WASS: an Open-Source Pipeline for 3D Stereo Reconstruction of Ocean Waves

Filippo Bergamasco[a,*], Andrea Torsello[a], Mauro Sclavo[b], Francesco Barbariol[b], Alvise Benetazzo[b]

[a] *DAIS, Università Ca'Foscari, Venice, Italy*
[b] *Institute of Marine Sciences, Italian National Research Council (ISMAR-CNR), Venice 30122, Italy*

Abstract

Stereo 3D reconstruction of ocean waves is gaining more and more popularity in the oceanographic community and industry. Indeed, recent advances of both computer vision algorithms and computer processing power now allow the study of the spatio-temporal wave field with unprecedented accuracy, especially at small scales. Even if simple in theory, multiple details are difficult to be mastered for a practitioner, so that the implementation of a sea-waves 3D reconstruction pipeline is in general considered a complex task. For instance, camera calibration, reliable stereo feature matching and mean sea-plane estimation are all factors for which a well designed implementation can make the difference to obtain valuable results. For this reason, we believe that the open availability of a well tested software package that automates the reconstruction process from stereo images to a 3D point cloud would be a valuable addition for future researches in this area.

We present WASS (`http://www.dais.unive.it/wass`), an Open-Source stereo processing pipeline for sea waves 3D reconstruction. Our tool completely automates all the steps required to estimate dense point clouds from stereo images. Namely, it computes the extrinsic parameters of the stereo rig so that no delicate calibration has to be performed on the field. It implements a fast 3D dense stereo reconstruction procedure based on the consolidated OpenCV library and, lastly, it includes set of filtering techniques both on the disparity map and the produced point cloud to remove the vast majority of erroneous points that can naturally arise while analyzing the optically complex nature of the water surface.

In this paper, we describe the architecture of WASS and the internal algorithms involved. The pipeline workflow is shown step-by-step and demonstrated on real datasets acquired at sea.

Keywords: Sea Surface Stereo Reconstruction, Stereo Reconstruction, 3D wave field, Camera Calibration, Open-source software

*Corresponding author
Email address: `filippo.bergamasco@unive.it` (Filippo Bergamasco)

1. Introduction

The 3D geometry is a crucial feature of ocean waves and it is important in many problems, covering both the scientific research, off-shore and coastal engineering applications. Indeed, natural ocean waves evolve in time spanning a defined sea surface region where the wave field is assumed statistically homogenous. However, at short-time scale (comparable to characteristic wave periods) the spatial wave field shows distinctive features (e.g. the short-crestedness) that cannot be retrieved by traditional one-point observational systems (for example wave probes, buoys, etc.) Instead, a spatio-temporal observational system is required.

Spatio-temporal wave measurements are typically somewhat hard to obtain in the open ocean, but significant progress has been made in recent years. Different techniques were used for this purpose, ranging from polarimetric imagery (Zappa et al., 2008), X-band wave radar (Borge et al., 2013; Young et al., 1985) to satellite-borne synthetic aperture radar (Hasselmann et al., 1985). A direct observation of the 3D wave field is also provided by the stereo imaging technique (Jahne, 1993). To give a general taxonomy of the spatial scales captured by each aforementioned technology, we can state that polarimetry captures short (wavelength range: $0.001-1m$) gravity-capillary and gravity waves, stereo imaging captures short to mid size wavelengths (wavelength range: $0.2-50m$), and X-band radar captures mid-length to long waves (wavelength range: $10-300m$).

The pipeline described here concerns specifically the processing of stereo images for 3D reconstruction of ocean waves. Indeed, stereo wave imaging has been proving to be an accurate technique for direct observations of 3D wave fields at scales important for many scientific and applicative problems, with the only limitation being the image processing that has to be performed a efficient and effective way. This is nowadays achieved given the advances in the field of computer vision. History of stereo wave imaging traces back of about one century, when (Schumacher, 1939) installed a stereo-photographic system on a ocean going ship. Afterwards, a major attempt in stereo-photography for ocean waves was the Stereo Wave Observation Project in 1954 (Cote et al., 1960). In the eighties, advances on stereo processing include the studies of (Shemdin et al., 1988; Shemdin and Tran, 1992; Banner et al., 1989). However, all early analysis processed manually the stereo pairs, resulting in a limited amount of image pairs to be processed. Only more recently has automatic computer-based processing of stereo images been used to investigate the 3D wave geometry.

Scholars used stereo 3D stereo reconstruction of ocean waves to explore a variety of relevant scientific questions. For example, in the field of extreme wave analysis, (Benetazzo et al., 2015) investigated the presence of rogue waves within space-time sample of sea surface elevations, and (Fedele et al., 2013) compared time and space-time statistics of high waves. (Banner et al., 1989) examined the shape of the wavenumber wave spectrum, while (Leckler et al., 2015) studied the shape of the wavenumber-frequency spectrum computed via direct Fourier transform of the stereo data (Benetazzo et al., 2012; Gallego et al., 2011). (Banner et al., 2014) verified, using laboratory data and field stereo ob-

servations, the slowdown of unsteady wave groups under wave crest maxima. (Sutherland and Melville, 2013) used stereo-processed infrared imagery to estimate the dissipation rate of breaking waves. (Campbell et al., 2014) studied the interaction of the wave field with different ice types. (Mironov et al., 2012) proposed a methodology to extract short-scale statistical characteristics of the sea surface topography. All studies so far used stereo system mounted over fixed (non-moving) structures. This condition eases the stereo processing but highly limits the conditions under which the 3D characteristic of the wave field are investigated. Hence, recently, studies have been conducted to verify limitations of stereo systems mounted over moving structures (typically a ship), the difference being the requirement of the motion parameters to correctly map the stereo data into a reference system coherent with the gravity field. The ship motion compensations for stereo data are obtained using external instruments (Brandt et al., 2010; Benetazzo et al., 2014), the horizon line (Bergamasco et al., 2016), or the 3D sea surface field itself (Benetazzo et al., 2016).

Stereo matching has been one of the most active area in computer vision in the recent past. A deluge of different methods have been proposed over the years building on the same common sequence of operations, namely: matching cost computation, cost aggregation, disparity computation and disparity refinement. Given the intrinsic ambiguity of the problem, spatial smoothness of the disparity map is often considered a good prior to obtain a physically reasonable surface reconstruction. In particular, many different strategies were studied for the first two steps depending whether the disparity computation is determined by matching intensity values over a finite window (*local methods*) or by considering the whole image to perform a global matching cost minimization (*global methods*). The first have the advantage to be faster in general but more prone to outliers for the local nature of the matching cost computation. Conversely, *global methods* usually result in a smoother and more accurate reconstruction but require far more computational resources.

