

Measurements of short water waves using stereo matched image sequences

Alvise Benetazzo *

Dipartimento di Ingegneria Idraulica, Marittima, Ambientale e Geotecnica (IMAGE), University of Padua, via Loredan 20, 35131 Padova, Italy

Received 8 June 2005; received in revised form 12 June 2006; accepted 21 June 2006

Available online 9 October 2006

Abstract

Image analysis techniques are used for retrieving water surface elevation fields spatially and temporally from CCD-images and CCD-image-sequences. The technique proposed herein utilizes binocular stereogrammetry to recover topographic information from a sequence of synchronous, overlapping video images. The method used differs from the traditional stereo-photogrammetric analysis of a single stereo-pair because the use of video allows for a continuous sequence of stereo-images to be digitally sampled and analyzed. For data acquisition two synchronized progressive-scan cameras were used.

A partially supervised 3-D stereo system (called WASS, Wave Acquisition Stereo System) is shown here. It is used to reconstruct the 3-D shape of water surface waves, acquired at frame rate, with small computational time needed. The stereo method is presented, including the derivation of a relationship relating the geometry of the stereo rig and the expected errors. Finally, the 3-D calculated scattered points give the complete spatio-temporal distribution of the water surface elevations. The measurable length-scales depend on the pixel resolution, the triangulation accuracy, and the acquisition frame rate. Limitations in the stereo measurements are also discussed.

Two experiments to test and to demonstrate the system took place: one on the Venice lagoon, north of the city of Venice in September 2004 and the second on the coast of California at San Diego in December 2005. For the second experiment, qualitative and quantitative intercomparisons of the stereo-matching and in-situ sensor measurements are presented. All the measurements of water surface waves indicate that the proposed approach is both accurate and applicable for measuring water surface waves. Moreover, shape estimates are accurately and extremely dense both in space and time, and the remote location of the instruments alleviates some difficulties associated with in situ instrumentation.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Water waves; Stereography; Image sequences; Field measurements

1. Introduction

Water wave elevations are usually measured with different methods, which range from 1-D instruments immersed into water, through a quasi-2-D field of 1-D wave gauges, to 2-D radar-based systems installed on either fixed positions (Dankert et al., 2003) or aircrafts (Hwang et al., 2000). In this scenario, methods of Computer Vision (Ma et al., 2004; Ballard and Brown, 1982) can be successfully exploited to obtain 3-D elevation maps of water waves by using a spatial-widespread and non-intrusive optical system.

Moreover, extending image processing from static individual images to image sequences of stereo-pairs opens a new world of information. In fact, the traditional point measuring probes only deliver time series of a quantity (i.e. the water surface elevation)

at a fixed or moving point in space. Single images give only spatial information (2-D in the case of a single image, 3-D after the stereo reconstruction), while image sequences contain both the temporal and spatial structure of the phenomena observed. Except for limitations in resolution in time and space and the size of the images, image sequences capture the events as completely as possible.

Images of the water surface in the visible range of the electromagnetic spectrum have for many years provided quantitative information on a number of physical parameters associated with the waves which ripple such a surface (Dugan et al., 2001; Stockdon and Holman, 2000). The application of stereo vision to measure the water surface topography starts from the conventional stereographic technique algorithms used to survey geodetical surfaces or static objects (Klette et al., 1998). The major differences depend on the fact that the water surface is a specular object in (rapid) movement. For this reason each stereo-pair is acquired

* Tel.: +39 049 827 5424; fax: +39 049 827 5446.

E-mail address: alvise.benetazzo@gmail.com.

simultaneously, and the geometry of the stereo system is such that the errors due to the water surface specular property are minimized. In this way, the first adaptations were implemented in an experiment with cameras mounted on an ocean going ship (Schumacher, 1939). More recently Côté et al. (1960) demonstrated, for long ocean waves, the use of stereophotography to measure sea surface topography. The effort required to extract 3-D elevation maps from an image pair limited the use of this technique in studying the dynamics of oceanographic phenomena until late 70s and early 80s (Sugimori, 1975; Holthuijsen, 1983). Nevertheless, the need to know directional information of the waves allowed the stereography to remain one of the investigation tools in the oceanographic studies. Shemdin et al. (1988) proposed the directional measurements of short ocean waves applying stereography. This experiment used a pair of cameras mounted on an oceanographic offshore tower near San Diego (USA) to create the 3-D elevation map of the sea surface and then, via spectral analysis, to extract directional information of the waves under inspection. In 1989, Banner et al. applied stereographic measurements in order to study wavenumber spectra of short gravity waves. The most recent integration of stereo graphic techniques into the field of oceanography was the WAVESCAN project (Santel et al., 2004) which contributed to the important extension in the time domain of the stereographic pairs.

The work presented here proposes a technique to estimate the shape of water waves using video image analysis that is inexpensive and partially automated with high spatial and temporal resolutions. This technique employs stereo vision based on two calibrated views (an extension to three views is possible with a reduction of the general errors) for providing time series of scattered 3-D points of the water surface from a stack of synchronous, overlapping images. This paper focuses on the stereo processing with error analysis, the instrumental aspects, data processing techniques, and quantification of water surface parameters, as phase speeds, wavenumbers, frequencies and wave heights.

The following section describes the principles of the stereo method, with a discussion of errors. Next, the data analysis techniques are presented, and finally the stereo method is applied to water wave 3-D mapping with results from two field experiments.

2. The Wave Acquisition Stereo System (WASS)

3-D reconstruction is one of the most popular problems in Computer Vision (Faugeras, 1993). One of the earliest methods in this field is stereo analysis, which uses two or more cameras to obtain the 3-D shape of the scene. Stereo analysis can be static if two still images are acquired and used for reconstruction, or dynamic when a movie of the scene is acquired.

The main problem in stereo analysis is how to find, given a point in the first image, the corresponding point in the second image. This problem is commonly known as the correspondence problem (Klette et al., 1998). Usually the two cameras are fixed and precalibrated (Zhang, 2000), and the search of correspondences is limited to epipolar lines (Pollefeys, 2000). If the scene is dynamic, i.e. a movie is acquired and objects in the scene are moving, motion can be used to help the research of

correspondences and correspondences can be used to infer the motion of rigid bodies in the scene (Kanade and Morris, 1998). This is not the case here, since water waves are definitely far from rigid bodies, and ad-hoc algorithm must be implemented.

2.1. The stereo camera model

In this project the perspective camera model is used. This corresponds to an ideal pinhole camera (Jähne, 1993). Three coordinate systems must be set (Fig. 1). The first coordinate system is fixed to the camera observing the scene, and the coordinates are denoted as $\mathbf{X}=[X,Y,Z]^T$, where the Z axis coincides with the camera line of sight, and the X and Y axes are parallel to the CCD axes (Charge Coupled Device, i.e. the sensible element of the camera). The camera coordinate system origin is the so-called focal point. Let $\mathbf{X}_0=[X_0, Y_0, Z_0]^T$ the 3-D spatial coordinates of a point relative to the Cartesian camera coordinate system, the second coordinate system defines the 2-D image coordinates (in pixels) $\mathbf{J}_0=[j_0, i_0]^T$ of the same point. With a CCD camera the relation between these two coordinate systems depends on the focal length, the size and shape of the pixels, and the position of the CCD chip in the camera-lens system. According to the pinhole approximation, the image coordinates of the point having 3-D coordinates given by $[X, Y, Z]^T$ are:

$$\mathbf{Z} \begin{bmatrix} j \\ i \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & o_x \\ 0 & f_y & o_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (1)$$

with f_x and f_y being the focal length measured in width and height of the pixels, and $[o_x, o_y]^T$ the principal point image coordinates, i.e. the coordinates of the intersection of the camera line of sight with the CCD plane.

A third coordinate system is related to the observed scene, and the coordinates are called world coordinates and denoted as

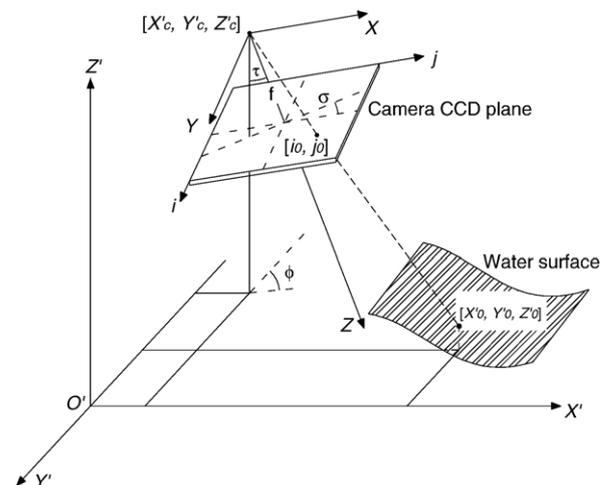


Fig. 1. Relationship between the camera focal point $[X'_c, Y'_c, Z'_c]$, i.e. the origin of the camera coordinate system, camera coordinates $[X, Y, Z]$, image coordinates $[i, j]$, world coordinates $[X', Y', Z']$, water surface point $[X_0, Y_0, Z_0]$, and rotation angles (ϕ, τ, σ) used in the orientation definition. f is the focal length of the camera-lens system. In figure, the camera CCD and the focal length are magnified.

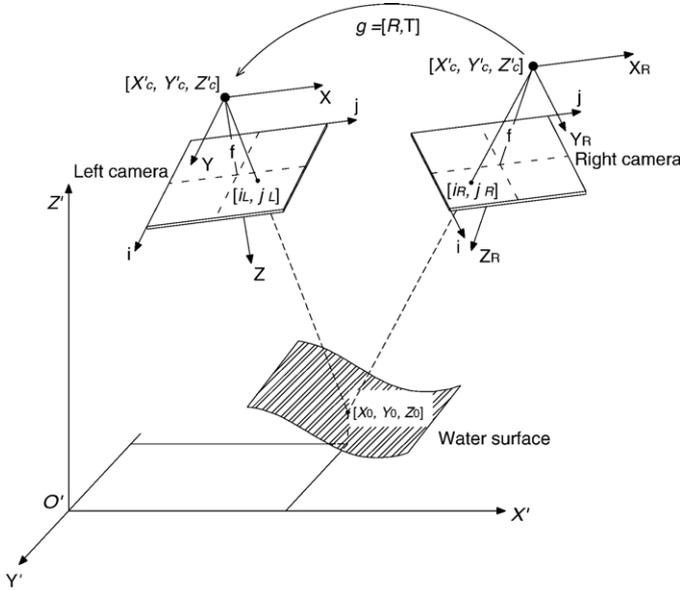


Fig. 2. Set-up of two camera (Left and Right in Figure) stereometric intersection which allows to determinate the three dimensions of a water surface point $[X_0, Y_0, Z_0]$. $[i_L, j_L]$ and $[i_R, j_R]$ are the Left and Right camera coordinates of the point $[X_0, Y_0, Z_0]$. $g=[R, T]$ is the rigid transformation between Right $[X_R, Y_R, Z_R]$ and Left $[X, Y, Z]$ camera coordinate systems. In figure, the cameras' CCDs and the focal lengths are magnified.

$\mathbf{X}'=[X', Y', Z']^T$. In this coordinate system, the Z' axis is parallel to the vertical direction, and the (X', Y') plane is parallel to the horizontal plane. The orientation between the camera and world coordinate systems is given by three successive rotations (ϕ , azimuth; τ , tilt; σ , roll) about the axes (Fig. 1). The complete transformation from world coordinate to camera coordinates (and vice versa) is expressed by the rigid-body motion $g=[R, T]$, which in matrix form becomes

$$g = \begin{bmatrix} R & \mathbf{T} \\ \mathbf{0}_3^T & 1 \end{bmatrix}. \quad (2)$$

and then

$$\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} R & \mathbf{T} \\ \mathbf{0}_3^T & 1 \end{bmatrix} \begin{bmatrix} X' \\ Y' \\ Z' \\ 1 \end{bmatrix} \quad (3)$$

where R is the rotation matrix (Ma et al., 2004) defined by the three angles (ϕ , τ , σ), and \mathbf{T} the translation vector between the origins of the camera and world coordinate systems.

The complete transformation from world coordinates to image coordinates can be then composed of elementary matrices:

$$Z \begin{bmatrix} j \\ i \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & o_x \\ 0 & f_y & o_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & \mathbf{T} \\ \mathbf{0}_3^T & 1 \end{bmatrix} \begin{bmatrix} X' \\ Y' \\ Z' \\ 1 \end{bmatrix} \quad (4)$$

which can be simplified to

$$Z \begin{bmatrix} \mathbf{J} \\ 1 \end{bmatrix} = K \Pi_0 g \begin{bmatrix} \mathbf{X}' \\ 1 \end{bmatrix} \quad (5)$$

The constant 3×4 matrix Π_0 represents the perspective projection. Equivalently

$$\begin{bmatrix} \mathbf{J} \\ 1 \end{bmatrix} \sim P \begin{bmatrix} \mathbf{X}' \\ 1 \end{bmatrix} \quad (6)$$

where the 3×4 matrix $P=K\Pi_0g$ is called camera projection matrix.

The calculation of the 3-D coordinates from corresponding points into the two image planes composing the stereo rig is based on a triangulation resolved by a least square solution. To do this it is necessary to manipulate the camera projection matrix P decomposing it in 4 sub-matrices:

$$P = \begin{bmatrix} D_{2 \times 3} & b_{2 \times 1} \\ q_{1 \times 3} & \gamma_{1 \times 1} \end{bmatrix} \quad (7)$$

so that the projection process can be written as

$$Z \begin{bmatrix} j \\ i \\ 1 \end{bmatrix} = \begin{bmatrix} D_{2 \times 3} & b_{2 \times 1} \\ q_{1 \times 3} & \gamma_{1 \times 1} \end{bmatrix} \begin{bmatrix} X' \\ Y' \\ Z' \\ 1 \end{bmatrix} \quad (8)$$

where Z is thought as a proportional coefficient.

