# IMPROVING PERFORMANCE OF A JELLY-TRACKING UNDERWATER VEHICLE USING RECOGNITION OF ANIMAL MOTION MODES

**Aaron M. Plotnik and Stephen M. Rock**
**Aerospace Robotics Laboratory, Stanford University**
Stanford, CA 94305
and
**Monterey Bay Aquarium Research Institute**
Moss Landing, CA 95039
Phone: 650-723-3608. Email: aplotnik@arl.stanford.edu, rock@arl.stanford.edu

## ABSTRACT

A vision-based automatic tracking system for gelatinous animals has been developed and demonstrated under a program of joint research between the Stanford University Aerospace Robotics Lab and the Monterey Bay Aquarium Research Institute (MBARI). In field tests using MBARI's ROV *Ventana* in the Monterey Bay, this system has demonstrated fully autonomous closed-loop control of *Ventana* to track a jellyfish for periods up to 1.5 hours. In these tests, conventional PID and Sliding Mode Control laws have both been used that rely primarily on the measurement of relative position errors derived from the vision-based system. This tracking system has been designed for both ROV and AUV deployments.

One difference between the logic embedded in this system and the way human pilots operate is that human pilots typically exploit their *a priori* knowledge of how a jellyfish moves in formulating their control commands. That is, they do not rely solely on lead information determined through differentiation. Presented here is a first step for incorporating this additional knowledge-based lead information into the automatic control system. The ultimate goal is determine if this can be used to improve the overall performance and robustness of the tracking task.

A key step towards quantification of motion behavior of gelatinous animals is a reliable capability to detect motion mode changes. This paper focuses on recognition of mode changes by applying techniques in real-time computer vision and supervised machine learning in the form of a support vector machine (SVM). Methods are presented to distinguish between active and resting modes, and to detect and measure rhythmic patterns in the body motions of these animals.

## 1 INTRODUCTION

A vision-based automatic tracking system has successfully tracked gelatinous animals in Monterey Bay, California as part of a joint research program between Stanford University and the Monterey Bay Aquarium Research Institute (MBARI). This was done using the MBARI tethered remotely operated vehicle (ROV) *Ventana*. Complete descriptions of this system are presented in [1]-[3].

A block diagram of the system's control architecture is shown in Fig. 1. This system measures relative position to the tracked animal using vision algorithms. Relative velocities are determined by differentiation. Control using PID and Sliding Mode Control laws command the vehicle thrusters, which are summed with the ROV pilot's joystick commands.

A key observation is that the gelatinous animals being tracked by this system exhibit motion behavior that is strongly modal. Human pilots performing the tracking task very naturally recognize this behavior and anticipate motions. They use this knowledge as lead information to help them operate an effectively low gain control loop that successfully keeps the animal in the limited field of view of the ROV cameras.

The current generation automatic tracking system does not exploit this knowledge of animal behavior. To add these capabilities, the autonomous system must model and quantify the animal's motion behavior. This requires recognition of the motion modes of the animals, and identification of dynamical models to associate with each mode. By introducing the additional lead information derived from this knowledge into the system, robustness of the automatic tracker may be enhanced, helping to achieve very long periods of uninterrupted tracking. An architecture for deriving and integrating a dynamical model of the animal's motion behavior is depicted in Fig. 2.

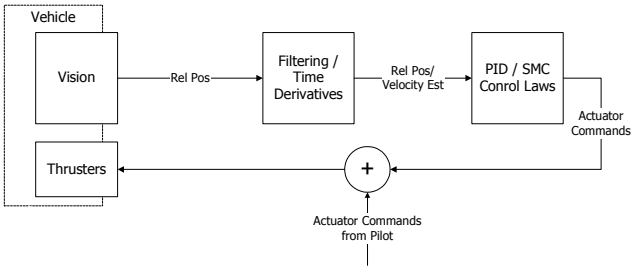



Fig. 1. Architecture for automatic jelly-tracking as mounted on the ROV *Ventana*.

A key step for an autonomous system to quantify the motion behavior of gelatinous animals is recognition of the mode of the animal. This paper presents methods to identify the motion mode of a tracked gelatinous animal using techniques in real-time computer vision and supervised machine learning in the form of a support vector machine (SVM). This paper builds upon recent work toward these goals presented in [4]. Section 2 describes the motion behavior of gelatinous animals and introduces a mode model that captures the key characteristics of that behavior. Section 3 presents algorithms for mode identification and the results from applying them to several underwater video samples.

## 2 MODELING OF MOTION BEHAVIOR

In order for an automated system to monitor the motion modes of gelatinous animals, a model of the modes must be established as a reference to compare with observed motion patterns. In this section, the motion behaviors exhibited by gelatinous animals are enumerated, and a mode model derived from those behaviors is established.

### 2.1 Motion Behavior of Gelatinous Animals

Gelatinous animals generally swim by deforming some part of their bodies (or the entire body) in a pulsing motion. Pumping water in and out of the bell portion of their bodies generates a thrust force on the surrounding water. For medusa jellyfish, a dynamic thrust equation given by (1) shows thrust, $T$, related to water density, $\rho$, and generated by the time rate of change in the volume of the bell, $V$, and the velar area, $A_v$, as proposed by Daniel [5].

