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 be obtained 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, limiting marine ecological research and application to aquatic management. The challenge of analyzing video from fixed cameras in observatories operating around the clock is particularly 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 neuromorphic-selective attention algorithm. Candidate locations are subject to a number of parameters and tracking, to mark detected events 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.

Above: MBARI annotator analyzing hours of ROV dive tape footage.

Video Collection and Annotation

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 years.

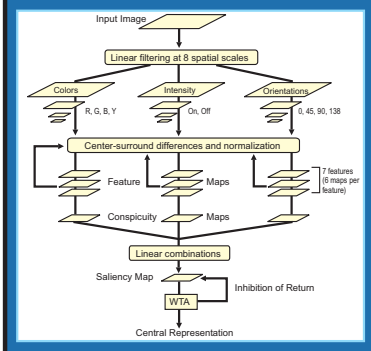
~2,000,000 individual observations in MBARI annotation database

Annotating video is time-consuming and tedious.

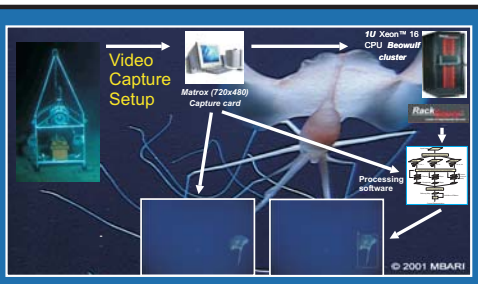
Can we supply tools to make the analysts more productive and efficient?

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.



Above Left: Flow diagram of a typical model for the control of bottom-up attention. This diagram is based on Koch and Ullman's hypothesis that a centralized two-dimensional saliency map can provide an efficient control strategy for the deployment of attention on the basis of bottom-up cues. The input image is decomposed through several pre-attentive feature detection mechanisms (sensitive to color, intensity, etc), which operate in parallel over the entire visual scene. Neurons in the feature maps then encode for spatial contrast in each of those feature channels. In addition, neurons in each feature map spatially compete for salience, through long-range connections that extend far beyond the spatial range of the classical receptive field of each neuron (here shown for one channel; the others are similar). After competition, the feature maps are combined unto a unique saliency map, which topographically encodes for saliency irrespective of the feature channel in which stimuli appeared salient. The saliency map is sequentially scanned by attention through the interplay between a winner-take-all network (which detects the point of highest saliency at any given time) and inhibition of return (which suppresses the last attended location from the saliency map, so that attention can focus onto the next most salient location). Top-down attentional bias and training can modulate most stages of this bottom-up model.



Processing

1. Record and Capture Video

- Video recorded by observatory camera, or broadcast / HDTV cameras on ROV (Digital BetaCam or HDTV Recorder).
- On shore, video is captured into data files that are further processed.

2. Pre-process Frames & Identify Salient Locations

- Smooth to remove scan lines.
- Subtract the sliding average of the last 10 frames to remove constant background.
- Salience based on low-level properties such as luminance contrast, local orientation contrast And color contrast (red-green and blue-yellow).
- Salient points are scanned by the interaction of a Winner-Take-All (WTA) neural network and Inhibition-Of-Return (IOR).

3. Track Salient Objects

- Track the x and y coordinate of the centroid with linear Kalman Filters (moving camera) or Nearest Neighbor (fixed camera).
- Assume constant acceleration (good assumption for projection of constant velocity motion onto the camera plane) for Kalman Filter.
- Data assignment for multiple target tracking made easy by sparseness of salient objects.
- Every 5 frames: Check for salient objects that are not yet tracked, and initialize new trackers.

4. Extract Binary Objects at Salient Locations

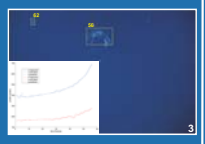
- Segment objects using image flooding with fixed thresholding.
- Extract a number of intermediate-level properties for each object ("object-token")
  - Area, centroid
  - Second moments
  - Major and minor axes, elongation
  - Major axis orientation
  - Maximum, minimum & average image Intensity within the object shape

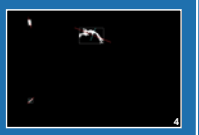
5. Decide which events are "Interesting"

- Initial approach: Decision was based on area.

Output interesting events into XML file (serial number, location,area, etc) .

Generate graphical output marking event in output frame.

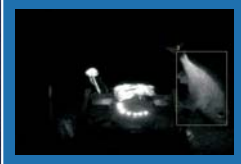




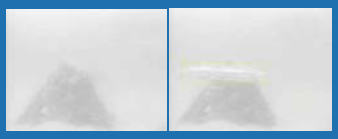
Results

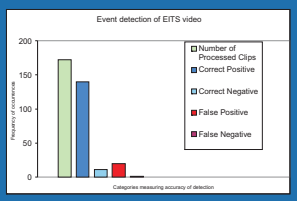
Below: Captured images depicting natural benthic scenes before and after AVED processing. Boxes drawn around **i) Rathbunaster californicus**, **ii) Parastichopus leukothele** and **iii) Microstomus pacificus** denote event detections.

Results



Above: Images from 2 fixed observatory cameras after AVED processing.
Left, from Eye-in-the-Sea intensified low-light video camera.
Right, from parked Ventana ROV HDTV camera.





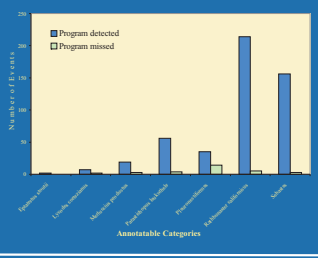
Above: Comparison of Eye-in-the-Sea video automated event detection with professional annotation for 172 video clips of various duration (1 to 20 minutes). A high rate of detection and a low rate of false detection and of misses are evident. The automated system correctly identifies video containing interesting events (Correct Positive) 81.4% as well as video not containing events (Correct Negative) 6.4% 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%.

Above: MARS Cable Observatory Test Bed. The Monterey Accelerated Research System (MARS) will allow scientists to perform a variety of long-term and real-time observations and experiments 900 meters below the surface of Monterey Bay. MARS will serve as an engineering, science and education test bed for even more extensive observatories in the USA (ORION) and Canada (NEPTUNE Canada).
Image: David Fierstein (c) 2005 MBARI.

Below Left: A comparison of event detections made by the AVED 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



Acknowledgements

We thank the David and Lucile Packard Foundation for their generosity in funding work at MBARI, and D. Walther, M.'A. Ranzato, C. Koch and P. Perona of the NSF Center for Neuromorphic Systems Engineering at Caltech and L. Itti of the Univ. Southern California for providing software and insight. This project was initiated at the 2002 NSF Neuromorphic Engineering Workshop, Telluride, Colorado; NSF Research Coordination Network (RCN) Institute for Neuromorphic Engineering (INE) supported collaborative travel. We thank E. Widder, E. H. Raymond (ORCA), B. Robison, R. Sherlock, L. Kuhnz (MBARI), C. Barans and G. Sedberry (SCDNR) for sample videos, annotations, and science insights, and MBARI video lab staff for their support and interest in our project. K. Salamy provided technical support.