Almost all the seminal oceanographic studies (Benetazzo, 2006; Wanek and Wu, 2006; Brandt et al., 2010; Bechle and Wu, 2011; de Vries et al., 2011; Kosnik and Dulov, 2011; Benetazzo et al., 2012; Leckler et al., 2015) adopted *local methods* to compute the disparity map and hence suffer the delicate trade-off between the disparity window size (which influence the match localization accuracy) and the required surface smoothness that is particularly important while reconstructing a continuous injective surface like the sea. The state-of-the-art global method applied to waves reconstruction was proposed by (Gallego et al., 2008, 2011) that use a variational approach to directly estimate a continuous smooth surface by optimizing the photometric consistency between the two views. Additionally, the formulation allows also to impose physical constraints in the recovered points elevations. Unfortunately, such approach is so computationally intensive that is unlikely to be used in practice especially for long sequences.

The open source pipeline proposed here is a refined implementation of the WASS system which has been extensively used and tested in the last five years (Benetazzo et al., 2012, 2014, 2015; Bergamasco et al., 2016). For stereo match-

ing, we make use of a *semi-global* approach that combines the good performances of local methods (in terms of CPU time) with the accuracy and robustness of global methods. Additionally, the proposed WASS pipeline incorporates all the features needed to perform the extrinsic calibration on the field which has been demonstrated to increase the calibration accuracy easing its installation at sea.

This manuscript is organized as follows. In Section 2 the proposed WASS software is described in detail, following the general workflow that is performed during the processing of the stereo frames. In Section 3 we give a theoretical discussion of the expected reconstruction errors together with the internal and external factors that influence the overall reconstruction accuracy. Finally, in Section 4 we demonstrate our pipeline with data we acquired from an off-shore oceanographic platform.

2. The WASS pipeline

Our proposed sea-waves reconstruction tool completely automates the creation of a sequence of dense 3D point clouds from stereo images providing three important functionalities.

First, WASS can automatically recover the extrinsic parameters of the stereo rig (up to scale) so that no delicate calibration has to be performed on the field. Our past experience (Benetazzo et al., 2016) has demonstrated that if the intrinsic calibration is easy to obtain with all the commodities of a dedicated laboratory, extrinsic calibration is not. This is due to the fact that, for common setups covering areas ranging from 20×20 to 70×70 meters, a distance of at least a couple of meters is required between the two cameras. Consequently, it's rather difficult to calibrate the extrinsics on the field, forcing the usage of a big calibration target in a possibly dangerous environment (ie. out of a ship deck, on an offshore platform, etc). Moreover, the auto-calibration offers the advantage of letting the rig geometry reconfigurable on-the-fly before each acquisition to accommodate different requirements. For instance, it may be reasonable to take the device closer to the sea when the waves are small, so a small but highly resolved sea surface region can be acquired. On the other hand, large waves demand a broader sea surface region, requiring the device to be repositioned farther from the surface. Finally, the "calibrate once and for all" strategy is not reliable since vibrations of the support and environmental factors, as wind, can modify the relative angle between cameras and jeopardize the reconstruction accuracy.

The second feature provided by WASS is a fast 3D dense stereo reconstruction procedure so that an accurate 3D point cloud can be computed from each stereo pair. We rely on the well consolidated OpenCV library (Bradski and Kaehler, 2008) both for the image stereo rectification and disparity map recovery as described in the following section 2.4.

Third, a set of 2D and 3D filtering techniques both on the disparity map and the produced point cloud are implemented to remove almost all the erroneous points that can naturally arise while analyzing the optically complex nature of

the water surface (examples are sun-glares, large white-capped areas, fog and water areosol, etc).

As with the vast majority of dense stereo reconstruction methods, we make the implicit assumption of the Lambertian response of the water surface. More specifically, to search for corresponding matching features in the two stereo images we expect that the optical appearance on the two cameras is mostly preserved in terms of pixels intensities. In general, this is true only if the surface intensity response depends only on the reciprocal position between the illuminant and the surface normal, as in the Lambertian model. Any other phenomena like specularities and refractions are function of the mixed contribution between the illuminant, the surface normal and the viewpoint. Since the appearance of sea water surface is essentially described by the Fresnel law, the Lambertian assumption is quite limiting for many aspects. Nevertheless, under good illuminant conditions (no direct sun, cameras almost parallel and angled with respect to the sea plane, etc) classical dense stereo 3D reconstruction still provides excellent results compared to other methods like mechanical buoys or radars (Benetazzo et al., 2012).

2.1. Software architecture

In the present version, WASS is composed by 4 highly optimized executables written in C++ whose parallel execution in a pipeline is orchestrated by a script (*WASSjs*) running on the node.js runtime environment (Tilkov and Vinoski, 2010). The four executables are named respectively: *wass_prepare*, *wass_match*, *wass_autocalibrate* and *wass_stereo*.

wass_prepare is responsible for initializing of each couple of stereo images placing them into a proper "working directory" together with the necessary calibration and configuration data. Images are also undistorted according to the provided intrinsic calibration data.

wass_match and *wass_autocalibrate* are used in case that the extrinsic autocalibration is needed. The first matches corresponding features of a stereo image pair with a robust state-of-the-art game-theoretic framework (Albarelli et al., 2012). After the feature matching, the essential matrix is recovered and the extrinsic parameters between the two cameras are factorized (up to translation scale). Matched features are also filtered according to epipolar consistency. *wass_autocalibrate* collects all the matches between multiple stereo frames and uses Sparse Bundle Adjustment (SBA) to simultaneously optimize the extrinsic camera parameters and the 3D points triangulated from all the matches. Final extrinsic parameters are saved to all the workspaces.

wass_stereo performs stereo rectification and dense stereo 3D reconstruction on a given stereo frame. After the reconstruction, multiple filters are applied to the point cloud to remove erroneous points. Finally, the mean plane is robustly fitted to the reconstructed data so that it can be further aligned to a common sea reference frame.

Each WASS component runs as a single-threaded process performing a specific task on a working directory. To exploit the intrinsic parallelism of the stereo

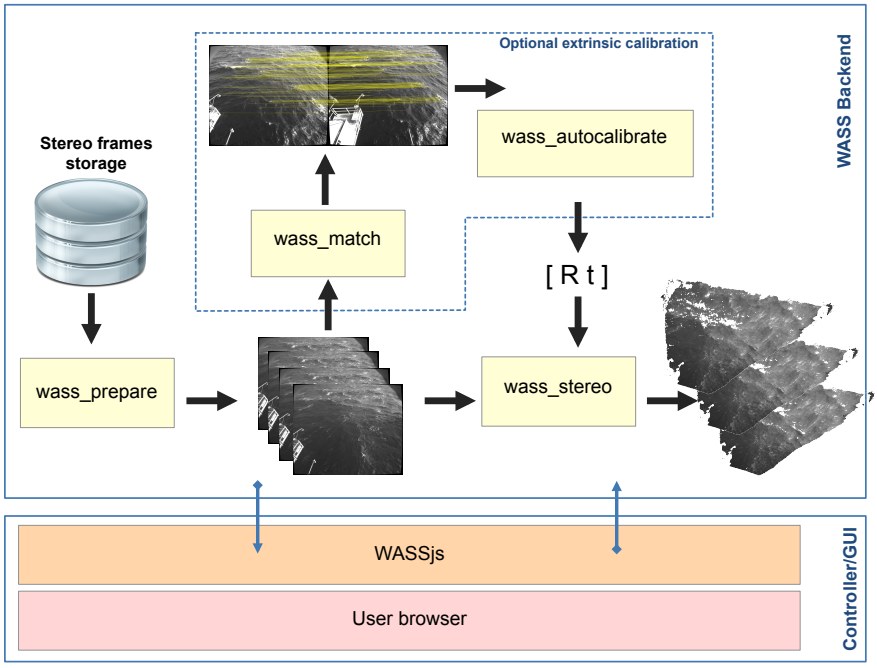


Figure 1: Components and workflow of the WASS reconstruction pipeline. User interacts with the pipeline via web browser communicating with the WASSjs component. *wass_prepare* and *wass_stereo* are responsible of the stereo frames undistortion and reconstruction respectively. The optional *wass_match* and *wass_autocalibrate* can automate the extrinsic parameters recovery of the stereo rig.

