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

Nicolas Paris Sylvain Marchand Petra Gomez-Krämer

L3i, La Rochelle University

Wednesday 19th August, 2026

Context: SeaLab, a semi autonomous sailboat

Why an "intelligent" sailboat?

- Wind powered: no carbon emissions
- Helping the crew make decisions

Context: SeaLab, a semi autonomous sailboat

Why an "intelligent" sailboat?

- Wind powered: no carbon emissions
- Helping the crew make decisions

Giving SeaLab "intelligence"

- regular marine sensors: anemometer, compass, GNSS...
- **Single consumer camera** fixed on mast, filming the sea surface at an oblique angle
- multiple inertial sensor units, computing power, blackbox capabilities...
- Should stay cheap and low power

Motivation

Estimating the sea state onboard

- Sea state affects vessel performance, safety and routing
- Sea state: agitation of the sea
- Key parameters: significant wave height H_s, peak frequency f_p, direction θ_p

Motivation

Estimating the sea state onboard

- Sea state affects vessel performance, safety and routing
- Sea state: agitation of the sea
- Key parameters: significant wave height H_s, peak frequency f_p, direction θ_p

Traditional sensors are limiting:

- Wave buoys: costly, not onboard
- Radar systems: energy-intensive
- Satellite SAR: not real-time

Motivation

Estimating the sea state onboard

- Sea state affects vessel performance, safety and routing
- Sea state: agitation of the sea
- Key parameters: significant wave height H_s, peak frequency f_p, direction θ_p

Traditional sensors are limiting:

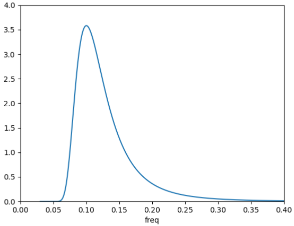
- Wave buoys: costly, not onboard
- Radar systems: energy-intensive
- Satellite SAR: not real-time

Our goal

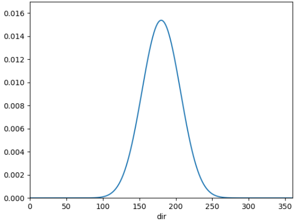
Estimate wave parameters from a **single onboard video camera**.

Ocean Wave Modeling: Wave spectrum

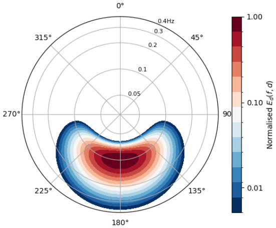
The wave field is described by the **directional spectrum** $S(f,\theta)$.



$S(f)$

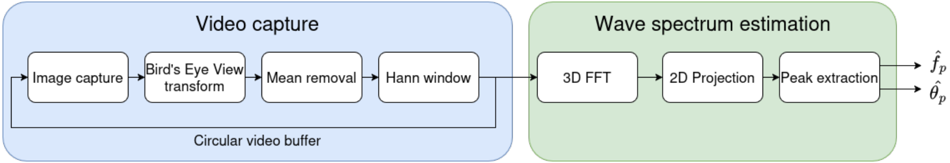


$G(\theta)$



$S(f,\theta)$: f_p=0.1 Hz, θ_p=180°

Method Overview



1. **Video stabilization** (not required for static cameras)
2. **Bird's Eye View (BEV) transform** to remove perspective distortion
3. **Preprocessing**: DC removal + 3D Hann windowing
4. **3D FFT**: spectral analysis over (x, y, t)
5. **Wave spectrum extraction**: convert to $S(f, \theta)$
6. **Peak extraction**: estimate $\hat{f}_p$ and $\hat{\theta}_p$

Step 2: Bird's Eye View Transform

Goal: suppress perspective effects so that wave frequencies are uniform across the image.

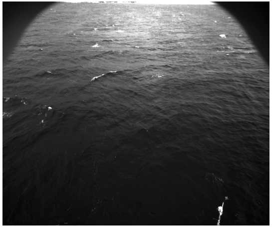
Uses the **pinhole camera model**:

$$u = KMX_w = PX_w$$

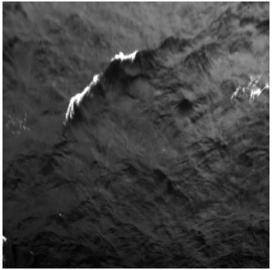
Inverse mapping onto the $z = 0$ plane:

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

Homography $H = P_2L(P_1L)^{-1}$ maps pixels from camera view to top-down view.



