WATER VIDEO ANALYSIS

Lisa Spencer and Mubarak Shah

Computer Vision Lab
University of Central Florida
Orlando, FL 32816

ABSTRACT

Many image processing and computer vision applications have difficulty dealing with a nonstatic background such as water waves, but this particular dynamic scene actually contains enough information to reveal much more than a static scene. In this paper, we propose a method to determine real world scale as well as other factors, including wave height, sea state, and wind speed, from uncalibrated water video. We do this by using low level image processing tools to extract a spatial frequency spectrum from individual frames of the video and temporal frequencies from the time dimension of the video, and applying known physics of water waves to find the high level properties of the scene. An example is presented to demonstrate and validate the process.

1. INTRODUCTION

At first glance, a topic that applies only to ocean water may seem quite narrow, but water covers the majority of our planet, and oceans comprise 92% of this water [8]. Ocean waves have been studied for hundreds of years. For early mariners, understanding of the ocean's patterns was a matter of life and death. While modern science has enabled us travel the seas in comfort and forecast the conditions with accuracy, ocean waves continue to be a source of irritation to the image processing and computer vision communities, due to their constantly changing appearance.

Many video analysis processes, like background subtraction and camera motion compensation, rely on the uninteresting parts of the scene being static. Dynamic components that are part of the background, including trees blowing in the wind and moving water, form video textures [7]. The graphics community models these phenomena for the purpose of simulating them, and work has been done in the vision community to classify temporal textures by their statistics [6]. Fitzgibbon [2] seeks to recover camera motion by treating the image sequence as a random process, and searching the space of possible motions at each frame to find the motion that results in the most efficient model of the random process. Doretto et al. [1] find static boundaries between regions with different spatio-temporal statistics.

In this paper, we deal specifically with water, not general temporal textures. We exploit the properties of water, such as the relationship between wavelength and wave period, and low level features of the video, such as spatial and temporal frequency spectra, to learn higher level properties of the environment, like real world scale, wave height, and even wind speed. Figure 1 shows a block diagram of our method. Fourier transforms of individual frames are used to find the energy at various spatial frequencies. Principal component analysis (PCA) of the whole video sequence followed by another Fourier transformation is used to find the energy at various temporal frequencies. Application of wave physics allows us to convert the temporal frequencies to real world wavelengths. Correlation of the spatial and temporal information leads to knowledge of the image dimensions in real world units, without the "up to a scale factor" disclaimer usually applied to imagery. Once the real world scale is known, the actual size of objects in the scene as well as their velocities can be found, with no prior knowledge about camera calibration.

Section 2 outlines the physical properties of water and waves that are needed for this application. Section 3 details the features that are extracted from individual frames of the video, and Section 4 describes the temporal properties. These pieces are integrated in Section 5 and applied to actual video in Section 6. Conclusions and future directions are listed in Sections 7 and 8.

2. WATER PHYSICS

Waves in the open ocean appear at first glance to be chaotic and unpredictable. However, there are some constants in the seemingly random ocean surface. We will focus on the behavior of open water. Shore effects, like waves breaking on the beach, have their own body of literature, and will not be dealt with here.

Once created, waves that depend on gravity can travel a long distance without significant loss of energy. Anyone who has thrown a pebble into a still pond has seen that the

0-7803-8554-3/04/$20.00 ©2004 IEEE.

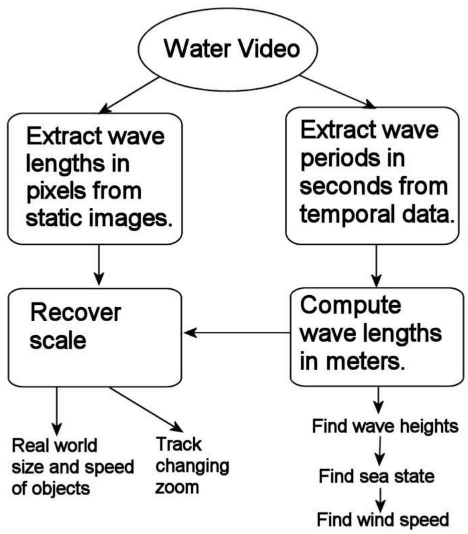


Fig. 1. Overview of water video analysis

ripples continue long after the pebble that caused them has sunk to the bottom. Likewise, if a storm causes rough seas, when the wind ceases, the waves will remain. The only aspect to change quickly is the tiny capillary waves, which vanish with the wind, since they are controlled by surface tension. Therefore, except for the highest frequencies, the overall characteristics of the water vary slowly over time. Observations of the same area of open water within a few hours are likely to capture fairly constant conditions [4].

