

Peak Wave Period and Direction Estimation Using 3D FFT on Monoscopic Videos

Nicolas Paris(✉), Sylvain Marchand, and Petra Gomez-Krämer

L3i Laboratory, La Rochelle University, La Rochelle, France
{nicolas.paris,sylvain.marchand,petra.gomez}@univ-lr.fr

Abstract. Reliable estimation of sea state parameters is essential for maritime operations, particularly for small vessels and sailing boats. Conventional approaches rely on in-situ sensors or radar systems, which are often costly, energy intensive, or limited in deployment. This paper investigates a vision-based method for estimating the peak wave period and direction from monoscopic video sequences using three-dimensional Fourier analysis. The proposed approach operates on stabilized video data and applies a bird's-eye-view transformation to reduce perspective effects prior to computing a 3D Fast Fourier Transform over spatial and temporal dimensions. A directional wave spectrum is then derived, from which the peak wave frequency and direction are extracted. The method is evaluated using a set of synthetically generated video sequences of ocean simulations, as well as real-world sea-surface imagery from a public dataset. Experimental results confirm that the wave direction can be estimated with consistent accuracy, while the frequency estimation shows sufficient precision but exhibits a systematic bias. These findings suggest that monoscopic video can provide a low-cost solution for onboard wave monitoring.

Keywords: Sea state estimation · Monoscopic video analysis · 3D Fourier transform

1 Introduction

Accurate sea state estimation plays a critical role in maritime operations, influencing vessel performance, safety, and navigation strategies. For sailing ships in particular, the sea state directly affects sailing performance, route optimization, and crew safety. Reliable onboard sea state estimation could help skippers and autonomous systems to make informed decisions, such as sail adjustments or course changes. The sea state is defined as the surface condition of the ocean and it can be linked to physical wave parameters such as significant wave height H_s, peak frequency f_p and direction θ_p, or the steepness of the wave. Traditionally, wave parameters are obtained using buoys [12], or radar [1]. However, these sensors are expensive, energy consuming or limited in coverage.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
M. De Marsico et al. (Eds.): ICPR 2026, LNCS 16821, pp. 542–556, 2027.
https://doi.org/10.1007/978-3-032-31438-3_36

The aim of this project is to enable a ship to sense wave information using a video camera which is readily available and cheap sensors. Some vision-based methods were tried, such as stereo-vision approaches [2,7] to reconstruct the 3D surface of the water. An attempt at estimating the sea spectrum has been made using a monoscopic camera by Loizou et al. [11]. They used a Kalman filter to follow pixels across the frame of a stabilized video to extract a sea spectrum, allowing the extraction of wave frequency and height.

In this paper, we propose a method that uses a 3D FFT on stabilized monoscopic video sequences to estimate the peak wave period and direction. The remainder of the article is organized as follows. In Sect. 1 we outline the ocean modeling methods, in Sect. 2 we present the workflow of our estimation method (Fig. 1), and in Sect. 3 we present the results obtained with this method on synthetic and real datasets. Finally, Sect. 4 concludes our work.

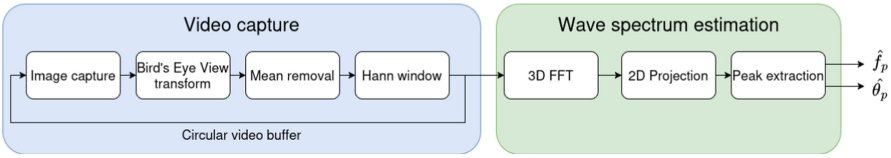


Fig. 1. Flowchart for the estimation method.

2 Modeling the Ocean Waves

The sea state is defined as the general condition of the surface water. It can be described by the sea wave spectrum $S(f, \theta)$. This spectrum gives the energy distribution of the sea as a function of wave frequency f and direction θ. The frequency spectrum $S(f)$ is referred to as the dimensionless spectrum and represents the distribution of energy as a function of wave frequency.

There exist several dimensionless spectral models such as the Phillips [15], Pierson-Moskowitz [16] and JONSWAP [10] spectra. Linking measured spectra to these models enables parameterization of sea conditions such as fetch-limited growth, wind input, and spectral peak sharpness and aids in both interpretation and forecasting of the ocean behavior.

The sea spectrum can be measured using various sensors, such as in-situ wave buoys enabling direct time-series recordings of surface elevation and derivation of spectral energy distributions [12]. Radars can also be used, typically X-band navigation radars which infer the directional wave spectrum by analyzing sea-clutter modulation or Doppler shifts [1]. Satellite-based altimeters and synthetic aperture radars (SAR) are used to estimate sea state parameters at large scale [19].

Some embedded methods to measure the sea spectrum are also used. The movement of the ship itself can be used to estimate the perceived sea spectrum by considering the ship as a buoy [4].

