



Research paper

Multi-view horizon-driven sea plane estimation for stereo wave imaging on moving vessels

Filippo Bergamasco^{a,*}, Alvisè Benetazzo^b, Francesco Barbariol^b, Sandro Carniel^b, Mauro Sclavo^b^a DAIS, Università Ca' Foscari, Venice, Italy^b Institute of Marine Sciences, Italian National Research Council (ISMAR-CNR), Venice 30122, Italy

ARTICLE INFO

Article history:

Received 31 January 2016

Received in revised form

21 July 2016

Accepted 22 July 2016

Available online 25 July 2016

Keywords:

Stereo vision

Horizon estimation

Sea surface reconstruction

3D wave field

ABSTRACT

In the last few years we faced an increased popularity of stereo imaging as an effective tool to investigate wind sea waves at short and medium scales. Given the advances of computer vision techniques, the recovery of a scattered point-cloud from a sea surface area is nowadays a well consolidated technique producing excellent results both in terms of wave data resolution and accuracy. Nevertheless, almost all the subsequent analyses tasks, from the recovery of directional wave spectra to the estimation of significant wave height, are bound to two limiting conditions. First, wave data are required to be aligned to the mean sea plane. Second, a uniform distribution of 3D point samples is assumed. Since the stereo-camera rig is placed tilted with respect to the sea surface, perspective distortion do not allow these conditions to be met. Errors due to this problem are even more challenging if the optical instrumentation is mounted on a moving vessel, so that the mean sea plane cannot be simply obtained by averaging data from multiple subsequent frames. We address the first problem with two main contributions. First, we propose a novel horizon estimation technique to recover the attitude of a moving stereo rig with respect to the sea plane. Second, an effective weighting scheme is described to account for the non-uniform sampling of the scattered data in the estimation of the sea-plane distance. The interplay of the two allows us to provide a precise point cloud alignment without any external positioning sensor or rig viewpoint pre-calibration. The advantages of the proposed technique are evaluated throughout an experimental section spanning both synthetic and real-world scenarios.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

In the recent years there has been a growing interest in remote and proximal observation of sea surface waves. New classes of instruments (e.g. radars, laser scanners, optical) have been placed aboard satellites, airplanes, or ships at sea, facing the need for new processing tools for specific tasks. In this context, observations of sea surface fields by means of optical-based stereo systems, after the pioneering work of Schumacher (1939), Holthuijsen (1979), and Banner et al. (1989), are nowadays becoming accessible to a large number of scientists given the advances in the processing of digital images (Kosnik and Dulov, 2011; Gallego et al., 2011; Benetazzo et al., 2012, 2015). The efforts made to develop both hardware and software for stereo processing has allowed stereo wave imaging to become a well consolidated and accurate tool for wave observations, that indeed is providing new insights into the

field of sea surface waves (Banner et al., 2014; Leckler et al., 2015; Yurovskaya et al., 2013). So far, snapshots of the sea surface have been mostly taken from fixed platforms at sea, as this condition greatly eases the cameras deployment and the image processing. Indeed, the time-constant rigid-motion that places all the acquired surfaces into a common geo-referenced frame can be estimated a priori exploiting the statistical nature of the waves (Benetazzo et al., 2016). However, this design for stereo systems is starting to become too restrictive for oceanographers, who need to collect wave data in as many different conditions and locations as possible, eventually where the use of a fixed platform is unfeasible. Thus, first applications of stereo systems mounted on oceanographic vessels is taking place, facing the nontrivial task of accounting for the camera motion during the stereo processing (Brandt et al., 2010).

In this study we present an analysis of spatio-temporal 3D wave fields that take advantage of a multi-view horizon estimate approach for compensating ship motion. The paper is organized as follows. We start with an overview of the related work in the field of horizon estimation, highlighting the novelties of our proposed

* Corresponding author.

E-mail address: filippo.bergamasco@unive.it (F. Bergamasco).

approach with respect to the literature. Then, in [Section 2](#) we describe the stereo system used to reconstruct the sea surface. In [Section 3](#) we present our multi-view horizon estimation approach to recover the sea surface plane normal. Then, in [Section 4](#) we describe how to use the recovered normal to estimate a homography relating to the camera and sea plane. Such homography is used to define a spatial density estimator to weight the 3D point samples and account for perspective distortion. This allows a robust recovery of plane distance and, consequently, a precise alignment of the scattered point data. An extensive experimental part ([Section 5](#)) demonstrates the strengths of our method comparing with both synthetic and real-world data acquired in a controlled environment (i.e. a fixed oceanographic platform). Furthermore, a use-case is presented to show the behaviour of the approach “in the wild” while investigating the wave statistics on a moving ship during a cruise. Finally, in [Section 6](#) we draw some concluding considerations.

1.1. Related work

Being able to recover the attitude and position of a moving vehicle subject to 6 degrees of freedom has been a topic of pivotal importance for both aerial and marine operations. With the former, the recent availability of inexpensive unmanned aerial vehicles to be used in dangerous or remote situations raised the need to improve the automatic control and stabilization beyond the accuracy obtainable with purely inertial sensors. In the latter, the precise estimation of ship motion is an aid for navigation safety ([Liu et al., 2008](#)), vessel detection ([Fefilatyeu et al., 2007](#)) or environmental sciences ([Williams and Howard, 2011](#)).

In this plethora of different applications, vision based approaches have gained an increased popularity due to their potential ability to hinge the estimation on well-localizable visual features occurring on the landscape. Among all, the horizon line is usually the weapon of choice as is expected being always visible if no other assumptions can be made to the application scenario. Even if seeing the horizon is not a sufficient condition to recover the full 6 degrees of freedom of a vessel (the estimation is recovered up to any rotation around an arbitrary axis orthogonal to the ground plane), position can be usually obtained with enough accuracy through GPS sensors and is less critical with respect to attitude recovery. In fact, by constraining to a physically plausible kinematics of a ship, statistical filtering methods can improve the accuracy of common marine GPS to tens of centimeters. Given the importance of the horizon as a visual feature, sophisticated techniques have been proposed to provide a precise estimation even when the acquired images are affected by cluttering objects or optical artifacts (sun-glares, fog, low-contrast, etc.).

Following the taxonomy introduced by [Shabayek et al. \(2012\)](#), horizon line estimation methods are essentially divided into two categories. In the former, a Sky/Ground segmentation is performed by means of statistical models over the feature space being either color ([Gallagher et al., 2004](#)), textures ([Cherian et al., 2009](#); [Schwendeman and Thomson, 2015](#)), polarization ([Shabayek et al., 2012](#)) or a combination of the previous ([Todorovic et al., 2003](#)). In the latter, the horizon line is directly estimated by extracting significant image edges (usually after a sequence of multiple filtering stages) and a statistically robust aggregation of contour segments composing the horizon or localizing vanishing points. For instance, [Wang et al. \(2009\)](#) propose an horizon extraction method from ocean observations by weighting the color channels to enhance sky and sea signals while reducing the effect of noise and highlights. Thereafter, contours are extracted with Canny edge detection and the predominant line clustered with a classical Hough transform. Similarly, [Williams and Howard \(2011\)](#) describe a rather sophisticated sequence of adaptive histogram thresholding,

region growing and Gaussian mixture model to preprocess the images before binarization and contour extraction.

When horizon estimation is used to control an UAV, temporal coherency can be exploited to aid the extraction considering multiple video frames ([Bao et al., 2003](#)). Moreover, optical flow can be extracted between subsequent images as an aid to cluster horizon edges from spurious clutter arising from natural or human-made features on the landscape ([Dusha et al., 2007](#)). When an inertial measurement unit (IMU) is available, its data can be used as an initial prior to limit the image analysis on a reduced area of the image ([Hugues et al., 2014](#)). Furthermore, [Angelino et al. \(2013\)](#) suggest the usage of an Unscented Kalman Filter (UKF) to describe the dynamic system non-linear equations to effectively fuse attitude information coming from GPS, IMU and cameras.

If stereo vision is available, the complete pose can be obtained as a natural consequence of multi-view triangulation. For instance, [Wang et al. \(2005\)](#) use stereo vision to recover the full position and attitude of a UAV flying near the ground whereas [Eynard et al. \(2010\)](#) use a hybrid stereo system composed of a central and a fish-eye camera to measure the altitude without any external sensor. However, stereo reconstruction can only be used for pose estimation if we assume a non-moving set of landmarks being present in a scene. This is obviously not the case when observing the continuously changing sea surface.

1.2. Contributions

In this paper we propose a novel horizon estimation technique to recover the stereo rig attitude with respect to the sea surface. Once the relative orientation of the rig and the sea plane is assessed (up to a rotation around the plane normal, that cannot be inferred by the horizon alone), its distance is estimated in a robust way that compensates the non-uniform distribution of all the reconstructed point samples.

With respect to the aforementioned existing approaches, our method introduces several novelties. To our knowledge, it presents the first intrinsically multi-view approach being able to fuse the horizon line estimates coming from different images. In our case, just the two images of the stereo rig are used but the method itself is flexible enough to comprise many different (pre-calibrated) image sources. Second, we use a prior given by the 3D reconstruction as a first guess to refine the estimated horizon line. In this sense, our method is more like an horizon refinement approach rather than a full estimator being able to locate the horizon anywhere in the image. Nevertheless, this prior allows us to pose the problem as an energy optimization spanning the entire image domain and avoiding voting schemes such as the commonly used Hough Transform. This both improves the accuracy of the method and reduces the computational effort. Third, our method is not based on a preliminary edge extraction step. Instead, it directly operates on image pixels in a way similar to active contours ([Kass et al., 1988](#)). This has the double benefit of getting rid of multiple edge extraction parameters and allowing a better line localization since no information is lost due to binarization.

The horizon line allows to recover the parameters defining the normal of the sea plane. This is of crucial importance for the alignment of the reconstructed data, that can be directly rotated so that the sea plane normal coincides with the z-axis. Also, we can exploit the relative angles to correct for perspective distortion in an optimal way to account for the non-uniform sampling of the reconstructed points with respect to the sea plane. This non-uniformity is typical of all the stereo methods that operate by matching corresponding pixels directly on the image space. These methods are the only feasible on a moving scenario (like the one investigated in this work) for which the pose of the cameras from the observed scene cannot be reliably provided a priori. On the

contrary, when the pose can be properly calibrated, methods that directly operate on the object space like the variational approach in Gallego et al. (2008) can estimate a smooth spatially uniform triangulated surface at a cost of being computationally intensive for long sequences.