reconstruction task, the *WASSjs* controller manages multiple parallel instances of each component. As a consequence, WASS can scale well in distributed memory machines (like virtualized environments) with the only requirement to have a consistent shared view of a common filesystem.

2.2. Image preparation

The first step performed by the pipeline is to organize the stereo images into an ordered sequence of working directories so that all the subsequent steps can be easily managed in parallel. Apart from the images, intrinsic calibration for each camera has to be provided by the user. The pipeline assumes a pinhole camera model for each camera with (optionally) non square pixels and zero skewness. We also support a polynomial (5-coefficients) radial distortion (Fryer and Brown, 1986) that may be provided or set to zero.

The user is free to use whichever method he prefers for intrinsic calibration. For our work, we used both the Bouguet's Matlab calibration toolbox, implementing the method of (Heikkila and Silven, 1997), or the method proposed in (Albarelli et al., 2010) that is less sensible to the manufacturing accuracy of the

calibration target. Once the calibration is performed, data must be saved in an *xml* format compatible with the OpenCV serialization standard (see the official documentation for more informations).

2.3. Automatic stereo-rig calibration

The auto-calibration of the stereo rig is a complex task that can be optionally executed if no extrinsic calibration is provided by the user. It's based on the assumption that the relative pose of the two cameras remains unchanged throughout the sequence so that it can be recovered up to a positive scale by relating multiple corresponding features that have to be reliably extracted between the two cameras.

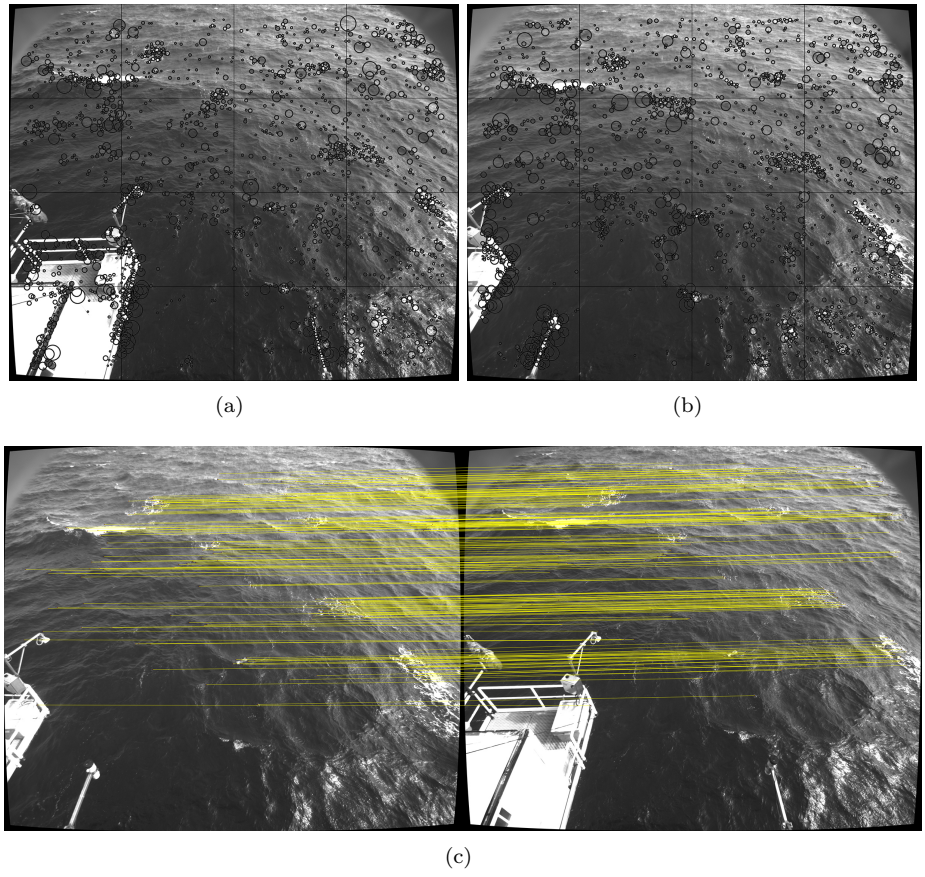
Supposing to have scene composed by a set of n 3D points projecting into the first and second camera as points $p_1 \dots p_n$ and $p'_1 \dots p'_n$ respectively, the following well-known epipolar constraint holds:

$$p_i = \mathbf{K}_1^{-T} \mathbf{E} \mathbf{K}_2^{-1} p'_i \ , \ i = 1 \dots n \tag{1}$$

where $\mathbf{E}$ is the essential matrix and $\mathbf{K}_1$ and $\mathbf{K}_2$ are the intrinsic camera matrices of the two cameras. This implies that, if a set of point-point correspondences can be matched from the two stereo images, the essential matrix can be recovered by solving equation (1). Then, the essential matrix can be further decomposed via Singular Value Decomposition (SVD) to recover the rotation $\mathbf{R}$ and translation direction $\vec{t}$ of the two views, as described in (Hartley and Zisserman, 2004).

WASS auto-calibration procedure is composed by the following steps:

1. A subset of all the available stereo frames are selected randomly as calibration images
2. A sparse set of *Speeded Up Robust Feature* (SURF) (Bay et al., 2008) is extracted from each image (Fig. 2 top row)
3. For each stereo pair, features are robustly matched with the approach of (Albarelli et al., 2012)
4. Erroneous matches are removed with a RANSAC-based epipolar filtering. (Fig. 2.c) As an outcome of this process, an initial estimate of the essential matrix is given for each stereo pair. The essential matrix associated to the pair with the highest number of feature matches is used as the initial coarse estimate for the next step.
5. All the matches between all the stereo frames of the calibration subset are used to simultaneously recover a 3D sparse point cloud of the matched features together with the camera poses with respect to a common unknown reference frame so that the global reprojection error is minimized. This effectively averages all the independent essential matrices estimations between each stereo pair used in the matching step. We perform this optimization as a Sparse Bundle Adjustment implemented as described in (Lourakis and Argyros, 2009)



(a) (b)

(c)

Figure 2: (a) Left and (b) right image captured by the stereo rig with the extracted SURF features superimposed. (c) Matched features (after epipolar filtering) computed by *wass_match* (best viewed in colors).