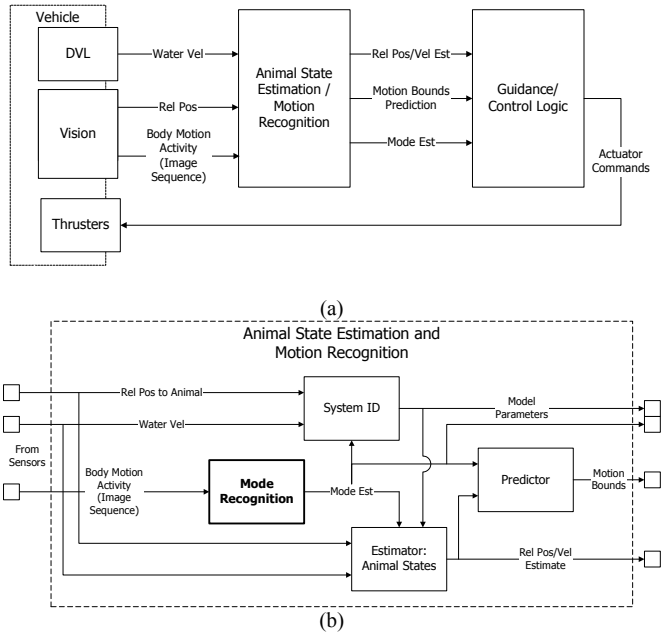
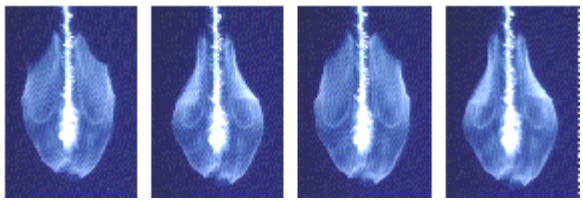
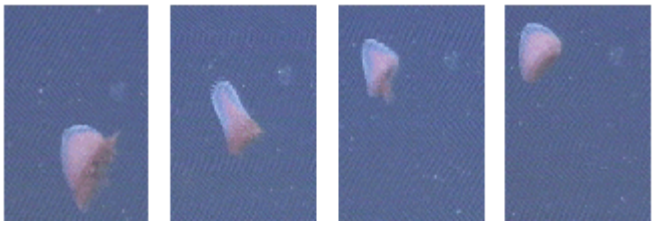


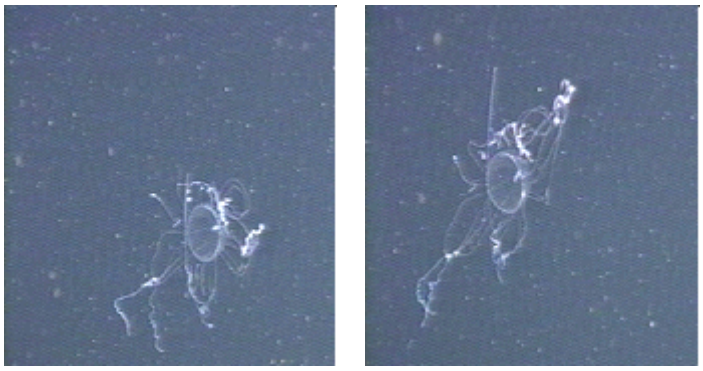

Fig. 2. Augmented architecture for improved jelly tracking. The guidance and control algorithms are provided with lead information derived from a model of the tracked animal's motion behavior and recognition of its motion mode. (a) is a top-level view, (b) shows a detailed view of the Animal State Estimation and Mode Recognition block of (a).



(a)
The head of a *Praya* siphonophore exhibiting repetitive pulsing motions with period of about 0.9 sec. Time-lapse at 0.43 sec intervals and stabilized. (Images courtesy of MBARI.)



(b)
A *Benthocodon* jellyfish making a single pulse swimming motion. Time-lapse at 0.27 sec intervals.



(c)
A *Colobonema* jellyfish changing its body configuration. Images taken 1.3 seconds apart.

Fig. 3. Examples of body motion behaviors.

$$T = -\frac{\rho}{A_v}\left(\frac{\delta V}{\delta t}\right) abs\left(\frac{\delta V}{\delta t}\right) \qquad (1)$$

Many species exhibit periodic bell pulsing behaviors in order to propel themselves through the water. Others exhibit these pulsing movements in non-periodic patterns. Table 1 gives some example statistics for several species in terms of body pulsing frequencies (gait) and the resulting cruise and maximum speeds.

| Species | Cruise (cm/s) | Max (cm/s) | Gait (Hz) | Ref. |
|---|---|---|---|---|
| P. camtschatica | 1-2 | 2 | 0.11-0.33 | [6] |
| Stomolophus meleagris | 4-12 | 16 | * | [7] |
| Solmissus incisa and Solmissus marshalli | 1.0-2.7 | * | 0.31-0.79 | [8],[9] |
| Aegina digitale | 2-5 | 30-40 | * | [10],[11] |
| Aegina grimaldi | 1.5-2.8 | * | 0.5-1.0 | [8] |

Table 1. Example statistics on motion behavior of several species of gelatinous animal. * indicates that data was not available in given references. (Table adapted from [3].)

Observable body motions often, but not always, lead to self-propulsion. Fig. 3 shows three image sequences with examples of these motions. In (a), the *Praya* siphonophore is swimming by the repetitive pulsing motions of its "head". In (b), a *Benthocodon* jellyfish makes a sudden swimming motion by pulsing its entire body once, then returns to resting. In (c), a *Colobonema* jellyfish changes the pose of its body, displaying a wide variation in its tentacle configuration over a short period of time. Unlike the first two examples, this body motion does not result in any self-propulsion.

It is useful to classify motions as belonging to four modes, as shown in Fig. 4. The four modes are resting, single pulsing, repetitive pulsing, and changing configuration; these modes make up a finite automaton representing the modal behavior of gelatinous animals. "Resting" is defined as the mode in which the body is not deforming significantly. This usually means the animal is passively floating with whatever currents are present or perhaps moving vertically due to a non-neutral buoyancy force. "Changing Config/Pose" means the body is deforming but not effecting any propulsion, as in Fig. 3(c). "Single Pulsing" represents an isolated pulsing motion, differentiated from "Changing Config/Pose" by the fact that the body returns to its original configuration at the end of the motion. "Repetitive Pulsing" signifies a sustained version of the "Single Pulsing" mode, and for most species represents a sustained swimming motion.

The behavioral model in Fig. 4 is a superset of possible body motion behaviors. For each species, some subset of this model is appropriate, in terms of both the modes and the transitions between them.

