VISUAL TRACKING OF JELLYFISH *IN SITU*

Jason Rife and Stephen M. Rock
Stanford University, CA 94305
and Monterey Bay Aquarium Research Institute, Moss Landing, CA, 95039

ABSTRACT

Marine biologists desire an automated system for observing gelatinous zooplankton. This study represents the first attempt at real-time robotic tracking of a gelatinous animal in the open ocean. Challenges stem from uneven lighting, encounters with other marine life and frequent out-of-frame events. Analysis of a large video library of human-controlled jelly tracking supported selection of an appropriate vision algorithm. The vision architecture was implemented as an ROV pilot aid and tested at 600m below the surface of Monterey Bay.

1. INTRODUCTION

Many deep ocean gelatinous animals are too fragile for transport, and must be studied *in situ*. Current research employs human pilots and remotely operated vehicles (ROVs) to observe deep ocean life. Vision technology enables development of a pilot aid to extend observations.

To hold station near a jelly requires a relative position sensor. The abbreviation jelly refers here to any instance of gelatinous marine animal (a definition that might horrify marine biologists, who meticulously differentiate among Cnidarians, Ctenophores, Appendicularians and other Phyla of gelatinous zooplankton). Visual sensors are preferred to sonar, as jellies reflect a negligible fraction of incident acoustic energy [6].

Several papers have touched on the topic of tracking in natural underwater environments, but none have discussed real-time tracking of a gelatinous animal *in situ*. Minami *et al.* tackled closed-loop tracking of a fish, confined to small tank [5]. Other workers have analyzed tracking of marine animals off-line, without closing a servo loop. Kocak *et al.* discuss vision techniques for off-line analysis of bioluminescent zooplankton data collected *in situ* [4]. Fan and Balasuriya tested a 20 Hz fish tracking technique off-line, using video collected in the open ocean [3]. Other investigators have focused on pattern recognition methods useful for detecting underwater targets [7,9].

This work presents an architecture specifically tailored to closed-loop tracking of gelatinous animals *in situ*. Section 2 describes the experimental apparatus for open-ocean testing. Section 3 lays out difficulties associated with the imaging environment. Section 4 introduces a vision system architecture for robust jelly tracking.

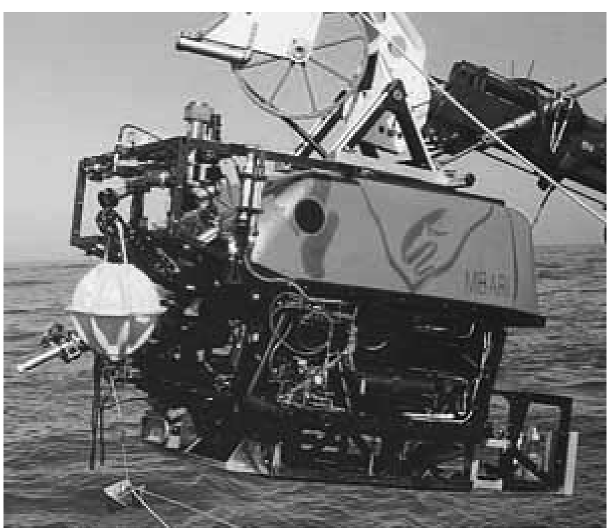


Figure 1. MBARI Remotely Operated Vehicle *Ventana*

2. FIELD TESTING

Because of the difficulty in developing systems for underwater use, experimental field-testing is required at each design stage. Initially, a large library of video was acquired using a human-operated submersible robot. This data was used to design and implement a vision system, which has completed initial testing in a series of dives in Monterey Bay, California.

Experiments occur in association with Monterey Bay Aquarium Research Institute (MBARI). ROV *Ventana*, an unmanned submersible, served as a platform for deep ocean data acquisition. ROV *Ventana* is the workhorse submersible for marine biology research in the Monterey Bay [1]. A 2000 m umbilical carries command signals via fiber to the submerged robot. Video images from a Sony HDC-750A camera return to the surface along the same umbilical. The umbilical also carries power to the ROV from the support ship.

Vision processing algorithms run on a 700 MHz Pentium III under Windows NT. Software operates on 128x120 images grabbed at camera frame rate (30 Hz). Computer output commands are routed through a pilot joystick and subject to a manual override, to ensure safe operation of the submersible.

3. VISION SYSTEM CHALLENGES

3.1. Control System Requirements

The jelly-tracking task requires position control accurate enough to envelop the target within the camera field of view. This loose positioning requirement permits monocular visual servoing, with control loops closed on target projected area and centroid location. Approximate range, up to a scale factor, may be inferred by comparing target area to a nominal area. Target bearing may be calculated as a function of its centroid and the image optical center. An appropriate transformation relates coordinates in the camera and vehicle reference frames.

