MULTI-SENSOR DATA FUSION FOR UNDERWATER ARCHAEOLOGICAL INVESTIGATION

Davide Moroni[a], Maria Antonietta Pascali[b], Marco Reggiannini[c], Ovidio Salvetti[d]

[a]davide.moroni@isti.cnr.it
[b]maria.antonietta.pascali@isti.cnr.it
[c]marco.reggiannini@isti.cnr.it
[d]ovidio.salvetti@isti.cnr.it

For all the authors:
Postal address: SI-Lab, ISTI (CNR), Via G. Moruzzi, 1, 56124 Pisa, Italy.
Fax number: +39 050 315 2810.

Abstract: *The primary purpose of this work consists in treating optical and acoustic signals in order to extract useful information for applications in underwater archaeology. Data are processed to assess the presence of geometrically regular elements, potentially indicating handmade objects lying on the seafloor. Geometrical elements are recognized by means of suitable algorithms and their statistical persistence in the data stream is employed as a descriptor. Multi-sensor data are processed by applying segmentation and classification procedures based on a geometrical pattern analysis, with the purpose of discerning different materials. We basically seek for meaningful features in the data in order to perform robust object recognition, also in case of unfavorable environmental conditions. Finally we define a unique data fusion model that can be exploited for exhaustive interpretation of the underwater scene.*

Keywords: *Cultural Heritage Safeguard, Autonomous Underwater Vehicles, Automatic Vision System, Side Scan Sonar, Geometrical Features Recognition, Image Segmentation, Image Classification, Multi-Sensor Data Integration*

1. INTRODUCTION

Among several areas of interest related to the underwater environment, the safeguard of cultural heritage is certainly undergoing significant consideration and development worldwide. The marine environment represents a challenging context for IT experts, both for what concerns the discovery of unknown sites and for the recovery and preservation operations. The hostile environmental conditions, unfit to human intervention, favored the increasing demand for artificial intelligence integration in Autonomous Underwater Vehicles (AUVs) in charge of survey operations. AUVs are commonly equipped with a set of sensors in order to acquire different types of signals from the surrounding environment. This multi-sensor system returns data that can be processed and combined for a reliable analysis of the environment. Actually there exist many different techniques (e.g. [1], [2]) for seafloor survey and underwater object detection. The multiple possible choices in terms of devices employed or environment settings (deep or shallow water, etc.), and the difficulty in each validation procedure have produced a wide family of techniques, but no settled standard (e.g. [3], [4]).

Our basic purpose is the full understanding of underwater scenes acquired by AUVs through the integration of acoustic and optical data. This kind of activity is based on collecting and processing data in two ways: *i) off-line*, in order to process in detail large amounts of data and *ii) on-board*, applying fast and efficient algorithms in order to automatize the survey operation, free the AUV intelligence unit from the operator intervention and send near real-time alerts to support the mission re-planning task. The idea of combining information from multiple sensors is not new and, since the applications of data fusion are disparate, it is quite difficult to build a *one-fits-all* framework; in particular underwater application is still an open problem. In our work we used a side-scan sonar and a stereoscopic vision system composed of two analog underwater cameras. Sonar and vision devices operate on different physical principles, provide different types of information and generally run at best in different conditions. The raw data captured by our multi-sensor system are videos from the optical device and chronologically structured 2D maps from the side-scan sonar.
In this paper we present a new method to integrate optical and acoustic data in order to collect the whole information acquired in one map, easier to use in order to perform terrain classification, object detection and recognition, hence to better understand the investigated scene. The paper is structured as follows: the first three sections introduce the experimental conditions, describe the 2D and 3D acoustic and optical images processing. Then in Section 4 we present our fusion model. In the conclusive section we describe our preliminary results based on a test performed at the Elba Island.

2. EXPERIMENTAL SETTING AND PRE-FILTERING ISSUES

The survey tools that will make part of the AUV sensor equipment have to be chosen in order to cope with the specific requirements of the experimental scenario. As known, optical and acoustic sensors are the most widely employed devices for underwater mapping purposes. Our multi-sensor set up has been conceived to assure a complete analysis of the seafloor. This means that it must be able to correctly “sense” the environment from a large scale point of view, in order to identify areas of potential archaeological interest, as well as to capture sufficient small-scale detail in order to completely describe the scene. To this end we chose a *side-scan sonar*, a device providing large-scale grey-level maps of the seabed. The acoustic device is particularly useful for underwater mapping since the acoustic wave decay length is typically an order of magnitude greater than in the case of optical wave. Indeed acoustic distances for optimal mapping are usually larger than 10-15 meters. This framework is surely tailored for the best performance of the acoustic device but turns out to be unsatisfactory in terms of resolution properties. Presuming that the sonar is traveling at an altitude of 20 meters with the *maximum response axis* bearing at 45° below the horizontal line, a typical resolution of $25x35\ cm^2$ is obtained. As a consequence the sonar will not be able to carry out a thorough analysis of small objects lying on the seafloor (pottery, amphorae, dishes, etc.). In order to correctly detect and map a wide variety of targets a stereo vision system (coupled camera system) has to be installed on the vehicle.