From it follows that:

$$Z \begin{bmatrix} j \\ i \\ 1 \end{bmatrix} = Z \begin{bmatrix} \mathbf{J} \\ 1 \end{bmatrix} = \begin{bmatrix} D\mathbf{X}' + b \\ q\mathbf{X}' + \gamma \end{bmatrix} \quad (9)$$

which can be decomposed in the system:

$$\begin{cases} \mathbf{J} = (D\mathbf{X}' + b)/Z \\ Z = q\mathbf{X}' + \gamma \end{cases} \quad (10)$$

The camera point \mathbf{J} is then expressed by the expression:

$$\mathbf{J} = \frac{D\mathbf{X}' + b}{q\mathbf{X}' + \gamma} \quad (11)$$

and then the system (6) can be rewritten as

$$\mathbf{J}(q\mathbf{X}' + \gamma) = D\mathbf{X}' + b \quad (12)$$

or

$$(\mathbf{J}q - D)\mathbf{X}' = b - \gamma\mathbf{J} \quad (13)$$

Knowing the image coordinates (\mathbf{J}) of the same world point (\mathbf{X}') on the two camera planes (L: Left camera; R: Right camera, as depicted in Fig. 2), the three coordinates $[X', Y', Z']^T$ are determined solving the system:

$$\begin{cases} (\mathbf{J}_L q_L - D_L)\mathbf{X}' = b_L - \gamma_L \mathbf{J}_L \\ (\mathbf{J}_R q_R - D_R)\mathbf{X}' = b_R - \gamma_R \mathbf{J}_R \end{cases} \quad (14)$$

The previous system in three unknown and four equations cannot be further simplified, and then it must be solved minimizing the squared errors, after each row of the system has been normalized, since the least square method minimizes the distances following a metric proportional to the modulus of each row.

The system (14) is written in terms of 3-D world coordinates \mathbf{X}' . In our case, the world coordinate system is a priori unknown. This means that in the stereo method here used, the reference coordinate system is set coincident with the left camera coordinate system of Fig. 2. This implies that rotation matrix R and the translation vector \mathbf{T} define the mutual position of the cameras' coordinate systems, say the rigid transformation from the right coordinate system to the left coordinate system: $\mathbf{X} = \mathbf{X}_L = R\mathbf{X}_R + \mathbf{T}$. This rigid motion is obtained by a camera calibration procedure (Bouguet, 2004). In the same way even the effective focal length and principal point coordinates of Eq. (1) are determined.

2.2. The stereo-matching algorithm

The stereo-method here proposed is pixel-correlation-based. Window size is kept fixed, since features of water that allow reconstruction (i.e. the water surface reflections) are quite small and homogeneous. A window size of 11×11 pixels centered in the analyzed pixel was found as an optimal compromise between reliability of the matched points and computational time. Only if the variance of the image gray values in this window is higher than a selected threshold, the search of the corresponding pixel starts. To automatically find the corresponding pixel in the stereo-images the principles of epipolar geometry are exploited (Ma et al., 2004). In particular, even if in an optical system the exact position of the 3-D point \mathbf{X}_0 is not known, it is bound to lie on the optical axis passing through the corresponding image point \mathbf{J}_0 (Fig. 1). In a stereo system (Fig. 2), if \mathbf{J}_L is the left image point of a 3-D point, the unknown corresponding right image point \mathbf{J}_R will be on the projection of the optical axis, relative to \mathbf{J}_L , on the right image plane. Manipulating the system (14), the equation of this projection (called epipolar line) can be derived (Benetazzo, 2006). On the epipolar line, the corresponding point is searched.

To find the correspondences, cross-correlation is used and a pyramidal search is implemented. The pyramidal process of stereo matching starts from few pixels on a regular grid on the left camera plane, and then iterates the process refining the regular grid in subsequent steps, to obtain a dense disparity map. At the first step, for each pixel of the grid, the correspondence is searched over all the epipolar line on the right camera plane. The right image pixel whose cross-correlation coefficient is highest is selected as the corresponding point. Only if this coefficient goes over a threshold value (set equal to 0.85) the position of the corresponding pixels in the two image planes and the cameras' projection matrices are used to determine the 3-D coordinates $[X, Y, Z]^T$ of the water surface at the selected point. In the first step, the few pixels of the grid allow only a first approximation of the water surface. Then, the hypothesis that the water surface can be locally approximated by a planar surface is exploited (Fig. 3). In the second step, the grid is refined and for each pixel of the new grid this hypothesis constrains the possible location of the corresponding point over a small portion of the epipolar line. This reduces the number of cross-correlations that must be computed and saves computational time. New correspondences are defined, and if the maximum cross-correlation coefficients go over the selected threshold, new 3-D position of the water

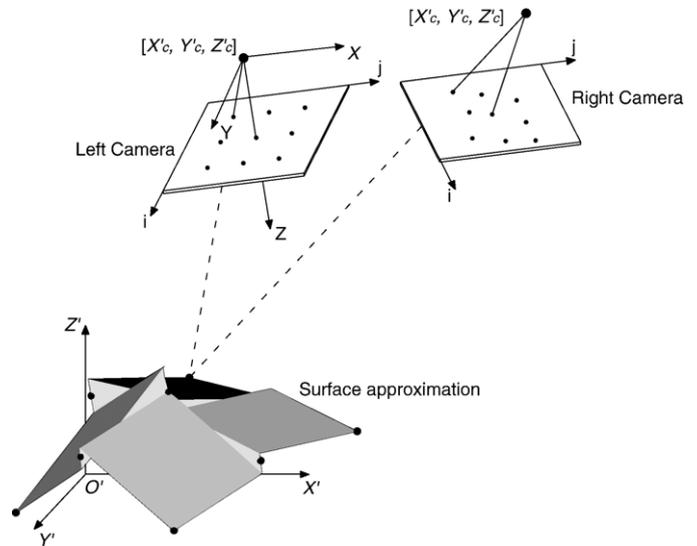


Fig. 3. First step of the pyramidal search. On the Left camera image, pixels on a regular grid are selected (black circles). The search algorithm determines the position of the corresponding pixels in the Right image (black circles). By exploiting the stereo method the 3-D positions (in the Left camera Cartesian coordinate system) of the selected points are determined. These give a first approximation of the surface being analyzed. Locally, the surface is approximated by a plane. The dashed line represents an example of the intersection of the cameras' lines of sight. In figure the cameras' CCDs and the 3-D surface are magnified.

surface are determined. The 3-D water surface shape is refined with new points, which allow to estimate a new local planar approximation. The process goes on, with a grid refinement until all the image pixels are stereo analyzed (Fig. 4). Details of the algorithm can be found in Benetazzo (2006). For each stereo-pair (left and right image), the final result is a sequence of 3-D coordinates, scattered in conformity with the 3-D water surface shape and the distribution of the matched pixels.

Since a movie, with a certain frequency, of the scene was acquired, the problem is a dynamic stereo reconstruction. With this amount of data, the 3-D location of corresponding points, obtained from the first couple of stereo-images, is exploited to reduce the region of epipolar search in the next frames. The reciprocal location of the corresponding points into the two stereo-images defines the disparity map. Once the disparity map at time $t=t_0$ is determined, the disparity map at time $t=t_0+t_s$ will be estimated. This estimation requires a preliminary evaluation of the characteristics of the waves (maximum speed and maximum elevation) to determine, in the next frame, which could be the maximum evolution of the disparity map. This evolution fixes a research range in the epipolar line. This constraint has been implemented in the stereo method and it allows an overall reduction in computational time of about 80%. However, if the frequency of acquisition is in the order of some frames-per-second (that will be the case for an industrial system, acquiring images many hours per day a scene far from the cameras), the disparity map estimation will fail. In this project, it is only at the end of the reconstruction that the movement of the water waves will be inferred, but during the acquisition little estimation can be done to connect points in

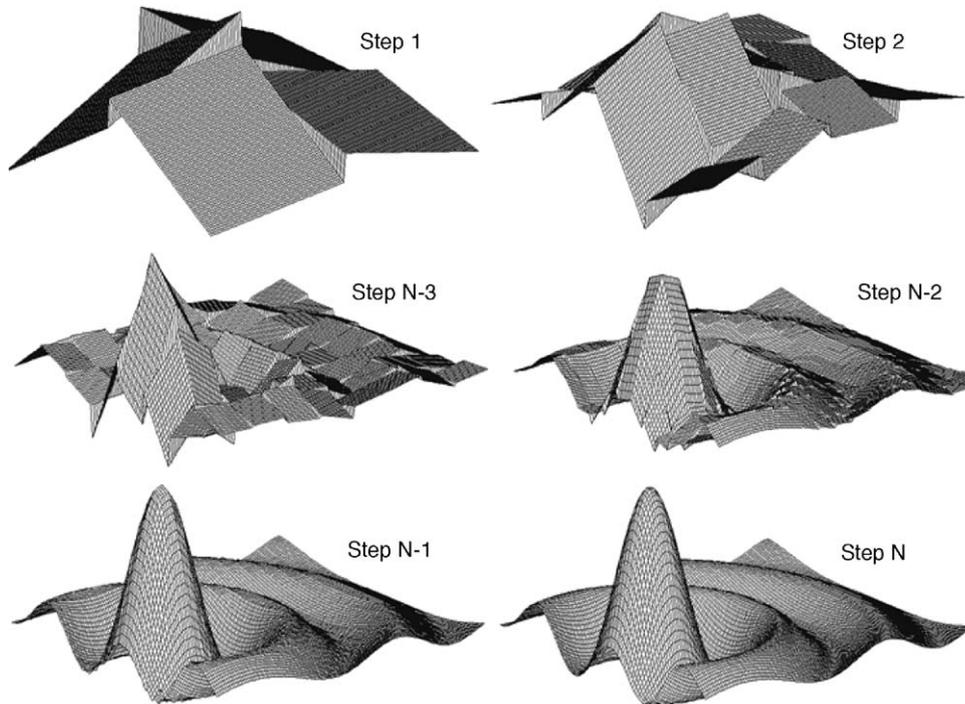


Fig. 4. Pyramidal search sketch. The final 3-D surface is obtained through approximations. At first step (Step 1), pixels on a regular grid are stereo-matched, and their 3-D coordinates are determined (Fig. 3). Then, around the known 3-D points the surface is approximated with planes, in order to obtain a first approximation of the final surface. In the second step the grid is refined, and 3-D locations of new corresponding pixels is expected to be close this approximation. This fact restricts the epipolar line portion where seek for the corresponding points. New vertices of the final 3-D shape are found (top-right panel, Step 2). The process is then refined until the complete matching (Step N-3 to Step N).

different frames. In that case the problem becomes much closer to static stereo reconstruction.

2.3. Camera-to-world coordinate transformation

After the triangulation process, the positions of the 3-D water surface points in space and time are given in the left camera coordinate system. Afterwards, to analyze these data it is necessary to rotate the 3-D points in order to set the Z axis aligned in the direction of the gravity acceleration. Generally, no known references that allow a reliable Z' axis to be computed are present in the images. A challenging application, at least from an operational point of view, is the transformation of 3-D coordinates from the camera coordinate system (image plane and camera line of sight) to world coordinate system (horizontal plane and vertical direction). This rigid transformation is completely described by a rotation matrix, R , and a translation vector \mathbf{T} , such that $\mathbf{X}' = R\mathbf{X} + \mathbf{T}$.

For a given sequence, the gravity direction can be found exploiting the hypothesis that the mean time water surface elevation map is a plane, representing the water surface without waves, orthogonal to gravity direction. In fact, it is fair to assume that over a long period of time (say, at least ten times the dominant period of the waves), the time–mean water surface elevation on a single point on the surface will be zero (if no long period waves are observed).

In the camera coordinate system, the equation of the mean plane ($aX + bY + cZ + d = 0$) is obtained through a least square solution. For each point, the water elevation (in the world

coordinate system) is calculated as the distance from the mean plane. Then, the last row of the rotation matrix, R , is given by

$$r_3 = \frac{[a, b, -1]^T}{\sqrt{a^2 + b^2 + 1}} \quad (15)$$

For a given $[X, Y, Z]^T$ point, the water surface elevation evaluated from the horizontal mean plane can then be expressed as

$$Z' = r_3 \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \frac{d}{\sqrt{a^2 + b^2 + 1}} \quad (16)$$

The other two rows are computed with the assumption that R is orthogonal and normalized:

$$r_1 = \frac{[a, -(1 + a^2)/b, -1]^T}{\sqrt{a^2 + ((1 + a^2)/b)^2 + 1}} \quad (17)$$

$$r_2 = \frac{[b, 0, ab]^T}{\sqrt{a^2b^2 + b^2}} \quad (18)$$

and finally

$$R = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} \quad (19)$$

where a and b are the parameters from the equation of the mean plane.

Table 1
Summary of the characteristics of the stereo set-up used in different investigation to measure the 3-D shape of water waves

| Reference | $ \mathbf{T} /Z_0$ | Size (m×m) | Z-Res (mm) | X/Y-Res (mm) |
|-----------------------|--------------------|-------------------|---------------|-----------------|
| Coté et al. (1960) | 0.6 | 900×600 | 150 | 600 |
| Dobson (1970) | n/a | 3.6×3.6 | 0.75 | n/a |
| Holthuijsen (1983) | ≈ 0.5 | (54–220)×(54–220) | n/a | n/a |
| Shemdin et al. (1988) | 0.4 | 2.5×2.5 | 3 | 8.5 |
| Banner et al. (1989) | 0.3–0.4 | 2.0×2.0 | 1 | n/a |
| Santel et al. (2004) | 0.25–1.0 | 200×200 | 40 | 40–80 |

Thus, the transformation between camera and world coordinate systems is given by

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = R \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \frac{d}{\sqrt{a^2 + b^2 + 1}} \end{bmatrix} \quad (20)$$

The individual elements R_{ij} of the 3×3 orthonormal rotation matrix are exploited to calculate the three successive rotations (ϕ , azimuth; τ , tilt; σ , roll) about the camera axes (Fig. 1).

2.4. Analysis of the errors

In this project a window stereo matching algorithm is used. Usually stereo matching works properly if the surface is not reflective and has Lambertian properties (Jähne, 1997). However it can be shown (Jähne, 1993) that some conditions in the position of cameras and characteristics of the surface allow the reflections to be considered as texture on the water surface. After stereo-matching and triangulation, elevations of the waves can be determined with a bias, which is due to the specular properties of the surface. Such bias depends on both the wavelength of the waves and their steepness. It is small for steep waves, i.e. when the angle formed by the slope with the horizontal plane is much larger than the inclination of the lines of sight of the cameras. Thus, this error can be reduced using a small angle between the lines of sight of the cameras; however, this increases the quantization error. A trade-off has to be found: the geometry of the camera set-up used for this work is characterized by a base to height ratio, $|\mathbf{T}|/Z_0$ (Fig. 2), of about 0.1, and the inclination of the line of sight of the cameras is about 0.05 radians. This is a very small value, if compared with typical maximum slopes of waves: therefore waves can be reconstructed with accuracy. A summary (after Jähne, 1993) of the characteristics of the stereo camera set-up used in recent investigations to measure the height of ocean waves is summarized in Table 1, where Res is the resolution. We justify the high values of the ratio $|\mathbf{T}|/Z_0$ given in Table 1, through the need, for the authors, to keep an high resolution. The water surface specularly was not a problem, since the stereo-correlation in all the works cited was done using an analytical plotter instead of an automatic algorithm as in the project presented herein.

