Integration of optical and acoustic sensors for 3D underwater scene reconstruction

N. Hurtós, X.Cufí, J.Salvi
Department of Computer Engineering, Universitat de Girona -17001 Girona, Spain
{nhurtos,xcuf,qsalvi}@eia.udg.es

***Abstract*— *Combination of optical and acoustic sensors to overcome the shortcomings presented by optical systems in underwater 3D acquisition is an emerging field of research. 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 system is presented.*

***Keywords*— *optical and acoustic fusion, multibeam sonar, underwater 3D reconstruction, calibration*

I. INTRODUCTION

Three-dimensional underwater reconstructions allow the analysis of the temporary evolution of marine ecosystems, as well as the morphology of the underwater seafloor. 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. 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 [6], [7], [2], [5], [4] combining visual data and some type of sonar data have been reviewed showing that data integration is performed at a feature level, basically through geometrical correspondences and registration. In this way a crucial problem of these sensor fusion techniques is the sensor calibration problem which allows data from one sensor to be associated with the corresponding data of the other sensor. Besides, the calibration procedures used by the reviewed approaches are all different, denoting that calibration algorithm is highly dependent on the sensors involved in the system. In section 2 we detail the proposal of an opti-acoustic system for 3D reconstruction. In section 3 we discuss a method to address the extrinsic calibration of the proposed system. Finally in section 4, we present the conclusions and we point out future work.

II. PROPOSAL OF OPTI-ACOUSTIC SYSTEM

A. System description

In the same line of the reviewed approaches, we want to take profit of an acoustic sensor to obtain seabed range information, while a camera is used to gather other features such as color or texture. Regarding the optical part of the system, we will consider a single video camera. Approaches which could involve more cameras such as an stereo system are discarded for the sake of simplicity and to decrease power consumption (since the system will be boarded on an AUV). Regarding the acoustic sensor, the number of possible choices is much higher. High-resolution acoustic cameras could achieve a good performance when combining its data with that of a video camera. Moreover, since they can provide a three-dimensional sonar image with a single shot they reduce the need to co-register the acoustic data using positioning and motion system equipment. However, acoustic cameras are approximately five times more expensive than other acoustic sensors such as multibeam systems. Unfortunately an acoustic camera is not economically viable for us. Hence it seems that the best option we have on hand is a multibeam sonar. A multibeam 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.

Therefore, the proposed system is constituted by a camera and multibeam sonar that will be attached to the vehicle rigid frame 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, as seen in figure 1.

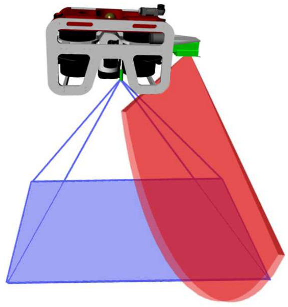


Fig. 1. Scheme of the system's proposed configuration.

B. Sensor Modeling

1) Camera: Within our framework we assume a standard calibrated pinhole camera model. Hence the mapping from 3-D world coordinates to 2-D coordinates in the image plane is defined by the perspective projection matrix $\tilde{P}$:

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

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.

2) Multibeam sonar: The multibeam sensor reports sonar readings which are range and bearing measurements of the intersection of each beam with the underlying surface. Its geometry (figure 2) can be modeled defining 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,

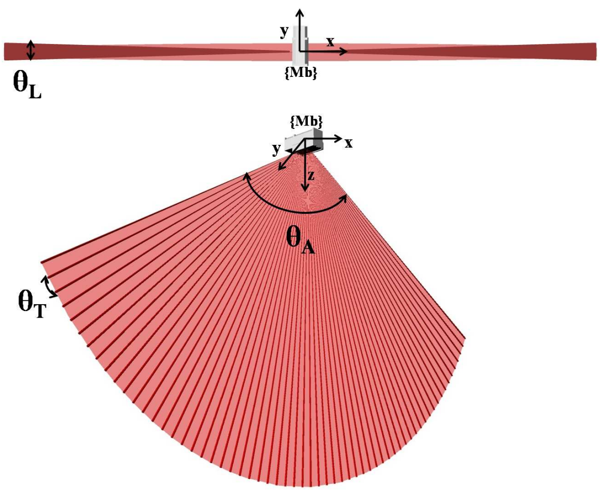


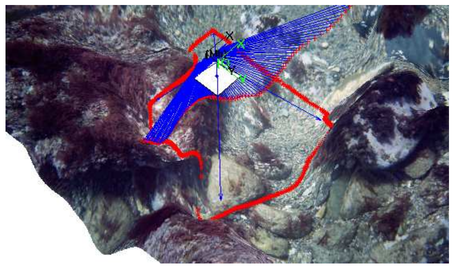
Fig. 2. Geometry of the multibeam model

In order to simplify 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 expected altitude of the surveys we want to carry on, this assumption is not far from reality.

