Underwater Human-Robot Interaction via Biological Motion Identification

Junaed Sattar and Gregory Dudek
Center for Intelligent Machines
McGill University
Montréal, Québec, Canada H3A 2A7.
Email: {junaed, dudek}@cim.mcgill.ca

***Abstract*— We present an algorithm for underwater robots to visually detect and track human motion. Our objective is to enable human-robot interaction by allowing a robot to follow behind a human moving in (up to) six degrees of freedom. In particular, we have developed a system to allow a robot to detect, track and follow a scuba diver by using frequency-domain detection of biological motion patterns. The motion of biological entities is characterized by combinations of periodic motions which are inherently distinctive. This is especially true of human swimmers. By using the frequency-space response of spatial signals over a number of video frames, we attempt to identify signatures pertaining to biological motion. This technique is applied to track scuba divers in underwater domains, typically with the robot swimming behind the diver. The algorithm is able to detect a range of motions, which includes motion directly away from or towards the camera. The motion of the diver relative to the vehicle is then tracked using an Unscented Kalman Filter (UKF), an approach for non-linear estimation. The efficiency of our approach makes it attractive for real-time applications onboard our underwater vehicle, and in future applications we intend to track scuba divers in real-time with the robot. The paper presents an algorithmic overview of our approach, together with experimental evaluation based on underwater video footage.**

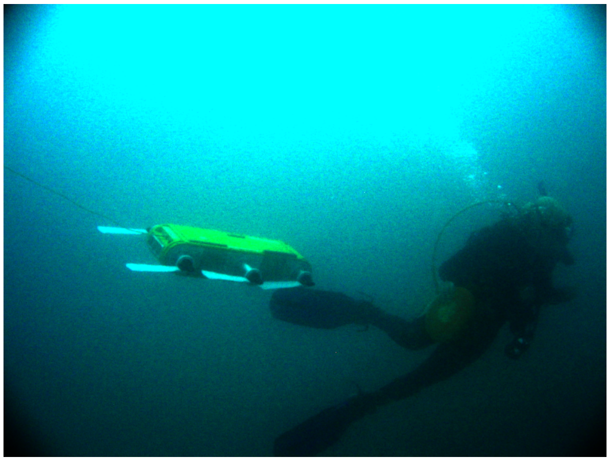


Fig. 1. An underwater robot servoing off a colored target carried by a diver.

I. Introduction

Motion cues have been shown to be powerful indicators of human activity and have been used in the identification of their position, behavior and identity. In this paper we exploit motion signatures to facilitate visual servoing, as part of a larger human-robot interaction framework. From the perspective of visual control of an autonomous robot, the ability to distinguish between mobile and static objects in a scene is vital for safe and successful navigation. For the vision-based tracking of human targets, motion patterns are an important signature, since they can provide a distinctive cue to disambiguate between people and other non-biological objects, including moving objects, in the scene. We look at both of these features in the current work.

Our work exploits motion-based tracking as one input cue to facilitate human-robot interaction. While the entire framework is outside the scope of this paper, an important sub-task for our robot, like many others, is for it to follow a human operator (as can be seen in Fig.1). We facilitate the detection and tracking of the human operator using the spatio-temporal signature of human motion. In practice, this detection and servo-control behavior is just one of a suite of vision-based interaction mechanisms. In the context of servo-control, we need to detect a human, estimate his image coordinates (and possible image velocity), and exploit this in a control loop. We use the periodicity inherently present in biological motion, and swimming in particular, to detect human scuba divers. Divers normally swim with a distinctive kicking gait which, like walking, is periodic, but also somewhat individuated. In many practical situations, the preferred applications of UAV technologies call for close interactions with humans. The underwater environment poses new challenges and pitfalls that invalidate assumptions required for many established algorithms in autonomous mobile robotics. While truly autonomous underwater navigation remains an important goal, having the ability to guide an underwater robot using sensory inputs also has important benefits; for example, to train the robot to perform a repetitive observation or inspection task, it might very well be convenient for a scuba diver to perform the task as the robot follows and learns the trajectory. For future missions, the robot can use the information collected by following the diver to carry out the inspection. This approach also has the added advantage of not requiring a second person teleoperating the robot, which simplifies the operational loop and reduces the associated overhead of robot deployment.

Our approach to track scuba divers in underwater video footage and real-time streaming video arises thus from the need for such semi-autonomous behaviors and visual human-robot interaction in arbitrary environments. The approach is computationally efficient for deployment on-board an autonomous underwater robot. Visual tracking is performed in the spatio-temporal domain in the image space; that is, spatial frequency variations are detected in the image space in different motion directions across successive frames. The frequencies associated with a diver's gaits (flipper motions) are identified and tracked. Coupled with a visual servoing mechanism, this feature enables an underwater vehicle to follow a diver without any external operator assistance, in environments similar to that shown in Fig. 2.

