A Variational Stereo Method for the 3-D Reconstruction of Ocean Waves

Guillermo Gallego, Anthony Yezzi, Francesco Fedele and Alvise Benetazzo

Abstract

We develop a novel remote sensing technique for the observation of waves on the ocean surface. Our method infers the three-dimensional wave form and radiance of oceanic sea states via a variational stereo imagery formulation. In this setting, the shape and radiance of the wave surface are given by minimizers of a composite energy functional that combines a photometric matching term along with regularization terms involving the smoothness of the unknowns. The desired ocean surface shape and radiance are the solution of a system of coupled partial differential equations derived from the optimality conditions of the energy functional. The proposed method is naturally extended to study the spatio-temporal dynamics of ocean waves and applied to three sets of stereo video data. Statistical and spectral analysis are carried out. Our results provide evidence that the observed omni-directional wavenumber spectrum $S(k)$ decays as $k^{-2.5}$ in agreement with Zakharov's theory (1999). Further, the three-dimensional spectrum of the reconstructed wave surface is exploited to estimate wave dispersion and currents.

Index Terms

Remote Sensing, marine technology, variational methods, stereo vision, image processing.

I. Introduction

Wind-generated waves play a prominent role at the interfaces of the ocean with the atmosphere, land and solid Earth. Waves also define in many ways the appearance of the ocean seen by remote-sensing instruments. Classical observational methods rely on time series retrieved from wave gauges and ultrasonic instruments or buoys to measure the space-time dynamics of ocean waves. Global altimeters, or Synthetic Aperture Radar (SAR) instruments are exploited for observations of large oceanic areas via satellites [24], [17], but details on small scales are lost. To complement the abovementioned instruments, this work develops a novel video observational system that relies on variational stereo techniques to reconstruct the 3-D wave surface both in space and time. The front end of the system consists of two or more camera views pointing at the ocean and providing space-time data whose statistical

G. Gallego and A. Yezzi are with the Department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332. E-mail: ggb@gatech.edu, ayezzi@ece.gatech.edu

F. Fedele is with the Department of Civil and Environmental Engineering, Georgia Institute of Technology, Savannah, Georgia 31407. E-mail: ffedele3@gtsav.gatech.edu

A. Benetazzo is with Protecno S.r.l, Via Risorgimento, Padova 35027, Italy. E-mail: alvise.benetazzo@gmail.com

content is richer than that of previous monitoring methods. Vision systems are non-intrusive and have economical advantages over their predecessors, but they require more processing power to infer information from the ocean.

This work relates to a vast body of literature because it covers both the topics of shape reconstruction and oceanic sea states. The three-dimensional reconstruction of an object's surface from stereo pairs of images is a classical problem in computer vision (see, for example [19], [13], [18], [26]), and it is still an extremely active research area. There exist many 3-D reconstruction algorithms in the literature and the reconstruction problem is far from being solved. The algorithms are designed under different assumptions and provide a variety of trade-offs between speed, accuracy and viability. Traditional *image-based* stereo methods typically consist of two steps: first image points are detected and matched across images by optimizing a photometric score to establish local correspondences; then depth is inferred by combining these correspondences using *triangulation* of 3-D points (*back-projection* of image points). The first step is significantly more difficult than the second one, but epipolar geometry between image pairs can be exploited to reduce stereo matching to a 1-D search along epipolar lines, as shown in recent systems [2], [32]. This approach is simple and fast, but it also has some major disadvantages that motivated the research on improved stereo reconstruction methods [7], [33], [16]. These disadvantages are: (*i*) Correspondences rely on strong textures (high contrast between intensities of neighboring points) and image matching gives poor correspondences if the objects in the scene have a smooth radiance. Correspondences also suffer from the presence of noise and local minima. (*ii*) Each space point is reconstructed independently and therefore the recovered surface of an object is obtained as a collection of scattered 3-D points. Thus, the hypothesis of the continuity of the surface is not exploited in the reconstruction process. The breakdown of traditional stereo methods in these situations is evidenced by "holes" in the reconstructed surface, which correspond to unmatched image regions [19], [2]. This phenomenon may be dominant in the case of the ocean surface, which, by nature, is generally continuous and contains little texture.

Modern *object-based* image processing and computer vision methods that rely on Calculus of Variations and Partial Differential Equations (PDE), such as Stereoscopic Segmentation [33] and other variational stereo methods [7], [1], [16], are able to overcome the disadvantages of traditional stereo. For instance, unmatched regions are avoided by building an explicit model of the smooth surface to be estimated rather than representing it as a collection of scattered 3-D points. Thus, variational methods provide dense and coherent surface reconstructions. Surface points are reconstructed by exploiting the continuity (coherence) hypothesis in the full two-dimensional domain of the surface. Variational stereo methods combine correspondence establishment and shape reconstruction into one single step and they are less sensitive to matching problems of local correspondences. The reconstructed surface is obtained by minimization of an energy functional designed for the stereo problem. The solution is obtained in the context of active surfaces by deforming an initial surface via a gradient descent PDE derived from the necessary optimality conditions of the energy functional, the so-called Euler-Lagrange (EL) equations.

In the context of oceanography, the first experiments with stereo cameras mounted on a ship were by Schumacher [25] in 1939. Later, Coté et al. [5] in 1960 demonstrated the use of stereo-photography to measure the sea topography for long ocean waves. Stereography gained popularity in studying the dynamics of oceanographic

phenomena during the 1980s due to advances progress in hardware. Shemdin et al. [28], [27] applied stereography for the directional measurement of short ocean waves. A more recent integration of stereographic techniques into the field of oceanography has been the WAVESCAN project of Santel et al. [23]. Recently, Benetazzo [2] successfully incorporated epipolar techniques in the Wave Acquisition Stereo System (WASS) and showed that the accuracy of WASS is comparable to the accuracy obtained from ultrasonic transducer measurements. Fig. 1 shows an example of a WASS system currently installed at the Acqua Alta platform that has been used to study space-time waves and spectra in the Northern Adriatic Sea [8]. An alternative trinocular imaging system (ATSIS) for measuring the temporal evolution of 3-D surface waves was proposed in [32]. More recently, in [11] it is shown how a modern variational stereo reconstruction technique pioneered by [7] can be applied to the estimation of oceanic sea states. References [20], [31], [12], [15] show that this is an active research topic.

Encouraged by the results in [2], [11], [9], in this paper we develop a novel variational framework for the recovery of the shape and radiance of ocean waves given stereo images acquired by calibrated cameras. In particular, motivated by the characteristics of the target object in the scene, i.e. the ocean surface, we first introduce the graph surface representation in the formulation of the reconstruction problem. Then, we present the new image processing algorithm in the context of PDEs and active surfaces. We validate the performance of the algorithm on experimental data and analyze the statistics of the reconstructed surface. Concluding remarks and future research directions are finally presented.

II. The variational geometric method

This paper is inspired by the works of [2], [11] and [33]. In particular, the variational approach of *stereoscopic segmentation* [33] is used to address the problem: the reconstructed surface of the ocean is obtained as the minimizer of an energy functional designed to fit the measurements of the ocean. In every 3-D reconstruction method, the quality and accuracy of the results depend on the calibration of the cameras. There are standard camera calibration procedures in the literature to characterize accurately the intrinsic and extrinsic parameters of the cameras [19]. We assume cameras are calibrated and synchronized, and we focus on the reconstruction of the water surface for a fixed time.

A. Graph surface representation

We consider S to be a smooth surface in $\mathbb{R}^3$ with generic local coordinates $(u, v) \in \mathbb{R}^2$. The geometry of the image formation process, which states how points in 3-D are mapped into points on the image plane, is described by the pinhole camera model [13]. Let $\{I_i\}_{i=1}^{N_c}$ be a set of images of a static (water) scene acquired by cameras whose calibration parameters are $\{\mathtt{P}^i\}_{i=1}^{N_c}$. Projective geometry in homogeneous coordinates provides a convenient framework to express such a projection mapping due to the linearity of the equations. A surface point (or, in general a 3D point) $\mathbf{X} = (X, Y, Z)^\top$ with homogeneous coordinates $\bar{\mathbf{X}} = (X, Y, Z, 1)^\top$ is mapped to point $\mathbf{x}_i = (x_i, y_i)^\top$ in the i-th image with homogeneous coordinates $\bar{\mathbf{x}}_i = (x_i, y_i, 1)^\top \sim \mathtt{P}^i \bar{\mathbf{X}}$, where the symbol $\sim$ means equality up to a nonzero scale factor and $\mathtt{P}^i = \mathtt{K}^i[\mathtt{R}^i \mid \mathbf{t}^i]$ is the 3×4 projection matrix with the intrinsic ($\mathtt{K}^i$) and extrinsic

$(\mathtt{R}^i, \mathbf{t}^i)$ calibration parameters of the i-th camera. These parameters are known under the hypothesis of calibrated cameras. The optical center of the camera is the point $\mathbf{C}_i = (C_i^1, C_i^2, C_i^3)^\top$ satisfying $\mathtt{P}^i \bar{\mathbf{C}}_i = \mathbf{0}$. Let $\pi_i : \mathbb{R}^3 \to \mathbb{R}^2$ note the perspective projection maps, $\mathbf{x}_i = \boldsymbol{\pi}_i(\mathbf{X})$, and $I_i(\mathbf{x}_i) \equiv I_i(\pi_i(\mathbf{X}))$ be the image intensity at $\mathbf{x}_i$.

We present a different approach to the reconstruction problem discussed in [33], [7], by exploiting the hypothesis that the surface of the water can be represented in the form of a graph or elevation map:

$$Z = Z(X, Y), \tag{1}$$

where Z is the height of the surface with respect to a domain plane that is parameterized by coordinates X and Y. Indeed, slow varying, non-breaking waves admit this simple representation with respect to a plane orthogonal to gravity direction. As a natural extension of previous methods, energy functionals can be tailored to exploit the benefits of this valuable representation. The surface can still be obtained as the minimizer of a suitable energy functional but now with a different geometrical representation of the solution.

The graph representation of the water surface presents some clear advantages over the more general level set representation in [11]. Surface evolution is simpler to implement since the surface is not represented in terms of an auxiliary higher dimensional function (the level set function). The surface is evolved directly via the height function (1) discretized over a fixed 2-D grid defined on the $X-Y$ plane. The latter also implies that for the same amount of physical memory, higher spatial resolution (finer details) can be achieved in the graph representation than with the level set. The $X-Y$ plane becomes the natural common domain to parameterize the geometrical and photometric properties of surfaces. This simple identification does not exist in the level set approach [33]. Finally, the graph representation allows for fast numerical solvers besides gradient descent, like Fast Poisson Solvers, Cyclic Reduction, Multigrid Methods, Finite-Element Methods (FEM), etc. In the level set framework, the range of solvers is not as diverse.

However, there are also some minor disadvantages. A world frame properly oriented with the gravity direction must be defined in advance to represent the surface as a graph with respect to this plane. This is not trivial *a priori* and might pose a problem if only the information from the stereo images is used [2]. This condition may not be so if external gravity sensors provide this information. It is also possible to choose an initial estimate for the plane and then update it with some feedback from the statistics of the reconstructed waves in time. Surface evolution is constrained to be in the form of a graph and this may not be the same as the evolution described for an unconstrained surface. As a result, more iterations may be required to reach convergence.

The reconstruction problem is mathematically stated in the following section. The desired surface is given by the solution of a variational optimization problem.