Steps 1 to 4 are performed by the *wass_match* executable that is run in parallel among all the randomly selected calibration images. Step 5 is performed by *wass_autocalibrate* as soon as the results of the previous matching process is available for all the calibration images. When *wass_autocalibrate* completes its job, the recovered rotation $\mathbf{R}$ and normalized translation $\vec{t}$ are saved on each workspace so that they can be used by the subsequent 3D reconstruction process. If the extrinsic parameters are provided by the user (either because they have been previously calibrated by the pipeline or with a calibration target) they can be automatically used to bypass the auto-calibration step described before.

2.4. 3D reconstruction and sea-plane estimation

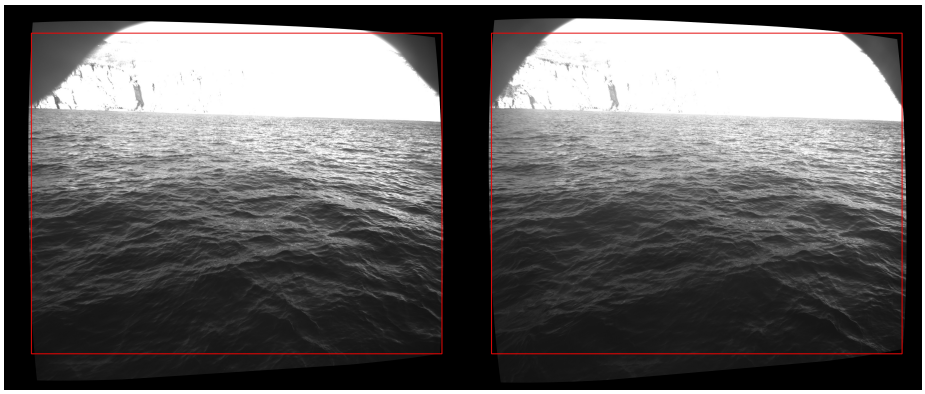
The 3D reconstruction process that computes a dense point cloud from the stereo frames is implemented in the *wass_stereo* executable. Again, we exploit

the temporal independence of the frames to execute the operation in parallel on all the available computation units so that the total computing time to reconstruct a whole sequence can be heavily reduced even on a common consumer workstation.

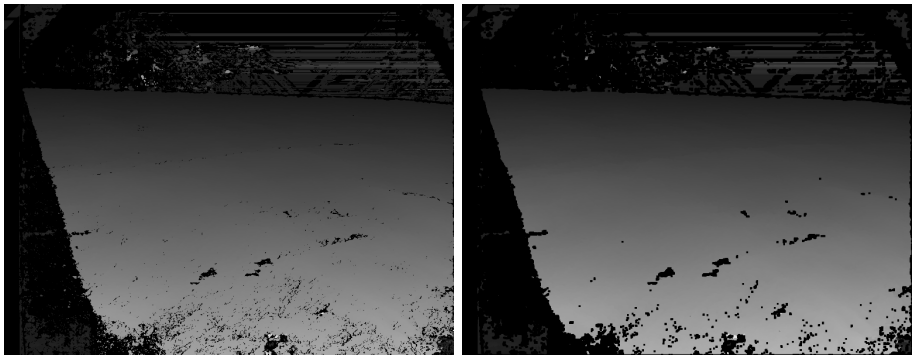
When *wass_stereo* is started, the two stereo frames are rectified using the extrinsic and intrinsic calibration data provided. At the end of this operation, all the corresponding epipolar lines will be perfectly aligned with the image rows (Fig. 3.a) so that the search of corresponding stereo pixels on the two images can be reduced along the same row during the dense stereo operation. The rectified stereo pair is also saved on each workspace for inspection purposes. In fact, the user can manually observe if the same corresponding visual features on left and right images are aligned on the same image row. If this happens, it gives a good feedback that both the intrinsic and extrinsic auto-calibration were performed with a reasonable accuracy.

After the rectification, a dense stereo algorithm is executed to obtain the disparity map between the two frames (Fig. 3.b). In our software we used the implementation provided by the OpenCV library (Bradski and Kaehler, 2008) that represent a de-facto standard reference implementation for many specialized computer vision algorithms. Specifically, we have chosen to use the semi-global dense stereo method described in (Hirschmüller, 2008) as it provides an excellent trade-off between the execution speed of a local method (ie. methods for which the cost aggregation step is performed locally along the disparity ranges) and the accuracy in low-textured areas of a global method (in which the disparity is recovered by means of an optimization involving all the image pixels). Since the obtained disparity map tends to exhibit errors at the boundaries of the matching regions, we perform some preliminary morphological filters directly on the disparity map to partially limit the outliers (Fig. 3.c). Specifically, we dilate the disparity map by n steps (usually 1 or 2 is sufficient) before eroding the results $m > n$ times. The initial dilation closes many of the small holes that may happen inside the observed sea surface due to sun glares or other artifacts. The subsequent erosion removes the disparity values at the boundary of the disparity map where most of the errors are located (since both the dense stereo cost computation and aggregation steps cannot be reliably performed).

With the cleaned disparity map, the initial unfiltered 3D point cloud is generated by triangulating all the corresponding pixels. In this step we filter all the points with a depth smaller than 1.0 (since $\vec{t}$ has unitary norm this essentially filters all the points nearest than the cameras baseline) and higher than *zmax* (ie. *zmax* times the camera baseline). Additionally, we filter all the points for which the angle between the two intersecting camera rays is lower than 20 degrees. Note that, since the corresponding points to be triangulated are always lying on the respective epipolar lines (the dense stereo algorithm operates along the rows of the rectified images) there is no need to use a computationally-expensive triangulation approach that minimizes the reprojection instead of the algebraic error. This greatly speeds up the 3D reconstruction that usually takes few seconds to process our 5 Mpixel images on a common Intel i7 CPU running



(a)



(b) (c)

Figure 3: (a) Rectified stereo-pair used as input for the dense stereo algorithm. Note that the corresponding visual features of the images are aligned on the same image row between the two frames. (b) Unfiltered disparity map obtained by the dense stereo algorithm. (c) Disparity map after the dilation and erosion filters. Note how the small holes at the bottom of the disparity map have been closed and the boundary of the disparity area is less noisy

at 2.5Ghz.

