

Quantification of Cyclic Motion of Marine Animals from Computer Vision

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 mounted on the ROV *Ventana* has successfully tracked jellyfish in Monterey Bay, California as part of a joint project between Stanford University and the Monterey Bay Aquarium Research Institute (MBARI). To enhance performance, improved lead information about the target's motion is desired. Human pilots derive lead information about that motion through the perception of motion modes of these animals and the propulsion associated with them. Although this kind of information is of a different nature than what is typically available to an automatic control system, performance gains could be achieved by incorporation of such information.

Jellyfish and related gelatinous animals exhibit very distinct modes of motion that are visually recognizable to human observers. For an autonomous animal-tracking system to interpret the motion of its target, and to generate lead information useful for control, it must first be able to identify the mode of motion of the animal under observation. This paper explores techniques in computer vision to detect and recognize the key motion modes and mode change events typical of these animals. Methods are presented to distinguish between active and resting modes, and to detect and monitor rhythmic patterns in the body motions of these animals.

I. INTRODUCTION

A vision-based automatic tracking system has successfully tracked gelatinous animals such as jellyfish in Monterey Bay, California. This experimental system is part of a joint project between Stanford University and the Monterey Bay Aquarium Research Institute (MBARI) and is installed on the MBARI tethered remotely operated vehicle (ROV) *Ventana*. This system extends the abilities of MBARI scientists to observe marine animals *in situ*. The automated version of the system takes the form of an ROV pilot assist that may be switched on and off under the supervision of the pilot. Complete descriptions of this system are presented in [1] and [2].

Further research and development is in progress to improve the performance of the automatic pilot assist and to investigate more autonomous solutions. One way to improve the performance of automatic control systems is to incorporate more lead information. Such information is typically calculated by approximating time derivatives of the controlled variables.

When a human pilot operates the ROV to track gelatinous animals, other kinds of lead information are utilized. A human operator can make observations about the animal, such as its species, and dynamic information about its pose and its mode of motion (e.g., resting vs. actively swimming). Understanding the motion capabilities of the animal and recognizing its motion modes allow bounds on the near-term

future motion to be approximated. For example, this style of lead information gives a more accurate sense of the probability of losing the animal from the field of view than is possible with conventional sensing techniques that filter a measured position. The human pilot uses that judgment to make decisions about how best to track the animal without losing it from view.

The motion of gelatinous animals of interest is generally separable into distinct modes that are recognizable from visual information. Because these animals are so strongly modal, understanding their motion fundamentally depends on the timely and correct identification of their mode. Adding the desired perceptive abilities to the autonomous tracker must therefore begin by developing techniques to infer its target's mode.

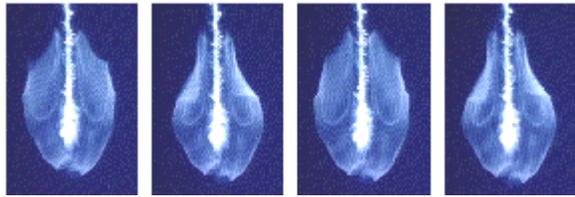
This paper describes methods to extract animal motion mode information from computer vision data. Section II examines the motion behavior of gelatinous animals and discusses how modes are inferred. Section III gives an overview of the algorithms chosen and the rationale behind them. Section IV describes the algorithms in detail and includes results from their application to underwater video samples.

II. MODELING AND OBSERVATION OF MOTION BEHAVIOR

A. Motion Behavior of Gelatinous Animals

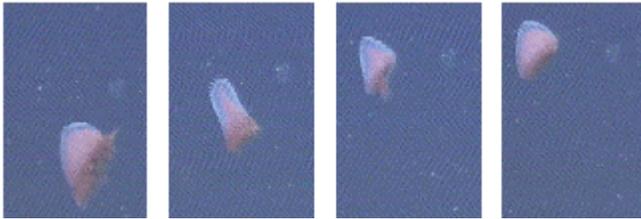
Before choosing algorithms to detect the mode of gelatinous animals, it is necessary to understand the kinds of motions they exhibit. Gelatinous animals generally swim by deforming some part of their bodies (or the entire body) in a pulsing motion. Fig. 1 shows three image sequences with examples of these motions. In (a), the 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 fairly short period of time. Unlike the first two examples, this body motion does not result in any self-propulsion.

