Automated Event Detection in Video of Benthic Habitat

Ishbel Kerkez, Flinders University of SA
Mentor: Duane Edgington
Summer 2004

Keywords: AVED, undersea vehicles, QVT, video, benthic transects, recognition

ABSTRACT

Quantitative video transects (QVTs) obtained with remotely operated vehicles (ROVs) are a sophisticated means of sampling that has recently replaced conventional methodologies. This technology has significantly advanced studies in animal diversity, distribution and abundance. The method currently used to analyse QVTs is labor intensive and costly, reducing the amount of data from the ROV dive and thus limiting marine ecological research. An automated system for detecting organisms and other objects visible in the videos would address these concerns. With this program, video frames are processed with a neuromorphic-selective attention algorithm, modeled after the human vision system. The candidate locations identified by the attention selection module are subject to a number of parameters. These parameters combined with successful tracking over several frames determine whether detected events are deemed "interesting" or "boring". "Interesting" events are identified and marked in the video frames. The visual characteristics of the ocean benthic habitat differ greatly from the midwater environment in which the system has previously been tested. In the benthic habitat the system yields a 92% correlation with professional annotations. Presented data details the comparison between automated detections and professional annotations.

INTRODUCTION

Remotely operated vehicles (ROVs) have revolutionized oceanographic research, replacing traditional technologies of towed nets and acoustics as tools for the assessment of animal diversity, distribution and abundance (Robison, 2000). Video equipment deployed on ROVs enable quantitative video transects (QVTs) to be obtained in midwater and benthic habitats at depths from 50 to 400 meters, providing high-resolution data at the scale of the individual organisms and their natural aggregation patterns (Robison et al., 1998; Robison, 2000; Robison, 2004).

With such a direct and non-invasive method of assessment in situ, high resolution video on dives are significantly advancing studies of behavior and physiology of deep sea organisms and correlating the physical, chemical and biological driving forces to spatial and temporal patterns of abundance and distribution (Robison, 2000; Robison, 2004; Ruhl and Smith, 2004). Gelatinous and other fragile organisms previously damaged and destroyed by sampling equipment of conventional methodologies have been grossly misrepresented in terms of diversity and abundance and the significant position they occupy in the food web. Video is the only accurate means of assessment for this large group of fauna.

Currently, the manual method employed by trained scientists analysing QVT video is very labor intensive and significantly reduces the amount of data processed from ROV dives. Application of this methodology to marine ecological research, therefore, is seriously limited by the ability to process the tapes of video.

Continued use of ROVs will provide great potential for QVTs to generate even more data, yet the time-consuming and tedious process of viewing and quantifying data remains a serious and prevalent constraint. It is proposed that the development of an automated system for detecting organisms and other objects (events) visible in the videos, would be beneficial in the analyses of ROV videos.

A neuromorphic-selective attention algorithm, modeled after the human vision system was used to process video frames (Itti and Koch, 1998). This algorithm is consistent with successful target detections in a variety of natural scenes (Itti and Koch, 1999; Itti and Koch, 2000). The candidate locations (tokens) identified by the attentional selection module are subject to a number of parameters that determine the likelihood of being deemed an “event” (Edgington et al., 2003). The combination of these tokens across video frames with token tracking enables the storage of potentially “interesting” events only after tracking has been successful over several frames. “Interesting” events are identified and marked in the video frames. A notion of “boring” video frames is a concept developed to represent sequences of underwater video in which no “interesting” events occur. Once “boring” frames are eliminated and tokens marked, it is believed that productivity and efficiency of professional annotators may be enhanced.

The attentional selection and tracking module is in an intial stage of development which has yet to be combined with an integrated automated video annotation system that will consist of an object classification module. The long-term objective for the automated visual event detection (AVED) program is to autonomously detect, recognize and classify organisms from the ROV database in real-time (Edgington et al., 2004).