The obtained point cloud is in general already quite clean but a limited amount of spurious points can still be present. In fact, the dense stereo algorithm is designed to be general purpose so it does not make any assumptions on the reconstructed scene. Conversely, we can take advantage of the fact that the sea surface is spatially continuous and smooth everywhere. To greatly reduce the number of point outliers we use an additional filtering step that exploits the surface smoothness along the camera optical z-axis. We start by building a graph with the vertices being the reconstructed 3D points and edges connecting each vertex with its 4-neighbors considering the adjacency relation of points induced by the image lattice topology. In other words, we consider two 3D points connected if they have been triangulated from two adjacent pixels in the right image. We compute a weight associated to each edge as the absolute differ-

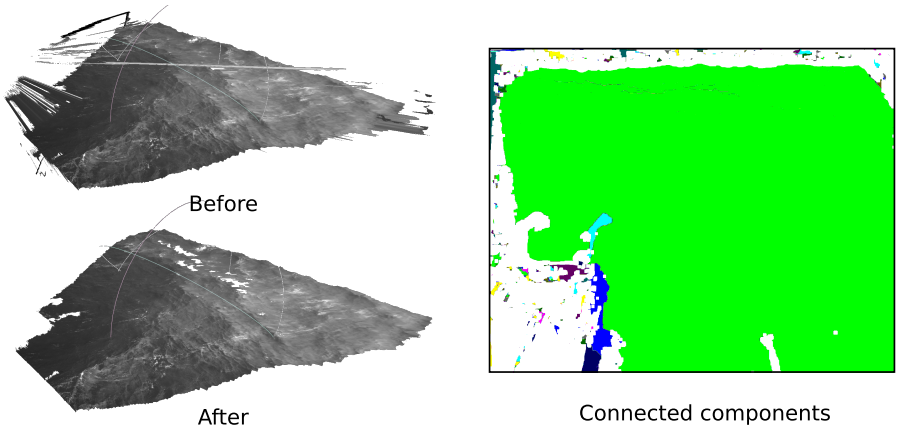


Figure 4: Effect of the graph-based filtering on the reconstructed point cloud. Left: Point cloud before and after the filtering. Right: Connected components extracted from the constructed graph. For the sake of visualization, a random color is assigned to each connected component except for the green which marks the biggest one.

ence along the camera z-axis of their respective points coordinates. Then, each edge whose weight is greater than the 98^{th} percentile of the weight distribution is pruned by the graph. The idea is to disconnect all the vertices exhibiting abrupt changes along the z-axis with respect to the neighbors. Finally, we filter out all the vertices and edges not belonging to the biggest connected component of the graph. Since the erroneous points are likely to exhibit abrupt changes with respect to the neighbours, they are efficiently removed from the principal observed surface. In Figure 4 we show an example of all the connected components extracted from the graph and the filtering result on the reconstructed point cloud. We can clearly observe that most of the reconstruction artifacts that happen at the boundary of the sea surface are effectively removed since they exhibit abrupt changes along the camera z-axis.

The final operation performed by *wass_stereo* is the estimation of the mean sea plane from the reconstructed point cloud. This allows a subsequent rotation of the 3D data so that the new Z-axis is oriented upward with respect to the earth surface. To limit the effect of possible point outliers that may have passed the previous filters, we perform an initial robust RANSAC (Fischler and Bolles, 1981) estimation of the mean plane before the least-squares fitting. Specifically, for a selected number of iterations we extract 3 random points from the point cloud and compute a candidate plane parameters. Then, we count the number of reconstructed points lying closer than a fixed threshold from the candidate plane. At the end of the RANSAC iterations, we keep the plane parameters corresponding to the candidate plane having the higher number of close points (inliers). This initial plane estimation is then refined with a least-squares fitting of all the inlier points by weighting each sample with a value proportional to the distance of the point from the camera. As previously assessed in (Bergamasco et al., 2016), this approach can improve the mean sea-plane recovery for wavy

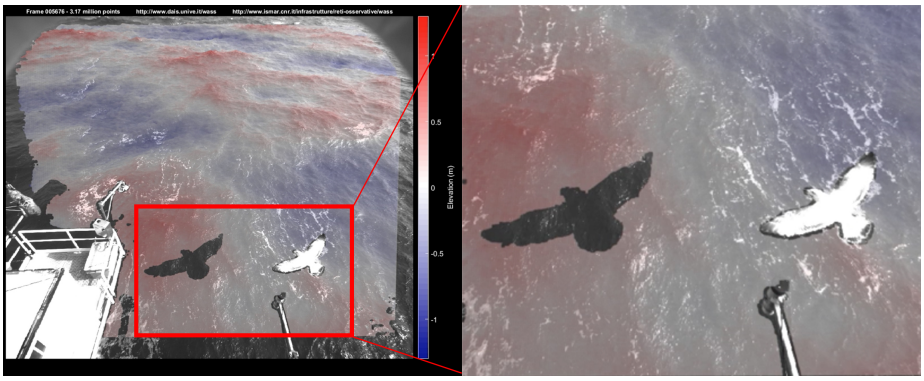


Figure 5: The implemented point cloud filtering approach is able to discard image portions clearly not belonging to the reconstructed sea surface area. Left: Example of a clutter object (a seagull in this case) in the field-of-view of the two cameras. Note how the reconstructed points corresponding to the seagull have been properly removed. The point removal is evident as it causes a seagull-shaped hole in the surface due to the shadowing produced by the seagull itself with the other camera field-of-view. Right: Magnification of the area highlighted in the red rectangle on the left. (Best viewed in colors)

seas especially if no other external aids are available (like horizon-line position or IMU data).

3. Expected Errors

Five principal uncertainties can be expected in any stereo method: the uncertainty in the internal parameter calibration (internal calibration error), the uncertainty in the external parameter calibration (external calibration error), the uncertainty in the determination of the corresponding pixels (matching error), the uncertainty in the recovery of 3D coordinates (quantization error), the uncertainty in the determination of the transformation between the camera reference system and the water reference system with the Z-axis vertical and pointing upward (camera orientation error). Following standard calibration procedures (Bouguet, 2004) we have verified that uncertainty in the internal calibration produces is of fractional of pixels, and we can consider them negligible for all applications. Yet this is not always the case for the external calibration. In previous WASS deployments (Benetazzo, 2006; Benetazzo et al., 2012), external calibration parameters were estimated by exposing an ad-hoc calibration target to both cameras, and by relating the known 3-D geometry of the target with its re-projection onto the image planes. However, even if this is the standard de-facto way to calibrate a stereo rig in laboratory conditions, this approach manifests several drawbacks when applied to stereo systems with large baseline, typical of field applications. In the latter the pipeline incorporates the auto-calibration procedure described in (Benetazzo et al., 2016), which has been proven to be as accurate as standard procedures. To determine the orientation and displacement between the camera reference system and the still sea water

surface, we have adopted the strategy proposed by (Benetazzo, 2006) that has been proved to be accurate on field (Benetazzo et al., 2012; Gallego et al., 2008) and synthetic (Benetazzo et al., 2016) data.