The period of a wave is the time for one complete cycle of a wave, that is, the time between two crests passing a stationary point. The wavelength is the spatial distance between two crests. The wave speed is then the ratio of the wavelength and the wave period. But the relationship between these quantities is not random. A wave with a longer period has a longer wavelength, as well as faster speed.

More precisely,

$$L = \frac{g}{2\pi}T^2 \tag{1}$$

$$c = \sqrt{\frac{gL}{2\pi}} \tag{2}$$

where L is the wavelength, T is the wave period, c is the wave speed, and g is the acceleration due to gravity (9.8 m/s^2) [4].

Of course, individual waves or wave trains are difficult to perceive in the open ocean. The ocean is not made up of waves of a single frequency, but of many frequencies and directions. While the surface of a particular patch of water is always changing, the frequency spectrum remains the same, allowing a constant statistical description of the water. Wave analysis for forcasting is based on the assumption of a stationary ergodic random process [5].

3. SINGLE IMAGE PERIODIC FEATURES

The spatial frequency spectrum of a patch of water can be computed from a single overhead image using a discrete Fourier transform [9]. One way to decompose the 2D frequency spectrum is to compute the energy as a function of radius or direction [3]. Direction is not relevant to our task, so we compute the spectrum $S(r)$ by averaging the spectrum power (magnitude squared) for each value of the radius.

Since spatial frequency is the inverse of wavelength, it is easy to convert between the energy at a given frequency and the energy at a given wavelength. Since the radius in Fourier space is the number of cycles in the image, $L = w/r$, where L is the wavelength in pixels, w is the image width in pixels, and r is the radius in pixels in Fourier space. The spatial wavelengths can only be determined in units of pixels without knowing the depth and zoom.

Two samples captured at different times should produce the same frequency spectrum, according to our model. We can increase the accuracy of our model by averaging the spectra from multiple images. In fact, samples (images) that have the same depth and zoom should have the same spectra, regardless of pan, tilt, roll, or time differences.

If the depth or zoom changes, it changes the scale of the sample, but not the underlying frequencies, so the new spectrum will be a scaled version of the old one, within the window of measurable frequencies. The discrete Fourier transform cannot capture wavelengths greater than the image size or smaller than two pixels.

4. VIDEO PERIODIC FEATURES

With video as an input, we can average the frequency spectra across multiple frames, as described in the previous section, but we also have the added dimension of time. While we can't extract real world length units like feet or meters directly from still images or video, we *can* determine time in seconds. By analyzing the time dimension of the video, we can find wave periods in seconds.

Assuming the camera is static, one way to do this is to first reduce the dimensionality of the data via principal component analysis (PCA) [2]. This reduces the 3D (x, y, t) data into a 2D array with a row for each frame, and columns containing the coefficients of each principal component. Summing the power in each row of the DFT of this matrix will give the energy at each temporal frequency.

Since we know the frame rate of the video, we now have the energy spectrum for different wave periods in real world units of seconds. The cycles can be converted to periods using $T = n/(30f)$, where T is the wave period in seconds,

Beaufort number	Wind (knots)	Wave height (feet)	Description
0	< 1	0	Like mirror
1	1 - 3	0	Ripples
2	4 - 6	0 - 1	Small wavelets
3	7 - 10	1 - 2	Scattered whitecaps
4	11 - 16	2 - 4	Numerous whitecaps
5	17 - 21	4 - 8	Some spray
6	22 - 27	8 - 13	More spray
7	28 - 33	13 - 20	White foam
8	34 - 40	13 - 20	Foam is blown
9	41 - 47	20	Dense streaks of foam
10	48 - 55	20 - 30	Overhanging crests
11	56 - 63	30 - 45	Covered with foam
12	64+	45+	Air filled with foam

Table 1. Beaufort Scale

n is the number of frames in the video, and 30 is the number of frames per second.

