Determining Scale and Sea State From Water Video

Lisa Spencer, *Student Member, IEEE*, Mubarak Shah, *Fellow, IEEE*, and Ratan K. Guha, *Member, IEEE*

***Abstract*—In most image processing and computer vision applications, real-world scale can only be determined when calibration information is available. Dynamic scenes further complicate most situations. However, some types of dynamic scenes provide useful information that can be used to recover real-world scale. In this paper, we focus on ocean scenes and propose a method for finding sizes in real-world units and the sea state from an uncalibrated camera. Fourier transforms in the space and time dimensions yield spatial and temporal frequency spectra. For water waves, the dispersion relation defines a square relationship between the wavelength and period of a wave. Our method applies this dispersion relation to recover the real-world scale of an ocean sequence. The sea state—including the peak wavelength and period, the wind speed that generated the waves, and the wave heights—is also determined from the frequency spectrum of the sequence combined with stochastic oceanography models. The process is demonstrated on synthetic and real sequences, validating the results with known scene geometry. This has wide applications in port monitoring and coastal surveillance.**

***Index Terms*—Fourier transforms, frequency domain analysis, image analysis, sea surface, water.**

I. Introduction

WATER is one of the basic requirements for life, and oceans cover a majority of the planet, yet its infinitely varying surface makes analysis of many scenes challenging. Water is a common element in many visual scenes, ranging from ponds and fountains to rivers, lakes, and oceans. Analysis of these difficult scenes is critical for a number of important tasks. Recent security concerns have increased the demand for surveillance of ports and coastlines. Emergency response teams have long dealt with determining the extent of flooding, as well as search and rescue operations over water.

It is common for vision algorithms to assume that most parts of the scene are static. Background subtraction methods typically model each background pixel with a Gaussian [1] or a mixture of Gaussians [2]. With the specular reflections and caustics that are characteristic of water, a color model that accounts for all the colors representing the water can easily encompass the foreground objects as well. An example of such a scene is shown in Fig. 1. The boat wake and the bright sun reflection on the water have very similar colors, and the boat pixels have the same color as the water with less intense sun reflection.

Optical flow methods use the brightness constancy constraint, which requires that the color of a pixel remain fixed and change only in location. Initially, the lighting was assumed to remain the same, and Lambertian surfaces were assumed. Extensions to this work have generalized the constraint for varying illumination [3] and dealing with illumination nonuniformity and specular reflection [4], but still assume rigid motion, at least in a small area. With the extreme nonrigid motion and dynamic specular reflections intrinsic to water, these tools are not suitable for use in this situation.

Fig. 1. Example of a water scene. The interesting parts (the boats and the wakes) have the same RGB values as the background in other parts of the image.

Even though static scenes are easier to analyze, there is no way to measure the actual size of objects in the scene without some additional knowledge, such as camera calibration or knowing the size of an object in the scene. Movie makers take advantage of this limitation, substituting smaller models for full sized ships or buildings. Most of the time, the replacement is not detectable, but sometimes it just does not look right. One cause for this is dynamic elements in the scene, such as water, smoke, or fire, that cannot be resized. We propose a novel method to recover the scale of objects in a scene containing ocean water without camera calibration. Our method also recovers the sea state, including the peak wavelength and period, the wind speed that created the waves, and the likely wave heights. Determining the size of a ship automatically would go a long way toward classifying its type in a coastal surveillance application. While camera calibration might be possible in this situation, there are other cases where it is not, such as an amateur video of a whale sighting given to a marine researcher.

A few researchers have tackled the problem of dynamic scenes, such as water. Voles *et al.* [5] and, later, Smith *et al.* [6] used frame differencing and statistical analysis of grey level histograms of patches within ocean images to segment ocean and object pixels. Their methods successfully filter out the water variations to detect potential obstacles.

Doretto *et al.* [7] and Fitzgibbon [8] model dynamic textures as an autoregressive moving average (ARMA) process. With a

Manuscript received October 7, 2004; revised June 4, 2005. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ivan W. Selesnick.

The authors are with the College of Engineering and Computer Science, University of Central Florida, Orlando, FL 32816-2450 USA (e-mail: lspencer@cs.ucf.edu; shah@cs.ucf.edu; guha@cs.ucf.edu).

Digital Object Identifier 10.1109/TIP.2006.871102

static camera, both used the learned model parameters to synthesize a longer sequence from the original. For a moving camera, Fitzgibbon registered the frames of the video by searching the space of possible camera motions and using the one that resulted in the most compact model. Monnet *et al.* [9] used the same ARMA model to perform background subtraction for scenes with a dynamic background. The model is used to predict the next frame, and the differences between the prediction and the observation are used for object detection.

In addition to these image-based techniques for simulating temporal textures such as water, the graphics community also uses model-based techniques. Fournier and Reeves [10] were among the first to propose such a model for rendering waves. They modeled multiple wave trains as Gerstner waves, which are based on sinusoids, but capture the elliptical motion of water particles. Mastin *et al.* [11] introduced the use of inverse Fourier transforms for synthesizing water for graphics. Tessendorf [12], [13] extended this technique. A frequency spectrum is created in Fourier space by adding noise to a spectrum based on models used in oceanographic research. To get the height field for a given time instant, each element is multiplied by an appropriate phase shift, then the inverse Fourier transformation is applied. Since the Fourier transform and its inverse are periodic, the result is a patch of ocean that can be seamlessly tiled to cover a large area. Instead of using a standard model, Thon and Ghazanfarpour [14] computed the model from either measured buoy data or by extracting the frequency spectrum from an image.

Fourier transforms have also been used to analyze water outside the graphics field. Sanderson *et al.* [15] identified a frequency template for the current sea state and detected regions that deviated from that template as maritime vessels.

The key that allows us to do more than mundane frequency analysis is the same concept that Mastin *et al.* and Tessendorf used to animate their waves for graphics: *There is a relationship between the spatial and temporal frequencies of water waves.* This relation is called the *dispersion relation* and will be described in detail in Section II. If we find the relationship between the temporal frequencies (in units of seconds $^{-1}$) and spatial frequencies (in units of pixels $^{-1}$) in a water video sequence, we can use the dispersion relation to solve for the pixels/meter scale factor, allowing us to recover real world scale from the water video. Additional knowledge of wave formation models allows us to ascertain the wave heights and wind conditions from the frequency spectrum.

Methods, such as grey-level histogram analysis, that ignore the frequency information cannot take advantage of the dispersion relation. Measurements of objects in these systems are limited to sizes in pixels, which cannot be converted to real world units without a calibrated camera. Our proposed method uses spatial and temporal frequency data to recover the sizes of objects in real world units, like meters.

The rest of this paper is organized as follows. Section II describes the terminology and physics of water waves needed for our analysis. In Section III, the process for applying the dispersion relation to determine real world scale is given. Section IV presents a method for determining the sea state in a video, including the typical wave height, dominant wavelength and period, and the speed of the wind that created the waves.

TABLE I
SYMBOLS USED IN WATER WAVE ANALYSIS

Symbol	Name	Definition	Units
λ	wavelength	Horizontal distance between crests	m
f_s	spatial frequency	$1/\lambda$	m^{-1}
H	wave height	Vertical distance from trough to crest	m
δ	wave steepness	H/λ	none
T	wave period	Time between crests passing a point	s
f_t	temporal frequency	$1/T$	s^{-1}
c	wave speed	λ/T	m/s
g	gravitational constant	9.8 $\mathrm{m/s}^2$	

We present results for both synthetic and real sequences in Section V, discussion in Section VI, and conclusion in Section VII.

