Stereo Reconstruction of a Submerged Scene*

Ricardo Ferreira[1], João P. Costeira[1], and João A. Santos[2]

[1] Instituto de Sistemas e Robótica / Instituto Superior Técnico
[2] Laboratório Nacional de Engenharia Civil

Abstract. This article presents work dedicated to the study of refraction effects between two media in stereo reconstruction of a tridimensional scene. This refraction induces nonlinear effects making the stereo processing highly complex. We propose a linear approximation which maps this problem into a new problem with a conventional solution. We present results taken both from synthetic images generated by a raytracer and results from real life scenes.

1 Introduction

Physical modelling is, still today, the main tool for testing and designing costal structures, specially rubble-mound breakwaters. One of the most important failure modes of this kind of structure is the armour layer hydraulic instability caused by wave action. Currently, to test the resistance of a proposed design to

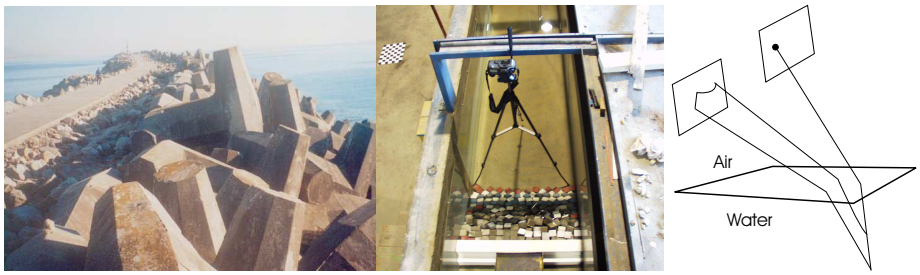

Fig. 1. Real and model breakwater.

this failure mode, a scale model of the structure is built in a wave tank or in a wave flume, such as the one shown in figure (1), and it is exposed to a sequence of surface waves that are generated by a wave paddle. One of the parameters that have proved of paramount importance in the forecast of the structure behaviour is the profile erosion relative to the initial undamaged profile. Thus, measuring and detecting changes in the structures envelope is of paramount importance.

* This work was supported by the Portuguese FCT POSI programme under framework QCA III and project MEDIRES of the AdI.

Laser range finders are one obvious and easy way of reconstructing the scene, however, since common lasers do not propagate in the water, the tank (or flume) have to be emptied every time a measurement is taken.

This is a quite expensive procedure, both in time and money resources. We propose to use a stereo mechanism to reconstruct a submersed scene captured from cameras placed outside of the water. This way we can monitor both the emerged and submerged part of the breakwater.

1.1 Problem definition

The problem tackled in this article is the reconstruction of a 3D scene with a stereo pair. Between the scene and the cameras there is an interface that bends light rays according to Snell's law.

The main difficulty here is that the known epipolar constraint, which helps reducing the search for a match, is not usable. Unlike conventional wisdom, straight lines underwater do not project as straight lines in the image. As figure 1.c illustrates, for each pixel in one image, possible matches are along a curve which is different for every point on the object. Essentially, this means that most stereo algorithms are unusable. We show that, if the incidence angle is small, the linear part of the Taylor Series expansion, which is equivalent to modifying camera parameters, is precise enough for our purpose. In other words current stereo algorithms can be used, provided the camera orientation parameters are within a certain range.

Though with a relatively straightforward solution, to our knowledge, this problem has not been addressed in the literature since most systems are placed underwater, thus eliminating the refraction issue.

2 Scene Reconstruction in the Presence of an Interface

2.1 Snell's law

Whenever an interface is involved, Willebrord Snell's Law will necessarily be spoken of. The law states that a light ray crossing an interface will be bent according to

$$k_1 \sin\varphi_i = k_2 \sin\varphi_r$$

where φ_i and φ_r are the angles the incident and refracted light rays have with respect to the normal of interface at the point of intersection. Considering a planar interface at $z = 0$ (see figure 1), a light ray emitted from a point above the interface will relate to its refracted ray by:

$$v_r^x(\mathbf{v_i}) = \frac{k_1}{k_2} v_i^x, \qquad v_r^y(\mathbf{v_i}) = \frac{k_1}{k_2} v_i^y \tag{1}$$

$$v_r^z(\mathbf{v_i}) = -\sqrt{\left(1 - \frac{(k_1)^2}{(k_2)^2}\right)\left((v_i^x)^2 + (v_i^y)^2\right) + (v_i^z)^2} \ .$$