The ability to track spatio-temporal intensity variations using the frequency domain is not only useful for tracking scuba divers, but also can be useful to detect motion of particular species of marine life or surface swimmers [6]. It is also associated with terrestrial motion like walking or running, and our approach seems appropriate for certain terrestrial applications as well. It appears that most biological motion underwater as well as on land is associated with periodic motion, but in this paper we concentrate our attention to tracking human scuba divers and servoing off their position. Our robot is being developed with marine ecosystem inspection as a key application area. Recent initiatives taken for protection of coral reefs call for long-term monitoring of such reefs and species that depend on reefs for habitat and food supply. We envision our vehicle to have the ability to follow scuba divers around such reefs and assist in monitoring and mapping of distributions of different species of coral.

The paper is organized in the following sections: in Sec. II we look at related work in the domains of tracking, oriented filters and spatio-temporal pattern analysis in image sequences, Kalman filtering and underwater vision for autonomous vehicles. Our Fourier energy-based tracking algorithm is presented in Sec. III. Experimental results of running the algorithm on video sequences are shown in Sec. IV. We draw conclusions and discuss some possible future directions of this work in Sec. V.

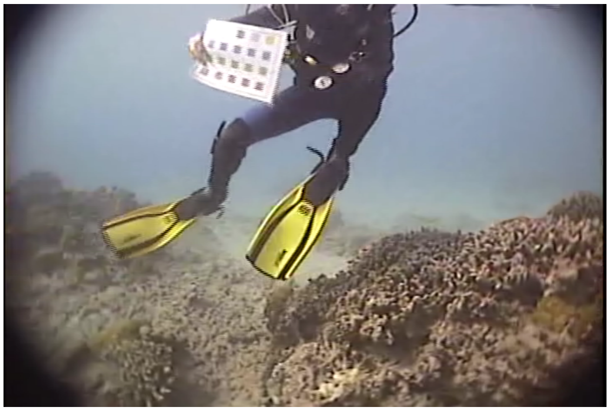


Fig. 2. Typical visual scene encountered by an AUV while tracking scuba divers.

II. Related Work

The work presented in this paper combines previous research in different domains, and its novelty is in the use of frequency signatures in visual target recognition and tracking, combined with the Unscented Kalman Filter for tracking 6-DOF human motion. In this context, 6-DOF refers to the number of degrees of freedom of just the body center, as opposed to the full configuration space. In the following paragraphs we consider some of the extensive prior work on tracking of humans in video, underwater visual tracking and visual servoing in general.

A key aspect of our work is a filter-based characterization of the motion field in an image sequence. This has been a problem of longstanding relevance and activity, and were it not for the need for a real-time low-overhead solution, we would be using a full family of steerable filters, or a related filtering mechanism [2, 3]. In fact, since our system needs to be deployed in a hard real-time context on an embedded system, we have opted to use a sparse set of filters combined with a robust tracker. This depends, in part, on the fact that we can consistently detect the motion of our target human from a potentially complex motion field. Tracking humans using their motion on land, in two degrees of freedom, was examined by Niyogi and Adelson [8]. They look at the positions of head and ankles, respectively, and detect the presence of a human walking pattern by looking at a "braided pattern" at the ankles and a straight-line translational pattern at the position of the head. In their work, however, the person has to walk across the image plane roughly orthogonal to the viewing axis for the detection scheme to work.

There is evidence that people can be discriminated from other objects, as well as from one another, based on motion cues alone (although the precision of this discrimination may be limited). In the seminal work using "moving light displays", Rashid observed [10] that humans are exquisitely sensitive to human-like motions using even very limited cues. There has also been work, particularly in the context of biometric person identification, based on the automated analysis of human motion or walking gaits [16, 7, 15]. In a similar vein, several research groups have explored the detection of humans on land from either static visual cues or motion cues. Such methods typically assume an overhead, lateral or other view that allows various body parts to be detected, or facial features to be seen. Notably, many traditional methods have difficulty if the person is walking directly away from the camera. In contrast, the present paper proposes a technique that functions without requiring a view of the face, arms or hands (either of which may be obscured in the case of scuba divers). In addition, in our particular tracking scenario the diver can point directly away from the robot that is following him, as well as move in an arbitrary direction during the course of the tracking process.

While tracking underwater swimmers visually has not been explored in great depth in the past, some prior work has been done in the field of underwater visual tracking and visual servoing for autonomous underwater vehicles. Naturally, this is closely related to generic servo-control. In that context, on-line real-time performance is crucial. On-line tracking systems, in conjunction with a robust control scheme, provide underwater robots the ability to visually follow targets underwater [14]. Previous work on spatio-temporal detection and tracking of biological motion underwater has been shown to work well [12], but only when the motion of the diver is directly towards or away from the camera. Our current work looks at motion in a variety of directions over the spatio-temporal domain, incorporates a variation of the Kalman filter and also estimates diver distance and is thus a significant improvement over that particular technique.

In terms of the tracking process itself, the Kalman filter is, of course, the preeminent classical methodology for real-time tracking. It depends, however, on a linear model of system dynamics. Many real systems, including our model of human swimmers, are non-linear and the linearization needed to implement a Kalman filter needs to be carefully managed to avoid poor performance or divergence. The Unscented Kalman Filter [5] we deploy was developed to facilitate non-linear control and tracking, and can be regarded as a compromise between Kalman Filtering and fully non-parametric Condensation [4].