B. Proposed vision-based energy functional.

A generative model of the images consisting of the joint estimation of the shape of the surface S and the radiance function on the surface f has been investigated. Consider the 3-D reconstruction problem from a collection of $N_c \geq 2$ input images (most of the time we will exemplify with $N_c = 2$). Let the energy functional be the weighted sum of

a data fidelity term E_{data} and two regularizing terms: a geometry smoothing term E_{geom} and a radiance smoothing term E_{rad}

$$E(S, f) = E_{\text{data}}(S, f) + \alpha E_{\text{geom}}(S) + \beta E_{\text{rad}}(f), \tag{2}$$

where $\alpha, \beta \in \mathbb{R}^+$. The data fidelity term is designed to measure the photo-consistency of the model: the discrepancy in the L^2 sense between the observed images I_i and the radiance model f,

$$E_{\text{data}} = \sum_{i=1}^{N_c} E_i, \qquad E_i = \int_{\Omega_i} \phi_i \, \mathrm{d}\mathbf{x}_i, \tag{3}$$

where the photometric matching criterion is

$$\phi_i = \tfrac{1}{2}\big(I_i(\mathbf{x}_i) - f(\mathbf{x}_i)\big)^2. \tag{4}$$

The region of the image domain where the scene is projected is denoted by Ω_i. Assuming that the surface of the scene (water) can be represented in the form of a graph $Z = Z(u, v)$, a point on the surface has coordinates

$$\mathbf{X}(u, v) = \big(u, v, Z(u, v)\big)^\top. \tag{5}$$

The chain of operations to obtain the intensity $I_i(\mathbf{x}_i)$ given a point $\mathbf{u} = (u, v)^\top$ in the parameter space of the surface is

$$\mathbf{u} \mapsto \mathbf{X}(\mathbf{u}) \mapsto \tilde{\mathbf{X}}_i = \mathtt{M}^i \mathbf{X} + \mathbf{p}_4^i \mapsto \mathbf{x}_i \mapsto I_i(\mathbf{x}_i), \tag{6}$$

where $\mathbf{X}(\mathbf{u}) \equiv S(\mathbf{u})$ are the world coordinates of a surface point, $\tilde{\mathbf{X}}^i = (\tilde{X}_i, \tilde{Y}_i, \tilde{Z}_i)^\top$ are related to the coordinates of the surface point $\mathbf{X}$ in the i-th camera frame, $\mathbf{x}_i = (x_i, y_i)^\top = (\tilde{X}_i/\tilde{Z}_i, \tilde{Y}_i/\tilde{Z}_i)^\top$ are the coordinates of the projection of $\mathbf{X}$ in the i-th image plane and $\mathtt{P}^i = [\mathtt{M}^i \,|\, \mathbf{p}_4^i]$, with $\mathtt{M}^i = \mathtt{K}^i\mathtt{R}^i$ and $\mathbf{p}_4^i = \mathtt{K}^i\mathbf{t}^i$. Also, $|\mathtt{M}^i| = \det(\mathtt{M}^i)$.

The radiance model f is specified by a function $\hat{f}$ defined on the surface S. Moreover, we consider its extension to the whole embedding space $\hat{f} : \mathbb{R}^3 \to \mathbb{R}$. There are many possible ways to define this extension; we will consider one that simplifies the equations of the model. Then, f in (4) is naturally defined by $f(\mathbf{x}_i) = \hat{f}(\pi_i^{-1}(\mathbf{X}))$, where π_i^{-1} denotes the back-projection operation from a point in the i-th image to the closest surface point with respect to the camera. With a slight abuse of notation, let us use f to denote the parameterized radiance $f(\mathbf{u})$, understanding that $f(\mathbf{x}_i)$ in (4) reads the back-projected value in $\hat{f}(\mathbf{X}(\mathbf{u})) = f(\mathbf{u})$.

Motivated by the common parameterizing domain of the shape Z and radiance f of the surface and to obtain the simplest diffusive terms in the PDEs derived from the necessary optimality conditions of the energy (2), let the regularizers be

$$E_{\text{geom}} = \int_U \tfrac{1}{2}\|\nabla Z(\mathbf{u})\|^2 \, \mathrm{d}\mathbf{u}, \tag{7}$$

$$E_{\text{rad}} = \int_U \tfrac{1}{2}\|\nabla f(\mathbf{u})\|^2 \, \mathrm{d}\mathbf{u}. \tag{8}$$

Now that all terms in (2) have been specified, some transformations are carried out to express the integrals over a more suitable domain. Integrals in (7) and (8) are already in a convenient domain, the parameter space. The data

fidelity term (3) can be expressed as an integral over the parameter space by means of a change of variables. Let the Jacobian of the change of variables be (see appendix A-A)

$$\mathtt{J}_i = \left| \frac{\mathrm{d}\mathbf{x}_i}{\mathrm{d}\mathbf{u}} \right| = -|\mathtt{M}^i| \tilde{Z}_i^{-3} (\mathbf{X} - \mathbf{C}_i) \cdot (\mathbf{X}_u \times \mathbf{X}_v). \tag{9}$$

Then, the data fidelity energy (3) becomes

$$E_i = \int_{\Omega_i} \phi_i \, d\mathbf{x}_i = \int_U \phi_i \mathtt{J}_i \, d\mathbf{u}, \tag{10}$$

where the last integral is over U: the part of the parameter space whose surface projects on Ω_i in the i-th image.

Furthermore, the data fidelity term can be expressed as a surface integral, according to the relationship between area measures (38) (see appendix A-B), with

$$E_i = -\int_S \phi_i(\mathbf{X}, f) |\mathtt{M}^i| \tilde{Z}_i^{-3} (\mathbf{X} - \mathbf{C}_i) \cdot \mathbf{N} \, \mathrm{d}A. \tag{11}$$

A visibility term (in the form of a characteristic function) that indicates what part of S should be integrated according to what part of S is visible from the i-th camera must also be included in the integrand of (11), but it has been omitted for the sake of clarity.

After collecting terms (7), (8), and (10), energy (2) is

$$E(\mathbf{X}, f) = \int_U L(\mathbf{X}, \mathbf{X}_u, \mathbf{X}_v, f, f_u, f_v, u, v) \, \mathrm{d}\mathbf{u}, \tag{12}$$

where subscripts indicate the derivative with respect to that variable, and the integrand, the so-called *Lagrangian*, is

$$L = L_{\text{data}} + \alpha L_{\text{geom}} + \beta L_{\text{rad}}, \tag{13}$$

with L_{data}, L_i, L_{geom} and L_{rad} being the Lagrangians for E_{data}, E_i, E_{geom} and E_{rad}, respectively.

C. Energy minimization. Optimality condition.

The energy (12) depends on two functions: the shape $\mathbf{X}$ and the radiance f of the surface. To find a minimizer of such a functional, we derive the necessary optimality condition by setting to zero the first variation of the functional. Before that, let us enforce the constraint on the representation of the surface as a graph (5). Fixing the parameterization $\mathbf{u}$, the energy functional (12) becomes a function of the height $Z(\mathbf{u})$:

$$E(Z, f) = \int_U L(Z, Z_u, Z_v, f, f_u, f_v, u, v) \, \mathrm{d}\mathbf{u}. \tag{14}$$

The first variation (Gâteaux derivative) of (14) is

$$\frac{\delta E}{\delta (Z, f)}(h, g) = \left. \frac{\mathrm{d}E\big((Z, f) + \epsilon(h, g)\big)}{\mathrm{d}\epsilon} \right|_{\epsilon=0}, \tag{15}$$

where h and g are arbitrary perturbations for Z and f, respectively.

Using the chain rule and integration by parts in (15), it follows that the first variation (16) has two terms: one in the interior of the integration region U in the parameter space and one boundary term (on ∂U), where $\boldsymbol{\nu} = (\nu^u, \nu^v)^\top$ represents the outward unit normal to U:

$$\frac{\delta E}{\delta(Z,f)}(h,g) = \frac{\delta E}{\delta Z}(h) + \frac{\delta E}{\delta f}(g), \tag{16}$$

where $\frac{\delta E}{\delta q}(w)$ is

$$\int_U \big(L_q - (L_{q_u})_u - (L_{q_v})_v\big) w \mathrm{d}\mathbf{u} + \int_{\partial U} \big(L_{q_u}\nu^u + L_{q_v}\nu^v\big) w \mathrm{d}\sigma.$$

Setting the first variation to zero for all possible perturbations (h, g) yields a coupled system of PDEs (EL equations) along with natural boundary conditions:

$$L_Z - (L_{Z_u})_u - (L_{Z_v})_v = 0 \qquad \text{in } U, \tag{17}$$

$$L_{Z_u}\nu^u + L_{Z_v}\nu^v = 0 \qquad \text{on } \partial U, \tag{18}$$

$$L_f - (L_{f_u})_u - (L_{f_v})_v = 0 \qquad \text{in } U, \tag{19}$$

$$L_{f_u}\nu^u + L_{f_v}\nu^v = 0 \qquad \text{on } \partial U. \tag{20}$$

After some calculations (see appendix A) equations (17), (18), (19), and (20) become

$$g(Z,f) - \alpha\Delta Z = 0 \qquad \text{in } U, \tag{21}$$

$$b(Z,f) + \alpha\frac{\partial Z}{\partial \boldsymbol{\nu}} = 0 \qquad \text{on } \partial U, \tag{22}$$

$$-\textstyle\sum_{i=1}^{N_c}(I_i - f)\mathtt{J}_i(Z) - \beta\Delta f = 0 \qquad \text{in } U, \tag{23}$$

$$\beta\frac{\partial f}{\partial \boldsymbol{\nu}} = 0 \qquad \text{on } \partial U, \tag{24}$$

where the non-linear terms due to the data fidelity energy are

$$\begin{aligned} g(Z,f) &= \nabla f \cdot \textstyle\sum_{i=1}^{N_c} |\mathtt{M}^i| \tilde{Z}_i^{-3} (I_i - f)(u - C_i^1, v - C_i^2), \\ b(Z,f) &= \textstyle\sum_{i=1}^{N_c} \phi_i |\mathtt{M}^i| \tilde{Z}_i^{-3} \big((u - C_i^1)\nu^u + (v - C_i^2)\nu^v\big), \end{aligned} \tag{25}$$

the Laplacians ΔZ and Δf arise from the regularizing terms (7) and (8), respectively, and $\partial * / \partial \boldsymbol{\nu}$ is the usual notation for the directional derivative along $\boldsymbol{\nu}$, the normal to the integration domain U in the parameter space.

The computations are involved, but a simple classification of the PDEs can be done as follows. For a fixed surface, (23) and (24) form a linear elliptic PDE (of the inhomogeneous Helmholtz type) with Neumann boundary conditions. On the other hand, for a fixed radiance, (21) and (22) lead to a nonlinear elliptic equation in the height Z with nonstandard boundary conditions.

Observe that if there was no regularizing term on the radiance ($\beta = 0$), equation (23) would be linear in f, and the solution would be a weighted average of the intensities at the image projections of the surface (42).

A common approach to solve difficult EL equations, such as the EL equation presented in (21), (22), (23), and (24), is to add an artificial time marching variable t dependency in the unknown functions (height, radiance) and

set up a gradient descent flow that will drive their evolution such that the energy (14) will decrease in time. Thus the solution of the elliptic PDEs (EL equations) is obtained as the steady-state of the gradient descent equations. This is the context of the so-called active surfaces. The gradient descent equations are:

$$Z_t = \alpha \Delta Z - g(Z, f), \tag{26}$$

$$f_t = \beta \Delta f - \textstyle\sum_{i=1}^{N_c} \mathtt{J}_i(Z)\, f + \sum_{i=1}^{N_c} I_i \mathtt{J}_i(Z). \tag{27}$$

To simplify the equations, we approximate the boundary condition (22) by a simpler, homogeneous Neumann boundary condition. This can be interpreted as if the data fidelity term vanished close to the boundary and it is a reasonable assumption since the major contribution to the energy is given by the terms in the interior of the discretized domain, not at the boundary.