With AVED being in such early stages of development, the potential in applying this system to a host of marine habitats and conditions remains largely unknown. Primary investigation of this systems application to ocean midwater habitat has demonstrated difficulties due to the low contrast of numerous translucent animals, in addition to the presence of debris in the water column (“marine snow”) cluttering the scene (Edgington et al., 2003).

In ocean benthic habitat, however, the systems parameters of color, size, contrast, orientation and movement are dramatically altered. Hence, any potential of the program in its ability to successfully detect objects of saliency is, as of yet, undetermined. The aims of this study are thus to examine the ability of the AVED program to detect objects of saliency in video of benthic transects and to conduct preliminary analyses required for further work on object recognition.

MATERIALS AND METHODS

Two ROVs are used at The Monterey Bay Aquarium Institute (MBARI), the *ROV Ventana* and the *ROV Tiburon* (Robison, 1993; Newman and Stakes, 1994). *ROV Ventana* uses a Sony HDC-750 HDTV (1035i30, 1920 x 1035 pixels) camera for video data acquisition, then recorded on a DVW-A500 Digital BetaCam video tape recorder (VTR). ROV Tiburon uses a Panasonic WVE550 3-chip CCD (625i50, 752 x 582 pixels) camera, with video also recorded on a DVW-A500 Digital BetaCam VTR.

Video was captured uncompressed as one minute AVI movie files at a resolution of 720 x 480 pixels and 30 frames per second using a Targa 3000 video capture card in a Pentium Xeon 2.0 GHz computer running the Windows XP operating system and Adobe Premiere 6.5. Video clips were converted to Quicktime movies and processed using the AVED processing system.

Before being tested for saliency, the individual video frames must undergo a series of pre-processing steps. Refer to Edgington et al. (2003) for a complete description of these steps. The video processing is performed on PCs running Red Hat Linux. In order to work with and be able to process large amounts of video data in a reasonable amount of time, a computer cluster was deployed with an 8 Rack Saver rs 1100 dual Xeon 2.4 GHz server, configured as a 16 CPU Gigabit Ethernet Beowulf cluster. Video clips were then burned on 4.7 GB DVDs using Veritas Record Now DX.

The video information management system (VIMS) at MBARI was queried to reveal a set of detailed annotations for two benthic transect dives conducted April 7 2003 and June 17 2004 at a depth of 200 meters in the Monterey Bay. Video processed with the AVED processing system was then compared with the professional annotations to denote whether detected events were in fact deemed “interesting” by the annotator. To investigate potential application of AVED to habitat classification, benthic video was obtained for a dive conducted November 16 1995 in the Monterey Bay.

RESULTS

i) Event detection

Results of testing the video processing capabilities of an automated visual event detection system (AVED) were promising for clips obtained from benthic dives of ROVs. For movie clips in which professionally trained annotators could identify numerous organisms, event detections made by the program coincided with those made by professional annotators approximately 92% of the time.

In a processed 45 minute segment of video, the total number of professional annotations was 236. The program detected 217 of those and missed 19 of the professionally annotated objects in the video segment. In two cases, the system detected animals that the annotator had missed.

The most abundant animal in the benthic transects was the seastar *Rathbunaster – californicus*, with a total of 123 recorded observations (figure 1). The program was successful in detecting these animals 100% of the time. The program performed poorly in detecting the category “larval fish/euphausia”, with only 3 out of 9 detections made.