III. Methodology

To track scuba divers in the video sequences, we exploit the periodicity and motion invariance properties that characterize biological motion. To fuse the responses of the multiple frequency detectors, we combine their output with an Unscented Kalman Filter. The core of our approach is to use periodic motion as the signature of biological propulsion and specifically for person-tracking, to detect the kicking gait of a person swimming underwater. While different divers have distinct kicking gaits, the periodicity of swimming (and walking) is universal. Our approach, thus, is to examine the amplitude spectrum of rectangular slices through the video sequence along the temporal axis. We do this by computing a windowed Fourier transform on the image to search for regions that have substantial band-pass energy at a suitable frequency. The flippers of a scuba diver normally oscillate at frequencies between 1 and 2 Hz. Any region of the image that exhibits high energy responses in those frequencies is a potential location of a diver. The essence of our technique is therefore to convert a video sequence into a sampled frequency-domain representation in which we accomplish detection, and then use these responses for tracking. To do this, we need to sample the video sequence in both the spatial and temporal domain and compute local amplitude spectra. This could be accomplished via an explicit filtering mechanism such as steerable filters which might directly yield the required bandpass signals. Instead, we employ windowed Fourier transforms on the selected space-time region which are, in essence, 3-dimensional blocks of data from the video sequence (a 2-dimensional region of the image extended in time). In principle, one could directly employ color information at this stage as well, but due to the need to limit computational cost and the low mutual information content between color channels (especially underwater), we perform the frequency analysis on luminance signals only.

We look at the method of *Fourier Tracking* in Sec. III-A. In Sec. III-B, we describe the multi-directional version of the Fourier tracker and motion detection algorithm in the XYT domain. The application of the Unscented Kalman Filter for position tracking is discussed in Sec. III-C.

A. *Fourier Tracking*

The core concept of the tracking algorithm presented here is to take a time-varying spatial signal (from the robot) and use the well-known discrete-time Fourier transform to convert the signal from the spatial to the frequency domain. Since the target of interest will typically occupy only a region of the image at any time, we naturally need to perform spatial and temporal windowing. The standard equations relating the spatial and frequency domain are as follows.

$$x[n] = \frac{1}{2\pi} \int_{2\pi} X(e^{j\omega}) e^{j\omega} d\omega \tag{1}$$

$$X(e^{j\omega}) = \sum_{n=-\infty}^{+\infty} x[n] e^{-j\omega n} \tag{2}$$

where $x[n]$ is a discrete aperiodic function, and $X(e^{j\omega})$ is periodic with length 2π and frequency ω. Equation 1 is referred to as the *synthesis* equation, and Eq. 2 is the *analysis* equation where $X(e^{j\omega})$ is often called the *spectrum* of $x[n]$ [9]. The coefficients of the converted signal correspond to the amplitude and phase of complex exponentials of harmonically-related frequencies present in the spatial domain.

For our application, we do not consider phase information, but look only at the absolute amplitudes of the coefficients of the above-mentioned frequencies. The phase information might be useful in determining relative positions of the undulating flippers, for example. It might also be used to provide a discriminator between specific individuals. Moreover, by not differentiating between the individual flippers during tracking, we achieve a speed-up in the detection of high-energy responses, at the expense of sacrificing relative phase information.

Spatial sampling is accomplished using a Gaussian windowing function at regular intervals and in multiple directions over the image sequence. The Gaussian is appropriate since it is well known to simultaneously optimize localization in both space and frequency space. It is also a separable filter, making it computationally efficient. Note, as an aside, that some authors have considered tracking using a box filter for sampling, but these produce undesirable ringing in the frequency domain, which can lead to unstable tracking. The Gaussian filter has good frequency domain properties and it can be computed recursively making it exceedingly efficient.

B. *Multi-directional motion detection*

To detect motion in multiple directions, we use a predefined set of vectors, each of which is composed of a set of small

