OPTO-ACOUSTIC DATA FUSION FOR SUPPORTING THE GUIDANCE OF REMOTELY OPERATED UNDERWATER VEHICLES (ROVs)

F. Bruno[1], A. Lagudi[1], G. Ritacco[1], M. Muzzupappa[1], R. Guida[2]

[1] Dept. of Mechanical, Energetics and Management Engineering, Università della Calabria, Rende (CS), Italy –
(fabio.bruno, antonio.lagudi, gerardo.ritacco, maurizio.muzzupappa)@unical.it
[2] Whitehead Sistemi Subacquei S.p.A., Pozzuoli (NA), Italy – ramona.guida@wass.it

Commission V

KEY WORDS: 3D opto-acoustic camera, optical and acoustic data fusion, stereovision system, system calibration, ROV guidance

ABSTRACT:

Remotely Operated underwater Vehicles (ROVs) play an important role in a number of operations conducted in shallow and deep water (e.g.: exploration, survey, intervention, etc.), in several application fields like marine science, offshore construction, and underwater archeology. ROVs are usually equipped with different imaging devices, both optical and acoustic. Optical sensors are able to generate better images in close range and clear water conditions, while acoustic systems are usually employed in long range acquisitions and do not suffer from the presence of turbidity, a well-known cause of coarser resolution and harder data extraction.

In this work we describe the preliminary steps in the development of an opto-acoustic camera able to provide an on-line 3D reconstruction of the acquired scene. Taking full advantage of the benefits arising from the opto-acoustic data fusion techniques, the system was conceived as a support tool for ROV operators during the navigation in turbid waters, or in operations conducted by means of mechanical manipulators.

The paper presents an overview of the device, an *ad-hoc* methodology for the extrinsic calibration of the system and a custom software developed to control the opto-acoustic camera and supply the operator with visual information.

1. INTRODUCTION

During the last few years, there has been growing interest for the development of efficient methodologies and systems in the underwater research area, in order to deal with the challenging problems of monitoring, survey and data gathering.

In the field of cultural heritage, the scientific community and the international Cultural Heritage safeguarding bodies have established the need to promote, protect, and preserve, possibly *in-situ*, the submerged artefacts and sites.

The CoMAS project, started in 2011, aims to develop new materials and technologies for supporting the restoration, conservation and documentation of underwater cultural heritage. The project goal is the definition of a conservation methodology that includes several stages: documentation, cleaning, restoration, maintenance and monitoring.

One of the most challenging aspects of this project is the set-up of a special ROV (Remotely Operated Vehicle) devoted to the monitoring, routine cleaning (through the use of a custom mechanical manipulator), and 3D mapping of the submerged archeological structures. These specific tasks require the accurate localization of the vehicle within the operational environment and, above all, a clear representation of the underwater scene in presence of low visibility conditions.

ROVs are usually equipped with different imaging devices both optical and acoustic. The acoustic systems typically give good results in long-range acquisition and do not suffer from the water turbidity, but the resulting data are affected by low resolution and accuracy. The optical systems, in contrast, are more suited for close-range acquisitions and allow for gathering high-resolution data and target details, but the results are constrained by a limited visibility range.

Hence, the fusion of data captured by these two types of systems stands as a promising technique in underwater applications, as it allows for compensating their respective limitations. Since the two categories of sensors are based on different physical principles, they provide, in general, different information of the scene to be acquired, and different methods are employed to process the data. Therefore, the integration of the two types of data gathered from these sensors is a very promising and interesting field that calls for new solutions.

Despite the difficulty of combining two modalities that operate at different resolutions, technological innovation and advances in acoustic sensors have progressively allowed for the generation of good-quality high-resolution data suitable for integration, and the related design of new techniques and systems for underwater scene reconstruction.

The aim of this work is to describe the preliminary steps in the development of an opto-acoustic camera able to provide on-line 3D reconstructions. Taking full advantage of the benefits arising from the opto-acoustic data fusion techniques, the system was conceived as a support tool for ROV operators during the navigation in turbid waters or in operation conducted by means of mechanical manipulators. The operator will be able to choose the most suitable visualization technique (optical, acoustic) according to the working conditions or, if needed, to fuse both of them in a single image where the missing parts of the acquired optical data are covered by the acoustic acquisitions.