Concerning the matching and quantization errors, when designing a stereo-camera system one must compromise to meet mainly the conflicting requirements of accurate 3D estimation and accurate image feature matching. In fact, as observed by (Mironov et al., 2012) the major difference between stereo wave imaging and the classical problem of stereo reconstruction is in the fact that the water surface reflectance is not Lambertian. However, (Benetazzo, 2006) shown that the matching error is small for highly-textured water surfaces, and (Jahne, 1993) indicates that the matching error is small when the wave slope is much larger than the inclination of the stereo cameras optical axis. The geometry of the stereo rig must be tuned so that the interplay between the light vector, the point of view and the surface normal is similar for the two cameras. Thus, we assume that the disparity of each corresponding pixel is dominated by the spatial position of the 3D surface point and not by the rather complex water surface bi-directional reflectance distribution function (BRDF). To provide a good 3D reconstruction, care has to be taken to choose different parameters, the principal being the camera cell size and pixel numbers, the focal length, the baseline, the camera reciprocal orientation, and the distance from the stereo-camera system to the scene of interest. Under general conditions, in order to keep small the range error due to quantization (hence to provide high accuracy in the 3D reconstruction) the baseline-to-distance ratio must be large; however, accurate feature points matching require that this ratio be small (Benetazzo, 2006; Jahne, 1993; Rodriguez and Aggarwal, 1990).

Based on our experience on stereo wave imaging at the sea, the photometric consistency between the two views can be rather good if the following general principles are taken with care. First, the stereo camera optical axes should be placed (almost) parallel in order to reduce the angle between them. Second, the baseline-to-distance should be kept small, around 0.10, which is much smaller than historical set-up used for field experiment like in (Jahne, 1993). Third, the camera should be oriented so that no severe sun glitters are visible in the field of view. Finally, the quantization error is alleviated by using subpixel (fractional pixel) correspondence as performed by our pipeline. Sub-pixeling reduces errors depending on the weight function adopted to estimate the cross-correlation map (Nobach and Honkanen, 2005), and has a large impact on the accuracy of the smaller wavelengths (see for instance Fig.7 of (Benetazzo et al., 2012)). To give an idea of the reconstruction error expected with the proposed pipeline, in Figure 6 we show a map of the quantization error for the same WASS setup described in (Benetazzo et al., 2015).

4. Example

We use a field data example to demonstrate the capability of the open-source pipeline for WASS. The pipeline has already been used and its performance assessed in different studies using stereo cameras mounted on oceanographic tow-

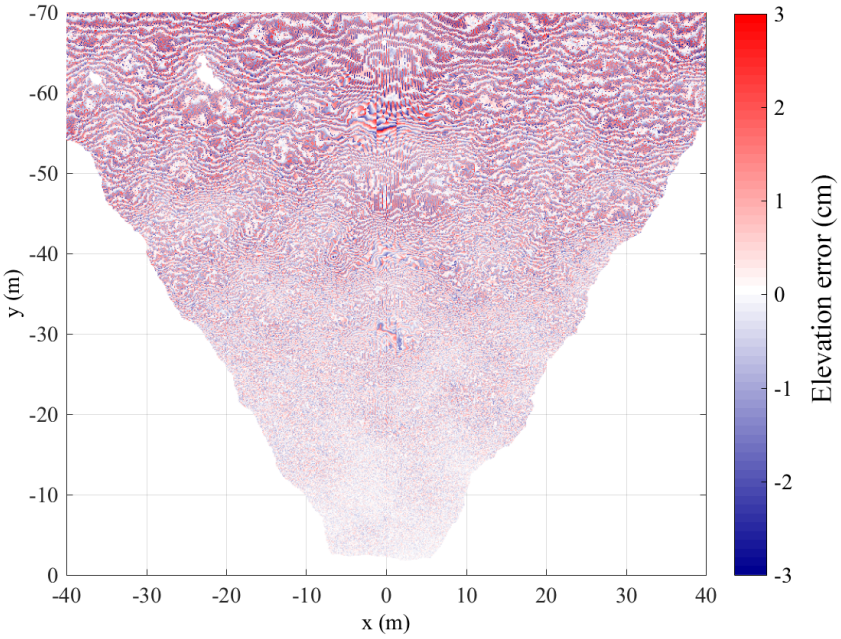


Figure 6: Map of the quantization error for sea surface elevations (Z-axis) within the horizontal field of view of the stereo cameras. WASS setup as in (Benetazzo et al., 2015)

ers (Benetazzo et al., 2015) and moving vessels (Benetazzo et al., 2016; Bergamasco et al., 2016; Falcieri et al., 2016). Here, the pipeline is tested on images acquired by a stereo wave imaging system ad-hoc installed on the oceanographic research platform Acqua Alta (North Adriatic Sea, Italy; see Figure 2). Image pairs were acquired with a 5 mega-pixel digital cameras (with 2456 columns by 2048 rows array of 3.45μm square active elements) and mounting 5 mm distortionless lenses, placed 2.5 m apart and on the roof of the platform.

The 3D wave fields were observed during a mature sea state generated by a north-easterly storm with average wind speed of 11 m/s, resulting in a local significant wave height of about 1.3 m. The WASS sequence comprises 26971 stereo-image pairs grabbed at 15 Hz. For internal parameter calibration of both cameras, we used the Bouguet's Matlab calibration toolbox (Bouguet, 2004).

Using the WASS pipeline, the automatic stereo rig calibration is performed selecting a subset containing 51 stereo frames (uniformly spaced along the image sequence). A total of about 20000 features are matched, and 89% of them retained after application of the RANSAC-based epipolar filtering. Then the rotation matrix $R = \begin{pmatrix} 0.998568 & 0.0297004 & -0.0444994 \\ -0.0311139 & 0.999022 & -0.0314147 \\ 0.0435229 & 0.0327543 & 0.998515 \end{pmatrix}$ and the normalized

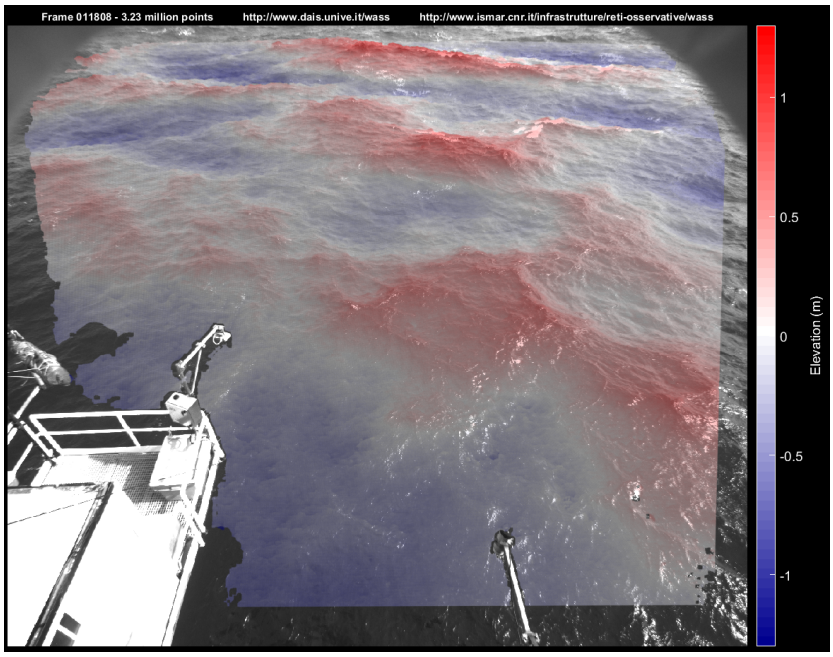


Figure 7: 3D point cloud result of the stereo triangulation superimposed to the left camera image of the stereo imaging system.