2.1 JONSWAP Spectrum

The JONSWAP (Joint North Sea Wave Project) spectrum [10] is a widely used parametric model for describing wind-generated seas. It is an extension of the Pierson-Moskowitz spectrum, it introduces a peak enhancement factor, which accounts for the growth stage of the wave field.

$$S(f) = \frac{\alpha g^2}{(2\pi)^4} \cdot f^{-5} \cdot \exp(-\frac{5}{4} \cdot \frac{f_p}{f}^4) \cdot \gamma^{\beta} \tag{1}$$

with

$$\alpha = 0.076 \cdot (g \cdot F \cdot U_{10}^{-2})^{-0.22}$$

$$f_p = \frac{3.5 \cdot g \cdot (g \cdot F \cdot U_{10}^{-2})^{-0.33}}{U_{10}}$$

$$\beta = \exp(-\frac{(f - f_p)^2}{2 \cdot \sigma^2 \cdot f_p^2})$$

$$\sigma = \begin{cases} 0.07 \text{ for } f \leq f_p \\ 0.09 \text{ for } f > f_p \end{cases}$$

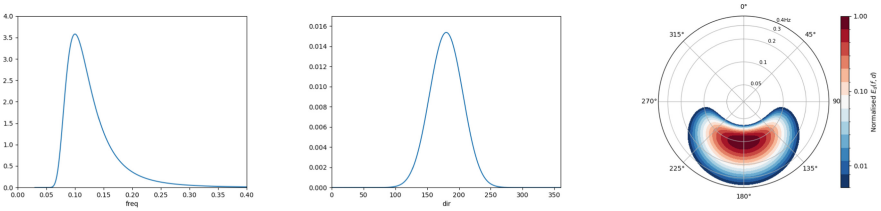
where F is the fetch, U_{10} is the wind speed at 10 m above the sea surface, f_p is the peak frequency, $g \approx 9.81$ m/s^2 the gravitational acceleration and γ is typically set between 1 and 7 depending on the sea development stage.

2.2 Cartwright Cosine-Squared Spread

In the JONSWAP spectrum given by Eq. (1), the direction of the waves is not taken in account. To add the direction component to the spectrum, one must use a directional spread function. The Cartwright cosine-squared spread function [5] is defined as follows

$$s = \frac{2}{(\sigma^2)} - 1 \tag{2}$$

$$G(\theta) = \cos^{2\,s} \frac{\theta - \theta_m}{2} \tag{3}$$



(a) $S(f)$ (b) $G(\theta)$ (c) $S(f, \theta)$

Fig. 2. JONSWAP spectrum with Cartwright cosine-squared spread function (here with peak frequency $f_p = 0.1$ Hz and direction $\theta_p = 180°$).

with σ as the spreading angle.

The final directional spectrum (Fig. 2) is given by

$$S(f, \theta) = G(\theta)S(f) \tag{4}$$

3 Estimation of the Wave Spectrum

The estimation of the wave spectrum consists of several steps. The steps will be described throughout this section. A video stabilization step might be required as the very first step, but in our case it is not needed as all videos are coming from static cameras.

3.1 Bird's Eye View Transform

The bird's eye view (BEV) transform is a perspective transformation used to suppress the effect of perspective. This requires applying a transform to all pixels of the image from camera C_1 to their corresponding pixels from camera C_2 as shown in Fig. 3.

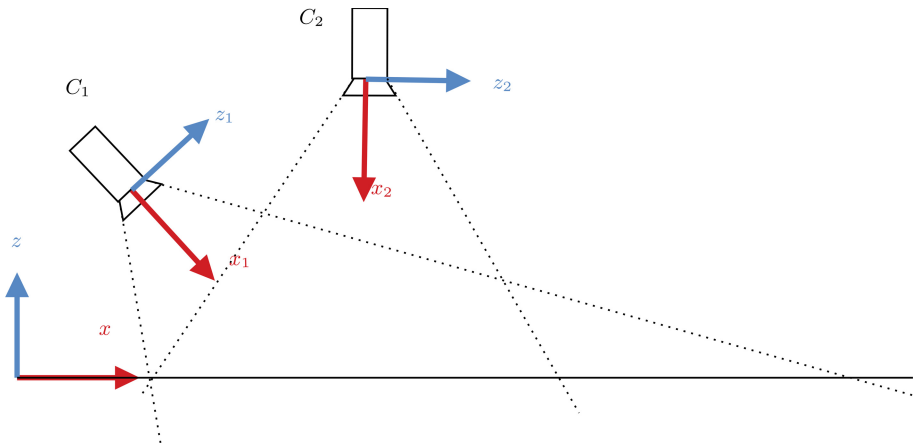


Fig. 3. Side view of the bird's eye view transform.

This step requires that the camera is calibrated, its pose and focal distance need to be known. The pinhole camera model [9] is used to map points X_w in world coordinates to points u on the image plane

$$u = PX_w = KMX_w \tag{5}$$

with P as the projection matrix and K as the intrinsic camera matrix

$$K = \begin{pmatrix} f_x & 0 & C_x & 0 \\ 0 & f_y & C_y & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{6}$$

where (C_x, C_y) denotes the image center in pixels and (f_x, f_y) the focal distances on the x and y axis. M is the extrinsic camera matrix defined as

$$M = \begin{pmatrix} R & -RT \\ 0 & 1 \end{pmatrix} \tag{7}$$

with R and T, the rotation and translation matrices of the camera.