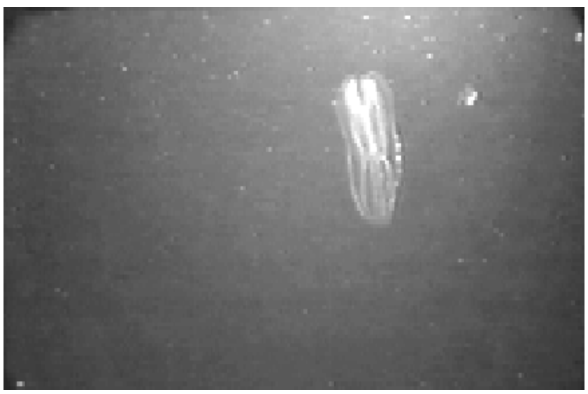


Figure 2. Intensity image depicting Ctenophore against a field of uneven illumination.

3.2. Limitations Imposed on Image Processing

Large Amplitude Motions:
Video recorded in Monterey Bay shows that jellies remain stationary for long periods, punctuated by bursts of acceleration. The submersible moves continuously, driven by tether forces, ocean currents, and closed-loop actuation. Combined jelly and ROV motions result in frequent out-of-frame events.

Three types of large amplitude motion carry the jelly outside the vision window. In the first type, the target moves laterally past the image border. The second type carries the jelly too close to the camera enclosure, inside the illumination zone. The third event type occurs when the jelly drifts too far away from the ROV. At range, the target grows too dim or small for the camera to resolve.

Intensity Processing:
Jellies do not appear as solitary bright spots against dark backgrounds. In the depths, illumination backscatters to the camera, producing a relatively bright and non-uniform background image. Figure 2 depicts a target jelly amidst veiling illumination. This diffuse light, created by the ROV, moves with the vehicle and remains nearly constant over time. Suspended organic particles, known as marine snow, introduce continual small fluctuations to this background image. Large changes occur infrequently, when the ROV pilot pans or zooms the camera (often as a diagnostic method to visually inspect ROV status).

Color Processing:
Most jellies have chromatic components matching seawater. A few species, like the *Atolla* of Figure 3, display prominent color patterns, usually red in nature. A red-component threshold easily identifies these jellies.

Gradient Processing:
When viewing opaque objects, gradients appear most prominently along edges. With translucent jellies, moderate gradients are distributed across the entire gelatinous body. Internal organs produce strong intensity gradients on the target interior (see Figures 2 and 3).

Hotspots on the camera enclosure produce strong gradient response. Lighting geometries that might result in these bright reflections are difficult to predict in advance, especially when ROV payloads change on a daily basis.

3.3. Identifying Jelly Features

Software must differentiate the target jelly from other marine animals. Wandering animals appear frequently in the video database, as often as every two minutes. Figure 3 depicts a close encounter between two jellies.

To distinguish among multiple targets and to reacquire targets following out-of-frame events, statistics need be computed for each segmented region. Some of the statistics may be position related, to aid in discrimination amongst multiple targets over short time intervals. The unpredictability associated with out-of-frame events necessitates that other statistics be position invariant. Position invariant statistics should be robust to bodily deformations and out-of-plane rotations. Tentacle motion, in particular, alters the jelly contour over very short time scales. Also, position invariant statistics should tolerate large range changes. In practice, range from target to camera can vary by a factor of more than five.

Figure 3. Image showing close proximity of two jellies, genus *Mitrocoma* (top) and *Atolla* (bottom).

4. A VISION SYSTEM ARCHITECTURE

Appropriate segmentation and pattern recognition methods alleviate constraints detailed in Section 3.

4.1. Evaluating Segmentation Techniques

The vision literature suggests many segmentation techniques, but few perform well for jelly tracking. Comparison of segmentation methods leveraged 20 hours of video collected during human piloted jelly tracking.

Large amplitude motions foiled two active contour methods, an ellipse fitting method and a snake method [8]. Frequent out-of-frame events required expensive and often unsuccessful re-initialization. Large tentacles produced additional problems for the active contours.