II. PROPERTIES OF WATER WAVES

In this section, we will review the oceanography background needed for our analysis, including definition of the basic terminology for describing water waves, the dependencies between these quantities for individual waves, statistical tools used for describing the random sea surface, and models used for describing the wave spectra.

A. Terminology

We will begin by defining some basic terminology, as it applies to water waves. The *wavelength*, λ, is the horizontal distance from crest to crest. The *wave height*, H, is the vertical distance from trough to crest. This is double the traditional *amplitude*, A, which is the vertical deviation from the surface at rest. Wave height is more commonly used than amplitude since sea level is difficult to determine for a dynamic ocean.

The *spatial frequency*, f_s, is the reciprocal of the wavelength. The *steepness*, δ, is the ratio of the wave height to the wavelength [16]. Similarly, in the time domain, *wave period*, T, is the time between two crests passing a stationary point and the *temporal frequency*, f_t, is the reciprocal of the wave period. The *wave speed*, c, the speed at which a wave crest advances, is then λ/T (not to be confused with the speed at which individual water particles move). These quantities and typical units are summarized in Table I.

B. Individual Wave Relations

As mentioned in Section I, these wave properties are not independent. Water waves are much longer than they are tall. Stokes sets the upper limit of the wave steepness, δ at $1/7 \approx 0.143$ [16]. This matches reasonably well with observations of $0.008 < \delta < 0.1$ [16]. If we can determine the wavelength, this makes the upper bound on the wave height

$$H < 0.1\lambda. \tag{1}$$

The general formula for the speed of waves is [17]

$$c = \sqrt{\frac{g\lambda}{2\pi} \tanh \frac{2\pi d}{\lambda}} \tag{2}$$

where d is the mean depth of the water and g is the acceleration of gravity (9.8 m/s^2 when λ and d are in meters and c is in m/s). This change in speed with water depth is what causes waves to break on the beach. For deep water, $(d/\lambda) > 0.5$ (i.e., the depth is more than half the wavelength), the $\tanh$ term approaches one, resulting in (3)

$$c = \sqrt{\frac{g\lambda}{2\pi}}. \tag{3}$$

Substituting (3) into the definition of wave speed, $c = (\lambda/T)$ yields

$$T = \frac{\lambda}{c} = \frac{\lambda}{\sqrt{\frac{g\lambda}{2\pi}}} \tag{4}$$

$$T^2 = \frac{\lambda^2}{\left(\frac{g\lambda}{2\pi}\right)} = \frac{2\pi\lambda}{g} \tag{5}$$

$$\lambda = \frac{g}{2\pi}T^2 = \frac{g}{2\pi f_t^2}. \tag{6}$$

Thus, (6) defines the dispersion relation for deep water, quantifying the link between the period and length of a water wave. The name "dispersion relation" comes from the fact that a set of waves that overlap will disperse over time, sorting themselves from longest to shortest due to the differences in speed. We will assume deep water for the remainder of this discussion.

C. Ocean as a Random Process

We have defined properties of a single wave, but in the ocean, it is difficult to isolate a single wave. It is more useful to use a stochastic model to describe the properties of the waves. For analysis, the ocean is assumed to be an ergodic stationary random process [16], [18]. This means that measuring one realization over a sufficiently long time is equivalent to measuring many realizations.

This stationary process model is possible because the characteristics of all but the smallest waves change slowly over time. When a breeze stirs over calm water, the first waves to appear are tiny capillary waves. The restoring force for these waves is surface tension, and if the wind ceases, these waves vanish, but this is the exception, not the rule. If the wind continues to add energy, larger waves form, whose restoring force is gravity. Once formed, a gravity wave can travel a long distance with minimal energy loss. Anyone who has watched the ripples from a pebble tossed into a pond has seen that the waves continue long after the pebble has sunk to the bottom. If energy continues to be added to the system in the form of wind, the waves will continue to grow until equilibrium is reached, where the energy added is equal to the energy dissipated by breaking waves. This condition is called a fully developed sea. The maximum heights and lengths of the waves depend on the wind speed [17]. If the wind suddenly ceases, the waves will remain. The use of a stationary random process to describe the ocean surface is based on the fact that the frequency spectrum characteristics of a given patch of ocean will change very slowly over time, even in the presence of variable winds. The fact that the surf gets extremely rough when a hurricane is still hundreds of miles offshore, even when the local winds are calm, is further evidence that the waves are not a function of the local wind.

TABLE II
BEAUFORT SCALE FOR SEA STATES 0–10 [17]

Beaufort Number	Wind (knots)	Nautical term	Form and height of waves in feet	Effects observed at sea
0	under 1	Calm		Sea like mirror.
1	1–3	Light air	Calm, glassy 0	Ripples with appearance of scales; No foam crests.
2	4–6	Light breeze	Rippled, 0–1	Small wavelets; crests of glassy appearance, not breaking.
3	7–10	Gentle breeze	Smooth, 1–2	Large wavelets; crests begin to break; scattered whitecaps.
4	11–16	Moderate breeze	Slight, 2–4	Small waves, becoming longer; numerous whitecaps.
5	17–21	Fresh breeze	Moderate, 4–8	Moderate waves, taking longer form; many whitecaps; some spray.
6	22–27	Strong breeze	Rough, 8–13	Larger waves forming; whitecaps everywhere; more spray.
7	28–33	Moderate gale		Sea heaps up; white foam from breaking waves begin to be blown in streaks.
8	34–40	Fresh gale	Very rough, 13-20	Moderately high waves of greater length; edges of crests begin to break into spindrift; foam is blown in well-marked streaks.
9	41–47	Strong gale		High waves; sea begins to roll; dense streaks of foam; spray may reduce visibility.
10	48-55	Whole gale	High, 20-30	Very high waves with overhanging crests; sea takes white appearance as foam is blown in very dense streaks; rolling is heavy and visibility reduced.

D. Ocean Models

The efforts to correlate wind speeds and the heights of fully developed seas date back at least to 1805, when Sir Francis Beaufort introduced a chart for this purpose [17]. Its original purpose was to help sailing captains determine whether to add or take in sail based on observation of the sea. After sails were replaced with engines, the scale continued to be used to forecast the sea conditions for a given wind speed. Although modern meteorology has replaced the Beaufort scale for forecasting, the scale is still useful in the absence of measurement instruments and provides an intuitive understanding of the conditions. A modern version of the Beaufort scale is given in Table II. We can use the table to relate wind speed and wave height, as well as to verify our results by matching visual observations, like the frequency of whitecaps, with the last column of the chart.