### 2.2 A Mode Model for Real-Time Monitoring of Animal Motions

The automaton of Fig. 4 cannot be applied as the reference model for real-time monitoring because some modes can only be identified in hindsight. Fig. 5 shows a

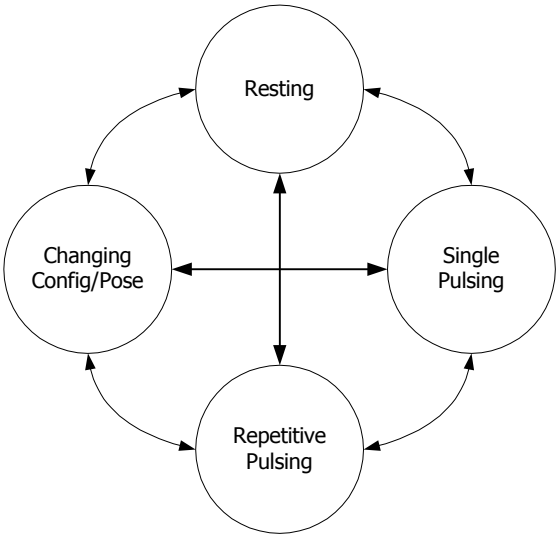



Fig. 4. Gelatinous animal body motion behavioral model.

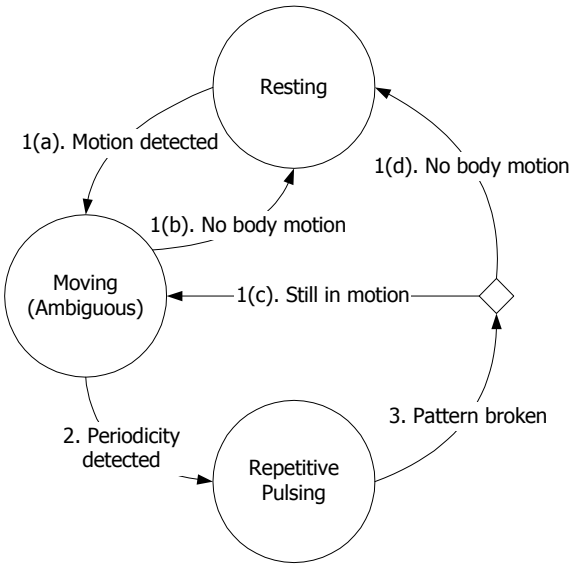



Fig. 5. Gelatinous animal body motion from point of view of observer.

finite automaton that expresses an observer's perception of mode in real-time. For instance, it is difficult to tell the difference between a single pulsing motion and a body configuration change until the motion is complete. Also, a periodic motion is not clearly periodic until multiple cycles are complete. Because of these ambiguities, Fig. 5 shows one "Moving" mode in place of "Single Pulsing" and "Changing Config/Pose". This body motion mode effectively means to the observer that the body is deforming, but it is unclear if there is a pattern to the deformations.

The model of Fig. 5 can be used as the reference model when monitoring a gelatinous animal. To identify which mode the subject is in, an observer must answer three questions. The switching events in Fig. 5 are numbered to match the question that must be answered for an observer to recognize the event.

1. 'Is the body deforming significantly?' If it is not, the animal may be considered to be resting. (Because this question is related to four switches in the mode logic diagram, those switches are labeled 1(a)-1(d).)
2. 'If so, is the motion rhythmic?' If it is, the animal is pulsing repetitively (which will usually be associated with swimming).
3. 'If it is pulsing repetitively, and its pattern changes, what is it doing?' This leads to the question, 'Is it moving in a different way or is it switching to a resting mode?'

Automated decision algorithms that answer each of these questions are introduced in Section 3.

## 3 ALGORITHMS AND EXPERIMENTAL RESULTS

For an automated sensing system to determine the body motion mode of a gelatinous animal, it must evaluate the

logic specified in Fig. 5. Three algorithms are required to perform that evaluation, one for each item in the bulleted list from the previous section. This section presents algorithms that use computer vision and supervised machine learning to achieve these objectives. They are based on analysis of self-similarity measures of the tracked animal's image over time.

Section 3.1 discusses the use of image self-similarity for analyzing body motion patterns. Section 3.2 describes the object tracking and stabilization methods used to generate sequences of images of the tracked animal. Then, the following three sections address the three questions introduced in Section 2.2. Section 3.3 provides results from applying a support vector machine to the binary classifier problem of deciding whether the animal being tracked is moving its body or not. Section 3.4 discusses detection of periodicity in the tracked animal's motion and calculation of period. Section 3.5 then discusses detection of deviations from the periodic pattern when periodicity is present. Fig. 6 depicts how these algorithms combine to determine the motion mode of the tracked animal.

### 3.1 Self-Similarity Metrics

The analytical techniques described in this paper to detect motion mode changes rely on a self-similarity metric. An example illustrates the concept of self-similarity. Fig. 3(a) shows some sequential images of a *Praya,* a type of siphonophore. This sequence shows the repetitive nature of typical pulsing motions, with the resulting images looking nearly identical every 0.9 seconds. In this sequence, the siphonophore's head has been tracked and stabilized, meaning that rigid-body translation, scaling and rotation of the subject in the image plane has been removed.

This example illustrates some basic properties of a stabilized sequence of images of a gelatinous animal. If, as in this example, relative three-dimensional rotations are small or slow, some general statements can be made. If the body is not significantly deforming over time, a short sequence will consist of images that look very similar. If it is deforming, each image in the sequence will look significantly dissimilar to recently preceding images. If there is a repetitive pulsing motion behind the deformations of the body with a period of $p$ samples, then every pair of stabilized images in the sequence that are $p$ frames apart should look very similar.