The relation (5) is not injective, meaning that one could determine the pixel u from a point in the space X_w, but from a pixel u it is only possible to find a line from the center of the camera to the point X_w.

To map a pixel u to a point X_w, we need to know where this point is located on a given axis. In our study we will assume that every pixel u maps to a point lying in the horizontal plane $(z = 0)$. Let L be a matrix that satisfies

$$u = PLX_w \tag{8}$$

thus

$$X_w = (PL)^{-1}u \tag{9}$$

The matrix L is defined so that the points X_w are in the $(z = 0)$ plane.

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{10}$$

The matrix H is defined as the homography mapping the pixels from the camera C_1 to the pixels of the camera C_2.

$$\forall (u_1, u_2) \exists H, u_2 = Hu_1 = P_2L(P_1L)^{-1}u_1 \tag{11}$$

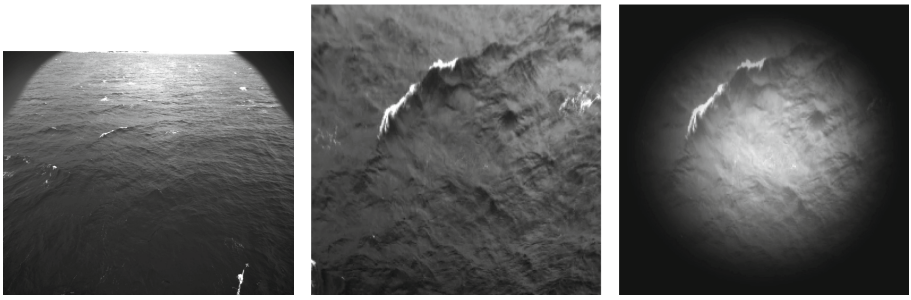
A drawback of this geometric method is that the computed homography is not surjective, some pixels u_2 are image of no pixel u_1 meaning that an interpolation is needed to achieve a complete image from C_2. In practice, the 'warpPerspective' method from the OpenCV library [3] was used with Lanczos interpolation (Fig. 4).

3.2 Frequency Analysis

To facilitate the frequency analysis, the DC component is removed by subtraction of the mean pixel value across frames. This operation is used to enhance the low-frequency precision. A 3D Hann window is also applied to the video to reduce spectral leakage. Frequency peaks will be better defined around the true frequency value.

A Discrete Fourier Transform (DFT) is applied directly to the preprocessed data cube $z(x, y, \omega)$. The Fourier series of $z(x, y, \omega)$ are

$$\mathcal{Z}(k_x, k_y, \omega) = \sum_{x=0}^{N_x} \sum_{y=0}^{N_y} \sum_{t=0}^{N_t} z(x, y, t).e^{-i(k_x x + k_y y + \omega t)} \tag{12}$$



(a) Original image. (b) Bird's eye view image. (c) DC component removal and windowing.

Fig. 4. Preprocessing steps.

$$a = |\mathcal{Z}|$$

with a as the amplitude of the resulting $\mathcal{Z}$ Fourier space. The amplitude can then be used to easily estimate period information such as peak wave direction $\hat{\theta}_p$ and frequency $\hat{f}_p$.

3.3 Wave Spectrum Extraction

The resulting Fourier space $\mathcal{Z}$ can be converted to a 2D wave spectrum $S(f, \theta)$ by summing values in the data cube for each f and θ in respect to a cylindrical representation where f is the height from the middle of the cube and θ is the azimuth. Nearest neighbor interpolation is used. This operation is represented on Fig. 5.

$$S(f,\theta) = \sum_{i=0}^{N/2} a(\lfloor i \sin\theta \rfloor + \frac{N_x}{2}, \lfloor i \cos\theta \rfloor + \frac{N_y}{2}, f + \frac{N_t}{2}) \tag{13}$$

This representation clearly shows the periods and directions in an easily displayable manner, at the expense of the loss of the x/y-frequency information (blue line on Fig. 5) used to compute the wave length if needed.

3.4 Peak Wave Information Extraction

The dimensionless spectrum can be easily derived from the directional spectrum:

$$S(f) = \int_{\theta} S(f,\theta) d\theta \tag{14}$$

From this dimensionless spectrum, the peak frequency f_p is extracted. The more commonly used peak period can then be computed $T_p = \frac{1}{f_p}$. The wave amplitude could be estimated from the spectrum if needed (see [11]).