2. The WASS pipeline

Before proceeding with the description of the sea plane alignment technique, we give a brief description of the 3D reconstruction machinery and algorithms used in this work. Our Wave Acquisition Stereo System (WASS) is essentially composed of two digital cameras firmly placed side-by-side looking toward the sea surface. Each camera is equipped with a 5.0 mm low radial distortion lens and a CCD with a pixel size of 3.45 μm producing images with a resolution of 2048 \times 2456 pixels. An electronic trigger ensures the synchronized acquisition of the frame pairs at a frequency that can be adjusted from 1 to 15 Hz. The internal properties of the lenses (i.e. focal-length, aperture, etc) are configured during the installation and therefore remain constant during the acquisition.

Both the cameras composing the stereo rig are tilted with respect to the sea surface so that the horizon is visible in the upper part of each image. This setup is quite common for this type of instrumentation as it offers a good trade-off between the size of the reconstructed area and the height of the stereo rig with respect to the sea, which is generally bounded by the characteristic of the ship.

Similarly to the system presented by Benetazzo et al. (2014), a target-less calibration technique allow us to recover the relative orientation (i.e. extrinsic parameters) of the two cameras whereas the focal length, optical center and the polynomial radial distortion parameters are estimated in the lab with a calibration target (Albarelli et al., 2010). After the intrinsic properties of the two cameras and their relative position is calibrated, stereo geometry can be used to obtain a dense reconstruction of the sea surface with respect to the reference frame of one of the two cameras. The outcome of the reconstruction process is a scattered cloud of 3D points which is further refined to filter out possible spikes and high-frequency noise.

Even if the produced cloud is more than adequate to extract valuable wave statistics, it cannot be used as is. Indeed, it has to be firstly aligned to a *sea reference frame* with the z-axis pointing upward the sea surface and then uniformly resampled into a regular surface grid so that two-dimensional spectral representation can be computed with FFT. As described by Gallego et al. (2008), the rotation \mathbf{R}_w and translation \vec{T}_v that align the reconstructed data can be obtained by the sea plane $n_x X + n_y Y + n_z Z + d = 0$ as

$$\mathbf{R}_w = \begin{pmatrix} 1 - \frac{(1 - n_z)n_x^2}{n_x^2 + n_y^2} & -\frac{(1 - n_z)n_x n_y}{n_x^2 + n_y^2} & -n_x \\ -\frac{(1 - n_z)n_x n_y}{n_x^2 + n_y^2} & 1 - \frac{(1 - n_z)n_y^2}{n_x^2 + n_y^2} & -n_y \\ n_x & n_y & n_z \end{pmatrix}, \quad \vec{T}_v = \begin{pmatrix} 0 \\ 0 \\ d \end{pmatrix}. \quad (1)$$

Notwithstanding the importance of obtaining a good estimate of sea plane parameters for all the subsequent wave analysis, this topic has been commonly sorted out in the literature by a simple linear least square fitting (LLS) of the plane to the point cloud samples. Given the quasi-Gaussian nature of sea waves elevation, this is a quite good approximation provided that a significant number of samples is given. This is true if the cameras are fixed (for instance, when mounted on a non-floating offshore platform),

so that the mean plane can be obtained by averaging the reconstructed clouds over time. In many cases this assumption does not hold. The data produced by a single stereo frame (i.e. at a specific time) is in general not sufficient to obtain a good estimate if the mean wavelength is similar to the size of the reconstructed area (Benetazzo et al., 2016). It is easy to imagine that, if we consider a scenario in which only a portion of a high wave is visible to the camera, the estimated plane would necessarily follow the slope of the surface. This can be even worse if the wave is moving towards the camera so that the falling part of the wave is occluded by the rising. This problem is usually solved by averaging all the planes obtained by many subsequent frames (Benetazzo et al., 2012) but this option is only viable if the camera rig position and attitude remain constant throughout the acquisition and the acquired area comprises at least 4×4 wavelengths (Benetazzo et al., 2016).

Second, least squares estimates are not robust to outliers or heteroscedastic errors (Seber and Lee, 2012). The former may rise from wrongly matched point correspondences during the stereo process. The latter is naturally occurring since the rig optical axis is not orthogonal to the sea plane. As a consequence, point cloud is denser and with less error closer to the camera than far away. Furthermore, closer points tend to be better localized along the xy-axes than on the z whereas, for distant points, the vice versa is valid.

Finally, even if we ignore reconstruction errors, the non-uniform spatial density of the point cloud produces an implicit different weighting of samples with respect of their distance. In Benetazzo et al. (2016) this is partially accounted considering a weighted linear regression whose weight is proportional of the point Z-coordinate expressed in camera world. This workaround partially mitigates the problem, although it was not supported by a theoretical motivation.

3. Sea plane normal estimation

The first step to align the reconstructed stereo data is to estimate the normal $\vec{n} = (n_x \ n_y \ n_z)^T$ of the sea plane φ with respect to the camera reference frame. We start by considering the case of a single camera oriented toward the surface so that the image plane is not parallel to φ (the trivial case in which they are parallel would not produce significant perspective distortion). In this context, the second camera of the rig is just used to perform 3D triangulation so that the reconstructed point cloud lies onto the first camera reference frame. Finally, we assume that the radial distortion has been calibrated, producing two undistorted images I_1 and I_2 for each frame together with the intrinsic camera matrices \mathbf{K}_1 and \mathbf{K}_2 and the rigid transformation \mathbf{R}_{21} and \vec{T}_{21} describing the relative motion that transforms the second camera reference frame onto the first.

We approximate the Earth surface as being planar so that the horizon line is located at the infinity. Since the Earth average radius is orders of magnitude greater than our reconstructed area (which is less than 100×100 m), this simplification allows us to use projective geometry concepts to compute \vec{n} while still producing accurate results. Indeed, if we suppose no atmospheric refraction, the Earth being spherical (with a radius of $r_e = 6371$ km) and a camera placed h meters above the sea level, we can derive the horizon distance h_d as

$$h_d^2 = (r_e + h)^2 - r_e^2 h_d = \sqrt{2r_e h + h^2} \approx \sqrt{2r_e h} = 3570 \sqrt{h} \quad (2)$$

see Fig. 1(a) for details. As a consequence, horizon distance is approximately 3570 times the square root of the camera height h ,

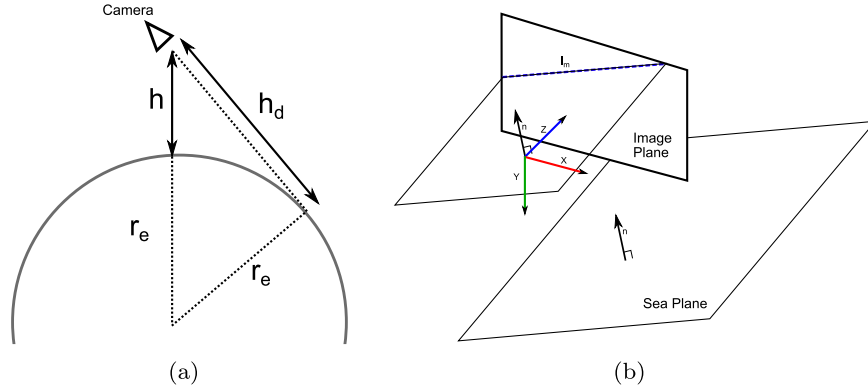


Fig. 1. (a) With some simplifications, the horizon distance can be computed from the observer height h and the Earth radius r_e . (b) Geometrical relations between the sea plane, the image plane, the horizon and a hypothetical plane parallel to the sea plane but passing through the optical center.

which is roughly 15 m above the sea level in a common WASS scenario. This places the horizon line ~ 14000 m away from the camera which is roughly 7000 times our baseline distance. Considering the optics and resolution of our cameras, this would produce a disparity of ~ 0.05 pixels between the two projections of an hypothetical point lying at the horizon, which is less than the accuracy we can pursue with the dense stereo estimation (an optimistic estimate is about 1/10 of a pixel).

Every plane in an affine 3-space intersects the plane at infinity in a unique line (Pedoe, 1970). Therefore, the intersection of the sea plane φ with the plane at infinity π_∞ would result in a line ℓ_∞ placed itself at infinity. Since the camera is calibrated and there exists a bijection between π_∞ and the image plane (Hartley and Zisserman, 2004), ℓ_∞ can be projected to produce a line ℓ_H (onto the image plane) called horizon. Algebraically, we can represent ℓ_H as a vector in \mathbb{P}^{2*} so that every image point $(x, y)^T$ (in pixels) lying on ℓ_H satisfies the equation $\ell_H^T(x \ y \ 1)^T = 0$.

Moreover, since every set of parallel planes would intersect into the same line at infinity (Hartley and Zisserman, 2004), to estimate \vec{n} we can restrict the analysis to a plane φ_0 which is parallel to φ but passing through the camera optical center O (Fig. 1(b)). This actually makes sense since we expect that the apparent position of the horizon on the image plane remains fixed if we just translate the camera without changing its orientation.

We parametrize φ_0 as the vector $(n_x \ n_y \ n_z \ d)^T \in \mathbb{P}^{3*}$ so that, for any projective point $p = (x, y, z, w)^T$ lying on φ_0 , the equation

$$\varphi_0^T p = 0 \quad (3)$$

holds. It is easy to demonstrate that:

$$\ell_H = \alpha \vec{n}^T \mathbf{K}_1^{-1} \quad (4)$$

where α is any non-zero scale factor. Indeed, let $p \in \mathbb{P}^3$ be a horizon point in homogeneous coordinates. By definition, p lies on φ_0 and projects on ℓ_H . So, p is a point satisfying:

$$\begin{cases} \varphi_0^T p = 0 & (a) \\ \ell_H^T \mathbf{K}_1(\mathbf{R}_1 \vec{T}_1) p = 0 & (b) \end{cases} \quad (5)$$

where \mathbf{R}_1 is the 3×3 identity matrix and $\vec{T}_1 = (0, 0, 0)^T$ since we are working on the reference frame of the first camera. By comparing Eqs. (5) we get:

$$\ell_H^T \mathbf{K}_1(\mathbf{R}_1 \vec{T}_1) p = \alpha \varphi_0^T p \quad (6)$$

$$\ell_H^T \mathbf{K}_1(\mathbf{R}_1 \vec{T}_1) p = \alpha \varphi_0^T p \quad (7)$$

$$\ell_H^T \mathbf{K}_1 = \alpha (n_x \ n_y \ n_z)^T \quad (8)$$

$$\ell_H = \alpha \vec{n}^T \mathbf{K}_1^{-1} \quad (9)$$

where α is any non-zero scale factor and Eq. (8) was obtained by dropping the last component since both d and w are always 0 (for the former, $d=0$ since φ_0 passes through the origin. For the latter, $w=0$ since $p \in \varphi_0$).