In terms of errors, a systematic source of error, intrinsic in the stereo process, is the resolution error, called quantization error. It arises from the pixels quantization operated by the CCD, and

from the triangulation process. The quantization error is not constant in the matched area. Only an estimation of its maximum value can be obtained. The 2-D model of the stereo rig geometry is shown in Fig. 5, where the \hat{Z} axis is perpendicular to the baseline b , and the \hat{X} axis is parallel to the baseline. The achievable 3-D accuracy is influenced by the object distance (\hat{Z}), the focal length (f), the camera resolution (i.e. the CCD physical size, equal to the number of pixels times the CCD cell size) and the cameras' mutual position (\mathbf{T}), as well as the angle between the cameras' lines of sight (α). The m and n axes are the image coordinates of the pixels on the first and the second image, relative to the principal points of the CCDs. In this 2-D model the CCD is approximated with a only one axis, i.e. it degenerates to a segment. If N is the number of pixels of this 1-D CCD, the quantization effect is maximum when $m = -N/2$, $n = N/2$. The maximum absolute error, $er_{\hat{z}}$, along the \hat{Z} axis is given, for a distance \hat{Z} from the baseline, by:

$$er_{\hat{z}} = \frac{\hat{Z}^2}{2|\mathbf{T}|N} \frac{\sin 2\beta}{\cos(\beta + \alpha)^2} \quad (21)$$

where β is half the angle of view of the cameras. The maximum errors on \hat{X} axis is only due to pixel resolution (differently from the error on \hat{X} axis, which is due to the triangulation), and it is equal to:

$$er_{\hat{x}} = \frac{\hat{Z}}{2N} \frac{\sin 2\beta}{\cos(\beta + \alpha)^2} \quad (22)$$

The extension to the third axis \hat{Y} (such that \hat{X} , \hat{Y} , \hat{Z} axes are an orthonormal space) gives for the maximum error the relation

$$er_{\hat{y}} = \frac{\hat{Z}}{2N} \frac{\sin 2\beta}{\cos(\beta)^2} \quad (23)$$

The \hat{X} , \hat{Y} , \hat{Z} axes are considered parallel to the X , Y , Z axes of Fig. 1 given the small values of α . This means that the previous formulae will be exploited to estimate the errors in the camera coordinate system.

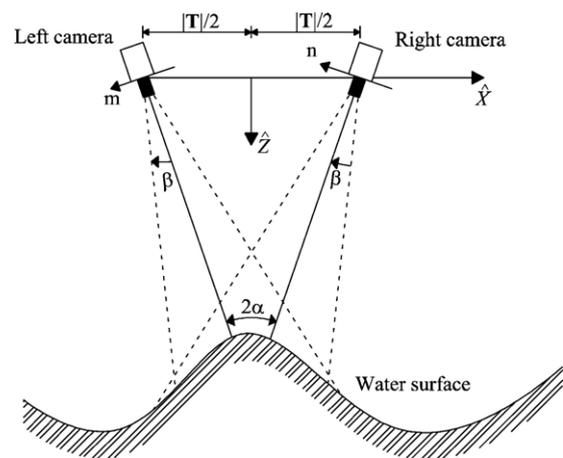


Fig. 5. 2-D projection of the acquisition set-up. The direction \hat{X} is defined by the line (baseline) connecting the two cameras' focal points. \hat{Z} is the direction orthogonal to the baseline and lying in the plane defined by the lines of sight of the cameras. The two cameras lines of sight are not parallel.

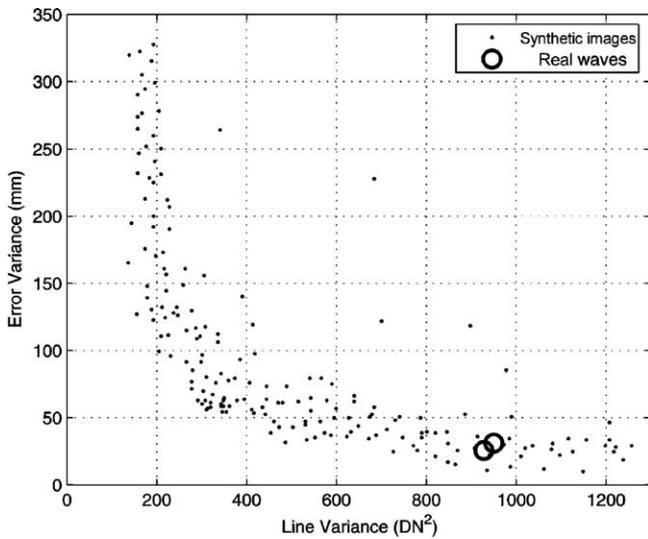


Fig. 6. Example of the error variance distribution experienced by synthetic images and images of real water surface waves. The horizontal axis of figure represents the epipolar line variance expressed in square Digital Number (DN), where the Digital number is the gray value representing the pixel brightness. In this example is used a 8 bit CCD, i.e. DN ranges between 0 and 255. For this example, the distribution of the error variance follows approximately an hyperbole.

The errors in the world coordinate system are calculated from the errors in the camera coordinate system applying the rotation matrix R which transforms the camera coordinate system to the world coordinate system (Section 2.3):

$$\begin{bmatrix} er_{x'} \\ er_{y'} \\ er_{z'} \end{bmatrix} = R \begin{bmatrix} er_x \\ er_y \\ er_z \end{bmatrix} \quad (24)$$

Another possible error in the stereo analysis is the matching error. There are two types of matching errors, false negative (or missed matches) and false positive (or erroneously made matches). Only the uncorrected matches are taken into consideration here. In order to estimate their effect in the 3-D stereo measurement, a series of virtual experiments are carried out. Two hundreds images reproducing natural and artificial subjects were created. Parallely some stereo-images of real water waves were taken into consi-

deration. The virtual stereo rig consists of a couple of 480×640 pixel cameras. The baseline, $|T|$, is set to 1.0 m and the mean distance between the baseline and the target, Z_0 , is 10.0 m. The angle of view of each camera is 43° , and the angle between the camera lines of sight is 4° . In this condition the maximum quantization error in the Z direction is approximately 0.065 m. Each virtual image is projected on a synthetic sinusoidal waveform (wavelength equal to 10 m, and wave amplitude equal to 0.5 m) as texture. The textured wave is then projected on the two camera planes forming the stereo rig using the projection matrix of the camera model. Then, images are corrupted with a Gaussian noise, with the characteristics of the signal-to-noise ratio of a real CCD. At the end, the stereo algorithm is applied. The water surface shape is reconstructed and the original synthetic wave is compared with its stereo reconstruction. For each point, the error in the water elevation measurement is computed as the difference between the measured elevation and the elevation of the original synthetic wave.

Without loss of generality, the position of the cameras has been chosen according to a simple rotation around the left camera Y axis and a translation parallel to the same axis. In real experiments the cameras will always be positioned close to this situation.

We note that an important parameter which describes the possibility to find reliable matches between images is the variance along the epipolar lines of images:

$$LV_i(Im) = E_{(j)}[(Im(i,j) - Im(i,j+1))^2] \quad (25)$$

where $Im(i,j)$ is the intensity value of the pixel at position (i,j) and $E_{(j)}$ is the expectation operator along the coordinate j of the epipolar line in the image coordinate system (Fig. 1).

Fig. 6 shows the relation between the variance of the errors (that include both the matching error and quantization error intrinsic in the triangulation process) and the epipolar line variance $LV_i(Im)$, relative to the virtual experiments described. The curve approximately follows an hyperbole. Assuming a normal distribution of errors (matching+quantization), the contribution of the matching error, for the virtual stereo rig used, in the definition of the maximum expected error ranges between 0.005 mm and 0.800 m, in dependence on LV_i parameter. We can note that the natural wave texture is such that the epipolar line variance is sufficiently high to consider this error generally negligible in the water elevation

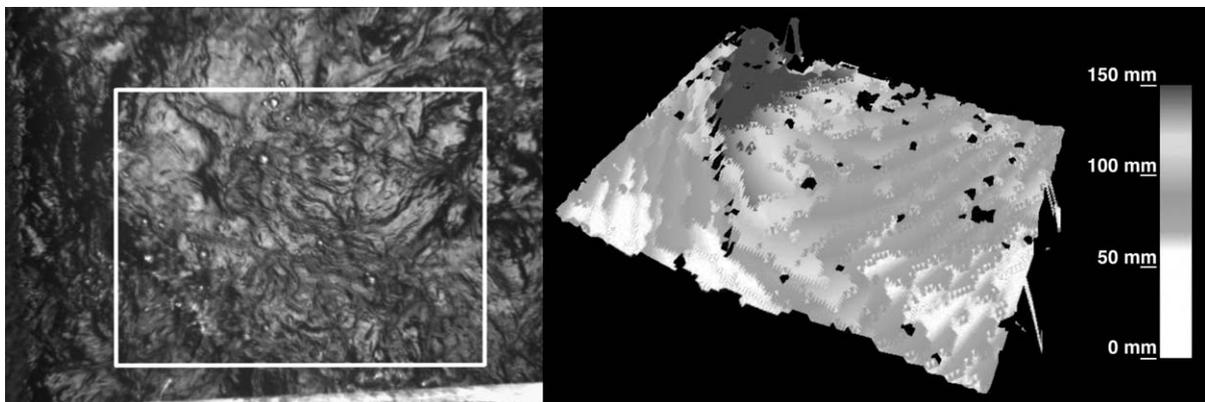


Fig. 7. Example of stereo-matching result. The right panel shows the left camera image, and the area selected for the stereo-matching (white rectangle). In the right panel the 3-D stereo-matched points are depicted. The scattered 3-D points are connected with triangles in order to relate the image color scale with the water elevation.

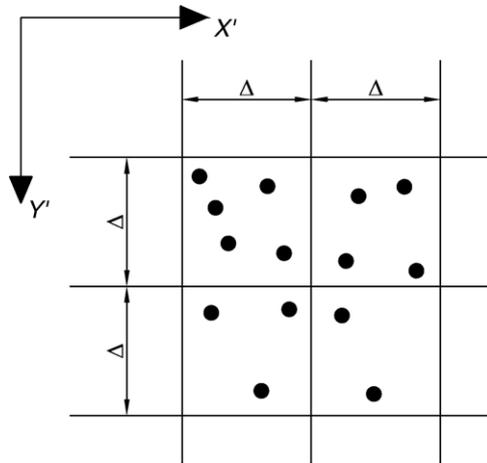


Fig. 8. t -plots extraction. The points in figure are the horizontal projection of the 3-D points. The sub-sampling of the 3-D scattered points on a regular grid is done by an average procedure. The numbers of points in each cell of the grid depends on the local availability of stereo-matched points. Δ is the (X', Y') ground spacing, equal for both X' and Y' world coordinates.

measurement with stereo techniques, under the condition of base to height ratio approximately 0.1.

In Fig. 7 an example of stereo analysis result is depicted. The distance from the focal points of the left camera (left image of Fig. 7) to the centre of the region of interest is approximately 1.5 m. The distance (baseline) between the cameras of the stereo rig is about 0.22 m, with a resulting base to distance ratio of about 0.15. A 0.95×0.75 m² region of interest (white rectangle in the same image) was selected for stereo-matching. The corresponding horizontal resolution in both the X' and Y' directions is 0.10 cm, and for a distance from the camera plane of 1.50 m the estimate vertical maximum error is 0.69 cm. About 90% of the points are stereo-matched. The right image of Fig. 7 shows the result of a meshing procedure on the 3-D scattered points result of the stereo-matching. The color scale is related to the measured water elevation. This image shows some aspects of the stereo analysis. First of all, the water surface is stepped: this is the effect of the quantization. The black holes represent the non-matched areas because of the poor local variance in the original camera images. These points are discarded as they do not ensure a reliable correlation between the windows of the two images of the stereo rig. Finally, the spikes are the false positives i.e. where the location of corresponding points is incorrect. These spikes are the effect of the matching error.

3. Data analysis

The photographic and the photogrammetric procedures provide the water surface elevation as a function of horizontal coordinates. The time sequence of the horizontally scattered 3-D points allows a tool for the dynamical and statistical measurements of the waves.

The treatment of this type of data usually (Stilwell, 1969; Shemdin and Tran, 1992; Dankert et al., 2003) follows both the principle of the wavenumber spectrum, and the more complete description offered by the wavenumber–frequency spectrum. A different approach, called t -plot analysis, is, at the first stage,

herein chosen as a tool to analyze the 3-D points result of the stereo triangulation. This artifice, described in the next section, gives a simple and compact way to analyze the wave parameters, in order to verify their coherency both in the space–time and in the wavenumber–frequency domains. In this scenario, simultaneous measurements with traditional wave probes (e.g. ultrasonic wave gauge) allow one to verify the reliability of the stereoscopic method as wave-meter.

3.1. Extraction of the t -plots

The complete set of 3-D scattered points in time (results of the stereo triangulation) are spatially sub-sampled on a squared regular grid (Fig. 8) in the Cartesian system of X' and Y' world coordinates. The grid ground spacing, Δ , is chosen in accordance with the quantization error (Section 2.4). In particular, given the maximum value, in the camera fields of view, among $er_{X'}$ and $er_{Y'}$, the value of Δ is four times this maximum value, i.e. twice the maximum error gap.