Original frame



After BEV transform

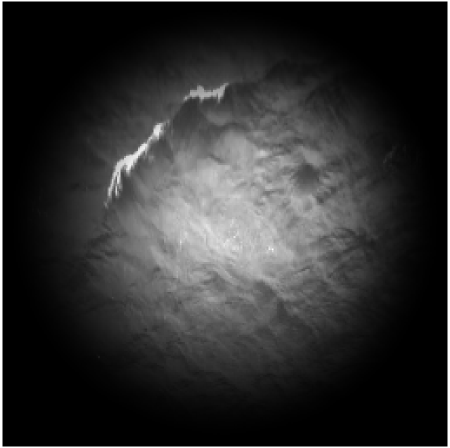
Step 3 & 4: Preprocessing and 3D FFT

Preprocessing of the video cube $z(x, y, t)$:

- **DC removal**: subtract mean pixel value across frames to enhance low-frequency precision
- **3D Hann window**: reduce spectral leakage, sharpen frequency peaks

3D Discrete Fourier Transform:

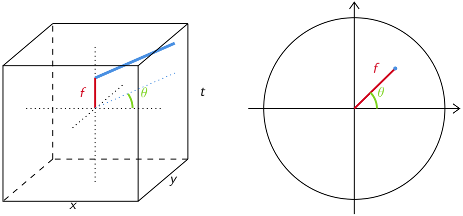
$$\mathcal{Z}(k_x, k_y, \omega) = \sum_x \sum_y \sum_t z(x, y, t)\, e^{-i(k_x x + k_y y + \omega t)}$$



After DC removal and windowing

Steps 5 & 6: Spectrum Extraction and Peak Detection

From 3D Fourier space to directional spectrum:

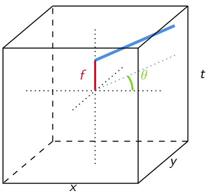


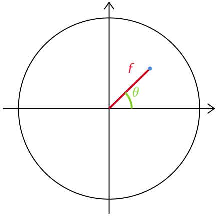
$$\sum_{i=0}^{N/2} a\left(\lfloor i \sin \theta \rfloor + \frac{N_x}{2},\ \lfloor i \cos \theta \rfloor + \frac{N_y}{2},\ f + \frac{N_t}{2}\right)$$

Cylindrical projection of the 3D amplitude cube. Wave-length information is traded for a compact (f, θ) representation.

Steps 5 & 6: Spectrum Extraction and Peak Detection

From 3D Fourier space to directional spectrum:



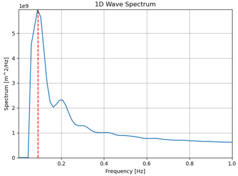


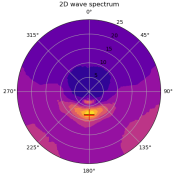
$$\sum_{i=0}^{N/2} a\left(\lfloor i \sin\theta \rfloor + \frac{N_x}{2},\ \lfloor i \cos\theta \rfloor + \frac{N_y}{2},\ f + \frac{N_t}{2}\right)$$

Cylindrical projection of the 3D amplitude cube. Wave-length information is traded for a compact (f, θ) representation.

Peak extraction:

- Peak frequency: $\hat{f}_p = \arg\max_f S(f)$, where $S(f) = \int S(f, \theta)\, d\theta$
- Peak period: $\hat{T}_p = 1/\hat{f}_p$
- Peak direction: $\hat{\theta}_p = \arg\max_\theta S(\hat{f}_p, \theta)$





1D and 2D estimated spectra

Evaluation: Synthetic Dataset

Blender simulations with Tessendorf ocean algorithm and JONSWAP spectrum.

- 684 sequences total (171 per sea state)
- 600 frames at 30 fps (20s each)
- Random peak direction $\theta_p \in [0^\circ, 360^\circ]$
- Random sun position: varying lighting

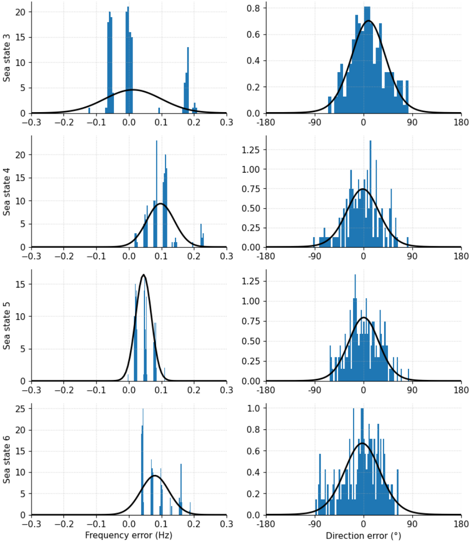
Sea state	f_p range	H_s range
3	0.16-0.18 Hz	0.65-0.80 m
4	0.11-0.13 Hz	1.35-1.65 m
5	0.09-0.10 Hz	2.25-2.75 m
6	0.07-0.08 Hz	3.50-4.50 m