D. Numerical solution.

An iterative, alternating approach is used to find the minimum of energy (2) via the evolution of the coupled gradient descent PDEs (26) and (27). During each iteration there are two phases: (1) evolve the shape, keeping the radiance fixed and (2) evolve the radiance, leaving the shape unchanged. The PDEs (26) and (27) are solved numerically after being discretized on a rectangular 2-D grid in the parameter space, with equidistant step size $h = \Delta u = \Delta v$ in both dimensions, i.e., along directions u and v of the integration region U. Forward differences in time and central differences in space approximate the derivatives, yielding an *explicit updating scheme*. The time step Δt is determined by the stability condition of the resulting PDE. In the case of the linear PDE in the radiance, (27), the time step for ℓ^2 stability satisfies

$$\Delta t \leq \big(\frac{4\beta}{h^2} + \frac{1}{2} \max \sum_{k=1}^{N_c} \mathtt{J}_k\big)^{-1}, \tag{28}$$

where $\mathtt{J}_k(Z) \geq 0$ and the maximum is taken over the 2-D discretized Jacobians for the current height function. Thus, the time step may change at every iteration, depending on the value of the evolving height. On the other hand, since equation (21) is a nonlinear PDE, the stability analysis is more complicated than in the linear case above. Nevertheless, we use the stability condition derived from the *linearized* PDE. The time step for ℓ^2 stability of (26) satisfies

$$\Delta t \leq \big(\frac{4\alpha}{h^2} + \frac{1}{2} \max |\dot{g}(Z)|\big)^{-1}, \tag{29}$$

where $\dot{g}(Z)$ is the derivative of (25) and the maximum is taken over the 2-D discretized grid at the current time. The maximum time step (29) may change at every iteration, as in the case of (28). In the experiments, the time steps used are a conservative proportion of the maximum allowable time steps: $0.8 \max \Delta t$.

The previous time-stepping methods are used as relaxation procedures inside a multigrid method [4] that approximately solves the EL equations. Multigrid methods are the most efficient numerical tools for solving elliptic boundary value problems. to detect stagnation and overcome it.

So far, the regularizing terms (7) and (8) have no physical meaning according to the dynamics of the water waves. They are the simplest smoothness penalties to support the conjecture that the problem is well posed and a

solution exists, without providing a rigorous proof. Since the regularizer on the shape of the surface (7) acts on a geometric object, a more sensible geometric choice that does not significantly complicate the model is to penalize the total area of the reconstructed surface:

$$E_{\text{geom}} = \int_S \mathrm{d}A = \int_U \sqrt{1 + Z_u^2 + Z_v^2}\,\mathrm{d}\mathbf{u}. \tag{30}$$

Surfaces that minimize the above energy are called minimal surfaces and they have the property of zero mean curvature. If (30) is used in (2), the diffusive term in the PDE (21), i.e. the Laplacian $\Delta Z = Z_{uu} + Z_{vv}$, is replaced by the mean curvature:

$$2H = \frac{(1 + Z_v^2)Z_{uu} - 2Z_u Z_v Z_{uv} + (1 + Z_u^2)Z_{vv}}{(1 + Z_u^2 + Z_v^2)^{3/2}}.$$

Calculations show that the new regularizer does not alter the homogeneous Neumann boundary condition. Assuming the explicit updating scheme is used to relax the modified non-linear PDE in the height, an ℓ^2 stability condition for the time step can be derived using Fourier analysis under reasonable approximations. The maximum time step has the same form as (29), but with $4\alpha/h^2$ replaced by $5\alpha/h^2$.

III. Experiments

a) Experiment 1: Images of "Canale della Giudecca" in Venice (Italy). After validating the numerical implementation of the proposed variational stereo method with synthetic data, some experiments with real data are carried out. Figs. 2 and 3 show an example of a reconstructed water surface from images of the Venice Canal. Cropped images in Fig. 2 are of size 600×450 pixels and show the region of interest to be reconstructed. Fig. 2 also displays the modeled images created by the generative model within our variational method. The data fidelity term compares the intensities of the original and modeled images in the highlighted region, in all images. As observed, the modeled images are a good match of the original images. Fig. 3 shows the converged values of the unknowns of the problem: the height and the radiance of the surface, as well as the 3-D representation of the reconstructed surface obtained by combining both 2-D functions. In this experiment, the values of the weights of the regularizers were empirically determined: $\alpha = 0.035$ and $\beta = 0.01$. At the finest of the 5-level multigrid [4] algorithm, the gradient descent PDEs are discretized on a 2-D grid with 129×513 points. The distance between adjacent grid points is $h = 5$ cm. Therefore, the grid covers an area of $6.45 \times 25.65\, m^2$. An example of a surface discretized at the finest grid level is shown in Fig. 3. Observe the high density of the surface representation, typical of variational methods. The step size h must be chosen so that it approximately matches the resolution in the images: a displacement of 1 pixel is observable at the finest grid level in the multigrid framework and it corresponds to a physical displacement of at least h. Due to perspective projection, the maximum value of h is determined by the grid points closest to the cameras.

b) Experiment 2: Image sequence I. The method proposed in this paper is naturally extended to process stereo video on a snapshot-by-snapshot basis by estimating the new surface shape and radiance based on the previously reconstructed surface. This sequential processing is the simplest way in which the method can be applied to stereo video imagery. We test the method on a different video data consisting of 10 consecutive snapshots (i.e., frames)

with images of size 1000×1000 pixels. A grid of size 513×513 points and with a step size $h = 1.5$ cm is selected. Thus, the grid covers an area of $7.7 \times 7.7\, m^2$. The deforming surface is initialized by the plane $Z = 0$. A multigrid method with 6 levels and 200 V-cycles (with 1 pre- and post-relaxation sweeps per level) is used to solve the problem at each snapshot. For the first frame, a full multigrid method (FMG) with 200 V-cycles per level is performed prior to entering the above processing schedule. In this experiment, the weights of the regularizers are $\alpha = 4{\cdot}10^{-2}$ and $\beta = 4{\cdot}10^{-3}$. Another reconstruction of the wave surface from video data collected by Benetazzo[2] is shown in Fig. 4. In the same figure we also report the the omni-directional spectrum $S(k)$ (averaged over the frames), computed by numerically integrating the 2-D spectrum $S(k_x, k_y)$ of the elevation map over all directions, where the wavenumber is $k = |\mathbf{k}| = \sqrt{k_x^2 + k_y^2}$. In agreement with Zakharov's theory [34], the spectrum tail decays close to $k^{-2.5}$. The results of the mean curvature diffusive term from (30) are a minor modification of the ones obtained with the Laplacian term.

c) Experiment 3: Image sequence II. We apply our variational method to a sequence of 2000 snapshots acquired at 10 Hz and at an off-shore platform near the southern seashore of the Crimean peninsula, in the Black Sea. Two cameras mounted 12 m above the mean sea level and with a baseline of 2.5 meters acquire images of size 1624×1236 pixels. Fig. 5 (left) shows a sample image from one of the cameras. A grid with 513×513 points and resolution $h = 2.5$ cm, covering an area of $13 \times 13\, m^2$, is used to discretize the graph of the surface. Fig. 5 (right) shows the approximate region of interest occupied by the projection of the reconstructed surface on one of the images. Roughly, 1 image pixel corresponds to a physical displacement of 1.06 cm (1.88 cm) for grid points near (resp. far from) the cameras. Both displacements are of the same order as h. The same multigrid processing scheme as in experiment 2 is used, but with 1000 V-cycles per level. The weights of the regularizers are $\alpha = 0.1$ and $\beta = 0.025$.

The four-dimensional reconstructed wave surface can be represented in the form of a space-time volume of wave heights, $V = Z(x, y, t)$, as visualized in Figure 6, where the oscillating pattern of the waves is evident by the oscillating color patterns. The spectra and statistics of the waves can be computed from the reconstructed surface.

The mean omni-directional spectrum $S(k)$, averaged over all 2000 snapshots, is shown in Fig. 7. According to the wave turbulence theory of Zakharov [34], the spectrum tail initially decays as $k^{-2.5}$ as a result of an energy cascade from large to small scales up to ~ 10 rad/m and then switch to a k^{-3} equilibrium range [22]. Fig. 7 clearly shows such behavior of the tail of the spectrum. More generally, Fourier analysis can be applied directly to the wave space-time volume $Z(x, y, t)$, resulting in a 3-D spectrum $\mathcal{Z}(k_x, k_y, \omega)$, where frequencies k_x, k_y and ω are the Fourier transformed variables of x, y and t, respectively. Recall that $\omega = 2\pi f$ is the angular frequency and $f \equiv f_t$ is the linear frequency, measured in Hertzs. Similarly, let us use $f_x = k_x/(2\pi)$ and $f_y = k_y/(2\pi)$ for the wavenumbers in units of cycles/m. The 3-D spectrum contains information of the propagation characteristics of the waves, such as their wavelengths, frequencies, and their directions and speeds of propagation. From a practical point of view, the 3-D spectrum of the reconstructed $513 \times 513 \times 2000$ wave height grid is computed by averaging the 3-D spectra of non-overlapping pieces of the grid. We split the wave space-time volume along the temporal dimension to compute the 3-D spectrum on a Fourier grid with $512 \times 512 \times 512$ points; thus, each piece consists of $N_t = 512$

snapshots. The Nyquist wavenumbers are $[f_x, f_y, f]_{\text{max}} = [h^{-1}, h^{-1}, \Delta t^{-1}] = [20\,\text{cycles}/m, 20\,\text{cycles}/m, 5\text{Hz}]$. The spectral resolutions are given by $\Delta f_x = \Delta f_y = 1/(Nh) \approx 0.078\,\text{cycles}/m$ and $\Delta f = 1/(N\Delta t) \approx 0.02\,\text{Hz}$ for the 3-D FFT with $N = 512$ points in each dimension. Fig. 8 shows the 3-D wave spectrum, and Fig. 9 shows two of its slices through the frequency axes: the frequency-wave number spectra ω-k_x and ω-k_y, respectively. The white curve in the vertical slices of $\mathcal{Z}(k_x, k_y, \omega)$ corresponds to planar projections of the linear dispersion manifold in deep water, namely $\sqrt{k_x^2 + k_y^2} = \omega^2/g$, where g is gravity acceleration. Other researchers [6] have measured the ω-k spectrum for long wave ranges at nearshore events to estimate surface currents and the water depth below the waves. Their measurements are also shown in comparison to the linear dispersion relation. At the Crimean platform, the water depth is approximately 30 meters. Therefore, for all practical purposes with respect to our wavenumber resolution, the depth can be regarded as being infinite. The components of the current $\mathbf{v}$ can be estimated from the observed deviations from the theoretical dispersion curve, as shown in Fig. 9, by a best fit of the wave-current dispersion relation $k = (\omega - \mathbf{k} \cdot \mathbf{v})^2/g$, where $k = |\mathbf{k}|$ (see [29], [14]). This yields $\mathbf{v} \approx (-0.17, -0.45)$ m/s, with the dominant component in the y direction. This propagation direction agrees with the one observed by visual inspection of the stereo video data. Figure 9 shows strong physical evidence to support the hypothesis that the variational graph method presented in this work is capturing real waves propagating in the observed direction.