The model used by Tessendorf [12], [13] to synthesize wind-driven waves in a fully developed sea for graphics is the Phillips spectrum

$$P_h(\mathbf{k}) = A\frac{\exp\left(\frac{-1}{\left(k\frac{V^2}{g}\right)^2}\right)}{k^4} \tag{7}$$

where A is a numeric constant, $\mathbf{k}$ is the wavevector, k is the magnitude of $\mathbf{k}$, and V is the wind speed. The wavevector, $\mathbf{k}$, is a horizontal vector that points in the direction of travel of the wave, and has magnitude $k = 2\pi/\lambda$. The maximum value for $P_h(\mathbf{k})$ (the waves with the most energy) occurs when $k = (g/V^2\sqrt{2})$, which is equivalent to

$$\lambda = \frac{2\pi V^2\sqrt{2}}{g} \tag{8}$$

$$T = \frac{2\pi V\sqrt[4]{2}}{g}. \tag{9}$$

For example, a wind speed of 5.1 m/s (equivalent to 10 knots, the upper end of the winds for Beaufort force 3) produces a sea with the most energy at a wavelength of 23.6 m and a period of 3.9 s. This agrees with Smith [17], who reports that a wind speed of 10 knots results in a fully developed sea where most of the energy is concentrated in waves with a period of 4 s. Thus, the Phillips spectrum allows us to predict the distribution of the waves as well as the most common wavelength and period with only the wind speed as a parameter.

III. Determining Scale

In our previous work [19], we did spatial and temporal analysis separately, and then correlated the peaks in the two to recover the scale factor. However, when the camera field of view was too narrow to capture the peak spatial frequencies, it was not possible to match the peaks. To overcome this difficulty, our current work uses the phase rate of change for all the frequencies, instead of relying on a peak magnitude.

Fourier analysis was used by Tessendorf [13] to validate the model he used for simulating water. He recounted an experiment in which a three-dimensional (3-D) Fourier transform was used to create a power spectral density (PSD) plot from a video sequence of ocean water. His results subjectively showed that the sample exhibits the square law expressed in the dispersion relation, but he did not extract any further measurements.

We propose to automatically find a scale factor β in units of pixels per meter, from an ocean video sequence. This scale factor can then be used to find the real world size of objects in the scene. The algorithm is summarized in Table III.

We can derive the expression for β as follows. Since the desired scale factor is in pixels/meter, we can start by finding the ratio of the wavelengths measured in pixels (measured from the video sequence) and in meters (in the real world)

$$\beta = \frac{\lambda_{\text{pixels}}}{\lambda_{\text{meters}}}. \tag{10}$$

TABLE III
Summary of Algorithm for Determining Scale From From Ocean Video

1) Find the direction containing the most energy in the FFT of the first frame.
2) Create a blank image with height equal to the length of the strip in this dominant direction, and width equal to the number of frames in the sequence.
3) For each frame, find the 2D FFT. Copy a strip from the dominant direction of the FFT to a column of the accumulated FFT image.
4) Find the magnitude of the FFT of each row of the completed accumulated FFT image.
5) Fit the best parabola $y = \alpha x^2$.
6) Solve for the scale factor, β, using (17).

Since the dispersion relation provides an association between wavelengths in meters and wave periods, we can also write

$$\beta = \frac{\lambda_{\text{pixels}}}{\frac{g}{2\pi}T^2} = \frac{2\pi\lambda_{\text{pixels}}}{gT^2}. \tag{11}$$

The unknowns at this point are the wavelengths and periods, both in units we can measure from the video sequence. Note that β does not just describe the relationship for one wavelength and its corresponding period, but for all wavelengths. A two-dimensional (2-D) Fourier transform is well suited to separate the different wavelengths in each frame. To reduce the memory requirements, we only collect a slice through the origin of the FFT from each frame. The idea of using a slice of each frame is not uncommon in video processing, and is used, for example, in film classification [20]. This slice contains all of the spatial frequencies that the FFT can measure, but in only one direction. The direction of the slice is chosen by averaging the energy in each direction in the first frame and choosing the angle with the highest energy. If the energy is evenly distributed, the choice of this direction does not make a big difference, but, if the waves are highly directional, samples in a direction orthogonal to the direction the waves are traveling may not contain enough energy to be useful. Since, as discussed in Section II-C, the frequency spectrum does not change rapidly, the dominant direction in the first image should suffice for all the frames in the sequence. The number of pixels in this slice will be denoted as L.

We also need to operate on the time dimension of the data to find the temporal periods for each spatial wavelength, i.e., how long it takes for the phase of each spatial frequency to complete a cycle. If we create a new image where column x contains the FFT slice from frame x, the FFT of row y will show the temporal frequencies exhibited by waves with spatial wavelength

$$\lambda_{\text{pixels}} = \frac{L}{y}. \tag{12}$$

Within row y, the magnitude of column x will show the energy of waves with period

$$T = \frac{\frac{N}{\nu}}{x} = \frac{N}{\nu x} \tag{13}$$

where N is the number of frames and ν is the frame rate. Substituting (12) and (13) in (11), we get

$$\beta = \frac{2\pi\left(\frac{L}{y}\right)}{g\left(\frac{\frac{N}{\nu}}{x}\right)^2} = \frac{2\pi L\nu^2x^2}{gyN^2}. \tag{14}$$

TABLE IV
SUMMARY OF ALGORITHM FOR DETERMINING SEA STATE FROM FROM OCEAN VIDEO

1) Find the direction containing the most energy in the FFT of the first frame.
2) Create a blank image with height equal to the length of the strip in the dominant direction, and width equal to the number of frames in the sequence.
3) For each frame, copy a strip in the dominant direction (found in the first step) from the original image to a column in the accumulated raw image.
4) Find the magnitude of the FFT of each row of the completed accumulated raw image.
5) Average each column of FFT magnitudes to get the energy at each temporal frequency.
6) Find the wind speed that minimizes the error between (7) and the result from the previous step.
7) Use the wind speed to estimate the wave heights from the Beaufort scale in Table II, the dominant wavelength from (8), the dominant wave period from (9), and maximum steepness from the dominant wavelength using (1).

By rearranging terms, we see that the energy image should show a parabola that intersects the origin

$$y = \frac{2\pi L\nu^2}{gN^2\beta}x^2 = \alpha x^2. \tag{15}$$

By fitting a parabola to the data, we can recover α. More specifically

$$\alpha = \arg\max_{\alpha} \sum_{x} \mathrm{Energy}(x, \alpha x^2) \tag{16}$$

where $\mathrm{Energy}(x, y)$ is the pixel value at row x, column y of the energy image. We can then find β, since all the other parameters are known, as

$$\beta = \frac{2\pi L\nu^2}{gN^2\alpha}. \tag{17}$$

IV. DETERMINING SEA STATE

The analysis in the previous section measures the differences between the propagation speeds for different wavelengths, but does not attempt to determine which wavelengths are present, much less which wavelengths are the most prominent. By finding the energy at the different wavelengths, we can determine the sea state, the size of the waves, and the wind velocity that created them. The algorithm is summarized in Table IV.

Since we have the scale, we could find the power in each spatial frequency in a single frame, then average the results of multiple frames to find the dominant spatial frequency. With the scale factor calculated according to Section III, this can be converted to a wavelength in meters. However, wavelengths less than two pixels or larger than the image size will not be correctly sampled. Even a modest Beaufort force 3 has a dominant wavelength of 16 m, so the water would have to be sufficiently distant for an image to cover multiple cycles of this wavelength. If we use the temporal axis instead, that same 16-m wavelength has a period of 3.2 s, so a 30-s video sequence would be able to capture more than nine cycles of the dominant wave period, regardless of the field of view. For this reason, our wave spectrum analysis uses the temporal frequencies.

In theory, sampling a single pixel over time will provide the temporal frequency spectrum. Averaging the spectrum for every pixel in the frame will increase the accuracy of the estimate. As a compromise, to keep memory requirements reasonable, we sample a slice of the original image from each frame, in the same direction as the slice of the FFT used for calculating the scale. These slices are collected in an image with the x coordinate (horizontal) equal to the frame number and the y coordinate (vertical) based on the pixel location in the original video frame.