The remainder of this paper is organized as follows: Section 2 presents the state-of-the-art concerning the integration of optical and acoustic sensors in the field of underwater applications, and the solutions adopted to fuse the data. Section 3 deals with the system configuration and the hardware specifications of the sensors that constitute the opto-acoustic camera. In Section 4 we formalize the extrinsic calibration problem of the system and provide a methodology to solve it. The conclusive section presents the main features and a first prototype of the user interface.

2. RELATED WORKS

Multisensor data fusion is a technology aimed to enable the combination of information coming from several sources, in order to form a unified picture. It represents a research field which has been extensively covered in the scientific

doi:10.5194/isprsarchives-XL-5-W5-47-2015

community, so that data fusion systems are now widely used in various areas such as sensor networks, robotics, video and image processing (Khaleghi et al., 2013).

In the field of underwater applications, one of the first efforts to merge optical and acoustic sensors is reported in (Singh et al., 2000). In this work, simple finite element warps may be used to merge optical photomosaics with microbathymetric maps (acquired through a pencil beam sonar), obtaining high-resolution quantitative opto-acoustic maps of the seafloor. The fusion is achieved by manually identifying corresponding points across the two representations and then, using a finite element warp, by warping the photomosaic onto the new mesh defined by the same corresponding points.

A system for automatically building 3D maps of underwater terrain through the fusion of visual data acquired by a single camera and range data from a multibeam sonar is presented in (Kunz and Singh, 2013). The system uses pose graph optimization, square root information smoothing and mapping framework ($\sqrt{\mathrm{SAM}}$) to simultaneously solve for the ROV's trajectory, the map, and the camera location in the ROV's frame. Matched visual features are treated within the pose graph as images of 3D landmarks, while multibeam bathymetry submap matches are used to impose the relative pose constraints by linking ROV poses from distinct tracklines of the dive trajectory. The system is initialized with a graph that captures the ROV's trajectory, as measured by the onboard navigation sensors, and then adds factor nodes representing visual feature observations and constraints induced by multibeam submap matches, in an incremental fashion. In this solution, the calibration of the navigation sensors and the extrinsic camera position and orientation is enabled through the use of specialized nodes in the pose graph itself.

A further example of tight integration between cameras and acoustic sensors is presented in (Fusiello and Murino, 2004), a work focused on local area imaging rather than the creation of large area maps. In this work, the integration of optical and acoustic sensors is proposed in order to improve the understanding of the underwater scenes and assist ROV pilots during the navigation. The integration of video and 3D data obtained from a single optical camera and a 3D acoustic camera (Coda Echoscope) is obtained by geometrically registering such data with respect to a well-known model of the observed scene. After some pre-processing steps, the acoustic data are registered with respect to a known model of the object acquired (an oil rig in this case) using the Iterative Closest Point (ICP) algorithm (Besl et al., 1992), while the optical alignment is performed by means of an algorithm developed by Lowe (Lowe, 1991).Once the poses of both sensors are calculated with respect to the observed object, the relative pose between the optical camera and the acoustic camera can be estimated without the need to use positioning and motion system equipment.

Although very interesting, the presented solution has the drawback of relying on a CAD model of the object acquired for the calibration of the system, which may not be available, thus linking the presented methodology to a specific application domain and a specific class of objects.

Finally, we have to mention a new paradigm of opto-acoustic stereo reconstruction, proposed in (Negahdaripour et al., 2009), which aims to apply the epipolar geometry to a stereo system composed by an optical camera and a 2D sonar (DIDSON), in order to improve the accuracy of 3D reconstruction in turbid waters. Due to the asymmetrical projection models for an opto-acoustic system, the epipolar geometry, in this case, differs in the two views. The derived equations show that for each point in the acoustic image, the relative correspondence in the optical image lies on a conic section. Moreover, the matching of each point in the optical image lies on specific trigonometric curves.

The extrinsic calibration of the optical and acoustic camera has been performed by using images of a planar grid (target), characterized by relevant optical acoustic features that have been manually associated. Therefore, the relative positions of the cameras were estimated through an optimization algorithm that minimizes the distances between 3D reconstructions of optical and acoustic matching projections.

3. SYSTEM DESCRIPTION

In this section we explain the setup of the proposed system.