Each position of the grid (cell) is "filled" with a significant value of surface elevation of 3-D points belonging to each horizontal cell, of known X' and Y' coordinates, and ground spacing Δ . The presence of possible mismatches and of quantization steps (Fig. 7) is avoided limiting the procedure to those points whose distance from the cell mean Z' value is less than twice the variance of the local elevations. In this sense, the t -plot extraction is a recursive procedure. The extraction is extended

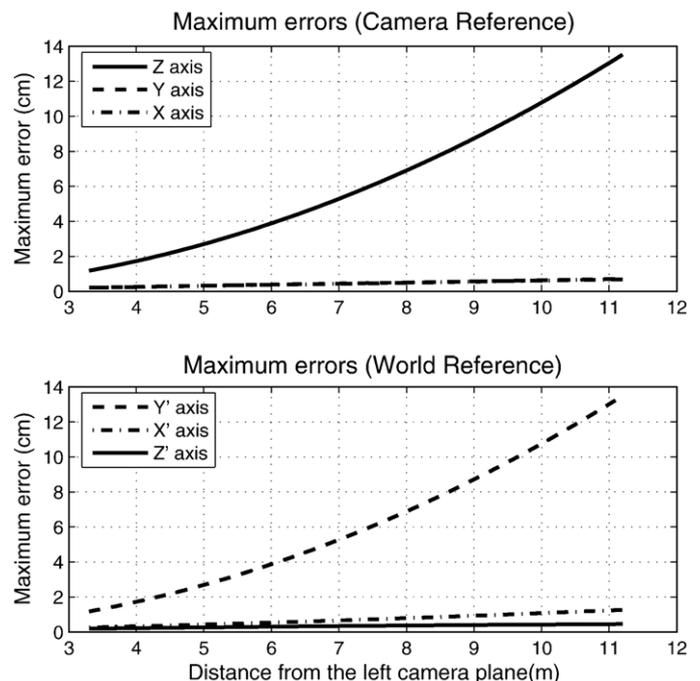


Fig. 9. Maximum quantization errors in camera (upper panel) and world (lower panel) coordinate systems. The maximum quantization error (proper to the Z axis in the camera coordinate system) increases quadratically with the distance between the Left camera focal point and the surface being framed. In the world coordinate system the error distribution depends on the rotation matrix (see Section 2.3) which transform the camera coordinate system to the world coordinate system. In particular, using high look angles the biggest error are expected in the horizontal plane, whereas the error in the vertical direction (Z' axis) is noticeably cut down.

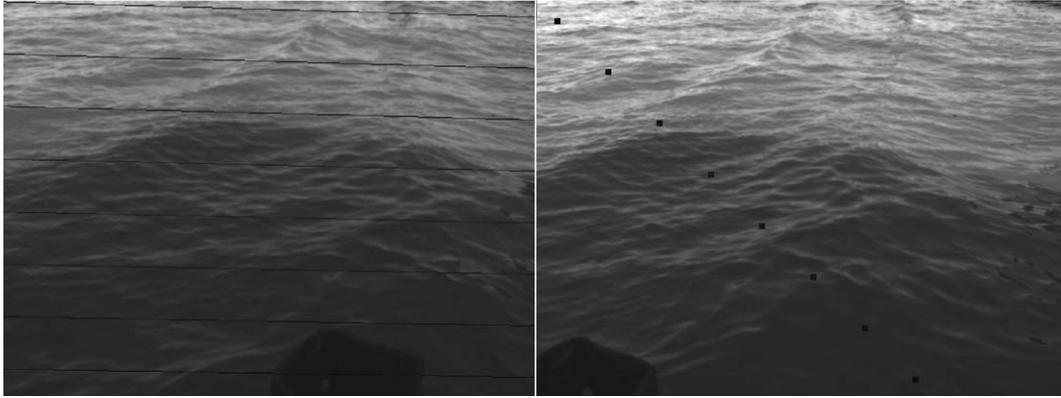


Fig. 10. The pixel coordinates in the Right image define the corresponding epipolar line in Left image. The search of the corresponding point is restricted to the epipolar line. This pair of matching image coordinates is used with the cameras' parameters to determine the 3-D coordinates (X, Y, Z) of the water surface at the selected point. To map the entire framed region, this process is repeated for each pixel.

to all the time-sequence of 3-D points. This method gives 1-D time series the length of which equals the number of images acquired, and the values of which are the local elevation on the sub-sampled grid. In such a way, every t -plot is considered as a single point probe, where the usual statistical and spectral analysis (Sections 3.2 and 3.3) available for 1-D data can be applied. Furthermore, the knowledge of the time series at different spatial locations is itself a way to extract directional (i.e. 2-D) information.

The following two section describe two techniques applied here to estimate phase speed, angle of propagation, and wavenumber of incident waves, by using a couple of adjacent t -plots. The two methods presented differ in their basic principles, and they aim to give two different ways to estimate the dynamics of the observed water surface waves.

3.2. Phase speed and angle

Measurements of the phase velocity of the waves with a pair of spatially separated sensors (the t -plots) depends on the estimate of the time the wave takes to pass between separated sensors (Lippmann and Holman, 1991). The time lag is here computed in the frequency domain.

In the frequency domain, mean time lag is manifested as a linear trend in the phase spectrum of the cross-spectral density function. In this work, the method described by Bendat and Piersol (2000) is used. Supposing that $s_1(t)$ and $s_2(t)$ are two zero mean value stationary random signals (time series), so that the attenuation factor between the two signals is 1. The two signals are representative of two adjacent positions. The cross-correlation function between $s_1(t)$ and $s_2(t)$ is given by

$$R_{12}(\tau) = \int_0^{\infty} G_{12}(f) e^{i2\pi f \tau} df \quad (26)$$

where $G_{12}(f)$ is the one-sided cross-spectral density function. The peak value of the estimated $R_{12}(\tau)$ is given by

$$\hat{R}_{12}(\tau)_{peak} = \hat{R}_{12}(\tau_0) = \int_0^{\infty} \hat{G}_{12}(f) e^{i2\pi f \tau_0} df \quad (27)$$

At the peak location the cross-correlation function must be maximum

$$\frac{\partial \hat{R}_{12}(\tau)}{\partial \tau} (\tau = \tau_0) = 0 = - \int_0^{\infty} 2\pi f |\hat{G}_{12}(f)| \sin[2\pi f \tau_0 - \hat{\theta}(f)] df \quad (28)$$

Since

$$\hat{\theta}(f) \approx 2\pi f \tau_0 \quad (29)$$

resolving for τ_0 the result is yielded by

$$\tau_0 \approx \frac{\int_0^{N_y} (2\pi f) |\hat{G}_{12}(f)| \hat{\theta}_{12}(f) df}{\int_0^{N_y} (2\pi f)^2 |\hat{G}_{12}(f)| df} \quad (30)$$

The range of integration of the frequency, f , is limited by the frequency defined by the Nyquist limit (Bendat and Piersol, 2000). It is important to note that in this procedure the individual frequencies are weighted in order to give higher importance to those frequencies with higher spectral coherence.

Introducing a Cartesian coordinate system (x, y) which axes lie on the horizontal plane, wave phase speed, c , and wave angle, α , are computed defining the vectorial quantity slowness (i.e. the inverse of the phase velocity), $s = [s_x, s_y]^T$, given by the ratio between the time lags (τ_x, τ_y) and the distances $(\Delta x, \Delta y)$ between the orthogonal sensors:

$$s_x = \frac{\tau_x}{\Delta x}$$

$$s_y = \frac{\tau_y}{\Delta y} \quad (31)$$

The magnitude of the phase speed and the wave angle are then defined by the orthogonal components of s :

$$|c|^2 = \frac{1}{s_x^2 + s_y^2} \quad (32)$$

$$\alpha = \arctan \frac{s_x}{s_y} \quad (33)$$

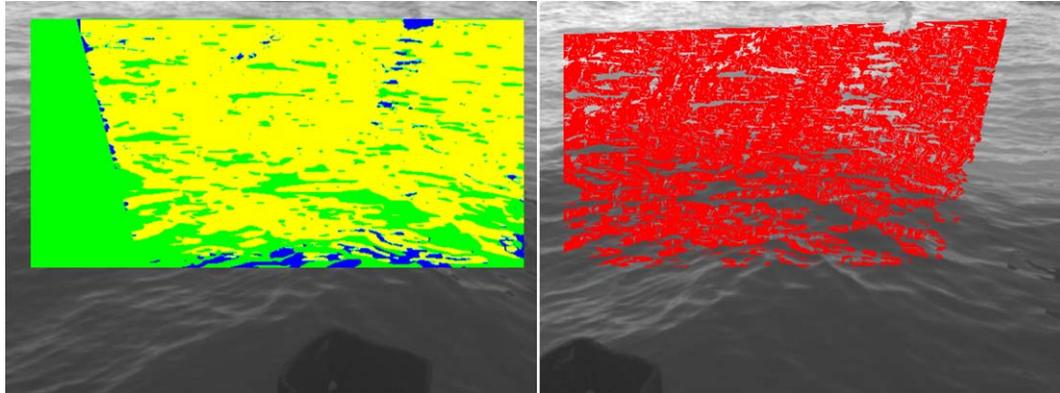


Fig. 11. Corresponding points. The green rectangle in the Left image signifies the selected area for stereo matching. All the pixels in this area are stereo-analyzed. The first examination is made on the variance of the area centered in each pixel. The area dimension is fixed in 11×11 pixels. Only if the variance is higher than 0.35 times the whole image variance, the epipolar line on the Right image plane is defined. The blue points in the Left image are the pixels not stereo-matched because of the poor cross correlation coefficient between areas centered in the corresponding pixel in Left and Right images. The threshold in the cross-correlation coefficient is set to 0.85. Finally, only the yellow marked pixels in the Left image are stereo-matched. To these pixels the search algorithm found the red marked pixels in Right image.

The resolution of c is done by

$$\frac{\Delta c}{c} = \frac{\Delta L}{L} - \frac{\Delta T}{T} \quad (34)$$

where L is distance separating the sensors, ΔL is the error in L (equal to the pixel resolution of the sub-sampled grid), T is the time lag, and ΔT is the error in the time lag estimate (equal to the reciprocal of the frequency of sampling).

3.3. Cross-spectral methods for the estimation of the wavenumbers

The wavenumbers are determined from cross-spectral analysis. A generic Cartesian coordinate system (x,y) lies on the horizontal plane is used. The determination of the wavenumber vector $\mathbf{k} = [k_x, k_y]^T$ from separated time series is

based on the fact that the phase shift $\Delta\phi$ between two contemporary signals is given by

$$\Delta\phi = \mathbf{k} \cdot \Delta \quad (35)$$

where Δ is the distance vector separating the two time series. The phase shift can be estimated by computing the cross-spectrum of the two time series. If $s_1(t)$ and $s_2(t)$ are two time series of adjacent positions in x direction, the estimation of the one-sided cross-spectrum, $\hat{G}_{12}(\omega)$, between the two signals separated by the distance Δx , is given by

$$\hat{G}_{\Delta x}(\omega) = \hat{G}_{11}(\omega) \hat{G}_{22}^*(\omega) \quad (36)$$

where $\hat{G}_{11}(\omega)$ and $\hat{G}_{22}(\omega)$ are the two one-sided auto-spectra of the original signals, ω is the angular frequency, and the index Δx refers to the distance between the two time series.

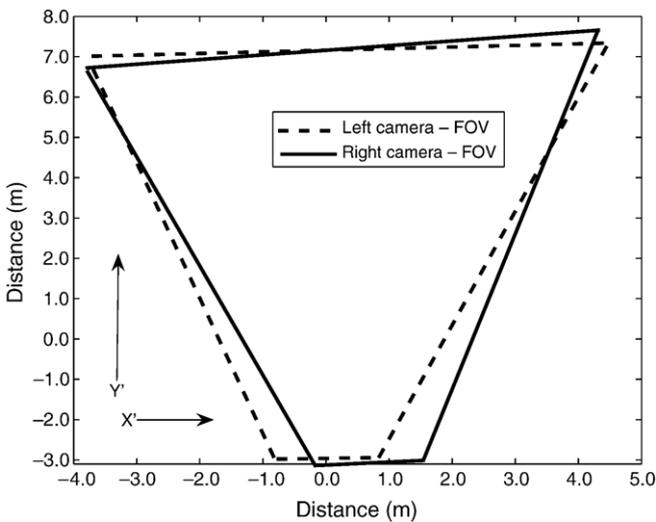


Fig. 12. Fields Of View of the search area of the stereo photographic setup at Venice experiment. For Left and Right cameras the trapezium bounds the respective water surface area of coverage. Given these 640×480 pixel images as viewed from cameras with a 41° angle lens with an oblique camera view the horizontal width of the field of view ranges between 1.5 m and 7.1 m, with a horizontal length of approximately 12 m.

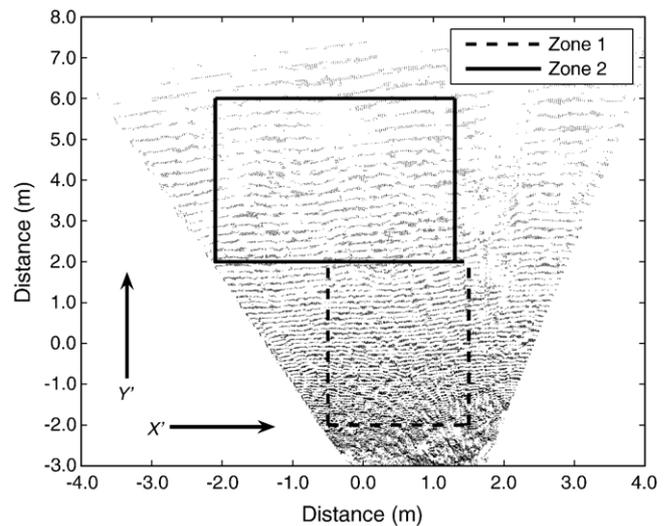


Fig. 13. Zones of analysis in the world reference frame. The black points are the horizontal projection of the matched points, in the Right camera coordinate system. The two zones analyzed differ for the distance from the baseline, and for the dimension of the ground spacing of the grid (see Table 2).