So, from (4), it is easy to recover the sea plane normal \vec{n} (expressed in the first camera reference frame) by first extracting the horizon line ℓ_H from \mathbf{I}_1 , then multiplying ℓ_H by \mathbf{K}_1 and finally normalizing the result to fix the scale to unitary norm. Since knowing two points is a sufficient condition to uniquely describe a two-dimensional line, one way to recover ℓ_H is to extract two vanishing points from a set of coplanar parallel lines in the image (Zhang and Koch, 2013). However, since we cannot exploit properties of parallel lines given the wavy nature of the sea surface, we make sure that the horizon line is always visible in the acquired frames and pose the problem of estimating \vec{n} as the problem of finding the best possible ℓ_H according to the image data.

3.1. Accurate horizon line fitting

The vast majority of techniques to extract salient linear features from images comprises a pre-processing step to extract significant edges from the boundaries of objects composing the scene. Then, with some pixel labelled as “edges” by a threshold or a more sophisticated classification, a clustering operation groups them maximizing their fitness to a linear model.

Edge extraction usually takes place by analyzing image partial derivatives. Indeed, it is natural to think of an edge as a local maximum with respect to the rate of change of pixel color or intensity. One example of this approach is the so popular “Canny edge detector” (Canny, 1986). Robust grouping of edges into lines can be performed with a RANSAC approach or a voting scheme (Duda and Hart, 1972).

Given the nature of our problem, we propose to tackle the horizon estimation from a different point of view. First, we observe that it is not necessary to extract all the significant lines from an image since we can easily obtain an initial horizon guess by fitting a plane to the scattered point cloud. This has been proved to be an acceptable approximation whenever the data cannot be averaged

among multiple frames (Benetazzo et al., 2014). Second, classifying a pixel of being either edge or not requires a tricky tuning of some image-dependent parameters and discards useful information like the direction of the image gradient. This can be even worse considering that the sharpness of the horizon line is affected by atmospheric conditions that may result in a over-smoothed perceived transition between sea plane and the sky.

Let $I(s) \in \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ be the image function giving the intensity (i.e. brightness) of the acquired scene at a pixel $s = (u, v)$. We formulate the horizon line estimation as the following constrained minimization problem:

$$\begin{aligned} \operatorname{argmax}_{\ell_H} \int_{\Omega} e^{-\frac{(s^T \ell_H)^2}{\sigma}} \nabla I(s)^T \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \\ \ell_H ds \text{ subject to } \ell_H^T \\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \ell_H \\ = 1 \end{aligned} \quad (10)$$

Before giving a possible numerical solution to the proposed optimization, we would like to spend some words to explain the motivations of our formulation. Ignoring the constraints, what we are trying to maximize in (10) is essentially the correlation between the image gradient $\nabla I(s)$ and a function giving the response (in terms of its partial derivative) of the line feature described by ℓ_H . Indeed, the function $f_{\ell_H}: \Omega \rightarrow \mathbb{R}^2$

$$f_{\ell_H}(p) = e^{-\frac{(p^T \ell_H)^2}{\sigma}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \ell_H \quad (11)$$

defines a vector field over Ω where each vector is orthogonal to the line ℓ_H with a magnitude inversely proportional to the distance between p and the line (Fig. 2, left). Similarly, the vector field $\nabla I(s)$ would exhibit a similar structure, with a strong gradient orthogonal to the horizon line fading toward the quasi-uniform brightness of the sky (Fig. 2, right). As a consequence, the scalar product of the two would result in a higher correlation the more ℓ_H is describing a line near the horizon. With our formulation, we explicitly consider the orientation of ℓ_h normal vector as we assume that the sea always appear darker than the sky. Additionally, since the distance between a point s and ℓ_H is scaled by the norm of the first two components of ℓ_H , a constraint is imposed to the optimization to forbid the resulting correlation value to increase just by scaling ℓ_H accordingly (recall that $\beta \ell_H$ is representing the same line for each non-zero scale factor β).

To solve (10) we start by observing that both the optimized function and its constraint have continuous first partial derivatives. Hence, using Lagrange Multipliers we re-write the problem as the unconstrained maximization:

$$\begin{aligned} \operatorname{argmax}_{\ell_H, \lambda} \Lambda(\ell_H, \lambda) \\ = \operatorname{argmax}_{\ell_H, \lambda} \int_{\Omega} e^{-\frac{(s^T \ell_H)^2}{\sigma}} \nabla I(s)^T \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \ell_H ds \\ - \lambda \left(\ell_H^T \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \ell_H - 1 \right) \end{aligned} \quad (12)$$

Setting $\nabla \Lambda = 0$, we obtain the system of equations:

$$\frac{\partial \Lambda}{\partial \ell_H} = -2\ell_H^1 \lambda + \int_{\Omega} E_{s, \ell_H} \nabla_x I(s) - \frac{2}{\sigma} s_x E_{s, \ell_H} s^T \ell_H \nabla I(s)^T U \ell_H ds \quad (13)$$

$$\frac{\partial \Lambda}{\partial \ell_H^2} = -2\ell_H^2 \lambda + \int_{\Omega} E_{s, \ell_H} \nabla_y I(s) - \frac{2}{\sigma} s_y E_{s, \ell_H} s^T \ell_H \nabla I(s)^T U \ell_H ds \quad (14)$$

$$\frac{\partial \Lambda}{\partial \ell_H^3} = - \int_{\Omega} \frac{2}{\sigma} E_{s, \ell_H} s^T \ell_H \nabla I(s)^T U \ell_H ds \quad (15)$$

$$\frac{\partial \Lambda}{\partial \lambda} = 1 - \ell_H^T \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \ell_H \quad (16)$$

where $E_{s, \ell_H} = e^{-\frac{(s^T \ell_H)^2}{\sigma}}$, $s = (s_x, s_y)^T$, $U = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ and $\ell_H = (\ell_H^1, \ell_H^2, \ell_H^3)^T$. Considering the image I as a discrete set of pixels $I_m = \{s_1 \dots s_n\} \subset \mathbb{R}^2$, the integral over Ω is reduced to a summation over I_m . The gradient ∇I_m can be computed numerically by convolving the image with a derivative of Gaussian kernel. At this point, we recover an initial guess for the line ℓ_H by applying Eq. (4) to the (approximate) plane normal estimated from the given point cloud. Then, we optimize iteratively the non-linear least squares fitting problem

$$\operatorname{argmin}_{\ell_H, \lambda} \|\nabla \Lambda(\ell_H, \lambda) - \vec{0}\|^2 \quad (17)$$

via Levenberg–Marquardt (Levenberg, 1944), with the initial value of ℓ_H computed from the point cloud and λ empirically set to 1.¹

In Fig. 3 we show an example of the result before and after the optimization process for the same input image shown in Fig. 2, right. The initial ℓ_H guess (blue line), computed from the 3D point cloud only, lies near the expected horizon line but still not close enough to provide a consistent alignment across multiple frames. After the optimization (red line), the correct orientation and position of the image is recovered even in the presence of high frequency noise, highlights and clutter (for example, the mountains on the left). After the line fitting process, we can go back to the sea plane normal by inverting Eq. (4).

3.2. Dealing with multiple views

Our formulation is posed in terms of a discrete vector field ∇I_m sampled at points $s_1 \dots s_n$. Those samples cover the entire set of pixels of an image but, in practice, we never make use of their regular grid distribution. Unlike methods based on an initial edge detection step, which implies the knowledge of the neighboring pixels of a point, our method would work even if $s_1 \dots s_n$ are not equally distributed or sampled at sub-pixel locations. We take advantage of this to provide multi-view horizon line estimation without modifying the core formulation of our approach.

We suppose to have two discrete images I_{m_1} and I_{m_2} . Since the horizon is assumed to be infinitely distant (i.e. it lies on π_{∞}), and the cameras are calibrated (both in the intrinsic and extrinsic parameters), all the I_{m_2} points can be projected into the first camera image space. This is essentially the same concept exploited by panorama-stitching techniques that compose multiple different images of the same scene into a single ultra-wide image. To produce correct results, panorama-stitching assumes that all the views are created just by rotating the camera around the optical center so that there is no observable disparity between objects. Under this condition, the transformations relating points among

¹ The initial value of λ has been empirically determined so that, for our tests, it was always close enough to the optimum to let the optimization converge to a satisfactory solution.

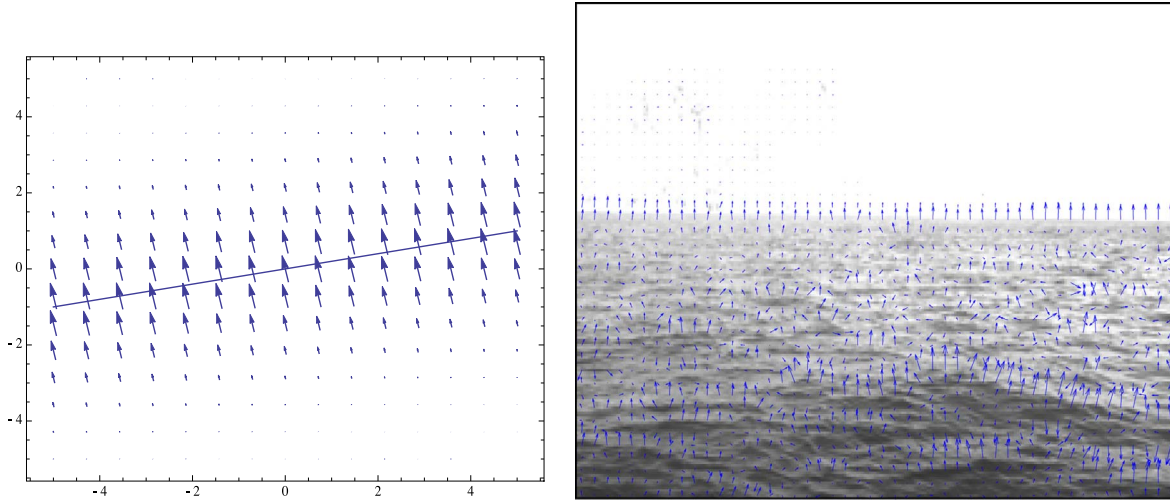


Fig. 2. Left: vector field generated by f_{ℓ_H} for $\sigma = 4$ and $\ell_H = (-\frac{1}{5}, 1, 0)$. All the vectors are orthogonal to the line and with a magnitude inversely proportional to the distance. Right: gradient of a horizon image in our dataset. Note that, since the sea appears darker than the sky, the gradient is oriented upward nearby the horizon line.