Time series at virtual probes. Statistical analysis. The rich content of the space-time reconstruction of the surface wave allows for the extraction of *time series* of wave displacements $Z_i(t) = Z(x_i, y_i, t)$ from the space-time volume V at virtual probes (x_i, y_i) in space, as illustrated in Figure 10. Several statistical and spectral parameters that characterize the sea states can be computed from such time series. The significant wave height and mean wave period are $H_s = 0.3$ m and $T_m = 2.77$ s, respectively. Fig. 11 shows the observed Power Spectral Density estimated from time series extracted from the wave space-time volume. An FFT with 2048 points was used, i.e., the spectral resolution is $\Delta f = 5 \cdot 10^{-3}$ Hz. If the tail of the wave number spectrum decays as $\tilde{F}(k) \propto k^{-2.5}$, the tail of the frequency spectrum decays as $F(f) \propto f^{-4}$. This behavior is observed in Figure 11, which also shows a verification of our variational method against an earlier WASS measurement technique based on epipolar geometry [2]. The peak at 2 Hz observed in the black curve is due to vibrations induced by fishermen walking on the Crimea platform while WASS was recording. The epipolar reconstruction [2] is purely based on the imaged data with no regularizing term as in the variational approach. The variational method unveiled the small-scale range of the spectrum improving the estimate at large wave numbers and frequencies.

By collecting the time waves observed at all the virtual probes indicated in Figure 10, one can estimate the wave height distribution, which is shown in Figure 12. A fair agreement with the Boccotti asymptotic form given by [3], [10]

$$P(\text{wave height} > H) \approx c \exp\left(-\frac{H^2}{4\sigma^2(1+\psi^*)}\right), \tag{31}$$

is observed. Here, the parameters c and $\psi^* \equiv |\psi(T^*)|$ both depend upon the first minimum of the wave covariance $\psi(T)$. In particular the mean values of c and ψ^* over the time series ensemble are $c \approx 1$ and $\psi^* \approx 0.52$.

IV. Conclusion

Building upon the multiple benefits of variational stereo methods over earlier traditional stereo methods, we develop a variational stereo method for the case of smooth surfaces representable in the form of a graph supporting a smooth radiance function. We successfully apply this method to reconstruct small regions of the ocean surface in several datasets (including video data) and begin to tailor the method for this particular problem, where the initially chosen regularizing terms (7) and (8) have no physical meaning according to the dynamics of the ocean waves. However, other regularizers such as (30) can be used in the variational framework to account for more physical properties of the waves. In future research we plan to elaborate on better choices for the regularizers as well as new ones that include global and/or local properties of the dynamics of ocean waves such as statistical distribution of wave heights, the wave equation, etc.

Departing from the simple snapshot-by-snapshot sequential temporal processing used in some of the experiments, the variational framework allows for better ways to enforce coherence in space-time of the reconstructed surface. This topic is now under investigation. Preliminary research shows that VWASS is a promising remote-sensing observational technology with a broader impact on ocean engineering since it will enrich the understanding of the oceanic sea states and wave statistics, enabling improved designs of off-shore structures and platforms.

Appendix A

Recasting the integral from the image domain to the parameter space

A. Jacobian of the change of variables

Let us derive an expression for the Jacobian of the change of integration variables from the image domain to the surface: $\mathtt{J}_i = \left|\frac{\mathrm{d}\mathbf{x}_i}{\mathrm{d}\mathbf{u}}\right|$. Applying the chain rule to (6), we have

$$\frac{\mathrm{d}\mathbf{x}_i}{\mathrm{d}\mathbf{u}} = \frac{\mathrm{d}\mathbf{x}_i}{\mathrm{d}\tilde{\mathbf{X}}_i}\frac{\mathrm{d}\tilde{\mathbf{X}}_i}{\mathrm{d}\mathbf{X}}\frac{\mathrm{d}\mathbf{X}}{\mathrm{d}\mathbf{u}} = \frac{1}{\tilde{Z}_i^2}\begin{pmatrix}\tilde{Z}_i & 0 & -\tilde{X}_i\\ 0 & \tilde{Z}_i & -\tilde{Y}_i\end{pmatrix}\mathtt{M}^i\,(\mathbf{X}_u, \mathbf{X}_v). \tag{32}$$

Using the fact that a point with homogeneous coordinates $\bar{\mathbf{C}}_i = (\mathbf{C}_i^\top, 1)^\top$ is the optical center of the i-th camera if it satisfies $\mathtt{P}^i\bar{\mathbf{C}}_i = \mathbf{0}$, i.e.,

$$\mathtt{M}^i\mathbf{C}_i + \mathbf{p}_4^i = \mathbf{0} \quad \Leftrightarrow \quad \mathbf{C}_i = -(\mathtt{M}^i)^{-1}\mathbf{p}_4^i, \tag{33}$$

the determinant of (32) becomes

$$\det\left(\frac{\mathrm{d}\mathbf{x}_i}{\mathrm{d}\mathbf{u}}\right) = |\mathtt{M}^i|\tilde{Z}_i^{-3}(\mathbf{X}-\mathbf{C}_i)\cdot(\mathbf{X}_u\times\mathbf{X}_v), \tag{34}$$

where $\mathtt{M}^i = (\mathbf{n}_1^i, \mathbf{n}_2^i, \mathbf{n}_3^i)^\top$ is the left 3×3 sub-matrix of the projection matrix $\mathtt{P}^i = (p_{kl}^i)$, and $\tilde{Z}_i = \mathbf{n}_3^i\cdot\mathbf{X} + p_{34}^i$ can also be expressed as

$$\tilde{Z}_i = \mathbf{n}_3^i\cdot(\mathbf{X}-\mathbf{C}_i). \tag{35}$$

Here, $\tilde{Z}_i > 0$ is the depth of the point $\mathbf{X}$ with respect to the i-th camera (located at $\mathbf{C}_i$), as is customary, in the direction of the normal $\mathbf{n}_3^i$ to the principal plane of the camera. We use the standard notation [13] that states that

the depth is positive for points in front of the camera. Finally, since the Jacobian is positive, it is the absolute value of (34).

Visibility of a surface point with respect to the camera can be included in the Jacobian. Recall that $\mathbf{X}_u \times \mathbf{X}_v$ is proportional to the outward unit normal to the surface at $\mathbf{X}(u, v)$:

$$\mathbf{N} = \frac{\mathbf{X}_u \times \mathbf{X}_v}{\|\mathbf{X}_u \times \mathbf{X}_v\|}, \tag{36}$$

Observe that $(\mathbf{X} - \mathbf{C}_i) \cdot \mathbf{N} < 0$ for neighborhoods of surface points (i.e., patches) pointing toward the camera and $(\mathbf{X} - \mathbf{C}_i) \cdot \mathbf{N} > 0$ for patches pointing away from the camera. The latter are occluded by the former from the viewpoint of the camera. Therefore,

$$\mathtt{J}_i = |\mathtt{M}^i| \tilde{Z}_i^{-3} \max\big(-(\mathbf{X} - \mathbf{C}_i) \cdot (\mathbf{X}_u \times \mathbf{X}_v), 0\big). \tag{37}$$

Beware that, for a given surface point $\mathbf{X}$, the condition of positive Jacobian is not sufficient for that point to be visible from the camera viewpoint since the surface may be self-occluded. Therefore, a positive Jacobian is a necessary visibility condition, but not a sufficient condition.

B. Area measures in the image and on the surface

With the expression of the Jacobian of the change of variables at hand (37), it is straightforward to derive a formula for the relationship between area elements in the image plane and on the surface: $\mathrm{d}\mathbf{x}_i = \mathtt{J}_i \, \mathrm{d}\mathbf{u}$. Since the surface area element is $\mathrm{d}A = \|\mathbf{X}_u \times \mathbf{X}_v\| \, \mathrm{d}u\mathrm{d}v$ and the outward unit normal to the surface at $\mathbf{X}(u, v)$ is (36), the relationship between area elements can be rewritten as

$$\mathrm{d}\mathbf{x}_i = |\mathtt{M}^i| \tilde{Z}_i^{-3} \max\big(-(\mathbf{X} - \mathbf{C}_i) \cdot \mathbf{N}, 0\big) \, \mathrm{d}A. \tag{38}$$

The term $(\mathbf{X} - \mathbf{C}_i) \cdot \mathbf{N}$ is proportional to the cosine of the angle between the unit normal to the surface at $\mathbf{X}$ and the *projection ray* (the ray joining the optical center of the camera and $\mathbf{X}$). One may observe the extreme cases: (*i*) If $(\mathbf{X} - \mathbf{C}_i) \perp \mathbf{N}$, the surface patch at $\mathbf{X}$ projects to a line in the image plane, hence $\mathrm{d}\mathbf{x}_i = 0$ (zero area) and that patch makes no contribution to the energy E_i. (*ii*) On the other hand, if the projection ray is parallel to the normal of the surface patch at that point, i.e., $(\mathbf{X} - \mathbf{C}_i) \parallel \mathbf{N}$, the surface patch projects onto a maximum area region $\mathrm{d}\mathbf{x}_i$. This qualitative behavior of the model agrees with our physical intuition.

To simplify calculations related to the evolution of the surface height and radiance according to the data fidelity term we will use the former expression for the Jacobian that does not take into account the necessary visibility condition, i.e.

$$\mathtt{J}_i = -|\mathtt{M}^i| \tilde{Z}_i^{-3} (\mathbf{X} - \mathbf{C}_i) \cdot (\mathbf{X}_u \times \mathbf{X}_v), \tag{39}$$

but we will bear in mind that if the surface point under consideration is not visible, it will not be allowed to evolve according to the data fidelity component.

APPENDIX B
EULER-LAGRANGE EQUATIONS

Here it is shown how to calculate the necessary optimality conditions to minimize the proposed energy functional (2). The variation of the energy with respect to the surface radiance will be presented first because it is easier to compute than the variation with respect to the shape.

A. Variation with respect to the surface radiance

Let us derive the PDE related to the first variation of the energy with respect to the radiance (19). Since E_{geom} does not depend on the radiance f, it has no effect on the aforementioned first variation. Straightforward calculations show that for the regularizer (8),

$$(L_{\text{rad}})_f - \big((L_{\text{rad}})_{f_u}\big)_u - \big((L_{\text{rad}})_{f_v}\big)_v = -f_{uu} - f_{vv}, \tag{40}$$

which is the Laplacian in (23). Focusing now on the data fidelity term, L_i does not depend on f_u, f_v. Therefore

$$(L_i)_f - \big((L_i)_{f_u}\big)_u - \big((L_i)_{f_v}\big)_v = -(I_i - f)\mathtt{J}_i. \tag{41}$$

It is straightforward to derive (23) by substituting (40), (41), and (13) in (19) and applying linearity. Observe that if $\beta = 0$ in (23), the optimal f is the weighted average

$$f = \textstyle\sum_{i=1}^{N_c} w_i I_i, \qquad w_i = \dfrac{\mathtt{J}_i}{\sum_{j=1}^{N_c} \mathtt{J}_j}, \tag{42}$$

where the weights w_i may not yield a convex combination because the non-negative Jacobians might all vanish for an occluded surface point.

a) Boundary condition for the PDE in the radiance of the surface: Neumann boundary conditions naturally arise from (20). The regularizer (8) yields the directional derivative of f along $\boldsymbol{\nu}$, the unit normal to ∂U:

$$(L_{\text{rad}})_{f_u}\nu^u + (L_{\text{rad}})_{f_v}\nu^v = f_u\nu^u + f_v\nu^v = \frac{\partial f}{\partial \boldsymbol{\nu}}. \tag{43}$$

Because L_i and L_{geom} do not depend on the gradient of f, the left hand side of (20) is $\beta\frac{\partial f}{\partial \boldsymbol{\nu}}$. If $\beta \neq 0$, it follows that (20) is equivalent to the boundary condition $\frac{\partial f}{\partial \boldsymbol{\nu}} = 0$ on ∂U.