This non-linear relation can be simplified by expanding $v_r^z(\mathbf{v_i})$ in its Taylor series (in the neighborhood of $\mathbf{v_i} = \begin{bmatrix} 0 & 0 & -1 \end{bmatrix}^T$) and retaining the first order term. This results in a much simpler (linear) transformation

$$\mathbf{v_r} \approx \begin{bmatrix} kv_i^x \\ kv_i^y \\ v_i^z \end{bmatrix} = \begin{bmatrix} k & 0 & 0 \\ 0 & k & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{v_i}, \quad \text{where } k = \frac{k_1}{k_2}. \tag{2}$$

2.2 Image Rectification

This approximation leads to a simple image rectification process, cancelling most of the distortion introduced by the interface. Using equation (2) and classic geometry, it can be shown that all light rays converge at a single point $\mathbf{p_1}$, as illustrated in figure 2. The relation between both focal points is done by:

$$\mathbf{p_1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{k} \end{bmatrix} \mathbf{p_2} \ . \tag{3}$$

This fact hints at the possibility of rectifying the image with refraction effects by only changing the extrinsic camera parameters. In other words, by approximating Snell's law, the problem with refraction is transformed into a typical stereo problem *"without"* air-water interface. All that remains to be done is to project the original image onto the $z = 0$ plane, and project it back to a virtual camera with projection center at $\mathbf{p_1}$. If $\mathcal{P}_2$ and $\mathcal{P}_1$ are, respectively, the original camera projection matrix and the virtual camera projection matrix, the rectification consists of a homography, given by:

$$\mathbf{H} = \mathcal{P}_1 \mathbf{M}(\mathbf{p_2}) \mathcal{P}_2^* . \tag{4}$$

Here, the operator $\{\cdot\}^*$ denotes matrix pseudo-inverse which projects a point in image coordinates onto the camera projection plane (at $z = 1$ in camera coordinates). Matrix $\mathbf{M}(\mathbf{p_2})$ projects a point onto the $z = 0$ plane using $\bar{\mathbf{p}}_\mathbf{2}$ as a projection center. It is defined by:

$$\mathbf{M}(\mathbf{p_2}) = \begin{bmatrix} -p_2^z & 0 & p_2^x & 0 \\ 0 & -p_2^z & p_2^y & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -p_2^z \end{bmatrix} \ . \tag{5}$$

The intrinsic parameters of the virtual camera are chosen to minimize information loss or any other criteria needed by the specific implementation. In particular in the case of stereo reconstruction, the image rectification process imposes a few constraints on these parameters.

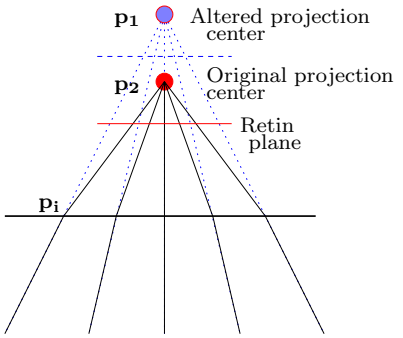

Fig. 2. Representation of the path followed by a beam of light when the first order snell approximation is used.

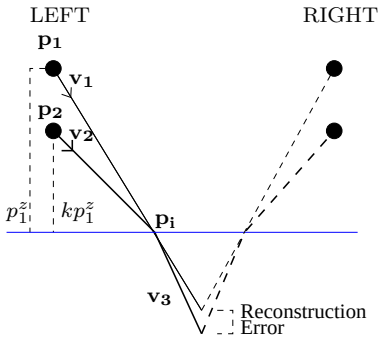

Fig. 3. Illustration of the correction needed to Snell's equations after image rectification.

2.3 Underwater Stereo Reconstruction

The previous rectification process changes the image in such a way that they become suitable to classic stereo reconstruction algorithms. Be advised though that no guarantee was made about epipolar lines. Generally, depending on the resolution used, baseline, and angle of incidence of the light rays, the epipolar constraint does not occur due to the effect of higher order terms, neglected by the Snell rectification. In case the rectification mentioned above is not accurate enough, two dimensional search must be done to match the images. In these circumstances, rectification can significantly narrow the band of search around the estimated epipolar line.