The body motion modes of gelatinous animals may be described by a discrete model. In general, the motions may be classified into four modes, as shown in Fig. 2. "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. "Changing Config/Pose" means the body is deforming but not effecting any propulsion, as in Fig. 1(c). "Single Pulsing" represents a



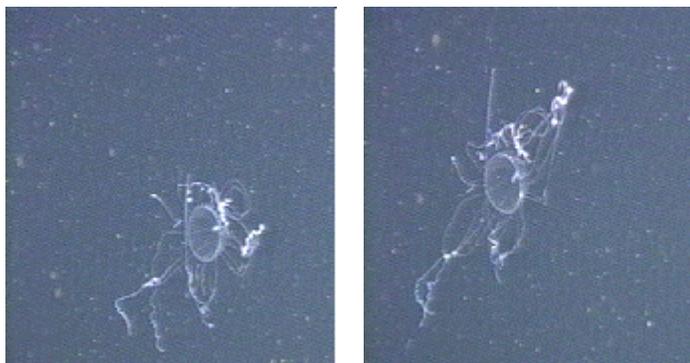
(a)

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. 1. Examples of body motion behaviors.

singular 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. 2 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.

B. Perception of Mode

Fig. 3 shows a discrete model that describes an observer’s perception of mode in real-time. Perception of mode is less distinct than what is depicted in Fig. 2 because some of those modes can be identified only in hindsight. For instance, until the motion is complete, it is difficult to tell the difference between a single pulsing motion and a body configuration change. Because of the real-time ambiguity between those two modes, in Fig. 3 they are replaced by one “Moving”

mode. Also, when the animal begins a repetitive pulsing motion, for the first period or more, the fact that a periodic motion is being witnessed is not necessarily clear, even to a human. Therefore this ambiguous body motion mode means to the observer that the body is deforming, but it is unclear if there is a pattern.

To identify which mode the subject is in, an observer must answer three questions. The switching events in Fig. 3 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.
2. ‘If so, is the motion rhythmic?’ If it is, the animal is pulsing repetitively (which will usually be associated with swimming). This leads to a secondary question, ‘If it is pulsing repetitively, what is the pulsing frequency?’ If the deformations are not identifiably rhythmic, the motion in progress is a configuration change, a single pulse, or the beginning of a repetitive motion.
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?’

Identification of the body motion mode of the animal as “Resting” mode then merely requires a “No” answer to the first question, ‘Is its body deforming?’ A subject that is perceived to be moving is reclassified as “Repetitively Pulsing” if the body motion is identified as rhythmic. Detection of transitions from “Repetitively Pulsing” to one of the other modes requires the observation that the pattern has been broken. Then, if body motion is still present, a new pattern may be forming, or it is possible that the animal may be just reconfiguring its body shape. Because the first question is related to four switches in the mode logic diagram, those switches are labeled 1(a)-1(d).

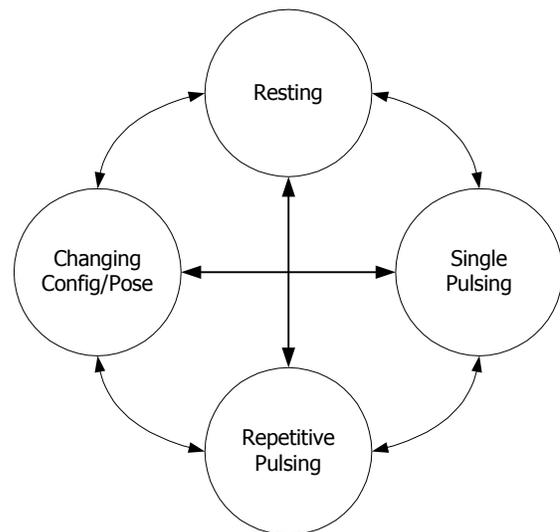


Fig. 2. Gelatinous animal body motion behavioral model.