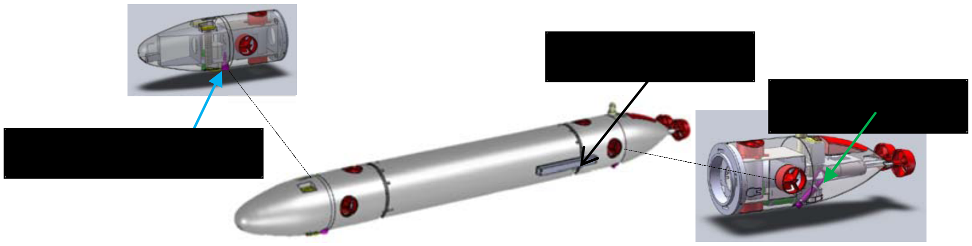

Figure 1: Sensors installation on the vehicle

In our particular experimental setting, the sensors will be properly placed on the same vehicle, with fixed coordinates with respect to the vehicle reference system. We assume that the data capture will be performed in an optimal way, i.e. with the vehicle moving at constant speed along its path, maintaining itself at a constant height with respect to the seafloor baseline. The devices placement is sketched in Figure 1. The data collected during underwater surveys are corrupted by many characteristic artefacts, which can be classified in two primary classes: the first refers to those undesired signal components that arise from the physics of the specific sensor device and that systematically affect every detection; the second instead is associated with the noise sources introduced by the surrounding environment. In the following some of the typical degradations involved in stereoscopic vision and side-scan sonar imagery are discussed; their removal requires preliminary corrections to allow further processing such as, in particular, by means of computer vision algorithms.

Vision Artefacts

- Optical image formation is affected by radial and tangential distortions due to the light propagation through the surrounding medium and the camera lens. This results in a distorted reproduction of the target geometry. This effect can be corrected by calibrating the optical stereoscopic system and by exploiting the estimated intrinsic and extrinsic parameters to rectify the images.
- Optical images captured in the underwater environment suffer from visibility degradation due to light attenuation, partial polarization and a hazing effect resulting from the light scattering inside the water medium. Suitable filtering techniques may restore the actual color properties of the recorded scene.

Side-scan Sonar Artefacts

- Back-scattered echo level suffers from propagation loss caused by multiple phenomena (*absorption, spherical divergence*). To work this problem out, analog data are usually processed by a Time Variable Gain (TVG) filter: this results into the enhancement of echoes coming from the furthest regions. Moreover the angular variability of the acoustic field in the vertical plane produces non-uniform ensonification of the seabed. That can be corrected by calculating the transducer directivity pattern and by multiplying the weakest back-scattered echoes by a proper amplification factor.
- Aiming at an accurate knowledge of the seafloor geometry, the piling up of successive scan strips is not sufficient: equidistant time samples do not correspond to equidistant range samples so a preliminary transform of the data (*Slant /Ground range*) must be performed.

The vehicle balance strongly depends on the surrounding environmental conditions (for example wave motion and currents) and that usually introduces significant distortions in the acquired sonograms. If the vehicle sensor equipment includes devices for navigation and motion control, this effect may be held down by estimating the sensor attitude and by referring the scan lines to a common reference system.

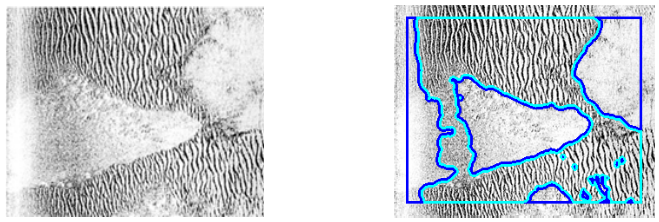

Figure 2: Side Scan Sonar segmentation example
(Imagery From A Klein Associates, Inc., 500 KHz Side Scan Sonar)

In the following sections we will assume that all the required preprocessing operations have been already carried out, and that every kind of noisy artefact or undesired signal component has been filtered out, leaving the optical and acoustic signals of interest.

3. 2D PROCESSING

In this section we suppose that the preliminary treatment of acoustical and optical data has been already successfully carried out.