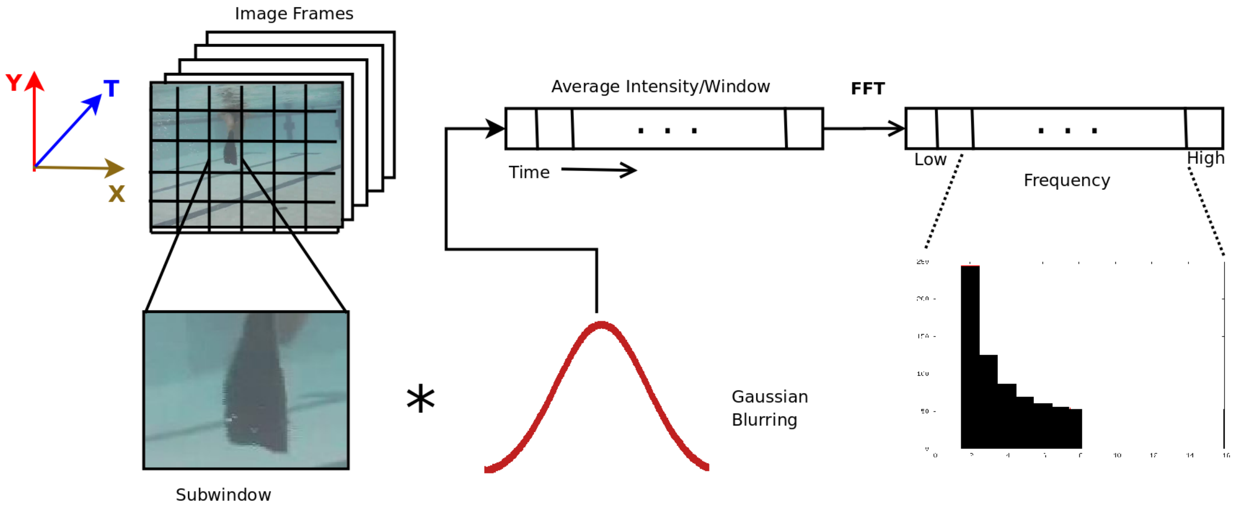


Fig. 3. Outline of the Directional Fourier motion detection and tracking process. The Gaussian-filtered temporal image is split into subwindows, and the average intensity of each subwindow is calculated for every timeframe. For the length of the filter, a one-dimensional intensity vector is formed, which is then passed through an FFT operator. The resulting amplitude plot can be seen, with the symmetric half removed.

rectangular subwindows in the spatio-temporal space. The trajectories of each of these subwindows are governed by a corresponding starting and ending point in the image. In any given time T, this rectangular window resides in a particular position along this trajectory and represents a Gaussian-weighted grayscale intensity value of that particular region in the image. Over the entire trajectory, these windows generate a vector of intensity values along a certain direction in the image, producing a purely temporal signal for amplitude computation. We weight these *velocity vectors* with an exponential filter, such that intensity weights of a more recent location of the subwindow have a higher weight than another at that same location in the past. This weighting helps to maintain the causal nature of the frequency filter applied to this velocity vector. In the current work, we extract 17 such velocity vectors (as seen in Fig. 4) and apply the Fourier transform to them (17 is the optimum number of vectors we can process in quasi-real time in our robot hardware). The space formed by the velocity vectors is a conic in the XYT space, as depicted in Fig. 5. Each such signal provides an amplitude spectrum that can be matched to a profile of a typical human gait. A statistical classifier trained on a large collection of human gait signals would be ideal for matching these amplitude spectra to human gaits. However, these human-associated signals appear to be easy to identify, and as such, an automated classifier is not currently used. Currently, we use two different approaches to select candidate spectra. In the first, we choose the particular direction that exhibits significantly higher energy amplitudes in the low-frequency bands, when compared to higher frequency bands. In the second approach, we precompute by hand an amplitude spectrum from video footage of a swimming diver, and use this amplitude spectrum as a true reference. To find possible matches, we use the Bhattacharyya measure [1] to find similar amplitude spectra, and choose those as possible candidates.

C. Position Tracking using an Unscented Kalman Filter

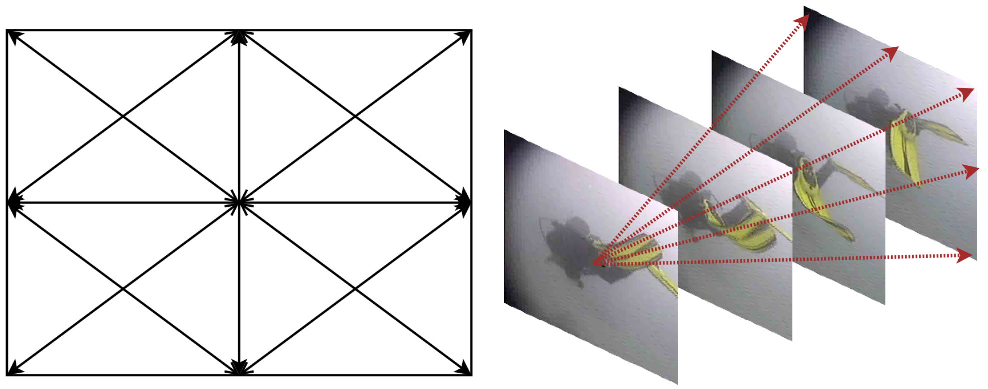
Each of the directional Fourier motion operators outputs an amplitude spectrum of different frequencies present in each associated direction. As described in Sec. III-B, we look at the amplitudes of the low-frequency components of these directional operators, the ones that exhibit high responses are chosen as possible positions of the diver, and thus the position of the diver can be tracked across successive frames.

To further enhance the tracking performance, we run the output of the motion detection operators through an Unscented Kalman Filter (UKF). The UKF is a highly effective filter for state estimation problems, and is suitable for systems with a non-linear process model. The track trajectory and the motion perturbation are highly non-linear, owing to the undulating propulsion resulting from flipper motion and underwater currents and surges. We chose the UKF as an appropriate filtering mechanism because of this inherent non-linearity, and also its computational efficiency.