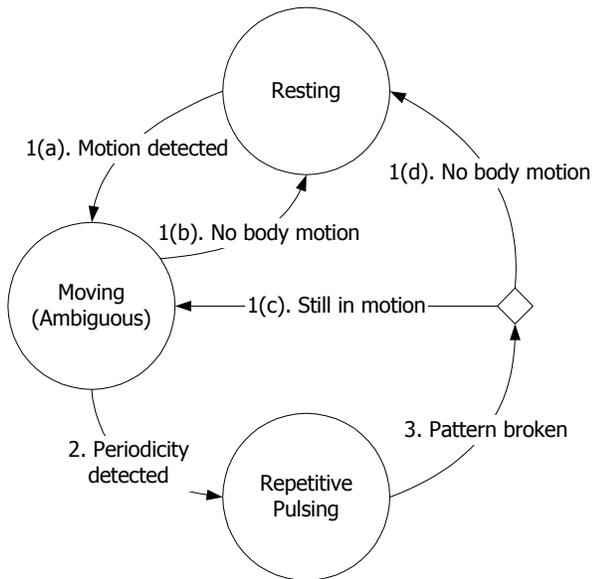


Fig. 3. Gelatinous animal body motion from point of view of observer.

III. RATIONALE FOR ALGORITHMS

A. Algorithms for Mode Determination

For an automated sensing system to determine the body motion mode of a gelatinous animal, it must evaluate the logic specified in Fig. 3. Three algorithms are required to perform that evaluation. An algorithm that detects the presence or absence of significant body deformations is required to evaluate the switching criteria labeled 1(a)-1(d). To identify repetitive pulsing (switching criterion 2), a means to detect periodicity in the motion is required. To recognize deviations from periodic motion (switching criterion 3), an algorithm is required that determines if the body is moving as expected based on the recent pattern.

The algorithms in this paper use computer vision to achieve these objectives. They are based on analysis of self-similarity measures of the tracked animal's image over time. To measure the evolution of self-similarity, stabilized image sequences that isolate the animal's body deformations from its overall rigid-body motions are required. The self-similarity data are then analyzed for indications of motion in general and for indications of periodicity.

B. Body Deformations in Sequences of Images

A human observer using primarily visual information can accurately determine which of the modes of Fig. 3 the animal is in with little difficulty or delay. For many of the animals of interest, the motions look distinct enough that they can be identified with numerical vision-processing techniques.

Fig. 1(a) shows some sequential images of a *Praya*, a type of siphonophore. This 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. Also, 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. (Note that because the stabilization process accumulates error and the objects of interest can and do rotate out of the image plane, that strong similarity may not hold for more than a period or two of past images.)

These properties of image sequences of a gelatinous animal provide a foundation for the algorithms required for mode identification. Quantitative methods that test those properties can be used for the recognition of body motion and the detection of periodicity in that motion.

C. Self-Similarity Metrics for Quantification of Body Motion

The stated properties of image sequences of a gelatinous animal are all based on the similarity between the images in the sequence. Analysis of the motion of an animal based on those properties requires a quantitative way to measure its self-similarity over time.

Area-based correlation metrics are chosen to provide this measurement of self-similarity. 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 [3]. This is important for this problem, most importantly because of the small pixel area that the tracked animals often occupy in the image frame.

Recent research by Cutler and Davis [4] used a similar approach 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 thresholding a one-dimensional power spectral density of the similarity between the current image and all other images in the sequence. The authors then found more consistent results by placing similarity data into a "similarity matrix" and computing a two-dimensional autocorrelation matrix from it. Period was then computed by fitting the best lattice over varying sizes and structures to the autocorrelation local peaks. Other methods to measure periodicity in image sequences were investigated in [5] and [6]. Those authors used approaches that were based on tracking the motion of individual pixels and point features through image and time space. While some advantages could be gained by such

approaches, the structure of gelatinous animals typically does not result in easily trackable point features.

This paper presents methods that are similar to those of [4], but with extensions to extract additional information from the self-similarity metrics. Recognition of periodic motion is important for this application, but identification of the modes of gelatinous animals also requires indications of the presence of non-periodic body motion and the rapid detection of deviations from repetitive motions. These are all accomplished through analysis of the animal’s self-similarity over time.