B. 3-D extensions of the radiance and the images

To simplify the calculations involved in the EL equations, let us define extensions of the radiance and image intensities to the whole three-dimensional space, $\hat{f} : \mathbb{R}^3 \to \mathbb{R}$ and $\hat{I}_i : \mathbb{R}^3 \to \mathbb{R}$, respectively. It is natural to define the latter as being constant along optical rays (projection rays) from the camera,

$$\hat{I}_i|_S \equiv \hat{I}_i(\mathbf{X}) \doteq I_i\big(\pi_i(\mathbf{X})\big). \tag{44}$$

The extension of the radiance $\hat{f}$ has been introduced in section II-A. Let us define the extension to be constant along the third dimension, i.e., the Z axis. In the considered world frame (where the parameter space of the surface is the plane $Z = 0$), equation this implies that $\hat{f}(X, Y, Z) \doteq f(X, Y)$. The photometric matching criterion (4) can

also be extended to the whole space, $\hat{\phi}_i : \mathbb{R}^3 \to \mathbb{R}$, by the definition: $\hat{\phi}_i(\mathbf{X}) \doteq \frac{1}{2}\big(\hat{I}_i(\mathbf{X}) - \hat{f}(\mathbf{X})\big)^2$. It is clear that for surface points the restriction of the extension satisfies $\hat{\phi}_i|_S \equiv \phi_i$.

C. Variation with respect to the shape of the surface

Now, let us compute each term in the left hand side of (17). Since L_{rad} does not depend on Z, it has no effect on (17). On the other hand, straightforward calculations show that the chosen regularizer (7) yields the Laplacian, ΔZ, as in (40). Let us focus now on the data fidelity term. The extensions defined in appendix B-B make it possible to rewrite the Lagrangian L_i in (13) as a function of $\mathbf{X}$, $L_i = \hat{\phi}_i(\mathbf{X}, \hat{f})\mathtt{J}_i(\mathbf{X}, \mathbf{X}_u, \mathbf{X}_v)$. The chain rule can be used to compute the left hand side of (17) for L_i because the derivatives in Z are projections of the ones in $\mathbf{X}$:

$$L_Z - (L_{Z_u})_u - (L_{Z_v})_v = \big(L_{\mathbf{X}} - (L_{\mathbf{X}_u})_u - (L_{\mathbf{X}_v})_v\big) \cdot \mathbf{e}_3, \tag{45}$$

where $\mathbf{e}_3 = (0, 0, 1)^\top$ is the direction of variation of the height. Now, it remains to calculate $(L_i)_{\mathbf{X}}$, $\big((L_i)_{\mathbf{X}_u}\big)_u$ and $\big((L_i)_{\mathbf{X}_v}\big)_v$. As is customary, let ∇ denote the spatial derivative, then image derivatives will arise in the calculations:

$$\nabla \hat{I}_i^\top = \nabla I_i^\top \frac{\partial \pi_i}{\partial \mathbf{X}}, \tag{46}$$

with $\partial \pi_i / \partial \mathbf{X}$ as in (32). As a space point $\mathbf{X}$ moves along the optical ray from a camera, the corresponding image point $\mathbf{x}_i = \pi_i(\mathbf{X})$ remains unchanged. This implies that

$$\frac{\partial \pi_i}{\partial \mathbf{X}} (\mathbf{X} - \mathbf{C}_i) = \mathbf{0}. \tag{47}$$

The proof is based on (33) and the formula for $\partial \pi_i / \partial \mathbf{X}$. Combining (46) and (47) one can show that, since the intensity of the extension $\hat{I}_i$ is constant along the projection ray, $\nabla \hat{I}_i$ lies in the plane orthogonal to such projection ray:

$$\nabla \hat{I}_i \cdot (\mathbf{X} - \mathbf{C}_i) = 0. \tag{48}$$

This result will lead to a simplification of the term $(L_i)_{\mathbf{X}}$ that will have an important consequence: no derivatives of the image data appear in the final EL equations. This desirable feature makes the algorithm less sensitive to image noise when compared to other variational approaches for stereo 3-D reconstruction. This feature is shared by the standard Mumford-Shah [21] formulation for direct image segmentation. In our case, it arises from the fact that the stereo discrepancy is measured in the image domain rather than on the surface [30]. Observe that it is a purely geometric result, thus independent of the choice of ϕ_i.

If the surface is sufficiently smooth such that $\mathbf{X}_{uv} = \mathbf{X}_{vu}$ (twice continuously differentiable), one can show that

$$\begin{aligned}
&(L_i)_{\mathbf{X}} - \big((L_i)_{\mathbf{X}_u}\big)_u - \big((L_i)_{\mathbf{X}_v}\big)_v \\
&= (\hat{\phi}_i)_{\mathbf{X}} \mathtt{J}_i - (\hat{\phi}_i)_u (\mathtt{J}_i)_{\mathbf{X}_u} - (\hat{\phi}_i)_v (\mathtt{J}_i)_{\mathbf{X}_v} \\
&\qquad + \hat{\phi}_i \Big((\mathtt{J}_i)_{\mathbf{X}} - \big((\mathtt{J}_i)_{\mathbf{X}_u}\big)_u - \big((\mathtt{J}_i)_{\mathbf{X}_v}\big)_v \Big), \\
&= -|\mathtt{M}^i| \tilde{Z}_i^{-3} \Big((\hat{I}_i - \hat{f})\, \mathbf{W} (\nabla \hat{I}_i - \nabla \hat{f}) \\
&\qquad + 3 \hat{\phi}_i \big((\mathbf{X}_u \times \mathbf{X}_v) - \tilde{Z}_i^{-1} \mathbf{W}(\mathbf{n}_3^i) \big) \Big),
\end{aligned} \tag{49}$$

where we define the vector

$$\begin{aligned}\mathbf{W}(\mathbf{b}) &= -(\mathbf{X}-\mathbf{C}_i)\cdot(\mathbf{X}_u\times\mathbf{X}_v)\,\mathbf{b}\\ &+(\mathbf{b}\cdot\mathbf{X}_u)\big(\mathbf{X}_v\times(\mathbf{X}-\mathbf{C}_i)\big)+(\mathbf{b}\cdot\mathbf{X}_v)\big((\mathbf{X}-\mathbf{C}_i)\times\mathbf{X}_u\big).\end{aligned}$$

Next, we show that this vector is proportional to the unit normal $\mathbf{N}$ and the value of $\mathbf{b}$ only affects its magnitude. Let $\mathtt{A}=(\mathbf{X}-\mathbf{C}_i,\mathbf{X}_u,\mathbf{X}_v)$, then from $\mathtt{I}=\mathtt{A}^{-\top}\mathtt{A}^{\top}$ we derive the matrix relationship

$$\begin{aligned}&\det(\mathtt{A})\mathtt{I}-\big(\mathbf{X}_v\times(\mathbf{X}-\mathbf{C}_i)\big)\mathbf{X}_u^{\top}-\big((\mathbf{X}-\mathbf{C}_i)\times\mathbf{X}_u\big)\mathbf{X}_v^{\top}\\ &=(\mathbf{X}_u\times\mathbf{X}_v)(\mathbf{X}-\mathbf{C}_i)^{\top},\end{aligned}$$

which can be used to obtain

$$\mathbf{W}(\mathbf{b})=(\mathbf{X}_u\times\mathbf{X}_v)(\mathbf{X}-\mathbf{C}_i)^{\top}\mathbf{b}. \tag{50}$$

Substituting (50) in (49) and using (35), (45), and (48), yields important simplifications: the term that is multiplying $\hat{\phi}_i$ vanishes and no image derivatives appear in the final expression. Therefore, the left hand side of (17) for L_i becomes

$$(L_i)_Z-\big((L_i)_{Z_u}\big)_u-\big((L_i)_{Z_v}\big)_v=|\mathtt{M}^i|\tilde{Z}_i^{-3}(\hat{I}_i-\hat{f})(\mathbf{X}-\mathbf{C}_i)\cdot\nabla\hat{f},$$

after substituting $\mathbf{X}_u\times\mathbf{X}_v=(-Z_u,-Z_v,1)^{\top}$ in $(\mathbf{X}_u\times\mathbf{X}_v)\cdot\mathbf{e}_3=1$. The freedom in the definition of $\hat{f}$ allows for further simplifications: $\nabla\hat{f}=(\nabla f^{\top},0)^{\top}$ implies that

$$\begin{aligned}&(L_i)_Z-\big((L_i)_{Z_u}\big)_u-\big((L_i)_{Z_v}\big)_v\\ &=|\mathtt{M}^i|\tilde{Z}_i^{-3}(I_i-f)(u-C_i^1,v-C_i^2)\cdot\nabla f,\end{aligned}$$

where $\mathbf{C}_i=(C_i^1,C_i^2,C_i^3)^{\top}$. The terms affected by Z are the image intensity at the current surface point, $I_i\equiv I_i(\mathbf{x}_i(\mathbf{X}(u,v,Z(u,v))))$, and the depth of the surface point with respect to the camera, $\tilde{Z}_i=\mathbf{n}_3^i\cdot\mathbf{X}(u,v,Z(u,v))+p_{34}^i$. Collecting terms for multiple images and regularizers, the EL equation (17) of the composite energy becomes (21). Observe that (21) does not depend on the image derivatives (∇I_i), as previously announced.

b) Boundary condition for the PDE in the height of the surface: The PDE (17) comes with natural boundary condition (18) because the surface is not closed. The geometric regularizer (7) yields the directional derivative of Z along $\boldsymbol{\nu}$, as in (43). The boundary condition arising from the data fidelity term is $b(Z,f)$ by the chain rule and previous results:

$$\begin{aligned}&(L_i)_{Z_u}\nu^u+(L_i)_{Z_v}\nu^v\\ &=\big(\nu^u(L_i)_{\mathbf{X}_u}+\nu^v(L_i)_{\mathbf{X}_v}\big)\cdot\mathbf{e}_3\\ &=\hat{\phi}_i|\mathtt{M}^i|\tilde{Z}_i^{-3}\Big(\nu^u\big((\mathbf{X}-\mathbf{C}_i)\times\mathbf{X}_v\big)\\ &\qquad+\nu^v\big(\mathbf{X}_u\times(\mathbf{X}-\mathbf{C}_i)\big)\Big)\cdot\mathbf{e}_3.\end{aligned}$$

Collecting expressions from all terms in (13) yields (22).

ACKNOWLEDGMENTS

Research supported by ONR grant BAA 09-012: “Ocean Wave Dissipation and energy Balance (WAVE-DB): toward reliable spectra and first breaking statistics”. The authors would like to thank both Aziz M. Tayfun and Harald Krogstad for useful discussions and suggestions.

REFERENCES

[1] L. Alvarez, R. Deriche, J. Sánchez, and J. Weickert. Dense disparity map estimation respecting image discontinuities : A pde and scale-space based approach. *Journal of Visual Communication and Image Representation*, 13:3–21, 2002.

[2] A. Benetazzo. Measurements of short water waves using stereo matched image sequences. *Coastal Engineering*, 53(12):1013–1032, December 2006.