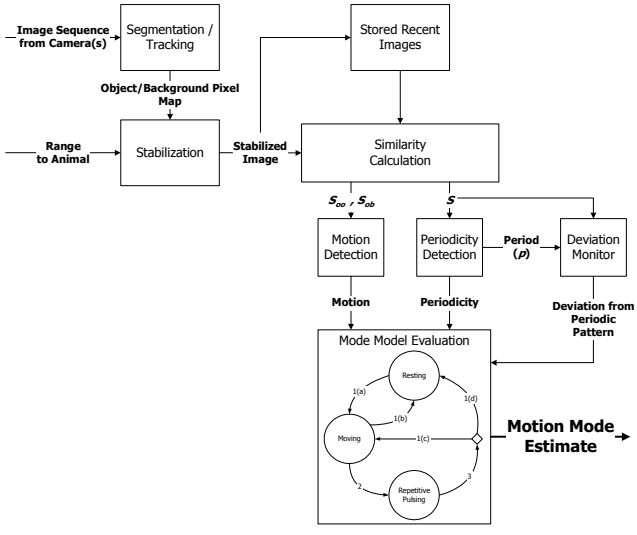



Fig. 6. Block diagram tracing steps taking an image stream of a tracked animal to determination of the body motion mode of the animal.

Cutler and Davis [12] used analysis of self-similarity of a tracked object over time to detect periodic motion in image sequences of people walking, dogs running, etc. Matching of patterns in the similarity data allowed computation of the period of the body changes and identification of the type of motion (e.g., walking vs. running, biped vs. quadruped). Periodicity was detected by placing similarity data into a "similarity matrix" and examining the "texture" of that matrix for regularity.

The computer vision and pattern recognition methods used in this paper differ from those of [12] in that the application also requires identification of non-periodic body motion and rapid recognition of deviations from repetitive motions. These are all accomplished through analysis of the animal's self-similarity over time.

To measure the similarity between two images of the tracked object, a normalized sum-of-squared-differences (SSD) is computed:

$$S(t_1,t_2) = \sum_{(x,y)}^{m,n} \left( I_{t_1}(x,y) - I_{t_2}(x,y) \right)^2 / (mn(255^2)) \qquad (2)$$

Here $I_t(x,y)$ represents the grayscale value at pixel coordinate $(x,y)$ of the stabilized image $I$ at time $t$ and $m$ and $n$ are the dimensions in pixels of the images. (Note that by using a measure of image differences, $S(t_1,t_2)$ is really a measure of dissimilarity. Comparing two identical images gives a score of zero and comparing a purely black image with a purely white image gives a score of one.) $S(t_1,t_2)$ is normalized to be a per-pixel average difference, in units of the maximum possible grayscale difference. The matrix with coordinates formed by pairs of time steps will be referred to as $S$.

The self-similarity metric used by this paper is an area-based correlation measure. Area-based correlation metrics are chosen to provide this measure. By comparing the entire area of an object's projection, many kinds of body deformations can be captured. Those include deformations that involve changes in the projected contour of the body, as well as deformations that are internal to its boundaries. Such changes might be thought of as variations in the object's visual "texture", and are not captured by tracking changes in object edges or contours. Also, good signal-to-noise ratio properties are associated with using the entire area of the object to measure changes in images [13]. Signal to noise ratio is a significant issue for this problem, because of the noise levels in typical underwater video and the small pixel area that the tracked animals often occupy in the image frame.

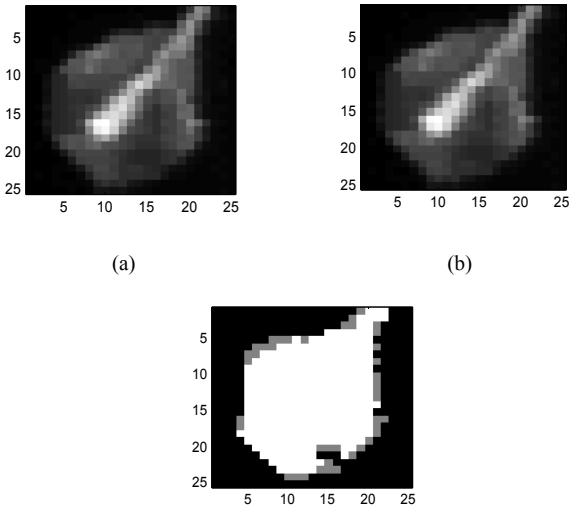



Fig. 7. Example of image regions when calculating similarity over regions defined by segmentation process. (a) and (b) are stabilized images 0.5 seconds apart. (c) shows the regions defined by comparing the segmented results from the two images. Color key: Object-object: white, background-backgournd: black, object-background: gray.

The intent of measuring self-similarity is to provide a quantitative measure of how much an animal's image projection has changed between two snapshots. By computing $S(t_1,t_2)$ as in equation (2), differences due to the background also contribute to this measurement, effectively as noise. The difference contribution for those pixels is typically quite small. Because the normalization in equation (2) includes these pixels, variations in the number of background pixels can have a strong effect on the magnitude of $S(t_1,t_2)$.

To measure changes in the animal's appearance in a more uniform way, segmentation results are also utilized in this process. Comparison measures are calculated for each of three pixel classification regions. The pixel area is divided into three regions; pixels that are segmented in both images (the object), pixels that are not segmented in both images (the background), and the remaining pixels that are classified differently by the segmenter in each image. This is illustrated in Fig. 7 for a pair of images. Self-similarity is computed using (2) separately for the pixels in these regions and will be referred to as $S_{oo}$, $S_{bb}$, and $S_{ob}$, where subscript $o$ refers to the object region and $b$ refers to the background region. The numbers of pixels in each region are defined as $n_{oo}$, $n_{bb}$, and $n_{ob}$, respectively.

Because confidence in longer-term stabilization results is low, a limit is placed on the range of time backwards for which similarity is computed. If $w$ is the maximum similarity horizon, then the similarity matrices $S$, $S_{oo}$, $S_{bb}$, and $S_{ob}$ are sparse, with entries only within $w$ of the main diagonal. For underwater clips, $w$ is typically chosen to limit comparisons to the last five to ten seconds worth of images.

