Detecting, Tracking and Classifying Animals in Underwater Observatory Video

Duane R. Edgington, Danelle E. Cline, Jerome Mariette, Ishbel Kerkez
Monterey Bay Aquarium Research Institute (MBARI) Moss Landing, CA

Abstract
We are developing software to analyze high resolution video imagery from cameras deployed on ocean observatories, enabling quantitative video analysis to beobtained at the scale of the individual organisms. Video survey advances studies in animal diversity, distribution and abundance. Analyzing video, however, is labor intensive and costly, making ecological research and application to aquatic management. The challenge of analyzing video from fixed cameras and deep ocean observatory video footage is daunting due to the enormous quantity of data.

To address this problem we developed an automated system for detecting and classifying organisms in which frames are processed with a neurocognitive-selective attention algorithm. Candidate locations are subject to a number of parameters and tracking, to marked identified as “interesting” or not. The “interesting” events undergo further processing with a statistical classifier utilizing a Gaussian mixture model to determine the abundance and distribution of a selected organism category.

Presented data detail the comparison between professional annotations and automated detection of organisms in coastal and deep ocean observatory video footage. We present automated classification of organisms in benthic video footage:

Video is professionally annotated to feed the MBARI Video Annotation and Reference System (VARS) database, which enables integration of annotation results and linking them to environmental data over many dives and over many years.

Application of Biomimic Models to Detection and Classification of Visual Events
• Humans and many animals are extremely good at attending to novel features in a scene. A model of attention was developed in 1985 by Koch and Ullman (MIT). It was based on the biology of human perception and visual system.
• The model was implemented as a computer program by Itti in the Koch lab at Caltech as a Ph.D. Thesis in the late 1990’s.
• The model has been applied to terrestrial surveillance, traffic surveillance and advertising copy. This research is the first application of the model to underwater video scenes.
• The “interesting” events undergo further processing with a statistical classifier utilizing a Gaussian mixture model to determine the abundance and distribution of a selected organism category.

Results
Below: Captured images depicting natural benthic scene before and after AVEO processing. Above: Graph illustrating the AVEO performance denoted by frequency of occurrences against professional annotations for 172 video clips of various duration (1 to 20 minutes). A high rate of detection and a low rate of false detection and misses are evident. The automated system correctly identifies video containing interesting events (Correct Positive) 81.4% as well as video not containing events (Correct Negative) 93% with few false alarms (False Positive) 11.6% and very few misses of video clips with one or more interesting events (False Negative) 0.6%.

Acknowledgements
This work was supported in part by funding from MBARI and NSF. We thank W. A. Barrett, A. M. Best, and B. A. Brieden for helpful feedback and support. This project was initiated at the 2003 NSF/Neuroinformatics Workshop, Telluride, Colorado. The authors wish to thank the Monterey Bay Aquarium Research Institute and Monterey Bay Aquarium for their generous support for this project. Portions of this work were performed at the NSF Center for Neuroinformatics System (SYNAPSE) in San Diego, 17th Ave. South, San Diego, CA for providing technical support and resources.

Below Left: A comparison of event detections made by the AVEO program against professional annotations for 85 minutes of processed benthic video.

Below Right: Graph illustrating program performance denoted by frequency of successful detection.

We analyzed 7.5 minutes of a benthic transect. We trained the classifier with grayscale square sub-images of segmented frames, each containing an example object. For testing, we extracted 210 events detected by our system (7250 images) The recognition module successfully classified 38 of 42 (90%) Rathbunaster tagged by professional annotators (90% recall). There were no instances in which any other events were falsely classified as Rathbunaster.