According to the UKF model, an N-dimensional random variable $\mathbf{x}$ with mean $\hat{\mathbf{x}}$ and covariance P_{xx} is approximated by $2N+1$ points known as the *sigma points*. The sigma points at iteration $k-1$, denoted by $\chi^{i}_{k-1|k-1}$, are derived using the following set of equations:

$$\begin{aligned}
\chi^{0}_{k-1|k-1} &= \mathbf{x}^{a}_{k-1|k-1} \\
\chi^{i}_{k-1|k-1} &= \mathbf{x}^{a}_{k-1|k-1} + \left(\sqrt{(N+\lambda)(P)^{a}_{k-1|k-1}}\right)_{i} \\
& \quad i = 1 \ldots N \\
\chi^{i}_{k-1|k-1} &= \mathbf{x}^{a}_{k-1|k-1} + \left(\sqrt{(N+\lambda)(P)^{a}_{k-1|k-1}}\right)_{i-N} \\
& \quad i = N+1 \ldots 2N
\end{aligned}$$

where $(\sqrt{(N+\lambda)(P)^{a}_{k-1|k-1}})_{i}$ is the i-th column of the matrix square-root of $((N+\lambda)(P)^{a}_{k-1|k-1})$, and λ is a predefined constant.



(a) Motion directions covered by the various directional Fourier operators, depicted in a 2D spatial arrangement. (b) Image slices along the time axis showing 5 out of 17 possible track directions while tracking a scuba diver.

Fig. 4. Directions of motion for Fourier tracking, also depicted in 3D in a diver swimming sequence.

For the diver's location, the estimated position $\mathbf{x}$ is a two-dimensional random variable, and thus the filter requires 5 sigma points. The sigma points are generated around the mean position estimate by projecting the mean along the X and Y axes, and are propagated through a non-linear motion model (*i.e.*, the transition model) f, and the estimated mean (*i.e.*, diver's estimated location), $\hat{\mathbf{x}}$, is calculated as a weighted average of the transformed points:

$$\chi^i_{k|k-1} = f(\chi^i_{k-1|k-1}) \quad i = 0 \dots 2N$$

$$\hat{\mathbf{x}}_{k|k-1} = \sum_{i=0}^{2N} W^i \chi^i_{k|k-1}$$

where W^i are the constant weights for the state (*position*) estimator.

As an initial position estimate of the diver's location for the UKF, we choose the center point of the vector producing the highest low-frequency amplitude response. Ideally, the non-linear motion model for the scuba diver can be learned from training using video data, but for this application we use a hand-crafted model created from manually observing such footage. The non-linear motion model we employ predicts forward motion of the diver with a higher probability than up and down motion, which in turn is favored over sideways motion. For our application, a small number of iterations (approximately between 5 and 7) of the UKF is sufficient for convergence.

IV. Experimental Results

The proposed algorithm has been experimentally validated on video footage recorded of divers swimming in open- and closed-water environments (*i.e*, pool and open ocean, respectively). Both types of video sequences pose significant challenges due to the unconstrained motion of the robot and the diver, and the poor imaging conditions, particularly observed in the open-water footage due to suspended particles, water salinity and varying lighting conditions. The algorithm outputs a direction corresponding to the most dominant biological motion present in the sequence, and a location of the most likely position of the entity generating the motion response. Since the Fourier tracker looks backward in time every N frames to find the new direction and location of the diver, the output of the computed locations are only available after a "bootstrap phase" of N frames. We present the experimental setup below in Sec. IV-A findings and the results in Sec. IV-B.

A. Experimental Setup

As mentioned, we conduct experiments offline on video sequences recorded from the cameras of an underwater robot. The video sequences contain footage of one or more divers swimming in different directions across the image frame, which make them suitable for validating our approach. We run our algorithm on a total of 2530 frames of a diver swimming in a pool, and 2680 frames of a diver swimming in the open-ocean, collected from open ocean field trials of the robot. In total, the frames amounted to over 10 minutes video footage of both environments. The *Xvid*-compressed video frames have dimensions of 768 × 576 pixels, the detector operated at a rate