The self-similarity metrics used here are close to those used in [12], but several extensions are made. The normalization of similarity scores to a per pixel scale, the additional statistics based on segmented image regions, and the short time horizon were added to address the additional analysis objectives and more challenging image stabilization issues.

Typical gelatinous animal motions register very clearly in a topographical view of a similarity matrix. This view of the motion is efficient to compute while providing a clear insight into the animal's body activity level given the right pattern identification tools. Fig. 8 shows an example similarity matrix from underwater footage of a *ptychogena* jellyfish. During the clip, the animal is resting initially, and then begins swimming by repetitive pulsing of its bell. After several seconds the animal stopped swimming, then started again for the last ten seconds of the clip. Very clear and regular textures are apparent during the periods of motion.

### 3.2 Object Tracking and Stabilization

Measurement of self-similarity requires that the subject's relative scaling and translational and rotational motions be corrected. The result is a sequence where the subject's apparent size and orientation does not change and its centroid does not move. Refer again to Fig. 3(a) for an example of a stabilized sequence.

The methods discussed in this paper require that the object be segmented and tracked in the images to correct for planar translations in the image frame. For details of the segmentation and tracking methods used in the Stanford University–MBARI jellyfish-tracking project, refer to [1]. The tracked portion of the image is then rescaled based on range measurements to normalize the object's size. Range calculated by stereo triangulation would be preferred to pixel area-based methods for this application because body deformations often cause false fluctuations in range measurements computed from object pixel area. (For

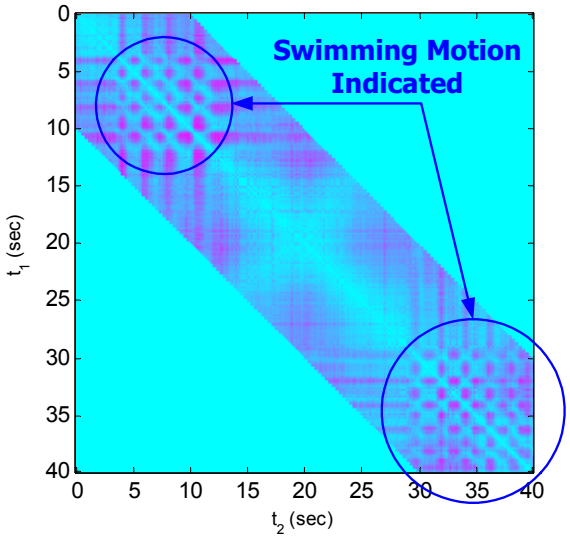



Fig. 8. Similarity matrix for *ptychogena* footage. This animal began swimming by repetitively pulsing its bell, paused for several seconds then began swimming again. The regions highlighted show a clear texture in the similarity matrix during repetitive pulsing motions.

experiments in this paper, area-based range is low pass filtered to approximate the signal quality of a stereo range.) After translation and scaling are corrected, rotations are corrected by testing rotational correlations over a small set of angles.

For this application, it is difficult to achieve precise stabilization due to the changes in viewpoint caused by the motions of the camera and the potentially unconstrained rotations of the subject in three dimensions. Therefore, it is important that the algorithms that use the results not require a long window backward in time.

### 3.3 Decision Criterion 1: Presence or Absence of Body Motion using a Support Vector Machine

This section uses the self-similarity metrics to create a decision criterion for the first of the three mode switch questions introduced in Section 2.2. Specifically, this decision queries whether or not the body of the animal is deforming significantly. The decision drives the transitions in the mode model of Fig. 5 between resting and moving modes, and the transition to either resting or moving after a deviation from repetitive pulsing is detected.

The method from [4] to detect the presence or absence of body motion yielded accurate results with a carefully chosen threshold, but did not facilitate the use of a fixed threshold when testing with different video clips, making it impractical for use on an autonomous vehicle. It was based on thresholding a moving average of $w$ similarity scores between the current image and $w$ most recent images. The algorithm was based on the premise that if the animal were not deforming significantly, a sequence of stabilized images of it would all appear to be very similar. This is quantified by equation (3),

$$\bar{s}_w(t) = \frac{1}{w}\sum_{k=0}^{w-1} S(t-kT,t) > T_m \rightarrow MOTION \qquad (3)$$

where $T$ is the image sample time period, n is the length in samples of the window backwards to consider self-similarity scores, and $T_m$ is a decision threshold.

In this paper, a supervised machine learning approach is chosen for this classification task. Supervised learning of a two-value classifier problem refers to algorithms that, given a set of $k$ feature vectors and their corresponding labeled values, find a decision function that is optimal for some criterion. That criterion might be to find the decision function that maximizes the number of test samples that are classified correctly or the maximum margin between samples and the decision boundary. In the case of detecting presence of body motion, the labels represent "moving" or "not moving", and are supplied by manually labeling several test image sequences.

Support vector machine (SVM) classifiers have been established as a mature and effective learning method in recent years [14][15]. Their structure is very general and numerous implementations are readily available, such as the LibSVM library utilized here [16]. One advantage of the SVM lies in its ability to efficiently learn in a very high (potentially infinite) dimensional space with feature vectors that can be expressed in a much smaller dimension by using kernel methods. A separating hyperplane is found in the high dimensional space that results in a maximum margin classifier. To accommodate outliers, explicit provision is made to allow some samples to be misclassified with a penalty. The resulting decision function is efficiently computed in the lower dimensional feature space of the training vectors.

The decision function of an SVM takes the form of:

$$h(x) = \sum_{i=1}^{m} \alpha_i y^{(i)} K(x^{(i)}, x) + b \qquad (4)$$

where $m$ is the number of support vectors, $K$ is a kernel function, $\alpha_i$ is a weight computed for the $i$th support vector, $y^{(i)}$ is the label of the $i$th support vector and $b$ is a computed constant offset for the hyperplane. Vectors $x$ and $x^{(i)}$ are feature vectors; $x$ is the sample being classified by $h(x)$, while $x^{(i)}$ is the $i$th support vector. The set of support vectors is the set of training samples closest to the separating hyperplane, and is typically much smaller than the training set. A decision is made by $h(x)$ classifying $x$ as having a predicted label of +1 if $h(x) > 0$, or a predicted label of -1 otherwise.