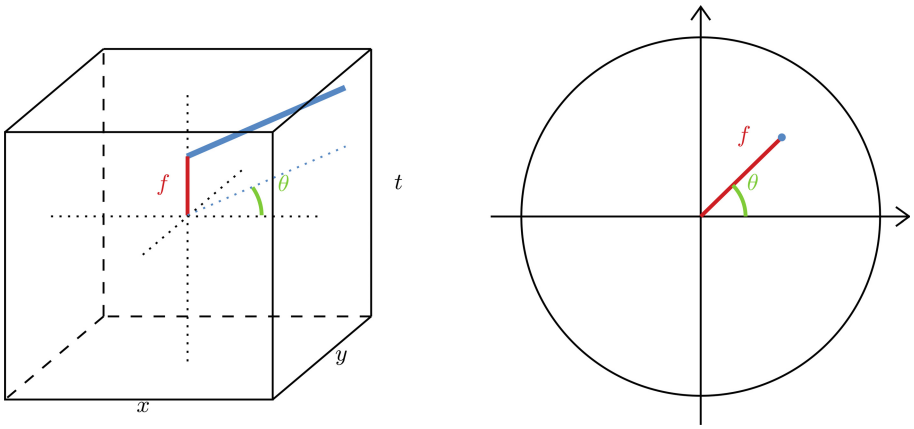
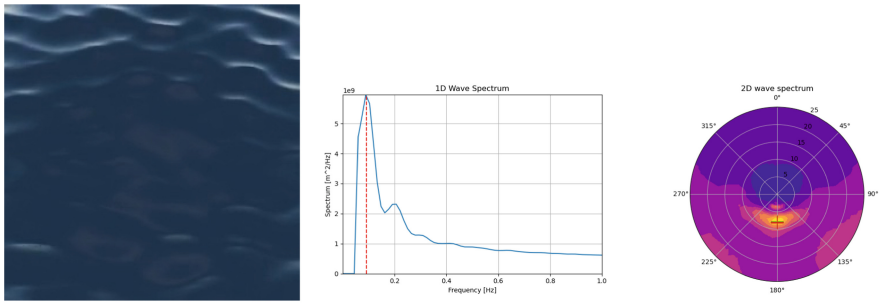


Fig. 5. Mapping between 2D directional spectrum in radial projection and 3D spectrum in Cartesian projection.

The peak direction is the direction at which the frequency is maximum (Fig. 6):

$$\theta_p = \arg\max_{\theta} S(f_p, \theta) \tag{15}$$



(a) Original image. (b) Dimensionless spectrum. (c) Directional spectrum.

Fig. 6. Peak wave information extraction.

4 Results

Our method was first validated using very simple examples of 2D sine waves mimicking the waves, which period and direction can be easily set. But of course we have to test it again on real data. Coastal water videos are not suitable for our needs as they exhibit non-linear dynamics and the waves are more likely to

break. Thus, to our knowledge, videos of waves with ground truth for period and direction are quite rare. In this paper we use a public dataset, presented in Sect. 4.2. In the future, we aim at acquiring our own dataset, with a camera fixed on the mast of a sailing boat. For now, we also propose to use a synthetic dataset [13], presented in Sect. 4.1.

4.1 Synthetic Data

Consider $s(x, y, t)$, the 3D volume of size N_x, N_y and N_t representing the video data over space (x, y) and time t.

Simulations were realized using blender [6] for its ease of use, the ability to render photo realistic scenes and its usage of the state of the art wave simulation algorithm [18]. This algorithm needs an input directional spectrum $S_i(f, \theta)$. The simulation will transform this 2D spectrum in a 2D height field using an Inverse Discrete Fourier Transform. This 2D height field can then be used to displace the vertices of a plane. The rendering is then done using the EEVEE rendering engine.

An added benefit of using a simulation is that we can choose to simulate a static scene, using a static camera. Or we can simulate a dynamic scene using a dynamic camera, attached on a boat floating on the ocean surface.

Simulations also allow us to test the presented method on many test scenes with different meteorological conditions without experimental constraints.



Fig. 7. Simulation examples.

A total of 684 video sequences have been generated. For each sea state, 171 sequences were rendered with the characteristics given by Table 1. Each sequence is composed by 600 images at 30 frames per second for a total of 20 s per sequence. The peak direction of the waves θ_p is chosen uniformly at random between 0 and 360°. The sun position is also randomized for each sequence, meaning that each sequence has a different lighting. Sun glare, sky and water colors are impacted by the sun position (Fig. 7).

f_p and H_s are manually set using wind speed U_{10} and fetch F, and the blender *scale* parameter respectively. The values of these parameters are chosen to correspond to the simulated sea state. Then, for each sequence these parameters will deviate at random by 10% maximum.

H_s is the significant wave height. It corresponds to the average of the highest third of the wave. It is given here as a basis of comparison but we do not try to estimate it in this work. The dataset generated for this study is available at [13].

Table 1. Synthetic dataset parameters.

