Calibration of optical camera coupled to acoustic multibeam for underwater 3D scene reconstruction

N.Hurtós, X.Cufí, J.Salvi
Department of Computer Architecture and Technology
University of Girona, Spain
Email: {nhurtos, xcuf, qsalvi}@eia.udg.edu

Abstract— **Combination of optical and acoustic sensors to compensate the strengths and weaknesses of each sensor modality is a topic of increasing interest in applications involving autonomous underwater vehicles (AUV). In this work, an opti-acoustic system composed by a single camera and a multibeam sonar is proposed, providing a simulation environment to validate its potential use in 3D reconstruction. Since extrinsic calibration is a prerequisite for this kind of feature-level sensor fusion, an effective approach to address the calibration problem between a multibeam and a camera is presented.**

I. Introduction

Three-dimensional underwater reconstruction allow to obtain the submarine cartography of certain areas of interest with applications in seashore resource management, environment affectation surveys and aquiculture. In order to obtain 3D information, scene key points from multiple underwater views (either supplied by multiple cameras or by a single moving camera) can be used to extract 3D estimates. However, while optical approaches provide high resolution and target details, they are constrained by limited visibility range. On the other hand, underwater sonars can operate in larger visibility ranges and provide 3D information even in presence of water turbidity conditions though at expense of a coarse resolution and harder data extraction. Hence, a promising emerging area of underwater 3D reconstruction has started to study the combination of data exploiting the complementary nature of optical and acoustic sensors. Despite the difficulty of combining two modalities that operate at different resolutions, technology innovations and advances in acoustic sensors have progressively allowed the generation of good-quality high-resolution data suitable for integration and consequently the related design of new techniques for underwater scene reconstruction. The main works combining some type of sonar (acoustic camera, single beam sounder, multibeam) and vision data [1] [2] [3] [4] [5] have been reviewed showing that in most of the approaches data integration is performed at a feature level, basically through geometrical correspondences and registration. The aim of this work is to propose an opti-acoustic system and solve the data alignment or sensor calibration problem which allows data from one sensor to be associated with the corresponding data of the other sensor.

The remainder of this paper is organized as follows. Next section introduces the system, formalizes the calibration problem and presents the approach to solve it. In section III we describe the simulation tests that have been done to evaluate the performance of the calibration method. In section IV results of calibration experiments performed with real data are presented. Finally an approach to validate the use of the system for 3D reconstruction is presented and discussed.

II. System description and calibration

A. System configuration

Our proposal aims to use a very simple opti-acoustic system in terms of sensors, taking profit of an acoustic sensor to obtain seabed range information while an optical camera is used to gather features such as color or texture. Regarding the acoustic sensor, we decided to work with a multibeam sonar since it offers much more resolution and coverage than a singlebeam echosounder; higher refresh rates and easier data treatment than a mechanically scanned profiler; and, in general, better bathymetric data (which is our main concern) than side-scan or forward looking imaging sonars which are more aimed at imaging tasks, besides being much more expensive.

Hence, the proposed system is constituted by a camera and a multibeam sonar that will be attached to an autonomous underwater vehicle in order to acquire images and profiles of the vehicle's underlying seafloor. In order to later combine information from both sensors its configuration must be such that part of the swath from the multibeam sonar intersects the projection area of the image (fig.1).
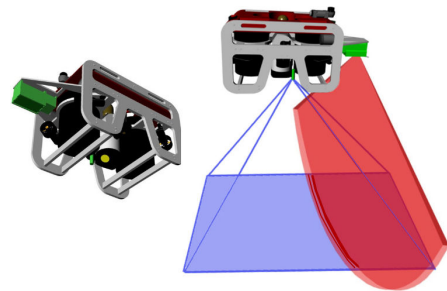

Fig. 1. Scheme of the system's proposed configuration.

Knowing the relative pose of the two sensors, an acoustic profile can be projected onto the optical image plane. Apart from the utility of this system for 3D reconstruction, the potential use of this mapping to robustly characterize features becomes evident, i.e a feature could be described by an interest

point descriptor from the image but also by a particular depth and/or a specific acoustic reflectivity.

The sensors of the real system are a Super SeaSpy camera from Tritech and a 837B Delta T multibeam sonar from Imagenex. The specifications of that equipments have been used to parameterize the models used within our framework.