5. PUTTING IT ALL TOGETHER

From individual images, we can get spatial frequencies (i.e., wavelengths) in pixels. From video, we can obtain wave periods in seconds. Since we're talking about water waves in the ocean, the relationships in Equations 1 and 2 apply, so we can convert the wave periods in seconds to wavelengths in meters. This gives us an energy spectrum for wavelengths in meters from the video, and an energy spectrum for wavelengths in pixels from the individual frames. Subject to the sampling limitions (short videos or zoomed in still frames will not be able to capture long wavelengths), the two spectra can be correlated, resulting in a pixels per meter conversion. With this conversion factor, many things are now possible, like determining the real world size and velocity of objects in the scene.

Wave heights can also be estimated. Wave steepness is defined as the ratio of the wave height (trough to crest) to the wave length. Observations of wave steepness range from 0.1 to 0.008 [4]. So the wave heights are likely to be in the range of $0.008L$ to $0.1L$, where L is the wavelength.

Knowledge about the wave heights leads to an estimate of wind speed. The Beaufort scale was introduced in 1805 by Sir Francis Beaufort to relate wave conditions and wind speeds [8]. It was originally used to gauge the wind speed by looking at the ocean, in order to decide whether to add or take in sail. In later years, it was used in the opposite fashion, to predict the wave heights based on the measured wind speed. It is no longer used in official forcasts, but still provides an intuitive description of the state of the ocean. Table 1 lists the wind speeds corresponding to different wave heights.

Fig. 2. The first frame of three videos. The first two are of the same lake, while the third is the ocean.

6. EXPERIMENTAL RESULTS

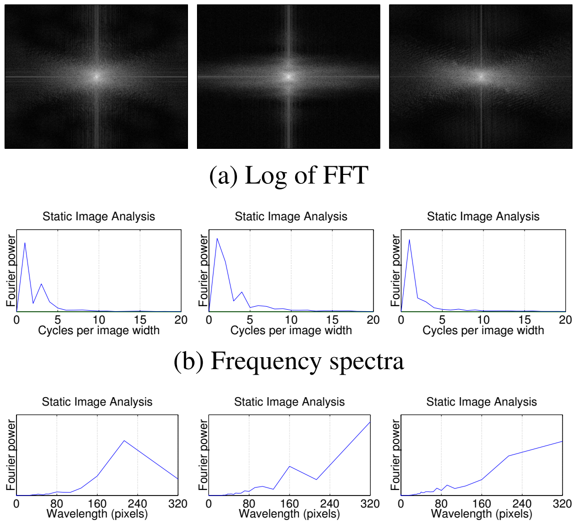
We analyzed three video sequences. Two video sequences were acquired by a static camera viewing the same large lake from the downwind shore on a windy day. The third was viewing the ocean from a pier. Each video is about 30 seconds (900 frames) in length. Figure 2 shows the first frame of each sequence. The first two were taken under the same conditions, but with different zoom, to capture different scales of the same water, while the third was a different place on a different day.

The first frames were analyzed as in Section 3, after correcting for perspective distortion. The Fourier spectra are shown in Figure 3a. The spatial frequencies $S(r)$ are shown in Figure 3b. The peak at one cycle per image is due to the fact that the right (top) edge does not match the left (bottom) edge. The wavelength spectra in units of pixels are shown in Figure 3c. The left pair in Figure 3b have roughly the same shape, but the horizontal scale differs. The resemblance is also present in Figure 3c, since the difference between b and c is just a transformation of the horizontal axis.

The sequences were then analyzed according to Section 4. The result of summing the rows of the FFT of the principal component coefficient matrix is shown in Fig. 4a, which shows the frequency data in terms of cycles per sequence (cycles per 30 seconds). The result of converting the frequency data to wave period is shown in Figure 4b. Applying Equation 1 gives us wavelengths in meters, shown in Figure 4c. The peaks appear at about the same wavelengths in the first two graphs. This time, there is no horizontal scale difference, since the scale is in real world units of meters. The third is different, reflecting the different conditions.

For the first pair of images, assuming the peak at two meters corresponds to the static image peaks at 220 pixels on the first image and 160 pixels on the second image, we have a conversion of 110 pixels/meter on the first, and 80 pixels/meter on the second. Going one step further, with a dominant wavelength of two meters, the wave heights are likely to range from $0.008 \times 2 = 0.016$m to $0.1 \times 2 = 0.2$m, or about 0.65 ft.