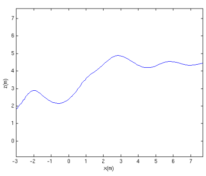
C. Simulations

A simulation environment has been created in order to perform tests using camera and multibeam data. This framework has been useful to validate the potential use of the opti-acoustic system as well as the performance of the calibration method proposed in next section. 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.

1) Sensors: System sensors have been simulated using the models described in the previous sections which have been parameterized with the values of the real sensors mounted in our AUV, a Tritech Super SeaSpy camera and a DeltaT multibeam sonar from Imagenex. Figure 3 shows the camera and multibeam coordinate systems placed at one point of the simulation environment and the corresponding image and profile acquired by the simulated sensors at this point.



(a) Camera and multibeam simulation



(b) Acquired profile

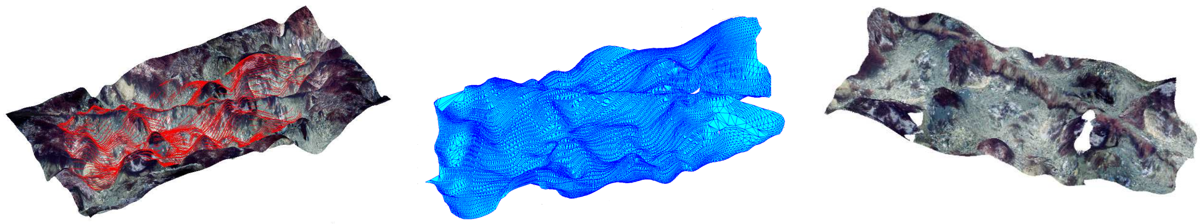
(c) Acquired image

Fig. 3.

2) Mapping: Given an acoustic profile, composed by a set of target points, each with a certain 3D position, we want to project it onto the optical image plane, obtaining the depth (and eventually also the reflectivity value) with reference to the image plane. Within simulation environment, this mapping is established fixing the rigid transformation matrix between the two sensors ${}^{mb}[\mathbf{R}|t]_{c}$. Applying this transformation as in (2) we obtain the mapping of the profile points (P_{mb}) in the camera reference system (fig. 4).

$$P_c = {}^{c}\mathbf{R}_{mb} \cdot P_{mb} + {}^{c}t_{mb} \quad (2)$$

Fig. 4. Mapping multibeam points to image frame.



(a) Profile points along the trajectory (b) Delaunay triangulation (c) Reconstruction

Fig. 5.

Since the intersected surface has 3D relieve, the profile does not describe a line with respect to the image plane (fig 4) and it can be seen how it is warped according to the shape of the underlying rock. The potential use of this mapping to characterize robust features can be perceived. A feature lying on the rock could be described by an interest point descriptor from the image but also by a particular depth (which will differentiate it of the features lying on the sandy bottom) and a specific acoustic reflectivity, as it will have a reflectivity value particular of the insonified material (rock, sand, algae, etc).

3) 3D reconstruction: A straightforward approach considering ideal calibration and navigation data has been implemented to demonstrate that a calibrated camera-sonar system can be used to obtain a 3D reconstruction of the seafloor. Given the profile points (figure 5(a)) along a trajectory a surface mesh is generated using Delaunay triangulation (figure 5(b)) and subsequently interpolated and resampled. Supposing the camera center of projection at the each of the corresponding locations with respect to the multibeam the images can be reprojected over the reconstructed bathymetry (figure 5(c)). It can be observed that the terrain along the trajectory is effectively recovered both in terms of relieve and visual information. However, such an straightforward approach would rarely have a good performance in real conditions. Even supposing an ideal calibration between the sonar and the camera, the 3D reconstruction will be constrained by the mismatch between navigation and sensor accuracy. Typical navigation systems rely on the combination of acoustic transponders and inertial navigation systems whose data accuracy far exceeds the intrinsic accuracy of the sonar. Some efforts [1] that exploit vehicle attitude and navigation information and enforce local and global consistency within sensor measurements have been developed, yielding to superior mapping results commensurate with sensor accuracy.

III. EXTRINSIC CALIBRATION BETWEEN A CAMERA AND A MULTIBEAM SONAR

In principle, to calibrate the sensors, 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 ${}^{mb}[\mathbf{R}|t]_c$ matrix relating the camera and the multibeam coordinate systems can be obtained using the well known Direct Linear Transform (DLT) algorithm [3].

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. We have seen that establishing opti-acoustic correspondences to calibrate a multibeam and a camera system is a nearly-impossible task due to the different resolution of both sensors and the noise of the acoustic data.

Hence we must seek for a method that does not rely on explicit opti-acoustic correspondences. Since we assume a simplified multibeam model, our problem might be considered similar to a calibration of a camera-laser system. Although in most of the cases laser points or stripes are visible to the camera, Zhang and Pless [8] proposed a calibration method to calibrate a camera and a invisible laser range finder. Their method is based on observing a planar pattern from several poses and establishing some 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 6.

