Calls (Based on 3 Channels)</th>
<th>CRA Average Calls (Based on 3 Channels)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall 2008</td>
<td>1500</td>
<td>50</td>
<td>75</td>
</tr>
<tr>
<td>Spring 2009</td>
<td>3000</td>
<td>300</td>
<td>350</td>
</tr>
</tbody>
</table>

Table 2. Estimated Number of Calls from HOG and CRA algorithms.
Discussion and Conclusions

The work herein combine high performance computing with novel methods for improving passive acoustic detection of a common type of long-range communication sound produced by the North Atlantic Right Whale. The performance of the proposed methods was evaluated using several large dataset consisting of sound clips and continuous sound recording in the Stellwagen National Marine Sanctuary. Three algorithms were successfully exercised on the High Performance Computing (HPC) system called DeLMA. Algorithms consisted of baseline methods, and two new approaches that rely on two standard machine learning technologies, histogram of oriented gradients (HOG) and continuous region analysis (CRA). Clips consisted of short audio samples, each being two seconds in duration, with a total combined duration of 88 hours. Area under the curve performance for the baseline methods was 0.72, while the new HOG and CRA algorithms achieved 0.96 and 0.93, respectively. The continuous datasets contain up to 10 channels of sound totalling 60,912 hours of recording. Since it would have been challenging and extremely time consuming to have complete ground truth for such large continuous datasets, validated detections from the Morano et al. [23] study served as a basis for estimating algorithm accuracy. The fall and spring months indicated that the animals, on average, had 400 and 1950 calls per month, respectively (a 1:5 fall:spring ratio). The baseline data showed an excess number of calling events on the diel plot, with little evidence of seasonal patterns. The HOG and CRA algorithms revealed a high degree of seasonal calling activity, with fewer in the fall and larger concentrations in the spring, similar to the Morano et al. [23] results. Comparing the seasonal calling ratios, baseline data resulted in a 1:2 ratio between fall and spring, HOG resulted in a 6:1 ratio between fall and spring and CRA resulted in a 1:4.7 between fall and spring.

These overall results demonstrate an effective method for detecting and classifying sounds in large-scale datasets. Baseline algorithms proved to be less accurate and less effective than the two new methods developed using some standard approaches taken from machine-vision. The application of high performance computing can allow scientists to begin studying these large archives of passive acoustic data, while at the same time providing a higher level of accuracy for bioacoustics signal detection. Challenges arise concerning the ability to validate and accept output from these complex systems with minimal human labelling, as discussed in this study. These techniques can also be expanded to other species, while relying on the convenience of having scalable, complex, distributed systems capable of exploring large acoustic datasets. Ultimately, the HPC system, along with accurate data mining technologies, will allow researchers the ability to efficiently process large datasets in order to study the spatial-temporal patterns of acoustically active species over ecologically meaningful spatial, temporal and acoustic scales.

References


