Aquaticus Datasets

August 2018

Michael Benjamin, mikerb@mit.edu
Michael Novitzky, novitzky@mit.edu
Paul Robinette, paulrobi@mit.edu
Department of Mechanical Engineering, CSAIL
MIT, Cambridge MA 02139

1 Aquaticus Datasets Overview **1**
1.1 Components of the Datasets 1
1.2 Tools for Playing Back and Analyzing Datasets 2
1.3 Where to Find Existing and Future Datasets 3

1 Aquaticus Datasets Overview

An Aquaticus *dataset* is a collection of data files from a single Aquaticus competition (typically 10 minutes). Each dataset is a bundle of several files, collected from all MOKAI (human) platforms, and all robotic platforms. A typical Aquaticus competition will have four human particpants and four robot participants. This document will discuss:

- The dataset components and their formats
- The tools for playing back or analyzing the datasets
- Where to find the existing and future datasets

There are several tools for managing data files, and further tools are under development for great ease in analyzing the data.

1.1 Components of the Datasets

Each four-on-four Aquaticus competition involves nine entities, each logging their own sets of data:

- Four MOKAI (human) platforms
- Four Robot platforms (Clearpath Heron USVs)
- The shoreside control station

By example, the structure of the competition `2018_08_10_C051`, is shown below, showing just the first two levels of the folder.

```
2018_08_10_C051/
    RECAP_VIDEO_C051.mov
    README
    mokai/
      red_one/
      red_two
      blue_one
      blue_two
    heron/
      red_three/
      red_four
      blue_three
      blue_four
    shoreside/
```

This folder will contain:

- Four video files of human players (MOKAI)
- Four alog files for human players (MOKAI)
- Four alog files, one for each robot
- Several audio files per human player (MOKAI)

The alog files are the standard log file format for the MOOS middleware and the MIT MOOS-IvP autonomy project has several tools for analyzing, modifying and replaying data in this format. For more on the alog tools for aquaticus see:
http://oceanai.mit.edu/aquaticus/datatools

1.2 Tools for Playing Back and Analyzing Datasets

The Alog Toolbox contains several command line tools for manipulating, analyzing or pruning raw data files, alog files, produced by the `pLogger` utility which is part of MOOS. To visually render or re-play the competition, the `alogview` tool is the tool you want. This is distributed as part of the MOOS-IvP public codebase and is launched from the command line.

For example, in the `2018_08_10_C051` dataset, `alogview` can be launched from the command line as follows:

```
$ cd 2018_08_10_C051/
$ alogview */*.alog
```

The term `"*/*.alog"` will match to all nine alog files, each with the file suffix `".alog"`, one level down from the main, top-level directory. This will open `alogview` with all eight vehicles, similar to the below recap video found on the `C051` splash page: http://oceanai.mit.edu/aquaticus/datasets/2018_08_10_C051:

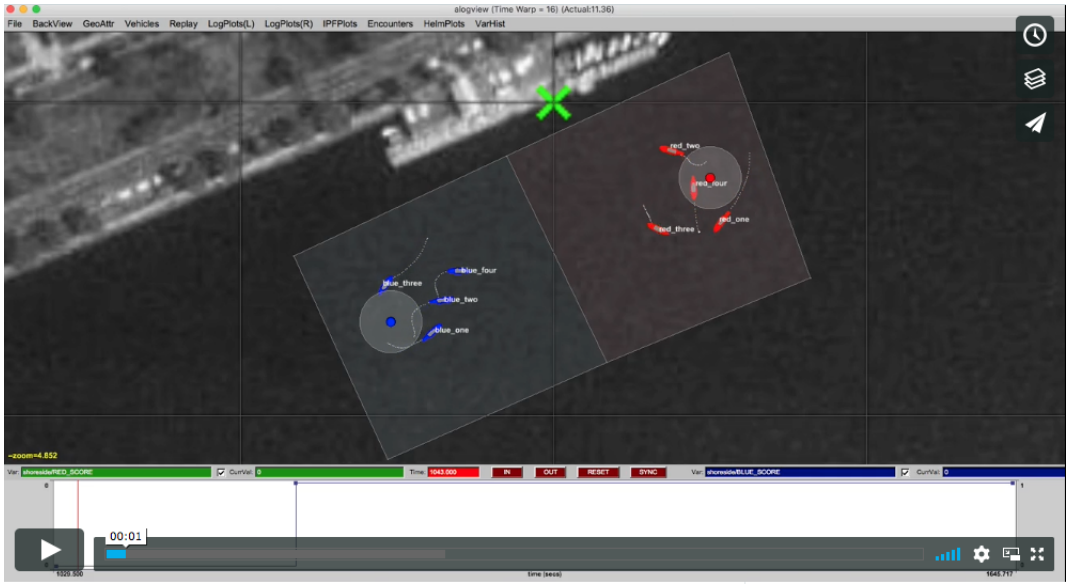

Figure 1: The time-lapse recap video for all eight platforms in competition 2018_08_10_C015, as seen by launching alogview on the alog files for all eight vehicles.
video:(0:58): https://vimeo.com/289136465

1.3 Where to Find Existing and Future Datasets

Currently there are ten datasets on line for public download. The portal for these datasets is:

http://oceanai.mit.edu/aquaticus/datasets

Visiting this portal will provide a link to each dataset sorted by date. Each dataset has its own splash page with the overview of the data, links to download the data, and typically a recap video like the one above in Figure 1 which is a time-lapse replay of the data using alogview.

Minimally there is a single tar file to download all the data. This tar file represents a subset of the raw data, prune by hand to remove any unintended data or video or to clip data down to just the window of the competition. The raw data is archived elsewhere.

In the future, each dataset will offer a reduced size tar file where the alog files are compressed, the video files are reduces to a low-res video, and the wav files are compressed to mp3 format. Very likely these low-res tar files will be 10% of the original tar file size and allow for quicker download and analysis. The full version tar files will always be available.