translation vector $\vec{t} = \begin{pmatrix} 0.999686 & 0.0101738 & 0.0229187 \end{pmatrix}^T$ between the two views are recovered. After the calibration and rectification of images, all stereo pairs are processed for 3D stereo reconstruction. The dense stereo algorithm execution find on average 3.3 million corresponding points, which are reduced of about 5% with the optimization and cleaning of the disparity map, including the graph-based filtering (see for example Figure 7). Along the whole sequence, about 80×10^9 3D sea-surface elevation points resulted from the stereo process. Then, as standard procedure, for each point cloud, elevations are transformed to a common earth reference frame (Benetazzo, 2006; Gallego et al., 2008). To this end, mean sea plane parameters are derived from each reconstructed point cloud, and then parameters averaged over the 26971 outputs. Finally, each reconstructed point cloud is scaled by a factor of 2.5 corresponding to the distance between the two cameras that have been empirically measured in the field. After this procedure, cameras resulted to be placed 12.5 m above the mean sea level, with an elevation angle of 50◦. Finally, to make stereo data useful for oceanographic applications, a patchwise planar surface is constructed by means of 2D Delaunay triangulation of each point cloud, resampled over a regular grid at uniform resolution of 0.2 m along the horizontal X- and Y-axis (in the world reference frame), spanning the horizontal region $x \in [-42.5m, 42.5m]$ and $y \in [-70.0m, -5.0m]$ (Fig.8).

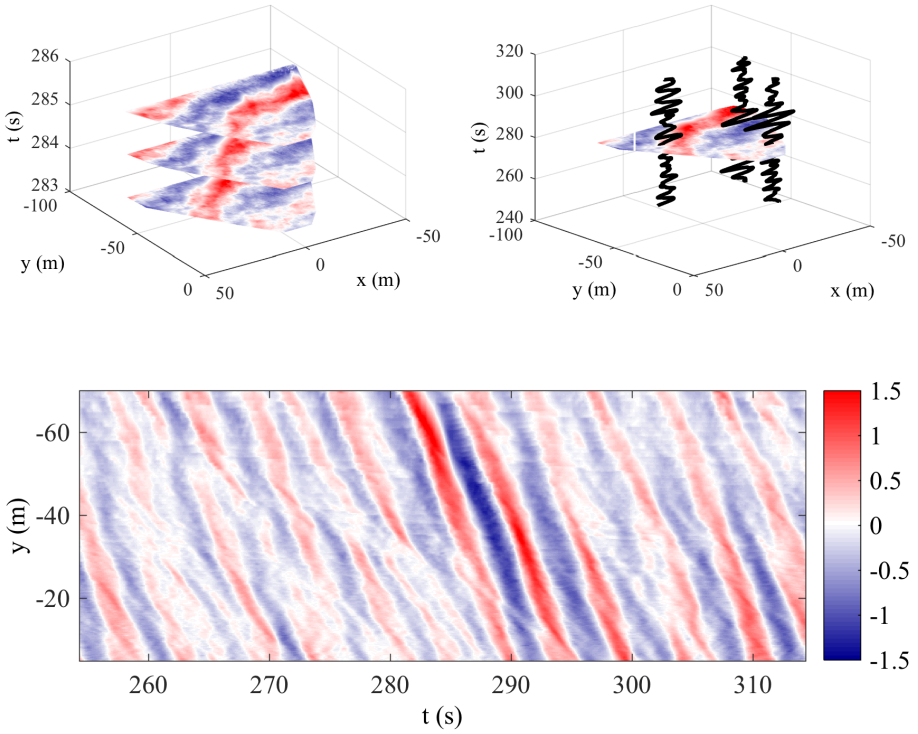


Figure 8: Representation of the 3D sea surface elevation fields measured by WASS. Top-Left: Slices on the XY-plane of the space-time volume. Top-right: Time series of sea elevations at different spatial positions. (bottom panel) Hovmoller diagram along the transect $x = 0$

5. Conclusions

We presented WASS, an open-source implementation of our sea waves 3D stereo reconstruction pipeline. WASS is designed to be fast and reliable when processing thousands of stereo frames and offers an intuitive interface that assist the user in all the common operations from dense stereo matching to stereo camera calibration. Tested and used extensively in our past works, we decided to release it to the whole oceanographic community in the hope that it can help improve the research in the study of sea waves.

We described the details of WASS inner working and demonstrated the performance and the expected errors of a typical WASS installation on a fixed oceanographic platform. We remand to the official website `http://www.dais.unive.it/wass` for additional technical details on installing and running our software on various platforms.

6. Acknowledgments

Authors gratefully acknowledge the funding from the Flagship Project RITMARE - The Italian Research for the Sea - coordinated by the Italian National Research Council and funded by the Italian Ministry of Education, University and Research within the National Research Program 2011–15

References

Albarelli, A., Rodolà, E., and Torsello, A. (2010). Robust camera calibration using inaccurate targets. *British Machine Vision Conference, BMVC 2010 - Proceedings*.

Albarelli, A., Rodolà, E., and Torsello, A. (2012). Imposing semi-local geometric constraints for accurate correspondences selection in structure from motion: A game-theoretic perspective. *International Journal of Computer Vision*, 97:36–53.

Banner, M. L., Barthelemy, X., Fedele, F., Allis, M., Benetazzo, A., Dias, F., and Peirson, W. L. (2014). Linking Reduced Breaking Crest Speeds to Unsteady Nonlinear Water Wave Group Behavior. *Physical Review Letters*, 112:114502, 1–5.

Banner, M. L., Jones, I. S. F., and Trinder, J. C. (1989). Wavenumber spectra of short gravity waves. *Journal of Fluid Mechanics*, 198:321–344.

Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Speeded-up robust features (surf). *Computer Vision and Image Understanding*, 110(3):346–359.

Bechle, A. J. and Wu, C. H. (2011). Virtual wave gauges based upon stereo imaging for measuring surface wave characteristics. *Coastal Engineering*, 58(4):305–316.

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

Benetazzo, A., Barbariol, F., Bergamasco, F., Torsello, A., Carniel, S., and Sclavo, M. (2015). Observation of extreme sea waves in a space-time ensemble. *Journal of Physical Oceanography*, 45(9):2261–2275.

Benetazzo, A., Barbariol, F., Bergamasco, F., Torsello, A., Carniel, S., and Sclavo, M. (2016). Stereo wave imaging from moving vessels: Practical use and applications. *Coastal Engineering*, 109:114–127.