IV. ALGORITHMS AND EXPERIMENTAL RESULTS

A. Object Tracking and Stabilization

The algorithms used to analyze the body motions of a gelatinous animal require a measure of self-similarity over time. 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. 1(a) for an example of a stabilized sequence.

By segmenting the object from each image and tracking its centroid’s position over time, translations in the image plane are corrected. Rescaling based on range measurements normalizes the object’s size. Range calculated by stereo triangulation is preferred to pixel area-based methods for this application because body deformations often cause false fluctuations in range measurements computed from object pixel area. After translation and scaling are corrected, rotations are corrected by testing rotational correlations over a small set of angles. For details of the segmentation and tracking methods used in the Stanford University–MBARI jellyfish-tracking project, refer to [1].

Many characteristics of the motion of underwater animals, the motion of the camera, and the underwater environment conspire to make object stabilization from this video a far from perfect process. The greatest difficulties are due to the changes in viewpoint caused by the motions of the camera and the potentially unconstrained rotations of the subject in three dimensions. Because of the inherent difficulties in achieving precise image stabilization in this environment, it is important that the algorithms that use the results not require a long window backward in time.

B. Computation of Self-Similarity

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

$$S(t_1, t_2) = \sum_{(x,y)}^{m,n} (I_{t_1}(x, y) - I_{t_2}(x, y))^2 / (mn(255^2)) \quad (1)$$

Here $I_i(x,y)$ 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

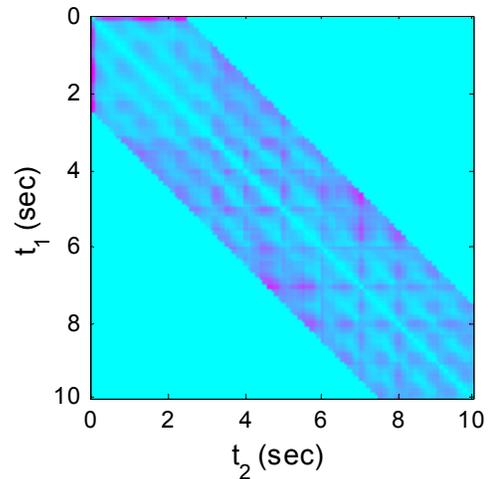


Fig. 4. Similarity matrix for siphonophore footage.

of zero and comparing a purely black image with a purely white image gives a score of one.) It is useful to store similarity data in a matrix with coordinates formed by pairs of time steps. This matrix will be referred to as S .

The normalization in (1) gives a sense of units to the similarity measurement. This facilitates choices of standard thresholds for analysis of similarity metrics that make sense from sequence to sequence.

Because confidence in longer-term stabilization results is low, a limit is placed on the range of time backwards for which similarity is computed. Let w be the maximum similarity horizon. Then the similarity matrix S is sparse, with entries only within w of the main diagonal.

Fig. 4 shows the similarity matrix from underwater footage of a siphonophore. The animal exhibits periodic body deformations with a period of approximately 0.9 seconds. This can be seen in the fairly uniform patterns in the matrix. Note the dark line down the main diagonal, indicating perfect similarity (zero) when comparing an image to itself. Moving upwards on each column results in periodic minima in the similarity score, indicating that at every p steps, the past image looks most similar to the current image.

The computation of self-similarity here is close to the method used in [4]. Better results were achieved for this application by substituting SSD for sum of absolute differences (SAD). The normalization of similarity scores and the short time horizon were added to address the additional analysis objectives and more challenging image stabilization issues, respectively.

C. Presence or Absence of Body Motion

Determination of the mode of the gelatinous animal under observation requires algorithms that can evaluate the switching criteria of the mode logic in Fig. 3. To evaluate each of the transitions labeled 1(a)-1(d), a method to detect the presence or absence of body motion is necessary. From this, mode changes between “Resting” and “Moving” are

detected. Also, after detection of an exit from “Repetitive Pulsing”, this allows an appropriate choice to be made between “Resting” and “Moving” modes.