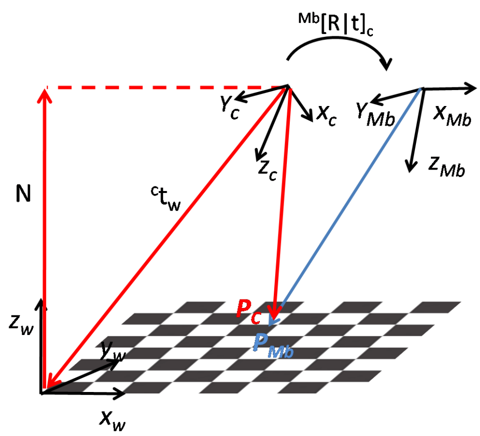


Fig. 6. Calibration method geometry

A planar calibration plane (i.e a checkerboard) must be placed in front of the camera-multibeam system. Since the method implies to locate the plane at different positions and orientations with respect to the sensors it is easier to place it on the floor of the water tank and move the rigidly attached sonar-video system from outside of the water tank. In this way, we avoid the presence of a diver performing the calibration movements underwater, which will cause additional noise and sources of error. 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 laying on the plane from the other points that impact with the water tank walls.

The method assumes that the camera has been already calibrated. Since the rotation and translation matrixes from world to camera frame (${}^{c}[\mathbf{R}|t]_{w}$) 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:

$$N = R_3(R_3^T(-t)) \tag{3}$$

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:

$$P_c = {}^{mb}\mathbf{R}_c^{-1}(P_{mb} - {}^{mb}t_c) \tag{4}$$

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

$$N \cdot {}^{mb}\mathbf{R}_c^{-1}(P_{mb} - {}^{mb}t_c) = \|N\|^2 \tag{5}$$

Equation 5 gives a constraint on ${}^{mb}R_c$ and ${}^{mb}t_c$ for each measured calibration plane parameters N and multibeam point P_{mb}. After some algebraic manipulations the constraints given by several measurements can be reshaped into a equation system. Finally, solving this system we are able to retrieve the ${}^{mb}[\mathbf{R}|t]_c$ transformation that relates the coordinate systems of the camera and the multibeam.

Some simulations have been performed in MATLAB to validate the method using the geometrical modeling of the sensors described in section II-B. To simulate the calibration procedure, we placed the calibration plane without loss of generality at the Z=0 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 ${}^{mb}[\mathbf{R}|t]_c$. This process is repeated along a trajectory defined over the calibration plane.

After the simulated acquisition of all the points, we solve the equation system 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, we must take a lot of points with several orientations in order to avoid degenerate cases.

In order to simulate more real conditions, additive white Gaussian noise has been introduced to the multibeam points simulating an accuracy of 5cm. In that case, Zhang and Pless propose to use the resulting matrices as an initial guess for a non-linear optimization problem which minimizes the Euclidean distances from multibeam points to the checkerboard planes. After this minimization, solved by Lavenberg-Marquardt algorithm in MATLAB, the results achieved show good tolerance to the introduced noise.

IV. CONCLUSIONS AND FUTURE WORK

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 simulations show that the method can robustly recover the position and orientation that relates the system's sensors.

The immediate future task is to test the proposed calibration approach in real conditions and perform a simple 3D reconstruction of real data. After the evaluation of the results obtained in the previous steps, a more complex 3D reconstruction approach should be devised. The most suitable option seems to integrate the opti-acoustic system into a visual SLAM framework. First of all we want to analyze how to build robust features from sonar and visual data. Then, an appropriate SLAM algorithm should be designed in order to enforce consistency within navigation data and the detected features to yield superior mapping results.

REFERENCES

[1] RM Eustice, O. Pizarro, and H. Singh. Visually Augmented Navigation for Autonomous Underwater Vehicles. *IEEE Journal of Oceanic Engineering*, 33(2):103–122, 2008.

[2] A. Fusiello and V. Murino. Augmented scene modeling and visualization by optical and acoustic sensor integration. *IEEE Transactions on Visualization and Computer Graphics*, 10(6):625–635, 2004.

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

[4] M. Johnson-Roberson, O. Pizarro, and S. Williams. Towards Three-Dimensional Heterogeneous Imaging Sensor Correspondence and Registration for Visualization. *OCEANS 2007-Europe*, pages 1–6, 2007.

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

[6] 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*, pages 206–211, 2000.

[7] S. Williams and I. Mahon. Simultaneous localisation and mapping on the great barrier reef. In *2004 IEEE Proceedings Int. Conf. on Robotics and Automation, 2004. ICRA'04*, volume 2, 2004.

[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*, volume 3.