Although the matching process gains considerably by assuming the simplification as valid, for greater reconstruction precision the nonlinear terms shouldn't be discarded. After the matching has been done, the true Snell deformation can be taken into account. In other words, equations 1 must be modified to include the rectification effect on the image coordinates. This is illustrated in figure 3. Note that $\mathbf{v_3}$ is the true trajectory of the underwater light beam and not $\mathbf{v_1}$. We known how to obtain $\mathbf{v_3}$ from $\mathbf{v_2}$, but now only $\mathbf{v_1}$ is available. Finding the intersection of the line through $\mathbf{p_1}$ tangent to $\mathbf{v_1}$ with the plane $z = 0$ yields $\mathbf{p_i}$

$$\mathbf{p_i} = \begin{bmatrix} p_1^x - \frac{p_1^z}{v_1^z} v_1^x & p_1^y - \frac{p_1^z}{v_1^z} v_1^y & 0 \end{bmatrix}^T . \tag{6}$$

As mentioned before, Snell's approximation changed the camera's focal point. Knowledge about the original camera's focal point ($\mathbf{p_2}$) allows us to find $\mathbf{v_2}$:

$$\mathbf{p_2} = \begin{bmatrix} p_1^x & p_1^y & kp_1^z \end{bmatrix}^T , \qquad \mathbf{v_2} = \mathbf{p_i} - \mathbf{p_2} = \begin{bmatrix} -\frac{p_1^z}{v_1^z} v_1^x & -\frac{p_1^z}{v_1^z} v_1^y & -kp_1^z \end{bmatrix}^T .$$

Replacing this expression of $\mathbf{v_2}$ in equation 1, we can represent $\mathbf{v_3}$ exclusively as a function of the virtual camera, that is:

$$\mathbf{v_3} \propto \begin{bmatrix} v_1^x & v_1^y & -\sqrt{\frac{1-k^2}{k^2}\left(\left(v_1^x\right)^2 + \left(v_1^y\right)^2\right) + \left(v_1^z\right)^2} \end{bmatrix}^T . \tag{7}$$

It is now possible to apply equations (6) and (7) to the left and right cameras to triangulate for the 3D point. Due to the discrete nature of the sensors the two lines do not usually intersect, so a least squares error approach is used.

2.4 Implementation notes

The location of the water plane is obtained during the calibration process using a floating checkered board. For a description on how to use this plane to calibrate the cameras' extrinsic (and intrinsic) parameters please see Bouguet's work [2] which is based on Zhang [3] and Heikkïla [4]. As stated before, the water plane is forced (calibrated) to be at $z = 0$. In order to facilitate point matching, the calibration data is then used to project the left and right images on a common plane making the epipolar lines horizontal [5]. These images are then processed by any classic stereo reconstruction algorithm. In our case we were interested in a dense stereo reconstruction so we used Sun's algorithm [6] based on dynamic programming.

Please note that what is discussed in this paper is valid only for underwater scenes. If the scene to be reconstructed is only partially submerged, two reconstructions should be performed. One valid for all the pixels corresponding to points over water, and another for the pixels corresponding to underwater points. Since the water plane is at $z = 0$, it can be written as $\mathbf{w} = [0\ 0\ 1\ 0]^T$ in projective coordinates. This plane can be easily described in disparity space as $\mathbf{w}_d = H^{-T}\mathbf{w}$, using the projective transformation

$$\mathbf{H} = \mathcal{D}\mathcal{E}, \quad \text{where} \quad \mathcal{D} = \begin{bmatrix} f & 0 & c_l^x & 0 \\ 0 & f & c_l^y & 0 \\ 0 & 0 & c_r^x - c_l^x & -Bf \\ 0 & 0 & 1 & 0 \end{bmatrix} .$$

$\mathcal{E}$ is the world to camera projective transformation and $\mathcal{D}$ is the camera to d-space transformation with f describing the focal length, c_i^j the j coordinate (x or y) of the principal point of camera i (left or right) and B is the baseline between left and right cameras (see for example [7]). It is then possible to know in a disparity map which camera pixels correspond to points under or above water.

3 Experiments

To validate the algorithm, two different experiments were made. First a synthetic scene with planes at different depths was created. Images rendered from this scene are completely known to us, allowing reconstruction errors to be measured. The second type of images are real world images from a model breakwater. Since we do not have "ground truth" we can evaluate performance only qualitatively.