First we describe the system configuration and then the hardware specifications of the sensors that constitute the opto-acoustic camera, i.e. the 3D acoustic camera and the stereo optical camera, respectively.

3.1 System Configuration

Our proposal aims at realizing an opto-acoustic camera capable of providing an on-line optical and acoustic 3D reconstruction of the scene to be acquired.

The opportunities offered by this kind of system are endless. In a typical application scenario, it could be used as a support tool by a ROV operator during the navigation in turbid waters, or in exploration / maintenance operations conducted by means of underwater manipulators. The operator, during the various stages of the intervention, may determine which one of the two display modes would be more suitable for a specific operation to be performed, or may also decide to display them simultaneously to represent both data types by means of data fusion techniques.

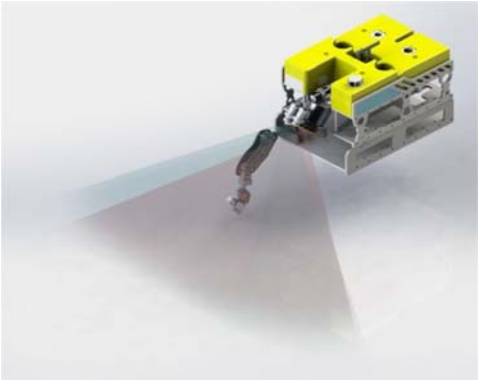

Figure 1. Scheme of the system's proposed configuration with Field-Of-Views of the optical cameras in stereo configuration (green) and of the 3D acoustic camera (red).

For which concerns the optical component of the system, a stereo camera, consisting of 2 ultra-compact digital cameras housed in custom-made waterproof cases, has been developed in the laboratories of DIMEG (Department of Mechanical, Energetic and Management Engineering), University of Calabria. Instead, the 3D acoustic camera is a custom system developed *ad-hoc* for the CoMAS project at the laboratories of Whitehead Sistemi Subacquei S.p.A. in Pozzuoli (Naples). The relative position of the stereo optical subsystem with respect to the acoustic one was defined during the planning stage of the ROV set-up in order to ensure the maximum overlap of both Field-Of-Views (FOVs) (Figure 1).

3.2 3D acoustic camera

The 3D acoustic camera is a custom system that provides high angular resolution images of the acoustic environment (figure 2).

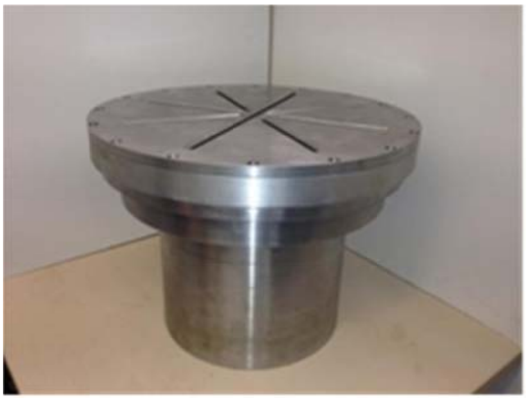

Figure 2. 3D acoustic camera.

It consists of two distinct subsystems: one containing the acoustic units for transmitting and receiving signals (TX / RX unit), and another one for processing the signals to be used in the beamforming process.

The transmission section of the acoustic unit is a logarithmically spaced linear array made of 40 analog channels, while the receiving section is made of 256 channels arranged in 4 logarithmically spaced linear arrays.

The acoustic unit generates beams which focus the energy in suitable directions, and acquires the backscattered signals coming from a target in the focused region.

The beamforming section processes the incoming signals at the digital level through beam-forming and beam-power evaluation algorithms in order to detect the polar coordinates of the acoustic echos (range and angle).

The key technical specifications of the system are presented in table 1.

Frequency	700 KHz
Number of beams	41x84x61 total
Maximum range	Focusing distance + 30 cm
Minimum range	Focusing distance - 30 cm
Update Rate	Up to 1 Hz
Beam spacing	1°
Angular coverage	60° x 40°
Dimensions (max diam., min diam., l)	410 mm, 260 mm, 287 mm
Weight	27 Kg (air), 5Kg (water)
Supply voltage	12 VDC
Power consumption	200 W ca.
Interface	Gigabit Ethernet

Table 1. Technical specifications of the 3D acoustic camera.