Dataset publicly available:

https://l3i-share.univ-lr.fr/2026OceanSimulationDataset

Simulation examples

Results: Synthetic Data



Error distributions by sea state

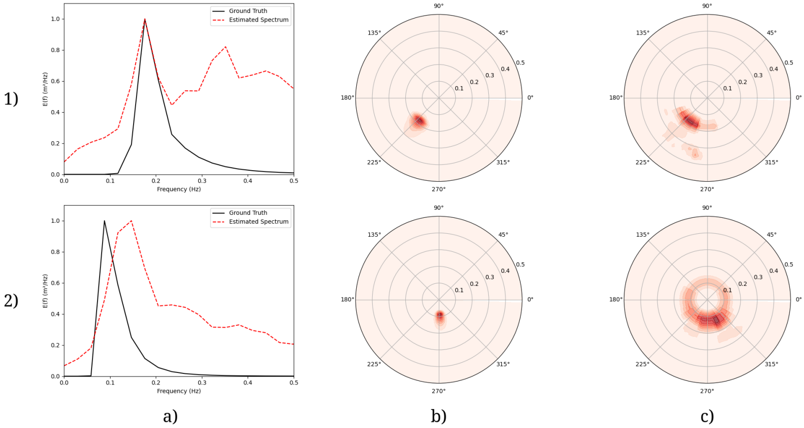
Direction estimation:

- Mean error $\approx 0^\circ$
- Std. deviation $\approx 30^\circ$

Frequency estimation:

- Bias $\approx 0.05\,\text{Hz}$
- Std. deviation $\approx 0.06\,\text{Hz}$

Results: Synthetic Spectra Examples



Row 1: sea state 3, minimal estimation errors.

Row 2: sea state 5, maximal objective errors, yet visually plausible.

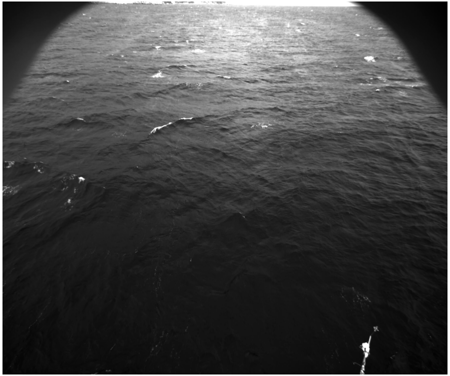
(a) 1D spectrum with ground truth (black) and estimate (red); (b) GT directional spectrum; (c) estimated directional spectrum.

Evaluation Real Dataset (Black Sea 2011)

Dataset: Guimarães et al. (2020) stereo-vision sequences with wave gauge reference (5-sensor star array).

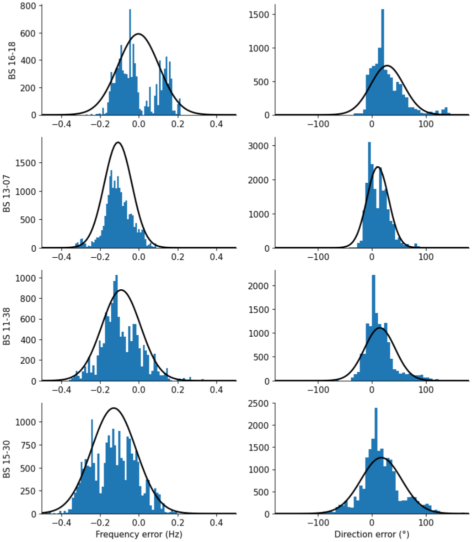
Time	fps	f_p	θ_p
16:18	15	0.33 Hz	WSW
11:38	12	0.38 Hz	WSW
13:07	12	0.33 Hz	SW
15:30	12	0.27 Hz	SW

$\sim$68 000 images, calm sea state ($H_s = 0.30$-0.55 m).



Example frame from real sequence

Results: Real Data



Error distributions, Black Sea sequences

Direction:

- Bias ≈ 15° toward South
- Std. deviation ≈ 20°

Frequency:

- Bias ≈ −0.1 Hz
- Std. ≈ 0.1 Hz
- Sequence 16:18 shows bimodal error: wind sea + swell

Computational Complexity

Dominant cost: 3D FFT

$$\mathcal{O}(N^3 \log N), \quad N = N_x \cdot N_y \cdot N_t$$

Experimental setup:

- Intel Core i9-12900H, 32 GiB RAM
- 1024 images at 256×256 px
- $N = 256 \times 256 \times 1024 = 2^{26}$

Per-iteration cost

2.9 s wall-clock time
2.4 GiB memory

Conclusion & Future Work

Summary

- 3D FFT on stabilised BEV video estimates $\hat{f}_p$ and $\hat{\theta}_p$
- Validated on synthetic and real dataset
- Direction: consistent accuracy
- Frequency: sufficient precision but systematic bias
- Runs in $\approx 3\,\mathrm{s}$ per window. Compatible with near-real-time use

Future work

- Fusion with onboard sensors: IMU, wind sensor, water speed
- FPGA deployment for real-time embedded processing
- **Real-world dataset** from a camera on a sailing boat mast

Thank you!