The weights $\alpha$ are computed by the following optimization:

$$\begin{aligned} \max_\alpha W(\alpha) &= \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{m} y^{(i)}y^{(j)}\alpha_i\alpha_j K(x^{(i)},x^{(i)}) \\ &\text{s.t. } 0 \le \alpha_i \le C, i = 1,\ldots,m \\ &\sum_{i=1}^{m}\alpha_i y^{(i)} = 0 \end{aligned} \qquad (5)$$

Here $C$ is a constant that allows regularization; i.e. by using this constraint, samples are allowed to be misclassified with a penalty. Without regularization, a hyperplane that perfectly separates the labeled data is found if one exists. Thus, if the training set is not separable, without regularization the optimization will not converge. Furthermore, whether or not the set is separable, regularization reduces sensitivity to outliers.

An SVM was trained and tested using a set of labeled sequences from clips of different animals in motion. Using an SVM requires choices to be made about what feature vector to process, which kernel function to use to map the feature vector to a high dimensional space, how to perform cross-validation and how to identify best choices for parameters.

The feature vector chosen for classification is made up of per-pixel similarity values separated for the two pixel regions of interest. Instead of using $S$ values, which difference entire stabilized images, segmentation results are incorporated, $S_{oo}$, $S_{ob}$ and $S_{bb}$. The feature vector is constructed using a set of weighted per-pixel averages of differences over two pixel regions, $S'$. A window of size w

is used to limit the number of images backward in time that are compared. The feature vector is for the *j*th sample is expressed by (6):

$$x^{(j)} = \begin{bmatrix} S'(j, j-T) \\ S'(j, j-2T) \\ \vdots \\ S'(j, j-wT) \end{bmatrix} \tag{6}$$

$$S'(t_1, t_2) = \frac{[n_{oo}(t_1,t_2)S_{oo}(t_1,t_2) \quad n_{ob}(t_1,t_2)S_{ob}(t_1,t_2)]^T}{n_{oo}(t_1,t_2) + n_{ob}(t_1,t_2)}$$

Note that the third region, $S_{bb}$, is disregarded; this is the region where the segmentation algorithm classified pixels as background in both images. This region contributes only noise to the process.

A Gaussian kernel function is a common choice for design of an SVM [16], and was used here. This kernel function is of the form:

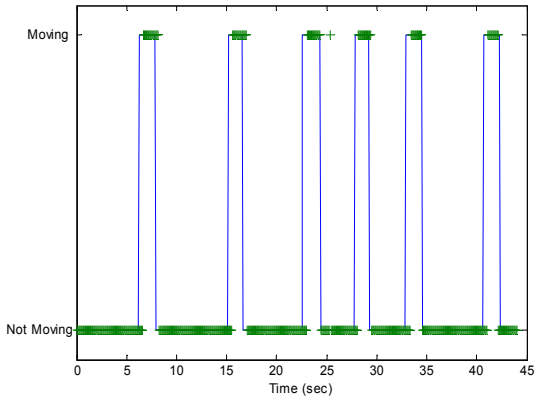
$$K(x_1, x_2) = \exp(-\gamma \|x_1 - x_2\|^2); \gamma > 0 \tag{7}$$

This kernel maps the feature vectors to an infinite dimensional space where the optimal separating hyperplane is calculated.

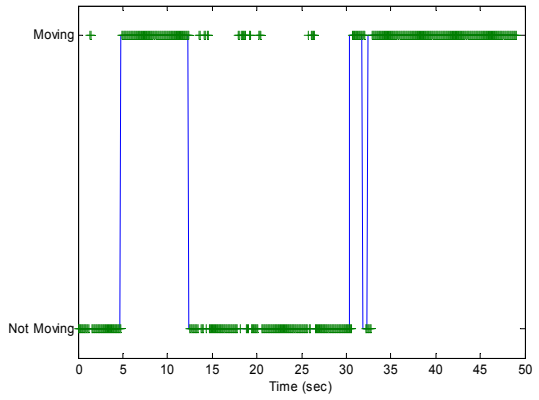
Cross-validation refers to the process of designing a classifier using a training set, then validating the results using a separate test set of data. Here this was performed by randomly choosing half of the training data and leaving that set out of the training process. This left out set became the test set. Then optimal values of $C$ and $\gamma$ were found by a designing an SVM using a grid of ($C$, $\gamma$) pairs and choosing the pair of values with the best cross-validation accuracy on the samples that were set aside for testing.

False mode switch indications are limited by requiring consecutive indications of a mode switch before validating the indication. Because the animals in question do not change modes at a high frequency, it makes sense only to trust that a mode change has occurred if it is identified for some small number of consecutive time steps. From a real-time perspective, this introduces a small amount of lag when identifying true mode changes, but can be worthwhile as a tradeoff in terms of rejecting false (extremely brief) mode change indicators. In terms of collecting statistics about the animal's motion behavior, no cost is paid by applying this extra criterion, since the first *m* mode switch indications can be relabeled after the fact.

Table 2 summarizes some results from applying this method to a set of four underwater clips of motion of four different animals. For the overall data set made up of all four clips, the best training rate before enforcing consecutive mode switch indications was 83.7%, with a training success rate of 92.4%. Note that the clip of the *solmissus* was particularly challenging in that the motion exhibited was significantly more difficult to see than for the other clips. Therefore the lower classification rate is not so surprising. Fig. 9 plots results for the first two clips in the table, both of which contained video of intermittent motions.