3.3 Stereo optical camera

The proposed stereo optical camera is the result of an intense research activity conducted at the DIMEG department of the University of Calabria in the field of the underwater stereo photogrammetry, both passive and active (Bruno et al., 2011; Bianco et al., 2013).

Figure 3. Stereo optical camera.

The system is composed of two cameras in stereo configuration, housed in custom-made waterproof cases (figure 3).

The digital cameras are two ultra-compact ImagingSource DFK 23G445 with a CCD (Charge Coupled Device) sensor size of 6.26 x 5.01 mm, a resolution of 1280 x 960 pixels and a frame rate of 30 fps. The devices are also equipped with a pair of Edmund Optics 1.8 mm wide-angle lenses.

Since the two digital cameras were not thought for their use in underwater environment, we have designed and constructed two waterproof housings. The body is made in aluminum in order to ensure efficient heat dissipation, while the flat port of the camera housing (dome port) is made in polycarbonate. This solution leads to a reduction of the angle of view caused by the refraction of the air-water interface, but its construction is easier. The camera is fixed within the case through an appropriate support, that also works as heat sink.

The layout of the stereo configuration has been designed to guarantee the maximum FOV with a minimal working distance. In table 2 we have reported the main technical specifications of the device.

Working distance (min.)	1 m (baseline 105 mm)
FOV (water) (diag.)	70°
Dim. point cloud	200.000 ca
Dimensions (w x h x l)	210 x 100 x 207 mm
Weight	2.6 Kg (air), 1.2 Kg (water)
Trigger	Yes
Supply voltage	11 – 13 VDC
Power consumption	4.8 W ca.
Interface	Gigabit Ethernet

Table 2. Technical specifications of the stereo optical camera.

This contribution has been peer-reviewed.
doi:10.5194/isprsarchives-XL-5-W5-47-2015

4. EXTRINSIC CALIBRATION OF THE SYSTEM

During the development of the proposed 3D opto-acoustic camera was the alignment of the optical and acoustic 3D data, defined as *sensor calibration problem*, which allows data from one sensor with the corresponding data of the other sensor.

Data alignment, that is, their transformation from each local frame into a common reference frame for both sensors, is of critical importance for the successful deployment of the opto-acoustic fusion system, since the 3D data generated by such sensors are so different from each other that they should be precisely matched in order to detect the spatial coordinates of a given point belonging to an object in both representations.

Up to now, the works presented in literature concerning the integration between several types of sonar (single beam sounder, multibeam, 3D acoustic camera) and optical cameras adopt a sensor fusion approach, which is mapping-oriented, according to the classification proposed in (Nicosevici et al., 2008). This means that the data acquired from the two sensors are described through geometric relationships (position and orientation), and the data fusion is performed by means of geometrical correspondences and registration.

The alignment of 3D data is usually solved by performing the extrinsic calibration of the integrated system, i.e. by searching for the fixed - but unknown - rigid transformation which relates the local reference frame of the stereo optical subsystem with that of the acoustic camera, thereby obtaining the relative pose (position and orientation) between the two sensors.

Now, let us assume that the two cameras have already been calibrated independently. In particular, with regard to the stereo optical system, the assumption is that both the intrinsic parameters of each optical camera (focal length, principal point, optical distortions and skew) and those of the extrinsic stereo pair (relative position between the two reference frames) have been determined. Downstream of this calibration process, it is possible to know the 3D coordinates of any point of the scene acquired by the optical subsystem with respect to its local reference frame (typically attached to the left camera of the stereo pair).

Therefore, assuming that a point $p_o = [x_o, y_o, z_o]^T$of the optical reference frame corresponds to a point $p_a = [\mathrm{x_a}, \mathrm{y_a}, \mathrm{z_a}]^T$of the acoustic reference frame, the rigid transformation that relates the two coordinate systems may be expressed as:

$$p_o = \boldsymbol{R} p_a + t \tag{1}$$

where $\boldsymbol{R}$ = 3 x 3 orthonormal rotation matrix from the acoustic camera to the stereo optical camera reference frame

t = 3D translation vector from the acoustic camera to the stereo optical camera reference frame