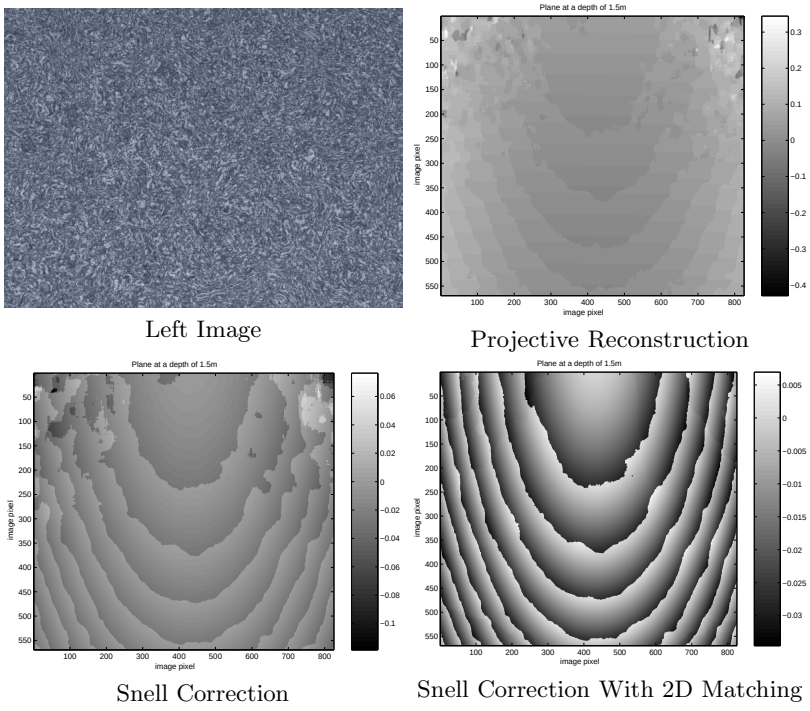

Fig. 4. Reconstruction error in depth (meters) for each pixel. The reconstructed scene consists of a textured plane at a depth of 1.5m as illustrated in the first image.

3.1 Synthetic Experiment

A few synthetic images were generated using povray[3] consisting of textured planes at various depths. The cameras are placed at 1.3m over the interface (looking slightly away from the perpendicular) with a baseline of 25cm . Please note that all of these reconstructions assume that the epipolar constraint is valid. This is clear in all the plane images since the matching algorithm starts to fail when the incidence angle becomes too great (noticeable in the top corners of the error images).

The first error image shown in figure 4 describes the reconstruction error when it is assumed that the disparity space is a projective reconstruction of the scenery. Note that Snell approximation is still used to help feature matching. The plane is reconstructed as a paraboloid (barely noticeable in the error images) due to the fact that higher order terms of Snell's law are discarded. This effect is much clearer in figure 5 where the actual plane reconstruction is shown. The top corners of the error image are poorly reconstructed due to the already mentioned failure in epipolar geometry.

[3] One of the oldest raytracers still used, which correctly models refraction effects.

The second error image shown in figure 4 uses equation 7 to correct the higher order distortion. Overall error is diminished but since nothing has been done to improve matching the top corners are still not corrected. For a clearer perception of the corrected distortion see figure 5 which shows the 3D reconstruction of the same plane (they are translated in relation to one another for visualization only) with (bottom plane) and without (top plane) use of equation 5. The plane reconstructed as a paraboloid effect mentioned earlier is clearly visible on the top plane. Although the planes are placed one above the other for comparisson purposes, they are both at the same depth (1.5m).

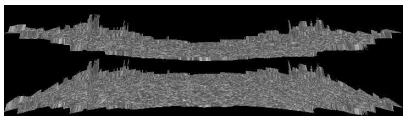

Fig. 5. 3D comparisson of plane reconstruction with snell correction applied and without it.