Sea state	f_p range	H_s range	θ_p range
3	[0.16, 0.18] Hz	[0.65, 0.80] m	$[0, 360]^\circ$
4	[0.11, 0.13] Hz	[1.35, 1.65] m	$[0, 360]^\circ$
5	[0.09, 0.10] Hz	[2.25, 2.75] m	$[0, 360]^\circ$
6	[0.07, 0.08] Hz	[3.50, 4.50] m	$[0, 360]^\circ$

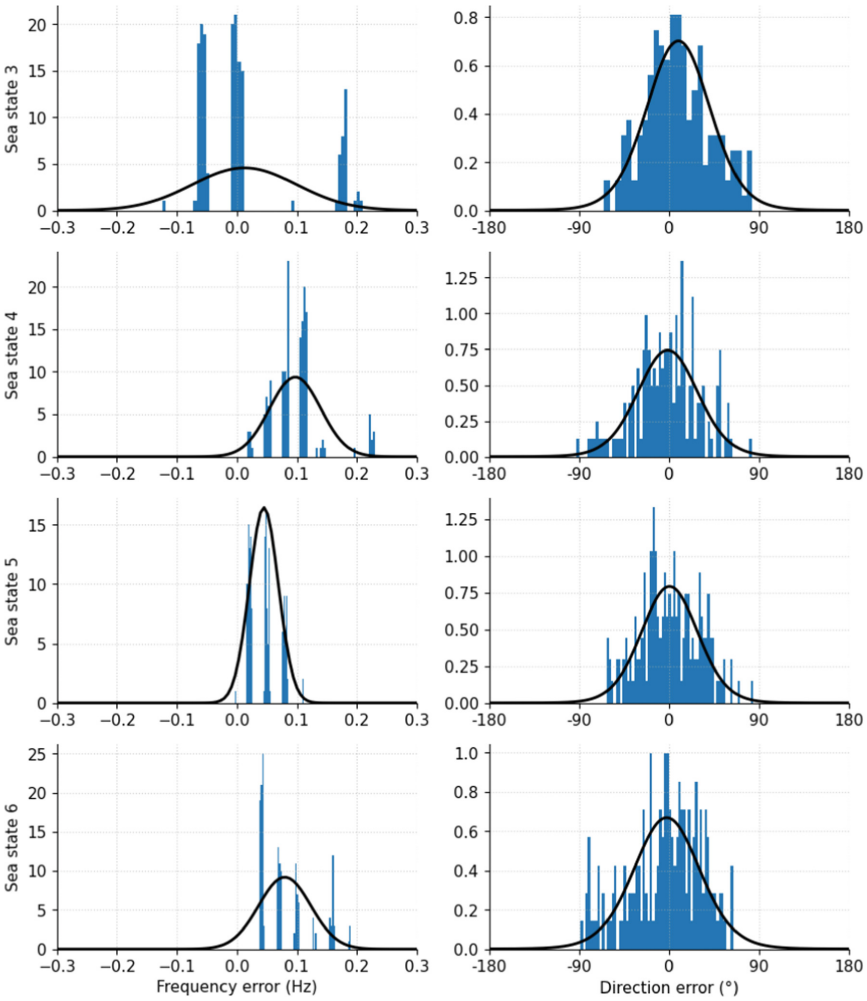


Fig. 8. Simulation error distribution.

For each sequence, the peak wave frequency and peak wave direction are estimated. Two estimation results are given as an example in Fig. 9. Figure 8 shows the error distribution for each considered sea state. The results show a satisfactory estimation for the direction estimation as the average error is centered in 0° and a standard deviation of about 30°.

However, the frequency error shows a sufficient precision but low accuracy of the estimate as the standard deviation is about 0.06 Hz but the average is about 0.05 Hz. A summary of the results is given in Table 3.

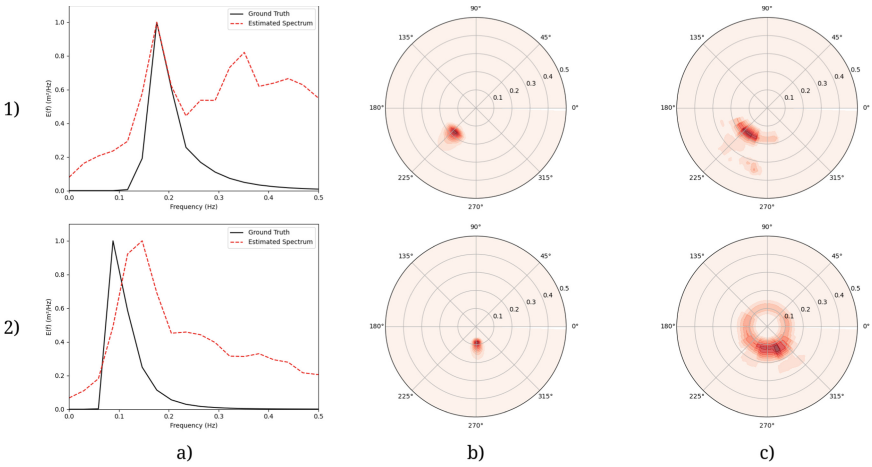


Fig. 9. Example estimation results from simulation sequences. a) represents the dimensionless spectrum with the ground truth in black and the estimate in red, b) the ground truth directional spectra, and c) the estimated directional spectra. The spectra of the top row 1) were obtained from a simulation from the sea state 3 where the estimation errors are minimal, while the spectra of the bottom row 2) were obtained with sea state 5 where these errors are maximal. However, the subjective errors (by visual inspection) seem better than the objective ones. (Color figure online)