Equivalently, indicating with $\widetilde{p_o}$ and $\widetilde{p_a}$ the homogeneous coordinates of the points p_o and p_a respectively, the aforementioned relation can be expressed as:

$$\widetilde{\mathrm{p_o}} = {}^{\mathrm{o}}\begin{bmatrix}\mathbf{R} & \mathrm{t} \\ 0 & 1\end{bmatrix}_{\mathrm{a}} \widetilde{\mathrm{p_a}} \tag{2}$$

where ${}^{o}\begin{bmatrix}\boldsymbol{R} & t \\ 0 & 1\end{bmatrix}_a = {}^{o}\mathbf{T}_a$ = 4 x 4 homogeneous transformation matrix from the acoustic camera to the stereo optical camera reference frame

Therefore, our goal is to develop a methodology for the calculation of the extrinsic parameters $\boldsymbol{R}$ and t, which define the orientation and the position of the acoustic subsystem against the optical one, respectively.

Usually, the calibration methods are highly dependent on the sensors that compose the system and especially on the type of data they provide. In our case, the particular nature of the data acquired by the systems (3D point clouds) makes it possible to implement a methodology for extrinsic calibration based on a "direct" computation of the rigid transformation matrix that relates the reference systems associated with the optical and acoustic sensors. This is achieved through a simple registration of each pair of optical and acoustic 3D point clouds of a planar pattern, which is used as a target in the calibration procedure and acquired in different poses. Finally, the ability to obtain multiple estimates of the transformation matrix allows for implementing an appropriate optimization technique, in order to obtain more accurate results.

The implemented methodology is composed of two separate data-processing threads that are related to the two sensory channels, which eventually merge in the last stages of the proposed solution.

Starting from the synchronous acquisition of the *n* poses of the calibration panel during the early stage of the process, the entire methodology is aimed to obtain *n* pairs of 3D point clouds, where the *n*-th pair is formed by the optical $\boldsymbol{P}_{o,n}$ and the acoustic $\boldsymbol{P}_{a,n}$ 3D point clouds, in order to calculate, by means of coarse and fine registration algorithms, *n* estimates ${}^{\mathrm{o}}\mathbf{T}_{a,n}$ of the rigid transformation matrix. At the end of the process, the final transformation matrix ${}^{o}\mathbf{T}_a^*$ is obtained by processing the dataset composed of *n* transformation matrices ${}^{o}\mathbf{T}_{a,n}$ obtained downstream of the previous registration stage.

4.1 Acoustic image processing

The 3D image provided by the acoustic camera can be corrupted either by false reflections caused by the secondary lobes of the receiving array or by the noise present in the acquisition phase of the backscattering signals. The latter is modelled as speckle noise. The secondary lobes are responsible for the blurring of the object, while the speckle noise causes a low response or no response at all within the object itself.

So it is evident that the operations of filtering (noise reduction and the elimination of possible outliers) and segmentation (differentiation of objects and background in the observed scene) are to be considered as preliminary and mandatory steps for the execution of all fusion algorithms to be applied to this specific type of data (Murino et al., 2000a).

While in literature there are a number of algorithms for filtering and segmentation with variable results and degrees of automation (Murino, 2001b), the solution adopted in this calibration method for the processing of the acoustic 3D point clouds representing the calibration panel in its different poses, provides for a completely manual filtering and segmentation procedure, performed through the open source software CloudCompare (CloudCompare, 2014), as the implementation of an automated procedure would require further, more focused research.

4.2 Optical stereo images processing

This subsection describes the procedures for the creation of a 3D point cloud representing the optical calibration panel in its different poses, starting from the acquisition of a pair of stereo images.

4.2.1 Image enhancement Underwater images are generally affected by degradations caused by the attenuation of light during its propagation in water (mostly due to absorption and scattering), such as low contrast, uneven lighting, blurring, limited visibility range, color attenuation (particularly in the red channel) and noise. All these issues have a strong impact on the quality of the results obtained in the application of any 3D reconstruction technique (Gallo, 2013).

Regarding the proposed calibration method, the adopted solution makes use of a methodology of image enhancement based on the technique of Histogram Stretching (Gonzalez et al., 1992) and on a subsequent manual color retouching procedure (HIST).

