

An Automated Visual Event Detection System for Cabled Observatory Video

Danelle E. Cline, Duane R. Edgington
Jérôme Mariette
Monterey Bay Aquarium Research Institute
7700 Sandholdt Road
Moss Landing, CA 95039 USA
{dcline,duane}@mbari.org
jerome.mariette@gmail.com

Abstract-- The permanent presence of underwater cameras on cabled observatories, such as the Victoria Experimental Network Under the Sea (VENUS) and Eye-In-The-Sea (EITS) on Monterey Accelerated Research System (MARS), will generate precious data that can move forward the boundaries of understanding the underwater world. However, sightings of underwater animal activities are rare, resulting in the recording of many hours of video with relatively few events of interest. Towards this end, an Automated Visual Event Detection System is in development at the Monterey Bay Aquarium Research Institute (MBARI) to address the problem of analyzing cabled observatory video. This paper describes the overall design of the development of a system to process video data and enable science users to analyze the results.

I. INTRODUCTION

Cabled observatory communities have the potential to unlock important discoveries with new and existing cabled cameras. Yet the burden of video management and analysis often requires reducing the amount of video recorded and later analyzed. Sometimes enough human resources do not exist to analyze the video; the strains on human attention needed to analyze video demand an automated way to assist in video analysis.

To help address this problem, the Automated Visual Event Detection (AVED) software has been under development at the Monterey Bay Aquarium Research Institute (MBARI) for the past 5 years and it has shown promising results when compared with human observers in 90% of single frame comparisons in MBARI Remotely Operated Vehicle Dives [1].

The first applications of AVED to cabled observatories are a deepwater video observatory for unobtrusive monitoring called the Eye-In-The-Sea (EITS) instrument [2], and a modified version of AVED is currently being developed for a proof-of-concept system to integrate with the Victoria Experimental Network Under the Sea (VENUS) observatory.

This paper first gives a technical overview of the AVED software, followed by a discussion of the AVED data flow for the VENUS and MARS observatories applications. Lastly, AVED future work is discussed.

II. AVED SOFTWARE OVERVIEW

The AVED software is a collection of custom software written in C++ and Java and designed to work on Linux enabled computers. This automated system for detecting visual events can be run three ways: through a Web Service, through a Condor managed pool of AVED-enabled computer, or locally on a single computer. The collection of software also includes a graphical user interface used to edit AVED results, or setup and execute AVED processing. Fig. 1 depicts the AVED software layers.

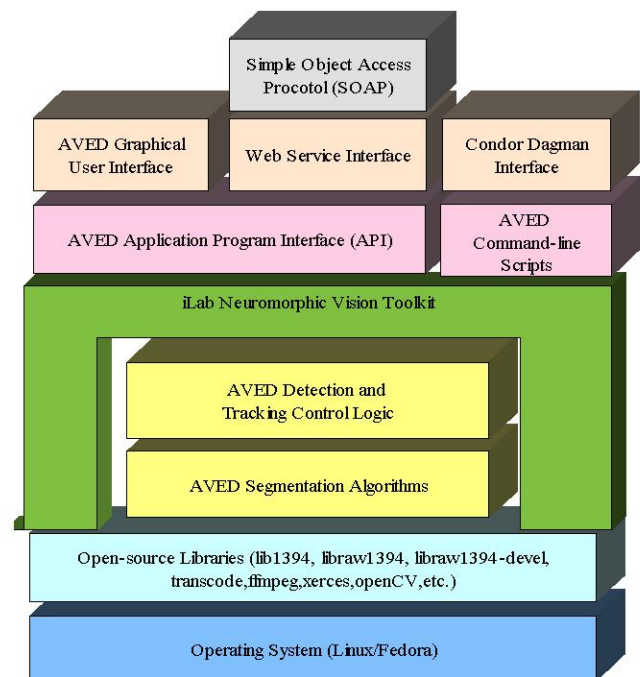


Figure 1 AVED Software Layers

A. The iLab Neuromorphic Vision Toolkit

Central to the AVED software design is the iLab Neuromorphic Vision C++ Toolkit developed by University of Southern California [3]. This toolkit provides many of the algorithms used in the AVED software including the neuromorphic modeled algorithms used in the detection component of AVED, and many of the basic image processing algorithms used. AVED directs attended locations in video by using a neuromorphic software model based on the human vision system from this toolkit. In this model, video frames are decomposed into specific feature maps that are combined into a unique saliency map as depicted in Fig. 2. This saliency map is then scanned to determine the most salient locations. The candidate salient locations are then segmented from the scene and tracked over multiple frames, resulting in the detection of visual events. For more information on the AVED algorithms see [4]