An algorithm for this purpose is based on the first property stated about a stabilized sequence: If the animal is not deforming significantly, a sequence of stabilized images will look very similar. This implies that the values of $S(t-T, t)$, $S(t-2T, t)$, etc. will all be very low. (T is the image sample time period.) Therefore an appropriate test for motion is to compare the average of the last n similarity values for the current image to a threshold, T_m , or:

$$\bar{s}_n(t) = \frac{1}{n} \sum_{k=0}^{n-1} S(t-kT, t) > T_m \rightarrow \text{MOTION} \quad (2)$$

Analysis of footage of a *Benthocodon* demonstrates this algorithm. During this sequence, the animal exhibits single pulse swimming motions intermittently. Application of (2) with a window of a half second for averaging successfully identifies body motion, as shown in Fig. 5(a). False indications were prevented without adding much delay by requiring consecutive indicators of a mode switch before validating the indication. Results for the siphonophore footage are plotted in Fig. 5(b). For that video, the animal is continuously in motion, which this algorithm correctly determines after a short lag.

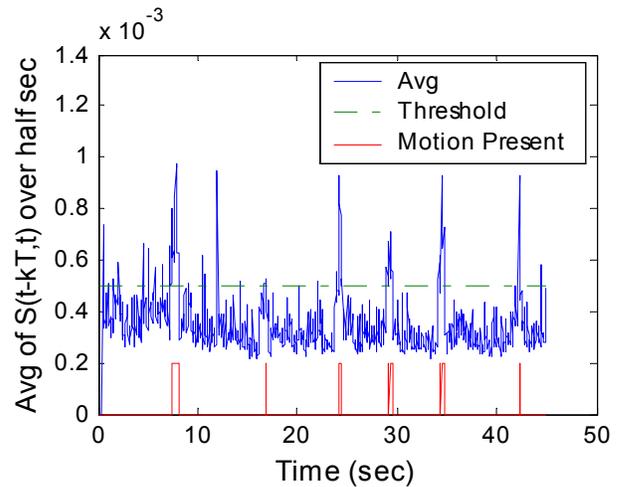
This approach results in very quick detection of the start of motion from rest. The first frame of significant change in pose will register as highly dissimilar to all of the last half second’s worth of images, so the average in (2) quickly crosses an appropriately chosen threshold. Indications that body motion has ceased are delayed longer because several highly similar images must be observed before the moving average will drop.

Further study is needed to determine how best to choose the threshold value and window for the moving average. One challenge lies in choosing an appropriate threshold for different objects, since one value may not give good results across many tracking subjects. Also, rigid body rotations out of the image plane, which cannot be removed by image stabilization, can be expected to register falsely as significant body deformations due to the dissimilar sequence of images that results. Further study is needed to find ways to discern between these rotations (still considered “Resting” by a human observer) and true changes in body configuration.

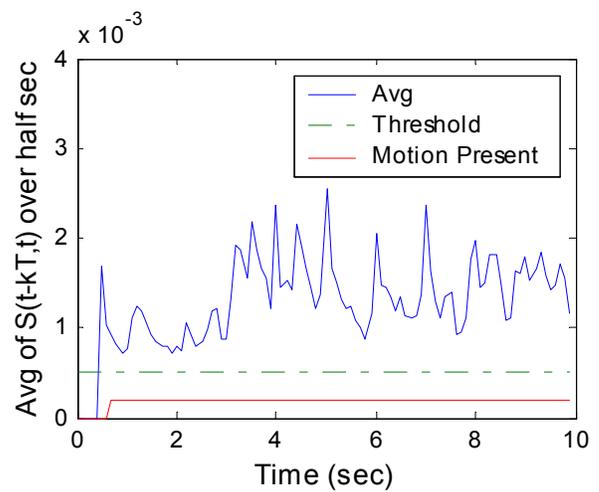
D. Detection of Cyclic Motion and Calculation of Period

To decide that the motion of a gelatinous animal is a repetitive pulsing motion (switching criterion 2 of Fig. 3), periodicity must be detected in the body motion. If periodicity is detected, the period of motion can be calculated.