4.2.2 Stereo optical calibration A calibration procedure is needed to compute the intrinsic parameters of each camera (focal length, coordinates of the principal points, radial and tangential distortions, pixel size) and the extrinsic parameters (translation and rotation with respect to a world coordinate system) of the camera and the stereo system. These parameters are computed using the well-known Camera Calibration Toolbox for Matlab (Bouguet, 2013).

4.2.3 Checkerboard 3D reconstruction Downstream of the calibration process, the acquired images were processed through well-known techniques of stereo photogrammetry in order to obtain the 3D point clouds of the calibration panel for each pose assumed during the acquisition stage.

The 3D reconstruction was carried out using the OpenCV (OpenCV 2014) and Libelas (Geiger et al., 2011) libraries. In particular, the OpenCV library was used for the rectification of the stereo pair, while Libelas was used for the implementation of the stereo matching algorithm and the generation of the disparity map. Finally, the 3D point cloud was determined as the product of the reprojection matrix obtained downstream of the calibration process and the disparity map itself.

4.1 Coarse and fine registration

This subsection describes the steps of the calibration methodology that allows for obtaining the *n* estimates ${}^{o}\mathbf{T}_{a,n}$ of the unknown rigid transformation matrix.

Considering the *n*-th pair $\boldsymbol{P}_{o,n}\boldsymbol{P}_{a,n}$ of acoustic and optical 3D point clouds, the associated ${}^{o}\mathbf{T}_{a,n}$ is determined as a composition of transformations obtained through coarse and fine registration algorithms.

Taking as a reference system the local reference frame of the stereo optical camera, the coarse registration stage was carried out through two operations:

1. Calculation of ${}^{1}\mathbf{T}_{a,n}$ by means of a manual orientation of the acoustic camera local reference frame, in such a way that the Z axis represents the depth of the scene, in line with the optical system;

2. Alignment of the pair of 3D point clouds $\boldsymbol{P}_{o,n}\boldsymbol{P}_{a,n}$ through an automatic algorithm implemented in CloudCompare that, from an estimate of the center of gravity of the two 3D point clouds, determines the translation vector *t* that relates them (assuming that the rotation matrix $\boldsymbol{R}$ is unitary).

   As a result of this operation, we obtain the transformation matrix ${}^{2}\mathbf{T}_{1,n}$.

For what concerns the step of fine registration, an implementation in CloudCompare of the well-known Iterative Closest Point algorithm (ICP) has been applied to the pair $\boldsymbol{P}_{o,n}\boldsymbol{P}_{a,n}$ aligned in the previous step, in order to obtain the transformation matrix ${}^{o}\mathbf{T}_{2,n}$.

Downstream of the previous operations, the unknown rigid transformation matrix ${}^{o}\mathbf{T}_{a,n}$ is obtained as (figure 4):

$$ {}^{o}\mathbf{T}_{a,n} = {}^{o}\mathbf{T}_{2,n} \times {}^{2}\mathbf{T}_{1,n} \times {}^{1}\mathbf{T}_{a,n} \qquad (3) $$

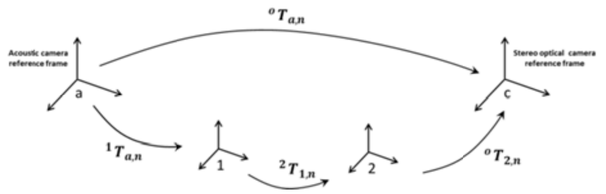

Figure 4. Sequence of transformations for computing the ${}^{o}\mathbf{T}_{a,n}$ matrix.

4.2 Statistical processing of the transformation matrices

A method of statistical processing has been implemented in order to estimate the rigid transformation matrix ${}^{o}\mathbf{T}_{a}^{*}$ from the matrices ${}^{o}\mathbf{T}_{a,n}$ obtained downstream of the operation of coarse and fine registration applied to pairs $\boldsymbol{P}_{0,n}\boldsymbol{P}_{a,n}$ of 3D point clouds.

The proposed solution is based on an algorithm implemented in MATLAB, which automatically determines the final transformation matrix ${}^{o}\mathbf{T}_{a}^{*}$, by selecting it from the ${}^{o}\mathbf{T}_{a,n}$ matrices included in the dataset.