The camera of the system is modeled using the standard pinhole model. Therefore the mapping from 3-D world coordinates to 2-D coordinates in the image plane is defined by the perspective projection matrix $\tilde{P}$:

$$\tilde{\mathbf{P}} = \mathbf{K} \cdot {}^c[\mathbf{R}|t]_w, \quad \mathbf{K} = \begin{pmatrix} \alpha_x & s & u_0 \\ \alpha_y & 0 & v_0 \\ 0 & 0 & 1 \end{pmatrix} \quad (1)$$

Here, $\mathbf{R}$ and t encode the coordinate transformation from world to camera frame and $\mathbf{K}$ is the known intrinsic camera calibration matrix where α_x and α_y are the pixel focal lengths in the x and y directions respectively, (u_0, v_0) is the principle point measured in pixels, and s is the pixel skew.

Regarding the multibeam sonar, we can define its geometry with the following parameters:

- An origin at the sonar position $\{Mb\}$.
- Along-track (longitudinal) aperture θ_L, which is the beamwidth in the horizontal plane usually narrow to insonify a thin strip of the terrain across-track of the vehicle.
- Across-track (transversal) beam aperture θ_T, which is the width of each beam in the vertical plane that provides the angular discrimination for reception beams.
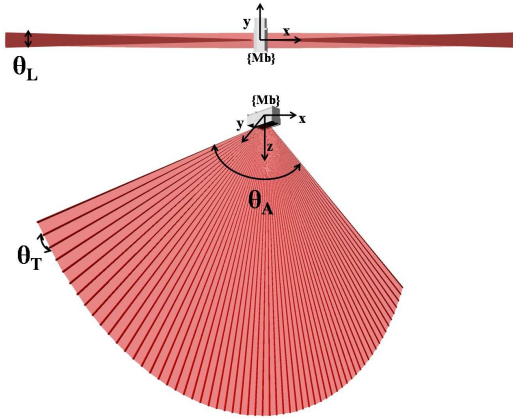- A maximum aperture of the sonar's fan θ_A,

Fig. 2.   Geometry of the multibeam model

The Delta T header offers up to 480 beams over a 120° interval giving an across-track angular resolution of 0.25° and 0.75° resolution along-track. For the simulations presented in this work we use a simplified model reduced to a number of rays equally distributed along the total aperture of the sonar and obviating the along-track aperture so that all the rays lay on the plane Y=0. Due to the narrow apertures of our real multibeam and the fact that we expect to employ the system

for performing surveys around 5m depth, this assumption stays close to the reality.

B. Extrinsic calibration of the system

Extrinsic calibration of two sensors consists in recovering the fixed but unknown rigid transformation that relates the two reference frames. Calibration methods are highly dependant on the sensors that compose the system and especially on the type of data they provide. Although there are some methods that address the calibration between optical and acoustical cameras[3] [6], to the best of the authors' knowledge it does not exist any method in the literature that specifically addresses the calibration between a multibeam sonar and a camera.

In principle, to calibrate the proposed system, a suitable object should be manufactured which gives raise to distinct features both in the acoustic profile and in the image. Provided with a set of 3D to image correspondences the matrix relating the camera and the multibeam coordinate systems can be obtained using the well known Direct Linear Transform (DLT) algorithm [7]. In order to assess how reliable would be to establish these opti-acoustic correspondences, we performed some tests with the real sensors in a water tank, trying to identify 3D corners such as vertexes of geometric polyhedrons in both acoustic profiles and images. However, establishing these direct opti-acoustic correspondences becomes a hard task due to the different resolution of both sensors and the noise of the acoustic data.

Hence, methods that rely on explicit opti-acoustic matches have been avoided. Since we assume a simplified multibeam model, our problem might be considered similar to a calibration of a camera-laser system. Our proposal is an adaptation of the method presented by Zhang and Pless [8] to calibrate a camera and an invisible laser range finder. This method is based on observing a planar pattern with both sensors at a same time from several poses and imposing a series of geometrical constraints through the data lying on that plane. Although acoustic data is noisier than laser data we have tried to adapt the calibration procedure to our problem and evaluate its suitability.