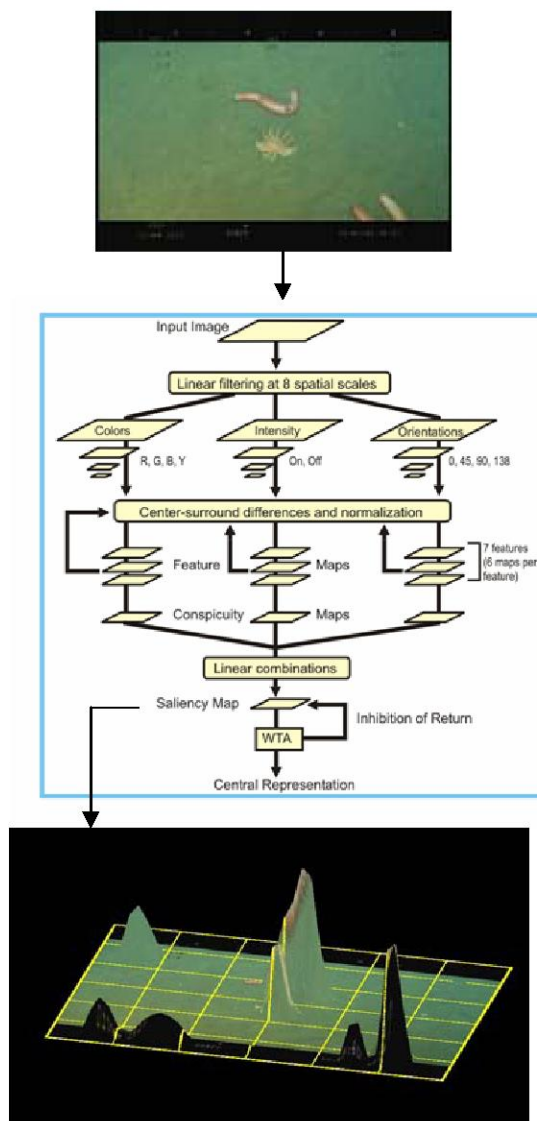


Figure 2 Saliency map from the iLab toolkit warped onto a 3-D map. Peaks in the map show points of high visual attention where the Rathbunaster and Leukothele are in the center image.

B. Parallel Program for Performance Improvement

The AVED software can execute on a single processor. AVED is also designed to execute on a parallel machine to support applications with large volumes of recorded video. For these high volume applications, performance gain was accomplished by functionally decomposing the processing algorithms into approximately equal execution time stages that were then executed in a pipeline fashion using a Message-Passing Interface (MPI) library called MPICH2. Using an optimized library in the Neuromorphic Vision Toolkit for the saliency map computation, we were able to further improve performance. For NTSC video recorded at 30 frames per seconds, processing rates of 5 frames per seconds were achieved using a 16 CPU Xeon™ Beowulf cluster. Processing rates varied slightly depending on the video content. Busy scenes generally took longer to process because of increased detection and tracking overhead.

C. Segmentation and Tracking Algorithms for Observatory Video

Underwater video poses some unique challenges given the low and often non-uniform lighting. The camera movement also poses challenges for tracking. To address these challenges, AVED contains a collection of different segmentation and tracking algorithms.

Two general categories may be considered in observatory video: cameras with little or no pan and tilt, and cameras with pan and tilt. For these two categories, through experiment we found certain paired segmentation and tracking options that improve the number of events detected.

In the case of a fixed observatory camera with minimum pan and tilt or zoom movement, an image average from a running image cache is used with a graph cut-based [5] algorithm to extract foreground objects from the video. Only pixels determined to be background versus detected foreground objects, are included in this image cache, thereby removing the objects weight on the background computation. To track visual events, a simple nearest neighbor tracking algorithm is used.

In the case of a fixed camera with pan and tilt and zoom movement, a segmentation algorithm based on an adaptive threshold and Otsu's method is used [6]. This method begins by building a histogram based upon the image, and then the threshold is determined by the value that maximizes the between-class variance of the gray level histogram. To track visual events, a Kalman filter is used for this case.