The algorithm operates in the following way:

1. for each ${}^{o}\mathbf{T}_{a,n}$, it applies this transformation to the $\boldsymbol{P}_{a,n}$ 3D acoustic point clouds, in order to align them with the corresponding 3D optical point clouds $\boldsymbol{P}_{o,n}$;

2. for each pair $\boldsymbol{P}_{o,n}\boldsymbol{P}_{a,n}$, the mean distance $d_{oa,n}$ between the points of the optical 3D cloud and the corresponding points of the acoustic 3D cloud is calculated;

3. calculates $\overline{d_n}$ as the mean of $\boldsymbol{d}_{oa,n}$;

4. selects $\overline{d_{n,min}}$ as the minimum value of $\overline{d_n}$;

5. assumes as ${}^{o}\mathbf{T}_{a}^{*}$ the transformation ${}^{o}\mathbf{T}_{a,n}$ corresponding to $\overline{d_{n,min}}$.

5. CAMERA CONTROL SOFTWARE

Designed with the aim of integrating the most recent data fusion techniques currently available in scientific literature, the proposed system is able to provide an on-line opto-acoustic 3D reconstruction of the acquired scene.

Concerning the acoustic subsystem, it provides a high angular resolution image by working in two different modalities:

1. *Detection mode*: in this modality the system performs an isotropic scan of the scene in order to determine the coarse position of a target;

2. *Classification mode*: after the target identification, the acoustic camera switches to this modality and performs a fine scan of its FOV (60˚ x 40˚). The classification mode is automatically activated when, in the previous modality, the system detects a target at a range less than 2 m or manually, by setting the focus distance.

On the other hand, the stereo optical subsystem allows, for visualizing the acquired scene through a digital camera of the stereo pair, and also to provide an on-line 3D reconstruction using different stereo matching algorithms.

A custom software is being developed for supporting the control of the image devices and for providing ROV operators with an easily interpretable representation of the visual information.

It was conceived and designed with the aim of meeting the following requirements:

1. a customizable interface to manage and organize the visual information according to different user profiles and working conditions;

2. a high flexibility such that the operator is able to set, in real time, the main parameters of the opto-acoustic camera in order to cope with the variable conditions typical of an underwater environment.

A first prototype of the user interface is show in figure 5.

The system is able to switch between different visualization modes. In this way, 2D and 3D information can be displayed on one screen, or in suitable display windows in which the visual information is enhanced through additional data that may help the ROV operator during specific work activities.

For example, Figure 6 shows a display window in which the depth of the scene is coded through a RGB scalar field.

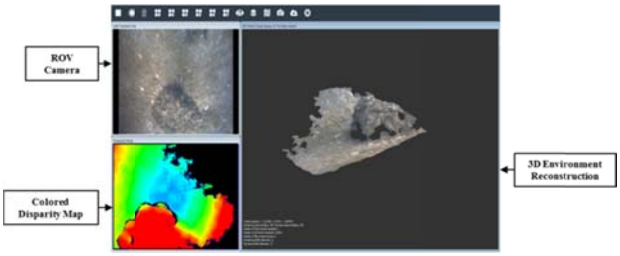

Figure 5. Prototype of the user interface and control window layout.

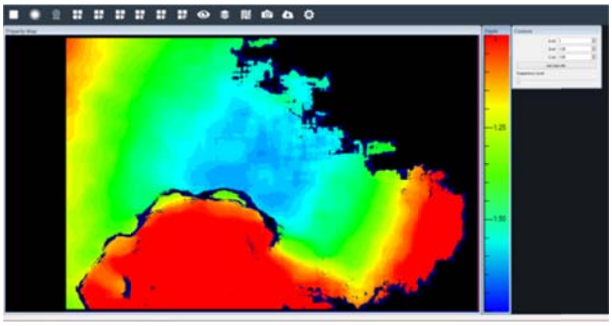

Figure 6. Display window with depth map coded through a RGB scalar field.