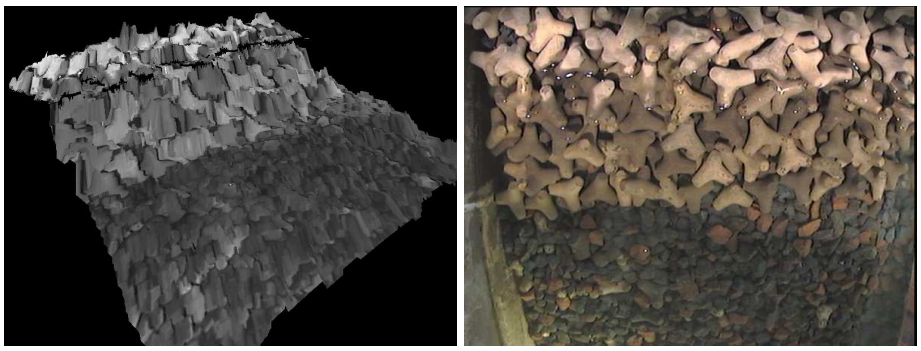

Fig. 6. 3D view and left image of a model breakwater partially submerged.

Finally, the result of using bi-dimensional matching is shown in the third error image of figure 4. Note that only a few pixels (depending on the resolution, baseline and depth of the scene) need be searched away from the epipolar line, and only where the angle of incidence is greater than a certain tolerance. The maximum error is now 3 centimeters for the plane at $z = -1.5$m, which is the expected error due to the discrete nature of the sensor at the given distance.

3.2 Real World Experiment

Figures 6 and 7 show two reconstructions of a real breakwater physical model. The first uses images taken with video low resolution PAL cameras with a baseline slightly below 40cm and about 1.2m above the water. The second uses images taken with a beam splitter mounted on a 6 megapixel still camera. The baseline is about 5cm at 1.2m above the interface. Notice in both reconstructions the

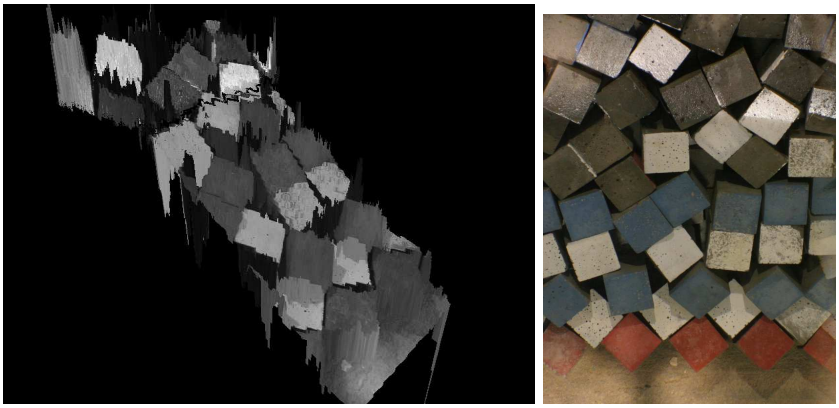

Fig. 7. 3D view and left image of another model breakwater partially submerged.

discontinuity near the top where the underwater and overwater reconstructions are fused. Unlike the synthetic images these are not so feature rich (for example dark shadows appear between rocks), resulting in some matching errors. Better results should be possible with algorithms that deal with occlusions and little texture.

4 Conclusion

We have shown how to diminish the refraction effect introduced by the presence of an interface between a stereo rig and the scene. The solution described allows for standard stereo matching algorithms to be used. The results show that the reconstruction error due to refraction is negligeable, provided the cameras are looking perpendicularly to the water surface.

References

1. G. Hough and D. Phelp. *Digital Imaging Procesing Techniques for the Aerial Field Monitoring of Harbour Breakwaters*, 1998.
2. http://www.vision.caltech.edu/bouguetj/
3. Zhang, Z. *Flexible Camera Calibration By Viewing a Plane From Unknown Orientations*, Microsoft Research, 1999.
4. Heikkild, J. and O. Silvin. *A Four-step Camera Calibration Procedure with Implicit Image Correction*, University of Oulu, 1997.
5. Pollefeys, M. *Tuturial on 3D Modeling from Images*, In conjuntion with ECCV 2000, Dublin, Ireland, Jun, 2000.
6. Sun. C. "Fast Stereo Matching Using Rectangular Subregioning and 3D Maximum-Surface Techniques", *International Journal of Computer Vision*, vol.47 no.1/2/3, pp.99-117, Mai, 2002.
7. Demirdjian, D. and T. Darrell. "Using multiple-hypothesis disparity maps and image velocity for 3-D motion estimation", Massachusetts Institute of Technology.