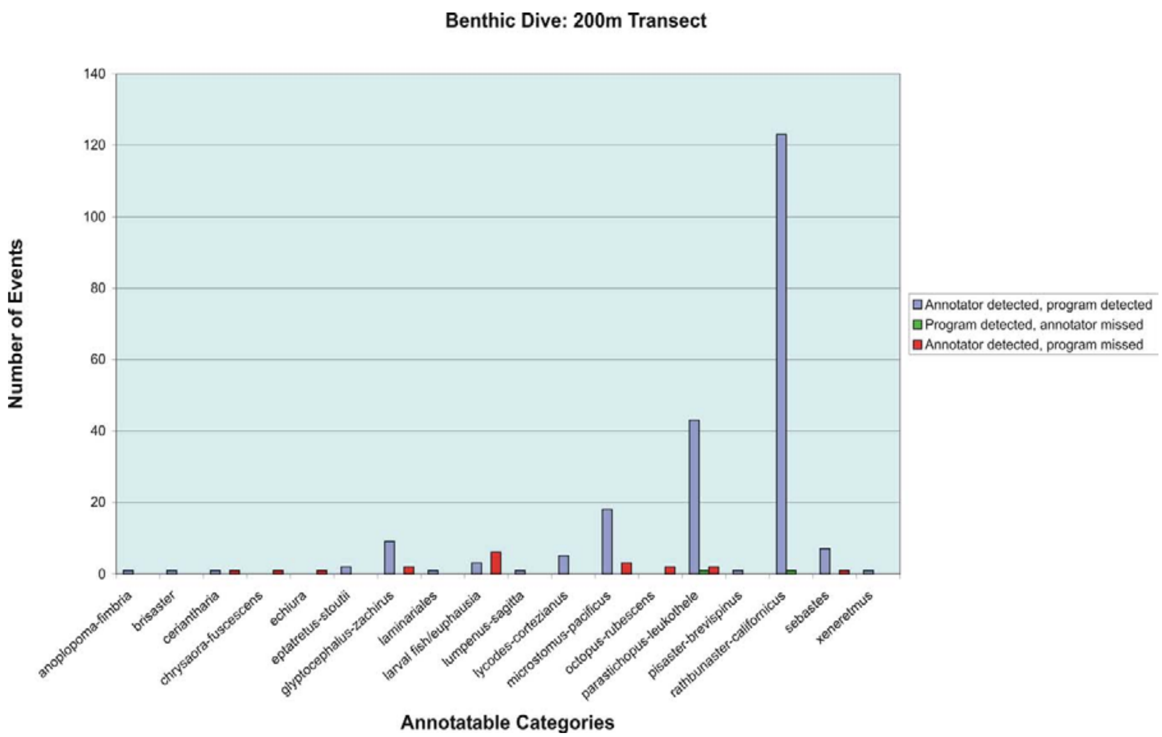

Figure 1. A comparison of event detections made by the AVED program against professional annotations for 45 minutes of processed benthic video.