This algorithm is subject to a number of objectives. One is to detect periodicity quickly, therefore from a limited amount of data. That goal is also driven by the constraint of the short window in time for which image stabilization results are trusted to be accurate. Another objective is accuracy and smoothness of the calculated period. While that is not necessary for detecting entry into a repetitive pulsing mode, it becomes important when detecting the exit from this mode.



(a)



(b)

Fig. 5. Body motion detection results for (a) *Benthocodon* jellyfish footage and (b) *Praya* siphonophore footage. The “Motion Present” indicator is superimposed below the data, with a non-zero value indicating body motion.

Similarity data may be analyzed from either one-dimensional or two-dimensional views. A one-dimensional approach examines the self-similarity of the current object image with its past images. A two-dimensional analysis of the data in the matrix S looks for patterns in the comparisons between many images and past images. Both approaches to detecting periodicity are explored here.

The one-dimensional view uses the last column of similarity matrix S , which contains the similarity measurements between the current image and recent past images. The motions of interest are typically at a single frequency. Fourier analysis is used via a one-dimensional power spectrum of that signal, which is examined with a criterion of the form:

$$P(f_i) > \mu_p + K\sigma_p \quad (3)$$

where $P(f_i)$ is the power spectrum at the i th frequency bin, μ_p is the mean of the power spectrum, σ_p is the standard deviation, and K is a constant typically in the range of 2 to 3. A frequency with energy exceeding that limit is considered to be significant. The time-dependent Fourier transform is used when computing the power spectrum, with data windowed by a Hanning window function.

A two-dimensional approach is based on the idea of using the texture apparent in the 2-D view of the similarity matrix. Rather than deciding if the current image has periodicity in similarity with past images, this method asks if all (or most) of the past n frames show strong similarity to their corresponding frames p frames ago. If so, the object has a body motion period of p time steps.

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 should form a lattice structure with points every p steps. Candidate lattice structures of spacing d are compared with the layout of the local peaks of A . Local peaks of A and the lattice points are matched one-to-one by closest proximity and a check is made that an acceptably high percentage of available peaks and lattice points are matched. If any lattice structure matches, periodicity is considered present, and the calculated period p is equal to the value of d that resulted in the minimum total matching distance error.

These periodicity detection methods are based on those used in [4]. There, lattice structures that span the entire area of the autocorrelation matrix A were used. However, tests on underwater video found that the peaks along the center axes were consistently spaced as expected, but often the off-axis peaks were less regularly spaced. Because of this, the matching requirements were relaxed somewhat to use a modified lattice structure that forms a “cross” shape along the axes of A . Fig. 6 shows an example result from the siphonophore footage. This figure shows a match to a “cross”-shaped lattice with spacing of 9 steps, indicating periodicity is present with a 0.9 second period.

Fig. 7 compares the calculated pulsing frequency over time determined by the 1-D and 2-D approaches. For this example, the 2-D algorithm better meets the requirements for smoothness and fast identification of periodicity. In this video, the siphonophore pulses with a frequency of about 1.1 Hz continuously throughout the 10-second clip. The 2-D method identifies periodicity after only one period of the pulsing, while the 1-D approach takes two to three periods.

To measure the accuracy of the period calculation, the similarity between the current image and the image taken p steps ago is compared for the two methods. The value of $S(t-p, t)$ should be very low if the period was correctly identified, meaning the reference image at time $t-p$ looks very close to the current image. For the same example, this quantity is plotted in Fig. 8. The 2-D texture-based approach outperforms the 1-D approach for this example.

E. Deviation from Expected Pose

To decide that the repetitive pulsing motion of a gelatinous animal has ended (switching criterion 3 of Fig. 3), a deviation from the motion pattern must be detected. If the pattern is broken, either transition 1(c) or 1(d) is then taken based on the motion detection algorithm described in Part C.

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. A somewhat degenerate case of this idea was already applied in detecting the start of body motions from rest in Part C.

