Audio Forensics for Maritime Recognition

Abhishek Kar
akar1@andrew.cmu.edu

Yi-Ching Lee
yichingl@andrew.cmu.edu

Ramesh Oswal
ramesho@andrew.cmu.edu

Andrew Tracy
atracy@andrew.cmu.edu

1 Problem Description

The United States Coast Guard (USCG) regularly receives distress calls via VHF radio that prompt a search and rescue (SAR) response. Out of these calls, a small percentage of them are hoax or prank calls (out of 16,343 responses in 2016, 161 were hoaxes [8]). Until it is determined that the call is a hoax, the USCG has a responsibility to respond as if there is an actual emergency.

Even though the number of hoax calls is small, the cost of the resulting response can be as high as $250,000 [2]. Even more importantly, the lives of USCG service people are put at risk every time a SAR effort is launched and critical assets are diverted from real emergencies.

One way to help identify hoax calls is to analyze the audio signal to determine the caller's environment. Specifically, by identifying machinery or maritime objects that can be heard in the background of the call, the USCG can use this information to locate the caller. For instance, if the sound of an engine in the background is determined to be from a particular model of a plane, the Coast Guard can cross reference known vehicles and roughly determine the caller's location. The same concept could be extended to many other maritime objects, such as animals, boats, or helicopters.

2 Literature Review

Analysis of hoax calls has been studied by Prof. Singh fairly extensively with a focus on caller identification [17].

There are several different appropriate methods to represent an audio signal so that it can be used as an input to classification. Some methods of feature extraction include mel-frequency cepstral coefficients (MFCCs), spectrograms, statistical functionals, and temporal dynamics [3]. The best performing signal processing pipeline usually involves a combination of these techniques. In the case where MFCCs fails, a combination of CNN and augmenting the data provides an alternative solution [15].

For classification of data, there are a wide variety of existing methods. For audio analysis, the classification tools include k-nearest neighbors [12], Support Vector Machines[11] [7] [6], and Convolutional Neural Nets [10]. Much of the recent research has focused on applying neural nets to classify a scene/environmental noise [3][14][15].

There are a number of existing methods for separating audio signal components, including non-negative matrix factorization [17], principle component analysis, and independent component analysis [18]. These methods vary in their complexity and optimality of results. There are commercial tools available for performing the same kind of analysis for music applications [16][9][13].

3 Proposed Solution

We propose a solution that will automatically analyze an audio signal, separate out different audio sources, and identify a maritime object in the components of the signal. As a stretch goal, we would also like the solution to provide other characteristics, such as the distance and speed of the object.

There will be three main components of the solution:

1. Data Collection - Compile a detailed reference table of the sound-emitting objects that can provide useful information in a forensic maritime setting. Do a study of the boats, planes, and helicopters we will be focusing on. The scope includes marine vessels (specifically boats and aircrafts) commonly found on shores through out the United States. Specifically, looking at details such as the engine specifications and sound pressure levels. Also research what other sorts of sounds might occur in a maritime environment. Examples include animals (birds, sea lions, whales) and industrial activities (oil drilling and air guns/seismic surveying) [1]. Scrape audio from various web sources such as Youtube [5] and Google's AudioSet [4] and perform labeling. Another source of audio data are online databases of birdcall [19], such as xeno-canto, that would, at the very least, help us detect the presence of oceanic birds.

2. Object Detection and Identification - Focus on building a robust classifier on real data collected from online sources. Apply ML techniques often used for audio source detection to identify maritime objects of interest in each pre-processed audio clip. Begin by trying to distinguish between general categories of maritime objects, such as helicopter and boats. After that task has been accomplished, attempt to distinguish between different variants of boats, etc. Possible classification tools include neural networks, bag-of-words with acoustic unit descriptors, and support vector machines.

3. Audio Segmentation - Use techniques such as non-negative matrix factorization (NMF) to segment foreground (boats/helicopters/other objects of interest) sounds from the background sounds. Automatically segregate audio from raw sources and identify noise from maritime objects of interest.

The performance metric of the classifier/detector will be measured by its accuracy. Accuracy in this context is defined by the fraction of audio clips which the ML algorithm predicts correctly. The performance metric of the audio segmentation will be done aurally (manually) as a start. Calculating the SNR (Signal-Noise Ratio) of the segmented audio may also act as a good performance metric.

References

[1] *Ocean noise and marine mammals.* National Academies Press, 2003.

[2] To thwart hoax distress calls, the coast guard r&d center has partnered with carnegie mellon researchers. *Command, Control, and Interoperability Center for Advanced Data Analysis*, Nov 2016.

[3] Wei Dai. Acoustic scene recognition with deep learning.

[4] Google. Audioset
https://research.google.com/audioset/about.html.

[5] Google. Youtube-8m
https://research.google.com/youtube8m/.

[6] Roger B. Grosse, Rajat Raina, Helen Kwong, and Andrew Y. Ng. Shift-invariance sparse coding for audio classification. *CoRR*, abs/1206.5241, 2012.

[7] Guodong Guo and Stan Z Li. Content-based audio classification and retrieval by support vector machines. *IEEE transactions on Neural Networks*, 14(1):209–215, 2003.

[8] Loretta Haring. Research, development, test and evaluation spotlight: Search and rescue hoax calls. *Coast Guard Compass*, Jul 2017.

[9] ISSE. Interactive source separation editor
http://isse.sourceforge.net/.

[10] Honglak Lee, Peter Pham, Yan Largman, and Andrew Y. Ng. Unsupervised feature learning for audio classification using convolutional deep belief networks. pages 1096–1104, 2009.

[11] Stan Z Li and Guo-dong Guo. Content-based audio classification and retrieval using svm learning. In *First IEEE Pacific-Rim Conference on Multimedia, Invited Talk, Australia*, 2000.

[12] Lie Lu, Hao Jiang, and HongJiang Zhang. A robust audio classification and segmentation method. pages 203–211, 2001.

[13] Melodyne. Celemony - a tool like no other.
http://www.celemony.com/en/melodyne/what-is-melodyne.

[14] Karol J Piczak. Environmental sound classification with convolutional neural networks. In *Machine Learning for Signal Processing (MLSP), 2015 IEEE 25th International Workshop on*, pages 1–6. IEEE, 2015.

[15] Justin Salamon and Juan Pablo Bello. Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal Processing Letters*, 24(3):279–283, 2017.

[16] Yann Salaün, Emmanuel Vincent, Nancy Bertin, Nathan Souviraà-Labastie, Xabier Jaureguiberry, Dung T. Tran, and Frédéric Bimbot. The Flexible Audio Source Separation Toolbox Version 2.0. ICASSP, May 2014. Poster.

[17] Rita Singh, Joseph Keshet, and Eduard Hovy. Profiling hoax callers. pages 1–6, 2016.

[18] Emmanuel Vincent. Blind audio source separation a review of state-of-the-art techniques
https://www.irisa.fr/metiss/members/evincent/keynoteicarn05.pdf.

[19] xeno canto. Sharing bird sounds from around the world
http://www.xeno-canto.org/about/xeno-canto.