III. AVED USER INTERFACE

The requirements for the AVED user interface were developed through discussions with potential science users and we expect further improvements to this interface with feedback from the EITS science users during the 2008 experiment on MARS.

The AVED interface is used to preview AVED metadata, setup processing options, and execute processing. The AVED processing XML metadata is imported into the interface, and used in concert with the input video to preview or edit the

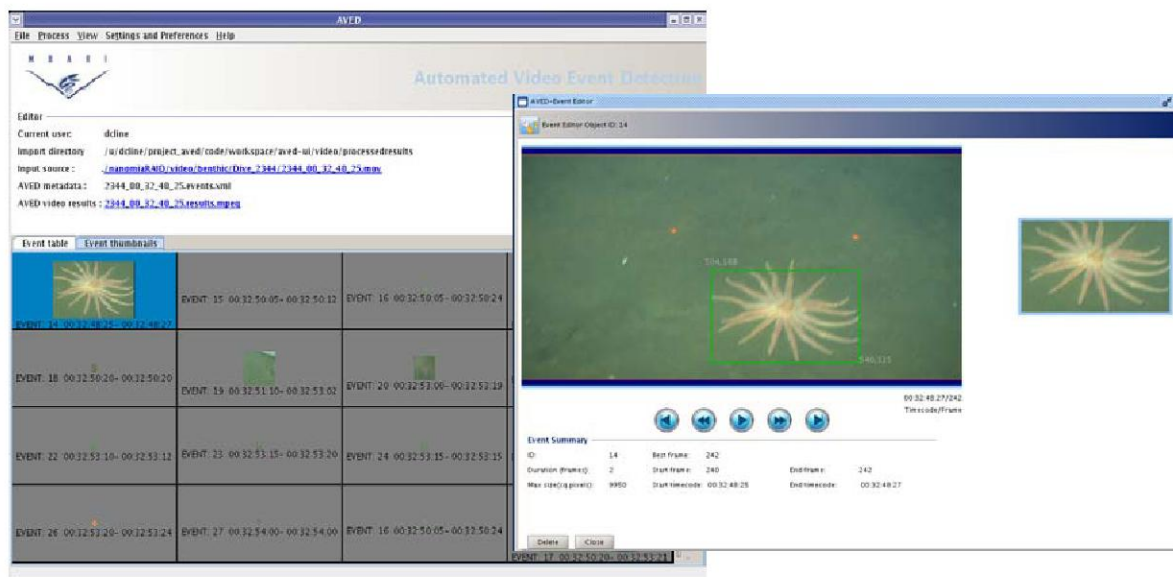


Figure 3 AVED User Interface: Event thumbnail and Event Editor Views

results. Screen captures of this interface are shown in Fig. 3. This same graphical interface can be used as a simple way to execute AVED. AVED can also be executed as a stand-alone series of steps using command line scripts for applications that do not need an interface.

D. Event Editing

The event editing capability of the interface includes a quick way to preview and edit or delete results either through an Event thumbnail view, Event table view, or an Event editor that can sequence through an interesting event frame-by-frame. This provides a way for the science user to preview relevant data, and delete false detections or other non-interesting events.

E. Processing Setup Options

AVED processing options, such as the segmentation and tracking options can be setup through the GUI. The options are written to a configuration XML file that is subsequently used in processing. One could, for example, setup and save setup options for benthic versus midwater footage which may require different processing algorithms and reuse these options.

IV. AVED DATA FLOW FOR EITS AND VENUS

The Eye-In-The-Sea (EITS) instrument planned for deployment on the MARS observatory in 2008 will have a video stream captured 24 hours a day 7 days a week over the course of a three months experiment. Video from this continuous stream will be analyzed for scientific research by trained science staff.

Fig. 4 shows the data flow of the EITS system on MARS. The MARS high bandwidth network enables digital video to be transmitted to shore. This digital video stream is then captured on shore and processed using a pool of compute resources, including an 8-node Beowulf cluster at MBARI. Processing is initiated by a simple periodic script that looks for newly recorded video clips. Recorded clips are then processed with AVED using a Condor [7] pool of computers. The AVED

software then finds interesting events, saves these events to a metadata XML file, and then a science annotator edits events in the AVED user interface for false detections or other non-interesting events. The edited XML metadata are then imported into a database for use with the Video Annotation and Reference System (VARs) that was developed at MBARI.