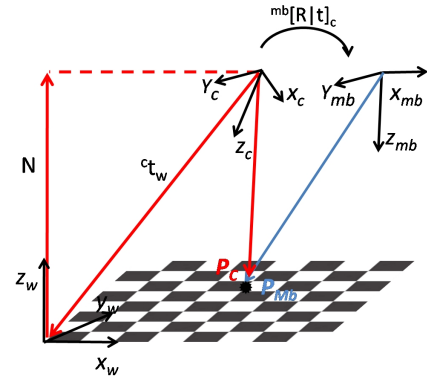The geometry of the calibration is shown in figure 3.

Fig. 3.   Geometry of the calibration method

A planar calibration plane (i.e a checkerboard) must be placed in front of the camera-multibeam system. In this case we must detect the multibeam points that lie on the plane, which is much easier than detecting an isolated 3D point. In order to enhance the robustness of the multibeam point detection, the calibration plane could be built from a material with a characteristic acoustic reflectivity which will provide another attribute to discriminate the points lying on the plane from the other points that impact with the water tank walls.

The method assumes that the camera has been already calibrated. Therefore, since the extrinsic parameters are known, the calibration plane can be parameterized in the camera coordinate system, by a 3 column vector N such that N is parallel to the normal of the calibration plane and its magnitude $\|N\|$ equals the distance from camera to the calibration plane.

Supposing a calibrated camera, and thus knowing the rotation and translation matrixes from world to camera frame ($^c[\mathbf{R}|t]_w$), we can derive that:

$$N = R_3(R_3^T(-t)) \tag{2}$$

where R_3 is the 3rd column of rotation matrix $^c\mathbf{R}_w$. Since the multibeam points must lie on the calibration plane estimated from the camera, we can establish a geometric constraint on the rigid transformation between the camera and multibeam coordinate systems. Given a multibeam point P_{mb} in the multibeam coordinate system we can determine its coordinates P_c in the camera reference frame as:

$$P_c = ^{mb}\mathbf{R}_c^{-1}(P_{mb} - ^{mb}t_c) \tag{3}$$

Since the point P_c is on the calibration plane defined by N, it satisfies that $N \cdot P_c = \|N\|^2$. Then we have:

$$N \cdot ^{mb}\mathbf{R}_c^{-1}(P_{mb} - ^{mb}t_c) = \|N\|^2 \tag{4}$$

For a measured calibration plane parameters N and multibeam point P_{mb}, equation 4 gives a constraint on $^{mb}R_c$ and $^{mb}t_c$.

Since all multibeam points are on the plane $Y = 0$ in the multibeam coordinate system, a multibeam point can be represented by $P_{mb} = [x, z, 1]^T$

Then we can rewrite equation 4 as:

$$N\mathbf{H}P_{mb} = \|N\|^2 \tag{5}$$

where $\mathbf{H}$ is a 3×3 transform matrix from the multibeam to the camera coordinate system of the form:

$$\mathbf{H} = ^{mb}\mathbf{R}_c^{-1} \begin{pmatrix} 1 & 0 & \\ 0 & 0 & -^{mb}t_c \\ 0 & 1 & \end{pmatrix} \tag{6}$$

Let $\mathbf{H} = [h_{ij}]_{3x3}$ and $N = [n_i]_{3x1}$, then we have a linear constraint on $\mathbf{H}$ for each measurement:

$$ah = \|N\|^2 \tag{7}$$

with $a = [n_1 x, n_1 z, n_1, n_2 x, n_2 z, n_2, n_3 x, n_3 z, n_3]$ and $h = [h_{11}, h_{12}, h_{13}, h_{21}, h_{22}, h_{23}, h_{31}, h_{32}, h_{33}]^T$. If a total number

of n multibeam points are observed, we can assemble n of the previous equations, obtaining:

$$\mathbf{A}h = B \tag{8}$$

where $\mathbf{A}$ is $nx9$ matrix and B is $nx1$ vector. If $n \geq 9$, we obtain a unique solution h since $h = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}B$. Once H is determined by solving for h, we can estimate camera relative orientation and position as follows:

$$^{mb}\mathbf{R}_c = [H_1, -H_1 \times H_2, H_2]^T \tag{9}$$
$$^{mb}t_c = -[H_1, -H_1 \times H_2, H_2]^T H_3 \tag{10}$$

where H_i is the ith column of matrix $\mathbf{H}$.

