A DETECTION, TRACKING, AND CLASSIFICATION SYSTEM FOR UNDERWATER IMAGES

Danelle E. Cline, Duane R. Edgington

Monterey Bay Aquarium Research Institute, 7700 Sandholdt Road, Moss Landing, CA 95039
USA, email:{dcline,duane}@mbari.org

Introduction

Traditional underwater images for ecology study often include time-lapse or brief video recordings over short periods of time. More recently, with the installation of under water cabled observatories that provide constant electrical power and data connections to the seafloor, long-term video recordings are possible for the first time.

These valuable recordings are essential for underwater ecology studies, particularly abundance and distribution studies but also behavioral studies. However, the analysis of these video and time-lapse recordings often becomes quickly overwhelming. In some cases, the number of underwater animal activities can be infrequent, resulting in recordings with many hours of video with few events of interest. Sometimes the human resources do not exist to analyze the recordings, or if resources are available, the strains on human attention quickly abate the efforts.

To help address this issue, over the past six years an automated detection, tracking, and classification software system called The Automated Visual Event Detection and Classification System (AVEDac) has been under development at the Monterey Bay Aquarium Research Institute (MBARI). The AVEDac system is a powerful aid that is currently being used to analyze Remotely Operated Vehicles (ROVs) and deep-water cabled observatory video recordings recorded in the Monterey Accelerated Research System (MARS) observatory. It has been recently modified to also process still images recorded from Autonomous Underwater Vehicles (AUVs) and stationary cameras on the sea floor.

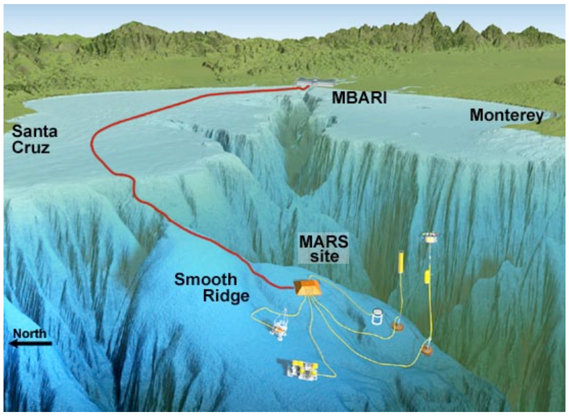

Figure 1. A 3-D perspective of the MARS cabled-to-shore observatory site on Smooth Ridge, at the edge of Monterey Canyon off the coast of California, USA.

The cornerstone of this system is its approach for detection. Taking cues from evolutionary systems, the AVEDac software analyzes each image searching for “interesting” events using a neuromorphic software model based on the human vision system. An “interesting” event is an object that can be tracked successfully over several frames. This approach has been shown to be remarkably effective. Once “interesting” events are detected using this approach, the system then tracks these events across multiple frames and then passes them to a Bayesian classifier utilizing a Gaussian mixture model to determine the lowest possible taxonomic category. The output of the AVEDac software is recorded into XML formatted data that can be edited in the AVEDac graphical editor for misclassifications, or exported to Excel format for further analysis in conjunction with ancillary data.

Detection

The AVEDac detection system uses a neuromorphic software model based on the human vision system from the iLab Neuromorphic Vision C++ Toolkit developed at the University of Southern California [1]. Using this toolkit, different methods are used for detecting "interesting" objects in a scene depending on whether still-frames and video. For still-frames, the local variance in 16x16 image patches of the input frame is calculated and input into a saliency map. For video, the input frame is decomposed into seven channels (intensity contrast, red/green and blue/yellow double color opponencies, and the four canonical, spatial orientations). In both cases, calculations are done at six spatial scales and stored into saliency maps. After iterative spatial competition for salience within each map, maps are combined into a unique saliency map. The saliency map is then scanned by the focus of attention in order of decreasing saliency, through the interaction between a winner-take-all neural and an inhibition-of return mechanism [2]. The peak points in this map are used to seed the tracking system.