The single images are limited in their ability to capture longer wavelengths, whereas the 30 second video can capture wave periods up to 15 seconds, which corresponds to a wavelength of 350 meters. The spectrum that was obtained from the first two videos has another peak at 10 meters. The

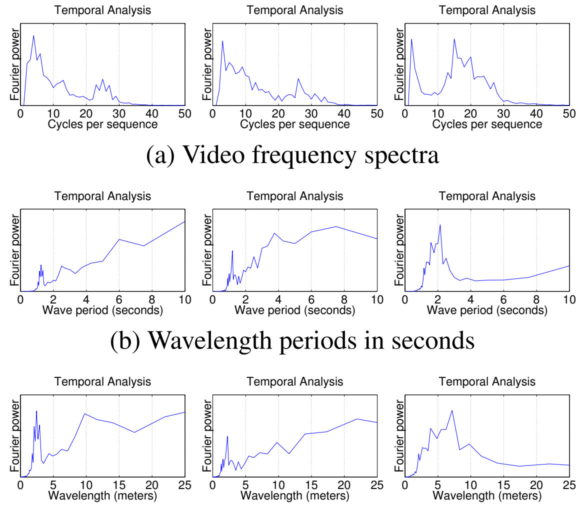


(a) Log of FFT

(b) Frequency spectra

(c) Wavelength spectra in pixels

Fig. 3. Frequency information extracted from the first frame of each sequence.



(a) Video frequency spectra

(b) Wavelength periods in seconds

(c) Wavelength spectra in meters

Fig. 4. Frequency information calculated from the temporal component of 30 seconds of video in each sequence.

waves at this wavelength are likely to be 0.08 to 1 meter high. This is consistent with Beaufort number 3, with winds from 7 to 10 knots, with scattered whitecaps. This, in fact, matched the conditions of the day fairly well.

The third video has the most energy in the waves with wavelength around 7 meters. Since a single frame was not wide enough to capture multiple cycles of this wavelength, the corresponding peak is not visible in Figure 3c. Thus, the method is better suited to wider camera views or smaller waves.

7. CONCLUSION

We have proposed and demonstrated a method to obtain real world scale from uncalibrated video of a water scene. Not only were we able to find the scale factor between pixels and meters, but we were able to judge wave heights and wind speed as well, with no prior calibration.

8. FUTURE WORK

Once the model is known for the patch of water, changes in camera zoom can be detected by monitoring the scale factor of the spatial frequency spectra in successive frames. It should also be possible to determine camera pan, tilt and translation by finding the nonperiodic components of the temporal spectrum. The spectrum directional energy $S(\theta)$ can be used for determining camera roll (rotation around the optical axis). In addition, the regions that do not fit the model can be found, to do a variant of background subtraction for water scenes. We intend to elaborate on and validate these ideas as our work continues.

9. REFERENCES

[1] Doretto, Cremers, Favaro, Soatto, "Dynamic Texture Segmentation", International Conference on Computer Vision, 2003.

[2] Andrew Fitzgibbon, "Stochastic rigidity: Image registration for nowhere-static scenes," International Conference on Computer Vision, 2001, pp. 662-669.

[3] Gonzalez, Woods, *Digital Image Processing,* 2nd. Ed., Prentice Hall, 2002.

[4] Blair Kinsman, *Wind Waves,* Dover 1984. (Reprint of Prentice-Hall, 1965.)

[5] Stanislaw Massel, *Ocean Surface Waves: Their Physics and Prediction,* World Scientific, 1996.

[6] Ramprasad Polana, Randal Nelson, "Temporal texture and activity recognition", *Motion-Based Recognition,*, Shah and Jain (eds.), Kluwer Academic Publishers, 1997.

[7] Schodl, Szeliski, Salesin, Essa, "Video Textures", Siggraph 2000, pp. 489-498.

[8] F. G. Walton Smith, *The seas in motion,* Thomas Y. Crowell Company, 1973.

[9] Thon, Ghazanfarpour, "Ocean waves synthesis and animation using real world information," Computers and Graphics, Vol. 26(1) Feb. 2002, pp. 99-108.