Hence we are able to retrieve the $^{mb}[\mathbf{R}|t]_c$ transformation that relates the coordinate systems of the camera and the multibeam.

III. CALIBRATION IN A SIMULATION ENVIRONMENT

Some simulations have been performed in MATLAB to validate the method using the geometrical modeling of the sensors described in section II-A. To simulate the calibration procedure, we placed the calibration plane without loss of generality at the Z=1 plane of the world coordinate system. Multibeam coordinate system is placed facing down towards the plane with a rotation of some small random angles. Then we fix the camera system with respect to the multibeam making use of the transformation $^{mb}[\mathbf{R}|t]_c$, which will be the ground truth for the extrinsic calibration. This process is repeated along a trajectory defined over the calibration plane.

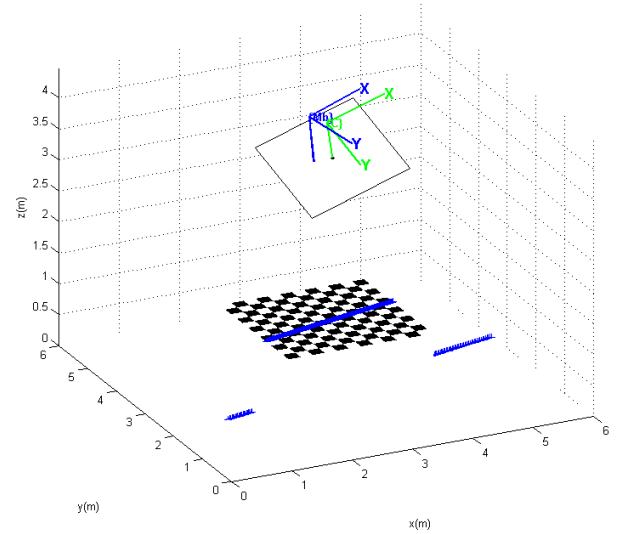Figure 4 shows an instant of the simulation where a pair profile/image is being acquired.

Fig. 4. Simulation of an acquisition during the calibration procedure.

After the simulated acquisition of all the points, we determine for each profile/image pair the multibeam points that are lying on the plane and the camera pose based on the acquired image. Having this information, we solve the

TABLE I
ERROR OF THE ESTIMATED EXTRINSIC PARAMETERS FOR DIFFERENT AMOUNTS OF NOISE

Noise Std Deviation (m)	Orientation error (degrees)	Position error (m)
0.02	0.195	0.006
0.05	0.286	0.008
0.1	0.893	0.025
0.2	2.034	0.078

equation system described in the equation 8 and we retrieve the matrix $^{mb}[\mathbf{R}|t]_c$. When dealing with ideal points, the rotation and translation matrices between both sensors are recovered precisely.

However, in order to simulate more real conditions, several sources of error have been introduced to the simulation. On one hand, we have introduced uniform noise in the range of the multibeam points, based on the range resolution of the sonar which according to the specifications corresponds to 0.2% of the range. In addition, gaussian noise has been introduced to the points with a standard deviation according to the tilt angle of the beams.

When introducing noise, the resulting matrices are used as an initial guess for a non-linear optimization problem which minimizes the Euclidean distances from every multibeam point j to its corresponding checkerboard plane i:

$$\sum_i \sum_j \left(\frac{N_i}{\|Ni\|} \cdot (^{mb}\mathbf{R}_c^{-1}(P_{ij}^{mb} - ^{mb} t_c)) - \|N_i\| \right)^2 \quad (11)$$

After this minimization, solved by Levenberg-Marquardt algorithm in Matlab, the obtained results show good tolerance to the introduced noise. Table I shows the results for different amounts of noise that have been introduced to the multibeam points. to perform simulated calibrations. To compute the error, the estimated extrinsic parameters are compared with the ground truth. We measure the error for the orientation $^{mb}\mathbf{R}_c$ by computing the angle between the estimate and the true orientation, and the error for the position $^{mb}t_c$ by computing the distance between the estimate and the true position. As Zhang and Pless remark, when increasing the number of profile/image pairs the error decreases. In the experiment summarized in table I a total of 25 profile/image pairs have been used in each calibration, with plane angles varying from -40 to 40 degrees. Theoretically, increasing the orientation angle of the checkerboard planes would decrease the error, however when the angle is big, less multibeam points fall onto the plane and its detection becomes more inaccurate due to the foreshortening.