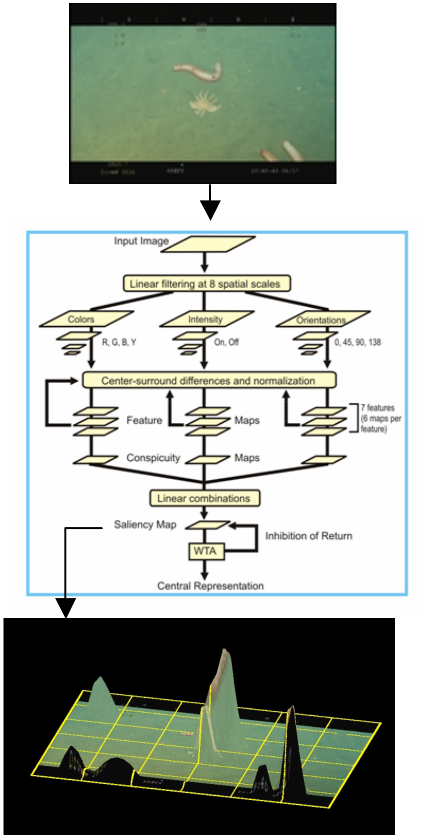

Figure 2. Saliency map from the iLab toolkit warped onto a 3-D map for a single video frame. Peaks in the map show points of high visual attention where the Rathbunaster and Leukothele are (near image center).

Segmentation and Tracking

A number of different algorithms are used for segmentation and tracking depending on whether processing ROV, AUV, or observatory camera images. These algorithms are defined in predefined profiles that are selected before processing.

For instance, in the case of video recorded from a fixed observatory, an image average from a running image cache is used with a graph cut-based [3] 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 in this case a simple nearest neighbor tracking algorithm is used. In the case of an ROV camera, a segmentation algorithm based on an adaptive threshold and Otsu's method is used [4]. 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 in this case, tracking is achieved with separate linear Kalman Filters for the x and y centroid of each tracked object.

Classification

For classification, each "interesting" even is segmented from the scene into square images. Because animals in underwater video can be in various poses, features from these images need to be invariant with respect to shift and rotation. We derive our features from "local jets" [5] and the invariants defined by Mohr and Schmid [6]. The first invariant is the local average luminance, the second one is the square of the gradient magnitude, and the fourth one is the Laplacian. These invariants are computed at four different scales. Classification is finally done using a Gaussian mixture model to the lowest taxonomic category represented by the labeled training set. In the case of video, a class is assigned for each video frame of an "interesting" event, then a majority-win voting scheme is used to decide the winning class. A further discussion of the classification can be found in the referenced paper [7] for which this work was based on.

Results and Future Work

In recent test, classification of an independent test set taken from bottom ROV video resulted in 100% accuracy for a type of deep-sea benthic echinoderm Rathbunaster californicus (a sea urchin) and 95% accuracy for the Parastichopus leukothele. In contrast, an independent test of still images from ocean bottom time-lapse recordings didn't perform well, with 24% accuracy for Echinocrepis rostrata (a sea urchin), and 26% accuracy for Benthocodon (a common jellyfish). Based on these results, future work includes exploring additional classifiers to improve still image classification results.

Conclusion

We present a real-word system for use in classifying animals in underwater video and still images. With the increasing importance and collections of digital images and video in oceanography, this software system is an example of how innovative software technology can be used to aid scientist in underwater ecology studies. This system is currently being used to analyze video and still images collected from ROVs, AUVs, and stationary cameras on the sea floor off the coast of California. Preliminary testing indicates AVEDac is suitable for classifying animals in video, yet improvements to the classification are needed to improve the accuracy to an acceptable level for day-to-day use in still image applications.

References

[1] iLab Neuromorphic Vision C++ Toolkit at the University of Southern California, viewed 30 March 2010, <http://ilab.usc.edu/toolkit>.

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

[3] N. Howe & A. Deschamps, “Better Foreground Segmentation Through Graph Cuts”, Technical report, viewed 30 March 2010, <http://arxiv.org/abs/cs.CV/0401017 >

[4] 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.

[5] Koenderink, J.J., van Doorn, J., 1987. Representation of local geometry in the visual system. Biological Cybernetics 55, 367-375.

[6] Mohr, R., Schmid, C., 1997. Local grayvalue invariants for image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (5), 530-535.

[7] Ranzato, M, P.E. Taylor, J.M. House, R.C. Flagan, Y. LeCun and P. Perona. “Automatic Recognition of Biological Particles in Microscopic Images”, Pattern Recognition Letters, Volume 28, Issue 1, 1 January 2007, pp. 31-39.