

AN AUTOMATED VISUAL EVENT DETECTION SYSTEM FOR CABLED OBSERVATORY VIDEO

Danelle E. Cline, Duane R. Edgington

Monterey Bay Aquarium Research Institute, 7700 Sandholdt Road, Moss Landing, CA 95039, USA

Jérôme Mariette

Monterey Bay Aquarium Research Institute, 7700 Sandholdt Road, Moss Landing, CA 95039, USA

Keywords: Cabled Observatory Video, Automated Visual Event Detection, AVED, Underwater Video Processing.

Abstract: This paper presents an overview of a system for processing video streams from underwater cabled observatory systems based on the Automated Visual Event Detection (AVED) software. This system identifies potentially interesting visual events using a neuromorphic vision algorithm and tracks events frame-by-frame. The events can later be previewed or edited in a graphical user interface for false detections, and subsequently imported into a database, or used in an object classification system.

1 PROJECT OVERVIEW

Ocean observatories and underwater video surveys have the potential to unlock important discoveries with new and existing camera systems. Yet the burden of video management and analysis often requires reducing the amount of video recorded and later analyzed. To help address this problem, the Automated Visual Event Detection (AVED) software has been under development for the past several years. The system has shown promising results when applied to video from video surveys conducted with video cameras on Remotely Operated Vehicles (Walther, 2003, 2004). Here we report the system's extension to cabled-to-shore observatory cameras.

Among the first applications of AVED to cabled observatories, include a deepwater video instrument called the Eye-In-The-Sea (EITS) instrument (Widder, 2005) to be deployed on the Monterey Accelerated Research System (MARS) observatory test bed in early 2008. Additionally, 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 an overview of the AVED system in general, followed by a discussion of the AVED system for the EITS experiment, and lastly,

preliminary results and future work is discussed.

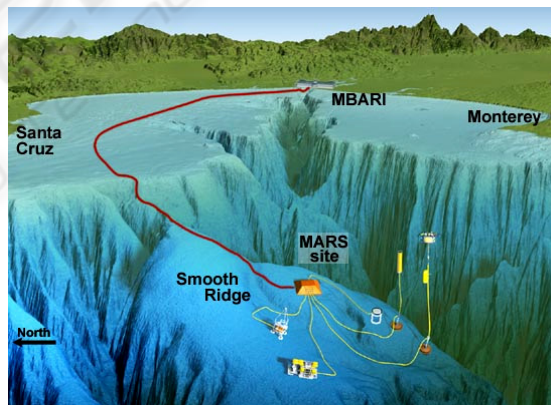


Figure 1: A 3-D perspective of the MARS cabled-to-shore observatory site on Smooth Ridge, at the edge of Monterey Canyon.

2 AVED OVERVIEW

The AVED software is a collection of custom software written in C++ and Java designed to work on Linux enabled computers. The collection of software includes a graphical user interface used to edit AVED results. To manage high compute demand applications, an optimized version of AVED for parallel execution runs on our 8-node Racksaver

rs1100 dual XEON 2.4 GHz servers configured as a Beowulf cluster.

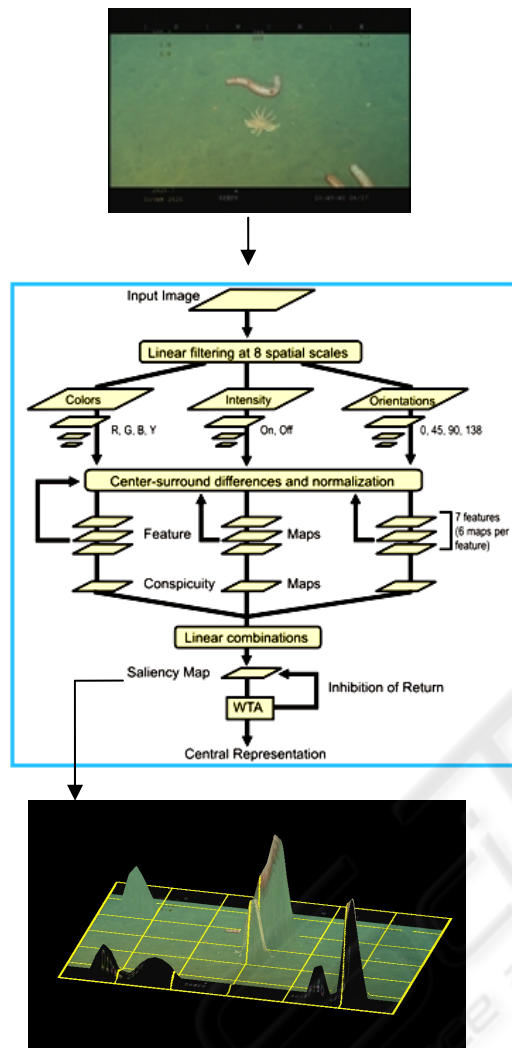


Figure 2: Saliency map from the iLab toolkit warped onto a 3-D map. Peaks in the map show points of high visual attention in the center of the image where the Rathbunaster and Leukothele are.