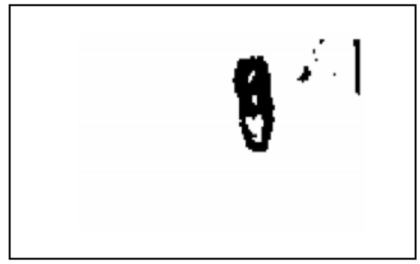
Intensity thresholding routines, even adaptive ones, proved unreliable. Gradients in the background image create overlap between target and background intensity values. In these cases, no unique threshold level exists. Region-merging methods also encounter difficulties that result from smoothly varying backgrounds and jelly transparency. Expansive regions belonging to the background were often misclassified as target regions, and vice versa. Nor did watershed methods give reliable results. Watershed methods were applied to the gradient image, using bright intensity patches to form initial markers [2]. Strong intensity gradients interior to the jelly create multiple watersheds for the same target. Attempts to merge these watersheds encountered difficulties similar to those observed for other region-merging methods.

Surprisingly, the most successful segmentation techniques were the simplest. The following section describes three simple, but highly robust methods.

4.2. Robust Segmentation

Gradient magnitude thresholding proved the most robust segmentation technique tested during video analysis. This filter relies on good separation between background and target gradient magnitude. Though the integrated effects are substantial across typical images, background gradient typically falls below one gray level per pixel. By comparison, gradient magnitude averages four or more gray levels per pixel across typical target jellies. Figure 4 plots the gradient separation for the Ctenophore and background first depicted by Figure 2. The plot gives percent misclassification of pixels (background or target) as a function of choice of gradient threshold level. If the threshold level is chosen at three gray levels per pixel, 9% of background pixels are misclassified as possible targets, and 12% of jelly pixels are incorrectly associated with the background. These results are representative for the image database. Most video yields good results for gradient thresholds between 2-7 gray levels per pixel.

Gradient magnitude images benefit from prefiltering. A morphological opening followed by a morphological closing removes stray gradients produced by small snow particles and tentacles. Subsequent convolution with a 3x3 uniform filter further smooths the gray image. The morphological gradient (local maximum minus local minimum) was found using the smoothed image. This operator outperformed the Cartesian gradient operator in terms of computational speed and noise.



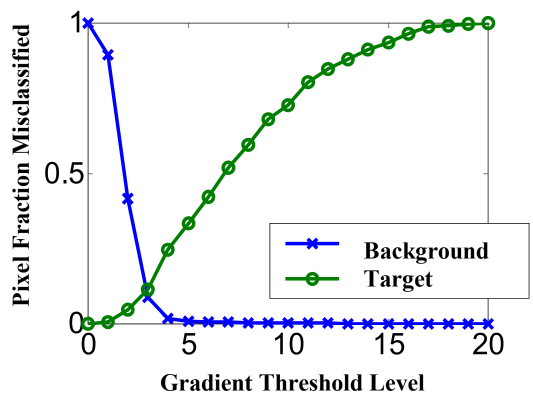


Figure 4. (a) Gradient segmentation of Figure 2, with threshold set at four. (b) Fraction of pixels misclassified as a function of threshold.

The assumption of target-background gradient separation weakens under two conditions. At close ranges, where the target outgrows 1/10th the viewable area, the horizontal scale of Figure 4 shrinks. Gradient levels for background and much of the jelly compress below one gray level per pixel. Gradient segmentation continues to function, but identifies target edges rather than the target body.

Gradient methods also fail for opaque jellies. At great depths, red coloration nearly matches transparency as a means of camouflage. Fortunately, thresholding the red color channel locates these dark red, opaque jellies. Typically, a 70% red cutoff gives good results.

A third segmentation type, based on background image differencing, also performed well. Background pixels were computed over time by averaging sequential images, excluding regions of high gradient or red color. Since the ROV moves with its light source, this method recovers a reasonable approximation of ambient lighting, as long as the target animal does not remain perfectly stationary in the field of view. A subtraction of the background image from the current image yields the background difference. The background difference image causes moving objects to stand out saliently. Results are moderately insensitive to threshold level, and thresholds between 4-10 gray levels function well.

Background differencing has the advantage of filtering out hotspots, and the disadvantage of a required re-initialized whenever the pilot adjusts camera angle. The segmentation technique also fails when the jelly occupies a large fraction of the viewing window. These limitations make background differencing most useful as a supplementary filter for hotspot elimination.

4.3. Feature Extraction

Frequent out-of-frame events and multiple animal encounters demand a robust vision algorithm that can distinguish the target from false signals. Wide variations of range, shape, and illumination spell difficulty for pattern recognition. Variations in range create the most significant problems for texture and feature based recognition. Features visible at proximity disappear below resolution limits with large changes in range. Jelly three-dimensionality and deformability introduce additional complexity for shape recognition routines. Correlation methods break quickly under these conditions.