6. CONCLUSION

In this work, we have presented a first step towards the development of an opto-acoustic camera, which is able to provide an on-line 3D reconstruction of the explored underwater scene. The system is composed of a custom stereo optical camera and 3D acoustic camera, and is designed to be installed on a ROV in order to improve its control capabilities by the operator during the navigation in turbid water conditions or in close range manipulation operations. In order to perform a correct sensor fusion, an extrinsic calibration methodology, aimed to determine the rigid transformation matrix that relates the reference systems associated with the two sensors, has been proposed. The calibration was conducted by using a planar pattern as a target, which has been positioned in different poses in front of the opto-acoustic system. The two sets of 3D point clouds provided by the two cameras have been aligned through a two-step process based on coarse and fine registration algorithms and then the final matrix was obtained through a statistical processing.

We developed a prototype of the software that controls the opto-acoustic camera and is able to provide and manage the 2D and 3D visual information supplied by the system.

This software allowed us to perform initial tests on the system and verify its usability. It also provided the operator with an easily interpretable representation of the visual information.

7. ACKNOWLEDGEMENTS

This work has been partially supported by the Project ‘CoMAS’ (ref. PON01_02140), financed by the MIUR under the PON ’R&C’ 2007/2013 (D.D. Prot. n. 01/Ric. 18.1.2010).

REFERENCES

Besl, P. J., & McKay, N. D., 1992. Method for registration of 3-D shapes. In: *Robotics-DL tentative* (pp. 586-606). International Society for Optics and Photonics.

Bianco, G., Gallo, A., Bruno, F., Muzzupappa, M., 2013. A comparative analysis between active and passive techniques for underwater 3D reconstruction of close-range objects. Sensors (Switzerland), Vol. 13, N.8, pp. 11007-11031.

Bouguet. Webpage. http://www.vision.caltech.edu/bouguet (June 2014).

Bruno, F., Bianco, G., Muzzupappa, M., Barone, S., and Razionale, A. V., 2011. Experimentation of structured light and stereo vision for underwater 3D reconstruction. ISPRS Journal of Photogrammetry and Remote Sensing, 66(4), pp. 508-518.

CloudCompare. Webpage http://www.danielgm.net/cc/ (June 2014).

Fusiello, A., and Murino, V., 2004. Augmented scene modeling and visualization by optical and acoustic sensor integration. *Visualization and Computer Graphics, IEEE Transactions on*, 10(6), pp. 625-636.

Gallo, A., 2013. Innovative methodologies for multi-view 3D reconstruction of Cultural Heritage, PhD Thesis.

Geiger, A., Roser, M., and Urtasun, R., 2011. Efficient large-scale stereo matching. In: *Computer Vision–ACCV 2010*, pp. 25-38. Springer Berlin Heidelberg.

Gonzalez R.C., and Woods R.E., 1992. Digital Image Processing. Addison-Wesley, Reading, MA.

Khaleghi, B., Khamis, A., Karray, F. O., and Razavi, S. N., 2013. Multisensor data fusion: A review of the state-of-the-art. *Information Fusion*, 14(1), pp. 28-44.

Kunz, C., and Singh, H., 2013. Map building fusing acoustic and visual information using autonomous underwater vehicles. *Journal of Field Robotics*, 30(5), pp. 763-783.

Lowe, D. G., 1991. Fitting parameterized three-dimensional models to images. *IEEE transactions on pattern analysis and machine intelligence*, 13(5), pp. 441-450

Murino, V., and Trucco, A., 2000a. Three-dimensional image generation and processing in underwater acoustic vision. *Proceedings of the IEEE*, 88(12), pp. 1903-1948.

Murino, V., 2001b. Reconstruction and segmentation of underwater acoustic images combining confidence information in MRF models. *Pattern Recognition*, 34(5), pp. 981-997.

Negahdaripour, S., Sekkati, H., and Pirsiavash, H., 2009. Opti-acoustic stereo imaging: On system calibration and 3-D target reconstruction. *Image Processing, IEEE Transactions on*, 18(6), pp. 1203-1214.

Nicosevici, T., & Garcia, R., 2008. Online robust 3D mapping using structure from motion cues. In: *OCEANS 2008-MTS/IEEE Kobe Techno-Ocean*, pp. 1-7. IEEE.

OpenCV. Webpage. http://opencv.org/ (June 2014)

Singh, H., Roman, C., Whitcomb, L., and Yoerger, D., 2000. Advances in fusion of high resolution underwater optical and acoustic data. In: *Underwater Technology, 2000. UT 00. Proceedings of the 2000 International Symposium on*, pp. 206-211. IEEE.