Benetazzo, A., Bergamasco, F., Barbariol, F., Torsello, A., Carniel, S., and Sclavo, M. (2014). Towards an operational stereo system for directional wave measurements from moving platforms. *Proceedings of the International Conference on Offshore Mechanics and Arctic Engineering - OMAE*, 8B: Ocean Engineering.

Benetazzo, A., Fedele, F., Gallego, G., Shih, P.-C., and Yezzi, A. (2012). Offshore stereo measurements of gravity waves. *Coastal Engineering*, 64(0):127 – 138.

Bergamasco, F., Benetazzo, A., Barbariol, F., Carniel, S., and Sclavo, M. (2016). Multi-view horizon-driven sea plane estimation for stereo wave imaging on moving vessels. *Computers and Geosciences*, 95:105–117.

Borge, J. N., Reichert, K., and Hessner, K. (2013). Detection of spatio-temporal wave grouping properties by using temporal sequences of x-band radar images of the sea surface. *Ocean Modelling*, 61:21 – 37.

Bouguet, J. Y. (2004). Camera calibration toolbox for Matlab. Technical Report, Calif. Inst. of Technol. Available at: www.vision.caltech.edu/bouguetj/calib doc/index.html. Technical report, Calif. Inst. of Technol.

Bradski, G. and Kaehler, A. (2008). *Learning OpenCV: Computer Vision with the OpenCV Library*, volume 1. Inc, OReilly Media.

Brandt, A., Mann, J. L., Rennie, S. E., Herzog, A. P., and Criss, T. B. (2010). Three-Dimensional Imaging of the High Sea-State Wave Field Encompassing Ship Slamming Events. *Journal of Atmospheric and Oceanic Technology*, 27:737–752.

Campbell, A. J., Beckle, A., and Wu, C. H. (2014). Observations of surface waves interacting with ice using stereo imaging. *Journal of Geophysical Research: Oceans*, 119(-):3266–3284.

Cote, L. F., Davis, J. ., Marks, W., McGough, R. F., Mehr, E., Pierson, W., Ropek, J., Stephenson, G., and Vetter, R. (1960). The directional spectrum of a wind generated sea as determined from data obtained by the Stereo Wave Observation Project,. *Meteorological Papers*, pages 1–88.

de Vries, S., Hill, D., de Schipper, M., and Stive, M. (2011). Remote sensing of surf zone waves using stereo imaging. *Coastal Engineering*, 58(3):239–250.

Falcieri, F. M., Kantha, L., Benetazzo, A., Bergamasco, A., Bonaldo, D., Barbariol, F., Malač?, V., Sclavo, M., and Carniel, S. (2016). Turbulence observations in the Gulf of Trieste under moderate wind forcing and different water column stratification. *Ocean Science*, 12(2):433–449.

Fedele, F., Benetazzo, A., Gallego, G., Shih, P.-C., Yezzi, A., Barbariol, F., and Ardhuin, F. (2013). Spacetime measurements of oceanic sea states. *Ocean Modelling*, 70:103–115.

Fischler, M. and Bolles, R. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395.

Fryer, J. G. and Brown, D. C. (1986). Lens distortion for close-range photogrammetry. *Photogrammetric Engineering and Remote Sensing*, 52:51–58.

Gallego, G., Benetazzo, A., Yezzi, A., and Fedele, F. (2008). Wave Statistics and Spectra via a Variational Wave Acquisition Stereo System. In *ASME 2008 27th International Conference on Offshore Mechanics and Arctic Engineering*, pages 801–808.

Gallego, G., Yezzi, A., Fedele, F., and Benetazzo, A. (2011). A variational stereo method for the three-dimensional reconstruction of ocean waves. *Geoscience and Remote Sensing, IEEE Transactions on*, 49(11):4445–4457.

Hartley, R. and Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*. Cambridge University Press.

Hasselmann, K., Raney, R. K., Plant, W. J., Alpers, W., Shuchman, R. A., Lyzenga, D. R., Rufenach, C. L., and Tucker, M. J. (1985). Theory of synthetic aperture radar ocean imaging: A marsen view. *Journal of Geophysical Research: Oceans*, 90(C3):4659–4686.

Heikkila, J. and Silven, O. (1997). A four-step camera calibration procedure with implicit image correction. In *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, CVPR '97, pages 1106–, Washington, DC, USA. IEEE Computer Society.

Hirschmüller, H. (2008). Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):328–341.

Jahne, B. (1993). *Spatio-Temporal Image Processing: Theory and Scientific Applications*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Kosnik, M. V. and Dulov, V. a. (2011). Extraction of short wind wave spectra from stereo images of the sea surface. *Measurement Science and Technology*, 22(1):015504, 1–9.

Leckler, F., Ardhuin, F., Peureux, C., Benetazzo, A., Bergamasco, F., and Dulov, V. (2015). Analysis and interpretation of frequency-wavenumber spectra of young wind waves. *Journal of Physical Oceanography*, 45(10):2484–2496.

Lourakis, M. A. and Argyros, A. (2009). SBA: A Software Package for Generic Sparse Bundle Adjustment. *ACM Trans. Math. Software*, 36(1):1–30.

Mironov, A., Yurovskaya, M., Dulov, V., Hauser, D., and Guérin, C. (2012). Statistical characterization of short wind waves from stereo images of the sea surface. *Journal of Geophysical Research*, 117(October):C00J35.

Nobach, H. and Honkanen, M. (2005). Two-dimensional Gaussian regression for sub-pixel displacement estimation in particle image velocimetry or particle position estimation in particle tracking velocimetry. *Experiments in Fluids*, 38(4):511–515.

Rodriguez, J. and Aggarwal, J. K. (1990). Stochastic Analysis of Stereo Quantization Error. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:467–470.

Schumacher, A. (1939). Stereophotogrammetrische Wellenaufnahmen. Wiss. Ergeb. Dtsch. Atlant. Exped. Forschungs Vermessung. Meteor. Technical report.

Shemdin, H. and Tran, H. M. (1992). Measuring short surface waves with stereophotography. *Photogram. Eng. Remote Sens.*, 93:311–136.

Shemdin, O. H., Tran, H., and Wu, S. (1988). Directional Measurement of Short Ocean Waves With Stereophotography. *Journal of Geophysical Research*, 93(C11):13891–13901.

Sutherland, P. and Melville, W. K. (2013). Field measurements and scaling of ocean surface wave-breaking statistics. *Geophysical Research Letters*, 40(12):3074–3079.

Tilkov, S. and Vinoski, S. (2010). Node.js: Using javascript to build high-performance network programs. *IEEE Internet Computing*, 14(6):80–83.

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

Young, I. R., Rosenthal, W., and Ziemer, F. (1985). A three-dimensional analysis of marine radar images for the determination of ocean wave directionality and surface currents. *Journal of Geophysical Research: Oceans*, 90(C1):1049–1059.

Zappa, C. J., Banner, M. L., Schultz, H., Corrada-Emmanuel, A., Wolff, L. B., and Yalcin, J. (2008). Retrieval of short ocean wave slope using polarimetric imaging. *Measurement Science and Technology*, 19(5):055503.