The art of seeing jellies

The oceans contain a wealth of living creatures that account for a large amount of the biomass on our planet. How can we assess the kinds and numbers of animals in the oceanic water column? For more than a century, the traditional approach has been to tow nets behind ships. This method is limited in its spatial resolution, and because of the design of the nets, gelatinous animals (such as jellies, previously known as jelly fish) are destroyed and, hence, undersampled. Today, remotely-operated underwater vehicles (ROVs) provide an excellent alternative to nets for obtaining quantitative data on the distribution and abundance of oceanic animals.[1]

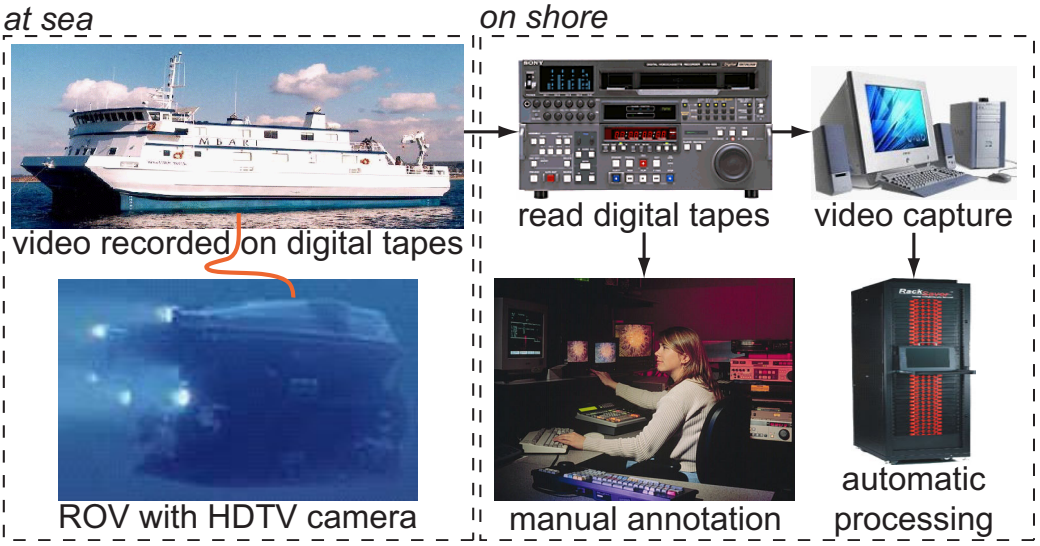

Figure 1. *The process flow for recording the video material at sea and processing it on shore either manually or automatically.*

Using video cameras, it is possible to make quantitative video transects (QVT) through the water, providing high-resolution data at the scale of the individual animals and their natural aggregation patterns. However, the current manual method of analyzing QVT video by trained scientists is very labor intensive and poses a serious limitation to the amount of data that can be obtained from ROV dives. To overcome this bottleneck in analyzing ROV dive videos we have developed an automated system for detecting and tracking animals for subsequent identification, based on neuromorphic vision algorithms.[2] These tasks are difficult due to the low contrast of many translucent animals and due to debris (known as 'marine snow') cluttering the scene.

Onboard the research vessel, the HDTV video signal from the ROV's broadcast-quality cameras is recorded on a digital BetaCam video deck. Back on shore, the videos are converted to a computer-readable format, and some generic pre-processing is performed for each frame, such as subtracting the background, smoothing scan lines, and global contrast enhancement.

For the crucial detection step, we use an extended version of the Itti & Koch saliency-based attention algorithm[3] (see also Itti's article in this issue). For this neuromorphic detection system for salient objects, each input frame is decomposed into seven channels for intensity contrast: red/green and blue/yellow double color opponencies, and four spatial orientations (0°, 45°, 90°, and 135°) at six spatial scales, yielding 42 'feature maps'. To improve the detection of faint, elongated animals, we introduced an additional across-orientation normalization step for the orientation filters, which is inspired by local interactions of orientation-tuned neurons in the primary visual cortex.

After iterative spatial competition for salience within each map, only a sparse number of locations remain active and all maps are combined into a unique 'saliency map'. This is scanned by the focus of attention in order of decreasing saliency, through the interaction between a winner-take-all neural network (which selects the most salient location at any given time) and an inhibition-of-return mechanism (transiently suppressing the currently-attended location from the saliency map).[3] Once salient targets have been detected, they are tracked from frame to frame using linear Kalman filters[4] for the *x* and the *y* coordinates of the apparent motion of the objects in the camera plane: this assumes motion with constant acceleration. This is a good assumption for the constant-speed-heading motion of ROVs while obtaining QVTs.

Normally, tracking multiple targets at the same time raises the problem of assigning measurements to the correct tracks. Since our neuromorphic detection algorithm only selects the most salient objects, however, we obtain a sparse number of objects whose predicted locations are usually separated far enough to avoid ambiguities. If ambiguities occur, we use a measure based on the distance of the objects from the predictions of the trackers and the size ratio of the detected and the tracked objects. Every couple of frames, the scene is again scanned for salient targets that are not already being tracked, and new trackers are initiated for these.

For each tracked object we obtain a binary mask that allows us to extract a number of low-level properties such as the object size, the second moments with respect to the centroid, the maximum luminance intensity, the average luminance intensity over the shape of the object, and its aspect ratio. We use these features to broadly classify the detected objects into those that are interesting for the scientists, and those that are debris. Since the occurrence of visible animals in the video footage is typically sparse in space and time, we can identify many frames that do not contain any objects of interest. By omitting these frames and marking candidate objects, we can enhance the productivity of human video annotators and/or cue a subsequent object classification module.

Our attentional selection and tracking system shows very promising results for transects from ROV dives that have been analyzed by human annotators already. This module is only the first step towards an integrated neuromorphic video annotation system that will consist of an object classification module and control modules for pan/tilt/zoom cameras: these in addition to the attentional module. This integrated system will be able to count the most common animals fully automatically.

Dirk Walther* and Duane Edgington†
*California Institute of Technology
E-mail: walther@caltech.edu
†Monterey Bay Aquarium Research Institute
E-mail: duane@mbari.org
http://www.mbari.org/aved

References

1. B.H. Robison, *The coevolution of undersea vehicles and deep-sea research,* **Marine Technology Society J. 33**, pp. 69-73. 2000.
2. D. Edgington, D. Walther, K.A. Salamy, M. Risi, R.E. Sherlock, and C. Koch, *Automated Event Detection in Underwater Video,* **Proc. MTS/IEEE Oceans Conf.,** San Diego, California, 2003.
3. L. Itti, C. Koch, and E. Niebur, *A model of saliency-based visual attention for rapid scene analysis,* **IEEE Trans. on Pattern Analysis and Machine Intelligence 20** (11), pp. 1254-1259, 1998.
4. R.E. Kalman and R.S. Bucy, *New Results in Linear Filtering and Prediction Theory,* **J. of Basic Engineering 83** (3), pp. 95-108, 1961.