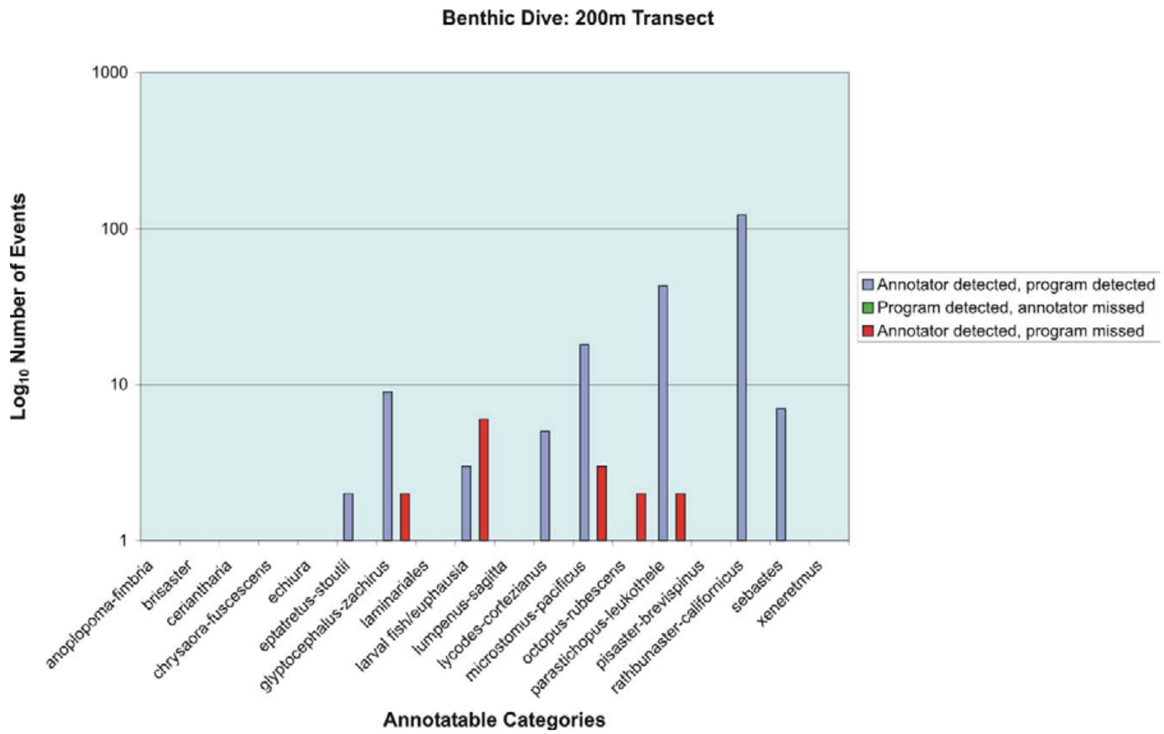

Figure 2. Logarithmic plot of event detections made by the AVED program against professional annotations for 45 minutes of processed benthic video.

ii) Habitat classification

Observation of one minute video clips of a complex rocky habitat yielded promising results. Detected events included individual rocks and rocky ledges and both grouped and individual animals separate from the rock they were inhabiting. It was also noted that there were few detections in open patches of sediment with relatively few “false hits”.

DISCUSSION

Results comparing events detected by the AVED system with professional annotations (figure 1) indicate a reliable performance of the computer vision system. The program was consistent in its ability to detect events deemed “interesting” by professional annotators and was in agreement with recorded annotations approximately 92% of the time.

Organisms with a higher rate of successful program detection included the seastar *Rathbunaster californicus* and the seacucumber *Parastichopus leukothele*. These animals were in high abundance throughout the analysed video segment and were detected by the program 100% and approximately 96% of the time respectively. Such a high rate of successful detection may be due to features of these animals conforming well to the pre-set program parameters. Both *Rathbunaster californicus* and *Parastichopus leukothele* are particularly large in size compared to midwater organisms previously used to test the program. They are also strongly coloured, of distinctive shape and well contrasted against a background of sediment. The AVED program also performed well detecting both groups of flatfish, *Microstomus pacificus* and *Glyptocephalus zachirus* (figure 2). Particularly promising were successful event detections of these animals despite individuals being buried in the sediment.

Results indicate a poor rate of detection for the category "larval fish/euphausia", with only 33% detected by the program. These midwater organisms appeared black against the dark background of sediment, thus they were poorly contrasted and often indistinct in shape. When observed, these animals were usually swimming around the camera lens of the ROV and thus often appeared out of focus. In addition, these animals were small in size.

In addition to the automated detection of benthic organisms, a further application of AVED could include habitat identification and classification. Observations of video clips of complex rocky habitat indicate a strong potential for development in this area. Of particular interest is the demonstrated ability of the program to detect rocks and animals irrespective of each other and to not group these together as one salient feature. Development of this program to classify habitat would broaden the scope of marine ecological research by enabling the direct investigation of associations between organisms and their habitat. This knowledge would contribute greatly to studies on the distribution and abundance of organisms, demonstrating yet another means of increasing data acquisition from video through an automated processing system.

This work constitutes a strong foundation for the development of the recognition module. Initial stages of testing would involve the collection of video frame grabs (n = 10) for each key species or those species occurring most frequently in the transects (i.e. *Rathbunaster californicus, Parastichipus leukothele*). Characteristics common to these individual groups of organisms can then be determined and specific traits such as colour, contrast, size and orientation will be averaged over the total number of clips and stored in the system. AVED can then be tested with these clips and its ability to identify an organism can be assessed.