[3] Paolo Boccotti. *Wave Mechanics for Ocean Engineering*. Elsevier Oceanography Series. Elsevier, Oxford, 2000.

[4] William L. Briggs, Van Emden Henson, and Steve F. McCormick. *A multigrid tutorial (2nd ed.)*. SIAM, Philadelphia, PA, USA, 2000.

[5] L.J. Coté, J.O. Davis, W. Marks, R.J. McGough, E. Mehr, W.J. Pierson, J.F. Ropek, G. Stephenson, and R.C. Vetter. The directional spectrum of wind generated sea as determined from data obtained by the stereo wave observation project. *Metereological paper*, 2(6), 1960.

[6] J. P. Dugan, C. C. Piotrowski, and J. Z. Williams. Water depth and surface current retrievals from airborne optical measurements of surface gravity wave dispersion. *Journal of Geophysical Research*, 106(C8):16903–16915, August 2001.

[7] Olivier D. Faugeras and Renaud Keriven. Variational principles, surface evolution, pdes, level set methods, and the stereo problem. *IEEE Transactions on Image Processing*, 7(3):336–344, 1998.

[8] F. Fedele, A. Benetazzo, and G.Z. Forristall. Space-time waves and spectra in the northern adriatic sea. In *ASME 2011 30th International Conference on Ocean, Offshore and Arctic Engineering OMAE2011*, June 19-24 2011. Rotterdam, The Netherlands. Paper no. OMAE2011-49924.

[9] F. Fedele, G. Gallego, A. Yezzi, and A. Benetazzo. Understanding extreme waves via a variational wave acquisition stereo system. In *Workshop ROGUE WAVES*, 2008. Oct 13-15, Brest, France.

[10] F. Fedele and A. Tayfun. On nonlinear wave groups and crest statistics. *Journal of Fluid Mechanics*, 620:221–239, 2009.

[11] G. Gallego, A. Benetazzo, A. Yezzi, and F. Fedele. Wave statistics and spectra via a variational wave acquisition stereo system. In *27th International Conference on Offshore Mechanics and Arctic Engineering OMAE2008*, June 2008. Lisbon, Portugal. Paper no. OMAE2008-57160.

[12] K.-R. Hao and Y.-S. Ding. Trinocular matching realized by a monocular stereovision sensor for parallel manipulator. In *10th International Conference on Control, Automation, Robotics and Vision - ICARCV*, pages 1436–1441, 2008.

[13] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004.

[14] Daniele Hauser, Kimmo K. Kahma, Harald E. Krogstad, Susanne Lehner, Jaak Monbaliu, and Lucy R. Wyatt. *European COST Action 714 (Measuring and Analysing the Directional Spectrum of Ocean Waves)*. European Commission/European Science Fundation, 2003.

[15] Y.-H. Hsiao and M.-C. Huang. Application of active contour model in tracking sequential nearshore waves. *China Ocean Engineering*, 23(12):251–266, 2009.

[16] H. Jin. *Variational methods for shape reconstruction in computer vision*. PhD thesis, Washington University, St. Louis, MO, USA, 2003. Director: Soatto, Stefano.

[17] Xiaoming Li, S. Lehner, and W. Rosenthal. Investigation of ocean surface wave refraction using terrasar-x data. *Geoscience and Remote Sensing, IEEE Transactions on*, 48(2):830 –840, 2010.

[18] Ye Lu, J.Z. Zhang, Q.M.J. Wu, and Ze-Nian Li. A survey of motion-parallax-based 3-d reconstruction algorithms. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 34(4):532 –548, nov. 2004.

[19] Y. Ma, S. Soatto, J. Kosecka, and S. Sastry. *An Invitation to 3D Vision: From Images to Geometric Models*. Springer Verlag, 2003.

[20] K.R. MacHutchon and P.C. Liu. Measurement and analysis of ocean wave fields in four dimensions. In *Proceedings of the International Conference on Offshore Mechanics and Arctic Engineering - OMAE*, volume 1, pages 923–927, 2007.

[21] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and variational problems. *Comm. Pure Appl. Math*, XLII(5):577–685, 1988.

[22] O.M. Phillips. The equilibrium range in the spectrum of wind-generated waves. *Journal of Fluid Mechanics*, 4(4):426–434, 1958.

[23] F. Santel, W. Linder, and C. Heipke. Stereoscopic 3d-image sequence analysis of sea surfaces. *Proceedings of the ISPRS Commission V Symposium*, 35(5):708–712, 2004.

[24] J. Schulz-Stellenfleth and S. Lehner. Measurement of 2-D Sea Surface Elevation Fields Using Complex Synthetic Aperture Radar Data. *IEEE Transactions on Geoscience and Remote Sensing*, 42:1149–1160, June 2004.

[25] A. Schumacher. Stereophotogrammetrische wellenaufnahmen. wiss ergeb. dtsch. atlant. exped. forschungs vermessung. meteor 1925-1927. *Ozeanographische Sonderuntersuchungen, Erste Lieferung,*, 1939.

[26] Steven M. Seitz, Brian Curless, James Diebel, Daniel Scharstein, and Richard Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 519–528, Washington, DC, USA, June 2006. IEEE Computer Society.

[27] O.H. Shemdin and H.M. Tran. Measuring short surface waves with stereography. *Photogrammetric Engineering and Remote Sensing*, 58(311-316), 1992.

[28] O.H. Shemdin, H.M. Tran, and S.C. Wu. Directional measurements of short ocean waves with stereography. *J. Geophys. Res.*, 93:13891–13901, 1988.

[29] R. Siegmund, M. Bao, S. Lehner, and R. Mayerle. First Demonstration of Surface Currents Imaged by Hybrid Along- and Cross-Track Interferometric SAR. *IEEE Transactions on Geoscience and Remote Sensing*, 42:511–519, March 2004.

[30] S. Soatto, A. Yezzi, and H. Jin. Tales of shape and radiance in multi-view stereo. In *ICCV*, pages 974–981, 2003.

[31] D. Stagonas and G. Muller. Wave field mapping with particle image velocimetry. *Ocean Engineering*, 34(11-12):1781–1785, 2007.

[32] Justin M. Wanek and Chin H. Wu. Automated trinocular stereo imaging system for three-dimensional surface wave measurements. *Ocean Engineering*, 33(5-6):723–747, April 2006.

[33] A. Yezzi and S. Soatto. Stereoscopic segmentation. *International Journal of Computer Vision*, 53(1):31–43, June 2003.

[34] Vladimir E. Zakharov. Statistical theory of gravity and capillary waves on the surface of a finite-depth fluid. *Eur. J. Mech. B - Fluids*, 18(3):327–344, 1999.

Guillermo Gallego is a Ph.D. student at the School of Electrical and Computer Engineering at the Georgia Institute of Technology. He majored in Telecommunications Engineering in 2004 at the E.T.S. Ingenieros de Telecomunicacion of the Universidad Politecnica of Madrid, Spain. In July 2005 he earned a Master's degree in Mathematical Engineering (Magister en Ingenieria Matematica) from the Universidad Complutense de Madrid. In 2005, he was a recipient of the Fulbright Scholarship to pursue graduate studies at the Georgia Institute of Technology. At this institution, he obtained an M.S. in Electrical and Computer Engineering in 2007, an M.S. in Mathematics in 2009. His research interests fall within the areas of geometry, optimization, image processing, computer vision and ocean engineering.

Professor Yezzi was born in Gainsville, Florida and grew up in Minneapolis, Minnesota. He obtained both his Bachelor's degree and his Ph.D. in the Department of Electrical Engineering at the University of Minnesota with minors in mathematics and music. After completing his Ph.D., he continued his research as a post-Doctoral Research Associate at the Laboratory for Information and Decision Systems at Massachusetts Institute of Technology in Boston, MA. His research interests fall broadly within the fields of image processing and computer vision. In particular he is interested in curve and surface evolution theory and partial differential equation techniques as they apply to topics within these fields (such as segmentation, image smoothing and enhancement, optical flow, stereo disparity, shape from shading, object recognition, and visual tracking). Much of Dr. Yezzi's work is particularly tailored to problems in medical imaging, including cardiac ultrasound, MRI, and CT. He joined the Georgia Tech faculty in the fall of 1999 where he has taught courses in DSP and is working to develop advanced courses in computer vision and medical image processing. Professor Yezzi consults with industry in the areas of visual inspection and medical imaging. His hobbies include classical guitar, opera, and martial arts.

Dr. Francesco Fedele is an Assistant Professor in the School of Civil and Environmental Engineering and he has an adjunct appointment in the School of Electrical and Computer Engineering at Georgia Tech. He received his Ph.D. in Civil Engineering from the University of Vermont in 2004 and earned his Laurea (magna cum laude) in Civil Engineering from the University Mediterranea, Italy, in 1998. Dr. Fedele joined the faculty at Georgia Tech in 2007 after a postdoctoral research position at the NASA Goddard Space Flight Center. His current research focus is on nonlinear wave phenomena, fluid mechanics, sustainable ocean energy, computational methods and inverse problems. The corresponding research thrusts are: wave turbulence and rogue waves, algorithms for biomedical tomographic imaging, mathematical modeling and experimentation on renewable devices to harness energy from tidal streams.

Dr. Alvise Benetazzo is Senior Engineer researcher PROTECNO S.r.l. Padua, Italy. He received his Ph.D. in Civil and Environmental Engineering from the University of Padua, Italy in 2006 and earned his Laurea in Civil Engineering from the same Italian University in 2000. Before joining PROTECNO in 2006, Dr. Benetazzo was a visiting engineering researcher at Scripps Institution of Oceanography, University of California, San Dieg, USA. His expertise is in mathematical and physical models of water phenomena, measurements and analysis of water waves by using remote sensors, and remote sensing models for detecting illegal land. While at PROTECNO, He has been involved in all the aspects of both mathematical and physical modeling of the MOSE structure at the Chioggia and Treporti inlets in Venice Italy.

LIST OF FIGURES

Figure 1. Left: off-shore platform "Acqua Alta" in the Northern Adriatic Sea, near Venice. Center: pair of synchronized cameras for monitoring the ocean climate from the platform. Right: WASS hardware installed at the platform for recording stereo videos of ocean waves.

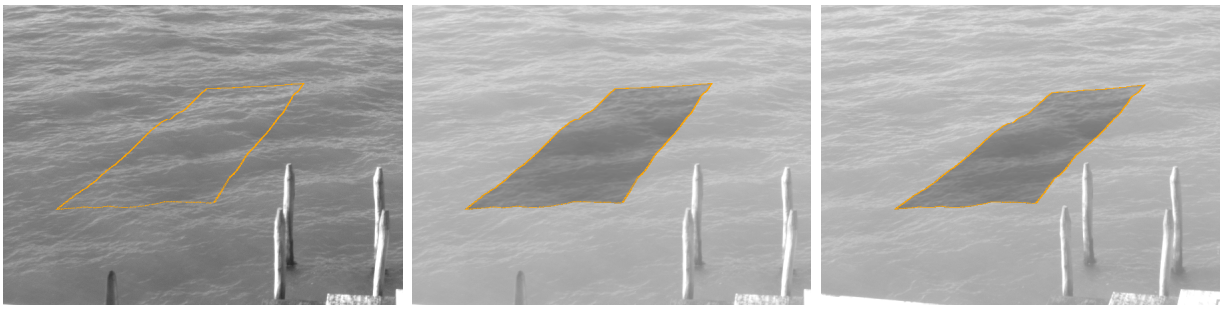

Figure 2. Experiment I (Venice). Left: projection on image 1 of the boundary of the estimated graph, which has been discretized by a grid of 129×513 points. Center: modeled image (computed form surface height and radiance) superimposed on original image 1. Right: modeled image 2 superimposed on original image 2.