4.2 Real Data

To our knowledge, the only publicly available dataset suitable for our needs has been published by Guimarães et al. [8]. This dataset is composed of several stereo-vision sequences across different stations. Unfortunately, only the 4 sequences taken from the Black Sea station are given with wave measurements, from a wave gauge cage array composed of 5 resistance wave recorders arranged in a star pattern [17]. This sensor array allows for a reconstruction of the directional spectrum, using the Extended Wavelet Directional Method [14]. This will serve as a reference for a comparison with our method. It is however sampled at 10 Hz while the videos are sampled at 12 Hz or 15 Hz. To compare the two

different sources, the same number of sample was taken for each iteration of the method starting from the same timestamp. The end timestamp of the 2 data streams is therefore different. The wave data is also measured a few meters away from the region of interest of the camera. However, it is assumed that the wave information stays relatively constant in space and time.

In total, about 68000 images are available for our evaluation. These images are taken from 4 different sequences and exhibit a similar rather calm sea state. The characteristics of the sequences are given in Table 2.

Table 2. BS2011 sequences characteristics. The direction is given in octants, with W for West (270°) and S for South (180°), as on a compass rose.

Time of capture	fps	Duration	f_p	H_s	θ_p
2011-10-01 16:18	15 Hz	12 min	0.33 Hz	0.30 m	WSW
2011-10-04 11:38	12 Hz	20 min	0.38 Hz	0.36 m	WSW
2011-10-04 13:07	12 Hz	30 min	0.33 Hz	0.45 m	SW
2011-10-04 15:30	12 Hz	30 min	0.27 Hz	0.55 m	SW

Table 3. Estimation errors.

Error type	Sequence	Average	Standard deviation
Frequency errors	sim state 3	0.001 Hz	0.087 Hz
	sim state 4	0.097 Hz	0.042 Hz
	sim state 5	0.045 Hz	0.024 Hz
	sim state 6	0.079 Hz	0.043 Hz
	16-18	−0.002 Hz	0.106 Hz
	11-38	−0.091 Hz	0.100 Hz
	13–07	−0.107 Hz	0.071 Hz
	15–30	−0.129 Hz	0.115 Hz
Direction errors	sim state 3	10.4°	34.4°
	sim state 4	−1.4°	32.3°
	sim state 5	0.8°	30.0°
	sim state 6	−3.2°	36.1°
	16–18	28.1°	30.7°
	11–38	14.5°	28.5°
	13–07	10.7°	20.1°
	15–30	17.3°	37.8°

The method was run on 1024 images in a circular buffer meaning that at each iteration the oldest frame was removed to make space for the next image

in the video file. Each image used was resized to 256×256. This combination was chosen as a trade-off between precision and run time.