It can be seen that for the introduced noise in the multibeam points the extrinsic calibration matrix can be recovered with tolerable errors. For example, considering the angle apertures of our real multibeam, the incidence area of the beams with the biggest tilt angle at 5 m depth is approximately a circle of 5 cm. Looking at the results, that means that operating at this depth we can keep the errors below the centimeter.

IV. CALIBRATION WITH REAL DATA

The calibration method has also been tested with real data gathered in the water tank of the Underwater Robotics Research Centre at the University of Girona (fig.5), using a pole-mounted system with the two sensors (fig.6). The pole has been fixed so that the end where the two sensors are attached remains still inside the water. A planar pattern (a checkerboard of 10x10 squares, each one of size 56mm x 56mm) has been moved in different positions and orientations under the system while recording a video sequence together with the sonar log file during approximately 3 minutes. Around 1300 profile/image pairs have been recorded.
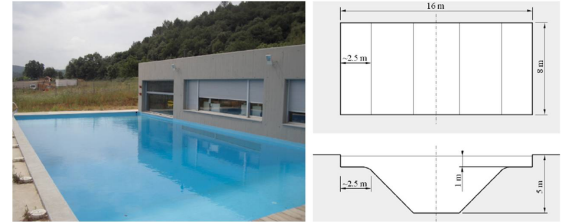

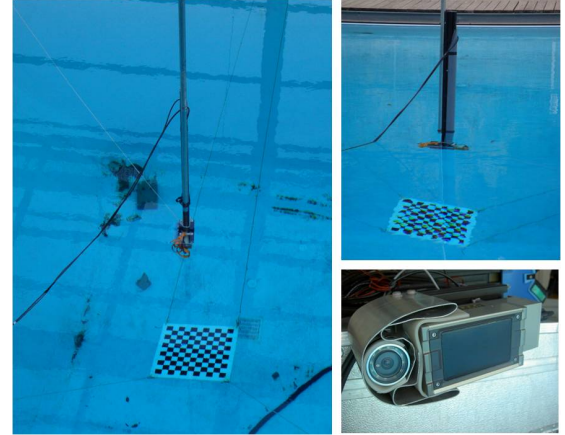Fig. 5.   Water tank at the university of Girona

Fig. 6.   Calibration setup in the real environment.

Given that the procedure is very sensitive to outliers, a strict filtering process is performed before taking any data into account. To identify for each multibeam swath the data that is lying on the checkerboard plane, we first perform a straightforward thresholding depending on the range values. Since the planar pattern was flying at mid-water, those beams that intersected with the plane report a much shorter range than those that reached the bottom of the water tank. In order to reduce the noise, a line has been fitted to the 3D points that fall on the plane using RANSAC algorithm, keeping as inliers those points within 1mm distance of the fitted line (fig. 7). On the other hand, those frames where the checkerboard does not appear entirely on the image or frames that offer a big reprojection error of the corners on the image plane, have been discarded together with its corresponding multibeam swath. It is worth to mention that since an important point

to ensure the robustness of the method is to have a great number of profile/image pairs, the calibration procedure has been implemented to be as automatic as possible, so that the user is not required to manually identify the corners for each frame in the data to extract the extrinsic parameters of the camera. To this end, we make use of the corner autoextraction algorithm developed by Martin Rufli and Davide Scaramuzza [9].

Figure 8 shows a calibration step where the points on the plane are filtered out from the acquired multibeam profile and the image corners are detected for the corresponding image frame.
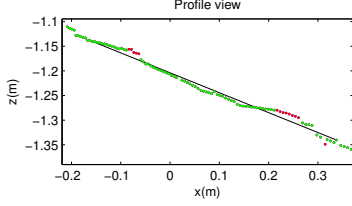
Fig. 7. Detail of the RANSAC line fitting for the multibeam points on the plane. Green points are considered inliers and red points outliers.
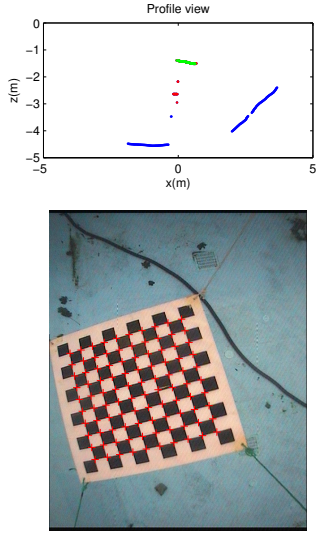

Fig. 8. Calibration step. Detection on the points lying on the plane and corners of the image frame.