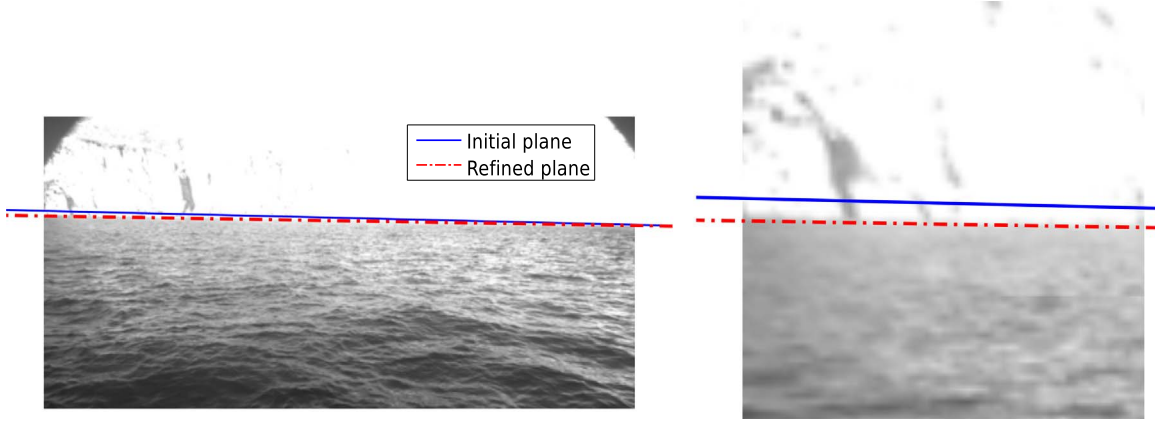


Fig. 3. Left: horizon line estimation process. The initial ℓ_H guess (blue line) gets refined by the optimization process according to the image gradient. The resulting optimized ℓ_H (red line) properly follows the correct horizon line. Right: magnification of a part of the left image. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

image spaces are planar homographies. In our case, we need a strong disparity between the cameras to reconstruct the point cloud. However, for all the points lying at the infinite, the disparity is negligible and so the behavior of points at infinity is similar to what happens for all other points in case of pure rotations. This is true since every set of parallel planes would intersect into the same line at infinity (see Section 3). Thus, since camera translation would not change the orientation of the plane, the projection of the horizon line on each camera depends only by its orientation (i.e. the rotation matrix) and not the spatial position (i.e. translation vector).

We perform our horizon line estimation working on the first camera image space. This time, before starting the optimization, we merge the vector field ∇I_{m_1} with ∇I_{m_2} transformed into I_{m_1} space. Let:

$$\nabla I_{m_2}(q_i) = \begin{pmatrix} \nabla_x I_2(q_i) \\ \nabla_y I_2(q_i) \end{pmatrix} \quad (18)$$

be the vector field of the second camera image gradient computed over the discrete set of points $q_1 \dots q_m \subset \mathbb{R}^2$. As before, ∇I_{m_2} is numerically obtained for each pixel by convolving the image with a derivative of Gaussian. All the I_{m_2} points $q_i = (u_i, v_i)^T$, $i = 1 \dots m$ can be transformed into the first camera space via the matrix of the infinite homography through the projection:

$$q'_i = \begin{pmatrix} u' \\ v' \\ w' \end{pmatrix}^T; \quad \begin{pmatrix} u' \\ v' \\ w' \end{pmatrix} = \mathbf{K}_1 \mathbf{R}_{21} \mathbf{K}_2^{-1} \begin{pmatrix} u_i \\ v_i \\ 1 \end{pmatrix} \quad (19)$$

Similarly, the gradient ∇I_{m_2} sampled at q_i can be transformed into the first camera space sampled at q'_i

$$\nabla I'_{m_2}(q'_i) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \mathbf{K}_1 \mathbf{R}_{21} \mathbf{K}_2^{-1} \begin{pmatrix} \nabla_x I_2(q_i) \\ \nabla_y I_2(q_i) \\ 0 \end{pmatrix}. \quad (20)$$

To summarize, we start by computing the image gradients ∇I_{m_1} and ∇I_{m_2} over the pixels of the first ($s_1 \dots s_n$) and the second ($q_1 \dots q_m$) image respectively. Then, we transform the vector field of the second camera into the first by using (19) and (20). Finally, we combine the vector fields ∇I_{m_1} and $\nabla I'_{m_2}$ to obtain the new vector field used for horizon line recovery as described in Section 3.1. Even if we restrict our application to stereo images, this approach could easily be extended to handle any number of different views.

4. Plane distance recovery

So far we showed that plane normal \vec{n} can be estimated without the reconstructed point cloud through the horizon line

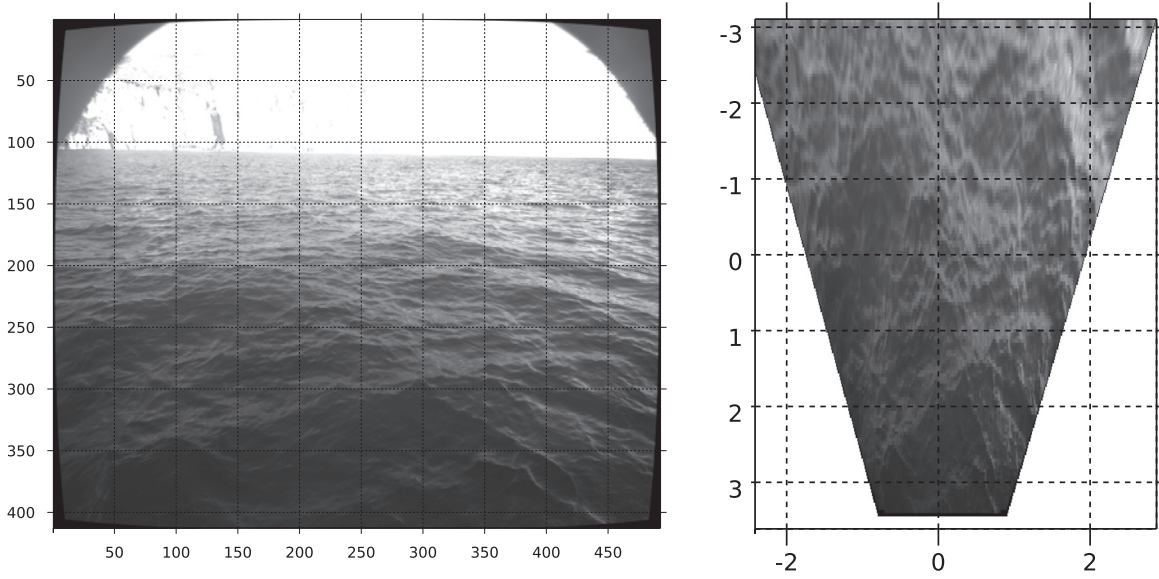


Fig. 4. Effect of the rectification of an image through the homography \mathbf{H}_r . Left: original image. Right: rectified image as seen by a camera with an optical axis orthogonal to the sea plane. Note that the scale of the right image is arbitrary (see text for details).

extracted from an image frame. From the normal \vec{n} , a transformation \mathbf{H}_r mapping points from the first camera image plane to an ideal camera whose optical axis is orthogonal to the sea plane can be recovered up to scale. This transformation is particularly interesting for two reasons: it allows the rectification of the acquired image to remove the perspective distortion (Fig. 4) and gives clues on how the space around a sea plane point get distorted when projected onto the image plane.

Following Hartley and Zisserman (2004), if we suppose to have two cameras looking at points $p_{i=1\dots k}$ lying on a plane Π , the projection of p_i on the first camera is related to its projection onto the second through the relation:

$$p_i^{(2)} = \mathbf{K}_2 \mathbf{H} \mathbf{K}_1^{-1} p_i^{(1)} = \mathbf{K}_2 \left(\mathbf{R} - \frac{\vec{t} \vec{n}^T}{d} \right) \mathbf{K}_1^{-1} p_i^{(1)} \quad (21)$$

where (\mathbf{R}, \vec{t}) is the relative pose of the first camera with respect to the second, d is the plane distance from the first camera and $p_i^{(1)}, p_i^{(2)}$ are the projections of p_i onto the first and second camera image planes respectively.

Since our only requirement is to transform to a camera with the optical axis orthogonal to Π , we can safely assume that $\vec{t} = \vec{0}$ (i.e. we rotate around the optical center), d is an undefined positive value and the new rectified camera has unitary focal length and principal point lying at the origin. Consequently, we let \mathbf{K}_2 be the identity matrix, $\mathbf{R} = \mathbf{R}_w$ and compute \mathbf{H}_r as

$$\mathbf{H}_r = \mathbf{R}_w \mathbf{K}_1^{-1}. \quad (22)$$

With \mathbf{H}_r and its inverse we can go back and forth between the first camera image plane and a normalized plane Π parallel to the sea plane. Specifically, given an image point $s = (u, v, 1)^T$ (where u, v are pixel coordinates), we can project to Π by a projection through \mathbf{H}_r .

The homography \mathbf{H}_r cannot be used to extract all the wave statistics that are usually computed to analyze a sea state. Indeed, the sea surface is not planar and the scale is arbitrary by construction since the horizon cannot provide us the distance between the camera rig and the sea plane. However, the projective transformation realized through \mathbf{H}_r can be used as a clue to quantitatively evaluate the amount of spatial distortion introduced during the reconstruction in a neighborhood of any image point s .

With this quantity in hand, we can weight all the reconstructed points to simulate a uniform spatial distribution and improve the plane recovery through LLS.

Our goal is to sample points on the sea plane at locations (X_1, X_2) with X_1 and X_2 being two uniformly distributed independent random variables whose joint pdf $f(u', v')$ is also uniformly distributed. The 2D location $(u', v')^T$ of a point on the sea plane with respect to its corresponding image plane coordinates $(u, v)^T$ is described by $h: \mathbb{R}^2 \rightarrow \mathbb{R}^2$:

$$\begin{pmatrix} u' \\ v' \end{pmatrix} = h(u, v) = \begin{pmatrix} h_1(u, v) \\ h_2(u, v) \end{pmatrix} = \frac{1}{\mathbf{H}_{r3} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix}} \begin{pmatrix} \mathbf{H}_{r1} \\ \mathbf{H}_{r2} \end{pmatrix} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \quad (23)$$

where $\mathbf{H}_{r1}, \mathbf{H}_{r2}$ and \mathbf{H}_{r3} are the three rows of \mathbf{H}_r .

Since our sampling space is the image plane, an effective way to simulate the uniform sampling is to compute the pdf over the image plane induced by $h(u, v)$ and use that density to weight each sample.