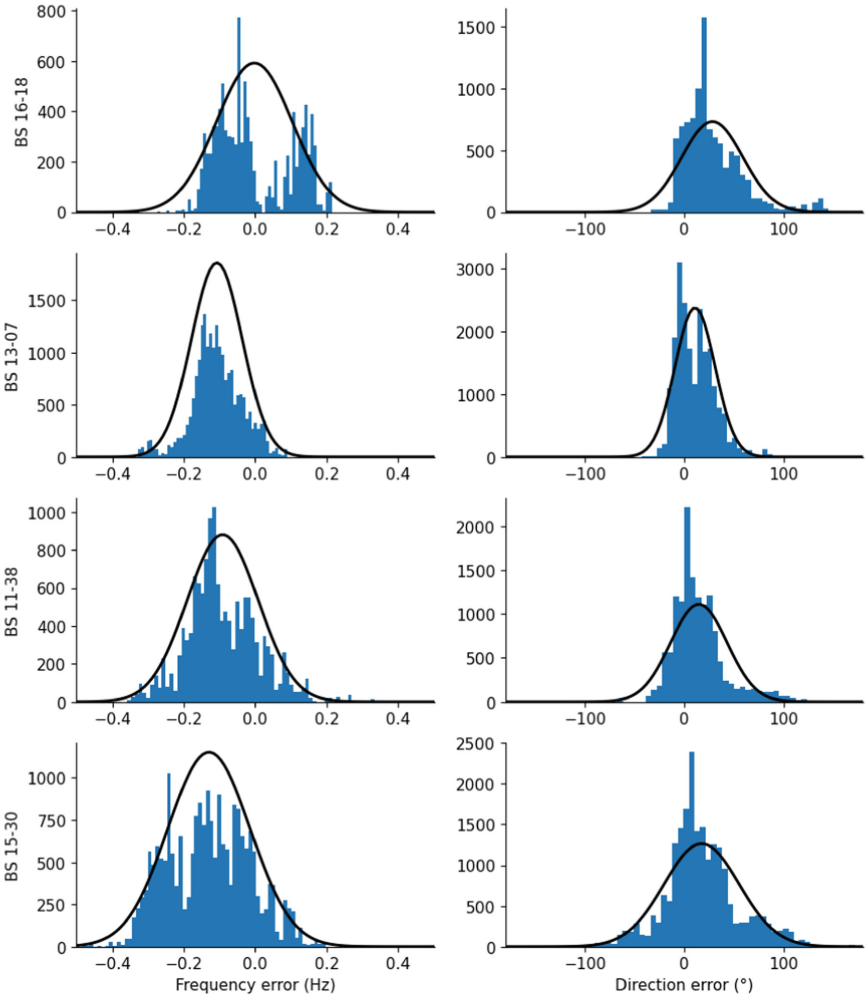


Fig. 10. Black Sea sequences error distribution.

Figure 10 shows the distribution error for both the frequency and direction errors. For sequences "13-07", "11-38" and "15-30", the error is relatively constant across the sequence, the standard deviation is no more than 0.1 Hz. However, the mean value is around -0.1 Hz, showing that the estimation is biased toward lower frequencies.

As for the direction error distribution, all sequences show a standard deviation of about 20° and an average error of about 10°. The bias toward the

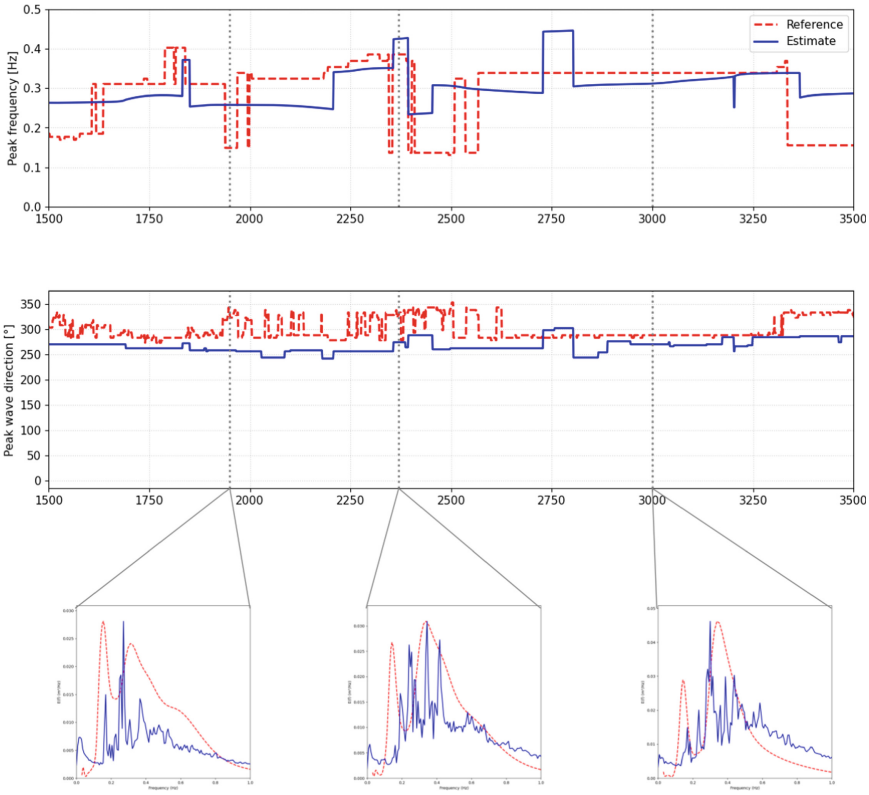


Fig. 11. Sequence BS 16-18. Above: time series of peak wave frequency and direction estimations. Estimates are shown in blue and their references in red. Below: directional wave spectra. (Color figure online)

apparent South is probably to be attributed to a misconfiguration of the camera pose as this error is not present in the simulated sequences where the camera pose is exactly known.

The frequency error for the "BS 16-18" sequence exhibits 2 distinct peaks. This can be explained by the nature of the sea state of this sequence as the ocean is composed of a wind driven wave component and a swell driven wave component. The estimation sometimes favors one or the other, depending on which is appearing the strongest. Fig. 11 shows this phenomenon.

4.3 Computational Time and Space

The computational complexity of this approach is dominated by the 3D FFT. The time complexity of the n-dimension FFT is $O(N^n \log(N))$. The N-dimension FFT is equivalent to N times (N-1)-dimension FFT. In our case, $N = N_x \cdot N_y \cdot N_t$.

Experiments were run on a laptop with an Intel Core i-9 12900H with 32 GiB of RAM. Using 1024 images of size 256 $\times$ 256, resulting in $N = 256 \times$

$256 \times 1024 = 2^{26}$. A single iteration of the method using Python 3.12 is taking on average 2.9 s and 2.4 GiB of memory.