After all the filtering process, 137 different profile/image pairs have been used, comprising plane orientations going from -40° to 40° degrees and resulting in a total number of 9713 multibeam points entered to the equation 8.

The calibration of the system using the described methodology reported the following translation and rotation between the multibeam and the camera coordinate frames (fig.9):

$$^{mb}t_c = (-0.0026; -0.1035; -0.055)$$

$$^{mb}\mathbf{R}_c = \begin{pmatrix} -0.9991 & -0.0414 & 0.0113 \\ 0.0410 & -0.9985 & -0.0363 \\ 0.0128 & -0.0358 & 0.9993 \end{pmatrix}$$
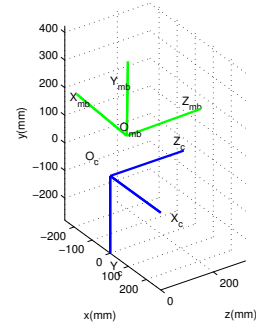

Fig. 9. Relative pose of the camera and multibeam coordinate systems .

Although we do not have the ground truth for the estimated $^{mb}[\mathbf{R}|t]_c$ matrix , the recovered offsets between the camera and multibeam are coherent with the real arrangement of the sensors during the experiment and show agreement with the coarse measurements that can be done over the system.

Using the obtained calibration matrix, multibeam points can be reprojected back to the image plane as shown in figure 10(a). Figure 10(b) shows the reprojection over the 3D calibration grid where it can be seen that the points effectively lie on the plane.
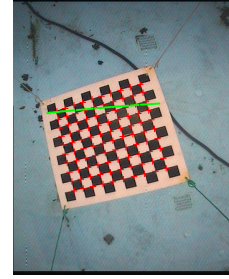
(a) Reprojection of the multibeam points over the image plane
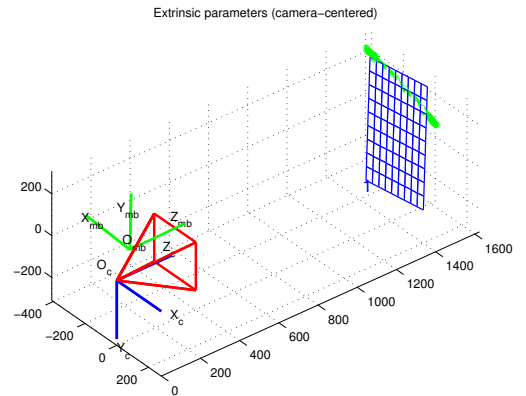
(b) 3D reprojection of the multibeam points over the calibration grid

Fig. 10.

When computing the distance of all the reprojected multibeam points to their respective calibration planes an average

error of 2.34cm is obtained, which demonstrates that the calibration can accurately relate the two coordinate frames.