(a)




(b)

Fig. 9. Body motion detection results for (a) *Benthocodon* jellyfish footage and (b) *Ptychogena* jellyfish footage. A value of 1 indicates motion is present, a value of -1 indicates that no motion is present. Solid line is the labeled value by human, x's are the result of the SVM decision function. A consecutive mode change indicator of 2 consecutive steps is enforced.

| Animal | Clip Length (sec) | Behavior during clip | Success Rate (%) | Success Rate (%) when requiring 2 consecutive indications |
|---|---|---|---|---|
| Benthocodon | 45 | Intermittent single pulses | 91.1 | 91.1 |
| Ptychogena | 50 | Intermittent repetitive pulsing | 86.3 | 91.4 |
| Siphonophore | 30 | Constant repetitive pulsing | 93.1 | 96.1 |
| Solmissus | 20 | Intermittent repetitive pulsing | 79.0 | 76.5 |

Table 2. Results from applying SVM to several clips. Success rate is the percentage of samples properly classified by the SVM throughout the clip. For these results, the SVM used a Gaussian kernel with $C$ = 8 and $\gamma$= 1 and window $w$ of 10 steps at 10 Hz sample rate. Test/training data separation done by splitting data set in two in random sequence.

### 3.4 Decision Criterion 2: Detection of Cyclic Motion and Calculation of Period

This section discusses an analytical technique to implement the second switching criterion of Fig. 5. To decide that the motion of a gelatinous animal is a repetitive pulsing motion (switching criterion 2 of Fig. 5), periodicity must be detected in the body motion. If periodicity is detected, the period of motion can be calculated.

The technique used for this implementation is based on a two-dimensional view of the similarity matrix. A two-dimensional analysis of the data in the matrix $S$ looks for patterns in the comparisons between many images and past images. This contrasts with a one-dimensional approach which examines the self-similarity of only the current object image with its past images.

A two-dimensional approach is based on the idea of using the texture apparent in the 2-D view of the similarity matrix. This method asks if all (or most) of the past $n$ frames show strong similarity to their corresponding frames $p$ frames ago. To answer this question, the 2-D autocorrelation of a local portion of the similarity matrix, $A$, is computed and searched for local peaks. If the motion is periodic with period $p$, the peaks will form a lattice structure with points every $p$ steps. Local peaks of $A$ and regularly spaced lattice points are matched. If any lattice structure matches, periodicity is considered present, and the calculated period $p$ is equal to the value the lattice spacing that resulted in the minimum total matching distance error. These periodicity detection methods are based on those used in [12]. However, to address the difficulties in achieving precise stabilization results, the lattice matching requirements were relaxed somewhat to use a modified structure that forms a "cross" shape along the axes of $A$.

Fig. 10 shows an example result from the *ptychogena* footage. This figure shows a match to a "cross"-shaped lattice with spacing of 20 steps, indicating periodicity is present with a 2 second period. Pulsing period calculated for this entire clip is shown in Fig. 11(a). Pulsing periods marked as zero indicate that no periodicity was detected at those times.

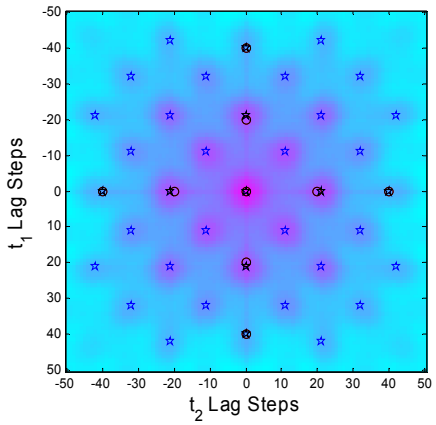



Fig. 10. Example of lattice-matching to local peaks of autocorrelation of S for *ptychogena* footage. "Cross" pattern matched to peaks at period of 2.0 sec. Stars represent local peaks in autocorrelation of the local portion of *S,* circles represent matched lattice points.

### 3.5 Decision Criterion 3: Deviation from Expected Pose

This section describes a technique to decide if a deviation from a repetitive pulsing motion has occurred (switching criterion 3 of Fig. 5). If the repetitive pattern is broken, either transition 1(c) or 1(d) is then taken based on the presence of motion detection algorithm described earlier in Section 3.3.

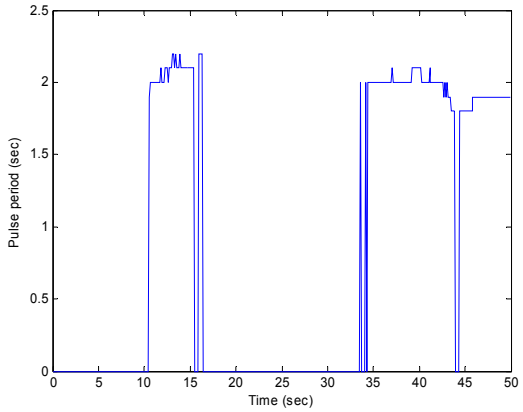
The algorithm uses a recent image as a prediction of what the animal should look like if the repetitive pulsing motion holds. When the animal is pulsing periodically, the sequence can be thought of as containing reference images that strongly resemble the current one. The computed period, $p$, is used to choose a reference image from the periodic motion image sequence. The reference image at time $t$-$p$ represents what the object is expected to look like if the pulsing pattern is still in effect. An implementation is given by:

$$r(t) = S(t-p,t) > T_r \rightarrow DEVIATION \qquad (8)$$

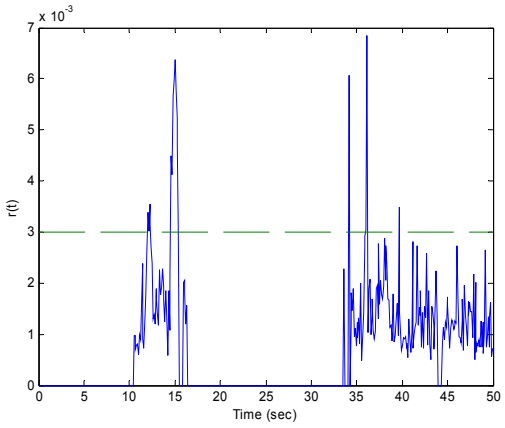
where $T_r$ is a threshold value. Note that for good performance, short-term variations in period must be tracked well by the periodicity analysis algorithm. That requirement was addressed by the design objectives defined in the previous section for that algorithm.