CONCLUSIONS AND RECOMMENDATIONS

With such a strong agreement between the AVED system and professional annotations being demonstrated, it is evident that the applicability of this system to video of ocean midwater can be extended to include that of the benthos. Not only is the program able to detect those species most frequently observed throughout the analysed video, it performs successful event detections with great consistency. In addition to the automated detection and recognition of organisms, applications for benthic ecology include habitat identification and classification. Initial observations provide promising results and indicate a strong potential for a further application of this system.

This work provides a strong foundation for the development of a recognition module which, when implemented in conjunction with the demonstrated attentional module, will significantly reduce both time and effort in the processing of tapes. This is particularly true for cameras deployed on underwater platforms and deep sea cables where "boring" frames can be omitted from processing procedures. Implementation of this computer vision system would provide a significant and ongoing contribution to marine ecological research by greatly increasing the acquisition of stored data and by enabling reannotation of video collected in the past.

Testing the applicability of the AVED system to the benthos is only in its early stages. To better develop the program and extend its uses and applications to ecological research, further comparison of this computer vision system with professional annotation is required. Processing more video with a greater number of species is also required to ensure adequate testing of the system, as is the tested application of this system under a variety of conditions. It would be desirable, for example, to test the program with video in which lighting and visibility is poor and frames are cluttered with numerous animals or physical features, such as rocky ledges or mussel beds. Results of this work would be imperative for the continued development of both the attentional and recognition components of the program. It would also provide a sound foundation for further developmental applications of the program to some scheme of habitat classification.

ACKNOWLEDGEMENTS

I wish to thank Duane Edgington for his supervision and involvement with the project and equally for his encouragement and support; Dorothy Oliver for her tireless efforts and assistance, and Karen Salamy for technical assistance and invaluable feedback on all aspects of the work. I also acknowledge George Matsumoto for his guidance and immense contribution to the MBARI summer intern program and to all 2004 interns for their support and comradeship.

References:

Edgington, D.R., D. Walther, K.A. Salamy, M. Risi, R.E. Sherlock, and C.Koch (2003). Automated event detection in underwater video. *Proc. MTS/IEEE Oceans Conference,* San Diego, CA.

Edgington, D.R. et al (2004). MBARI internal project proposal, *Moss Landing, California*.

Itti, L., and C. Koch (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research,* **40(10-12)**: 1489-1506.

Itti, L., C. Koch, and E. Niebur (1998). A model for saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **20(11)**: 1254-1259.

Itti, L., and C. Koch (1999). Target detection using saliency-based attention. *In Proceedings RTO/SCI-12 Workshop on Search and Target Acquisition (NATO Uncalssified),* Utrecht, The Netherlands.

Newman, J.B. and D. Stakes (1994). Tiburon, development of an ROV for ocean science research. *In Proceedings MTS/IEEE Oceans*, Brest, France.

Robison, B.H. (2004). Deep pelagic biology. *Journal of experimental marine biology and ecology,* **300**:253-272.

Robison, B.H. (2000). The coevolution of undersea vehicles and deep-sea research. *Marine Technology Society Journal,* **33**: 69-73.

Robison, B.H., K.P. Reisenbichler, R.E. Sherlock, J.M.B. Silguero, and F.P. Chavez (1998). Seasonal abundance of the siphonophore, *Nanomia bijuga*, in Monterey Bay. *Deep-sea Research,* **22**: 1139-1153.

Robison, B.H. (1993). Midwater research methods with MBARI's ROV. *Marine Technology Society Journal,* **26**: 32-39.

Ruhl, H.A, and K.L. Smith (2004). Shifts in deep-sea community structure linked to climate and food supply. *Science,* **305**: 513-515.