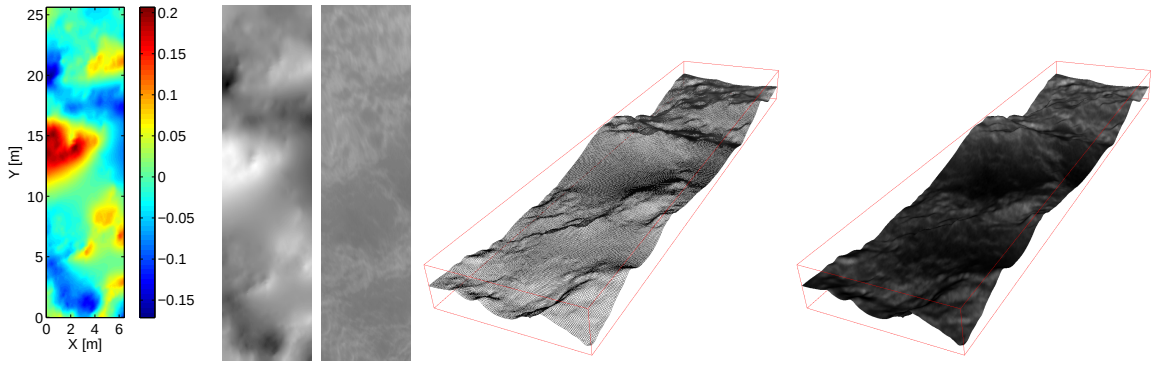

Figure 3. Experiment I (Venice). Form left to right: (1) estimated height function $Z(u, v)$ (shape of the water surface) in pseudo-color; (2) height represented by greyscale intensities, from dark (low) to white (high); (3) estimated radiance function $f(u, v)$ (texture on the surface); (4) perspective, three-dimensional wire-frame representation of the estimated surface shape (height) according to grid points; (5) texture-mapped surface obtained by incorporating the radiance function in the wire-frame model. In (4) and (5) the vertical axis has been magnified by a factor of 5 with respect to the horizontal axes for visualization purpose.

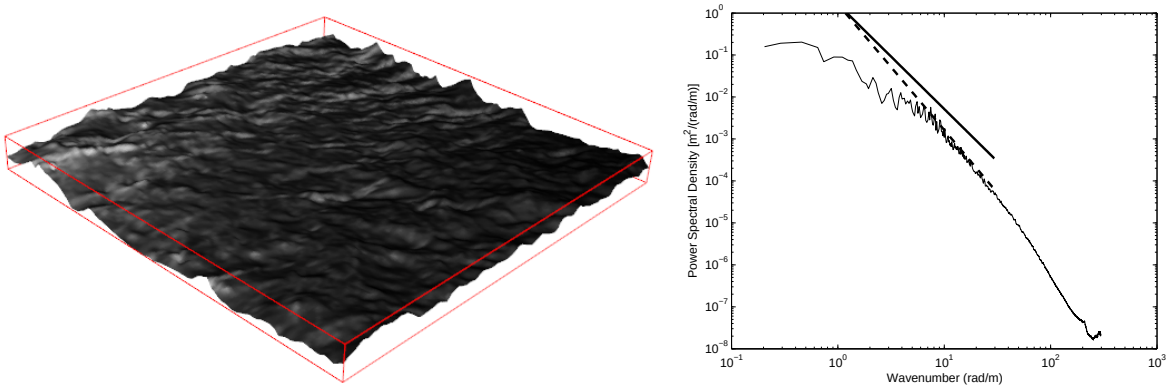

Figure 4. Experiment II (San Diego). Left: Surface reconstruction from a snapshot of the data in [2]. Right: Average omni-directional wave number spectrum. Straight lines:$k^{-2.5}$ (solid), k^{-3} (dashed).

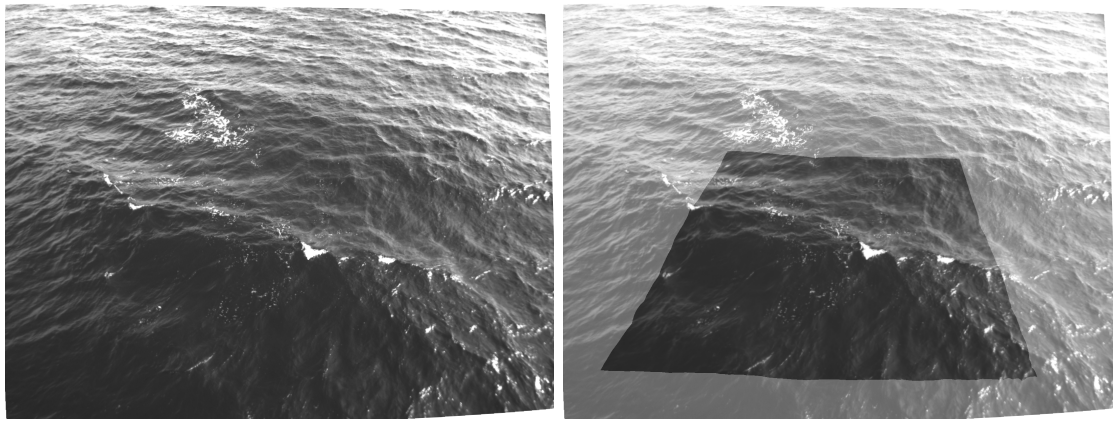

Figure 5. Experiment III (Crimea): Original image (left) and modeled image superimposed on original image (right).

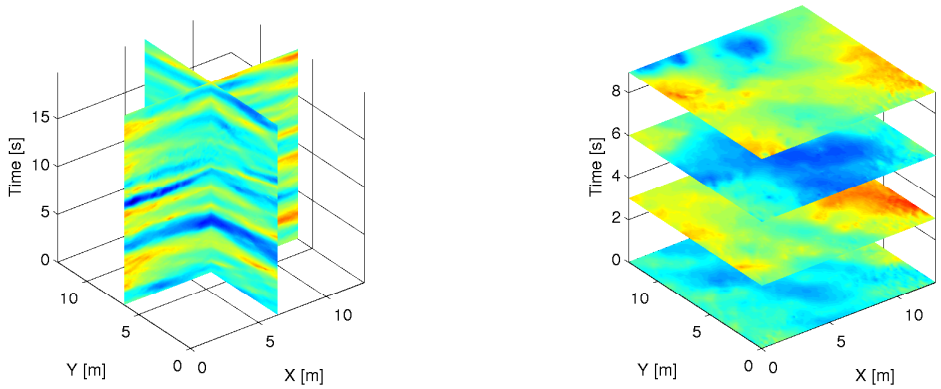

Figure 6. Experiment III (Crimea). Vertical (left) and Horizontal (right) slicing of the wave space-time volume $Z(x, y, t)$: reconstructions for fixed x-t or y-t planes. The wave heights Z have been pseudo-colored from blue (low) to red (high).

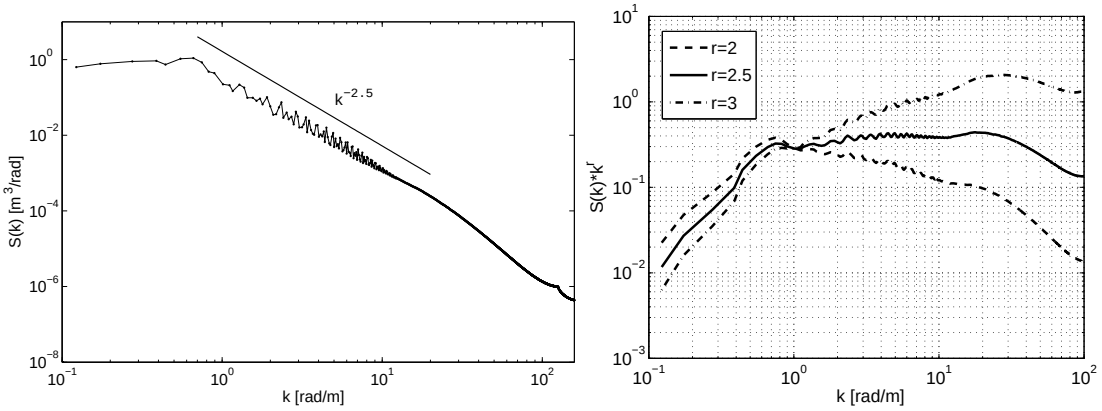

Figure 7. Experiment III (Crimea). Left: Mean omni-directional spectrum $S(k)$ averaged over 2000 snapshots. Right: mean saturation spectrum $S(k)k^r$ for $r = \{2, 2.5, 3\}$.

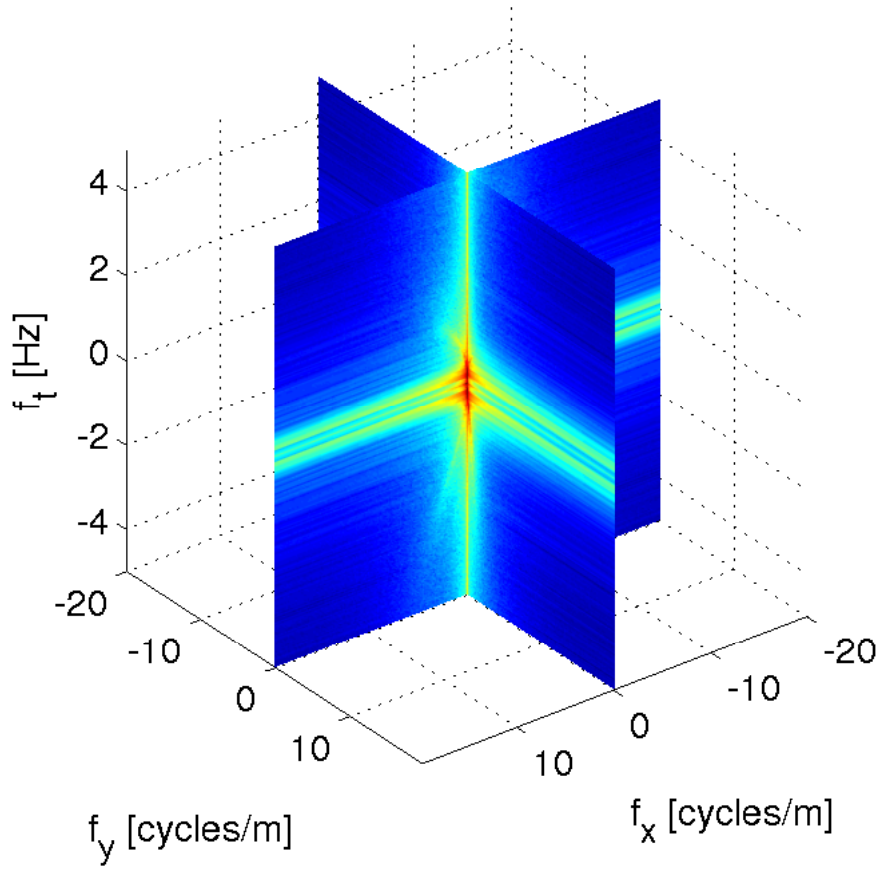

Figure 8. Experiment III (Crimea). Two orthogonal slices of the wave 3-D spectrum $\mathcal{Z}(k_x, k_y, \omega)$ through the frequency origin. Slices correspond to the values of $\mathcal{Z}(k_x, k_y, \omega)$ at planes $k_x = 0$ and $k_y = 0$. Axes f_x, f_y and f_t stand for $k_x/(2\pi)$, $k_y/(2\pi)$ and $\omega/(2\pi)$, respectively.