Specifically, each sample in the image is assumed being extracted at the coordinates given by two random variables (Y_1, Y_2) whose joint distribution is

$$g(u, v) = |\mathbf{J}| f(h_1(u, v), h_2(u, v)) \quad (24)$$

where $|\cdot|$ is the determinant operator and \mathbf{J} is the 2×2 Jacobian of h . Since f is constant everywhere inside the point sampling area we can discard that term as it would produce a simple scaling of all the sampling weights. Consequently, a better estimate of the sea plane distance given the reconstructed 3D point cloud can be obtained by weighting each sample s at image coordinates (u, v) with the determinant of the Jacobian of h evaluated at (u, v) . Specifically, the plane distance is obtained as the dot product between the centroid of the point cloud weighted with the proposed correction and the plane normal recovered with the horizon.

To summarize, supposing that the reconstructed 3D point cloud is composed of a set of $p_1 \dots p_N$ 3D-points whose projections on the first camera image plane are the 2D-points $s_1 \dots s_N$, the density-corrected plane distance can be computed with the formula:

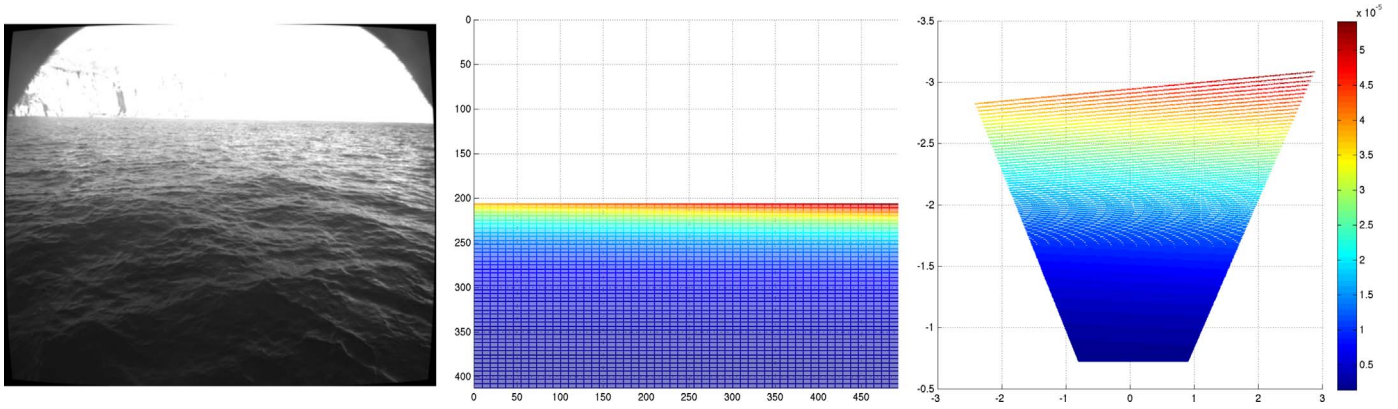


Fig. 5. Left: example image taken from a stereo rig. Center: weights computed with the proposed approach as seen by the camera. Right: the same weights as would be seen by a camera facing downward the sea.

$$d = \left\langle \frac{\sum_{i=1}^n \mathbb{J}(s_i) p_i}{\sum_{i=1}^n \mathbb{J}(s_i)}, \vec{n} \right\rangle \quad (25)$$

where $\mathbb{J}(s_i)$ is the determinant of the Jacobian of h evaluated at the 2D coordinates of the point s_i and $\langle \cdot, \cdot \rangle$ represents the dot product operation.

In Fig. 5 we show an example of the weights computed for the sea plane observed in Fig. 4 (left). It can be observed that the farther a point from the camera, the higher its associated weight would be. This makes sense since the reconstructed cloud density decreases with the camera distance. This is also coherent with the approximation made during previous WASS installations, where each sample was simply weighted by the z -component of each point as a fast approximation of the distance with respect to the camera center.

The proposed weighting criterion is theoretically correct to simulate a uniform random sampling on the reconstructed plane but has the drawback to amplify the noise since also the depth uncertainty increases with the distance. Therefore, we face two competing terms: with a weight proportional to the point distance we obtain a quasi-uniform distribution but, on the other hand, we are giving more weight to the points that are intrinsically more uncertain because they are far away from the camera. If the reconstructed surface were a real plane (i.e. a sea with no waves at all), it would certainly be better to avoid the proposed weight criterion to limit the effect of depth uncertainty on the least-squares fitting. However, when the variability of the reconstructed point elevation clearly dominates the variability due to the depth uncertainty (i.e. when observing a wavy sea surface), the proposed weighting effectively increases the plane fitting accuracy. We demonstrate this phenomenon with synthetic tests in the experimental section.

Finally, we mind the reader that the considerations made so far consider the sample density along the sea or image plane. However, point lying far away from the plane (either for being outliers or belonging to rogue wave events) may receive an associated weight not reflecting the actual space density in the neighborhood. In practice, this difference is negligible and does not hinder the improved plane estimation accuracy on the tests we made so far.

5. Experimental section

In this section we test the proposed approach under different conditions. We start with some synthetic cases to quantitatively evaluate the sea plane recovery accuracy in a simulated scenario. In Section 5.3 we test the horizon estimation in a controlled real-

world environment during the acquisition of 3D data from a fixed platform. Then, in Section 5.4 we present a real-case study of the data acquired during a cruise over the Garda Lake, Italy.

5.1. Synthetic tests

To evaluate how the plane distance recovery presented in Section 4 leads to a better aligned sea surface reconstruction, we implemented a test based on purely synthetic data.

We started by generating a sequence of 100 consecutive frames of a theoretical sea surface (with significant wave height of 1 m) by simulating an evolving linear sea surface based upon the random-phase/amplitude model (Pierson and Moskowitz, 1964). Such a model assumes that a Gaussian sea surface is the result of the summation of independent harmonic components of amplitude, angular frequency, direction, and phase. Operatively, we used the WAFO toolbox for MATLAB® (Brodtkorb et al., 2000), which has been extensively applied for simulations of random seas (Gemmrich and Garrett, 2008). Each sea surface was generated so that the mean sea plane lies exactly on the XY -plane and the mean elevation of all the surface points is 0. Thereafter we considered a virtual non-distorted pinhole camera with a resolution of 500×420 pixels, a focal length of 350 pixels and a principal point located at the image center. The camera is initially faced downward (so that the synthetic sea plane is parallel to the image plane) at a distance of $d = 100$ pixels from the sea plane.

We simulate the perspective distortion occurring for stereo reconstruction by rotating the synthetic sea surface around the camera x -axis with an angle of θ radians and computing the 2D intersection coordinates on the sea reference frame of each ray exiting the camera with the sea plane. This projective-distorted samples are then bilinearly interpolated with the sea surface elevation to obtain a 3D point cloud that resembles a real stereo acquired point cloud (we do not consider shadowing effects or triangulation errors). Fig. 6 (left) gives an example of how the distorted synthetic sea plane would result after a rotation of 20° , 40° and 50° respectively.

For any θ , the homography \mathbf{H}_r is known given the plane normal $\vec{n} = (0 \sin\theta \cos\theta)^T$ and distance d . We compare in terms of absolute mean error the 3D plane fitted on the distorted point cloud by weighting the points with the determinant of the Jacobian (Eq. (24)) with respect to the unweighted case. In Fig. 6 (right) we observe how the point weighting based on sample density (“Density sample” in the plot) almost always result in a lower elevation error on the sea surface caused by a more accurate estimation of the mean plane distance. As expected, the more the sea plane is tilted with respect to the camera, the more the density weighting is effective to provide a precise sea distance estimation

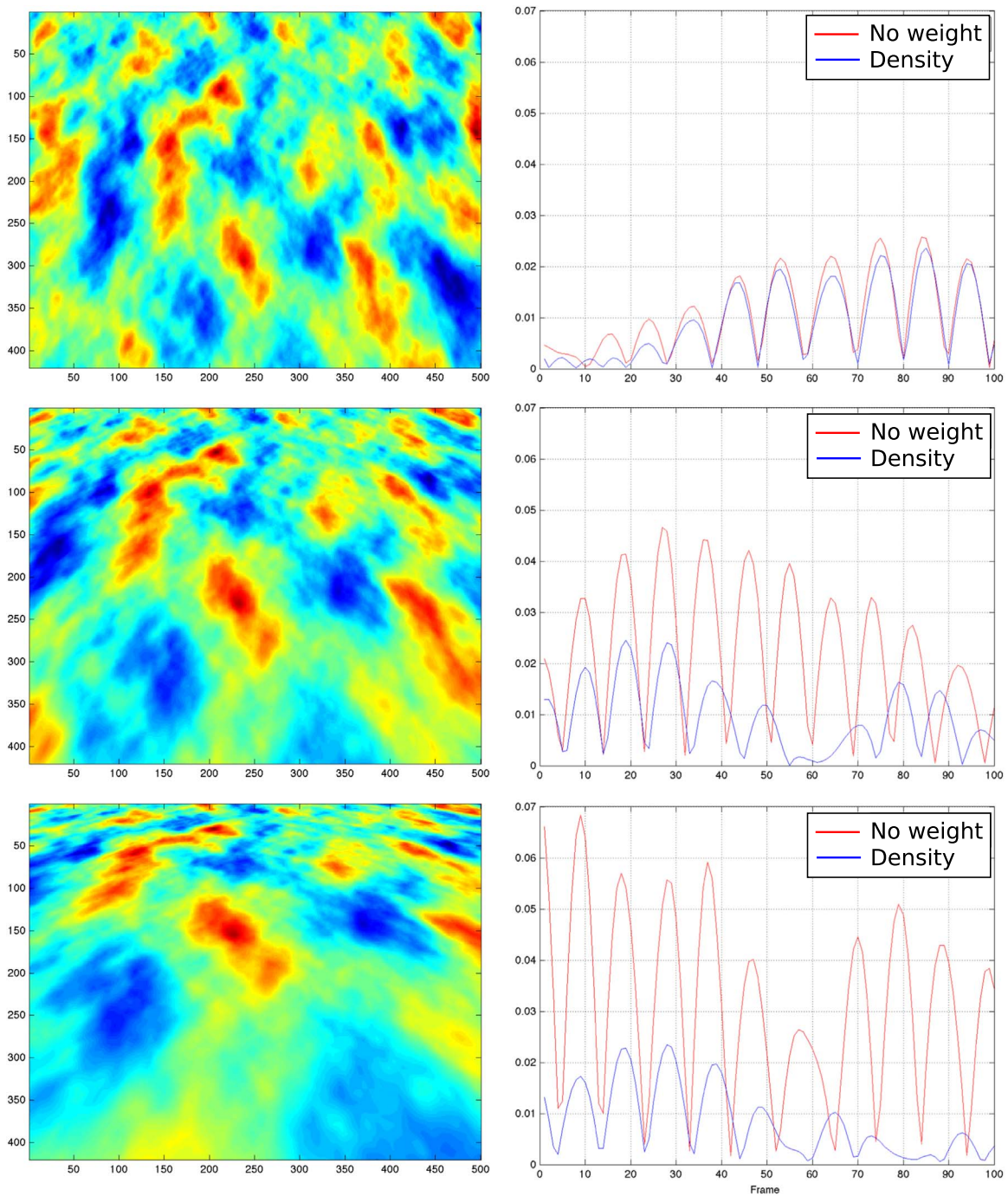


Fig. 6. Different LLS fits on sea plane estimation for the density weighted and un-weighted (i.e. uniform weight) case. Left: example of the synthetically generated sea surface varying the observation angle $\theta = 20^\circ$ (top), 40° (middle), 50° (bottom). Right: absolute elevation error (in meters) between the generated sea surface and the rotated sea surface after the plane estimation for the two cases.

that is otherwise biased by the samples nearby the camera. Finally, for both the methods the error is oscillating as the plane estimation is affected by the wavy nature of the generated sea surface. This was also expected, as it has been already demonstrated in the literature that is impossible to obtain a consistent sea plane estimation by observing a limited (in terms of number of wave-lengths) part of the whole sea surface.