Although the mode switch event from "Repetitive Pulsing" can also be determined by the loss of indication that periodicity is present, this predictor of animal pose reduces the lag in identifying the change. This method can typically detect the change in pattern within only a few time steps, substantially faster than it takes for the periodicity indicator to cease to indicate periodicity. In the case of a switch from "Repetitive Pulsing" to "Resting", the body motion detection algorithm can also identify the event by detecting a lack of motion. Again however, use of this predictor can reduce the lag involved.

Results of the period calculation along with the deviation from period pattern results for the *ptychogena* footage are plotted in Fig. 11(a) and (b).




(a)




(b)

Fig. 11. (a) Period and (b) deviation from periodic pattern results for *Ptychogena* jellyfish footage. In (a), a period of zero indicates that no periodicity is detected at that time. In (b) the dotted line shows the threshold value $T_r$; when $r(t)$ is above $T_r$ , a pattern deviation is detected.

The results of combining the motion, periodicity and pattern deviation indications are plotted in Fig. 12. Here the overall mode indication from evaluating the mode model using the indications of each of the algorithms is illustrated.

## 4 CONCLUSION

With the mode model and mode identification algorithms established in this paper, a foundation is in place to improve the current generation jelly tracker through the use of a hybrid dynamical model of jellyfish motion behavior. Additional testing is required using more video footage to confirm the effectiveness of the SVM approach to motion detection across a wider variety of different animals, lighting and camera configurations. These mode identification algorithms on their own could provide the basis for automated data collection on the motion behavior of marine animals either by the animal tracking system or via off-line processing.

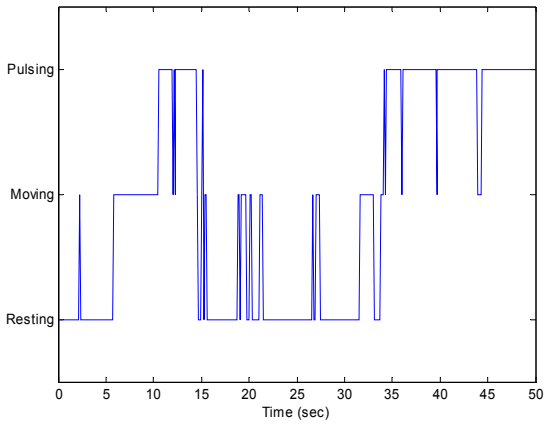



Fig. 12. The overall mode indication based on all indicators for *Ptychogena* jellyfish footage.

## REFERENCES

[1] J. Rife and S. Rock, "A Pilot-Aid for ROV Based Tracking of Gelatinous Animals in the Midwater", *Proc. IEEE/MTS OCEANS 2001*, vol. 2, pp. 1137-1144, Nov. 2001.

[2] J. Rife and S. Rock, "Field Experiments in the Control of a Jellyfish Tracking ROV", J. Rife and S. Rock, "A Pilot-Aid for ROV Based Tracking of Gelatinous Animals in the Midwater", *Proc. IEEE/MTS OCEANS 2002,* Nov. 2002.

[3] J. Rife, *Automated Robotic Tracking of Gelatinous Animals in the Deep Ocean*, thesis in *Aeronautics and Astronautics*. To be published 2003, Stanford University: Stanford, CA.

[4] A. Plotnik and S. Rock, "Quantification of Cyclic Motion of Marine Animals from Computer Vision," *Proc. IEEE/MTS OCEANS 2002,* pp. 1575-1581, Nov. 2002.

[5] T.L. Daniel. "Mechanics and energetics of medusan jet propulsion." *Canadian Journal of Zoology*, 61:1406-1420, 1983.

[6] S.W. Strand and W.M. Hamner. "Predatory behavior of Phacellophora-camtschatica and size-selection predation upon Aurelia-aurita (scyphozoa, cnidaria) in Saanich Inlet, British Columbia." *Marine Biology*, 99:409-414, 1988.

[7] R.J. Larson. "Costs of transport for the scyphomedusa Stomolophus meleagris L. Agassiz,"*Canadian Journal of Zoology*, 65:2690-2695, 1987.

[8] R.J. Larson, C.E. Mills, and G.R. Harbison. "In situ foraging and feeding-behavior of narcomedusae (cnidaria, hydrozoa)," *Journal of the Marine Biological Association of the United Kingdom*, 69:785-794, 1989.

[9] C.E. Mills and J. Goy. "In situ observations of the behavior of mesopelagic Solmissus arcomedusae (Cnidaria, Hydrozoa)," *Bulletin of Marine Science*, 43(3):739-751, 1988.

[10] S.P. Colin and J.H. Costello. "Morphology, swimming performance and propulsive mode of six co-

occurring hydromedusae," T*he Journal of Experimental Biology*, 205:427-437, 2002.

[11] S. Donaldson, G.O. Mackie and A. Roberts. "Preliminary observations on escape swimming and giant neurons in Aglantha digitale (Hydromedusae: Trachylina)," *Canadian Journal of Zoology*, 58:549-552, 1980.

[12] R. Cutler and L.S. Davis, "Robust Real-Time Periodic Motion Detection, Analysis, and Applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 781-796, August 2000.

[13] R. Marks, *Experiments in Visual Sensing For Automatic Control of an Underwater Robot*, thesis in *Aeronautics and Astronautics*. 1995, Stanford University: Stanford, CA.

[14] B. Boser, I. Guyon, and V. Vapnik. A training algorithm for optimal margin classifiers. *Proceedings of the Fifth Annual ACM Conference on Computational Learning Theory*, pages 144-152, Pittsburgh, PA, July 1992.

[15] V. Vapnik. *Statistical Learning Theory*. Wiley, 1998.

[16] Chih-Chung Chang and Chih-Jen Lin, LIBSVM: a library for support vector machines, 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.