Given the challenges associated with the imaging environment, simple features prove most useful. For this work, the pattern vector consists of two feature sets, those describing spatial relations (x-y coordinates, size, aspect ratio) and those describing the jelly in a position-invariant way (color, reflectance).

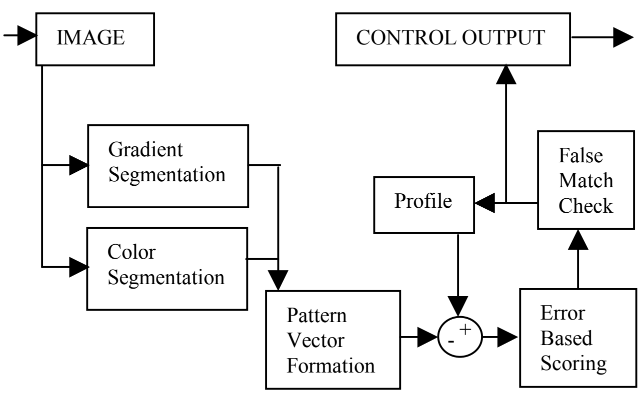


Figure 5. Block Diagram for jelly tracking system.

4.4. The Complete Architecture

Gradient and color thresholding methods run in parallel, segmenting the search space. Statistics for each segmented region form a pattern vector. In turn, pattern vectors are differenced from a historical profile to produce an error measure. With error sufficiently low to preclude a false match, a weighting function, based on confidence of a correct match, updates the profile. A training phase generates the initial profile automatically. During this phase a human pilot points the camera at the target.

4.5. Experiments

Preliminary experiments occurred in Monterey Bay during 2000. Ocean experiments employed a prototype architecture as a sensor for closed loop PD control. Actuating in vertical translation and yaw rotation, the controller enabled ROV *Ventana* to automatically track two targets. The system tracked the first jelly, of genus *Mitrocoma,* for one minute. The ROV tracked a second target, *Phacellophora Camtschatica*, for consecutive ten minute periods, interrupted by the appearance of a large school of squid. Longer tracking times are expected when the full software becomes available for summer, 2001.

5. CONCLUSION

This paper describes automated visual jelly tracking. Vision requirements were specified using a video library recorded during ocean operations. Algorithm robustness proved particularly sensitive to choice of segmentation method. The wide range of conditions encountered by the jelly tracker implies that simple methods, including threshold segmentation and pixel-based feature acquisition, outperform several more complex techniques. Prototype software was tested *in situ* using MBARI ROV *Ventana.* The system automatically tracked a single jelly for an uninterrupted period of ten minutes.

We thank MBARI and Packard Foundation grants 98-3816 and 98-6228 for supporting this work.

6. REFERENCES

[1] R. Barber. “Deep Ocean Research Using Computer Controlled ROVs”, *Sea Technology*, 30, pp. 47-49, 1989.

[2] E. Dougherty (ed.). *Mathematical Morphology in Image Processing*, Marcel Dekker. 1992.

[3] Y. Fan and A. Balasuriya. “Autonomous Target Tracking by AUVs using Dynamic Vision”, *Proc. of the 2000 International Symposium on Underwater Technology.* pp. 187-192, 2000.

[4] D. Kocak, N. da Vitoria Lobo, E. Widder. “Computer Vision Techniques for Quantifying, Tracking, and Identifying Bioluminescent Plankton”, *IEEE Journal of Oceanic Engineering*, 24, v. 1, pp. 81-95, 1999.

[5] M. Minami, J. Agbanhan, T. Asakura. “Manipulator Visual Servoing and Tracking of Fish Using a Genetic Algorithm”, *Industrial Robot*, v. 26, 4, pp. 278-289, 1999.

[6] E. Mutlu. “Target Strength of the Common Jellyfish (*Aurelia Aurita*): a Peliminary Experimental Study with a Dual-Beam Acoustic System”, *ICES Journal of Marine Science*, 53, pp. 309-311, 1996.

[7] X. Tang and W. K. Stewart. “Plankton Image Classification Using Novel Parallel-Training Learning Vector Quantization Network”, *Proc. IEEE/MTS OCEANS ‘96.* v.*3,* pp. 1227-1236. 1996.

[8] E. Trucco and A. Verri. *Introductory Techniques for 3-D Computer Vision*. Prentice Hall. pp. 108-112. 1998.

[9] X. Yuan, Z. Hu, J. Chen, R. Chen, P. Liu. “Online Learning and Object Recognition for AUV Optical Vision”, *IEEE Int. Conf. on Systems, Man, and Cybernetics, 1999.* v.6, pp. 857-862. 1999.