Considering two time series in the y direction, an analogous result is obtained:

$$\hat{G}_{\Delta y}(\omega) = \hat{G}_{11}(\omega) \hat{G}_{22}^*(\omega) \quad (37)$$

Supposing that the $s_1(t)$ time series is a phase shifted copy of the $s_2(t)$ time series, the cross spectrum along the x direction can be rewritten as

$$\hat{G}_{\Delta x}(\omega) = |\hat{G}_{11}(\omega)|^2 e^{i\Delta\phi} \quad (38)$$

where the phase shift is $\Delta\phi = k_x \Delta x$. Considering two orthogonal directions (x, y), the two components of the phase shift are then given as the arc tangent of the ratio of the imaginary (quad-spectrum) and real (co-spectrum) parts of the complex cross spectra. The wavenumbers, function of the frequency, are directly calculated as the ratio between the directional phase shift and the distance separating the t -plots:

$$k_x(\omega) = \frac{\Delta\phi_x(\omega)}{\Delta x}$$

$$k_y(\omega) = \frac{\Delta\phi_y(\omega)}{\Delta y} \quad (39)$$

The wave angle, α , and the phase speed, c , can be calculated using the two orthogonal components of \mathbf{k}

$$\alpha(\omega) = \arctan \frac{k_x(\omega)}{k_y(\omega)} \quad (40)$$

$$|c(\omega)| = \sqrt{c_x^2(\omega) + c_y^2(\omega)} = \omega \sqrt{\frac{\Delta x^2}{\Delta\phi_x^2(\omega)} + \frac{\Delta y^2}{\Delta\phi_y^2(\omega)}} \quad (41)$$

Then, the wavenumbers relative to the peak of the frequency response are selected, and the relative values α and c are calculated.

The resolution of c for the cross-spectral method coincides with the resolution of c given in the previous section.

4. Experiments

4.1. Venice experiment

The data were acquired on September 9 2004 at the delle Sacche channel, parallel and contiguous to the northern bank of the city of Venice (Italy), in the Venice lagoon. This channel is approximately 6 m wide and 2 m deep. The channel was furrowed by surface waves caused by little private boats (length less than 6 m, and velocity approximately 1.5 m/s). In these conditions the image data were acquired. No wind was present.

Table 2
Characteristics of the two zones analyzed

| | Z_0 | Grid points (X', Y') | Covered area |
|--------|------------|--------------------------|--------------------------|
| Zone 1 | 3.3–7.2 m | 67 × 134 | 2.0 × 4.0 m ² |
| Zone 2 | 7.1–11.2 m | 49 × 58 | 3.4 × 4.0 m ² |

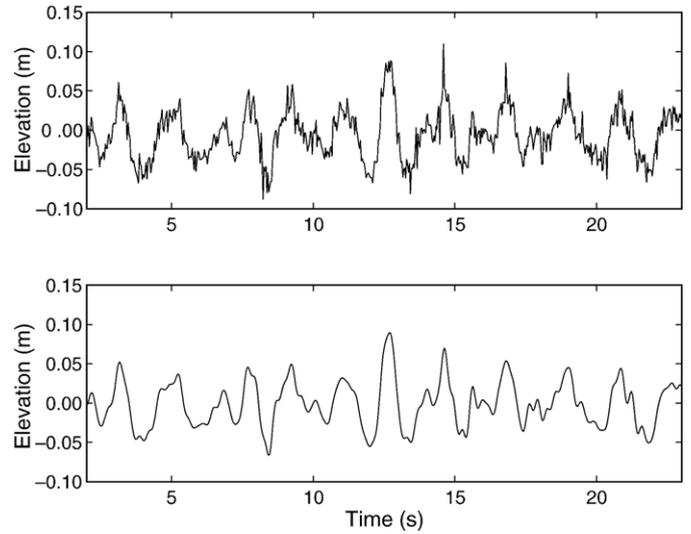


Fig. 14. Example of t -plot extracted from the time-stack of 3-D scattered points in position $(X', Y') = (0, 0)$ of Fig. 13, before (upper panel), and after the operation of filtering and detrending (lower panel).

In this test no in-situ sensors were present in the site to compare the stereo measurements.

Video data were collected using two synchronized progressive-scan CCD cameras (JAI CV-A11) with 6.3-mm lenses. Each CCD-sensor had a radiometric resolution of 8 bit greyscale (monochrome) and a geometric resolution of $7.4 \times 7.4 \mu\text{m}^2$ per pixel. The signal-to-noise ratio was greater than 56 dB. The sensor size was 640×480 pixels, and the frame rate, in continuous mode, was 30 frames per second. Each camera was connected to a separate image acquisition device (National Instruments PCI-1409 frame

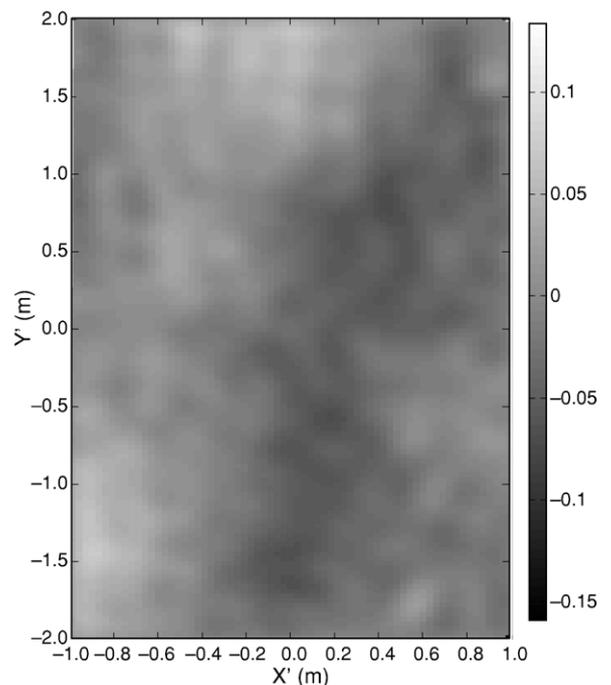


Fig. 15. 3-D elevation map of water surface elevation determined from gridded stereo 3-D data in zone 1, September 9, 2004, Venice. Intensity in the frame represents the water surface elevation in m.

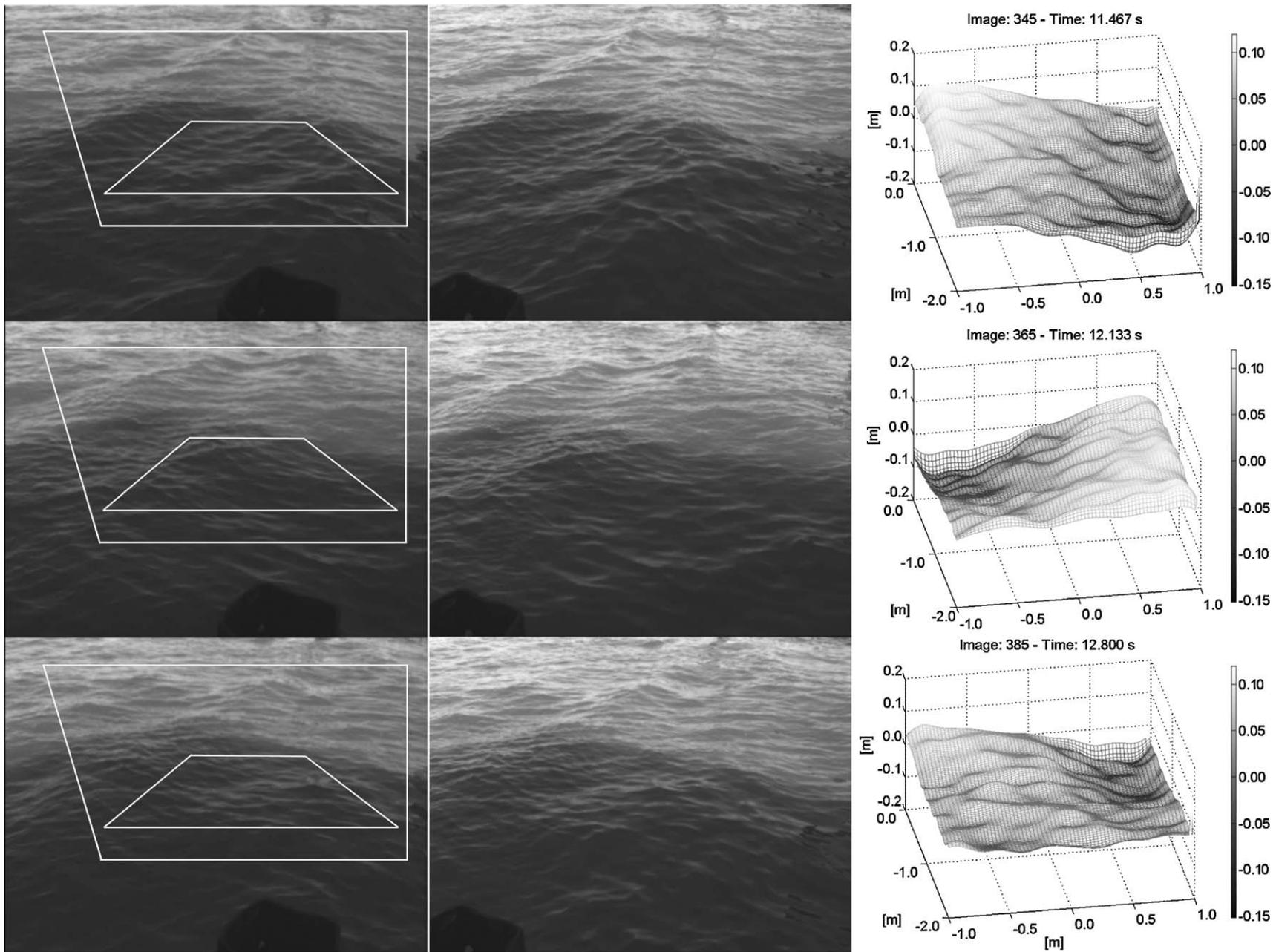


Fig. 16. Sequence of water surface 3-D elevation maps (right column) from sub-sampled stereo-pair frames 345, 365, and 385. The left column shows the corresponding sequence of 2-D images acquired from the left camera. The sequence in the middle column is the equivalent for the right camera. On the left camera image the stereo-matched area (white outer region) and zone 1 (white inner region) are superimposed.

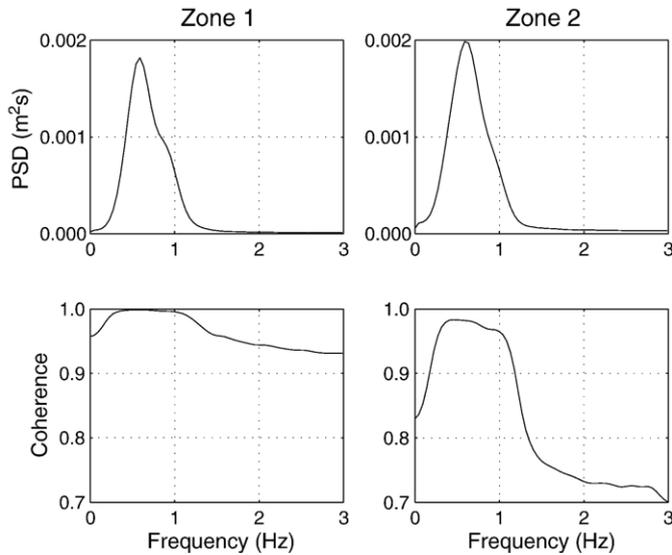


Fig. 17. Power Spectral Density (PSD) and Coherence spectra obtained by ensemble averaging. The coherence spectrum is calculated for t -plots separated by 30 mm (zone 1, left panel) and 70 mm (zone 2, right panel) in Y' direction. All spectra are computed for 16 degrees of freedom.

grabber) to convert the analogical signal coming from the cameras to a digital form. The acquisitions boards also controlled the synchronicity of each stereo-pair. The time delay between the stereo-images was approximately 2 ms, less than the maximum permissible time interval (Holthuijsen, 1983). In this first version of the acquisition system, during the acquisition, all the images were stored into the RAM memory and, when it was full, the acquisition was ended, and the images were saved on the Hard Disk. The system allowed for a maximum observation period of approximately 45 s, the bottle-necks were the camera number of frames per second (30) and the current RAM memory. The whole acquisition system was managed in LabView language.

In the stereo system used here, the shutter speed was generally determined in accordance with two principles. The first goal was to avoid blurred images, due to the motion into the framed scene. This was obtained by increasing the shutter speed. Nevertheless, to have a high epipolar line variance, the image histogram (i.e. the distribution of the grey values in the image) had to be centred at half dynamic range of the camera (e.g. with an 8 bit camera, the goal is to set the image mean grey value at 128). In order to obtain these conditions the shutter speed in the experiment herein described was 1/400 s. The depth of field was kept within the range of 1.8 m (hyperfocal distance) to infinity to insure continuous focus along the wave profile. The lens geometrical aberration was reduced via software after the calibration (Bouquet, 2004).

The geometry of the point of view is described by the rigid motion (Eq. (20)), which transforms the camera coordinate system to the world coordinate system. In order to set correct water wave elevations, the Z axis of the world coordinate system must coincide with the gravity direction. The rigid motion of this experiment was calculated using the method described in Section 2.3. The derived three rotation angles about the camera axes were $\phi = -5^\circ$, $\sigma = -3^\circ$, $\tau = 67^\circ$. The angle between the camera lines of

sight (2α , see Fig. 5) was set in order to avoid the correspondence problem at specular surfaces (Jähne, 1993), i.e. a feature seen in the first image may not appear on the other image, resulting in an angle of approximately 4° .

The distance, Z_0 (Fig. 2), between the left camera (used as reference coordinate system) focal points and the framed region ranged between 3.3 m and 11.2 m. The baseline, $|T|$ (Fig. 2), between the camera principal points was 0.58 m. This photographic set-up results in a base to height ratio, $|T|/Z_0$, ranging between 0.18 and 0.05, in accordance with the value suggested in Section 2.4.

Fig. 9 shows the maximum quantization error; clearly, the error increases with the distance from the camera. On the upper panel the error is relative to the camera coordinate system; the largest error is along Z axis, that is the line of sight of the camera. On the lower panel the errors are given in the world coordinate system, as defined by Eq. (24). Since the angle between the Z axis of the camera coordinate system and the Z' axis of the world coordinate system was large ($\sim 67^\circ$), the largest errors in the world reference followed the Y' axis while the error on the Z' axis (that is the direction of the wave elevations) was smaller.

Fig. 10 shows the epipolar constraint: points on the left camera image correspond to lines on the right camera image. The quality of the calibration can be appreciated: each epipolar line in the right image passes exactly through the expected corresponding point (selected in the left image), the exact position of which is the first goal of the stereo-matching algorithm.

In Fig. 11 the area covered by the stereo matching is superimposed to a stereo-pair: the green rectangle in the right image bounds the search area, corresponding to the red bounded area in the left image, where the conjugated matched points are highlighted.