To find the temporal frequencies, we compute the FFT power for each row, resulting in an image where the y coordinate (row) still represents the pixel location, and the x coordinate (column) shows the energy at a given temporal frequency. Averaging each column results in a row vector showing the strength at each temporal frequency. The wave period corresponding to each element in the vector is given by (13), and the wavelength can be computed from the period using (6).

We want to extract information about the sea state from this energy spectrum. One way would be to find the peak period, then use (9) to find the corresponding wind speed. However, noise could easily distort the peak. Instead, we use the energy at all the measured frequencies to find the wind speed that minimizes the error between the Phillips spectrum model from (7) and our data (this may not be the same as the current wind speed, since the waves change much more slowly than the wind). More formally, we want to find the wind speed V, such that

$$V = \arg\min_{V} \sum_{k} (S(k) - P(k))^2 \tag{18}$$

where $S(k)$ is the PSD measured from the video sequence as a function of the wavenumber magnitude k, and $P(k)$ is the Phillips spectrum from (7). At each value of V, linear least squares was used to find the Phillips amplitude A that best fit the data, which was then used to compute the error.

Once the wind speed is known, the Beaufort chart (Table II) provides the probable wave heights and overall marine conditions. Equations (8) and (9) can be used to find the dominant wavelength and period, and (1) provides an upper limit on the dominant wave height.

V. RESULTS

We tested this process on both synthetic and real sequences. The synthetic sequences used Tessendorf's inverse Fourier transform technique [12], [13] to generate water scenes. Since the synthetic data was created using the deep water dispersion relation and the Phillips frequency spectrum with known parameters, experiments on the synthetic data verify that our method can successfully extract these parameters. The real sequences test whether the scale and sea state can also be recovered under imperfect conditions in a real situation.

A. Synthetic Data

The steps in the process are illustrated in Fig. 2. All the synthetic sequences used a 300×300 pixel viewport with a 30° field of view. The first synthetic sequence had a length of 60 s at 30 f/s (1800 frames). The viewpoint was positioned 50 m above the water looking down. In this configuration, the 150 pixels in

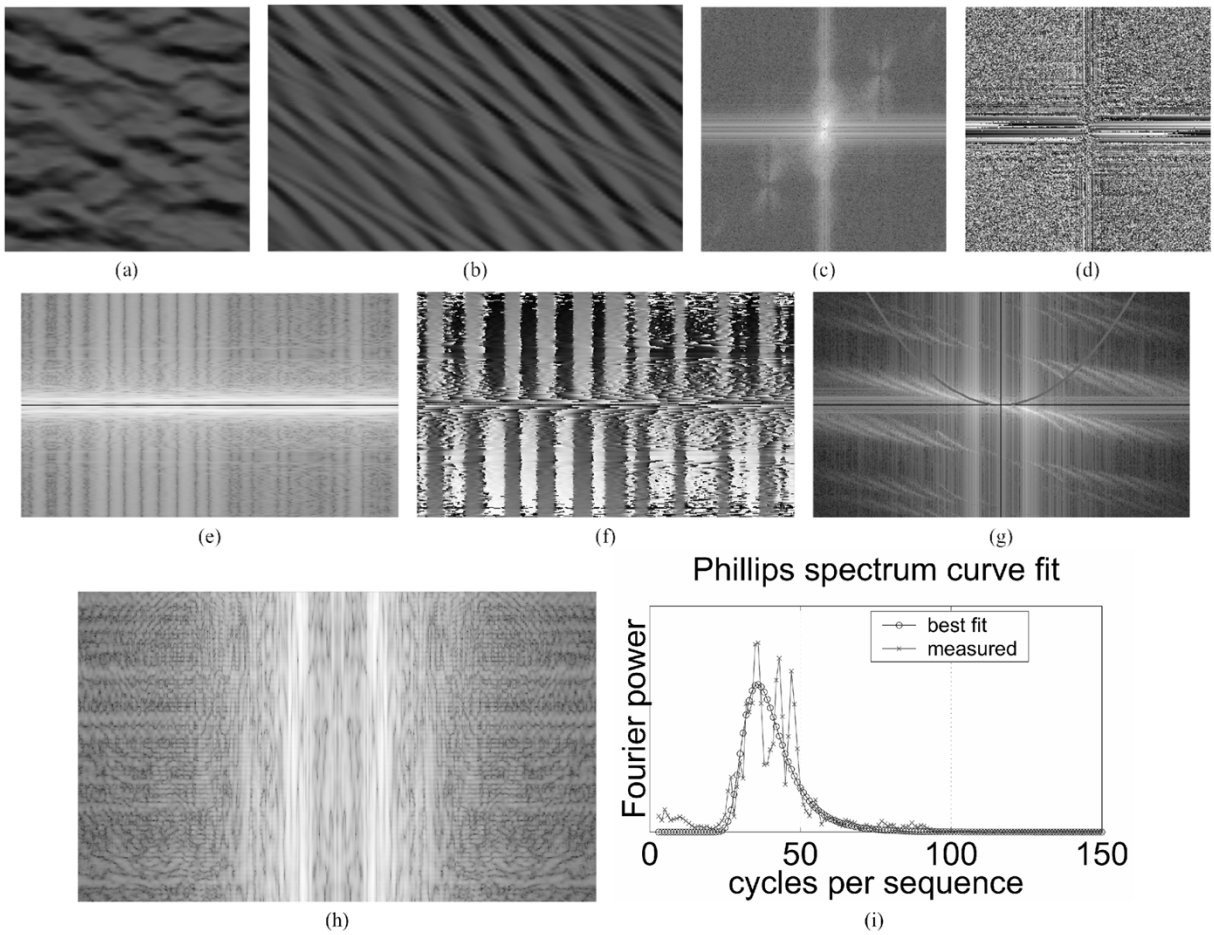

Fig. 2. Steps in the process for determining scale for a synthetic water sequence. All FFT magnitude plots actually show the log of the magnitude to make details easier to see. All FFT phase plots show 0° as black and 360° as white. (a) First frame. (b) One column from each frame. (c) First frame FFT magnitude. (d) First frame FFT phase. (e) Temporal magnitude image containing one column from the FFT of each frame. (f) Temporal phase image. (g) Magnitude of the FFT of the temporal image. (h) Magnitude of the FFT of the rows of the raw collection image. (i) Energy at each temporal frequency.

half the field of view cover $50 \tan 15^\circ = 13.4$ m, so the ground truth for β is 11.20 pixels/m. A wind speed of 2.6 m/s (about 5 knots) was used to generate the waves, which is typical of Beaufort force 2. According to (9), this should correspond to a peak wave period of 2.0 s.

Fig. 2(a) shows the first frame of the sequence. Fig. 2(c) and (d) shows the magnitude and phase respectively of Fig. 2(a). The zero frequency values are at the center of each. The magnitude plot shows directionality of the waves, with the majority of the energy concentrated at the lower frequencies and most of the energy is in waves traveling vertically in the video. The phase plot looks like random noise because random phases were used to generate the frequency spectrum in the simulation.