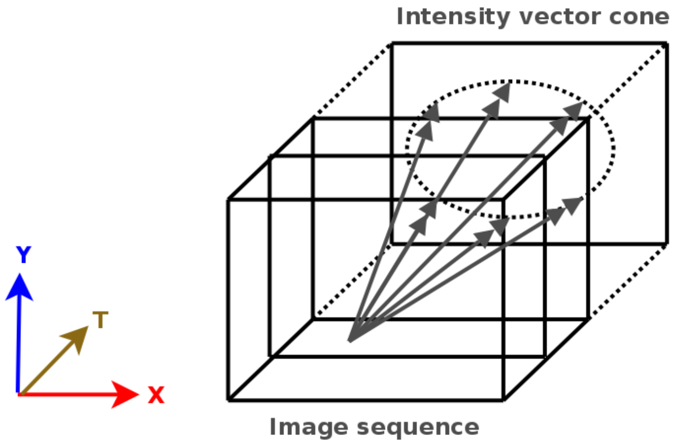
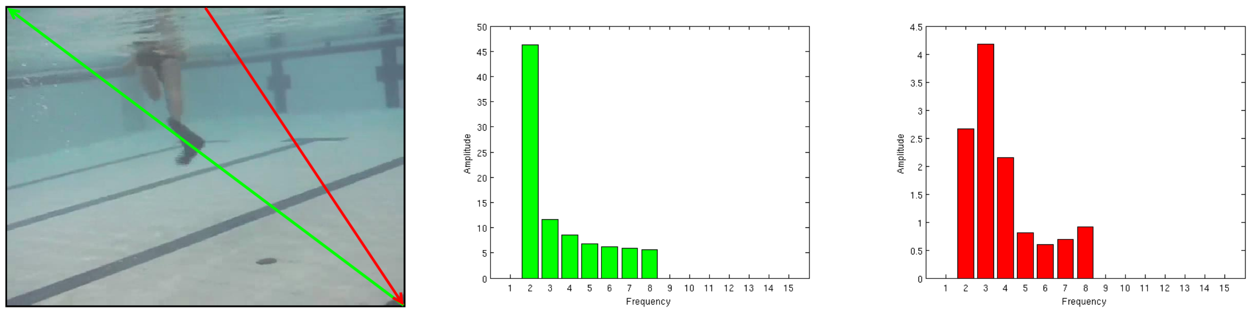


Fig. 5. Conic space covered by the directional Fourier operators.



(a) Snapshot image showing direction of diver motion (in green) and an arbitrary direction without a diver (in red).

(b) Frequency responses along the motion of the diver.

(c) Frequency responses along the direction depicted by the red arrow. Note the low amplitude values.

Fig. 6. Contrasting frequency responses for directions with and without diver motion in a given image sequence.

of approximately 10 frames *per* second, and the time window for the Fourier tracker for this experiment is 15 frames, corresponding to approximately 1.5 seconds of footage. Each rectangular subwindow is 40 × 30 pixels in size (one-fourth in each dimension). The subwindows do not overlap each other on the trajectory along a given direction.

For visually servoing off the responses from the frequency operators, we couple the motion tracker with a simple Proportional-Integral-Derivative (PID) controller. The PID controller accepts image space coordinates as input and provides as output motor commands for the robot such that the error between the desired position of the tracked diver and the current position is minimized. While essential for following divers, the servoing technique is not an integral part of the motion detection algorithm, and thus runs independently of any specific visual tracking algorithm.

B. Results

Figure 6(a) shows a diver swimming along a diagonal direction away from the camera, as depicted by the green arrow. No part of the diver falls on the direction shown by the red arrow, and as such there is no component of motion present in that direction. Figure 6(b) and 6(c) show the Fourier filter output for those two directions, respectively (the green bars correspond to the response along the green direction, and similarly for the red bars). The DC component from the FFT has been manually removed, as has the symmetric half of the FFT over the Nyquist frequency. The plots clearly show a much higher response along the direction of the diver's motion, and almost negligible response in the low frequencies (as a matter of fact in all frequencies) in the direction containing no motion component (as seen from the amplitude values). Note that the lane markers on the bottom of the pool (that appear periodically in the image sequence) do not generate proper frequency responses to be categorized as biological motion in the direction along the red line.