The data flow for VENUS is similar to EITS, except AVED metadata XML files are imported into the Data Management and Archive System (DMAS) currently in development at NEPTUNE Canada DMAS system [8]. Also, video data from VENUS are collected at a lower rate of 1 minute of video every hour, and thus can be processed on a single computer instead of a Condor pool.

V. TOWARD A GRID-ENABLED AVED

Looking to the future, we see Grid technology as an important element to addressing the problem of accessing and processing large video data sets. It's unknown how many digitized video data sets exist in the oceanographic community, but we suspect that many remain under analyzed due to lack of good tools or human support. Tools such as Montage [9], which provide a Grid-enabled toolkit to help the Earth and space astronomy communities process large distributed image data, serve as a model to the oceanographic community with its growing video data sets. We see our first steps towards using Grid aware tools for scheduling, workflow execution, and enabling AVED as a web service as a logical progression towards developing a Grid-enabled version of AVED.

A. Scheduling and Workflow Execution

Processing video with AVED is a compute and data intensive activity that involves a series of processing steps. These can be described as a workflow, where each step in the workflow has data input, output and/or control dependencies and each step in this workflow doesn't necessary need to execute on the same computer.

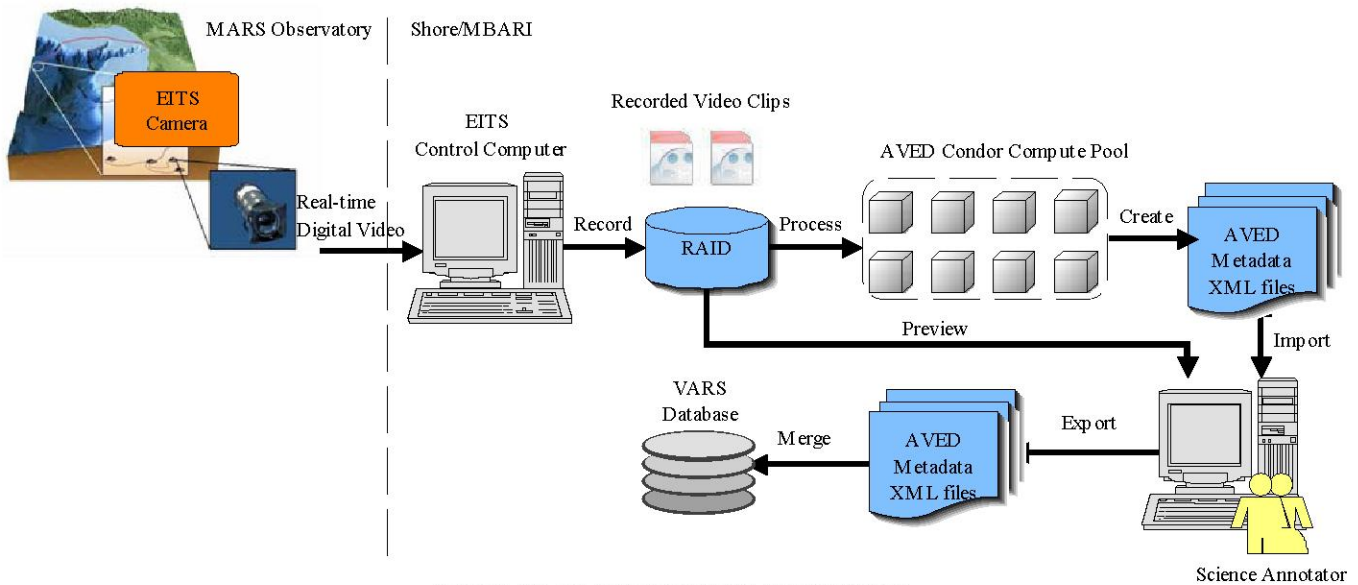


Figure 4 The Eye-In-The-Sea (EITS) AVED data flow

To execute and manage this workflow, we use a specialized workload management system for compute and data intensive jobs called Condor developed by the University of Wisconsin Madison. Condor provides scheduling queuing and resource management. Condor also provides powerful capabilities like matching job request with machines. Condor is planned for use in the EITS experiment in conjunction with the parallel AVED program to process the large volume of video generated.

These capabilities allow for users to harness the power of their computing resources by installing AVED on as many machines as needed and configuring Condor policies to manage the machines as desired. For example, Condor can be configured to only use desktop machines when the keyboard and mouse are idle, or it can be configured to always use a pool of dedicated machine such as a Beowulf cluster.