For each frame, a slice from the raw image and a slice from the FFT of the frame were collected, both in the dominant direction calculated from the first frame (vertical). Each accumulated image is 300 pixels high (the slice length L) by 1800 pixels wide (the number of frames N). The resulting raw image is real and the first 500 frames are shown in Fig. 2(b). The FFT image is complex. The magnitude and phase are depicted in Fig. 2(e) and (f), respectively. Time increases from left to right in both, and the lowest spatial frequencies are in the middle rows. The horizontal stripes in the magnitude image show that the average energy of the waves having a given wavelength did not change much over time. It also echoes the observation from the single frame FFT magnitude, that most of the energy is at lower spatial frequencies. In the phase image, 0° is shown as black and 360° is shown as white. While the phases at the various spatial frequencies in the single image were random, Fig. 2(f) shows that they change smoothly. More importantly, the lower frequencies change more slowly than the higher frequencies. This was predicted by the dispersion relation.

Finding the scale factor requires us to quantify the rate at which the phases change for each spatial frequency, so we found the FFT of each row. The magnitude of the result is shown in Fig. 2(g). The brightest pixels form half a parabola through the origin in the center of the image. The other half is missing here because waves traveling in the direction opposite the wind were suppressed in the simulation. The parabola that best fit this image was found to be $y = 0.0103x^2$. This parabola is drawn superimposed on Fig. 2(g). The integer multiples, or harmonics, of the base frequencies can also be seen in the plot. Using (17), we find $\beta = 11.24$ pixels/m, which is very close to the ground truth value of 11.20 pixels/m, with an error of 0.36%.

To quantify the sea state, we need to find the wind speed that best fits our data. The magnitude of the FFT in the time dimension, along the rows of Fig. 2(b), is shown in Fig. 2(h). The ver-

TABLE V
RESULTS FROM SYNTHETIC SEQUENCES

Beaufort Number	Ground Truth β (pix/m)	Ground Truth Wind Speed (m/s)	Calc. β (pix/m)	Calc. Wind Speed (m/s)	β Error (%)	Wind Speed Error (%)
2	11.20	2.6	11.24	2.2	0.35	15.4
3	7.00	4.2	7.09	4.2	1.29	0.0
4	4.67	7.0	4.42	7.2	5.35	2.9
5	1.87	9.7	1.83	9.6	2.14	1.0
6	1.40	12.5	1.29	12.5	7.86	0.0

tical stripes indicate that the energy spectra at each measured pixel were similar. Averaging each column produces the energy at each temporal frequency, and is shown in Fig. 2(i). The best wind speed was calculated to be 2.2 m/s (4.3 knots), 15% away from the 2.6 m/s expected. This was found by using (18) to test wind speeds from 0.5 to 25.0 m/s at increments of 0.1 m/s. The Phillips spectrum generated by a wind speed of 2.2 m/s is also shown in Fig. 2(i).

This wind speed is consistent with sea state 2 on the Beaufort scale. Table II leads us to expect wave heights less than a foot (0.3 m). Equations (8) and (9) show that the peak wavelength should be 4.4 m, and the peak wave period should be 1.7 s. The wave steepness from (1) limits the maximum wave height to 0.44 m.

The analysis was done on simulated oceans ranging from sea state two through sea state six. The results are summarized in Table V. In all of these trials, analysis of the synthetic data yields a scale factor within 8% and wind speed within 15% of the ground truth values. This validates our premise that this algorithm can recover these quantities from ocean video.

To test the robustness to noise, we added random zero mean Gaussian noise to each 8-bit channel of the sea state 2 synthetic video, with the standard deviation of the noise varying from 10 to 100. The recovered scale and wind speed varied by only 2%, indicating that the method is not sensitive to noise.

B. Real Data

In the real world, it is difficult to capture ocean video with a stationary camera from directly overhead. For the first sequence, dubbed "surfer1," perspective correction was applied to video acquired using a tripod on a pier to generate an overhead view. The visible horizon and knowledge of the camera focal length facilitates calculation of the camera slant angle [21], which, in this case, was 80° (10° below the horizon). Fig. 3(a) shows the first frame of the first real sequence, and Fig. 3(b) shows the overhead view generated. The resolution of the corrected view is 360×360 pixels, and the video is 45 s (1363 frames) long. There is a surfer in the center of the raw image, which is at the top edge of Fig. 3(b), and was used to verify our scale calculation. The length of the surfboard in the video is 46 ± 2 pixels, and the board was estimated to be 8 or 9 ft (2.4 to 2.7 m) long. This gives a ground truth scale factor of between 16.3 and 20.0 pixels/m for this video.

Fig. 3(d) shows the FFT of Fig. 3(b). The dominant direction in the first frame was found to be 18° from horizontal, which agrees with the observation that the waves primarily traveled from left to right in the images, toward the shore. Fig. 3(e) and (f) depict changes in the magnitude and phase respectively of the slices over the first 500 frames. The magnitude image shows the maximum magnitude at one cycle per image in each frame, with the magnitude decreasing as the frequency increases, indicating that the field of view (or the portion of it used in the perspective correction) was probably not wide enough to capture the actual peak spatial frequency. This is not a problem, since we use the temporal frequencies to find the peak frequency. The phase image clearly shows the phase offset between the top half (positive frequencies) and bottom half (negative frequencies). It also shows that the phase changes more slowly in the rows closer to the center (i.e., lower frequencies). The magnitude of the FFT of the rows the temporal image and the best fit parabola are shown in Fig. 3(g). The parabola is not as obvious, but is visible near the center. The scale factor found by fitting a parabola was $\beta = 18.1$ pixels/m. This agrees with our estimate using the surfboard length, which was 16.3 to 20.0 pixels/m.

Fig. 3(c) shows a column from each of the first 500 frames of the raw sequence, with time increasing from left to right. The next step of our algorithm for determining the sea state requires finding the FFT of each row of this image. The magnitude of the result in shown in Fig. 3(h), with the lowest frequencies in the center columns. The column averages are displayed in Fig. 3(i). The wind speed that fits this data the best is 2.3 m/s (4.5 knots), and the corresponding Phillips spectrum is also shown in Fig. 3(i). The dominant wavelength is, therefore, 4.8 m, and the wave period is 1.8 s.

We were able to recover this peak period from the temporal dimension, even though a single frame did not cover enough area to capture this peak wavelength. This is because the time dimension spanned 45 s, or 25 cycles of the peak period. The spatial dimension only covered 20 m (360 pixels/18.1 pixels/m), which is four cycles of the peak wavelength.

A wind speed of 4.5 knots qualifies as Beaufort force 2. This is consistent with the observations of the day, where breaking waves were rare (the surfer did not find any suitable waves before leaving). Wave heights can be expected to be less than 0.3 m, according to the Beaufort scale, and wave steepness limits the wave heights to 0.36 m.

To validate our claim that the direction used for the slice is not critical, we processed the sequence using a variety of different directions. The scale factor calculated for each is shown in Table VI. For directions between $-60°$ and $+60°$, all scale factors are within 10% of the expected range. It is only the directions nearly orthogonal to the dominant direction that produce erroneous results.

An additional sequence, "surfer2," filmed immediately after "surfer1," with doubling the zoom factor as the only change, produced a scale factor of 9.2 pixels/m and the same wind speed of 2.3 m/s. The scale factor is half of that for the "surfer1" sequence, which is the expected result when the field of view is cut in half. Since the water did not change, the same wind speed was found.