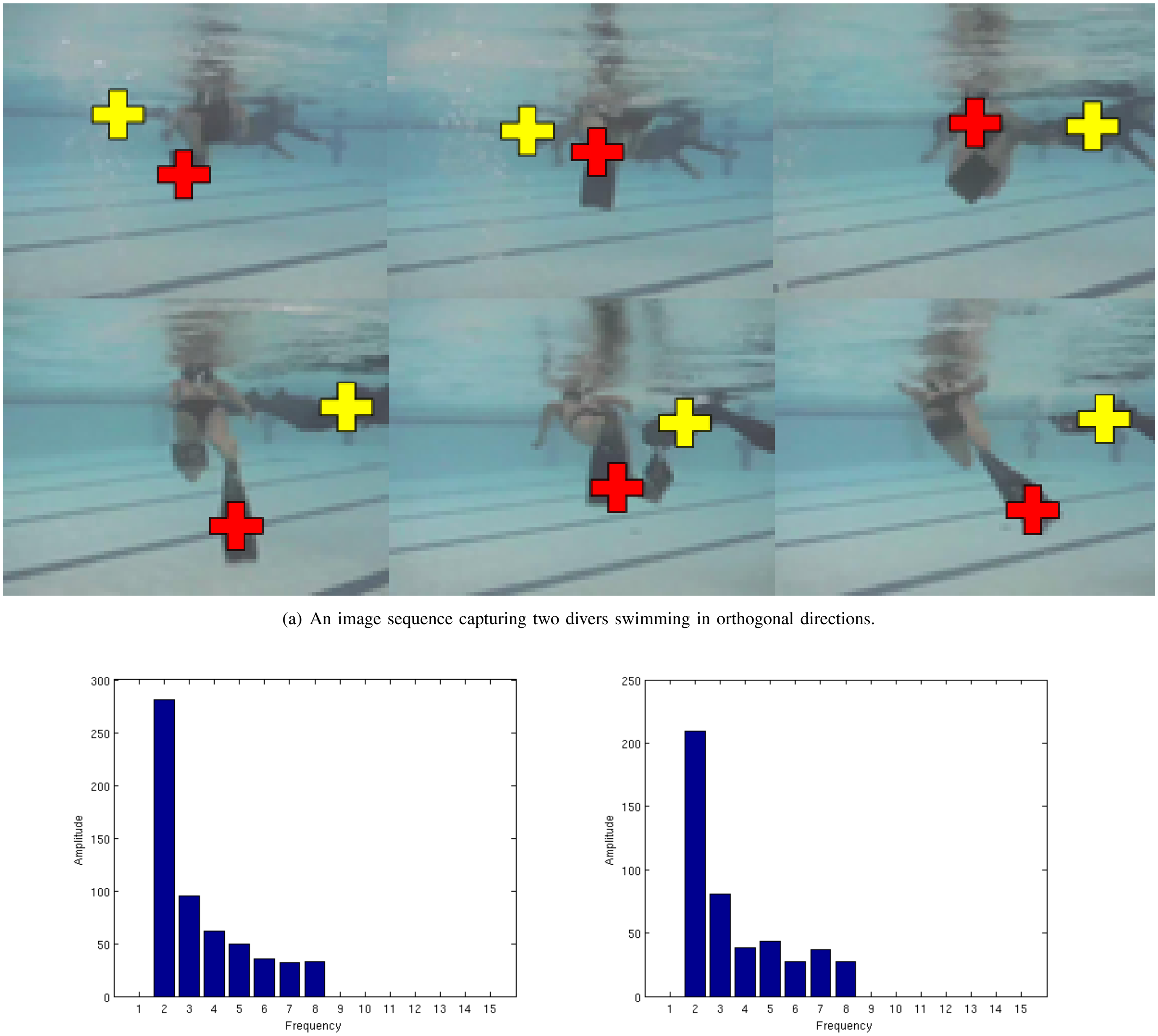
In Fig. 7(a), we demonstrate the performance of the detector in tracking multiple divers swimming in different directions. The sequence shows a diver swimming in a direction away from the robot, while another diver is swimming in front of her across the image frame in an orthogonal direction. The amplitude responses obtained from the Fourier operators along the directions of the motion for the fundamental frequency are listed in ascending order in Tab. I. The first two rows correspond to the direction of motion of the diver going across the image, while the bottom three rows represent the diver swimming away from the robot. As expected, the diver closer and unobstructed to the camera produces the highest responses, but motion of the other diver also produces significant low-frequency responses. The other 12 directions exhibit negligible amplitude responses in the proper frequencies compared to the directions presented in the table. The FFT plots for motion in the bottom-to-top and left-to-right direction are seen in Figs. 7(b) and 7(c), respectively. As before, the FFT plot has the DC component and the symmetric half removed for presentation clarity.

Direction	Lowest-Frequency Amplitude response
Left-to-right	205.03
Right-to-left	209.40
Top-to-bottom	242.26
Up-from-center	251.61
Bottom-to-top	281.22

TABLE I

LOW-FREQUENCY AMPLITUDE RESPONSES FOR MULTIPLE MOTION DIRECTIONS.

An interesting side-effect of the Fourier tracker is the effect of the diver's distance from the robot (and hence the camera) on the low-frequency amplitude. Figure 8 shows two sequences of scuba divers swimming away from the robot, with the second diver closer to the camera. The amplitude responses have similar patterns, exhibiting high energy at the low-frequency regions. The spectrum on top, however, has more energy in the low-frequency bands than the one on the bottom, where the diver is closer to the camera. The close proximity to the camera results in a lower variation of the intensity amplitude, and thus the resulting Fourier amplitude spectra shows lower energy in the low-frequency bands.



(a) An image sequence capturing two divers swimming in orthogonal directions.

(b) Frequency responses for the diver swimming away from the robot (red cross) in Fig. 7(a).

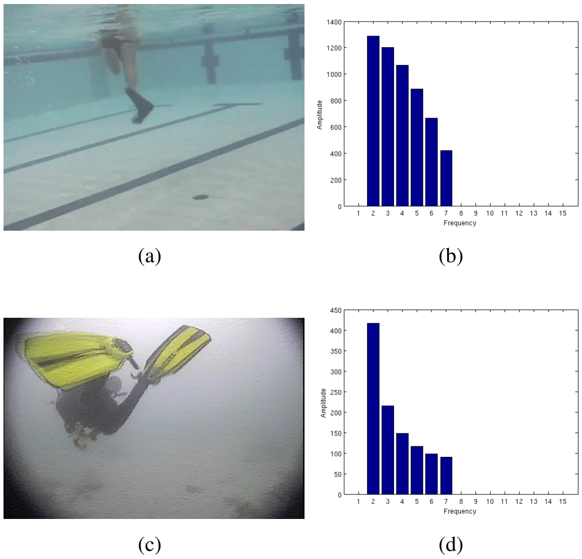
(c) Frequency responses for the diver swimming across the robot (yellow cross) in Fig. 7(a).