5.2. Evaluating the weighting effect on depth uncertainty

To study the tradeoff described in Section 4 for which the proposed weighting criterion would inevitably amplify the noise

due to depth uncertainty we decided to evaluate its effect in a purely numerical way. Since our reconstruction pipeline is composed of many different filters and optimizations, it would be too limiting to model its behavior as an ideal set of undistorted calibrated cameras. As a consequence, we rendered a synthetically generated evolving wave surface using the same intrinsic and extrinsic coefficients calibrated on the WASS device in the “Acqua Alta” platform (See Section 5.3) and run the whole pipeline to compare the reconstructed point cloud with the known generated ground truth. The effect of an erroneous horizon estimation is eliminated in this test since the stereo rig orientation with respect to the sea plane is known so that the image-sea plane homography

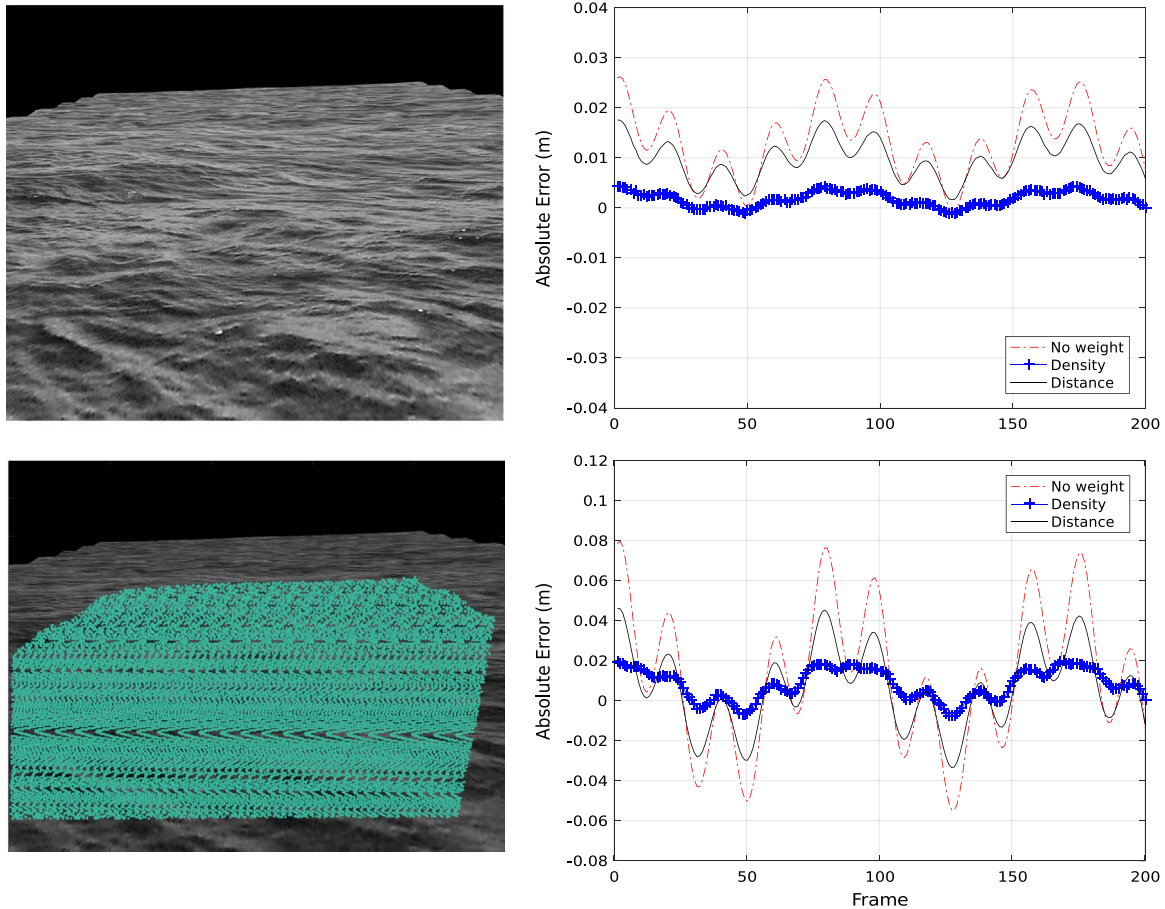


Fig. 7. Comparison of different weighting schemes on a generated sea surface rendering to include the effects of depth uncertainty in a simulated scenario. Top-left: example of the generated sea surface image. Bottom-left: reconstructed area. Right column: absolute error of the estimated plane distance for different weightings on a generated sea with wave amplitudes of 0.5 m (top) and 2.5 m (bottom).

H_r is known a priori.

We rendered a sequence of 200 stereo images (Fig. 7, top-Left) and run the reconstruction pipeline with the same parameters used with real-world tests to obtain a sequence of point clouds which were cropped to limit the reconstruction area to $51 \times 44 \text{ m}^2$ as shown in Fig. 7 (bottom-Left). Then, we estimated the sea-plane distance using three different weighting criteria: the proposed density correction, a weight proportional with the point distance and no weight (i.e. uniform weighting). We performed this test twice, one with a synthetically generated sea with a maximum sea wave amplitude of 0.5 m (Fig. 7, top-right) and one with 2.5 m (Fig. 7, bottom-right) to plot the estimated sea plane distance absolute error with respect to the ground truth. In both the cases, our proposed plane distance estimation consistently give lower error and less variability. As proposed in Benetazzo et al. (2016), the distance weighting criterion give better results than uniform weighting as it partially compensates the non-uniform sampling. In the considered reconstruction area the effect of the depth uncertainty does not hinder the estimation even if we give more weight to distant points.

5.3. Real world scenario

The approach discussed in Sections 2 and 4 has been tested on real world stereo data collected in controlled conditions. Specifically, we mounted the WASS device described in Section 2 at 12.5 m above the sea level on the “Acqua Alta” oceanographic platform located in the northern Adriatic Sea (Italy) with an angle of 60° with respect to the sea plane normal.

Using a fixed platform allows us to test our algorithms with real-world conditions (illumination changes, noise, atmospheric effects, etc) but with the additional guarantees that the stereo-camera rig is not moving during the acquisition. As a consequence, a ground truth sea plane orientation can be obtained by averaging the acquired sea surfaces over a long period of time.

In this experiment, we oriented the stereo rig so that the horizon line is dividing the sea from the sky at about 1/6 of the entire image seen from the left camera (see Fig. 8, bottom-left). We acquired a sequence of 2000 frames and processed the whole data to obtain a set of 3D point clouds. For each frame, we compare the original pipeline in which the mean sea plane is fitted from the 3D point cloud as linear least squares weighted with the point distance (3D fit in the figure) with the proposed horizon approach. In Fig. 8 the dihedral angles of the sea plane for the two methods are compared (top-row) together with the estimated plane distance (bottom-right plot). Since the stereo rig is fixed, we expect the recovered sea plane parameters being constant over time.

From the plots, two considerations can be made. First, the variability of the plane estimation (both by means of plane angle and plane distance) with the aid of the horizon is significantly lower with respect to the 3D fitting alone. This was expected since it has already been demonstrated that the accuracy of the sea plane estimation from 3D data depends on the number of wavelengths observed in the reconstruction area. With these sea conditions, using the horizon from the sea plane recovery can greatly improve the alignment between each surface. Second, we observe a little bias with the parameters estimated from the horizon with respect to the average of the planes fitted with LLS. In the specific,