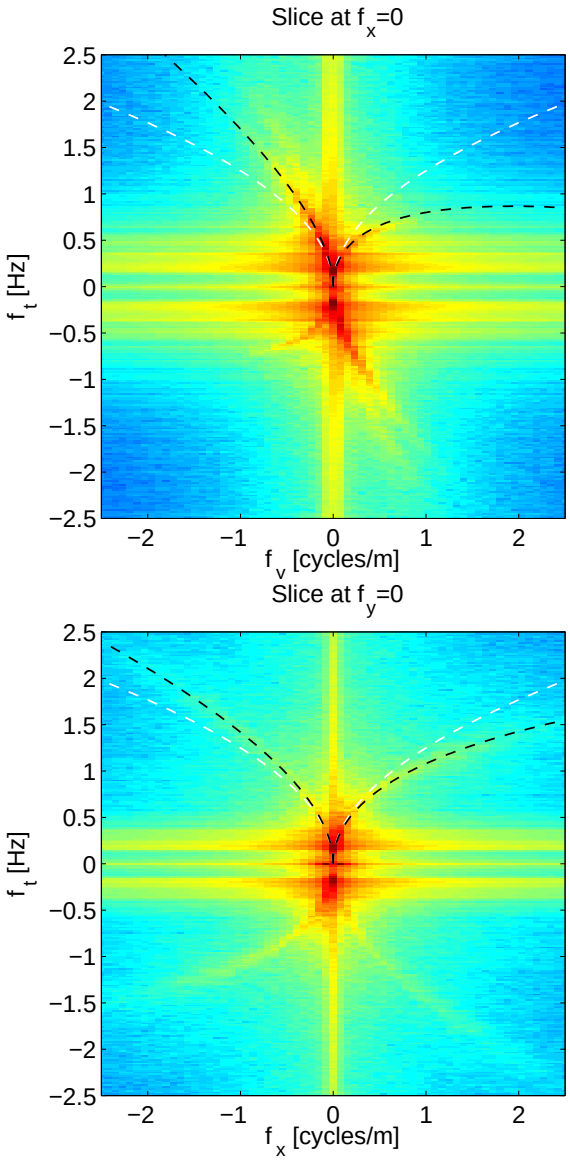

Figure 9. Experiment III (Crimea). Vertical slices of the 3-D wave spectrum at frequencies $k_x = 0$ (top) and $k_y = 0$ (bottom). Superimposed on top half of both plots: (white curve) vertical slice of the linear wave dispersion manifold $|\mathbf{k}| = \omega^2/g$, with $\omega = 2\pi f_t$, and (black curve) vertical slice of the wave-current dispersion manifold $|\mathbf{k}| = (\omega - \mathbf{k}\cdot\mathbf{v})^2/g$, with $\mathbf{v} \approx (-0.17, -0.45)$ m/s. Axes f_x, f_y and f_t stand for $k_x/(2\pi)$, $k_y/(2\pi)$ and $\omega/(2\pi)$, respectively.

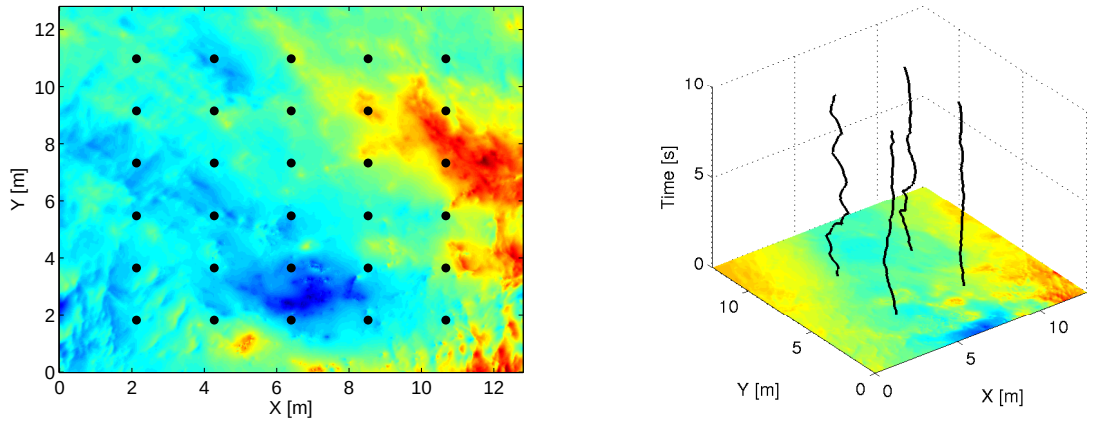

Figure 10. Experiment III (Crimea). Left: Location of the virtual probes. Right: Illustration of extracted time series at probe points within the space-time volume $Z(x, y, t)$.

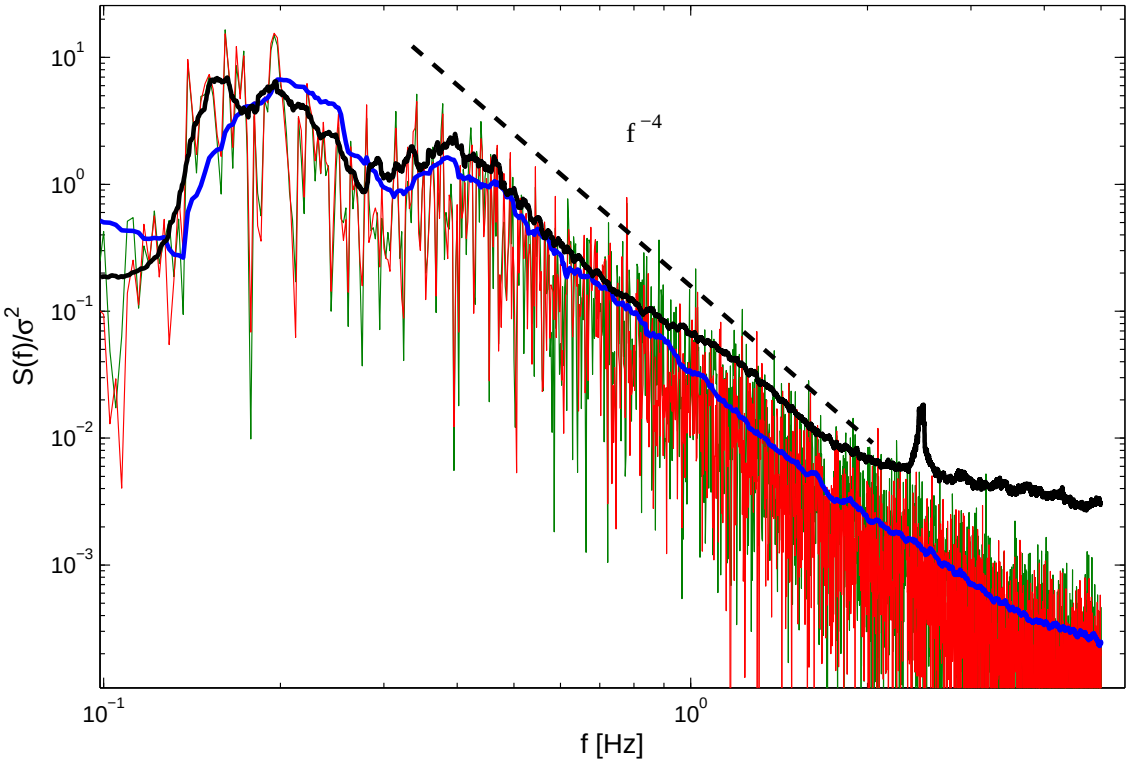

Figure 11. Experiment III (Crimea). Normalized frequency spectrum (σ^2 is the variance of the wave surface) averaged over all virtual probes (blue line) and estimated counterpart using classical epipolar method (black line). Note that the Nyquist frequency (half of the sampling frequency) is 5 Hz, according to the snapshot (e.g. frame) rate.

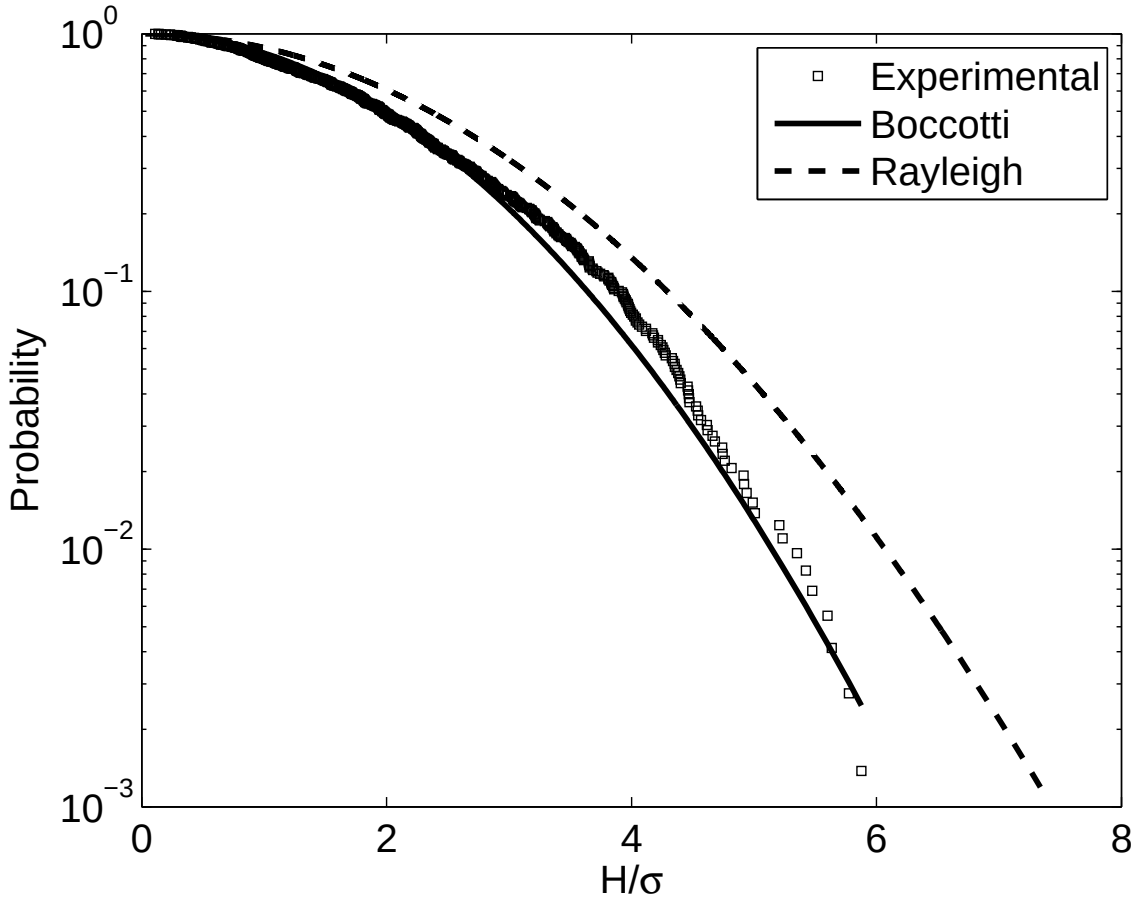

Figure 12. Experiment III (Crimea). Wave height exceedance probability estimated from all time series at virtual probes, compared to Rayleigh's distribution and Boccotti's distribution (31) (σ is the standard deviation of the wave surface).