2.1 Image Pre-processing

Underwater video often contains artifacts like lens glare, visual obstructs such as instrumentation equipment, or introduced artifacts such as time code video overlays. Some simple algorithms are employed to remove these artifacts. To remove the lens glare and transient equipment, a simple background subtraction scheme is used whereby the average from a running image cache is subtracted from the input image. To remove time code overlays

or stationary equipment in a scene, a simple mask is applied that removes the areas before the detection and tracking steps.

2.2 Detection and Tracking

2.2.1 Neuromorphic Event Detection

Central to the AVED software design is the detection step, where candidate events are identified using a neuromorphic vision algorithm developed by the Itti and Koch (Itti, 1998). In the saliency model, each input video frame is decomposed into seven channels (intensity, contrast, red/green and blue/yellow double color opponencies, and four canonical, spatial orientations at six spatial scales, yielding 42 feature maps. After iterative spatial competition for saliency within each map, maps are then combined to form a unique saliency map. This saliency map is then scanned for the topmost salient locations using a winner-take-all neural network.

Figure 2 illustrates an example saliency map from the iLab toolkit warped onto a 3-D map for a single underwater video frame. Peaks in the map show points of high visual attention. Objects are then segmented around these peak points and then tracked frame-by-frame to form a visual event. Events that can be tracked over several frames are stored as “interesting”; otherwise they are designated as “boring” and removed from tracking.

This AVED saliency-based detection algorithm and many of the basic image processing algorithms used in AVED are provided by the iLab Neuromorphic Vision C++ from the University of Southern California.

2.2.2 Fixed Camera Object Tracking

In the case of a fixed observatory camera with minimum pan and tilt or zoom movement such as the EITS camera, an average from a running image cache is used with a graph cut-based (Howe, 2004) 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. This segmentation scheme results in better segmentation of faint objects. To track visual events, a nearest neighbor tracking algorithm is used.

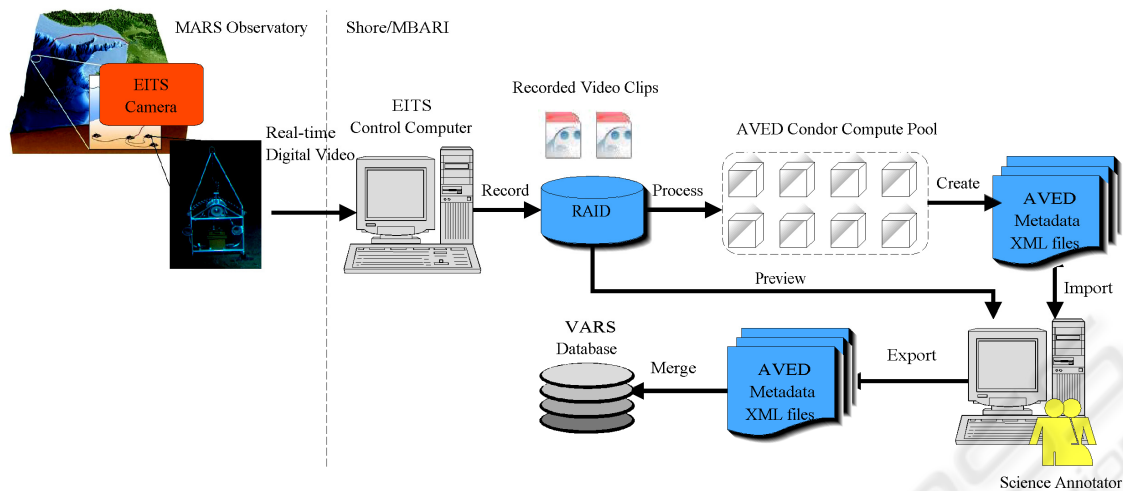


Figure 3: EITS-AVED Data Flow.

3 AVED DATA FLOW FOR EITS

Figure 3 shows the end-to-end data flow for the EITS camera 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 into individual clips. To execute and manage this workflow, we use Condor, a specialized workload management system for compute and data intensive jobs developed by the University of Wisconsin Madison. (<http://www.cs.wisc.edu/condor/>). Condor provides scheduling queuing and resource management. Video clips are then submitted for processing in a pool of Condor-enabled compute resources, including an 8-node, 16 CPU Beowulf cluster. The AVED software finds interesting events, saves these events to a metadata XML file. A science annotator then 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 forms a catalogue of the clips as well as the annotations of interesting events by AVED.