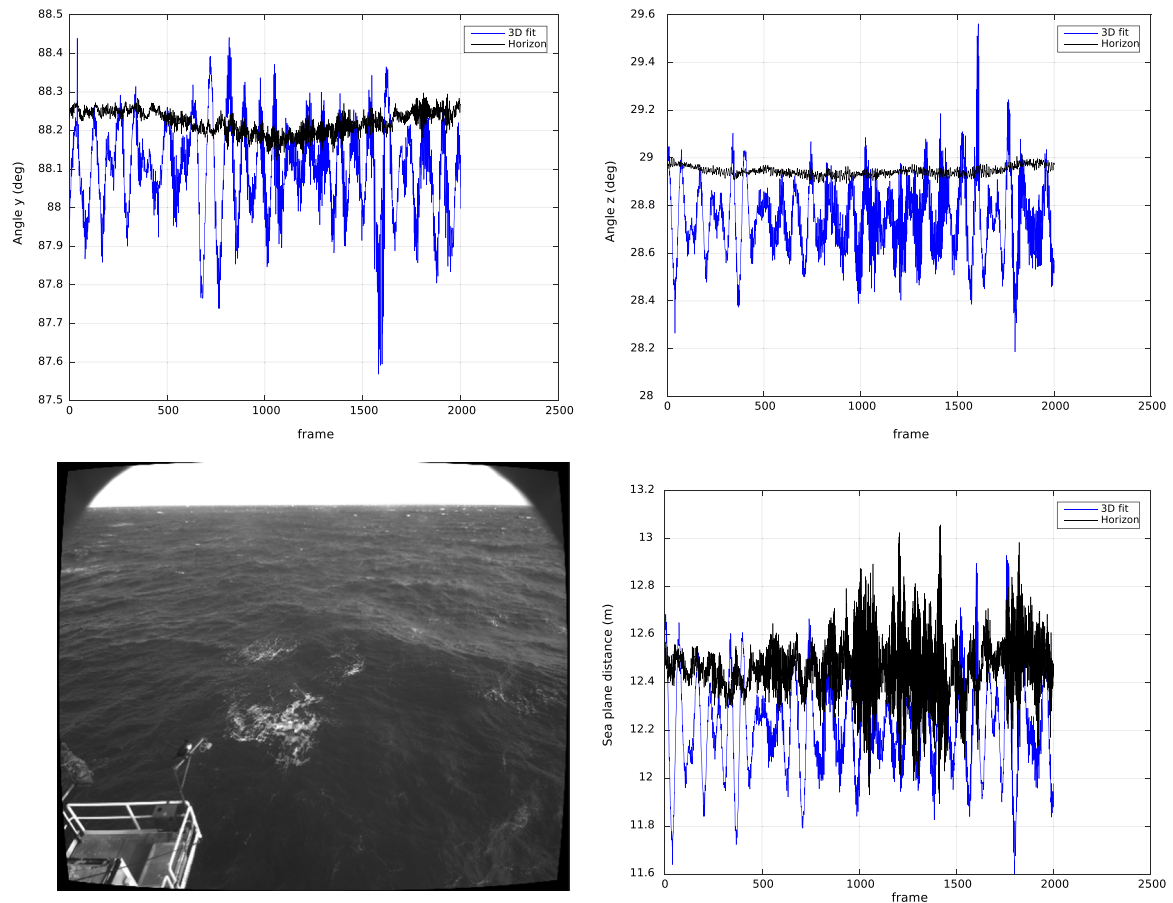


Fig. 8. Time series comparison between the mean sea plane estimation performed as a linear least squares of each 3D point cloud and the proposed approach. Top: plane angle over time with respect to camera z- and y-axes respectively. Bottom-right: plane distance estimation over time. Bottom-left: example of an image acquired during the experiment.

the angle with respect to the y -axis is overestimated of about 0.12° , with respect to the z -axis of about 0.15° and the plane distance of ≈ 20 cm. This phenomenon is probably caused by small calibration errors in the camera intrinsics. Overall, the error introduced by this small bias is less than the intrinsic variability of the plane parameters recovered with the 3D data alone. This experiment suggests that the horizon line can greatly improve the sea plane recovery in a moving vessel scenario for which no additional attitude estimation aids are available.

5.4. Case study: wave analysis from a moving ship

As a case study, we tested the wave data acquired from a moving ship during a cruise (onboard the vessel Motonave San Marco, about 32 m long) over the Garda Lake (Italy). Here, the stereo-camera system mounted had same features as the WASS installed on the “Acqua Alta” platform (Section 5.3), apart from the baseline that was reduced ($t = 1.07$ m) as onboard cameras were deployed at a smaller height, about 3.3 m above the mean sea level.

An experiment was conducted starting at 11:03 UTC on December 09, 2015; at that time the wind (measured on the vessel at about 6 m height) was blowing at approximately 5 m/s along the lake main axis (oriented southwest-northeast). As in a previous installation on a vessel (Benetazzo et al., 2014), WASS collected stereo images at 15 Hz, and it was synchronized with a compass (that provided the heading with respect to the true north) and a global positioning system (GPS) to retrieve the WASS location on the world reference frame (i.e., longitude–latitude oriented). From

the entire sequence of stereo images, the first 1000 pairs have been used to assess the proposed approach to recover the attitude of the stereo rig using the horizon line.

In this respect, we have firstly fitted, in a linear least squares sense, the 3D data in the camera reference system to retrieve, at each frame, a guess of the horizontal plane attitude and position, which are required (Benetazzo, 2006) to map the stereo data onto the sea reference frame (Fig. 9). As the stereo system onboard the vessel qualitatively encompassed four wavelengths and four wave crests, the study of Benetazzo et al. (2016) demonstrated that the fitting procedures yielded, at each instance, a reliable estimation of the mean sea plane coefficients, that is its normal (n_x, n_y, n_z) and distance d . Second, as a horizontal line was visible on both views (Fig. 9), this was exploited to retrieve an additional set of coefficients. In accordance with the two methodologies just defined above, some of the mean sea plane parameters are depicted in Fig. 10 for both cases. The rotations around the camera y -axis are pretty close with an average difference between the dihedral angles of 0.1° , while the average difference is smaller than 0.01° insofar as rotation around the x -axis is concerned. Also, even though the fitting procedure captures slightly larger oscillations of the distance (heave) of the cameras from the mean sea planes, the average difference between the heaves is small, about 0.01 m.

Along with GPS and compass data, these two sets of coefficients permitted to get two distinct space–time ensembles of sea surface elevations mapped onto the world reference frame (Benetazzo et al., 2016). The two sea elevation ensembles have small differences, in the order of few centimeters at most, whereas the mean absolute difference is 0.6 cm. Not surprisingly then, comparison

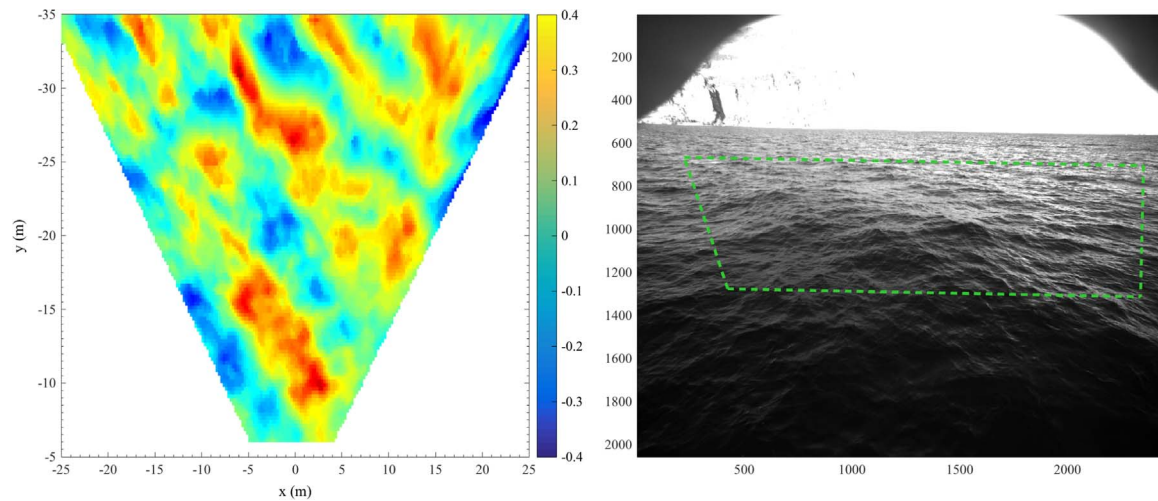


Fig. 9. Example of an acquired camera image (right) and its 3D reconstruction (left) in the sea reference frame.

suggests that the two dataset are very similar, and it indicates that the horizon permitted to properly orient the cameras with respect to the sea reference frame. Further confirmation is provided by the results in Fig. 11 that shows the time series extracted from the two ensembles at the position $(x, y) = (68 \text{ m}, 49 \text{ m})$ of the geographical frame. The two time series show similar values, with a bias of 0.3 cm, with the standard deviation of both series equal to about 10.5 cm.

6. Conclusions

In this paper we proposed a novel technique that exploits the observed horizon to provide a better alignment of the reconstructed 3D sea surface with respect to the mean sea plane. We avoid the combinatorial search of the horizon line over an image assuming that a 3D reconstruction of the surface have already been produced by the stereo vision machinery and refine the initial coarse estimate by means of a multi-view energy optimization involving all image pixels. Thus, this approach recovers the sea plane orientation with higher accuracy without pre-processing the image to extract and classify edge features. Moreover, we propose

a point weighting criterium that compensates the non-uniform spatial distribution of the reconstructed point cloud due to the relative angle between the stereo rig and the sea surface. We empirically showed that such point weighting improves the sea plane distance estimation (that cannot be recovered by the horizon line itself) over the common unweighted Linear Least Squares and the previously adopted point distance weighting even considering that such weighting would inevitably amplify also the depth error.

We tested the proposed method against real-world and synthetically generated data. From the synthetic test we assessed how the weighting scheme improves the sea plane distance estimation. The improvement has a significant impact that if the cameras are tilted with respect to the sea plane that is an unavoidable condition if the horizon has to be visible on the images. Furthermore, from our test on the oceanographic research platform “Acqua Alta” we observed that the variability of the estimated sea-plane parameters can be greatly reduced if the horizon is taken into account. As a limited drawback, such estimation may be biased by the stereo rig intrinsic and extrinsic parameters that must be precisely calibrated. Finally, we tested the proposed method with data acquired on a real vessel to demonstrate its feasibility in a real-world

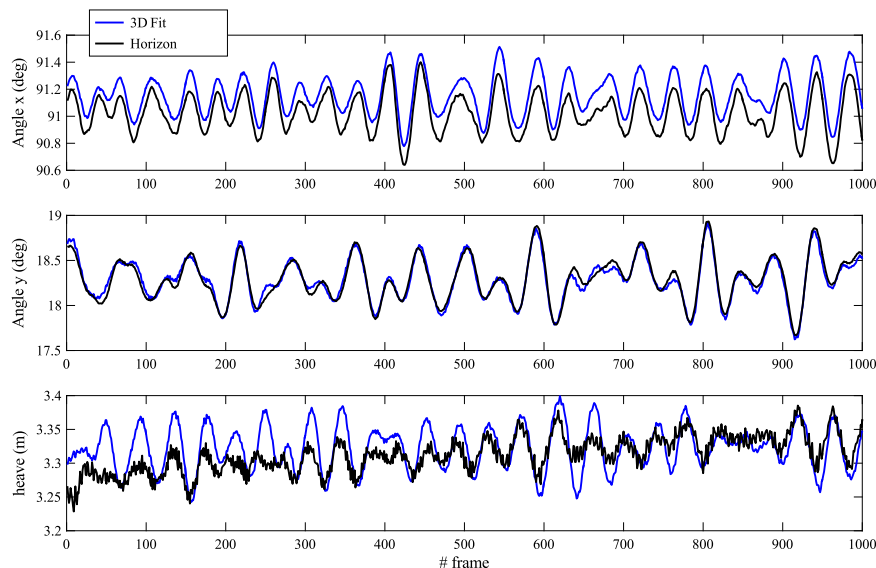


Fig. 10. Sea plane parameters evolution over time. Top: dihedral angle around the camera x -axis. Middle: dihedral angle around the camera y -axis. Bottom: camera distance (heave) from the mean sea plane. Parameters are shown for the two sets of planes obtained fitting the 3D data (blue) and using the horizon line (black). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

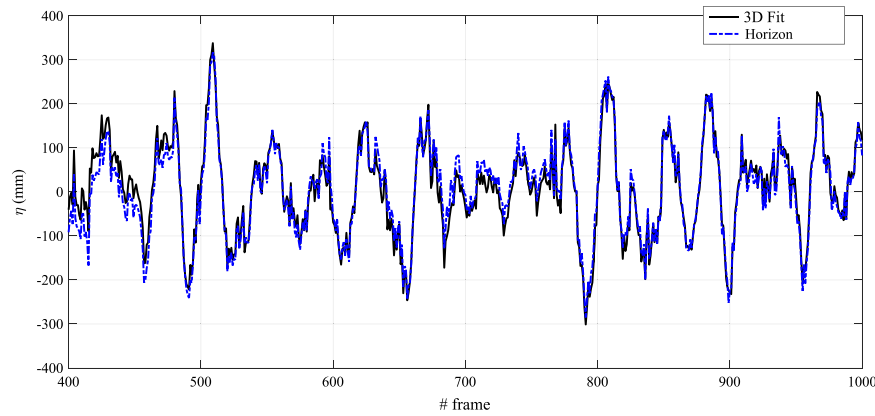


Fig. 11. Time series comparison of the sea surface elevation observed at the horizontal point of coordinates $(x, y) = (68 \text{ m}, 49 \text{ m})$. Data are extracted from the space–time ensembles computed fitting the 3D data in the camera reference system (black) and using the horizon line (blue). Data are plotted for frames 400–1000, when the stereo-camera field of view spanned the point $(x, y) = (68 \text{ m}, 49 \text{ m})$. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

scenario. Overall, we think that this technique can offer a significant advantage when operating on a moving vessel and no additional position and attitude recovery devices are available.