In the right image, the green points mark those points not investigated, since their local variance (11×11 pixel window) was too small (smaller than 0.35 times the whole image variance).

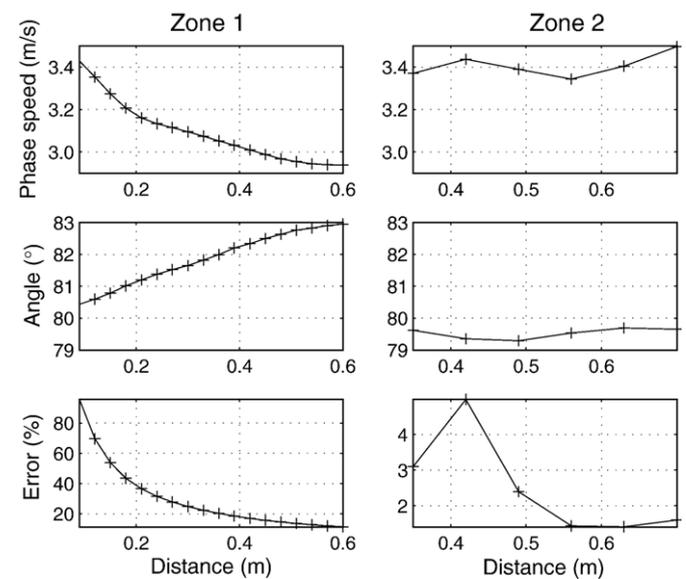


Fig. 18. Mean phase speed (upper panel), and mean wave angle (middle panel) in zone 1 (left column) and zone 2 (right column). In the lower panel the errors in the phase speed calculation due to the resolution of c are plotted. All the measurements are calculated for different distances separating the t -plots.

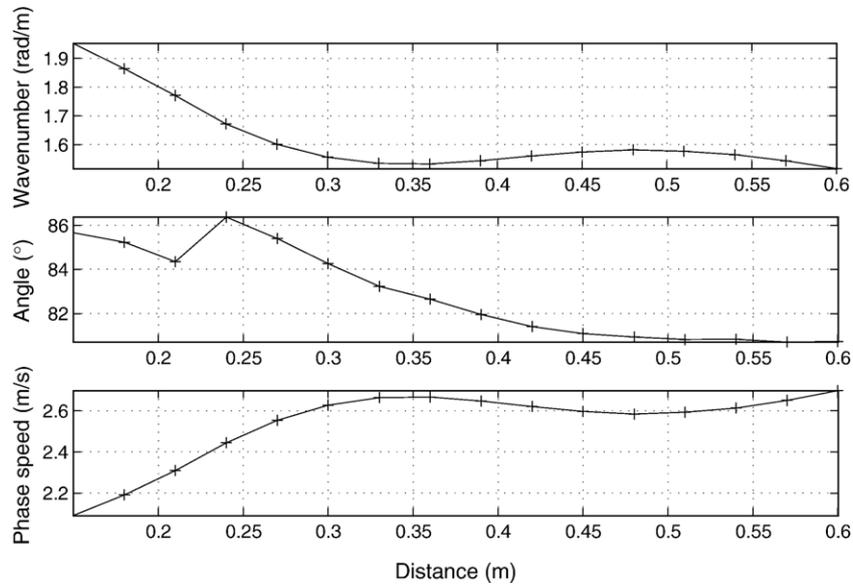


Fig. 19. Wavenumber magnitude, wave angle, and phase speed in zone 1, computed by using the cross-spectral analysis. All the measurements are calculated for different distances separating the *t*-plots.

Instead, the blue points are the not-matched points because the second threshold (set to 0.75) on the cross-correlation coefficient was not passed. From this, it follows that the matched points percentage was about 75% of the total pixels analyzed. The computational time for a time sequence of 24 s (corresponding to 720 stereo-pair) of a 500×400 pixel image area was approximately 220 s, exploiting both the pyramidal search and the time evolution implemented in the stereo-matching algorithm.

The Field Of View (FOV) of the search area of the photographic setup is shown in Fig. 12.

In terms of *t*-plots, two different zones (Fig. 13) were considered. In these zones, *t*-plots on a regular grid were extracted. These two zones differed by the mean distance from the left camera plane, therefore they were characterized by different statistics of quantization errors. Moreover, for these two zones a different nominal ground spacing between *t*-plots was chosen, in order to

follow the increase of the quantization errors as the distance from the cameras increases. Because of that, the ground spacing Δ equalled 30 mm in the zone 1, and 70 mm in the zone 2. The geometric characteristics of the two zones are summarized in Table 2, where Z_0 is the distance from the left camera plane, and *Gridpoints* is the dimension of the *t*-plot grid. Each *t*-plot was frequency domain filtered (Parks and Burrus, 1987) by using a lowpass filter with cutoff at 4 Hz, and linear detrended with a least-squares fit, in order to reduce the high-frequency noise and the long period effects (Fig. 14). Clearly, the filtered *t*-plot is less sensitive to the quantization error.

Considering at the same time the water elevation for the entire the *t*-plot set, a 3-D water elevation map was created. Such a map is shown in Fig. 15, where the horizontal water surface area is 4.0 m by 2.0 m. The complete time distribution of *t*-plot water elevations provided a time sequence of 3-D elevation maps, sub-

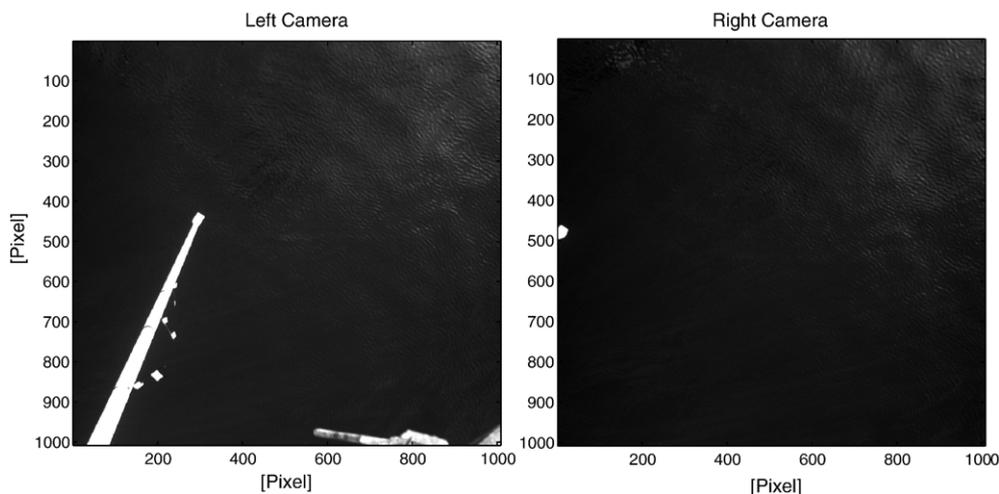


Fig. 20. Images from Left (left panel) and Right (right panel) cameras, December 14, 2005, San Diego, CA. On the Left camera can be seen the beam carrying the ultrasonic gauge used to validate the stereo measurements.

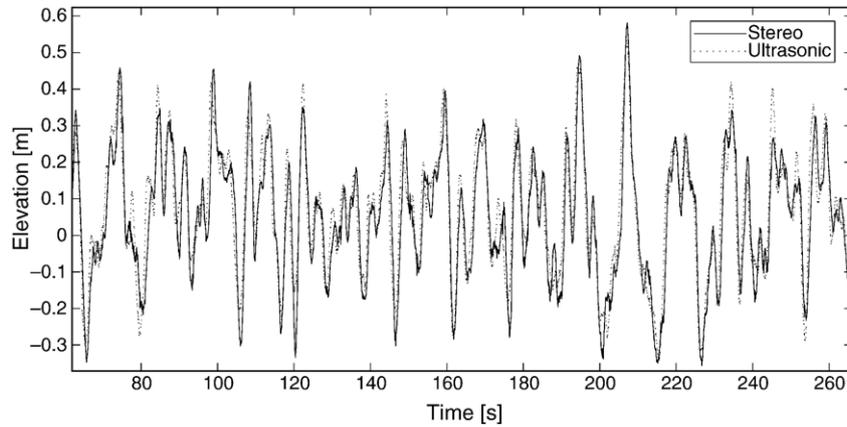


Fig. 21. Comparison in the time domain of image matching results (continuous line) and gauge data (dashed line). The time subset (of duration equals to 195 s) of the complete 1200 s record shows the agreement between the stereo and ultrasonic gauge sea elevations.

sampled on a regular grid. Fig. 16 shows the left and right camera images in three different instants, and their 3-D elevation maps, characteristic of zone 1 in Fig. 13.

In order to compute the phase relationship of the progressive incident waves, cross-spectra between spatially separated t -plots was computed. The cross-spectra and the auto-spectra were estimated by ensemble averaging (Welch, 1967): each time-series was treated with 16 degrees of freedom and 50% of overlapping. The side-lobe leakage, due to the discontinuities at the boundaries of the data, was suppressed with a full cosine tapering window (Bendat and Piersol, 2000). Fig. 17 shows the power spectrum (upper panel), and the coherence spectrum (lower panel) relative to two equally spaced t -plots in Y' direction. The t -plots were separated by a distance of 30 mm (left column, zone 1), and 70 mm (right column, zone 2).

Estimates of the wave propagation mean travel times were made using the spectral method (Section 3.2). Phase speed magnitudes and wave angles were computed from orthogonal slowness components derived from pairs of adjacent t -plots in X' and Y' perpendicular direction, choosing $\Delta x = \Delta y$ in Eq. (31). A sensitivity analysis was performed in order to verify the depen-

dence of the calculated phase speeds and propagation angles on the distance separating the t -plots used for the calculation.

Therefore, the phase speeds and the propagation angles were calculated for different space lag. The results for different space lags in zones 1 (left column) and 2 (right column) are shown in Fig. 18. In the same figure, the lower panel shows the errors in the phase speed calculation, for different space lags. The measurements shown in Fig. 18 are made for distances between t -plots when the spectral coherence is greater than 0.95. In the figure the dependence of the measurements on the space lag is highlighted. In particular, a convergence value is visible for distances such that the resolution in c measurements is less than 10%. For c this convergence value is 2.9 m/s. The differences between zones 1 and 2 are ascribed to the different water depth, since zone 1 is landward and zone 2 is lagoonward.

The wavenumbers were calculated here following the cross-spectral analysis (Section 3.3): the orthogonal components of the local wavenumber were calculated from the phase shift between adjacent t -plots along X' and Y' directions. This method requires the selection of the incident wave frequency peak in the power spectrum (Fig. 17). This allows one to calculate the wavenumbers

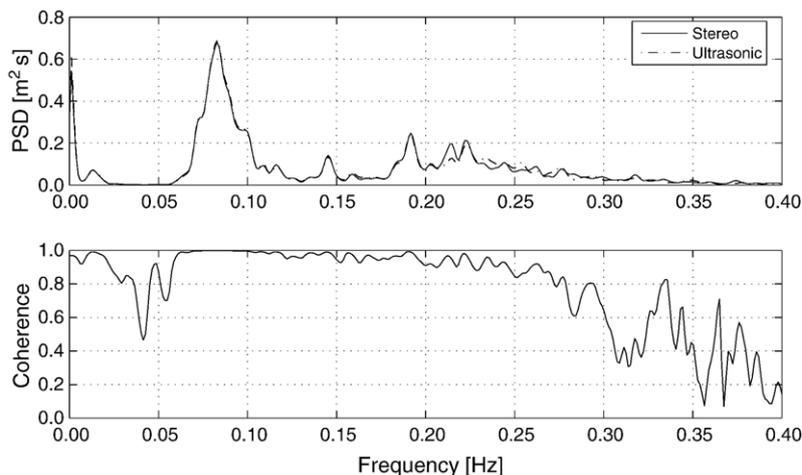


Fig. 22. Comparison in the frequency domain of image matching results (continuous line) and gauge data (dash-dotted line). In the frequency range of the incident waves the spectral coherence is approximately 1.

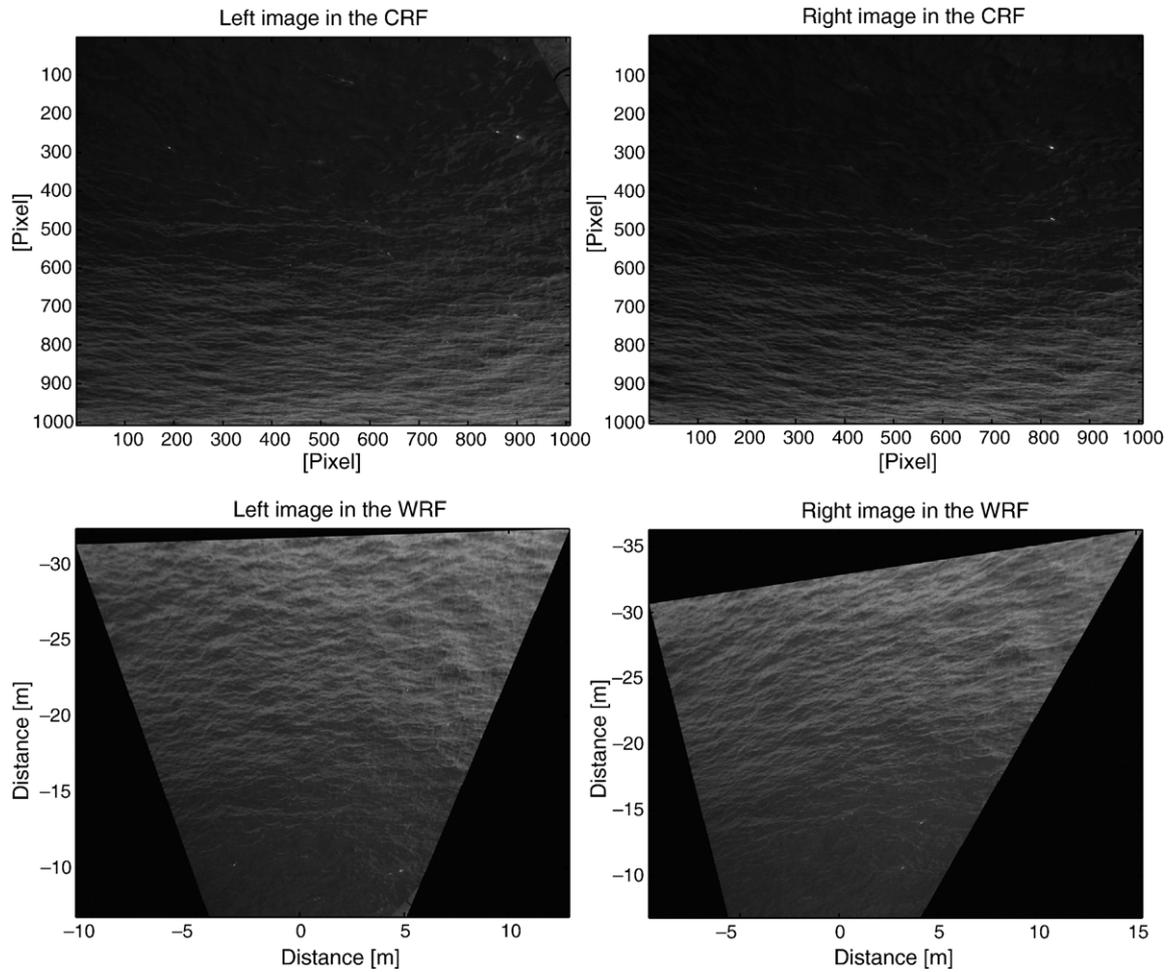


Fig. 23. Images from Left (left column) and Right (right column) cameras, in the camera reference system (CRF) and in the world reference system (WRF), December 14, 2005, San Diego, CA. The rotation of the baseline allows to increase the field of view of the cameras. Since the original image was sampled at discrete pixel locations, some interpolation of image intensities is required to produce a smooth rectified image in the world reference frame.