This algorithm uses the computed period, p , 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:

$$r(t) = S(t - p, t) > T_r \rightarrow \text{DEVIATION} \quad (4)$$

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 no longer register 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

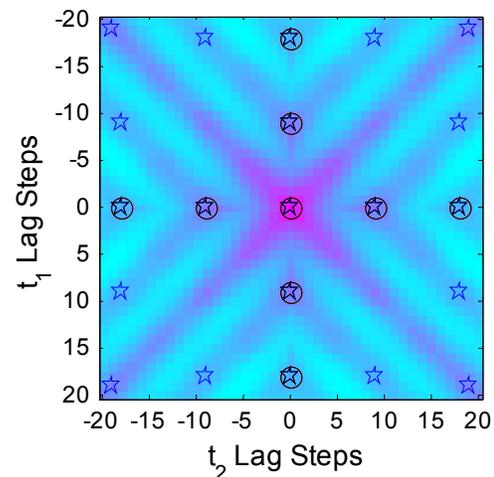


Fig. 6. Example of lattice-matching to local peaks of autocorrelation of S for siphonophore footage. “Cross” pattern matched to peaks at period of 0.9 sec. Stars represent local peaks in autocorrelation of the local portion of S , circles represent matched lattice points.

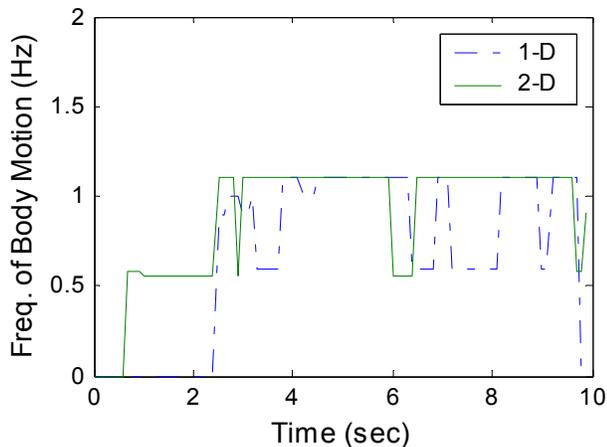


Fig. 7. Periodicity results for siphonophore footage.

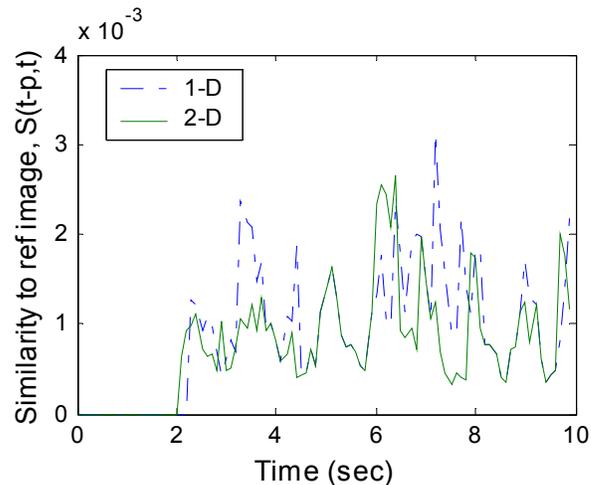


Fig. 8. Similarity measure of image at time t to reference image one calculated period backwards in time for siphonophore footage. Lower values indicate a better match to the reference image.

motion. Again however, use of this predictor can reduce the lag involved.

V. CONCLUSION

The current generation jellyfish tracker incorporates no model of jellyfish behavior. Using a discrete behavior mode model to generate lead information can improve the performance of the tracking system. Basic tools to identify body motion modes of gelatinous animals from computer vision were explored and tested. Further refinement of the algorithms proposed here will be explored to improve performance and robustness of mode identification capabilities. With timely and accurate determination of target mode, a degree of motion bound prediction and improved control approaches may be possible by the underwater tracking system.

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] R. Marks, *Experiments in Visual Sensing For Automatic Control of an Underwater Robot*, thesis in *Aeronautics and Astronautics*. 1995, Stanford University: Stanford, CA.
- [4] 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.

- [5] R. Polana and R. Nelson, "Detection and Recognition of Periodic, Non-Rigid Motion," *Int'l Journal of Computer Vision*, vol. 23, no. 3, pp. 261-282, June/July 1997.
- [6] S. M. Seitz and C.R. Dyer, "View-Invariant Analysis of Cyclic Motion," *Int'l Journal of Computer Vision*, vol. 25, no. 3, pp. 1-23, 1997.