References

- Albarelli, A., Rodolà, E., Torsello, A., 2010. Robust camera calibration using inaccurate targets. In: *Proceedings of the British Machine Vision Conference*, BMVA Press, Aberystwyth, UK, pp. 16.1–16.10. <http://dx.doi.org/10.5244/C.24.16>.
- Angelino, Baraniello, V., Cicala, L., 2013. High altitude UAV navigation using IMU, GPS and camera. In: *2013 16th International Conference on Information Fusion (FUSION)*, pp. 647–654.
- Banner, M.L., Barthelemy, X., Fedele, F., Allis, M., Benetazzo, A., Dias, F., Peirson, W.L., 2014. Linking reduced breaking crest speeds to unsteady nonlinear water wave group behavior. *Phys. Rev. Lett.* 112, 114502.
- Banner, M.L., Jones, I.S.F., Trinder, J.C., 1989. Wavenumber spectra of short gravity waves. *J. Fluid Mech.* 198, 321–344.
- Bao, G., Zhou, Z., Xiong, S., Lin, X., Ye, X., 2003. Towards micro air vehicle flight autonomy research on the method of horizon extraction. In: *Instrumentation and Measurement Technology Conference, 2003. IMTC '03. Proceedings of the 20th IEEE*, vol. 2, pp. 1387–1390.
- Benetazzo, A., 2006. Measurements of short water waves using stereo matched image sequences. *Coast. Eng.* 53, 1013–1032.
- Benetazzo, A., Barbariol, F., Bergamasco, F., Torsello, A., Carniel, S., Sclavo, M., 2015. Observation of extreme sea waves in a spacetime ensemble. *J. Phys. Oceanogr.* 45 (9), 2261–2275.
- Benetazzo, A., Barbariol, F., Bergamasco, F., Torsello, A., Carniel, S., Sclavo, M., 2016. Stereo wave imaging from moving vessels: practical use and applications. *Coast. Eng.* 109, 114–127.
- Benetazzo, A., Bergamasco, F., Barbariol, F., Torsello, A., Carniel, S., Sclavo, M., 2014. Towards an operational stereo system for directional wave measurements from moving platforms. In: *Proceedings of the ASME 33th International Conference on Offshore Mechanics and Arctic Engineering, OMAE 2014, June 19–24, 2014, Nantes, France. ASME*.
- Benetazzo, A., Fedele, F., Gallego, G., Shih, P.-C., Yezzi, A., 2012. Offshore stereo measurements of gravity waves. *Coast. Eng.* 64, 127–138.
- Brandt, A., Mann, J.L., Rennie, S.E., Herzog, A.P., Criss, T.B., 2010. Three-dimensional imaging of the high sea-state wave field encompassing ship slamming events. *J. Atmos. Ocean. Technol.* 27, 737–752.
- Brodtkorb, P., Johannesson, P., Lindgren, G., Rychlik, I., Rydén, J., Sjö, E., 2000. WAFO—a Matlab toolbox for the analysis of random waves and loads. In: *Proceedings of the 10th International Offshore and Polar Engineering Conference, ISOPE, Seattle, USA*, vol. 3, pp. 343–350.
- Canny, J., 1986. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell. PAMI-8* (6), 679–698.
- Cherian, A., Morellas, V., Papanikolopoulos, N., 2009. Accurate 3d ground plane estimation from a single image. In: *IEEE International Conference on Robotics and Automation, 2009. ICRA 2009*, pp. 2243–2249.
- Duda, R.O., Hart, P.E., 1972. Use of the Hough transformation to detect lines and curves in pictures. *Commun. ACM* 15 (1), 11–15.
- Dusha, D., Boles, W.W., Walker, R.A., 2007. Attitude estimation for a fixed-wing aircraft using horizon detection and optical flow. In: *9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications, Adelaide, Australia. IEEE*, pp. 485–492.
- Eynard, D., Vasseur, P., Demonceaux, C., Fremont, V., 2010. UAV altitude estimation by mixed stereoscopic vision. In: *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 646–651.
- Fefilat'ev, S., Goldgof, D.B., Langebrake, L., 2007. Toward detection of marine vehicles on horizon from buoy camera. <http://dx.doi.org/10.1117/12.747512>.
- Gallagher, A., Luo, J., Hao, W., 2004. Improved blue sky detection using polynomial model fit. In: *2004 International Conference on Image Processing, 2004. ICIP '04*, vol. 4, pp. 2367–2370.
- Gallego, G., Benetazzo, A., Yezzi, A., Fedele, F., 2008. Wave statistics and spectra via a variational wave acquisition stereo system. In: *Proceedings of the ASME 27th International Conference on Offshore Mechanics and Arctic Engineering, OMAE 2008, June 15–20, 2008, Estoril, Portugal, ASME*, pp. OMAE2008–57160.
- Gallego, G., Yezzi, A., Fedele, F., Benetazzo, A., 2011. A variational stereo method for the three-dimensional reconstruction of ocean waves. *IEEE Trans. Geosci. Remote Sens.* 49 (11), 4445–4457.
- Gemmrich, J., Garrett, C., 2008. Unexpected waves. *J. Phys. Oceanogr.* 38, 2330–2336.
- Hartley, R.L., Zisserman, A., 2004. *Multiple View Geometry in Computer Vision*, second edition. Cambridge University Press, ISBN: 0521540518.
- Holthuijsen, L., 1979. *Stereophotography of ocean waves*.
- Hugues, O., Cieutat, J.-M., Guitton, P., 2014. Real-time infinite horizon tracking with data fusion for augmented reality in a maritime operations context. *Virtual Real.* 18 (2), 129–138.
- Kass, M., Witkin, A., Terzopoulos, D., 1988. Snakes: active contour models. *Int. J. Comput. Vision.* 1 (4), 321–331.
- Kosnik, M.V., Dulov, V.A., 2011. Extraction of short wind wave spectra from stereo images of the sea surface. *Meas. Sci. Technol.* 22 (1), 015504.
- Leckler, F., Arduin, F., Peureux, C., Benetazzo, A., Dulov, V., 2015. Analysis and interpretation of frequency-wavenumber spectra of young wind waves. *J. Phys. Oceanogr.* <http://dx.doi.org/10.1175/JPO-D-14-0237.1>
- Levenberg, K., 1944. A method for the solution of certain non-linear problems in least-squares. *Q. Appl. Math.* 2 (2), 164–168.
- Liu, T., Guo, H., Wang, X., Abraham, A., Liu, H., 2008. A machine vision approach for jacket launching angle measurement. In: *Third International Conference on Digital Information Management, 2008. ICDIM 2008*, pp. 227–232.
- Pedoe, D., 1970. *Geometry: A Comprehensive Course*. Dover Books on Advanced Mathematics. Dover Publications, Mineola, New York, US.
- Pierson Jr., W.J., Moskowitz, L., 1964. A proposed spectral form for fully developed wind seas based on the similarity theory of S. A. Kitaigorodskii. *J. Geophys. Res.* 69, 5181–5190.
- Schumacher, A., 1939. Stereophotogrammetrische Wellenaufnahmen. *Wiss. Ergeb. Dtsch. Atlant. Exped. Forschungs Vermessung. Meteor. Technical report*.
- Schwendeman, M., Thomson, J., 2015. A horizon-tracking method for shipboard video stabilization and rectification. *J. Atmos. Ocean. Technol.* 32, 164–176.
- Seber, G., Lee, A., 2012. *Linear Regression Analysis*. Wiley Series in Probability and Statistics. Wiley.
- Shabayek, A.E.R., Demonceaux, C., Morel, O., Fof, D., 2012. Vision based UAV attitude estimation: progress and insights. *J. Intell. Robot. Syst.* 65 (1–4), 295–308.
- Todorovic, S., Nechyba, M., Ifju, P., 2003. Sky/ground modeling for autonomous MAV flight. In: *Proceedings of the IEEE International Conference on Robotics and Automation, 2003, ICRA '03*, vol. 1, pp. 1422–1427.
- Wang, L., Hsieh, S.-C., Hsueh, E.-W., Hsiao, F.-B., Huang, K.-Y., 2005. Complete pose determination for low altitude unmanned aerial vehicle using stereo vision. In: *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2005, IROS 2005*, pp. 108–113.
- Wang, Y., Liao, Z., Guo, H., Liu, T., Yang, Y., 2009. An approach for horizon extraction in ocean observation. In: *2nd International Congress on Image and Signal Processing, 2009, CISP '09*, pp. 1–5.
- Williams, S., Howard, A., 2011. Horizon line estimation in glacial environments using multiple visual cues. In: *2011 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 5887–5892.
- Yurovskaya, M.V., Dulov, V.A., Chapron, B., Kudryavtsev, V.N., 2013. Directional short wind wave spectra derived from the sea surface photography. *J. Geophys. Res.: Ocean.* 118 (9), 4380–4394.
- Zhang, L., Koch, R., 2013. Vanishing points estimation and line classification in a Manhattan world. In: *Lee, K., Matsushita, Y., Rehg, J., Hu, Z. (Eds.), Computer Vision ACCV 2012*, vol. 7725, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, pp. 38–51.