The next real sequence, named "east1," was recorded on the same day, and from the same pier as before, facing out to sea toward the east. The camera was on a tripod braced against the pier railing and tilted down 60° to 70° from horizontal. This was as close to vertical as we could get while keeping the pier itself

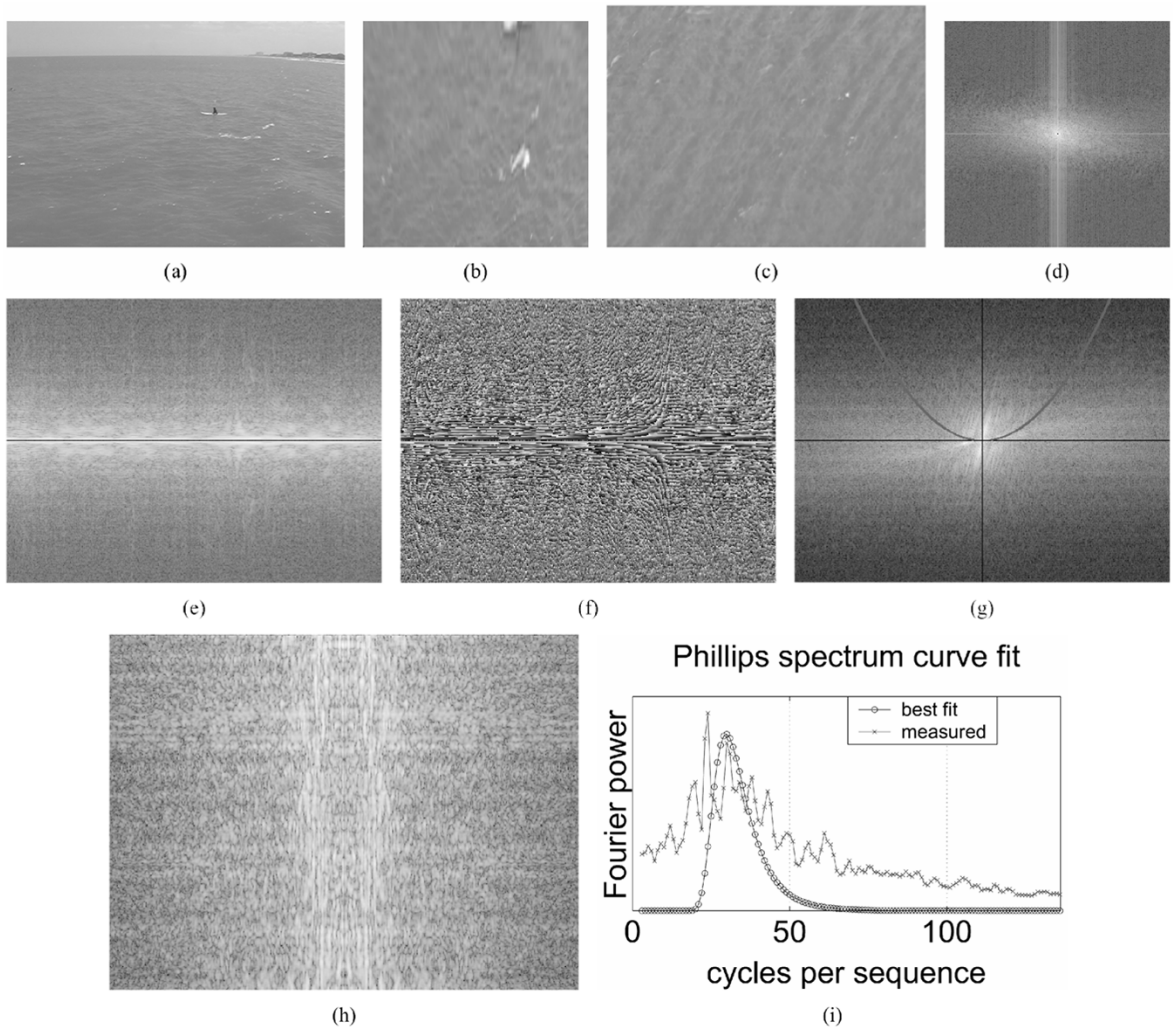

Fig. 3. Determining scale for the "surfer1" ocean sequence. The surfboard visible in the first frame was used to validate the results. (a) First frame. (b) Overhead view. (c) One column from each of the first 500 frames. (d) First frame FFT magnitude. (e) Temporal magnitude image containing one column from the FFT of the first 500 frames. (f) Temporal phase image. (g) Magnitude of the FFT of the temporal image. (h) Magnitude of the FFT of the rows of the raw collection image. (i) Energy at each temporal frequency.

TABLE VI
SCALE FOR THE "SURFER1" SEQUENCE FOR VARIOUS SLICE DIRECTIONS

Slice direction	β (pix./m.)
$-80°$	3.42
$-60°$	19.7
$-40°$	21.9
$-20°$	16.5
$0°$	15.8
$20°$	17.0
$40°$	18.4
$60°$	17.9
$80°$	2.63

out of the camera field of view. No perspective correction was used. Due to the optical characteristics of the water, the waves are less well defined in this video, and there is more sun glint. The physical patch of water observed is much smaller, but the analysis was still able to find the scale factor. We used the first 60 s of the video, so the sequence length was 1800 frames.

Fig. 4(a) shows the first frame of the video. The magnitude and phase of the FFT of this image are in Fig. 4(c) and (d). The FFT magnitude looks very similar to the previous sequence, which was expected since the statistical properties of the water change very slowly. Fig. 4(e) and (f) shows the magnitude and phase of the temporal image. The longer periods at lower frequencies in Fig. 4(f) are only visible in the lowest frequency rows. Fig. 4(g) shows the magnitude of the FFT of the rows of the temporal image (a slice from the 3-D FFT). We have drawn the best fitting parabola in the top half. The resulting scale factor is 74.6 pixels/m, and this is consistent with the experimental setup.

To determine the sea state, we collected a slice from each raw frame, shown in Fig. 4(b). The temporal frequencies in each row are shown in Fig. 4(h), and the average energy at each temporal frequency over all the rows is plotted in Fig. 4(i). The calculated wind speed is 2.5 m/s (4.9 knots), which differs from the value found in the previous video by less than 10%, and still qualifies as Beaufort force 2. The dominant wavelength is 5.7 m and the dominant wave period is 1.9 s. The Beaufort scale leads us to expect wave heights around 0.3 m, and the wave steepness limits the most common waves to 0.57 m.