4 RESULTS

Figure 4 shows the comparison of EITS video processed by AVED with professional annotation for 172 previously recorded video clips of varying lengths from 1 to 20 minutes. A high rate of

detection and a low rate of false detection and of misses are evident. The automated system correctly identified video containing interesting events (Correct Positive) 81% as well as video not containing events (Correct Negative) 6% with few false alarms (False Positive) 12% and very few misses of video clips with one or more interesting events (False Negative) 1%.

5 CONCLUSIONS

A system for detecting and visual events in an observatory using the AVED software is in development and planned for deployment on the MARS observatory in 2008. This automated system for detecting visual events includes customized tracking and detection algorithms tuned for underwater still cameras. Analysis of video clips from previous deployments of the Eye-in-the-Sea camera system processed by AVED demonstrate its potential to correctly identify events of interest, as well as clips of low interest that can be skipped.

6 FUTURE WORK

Preliminary work has been done on a computer classification program used in conjunction with AVED to classify benthic species (Edgington, 2006). Future work includes further improvements to this classification software and full integration with the AVED software.

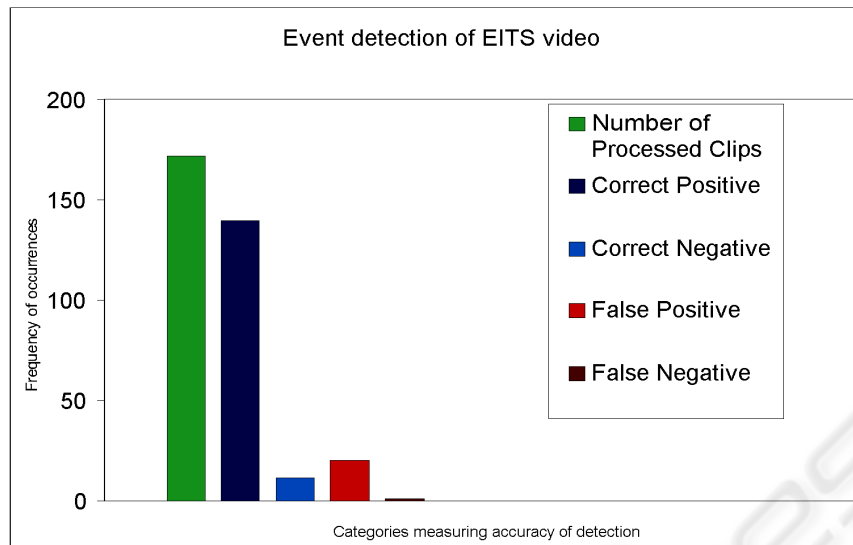


Figure 4: The EITS AVED detection results compared with professional annotators.

ACKNOWLEDGEMENTS

We thank the David and Lucile Packard Foundation for their continued generous support. 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

- Condor High Throughput Computing, The University of Wisconsin, Madison, viewed 10 August, 2007, <<http://www.cs.wisc.edu/condor/>>.
- Edgington, D.R., Cline, D.E., Davis, D., Kerkez, I., and Mariette, J. 2006, 'Detecting, Tracking and Classifying Animals in Underwater Video', in *MTS/IEEE Oceans 2006 Conference Proceedings, Boston, MA, September*, IEEE Press.
- Howe, N. & A. Deschamps, 2004, 'Better Foreground Segmentation Through Graph Cuts', technical report, viewed 18 September, 2007, <<http://arxiv.org/abs/cs.CV/0401017>>.
- iLab Neuromorphic Vision C++ Toolkit at the University of Southern California, viewed 18 September, 2007, <<http://ilab.usc.edu/toolkit/>>.

- Itti, L., C. Koch, and E. Niebur, 1998. 'A model of saliency-based event visual attention for rapid scene analyses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(22): p 1254-1259.
- Otsu, N. 1979, 'A Threshold Selection Method from Gray-Level Histograms', *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 9, No. 1, pp. 62-66.
- Video Annotation and Reference System (VARs), viewed 12 November, 2007, <<http://www.mbari.org/vars/>>.
- Walther, D., D.R. Edgington, K A. Salamy, M. Risi, R.E. Sherlock, and Christof Koch, 2003, 'Automated Video Analysis for Oceanographic Research', *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, demonstration, Madison, WI.
- Walther, D, D.R. Edgington, C. Koch, Detection and Tracking of Objects in Underwater Video, 2004, *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, Washington, D.C.
- Widder, E.A., B.H. Robison, K.R. Reisenbichler, S.H.D. Haddock, 2005, 'Using red light for in situ observations of deep-sea fishes', *Deep-Sea Research*, I 52:2077-2085.