B. Web service for DMAS Integration

To interface to the VENUS data management system the Data Management and Archive System (DMAS), an interface to AVED in the form of a Web service is provided. The AVED Web service is described by an Extensible Markup Language (XML) format called Web Services Description Language (WSDL). The AVED WSDL format describes, for example, the location of the input video file to process. Communication between DMAS and AVED is done through XML messages using the Simple Object Access Protocol (SOAP) standard. The DMAS system initiates AVED processing by sending an XML request message to the AVED Web Service.

VI. SUMMARY AND FUTURE WORK

A system for detecting and visual events in an observatory using the AVED software is in development and planned for deployment on the MARS and VENUS observatories in 2008. This automated system for detecting visual events includes a collection of custom and open source software that can be run

three ways: through a Web Service, through a Condor managed pool of AVED enabled compute servers, or locally on a single computer. The collection of software also includes a graphical user interface to preview or edit detected results and to setup processing options. To optimize the compute-intensive AVED algorithms, a parallel program has been design for high-data rate applications like the EITS instrument on MARS.

Preliminary work has been done on a computer classification program used in conjunction with AVED to classify benthic species [10]. Future work includes further improvements to this classification software and full integration with the AVED software.

To fully enable AVED for grid execution, investigation into various Grid software will be needed, including Condor-G, the Globus toolkit, and software package called Pegasus that maps workflow execution onto Grids.

ACKNOWLEDGMENT

We thank the David and Lucille Packard Foundation for their continued generous support. We thank the University Of Victoria, Canada and the National Science Foundation for their support for Jérôme Mariette's work to integrate AVED with the VENUS observatory. This project originated at the 2002 Workshop for Neuromorphic Engineering in Telluride, Colorado, USA in collaboration with Dirk Walther, California Institute of Technology, Pasadena, California, USA. We thank Karen Salamy for her technical assistance and the MBARI video lab staff for their interest and input on the AVED user interface. We thank Edith Widder, Erika Raymond, and Lee Frey for their support and interest in using AVED for the EITS instrument.

REFERENCES

-
- [1] D. Walther, D., D. Edgington, K. A. Salamy, M. Risi, R.E. Sherlock, and Christof Koch, "Automated Video Analysis for Oceanographic Research", *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, demonstration, Madison, WI, 2003.
 - [2] E.A. Widder, B.H. Robison, K.R. Reisenbichler, S.H.D. Haddock, "Using red light for in situ observations of deep-sea fishes", *Deep-Sea Research* 1 52 2077-2085.
 - [3] iLab Neuromorphic Vision C++ Toolkit at the University of Southern California, available from World Wide Web: <http://ilab.usc.edu/toolkit>.
 - [4] D. Walther, D. R. Edgington, C. Koch, "Detection and Tracking of Objects in Underwater Video", *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, demonstration, Washington, D.C., 2004.
 - [5] N. Howe & A. Deschamps. Better Foreground Segmentation Through Graph Cuts, Tech report, available from the World Wide Web, <http://arxiv.org/abs/cs.CV/0401017>, 2004.
 - [6] Otsu, N., "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 9, No. 1, 1979, pp. 62-66.
 - [7] Condor High Throughput Computing, The University of Wisconsin, Madison, available from the World Wide Web, <http://www.cs.wisc.edu/condor/>.
 - [8] C.R. Barnes, M.M.R. Best, B.D. Bornhold, S.K. Juniper, B. Pierenne, P. Phibbs, "The NEPTUNE Project –a cabled ocean observatory in the NE Pacific: Overview, challenges and scientific objectives for the installation and operation of Stage I in Canadian waters", *IEEE International Symposium on Underwater Technology* 2007.
 - [9] J. C. Jacob, D. S. Katz, T. Prince, G. B. Berriman, J. C. Good, A. C. Laity, E. Deelman, G. Singh, and M. Hui Su, "The Montage Architecture for Grid-Enabled Science Processing of Large, Distributed Datasets", *Earth Science Technology Conference*, 2004.
 - [10] Edgington, D.R., Cline, D.E., Davis, D., Kerkez, I., and Mariette, J., Detecting, Tracking and Classifying Animals in Underwater Video (060331-207). *MTS/IEEE Oceans 2006 Conference Proceedings*, Boston, MA September 2006. IEEE Press