Fig. 7. Frequency responses for two different directions of diver motion in a given image sequence.

V. Conclusions and Future Work

In this paper, we present a technique for robust detection and tracking of biological motion underwater, specifically to track human scuba divers. We consider the ability to visually detect biological motion an important feature for any mobile robot, and especially for underwater environments to interact with a human operator. In a larger scale of visual human-robot interaction, such a feature forms an essential component of the communication paradigm, using which an autonomous vehicle can effectively recognize and accompany its human controller. The algorithm presented here is conceptually simple and easy to implement. Significantly, this algorithm is optimized for real-time use on-board an underwater robot. In the very near future, we aim to focus our experiments on our platform, and measure performance statistics of the algorithm implemented on real robotic hardware. While we apply a heuristic for modeling the motion of the scuba diver to feed into the UKF for position tracking, we strongly believe that with the proper training data, a more descriptive and accurate model can be learned. Incorporating such a model promises to increase the performance of the motion tracker.

While color information can be valuable as a tracking cue, we do not look at color in this work. Hues are affected by the optics of the underwater medium, which changes



(a) (b)

(c) (d)

Fig. 8. Effect of diver's distance from camera on the amplitude spectra. Being farther away from the camera produces higher energy responses(Fig. 8(b)) in the low-frequency bands, compared to divers swimming closer(Fig. 8(d)).

object appearances drastically. Lighting variations, suspended particles and artifacts like silt and plankton scatter, absorb or refract light underwater, which directly affects the performance of otherwise-robust tracking algorithms [11]. To reduce these effects and still have useful color information for robustly tracking objects underwater, we have developed a machine learning approach based on the classic Boosting technique. In that work, we train our visual tracker with a bank of *spatio-chromatic* filters [13] that aim to capture the distribution of color on the target object, along with color variations caused by the above-mentioned phenomena. Using these filters and training for a particular diver's flipper, robust color information can be incorporated in the Fourier tracking mechanism, and be directly used as an input to the UKF. While this will increase the computational cost somewhat, and also introduce color dependency, we believe investigating the applicability of this machine learning approach in our Fourier tracker framework is a promising avenue for future research.

References

[1] A. Bhattcharyya. On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin Calcutta Math Society*, 35:99–110, 1943.

[2] D. J. Fleet and A. D. Jepson. Computation of component velocity from local phase information. *Internation Journal of Computer Vision*, 5(1):77–104, August 1990.

[3] W. T. Freeman and E. H. Adelson. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(9):891–906, 1991.

[4] M. Isard and A. Blake. Condensation – conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1):5–28, 1998.

[5] S. Julier and J. Uhlmann. A new extension of the kalman filter to nonlinear systems. In *International Symposium on Aerospace/Defense Sensing, Simulation and Controls*, Orlando, FL, USA, 1997.

[6] M. F. Land. Optics of the eyes of marine animals. In P. J. H. A. K. C. M. W. L. maddock, editor, *Light and life in the sea*, pages 149–166. Cambridge University Press, Cambridge, UK, 1990.

[7] M. S. Nixon, T. N. Tan, and R. Chellappa. *Human Identification Based on Gait*. The Kluwer International Series on Biometrics. Springer-Verlag New York, Inc. Secaucus, NJ, USA, 2005.

[8] S. A. Niyogi and E. H. Adelson. Analyzing and recognizing walking figures in xyt. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 469–474, 1994.

[9] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab. *Signals & systems (2nd ed.)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1996.

[10] R. Rashid. Toward a system for the interpretation of moving light display. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(6):574–581, November 1980.

[11] J. Sattar and G. Dudek. On the performance of color tracking algorithms for underwater robots under varying lighting and visibility. In *Proceedings of the IEEE International Conference on Robotics and Automation ICRA2006*, pages 3550–3555, Orlando, Florida, May 2006.

[12] J. Sattar and G. Dudek. Where is your dive buddy: Tracking humans underwater using spatio-temporal features. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems IROS2007*, pages 3654–3659, San Diego, California, October 2007.

[13] J. Sattar and G. Dudek. Robust servo-control for underwater robots using banks of visual filters. In *Proceedings of the IEEE International Conference on Robotics and Automation, ICRA2009*, pages 3583–3588, Kobe, Japan, May 2009.

[14] J. Sattar, P. Giguere, G. Dudek, and C. Prahacs. A visual servoing system for an aquatic swimming robot. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS2005*, pages 1483–1488, Edmonton, Alberta, Canada, August 2005.

[15] H. Sidenbladh and M. J. Black. Learning the statistics of people in images and video. *Int. J. Comput. Vision*, 54(1-3):181–207, 2003.

[16] H. Sidenbladh, M. J. Black, and D. J. Fleet. Stochastic tracking of 3d human figures using 2d image motion. In *ECCV (2)*, pages 702–718, 2000.