V. SIMULATION ENVIRONMENT FOR 3D RECONSTRUCTION

A straightforward approach considering ideal calibration and navigation data has been implemented to demonstrate the applicability of the calibrated system to obtain a 3D reconstruction of the seafloor. We have created a simulation environment which allows to define a trajectory over a virtual terrain and gather synthetic images and multibeam data. It is worth to underline that simulations concentrate only in the geometry of the system, disregarding camera photometric issues (i.e lighting conditions) or acoustic reflectivity parameters. Given a trajectory of camera poses over the terrain, the multibeam coordinate system is placed at each of the corresponding locations with respect to the camera. In this way, image data can be properly reprojected over the bathymetry giving a 3D reconstruction of the terrain which comprises both range and visual information (fig. 12).
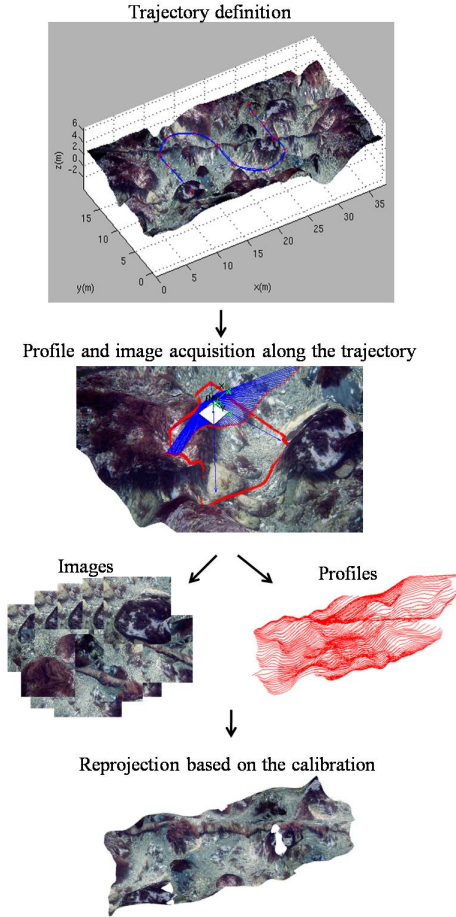

Fig. 11.   3D reconstruction in the simulated environment

It can be observed that the terrain along the trajectory is effectively recovered both in terms of relieve and visual information. A degradation on the visual information can also be perceived because of the low resolution of the simulated

images and the final interpolation. However such a direct approach would rarely have a good performance in real conditions since data accuracy of typical navigation systems far exceeds the intrinsic accuracy of the sonar. Hence, an appropriate algorithm should be designed in the future in order to enforce local and global consistency within navigation data and sensor measurements to achieve satisfactory mapping results. However, this first approach is helpful to validate the calibration results when the positions of the system sensors are well known.

VI. 3D RECONSTRUCTION WITH REAL DATA

The approach described in the previous section has been followed as a first attempt to reconstruct real data, making use of well known positions to establish a mapping with the help of the calibration data. Thus, besides of the 3D reconstruction, this experiment constitutes another way to validate the calibration of the system.

A checkerboard pattern has been placed at the bottom of the water tank and the pole-mounted system has been hung from one end so that the camera and the multibeam sonar stay inside the water and can be moved easily by pulling some ropes.
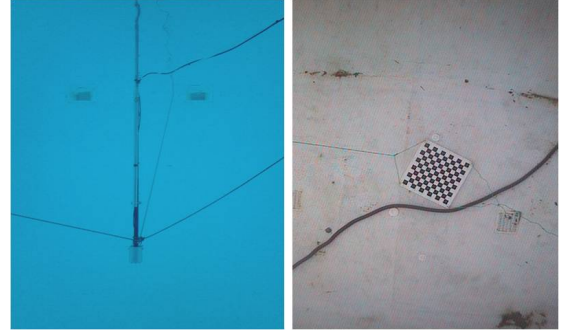

Fig. 12.   Setup of the experiment in the water tank with the pole-mounted system and the calibration pattern at the bottom

Since we have the camera calibrated, the checkerboard at the bottom of the tank allows us to extract the pose of the camera with respect to it with relatively good accuracy. Knowing the camera pose and knowing the calibration of the camera-multibeam system obtained in section IV, we can place the multibeam coordinate system for each camera pose. This allows us to reproject the profile on to the corresponding image and thus obtain colour information for each 3D profile point.

Therefore, the pole-mounted system has been moved in several directions and orientations while recording images and multibeam data. Figure 13 shows the estimated camera poses with respect to the calibration pattern, while figure 14 shows the walls and the bottom that has been reconstructed.