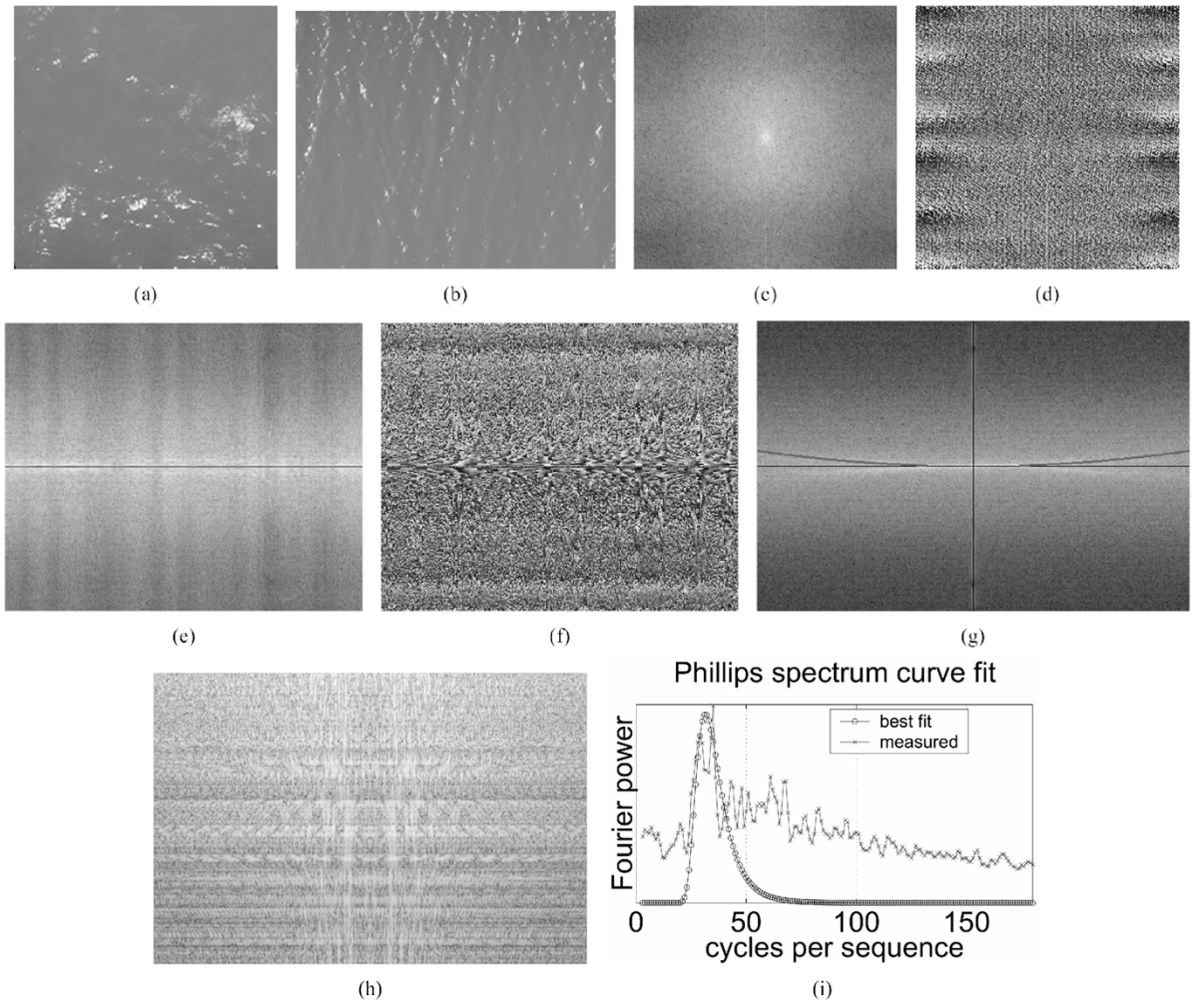

Fig. 4. Determining scale for the "east1" ocean sequence. (a) First frame. (b) One column from each frame. (c) First frame FFT magnitude. (d) First frame FFT phase. (e) Temporal magnitude image containing one column from the FFT of each frame. (f) Temporal phase image. (g) Magnitude of the FFT of the temporal image with the best parabola superimposed in one half. (h) Magnitude of the FFT of the rows of the raw collection image. (i) Energy at each temporal frequency.

The "surfer1" and "east1" sequences had very different visual appearance due to changes in the view angle and sun angle. The first sequence had a wave pattern that was easy to see, whereas the second sequence showed mostly sun glint. Our method for finding the scale factor gave reasonable results in both cases, and our analysis of the sea state produced comparable results for both, in spite of the large visual differences.

VI. Discussion

The methods proposed here work in a wide variety of conditions, but as with most things, there are limits. The sequence must have a frame rate fast enough to capture the temporal frequencies present in the current sea state without aliasing, and be long enough to cover a number of cycles of these frequencies. The dominant periods range from 2 s for sea state 2 to 20 s from sea state 10. In sea state 2, the wave periods with significant energy range from 1 to 3 s. According to the Nyquist–Shannon sampling theorem, we need a sampling rate double the highest frequency to recover the frequency spectrum, which is 2 Hz. for this case. For reliable results, the sequence should cover at least 20 cycles of the dominant period, which translates to a minimum length of 40 s for seastate 2, increasing to 400 s for seastate 10.

We tested various shorter clips from the sea state 2 synthetic video, and found that the scale could still be recovered within 16% of the ground truth value for durations as short as 4 s (twice the dominant period). The calculated wind speed varied by as much as 36% from the ground truth for durations of more than 5 s.

In the spatial domain, the dominant wavelengths vary from 6 m in seastate 2 to 600 m in seastate 10. Within seastate 2, the wavelengths with significant energy range from 1 to 12 m. Video with at least 2 pixels/m should avoid spatial aliasing at all sea states. Since the peak wavelength is derived from the peak period using the temporal domain, there are no requirements for the camera field-of-view to cover a specified number of wavelengths.

We tested the robustness of the method to the scale factor by generating synthetic sequences for sea state 2 from a range of altitudes such that β ranged from half a pixel per meter to over 100 pixels/m. The errors in the recovered scale factor were less than 2% for $\beta <$ 40, but got worse as β increased beyond 40. We attribute these errors at low altitudes to fewer spatial wavelengths being visible in each frame and the polygons in the synthetic model getting larger.

Our procedure does not require excessive amounts of memory. One minute of uncompressed video at a modest 300×300 resolution adds up to over 160 million pixels. When 8 bytes are needed for each pixel (32-bit floats for both the real and imaginary parts) and multiple copies are required (source and destination of an operation), the memory capacity of even high-end machines can quickly be exceeded if all the data is required at once (such as for a 3-D FFT). Since we operate on one frame at a time and accumulate intermediate results in a 2-D image, we can handle higher resolutions and longer durations of video without running out of memory.

More precisely, the storage requirements of our algorithm are order $O(LN)$, where L is the number of pixels in a slice and N is the number of frames in the sequence. This is the size of the accumulated FFT image used to find the scale, as well as the size of the collected strips used to find seastate. Subsequent FFTs of these image in both algorithms produce results of the same size.

As for computational complexity, the scale calculation is dominated by finding the 2-D FFT of each frame. FFT in one dimension is $O(n \log n)$, where n is the number of elements. FFT in 2-D consists of a one-dimensional (1-D) FFT for each row, followed by a 1-D FFT for each column, or $O(n^2 \log n)$, for an $n \times n$ image. Performing this operation N times produces a computational complexity of $O(Nn^2 \log n)$ for finding the scale. Sea state computation requires an FFT on each of L rows, each having N elements, resulting in $O(LN \log N)$ (remember, L is the number of pixels in a slice, and varies from n to $\sqrt{2}n$, depending on the direction of the slice).

VII. Conclusion

We have used Fourier transforms to exploit the dispersion relation property of water waves to find the real world scale (in pixels per meter) for synthetic and real sequences with no prior camera calibration. This well-known (in oceanography circles) equation quantifies the square relationship between the wavelength and period of a wave. We used the known time axis of the video to find the unknown scale of the scene. The results were validated by the ground truth in the synthetic case, and known scene geometry for the real sequences. We are not aware of any previous work in the image processing community that has used the dispersion relation to recover the scale.