(Fig. 19) following Eq. (39), with $\Delta x = \Delta y$. Consequently, an estimation of wave propagation angle and the phase speed is made. All these measurements were calculated for different distances between t -plots relative to zone 1 (Fig. 13). The resolution in c calculation is not shown since it coincides with the resolution given in the previous paragraph.

Different phase speeds were obtained by using the two methods described. These discrepancies are intrinsic to two methods described in Sections 3.3 and 3.2. In fact, while the first method is based on the estimation of a mean time lag between signals, the second method identifies the behaviour of the most energetic wave. Moreover, the differences belong to the errors expected in the phase speed calculation.

4.2. San Diego experiment

The system used in the second experiment consisted of a couple of synchronized Pulnix 1040 video cameras (1008×1008 pixels) with 13-mm lenses mounted on a boom atop an 11-m height permanent pier at the Scripps Institution of Oceanography in San Diego, California. All the images were collected on

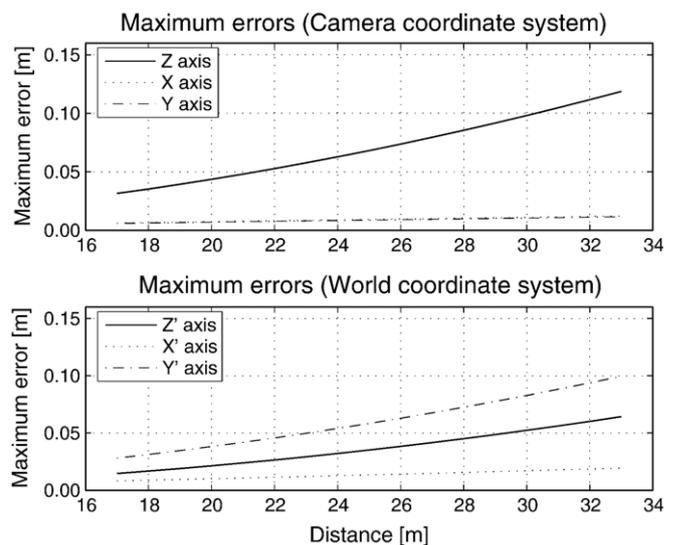


Fig. 24. Expected maximum quantization errors, function of the distance of the water surface from the Left camera focal point. The upper panel refers to the camera coordinate system, where the maximum error lies on the Z axis, coincident with the Left camera line of sight. In the world coordinate system, where the Z' axis is parallel to the vertical direction, the maximum theoretical error in the water elevation measurement is less than 7 cm.

December 14, 2005. The images' frequency of acquisition was 20 Hz. The maximum duration of the video image was 20 min, resulting in a sequence of 24,000 images for each camera. Video signals from the camera were digitized into 10-bit gray-scale images using a couple of frame grabbers housed in a personal computer. The 3-D points result of the stereo procedure were compared with data available from traditional wave gauges (ultrasonic gauge). The ultrasonic gauge was mounted on the same boom. Two series of tests were performed. In the site, a gentle sea breeze blown on swell waves propagating perpendicularly to the shore line, i.e. eastward.

This experiment aims at describing the resolution of the stereo rig, comparing the wave elevation stereo data and the measurements from the ultrasonic wave gauge. Moreover, this experiment focuses on the stereo system as 2-D wave meter.

In the first test, the distance between the cameras was fixed at 1.88 m. The cameras were positioned in order to have the lines of sight nearly vertical (Fig. 20), resulting in a field of view approximately square ($\approx 7.5 \times 8.0 \text{ m}^2$). It resulted in a maximum quantization error in the Z' direction of 0.023 m, nearly uniform in the whole image. The horizontal resolutions were about 0.004 m (Y' direction) and 0.004 m (Y' direction). The shutter speed was 1/250 s, and the lens aperture size was f-4.

To compare the water elevation calculated via stereo-matching with the gauge data, a t -plot grid (ground spacing equals to 0.10 m) was considered. Fig. 21 compares the image matching result and the gauge data over a period of 195 s (time subset of the entire sample of 1200 s). The horizontal axis of the same figure represents the time in seconds. The vertical axis shows the water surface elevation in meters. As can be seen, the elevation values determined by image matching correspond to the gauge data. However, the image matching results show some high-frequency noise because of the quantization error. The root-mean-square error of the differences between the two measurements was 6.7 cm over the 1200-second period, with a mean absolute difference error of 4.7 cm, and a mean difference error of 0.9 cm. A comparison in the frequency domain shows an excellent agreement in terms of Power Spectral Density, and spectral Coherence, as depicted in Fig. 22. As shown, the stereo results compared favourably with the expectation (the time series from the ultrasonic gauge), as suggested by the small quantization error.

A phase speed of 5.5 m/s was measured by using adjacent t -plots (Section 3.2) with an error of 0.4 m/s. This phase speed was a weighted value over all the frequencies. Limiting the range of frequencies where the cross-spectrum was maximum

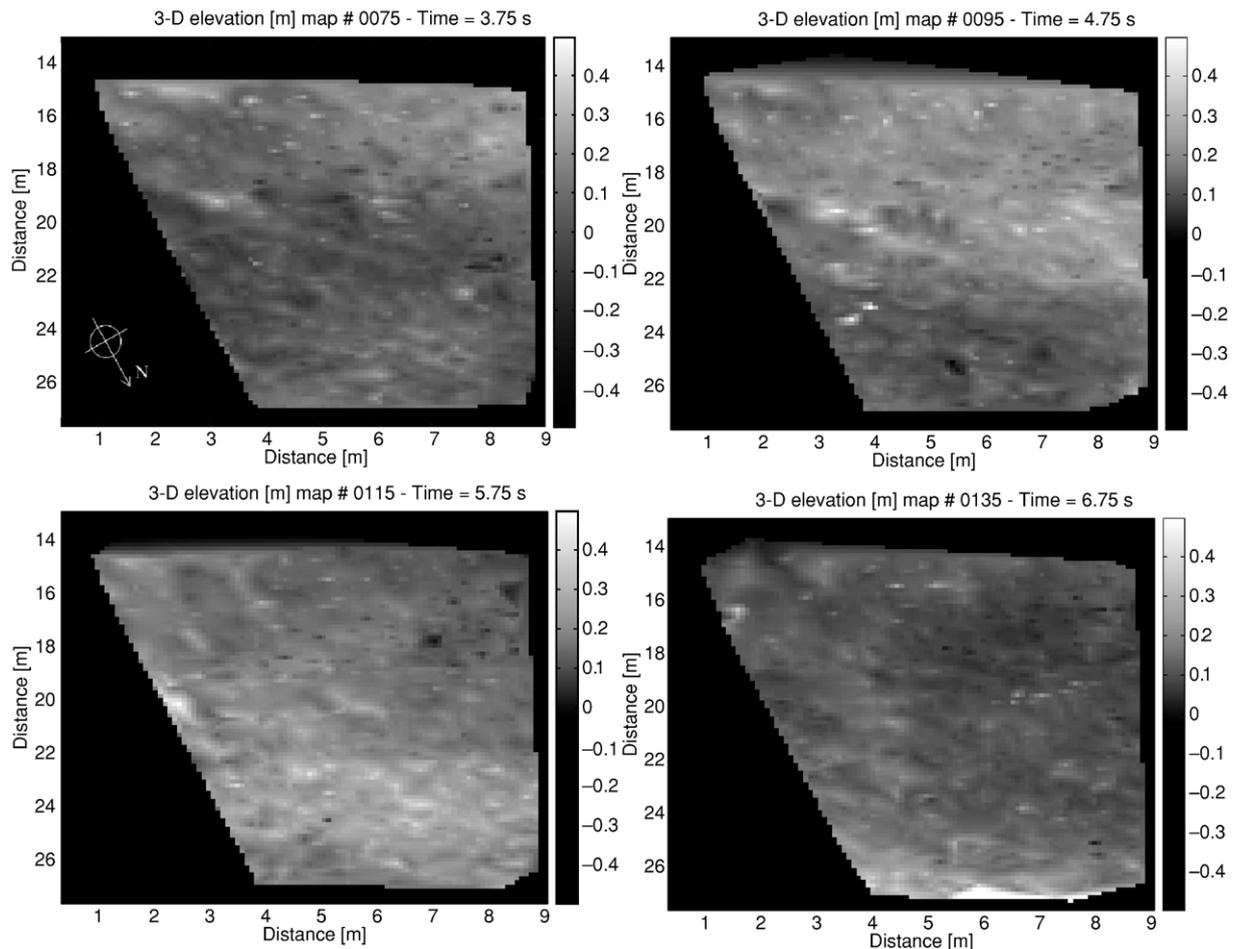


Fig. 25. 3-D water elevation maps. The image size is about $5\text{--}9 \text{ m} \times 13 \text{ m}$, with a pixel size of $9 \text{ cm} \times 14 \text{ cm}$. The images show eastward travelling nearshore ocean waves of different wavelengths with a dominant wave of approximately 4.5 s. The grey scale in the images is related to the water surface elevation in meters.

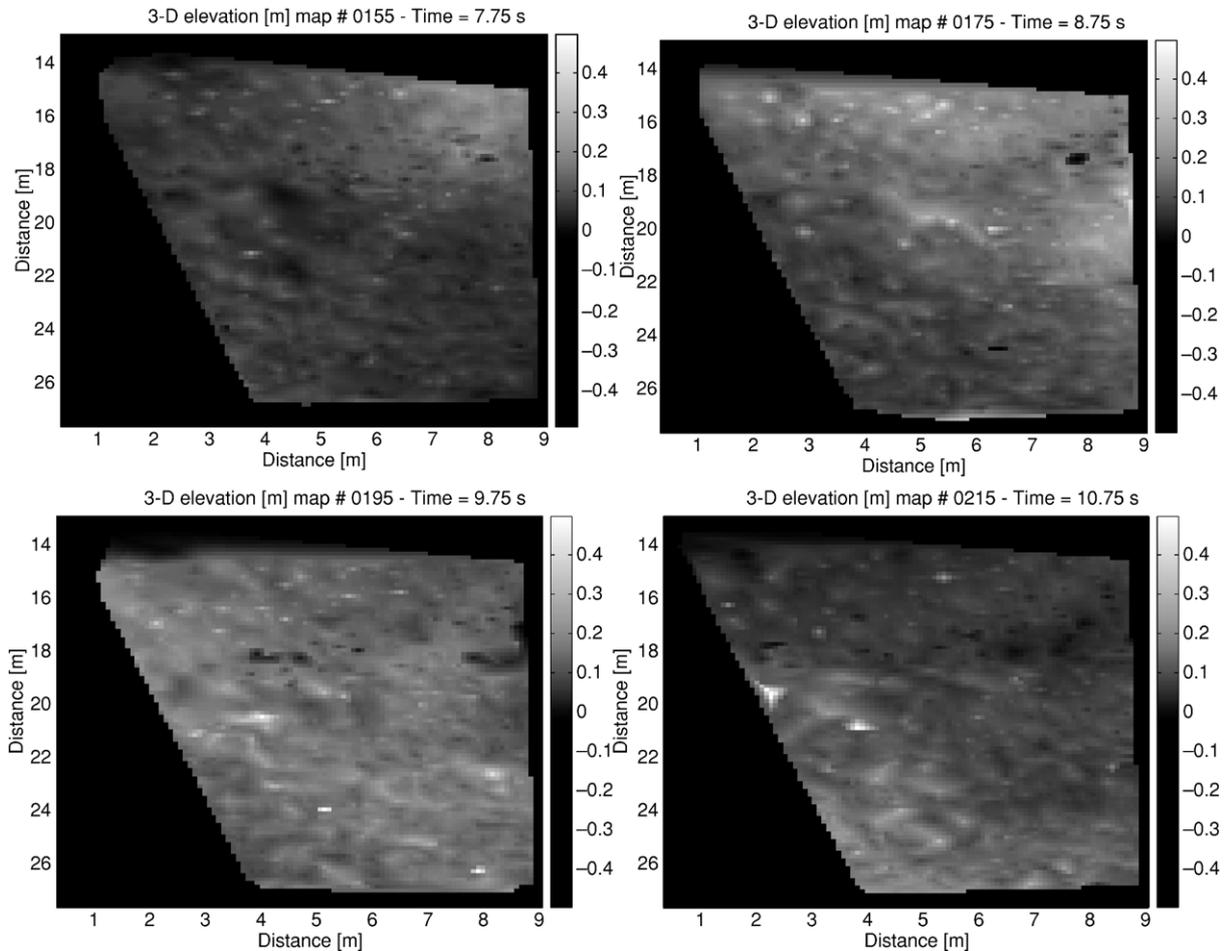


Fig. 25 (continued).