It can be seen that the profiles are correctly orientated defining the 'U' shape of the water tank. However since the field of view of the camera is smaller than the multibeam aperture, the colour information could only be mapped in the central part of the profiles. The checkerboard pattern as well as the cable that was lying on the bottom of the tank can
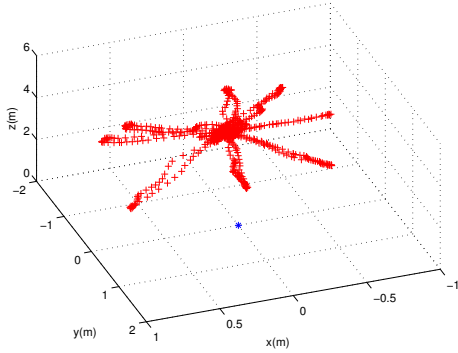
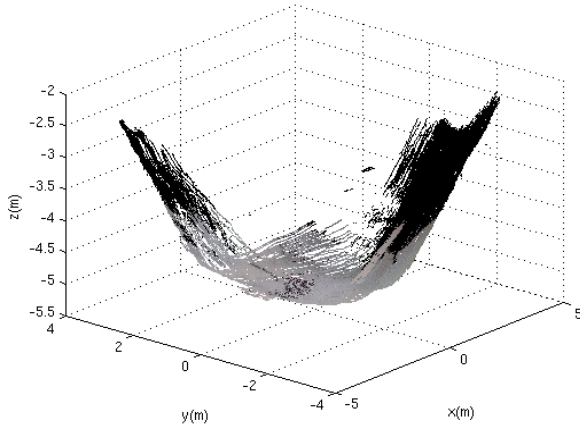Fig. 13. Camera poses with respect to the calibration plane.

Fig. 14. 3D reconstruction of the water tank bottom.

be slightly appreciated in the mapped profiles. However, the reconstruction lacks of consistency probably due to errors in the estimation of the camera pose and the fact that the dataset is far from being dense.

VII. CONCLUSION

In this work, we have presented a first step towards the integration of optic and acoustic information for the three-dimensional reconstruction of underwater scenes. An opti-acoustic system composed of a camera and a multibeam sonar has been proposed, providing simulations to validate its potential use both in the establishment of robust features and the 3D reconstruction of environments. In order to calibrate the system an approach originally developed for a calibration of a laser range finder and a camera has been considered and tested both in simulation and in real conditions showing that the relative pose between the two sensor frames can be accurately recovered. However, in order to use the system for 3D reconstruction a more complex reconstruction approach should be devised. The most suitable option seems to integrate the opti-acoustic system into a visual SLAM framework in order to enforce consistency within navigation data and the sensor measurements to yield satisfactory mapping results.

REFERENCES

[1] H. Singh, C. Roman, L. Whitcomb, and D. Yoerger, "Advances in fusion of high resolution underwater optical and acoustic data," in *Proc. IEEE Int. Symp. Underwater Technology*, pp. 206–211, 2000.

[2] S. B. Williams and I. Mahon, "Simultaneous localisation and mapping on the great barrier reef," in *IEEE Proceedings Int. Conf. on Robotics and Automation, 2004. ICRA'04*, pp. 1771–1776, 2004.

[3] A. Fusiello and V. Murino, "Augmented scene modeling and visualization by optical and acoustic sensor integration," *IEEE Transactions on Visualization and Computer Graphics*, vol. 10, no. 6, pp. 625–635, 2004.

[4] S. Negahdaripour, H. Sekkati, and H. Pirsiavash, "Opti-acoustic stereo imaging, system calibration and 3-d reconstruction," in *CVPR*, 2007.

[5] M. Johnson-Roberson, O. Pizarro, and S. Willams, "Towards large scale optical and acoustic sensor integration for visualization," in *OCEANS 2009-EUROPE, 2009. OCEANS '09.*, pp. 1–4, May 2009.

[6] S. Negahdaripour, H. Sekkati, and H. Pirsiavash, "Optiacoustic stereo imaging, system calibration and 3D reconstruction," *Proc. IEEE Beyond Multiview Geometry*, 2007.

[7] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge Univ Pr, 2003.

[8] Q. Zhang and R. Pless, "Extrinsic calibration of a camera and laser range finder," in *IEEE/RSJ Proceedings Int. Conf. on Intelligent Robots and Systems, 2004. IROS 2004*, vol. 3, pp. 2301–2306, 2004.

[9] M. Rufli, D. Scaramuzza, and R. Siegwart, "Automatic detection of checkerboards on blurred and distorted images," in *IEEE/RSJ International Conference on Intelligent Robots and Systems, 2008. IROS 2008*, pp. 3121–3126, 2008.