5 Conclusion

The presented work proposes a vision-based method for estimating peak wave period and direction from monoscopic video sequences using a 3D Fourier analysis. The method was validated on synthetic data and a few sequences from a public dataset, albeit with some bias.

As the method relies mainly on the FFT, hardware acceleration on devices such an FPGA is straightforward, enabling the algorithm to run in real-time, possibly in an embedded system onboard a ship. The results highlight the potential of monoscopic video as a low-cost, energy-efficient alternative for onboard sea state estimation.

Future work include integration of other sensors that are present in a ship such as wind sensor, inertial navigation systems or water speed sensor. The resulting estimation system will also be embedded in a sailing boat to inform sailors about the sea state.

Acknowledgments. This work benefited from access to the computing resources of the L3i laboratory, operated and hosted by the University of La Rochelle. It is funded by the French government and the Region Nouvelle-Aquitaine.

References

1. Barrick, D.: Remote sensing of sea state by radar. In: Ocean 72-IEEE International Conference on Engineering in the Ocean Environment, pp. 186–192. IEEE (1972)
2. Benetazzo, A., Barbariol, F., Bergamasco, F., Torsello, A., Carniel, S., Sclavo, M.: Observation of extreme sea waves in a space–time ensemble. J. Phys. Oceanogr. **45**(9), 2261–2275 (2015). https://doi.org/10.1175/JPO-D-15-0017.1
3. Bradski, G.: The OpenCV library. Dr. Dobb's J. Softw. Tools (2000)
4. Brodtkorb, A., Nielsen, U., Sørensen, A.: Online wave estimation using vessel motion measurements. IFAC-PapersOnLine **51**, 244–249 (2018). https://doi.org/10.1016/j.ifacol.2018.09.510
5. Cartwright, D.E.: The Use of Directional Spectra in Studying the Output of a Wave Recorder on a Moving Ship. Prentice Ball (1963)
6. Community, B.O.: Blender - a 3D modelling and rendering package. Blender Foundation, Stichting Blender Foundation, Amsterdam (2018). http://www.blender.org
7. Greenberg, O., Ben-Moshe, B.: Real-time stereo-based ocean surface mapping for robotic floating platforms: concept and methodology. Sensors **23**(8), 3857 (2023). https://doi.org/10.3390/s23083857. publisher: MDPI AG
8. Guimarães, P.V., et al.: A data set of sea surface stereo images to resolve space-time wave fields. Sci. Data **7**(1) (2020). https://doi.org/10.1038/s41597-020-0492-9, https://sextant.ifremer.fr/Donnees/Catalogue#/metadata/af599f42-2770-4d6d-8209-13f40e2c292f

9. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press (2004). ISBN 0521540518
10. Hasselmann, K., et al.: Measurements of wind-wave growth and swell decay during the joint north sea wave project (jonswap). Ergaenzungsheft zur Deutschen Hydrographischen Zeitschrift Reihe A **8**, 1–95 (1973)
11. Loizou, A., Christmas, J.: Sea state estimation from uncalibrated, monoscopic video. SN Comput. Sci. **2**(4), 328 (2021). https://doi.org/10.1007/s42979-021-00727-0
12. Mitsuyasu, H., et al.: Observations of the directional spectrum of ocean waves using a cloverleaf buoy. J. Phys. Oceanogr. **5**(4), 750–760 (1975)
13. Paris, N., Marchand, S., Gomez-Krämer, P.: L3I ocean simulation dataset (2026). https://l3i-share.univ-lr.fr/2026OceanSimulationDataset/. Accessed 26 Apr 2026
14. Peláez-Zapata, D., Dias, F.: EWDM: a wavelet-based method for estimating directional spectra of ocean waves. J. Open Sour. Softw. **10**(109), 7942 (2025). https://doi.org/10.21105/joss.07942
15. Phillips, O.M.: On the generation of waves by turbulent wind. J. Fluid Mech. **2**(5), 417–445 (1957). https://doi.org/10.1017/S0022112057000233
16. Pierson Jr., W.J., Moskowitz, L.: A proposed spectral form for fully developed wind seas based on the similarity theory of S. A. Kitaigorodskii. J. Geophys. Res. (1896-1977) **69**(24), 5181–5190 (1964)
17. Shokurov, M.V., Dulov, V.A., Skiba, E.V., Smolov, V.E.: Wind waves in the coastal zone of the Southern Crimea: assessment of simulation quality based on in situ measurements. Oceanology **56**(2), 214–225 (2016). https://doi.org/10.1134/S0001437016020181
18. Tessendorf, J.: Simulating ocean water. SIG-GRAPH'99 Course Note (2001)
19. Ulaby, F.T., Moore, R.K., Fung, A.K.: Microwave Remote Sensing: Active and Passive. Volume 1: Microwave Remote Sensing Fundamentals and Radiometry. Addison-Wesley, Reading (1981)