(spectral width equals to 0.04 Hz), an estimation of the phase speed of the dominant wave is given. In such a case, the wave celerity was 7.4 m/s (± 0.6 m/s), with a direction of propagation of 94° (measured clockwise), with respect to the X' direction in the world coordinate system. This phase speed is consistent with the celerity calculated using a cross-spectral technique (Section 3.3). In fact, this formulation gives a dominant phase speed of 7.8 m/s, and a dominant wavenumber of 0.07 rad/m, corresponding to a wavelength of about 90 m. Considering a depth of about 8–9 m, a theoretical linear phase speed (Svendsen and Jonsson, 1980) of about 8.5 m/s results. The discrepancy could be imputable to the limits of the linear theory, not strictly valid since the wave height is the same order of magnitude of the depth. A statistical analysis in the time domain gave a significant wave height, from a zero-crossing point of view (IAHR, 1986), H_s , of 0.65 m both for the ultrasonic gauge, and the stereo-matching results. The maximum wave height, H_{\max} , measured 1.15 m, for either sensor.

The second test of San Diego experiment used the same stereo rig used in the first test. Nevertheless, the geometry of the system was changed in order to increase the field of view of the cameras (Fig. 23).

For this reason, the cameras were rotated in order to raise the angle between the camera lines of sight and the vertical direction.

As the mean distance between the cameras and the framed water surface increased, to preserve the resolution of the measurements, the distance (baseline) between the cameras was increased till 3.04 m. Estimation of the rotation matrix R (Section 2.3) by using the 3-D calculated points gave a maximum quantization error in the vertical direction of 6.2 cm, and a maximum horizontal resolution of 1.8 cm and 9.5 cm, on the X' and Y' direction respectively (Fig. 24). The shutter speed, with respect to the previous test, was reduced to 1/1000 s in order to avoid saturation of those pixels far away from the pier where the sun was reflected.

A t -plot gave a peak frequency of about 0.085 Hz, and an energetic zone around 0.20 Hz. The significant wave height was approximately 0.60 m, and the maximum wave height was 1.10 m.

In Fig. 25 a sequence (time gap of 1 s between subsequent elevation maps) of height 3-D elevation maps is shown. Each map was obtained from the 3-D scattered points. All the 3-D scattered points were filtered in order to reduce the quantization error, thresholding the points with a spatial filter. In particular, the filter worked so that were discarded those points whose 3-D surface slope (computed as the mean slope around each point) is greater than 0.2. The remaining 3-D points were gridded through a triangle-based cubic interpolation (Barber et al., 1996). To interpolate the water surface elevations a horizontal regular grid of ground spacing of 10 cm was used. The 3-D

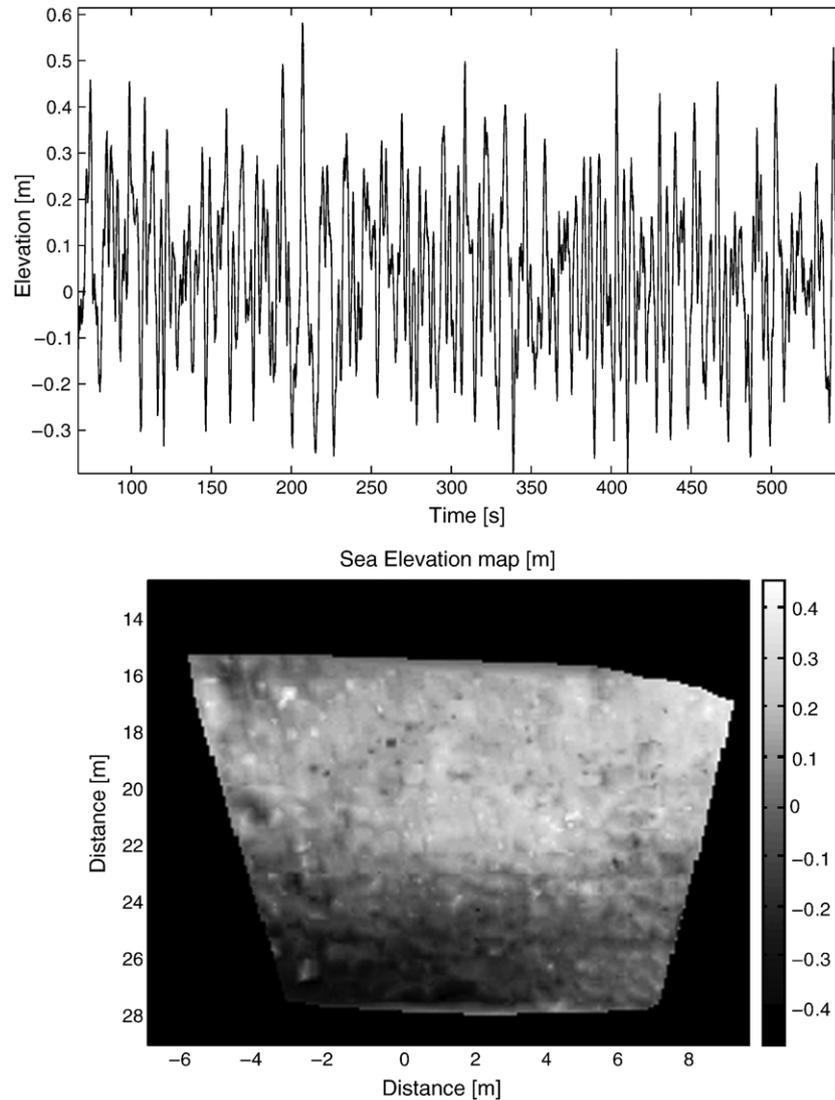


Fig. 26. Results of the stereo analysis. Upper panel: 1-D t -plot over an area of $10 \text{ cm} \times 10 \text{ cm}$. Lower panel: 2-D water surface elevation map.

water elevation map sequence in figure shows a wave train of period approximately 4.5 s. This period is consistent with the frequency peak shown in the frequency domain.

5. Remarks and conclusions

The synchronized optical vision from two calibrated vantage-points provides the opportunity to develop a technique for the remote measuring of water surface elevations from a sequence of overlapping images. The resulting method consists of a partially supervised algorithm for the 3-D location of corresponding image points. The stereo-processing is based on a pyramidal pixel-based correlation method, and it was derived from methods used in the field of Computer Vision, but adapted here to allow for greater precision, greater flexibility, and small computational time (precious when a large amount of data must be analyzed) in the oceanographic field of application. In this scenario, stereo-technique can be favourably exploited to measure the 3-D water elevation field wherever a fix location allow the cameras being fixed (i.e. a lighthouse, an oil platform).

Accuracy of the technique depends on the geometry of the stereo rig, and the cameras' resolution. The distance between the cameras and the inclination angle of the cameras' lines of sight strongly effect the errors in the 3-D measurements. Performance of stereo video imaging of waves is expected to degrade with increased distance between the cameras and the water surface. An optimal condition was found in a base to height ratio of 0.1. Certain wave conditions are more advantageous for use with stereo analysis. A rogue sea condition with sky diffuse lightning increases the image texture, making the task of identifying the corresponding pixels straight forward.

Global error variance has been related to the characteristics of the examined scenes through the epipolar line variance: high local variances in water movements, associated with high image-line-variance values, led to optimal performances of the proposed algorithm.

From field and synthetic tests it was determined that by using inclined cameras (with respect to the vertical direction) the errors in the vertical direction decrease. This configuration must

be carefully used because it stretches the image fields, i.e. the pixel footprint, increasing the matching error in the stereo recognition.

In terms of physical parameters of the waves, the resulting 3-D points offer a wide range of study methods. The simplest is based on the extraction of the t -plots (Fig. 26). Simultaneous t -plots on a regular grid allow one to calculate phase speeds and wavenumbers. In this case, the resolution and uncertainty will be dependent on the t -plot spatial separation and the time sampling interval. This procedure has shown a high potentiality in examining the wave dynamics, giving, besides this, all the results in a compact support.

A comparison of the stereo-measured water elevations with in-situ point sensors showed agreed. The bias for data collected in San Diego experiment was characterized by a mean absolute difference error of 4.7 cm, belonging to the expected quantization error in the vertical direction. The spectral coherence between the two signals, in the frequency range of the incident waves, was greater than 0.95.

2-D water elevation maps (Fig. 26) were obtained interpolating the 3-D scattered points result of the stereo procedure. This application typically involves the transformation of 3-D camera coordinates to 3-D world coordinates. The latter transformation requires only the hypothesis that over a long period the measured water elevation over a point is zero, setting the world coordinate system only by using the stereometric measurements.

In conclusion, the techniques described herein showed certainty for remotely measuring the dynamics of waves with high resolution in space and time. The errors, in the range of 1–100 m from the cameras, are limited in the practical range of sea wave measurements.

Acknowledgement

The author is grateful to Professor Attilio Adami who first devised this project, and to Emanuele Simioni who formalized the stereo algorithm. The author is grateful to Professor Ken Melville, of the Scripps Institution of Oceanography, San Diego, for the support received. A personal thanks to the reviewers for their comments.

References

- Ballard, D.H., Brown, C.M., 1982. *Computer Vision*. Prentice-Hall, Englewood Cliffs, NJ.
- Banner, M.L., Jones, S.F., Trinder, J.C., 1989. Wavenumber spectra of short gravity waves. *Journal of Fluid Mechanics* 198, 321–344.
- Barber, C.B., Dobkin, D.P., Huhdanpaa, H.T., 1996. The Quickhull Algorithm for Convex Hulls. *ACM Transactions on Mathematical Software* 22 (4), 469–483.
- Bendat, J.S., Piersol, A.G., 2000. *Random Data Analysis and Measurements Procedures*, 3rd ed. Wiley-Interscience, New York.
- Benetazzo, A., 2006. Measurement of water surface wave fields using a non-intrusive method based on a stereo-matching algorithm. PhD Diss., University of Padua.
- Bouguet, J.Y., 2004. Camera Calibration Toolbox for Matlab. Personal web-site: <http://www.vision.caltech.edu/bouguetj/calibdoc/>.
- Coté, L.J., Davis, J.O., Marks, W., McGough, R.J., Mehr, E., Pierson, W.J., Ropek, J. F., Stephenson, G., Vetter, R.C., 1960. The directional spectrum of wind generated sea as determined from data obtained by the stereo wave observation project. *Metereological Paper*, vol. 2, No. 6. New York University, College of Engineering.
- Dankert, H., Horstmann, J., Lehner, S., Rosenthal, W.G., 2003. Detection of wave groups in SAR images and radar image sequences. *IEEE Transactions on Geoscience and Remote Sensing* 41, 1437–1446.
- Dobson, E.B., 1970. Measurement of the fine-scale structure of the sea. *Journal of Geophysical Research* 75, 2853–2856.
- Dugan, J.P., Piotrowsky, C.C., Williams, J., 2001. Water depth and surface current retrievals from airborne optical measurements of surface gravity dispersion. *Journal of Geophysical Research* 106 (C8), 16903–16915.
- Faugeras, O., 1993. *Three-dimensional Computer Vision: A Geometric Viewpoint*. MIT Press.
- Holthuijsen, L.H., 1983. Observations of the directional distribution of ocean wave energy. *Journal of Physical Oceanography* 13, 816–827.
- Hwang, P., Wang, D., Walsh, E., Krabill, W., Swift, R., 2000. Airborne measurements of the wavenumber spectra of ocean surface waves, part i: spectral slope and dimensionless spectral coefficient. *Journal of Physical Oceanography* 30, 2753–2767.
- IAHR, 1986. List of sea state parameters. Supplement to Bulletin, vol. 52.
- Jähne, B., 1993. *Spatio-Temporal Image Processing — Theory and Scientific Applications*. Springer-Verlag, Berlin Heidelberg.
- Jähne, B., 1997. *Practical Handbook on Image Processing for Scientific Applications*. CRC Press, New York.
- Kanade, T., Morris, D.D., 1998. Factorization methods for structure from motion. *Philosophical Transactions - Royal Society, Mathematical, Physical and Engineering Sciences*, the Royal Society 356, 1153–1173.
- Klette, R., Schlus, K., Koschan, A., 1998. *Three-Dimensional Data from Images*. Springer, Singapore.
- Lippmann, T.C., Holman, R.A., 1991. Phase speed and angle of breaking waves measured with video techniques. *Coastal Sediments* 542–556.
- Ma, Y., Soatto, S., Kosecka, J., Shankar Sastry, S., 2004. *An invitation to 3-D Vision From Images to Geometric Models*. Springer-Verlag, New York.
- Pollefeys, M., 2000. Tutorial on 3-D modelling from images. *Lecture Notes*.
- Parks, T.W., Burrus, C.S., 1987. *Digital Filter Design*. John Wiley and Sons, New York.
- Santel, F., Linder, W., Heipke, C., 2004. Stereoscopic 3D-Image Analysis of sea surface. *Proceedings of the ISPRS Commission V Symposium*, July 12–23, 2004, Istanbul, Turkey, vol. XXXV, Part 5, pp. 708–712.
- Schumacher, A., 1939. Stereophotogrammetrische Wellenaufnahmen. *Wiss. Ergeb. Dtsch. Atlant. Exped. Forschungs Vermessung. "Meteor" 1925–1927, Ozeanographische Sonderuntersuchungen, Erste Lieferung*, Berlin.
- Shemdin, O.H., Tran, H.M., 1992. Measuring Short Surface Waves with Stereography. *Photogrammetric Engineering and Remote Sensing* 58, 311–316.
- Shemdin, O.H., Tran, H.M., Wu, S.C., 1988. Directional measurements of short ocean waves with stereography. *Journal of Geophysical Research* 93, 13891–13901.
- Stilwell, D., 1969. Directional energy spectra of the sea from photographs. *Journal of Geophysical Research* 74, 1974–1986.
- Stockdon, H.F., Holman, R.A., 2000. Estimation of wave phase speed and nearshore bathymetry from video imagery. *Journal of Geophysical Research* 105 (C9), 22,015–22,033.
- Sugimori, Y., 1975. A study of the application of the holographic method to the determination of the directional spectrum of ocean waves. *Deep-Sea Research* 22, 339–350.
- Svendsen, I.A., Jonsson, I.G., 1980. *Hydrodynamics of coastal regions*. Den private Ingeniørfond. Tech. Univ., Denmark.
- Welch, P.D., 1967. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics* AU-15, 70–73.
- Zhang, Z., 2000. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (11), 1330–1334.