In addition, we used the temporal frequency spectrum together with known stochastic models of ocean frequency spectra to determine the sea state, including the dominant wavelength and period, the wind speed that created the waves, and the probable wave height. All of these properties were recovered from the video with no prior knowledge of the camera parameters or scene geometry.

This process can be used in tasks such as coastal surveillance to determine the size of ships, or in marine research to find the size of objects in an ocean surface video when no scale reference is available.

Ocean scene analysis is not as specialized a field as it first appears, due to the abundance of ocean on our planet, and the importance of security and monitoring of coastlines, since they usually coincide with international borders. We have used properties that are unique to water for this process, but we seek to contribute to the overall effort of understanding dynamic natural phenomena.

References

[1] C. R. Wren, A. Azarbayejani, and A. Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 780–785, Jul. 1997.

[2] C. Stauffer and W. Grimson, "Learning patterns of activity using real time tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 747–767, Aug. 2000.

[3] L. Zhang, B. Curless, A. Hertzmann, and S. M. Seitz, "Shape and motion under varying illumination: Unifying structure from motion, photometric stereo, and multi-view stereo," in *Proc. 9th IEEE Int. Conf. Computer Vision*, Oct. 2003, pp. 618–625.

[4] S. Negahdaripour and C.-H. Yu, "A generalized brightness change model for computing optical flow," in *Proc. IEEE 4th Int. Conf. Computer Vision*, May 1993, pp. 2–11.

[5] P. Voles, M. Teal, and J. Sanderson, "Target identification in a complex maritime scene," in *Proc. IEE Colloq. Motion Analysis and Tracking*, May 1999, pp. 15/1–15/4.

[6] A. Smith, M. Teal, and P. Voles, "The statistical characterization of the sea for the segmentation of maritime images," in *Proc. 4th EURASIP Conf. Video/Image Processing and Multimedia Communications*, Jul. 2003, pp. 489–494.

[7] G. Doretto, A. Chiuso, Y. Wu, and S. Soatto, "Dynamic textures," *Int. J. Comput. Vis.*, vol. 51, no. 2, pp. 91–109, 2003.

[8] A. Fitzgibbon, "Stochastic rigidity: Image registration for nowhere-static scenes," in *Proc. IEEE Int. Conf. Computer Vision*, Jul. 2001, pp. 662–669.

[9] A. Monnet, A. Mittal, N. Paragios, and V. Ramesh, "Background modeling and subtraction of dynamic scenes," in *Proc. 9th IEEE Int. Conf. Computer Vision*, Oct. 2003, pp. 1305–1312.

[10] A. Fournier and W. Reeves, "A simple model of ocean waves," *Proc. ACM SIGGRAPH*, vol. 20, no. 4, pp. 75–84, Aug. 1986.

[11] G. Mastin, P. Watterberg, and J. Mareda, "Fourier synthesis of ocean scenes," *IEEE Comput. Graph. Appl.*, vol. 7, no. 3, pp. 16–23, 1987.

[12] J. Tessendorf, "Simulating ocean water," in *Proc. Siggraph Course 26 Notes*, Aug. 1999, pp. 3.1–3.18.

[13] ——, "Simulating ocean water," in *Proc. Siggraph Course 47 Notes*, 2001, pp. 3.1–3.61.

[14] S. Thon and D. Ghazanfarpour, "Ocean waves synthesis and animation using real world information," *Comput. Graph.*, vol. 26, pp. 99–108, 2002.

[15] J. Sanderson, M. Teal, and T. Ellis, "Characterization of a complex maritime scene using fourier space analysis to identify small craft," in *Proc. 7th Int. Conf. Image Processing and Its Applications*, Jul. 1999, pp. 803–807.

[16] B. Kinsman, *Wind Waves*. Mineola, NY: Dover, 1984.

[17] F. G. W. Smith, *The Seas in Motion*. New York: Crowell, 1973.

[18] S. R. Massel, *Ocean Surface Waves: Their Physics and Prediction*, ser. Advanced Series on Ocean Engineering, Singapore: World Scientific, 1996, vol. 11.

[19] L. Spencer and M. Shah, "Water video analysis," presented at the IEEE Int. Conf. Image Processing, Singapore, Oct. 2004.

[20] Z. Rasheed, Y. Sheikh, and M. Shah, "On the use of computable features for film classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 1, pp. 52–64, Jan. 2005.

[21] E. Ribeiro and E. Hancock, "Estimating the perspective pose of texture planes using spectral analysis on the unit sphere," in *Pattern Recognit.*, 2002, vol. 35, pp. 2141–2163.

Lisa Spencer (S'02) received the B.S. degree in computer engineering from the University of Arizona, Tucson, in 1984, and the M.S. and Ph.D. degrees in computer science from the University of Central Florida (UCF), Orlando, in 2002 and 2006, respectively.

She has worked for 20 years in the graphics and simulation industries with companies including Quantum3D, Real 3D, Lockheed Martin, Martin Marietta, and GE.

Mubarak Shah (M'81–F'03) received the B.E. degree in electronics from Dawood College of Engineering and Technology, Karachi, Pakistan, in 1979. He completed the EDE diploma at Philips International Institute of Technology, Eindhoven, The Netherlands, in 1980, and received the M.S. and Ph.D. degrees in computer engineering from Wayne State University, Detroit, MI, in 1982 and 1986, respectively.

Since 1986, he has been with the University of Central Florida, Orlando, where he is currently the Agere Chair Professor of Computer Science and the Director of the Computer Vision Laboratory. He has served as a Project Director for the national site for Research Experience for Undergraduates (REU) in Computer Vision, funded by the U.S. National Science Foundation since 1987. He is the coauthor of the books *Motion-Based Recognition* (Kluwer, 1997) and *Video Registration* (Kluwer, 2003). He has published more than 150 papers in leading journals and conferences on topics including visual motion, tracking, video registration, and activity and gesture recognition.

Dr. Shah was awarded a five year Quaid-e-Azam (Father of Nation) scholarship for the Ph.D.degree. he was an IEEE Distinguished Visitor Speaker for 1997 to 2000. He received the Harris Corporation Engineering Achievement Award in 1999; the TOKTEN awards from UNDP in 1995, 1997, and 2000; the Teaching Incentive Program award in 1995 and 2003; the Research Incentive Award in 2003; and the IEEE Outstanding Engineering Educator Award in 1997. He is an Editor of an international book series on video computing, Editor-in-Chief of *Machine Vision and Applications*, and an Associate Editor of *Pattern Recognition*. He was an Associate Editor of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE and a Guest Editor of the *International Journal of Computer Vision* special issue on video computing.

Ratan K. Guha (S'67–M'70) received the B.Sc. degree (with honors) in mathematics and the M.Sc. degree in applied mathematics from Calcutta University, Calcutta, India, and the Ph.D. degree in computer science from the University of Texas at Austin in 1970.

He is currently a Professor of computer science at the University of Central Florida, Orlando. He has authored over 100 papers published in various computer journals, book chapters, and conference proceedings. His research interests include distributed systems, computer networks, security protocols, modeling and simulation, and computer graphics. His research has been supported by grants from ARO, NSF, STRICOM, PM-TRADE, and the State of Florida.

Dr. Guha has served as a member of the program committee of several conferences, as the General Chair of CSMA'98 and CSMA'2000, and as the Guest Co-Editor of a special issue of the *Journal of Simulation Practice and Theory*. He is a member of ACM, IEEE, and SCS and is currently serving as a member of the Board of Directors of SCS.