Texture Segmentation/Classification: Texture analysis can be applied to optical and acoustic images in order to discriminate between different areas of the seabed. A procedure for texture segmentation based on Gabor filtering has been implemented: we performed the convolution of the image with a bank of Gabor filters having different frequency and orientation settings. Areas of the image exhibiting regular texture properties have different responses with relation to a particular choice of the Gabor filters settings. Varying these settings we have analyzed the filtered outputs, hence we have clustered similar regions by using the *K-means* algorithm. This processing stage produces an output image in which every area of the seabed has been identified as belonging to a certain class and has been properly marked by a specific color. That can be further interpreted by means of texture classification procedures, for instance aiming at detecting specific sediment categories (rock, sand, ripple, etc.) as discussed in [5] and [6]. This procedure has been applied to a side scan sonar picture showing different types of seabed regions (sand, rock, etc…): the result can be seen in Figure 2.

Figure 3: Curve detection from data captured at Elba Island

2D Analysis: This section is devoted to the 2D analysis of acoustic and optical data. In the current literature we observed a lack concerning the development of methods and algorithms specifically conceived for the real time detection of man-made and archaeological objects, through the analysis of their geometrical attributes. Our approach on geometrical curve detection is based on the assumption that a high concentration of regular curves represents a marker for the presence of manmade objects or shipwrecks. The technique that is employed as a method for assessing the wealth of regular geometric shapes in a video stream is described in [7]. It is an application of the ELSD algorithm, inspired by Gestalt Theory (see [8]), to man-made object detection in underwater environment. The preliminary results we obtained by applying our method on experimental acquisitions show a nice correlation between the detections performed by the algorithm and the ground-truth. The method we outline here, allows to give emphasis to objects in the scene that exhibit geometrical regularity features, in contrast with an unstructured and chaotic surrounding environment. In [7] this method has been applied to optical data; here its application is extended also to the case of acoustic data.

The ELSD detection algorithm is based on a three-stage process: the first stage (*candidate selection*), starts with the evaluation of the intensity gradients at every pixel, followed by a clustering step in which groups of pixels sharing similar orientation properties are gathered into rectangles (for lines) or chains of rectangles fulfilling proper curvature constraints (ellipses and circles). The sensitivity of the detection depends on some relevant internal parameters such as ρ (the smallest detectable gradient magnitude), τ (an angle tolerance which defines the angular region where pixels are to be considered aligned) and D (a lower threshold value on the density of aligned pixels inside a rectangle). In a second stage (*validation* stage), the candidates are further analyzed in order to decide whether they are meaningful structured groups of pixels or if they represent noisy clutter. This is an important step since it allows the rejection of false positives by comparison with an appropriate *detection threshold*, which is automatically computed by the algorithm. Indeed, the estimation process is based on the so called *Helmoltz perception principle*: it essentially states that there is no perception in white noise. In the third and final stage (*model selection*) the candidates are classified as belonging to a specific model (line, circle, ellipse) by considering the most suitable model as the one producing fewer false alarms.

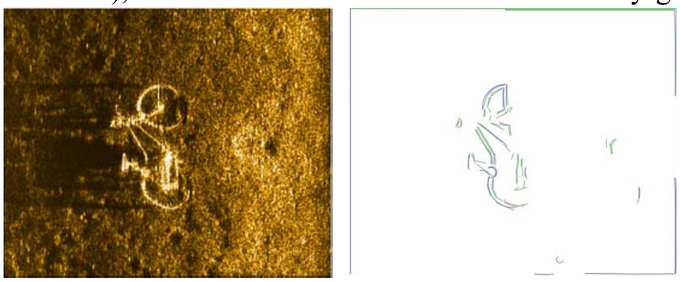

Figure 4: ELSD curve detection result on a sonar map
Imagery from jwfishers.com

Depending mainly on contrast, dimension, and lighting the performance can vary widely, but it typically ensures robust performances in case of Gaussian noise.

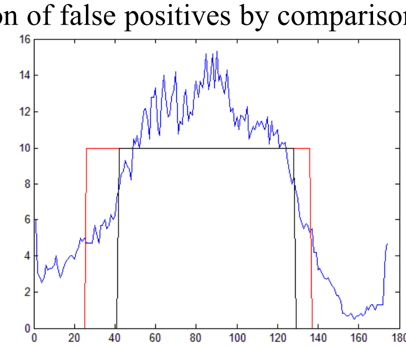

Figure 5: Detections (blue), ground-truth (red) and supra-threshold (w.r.t. mean) value (black)

On the basis of ELSD algorithm, an algorithm for curve detection specifically tailored for real-time operations has been implemented. More in detail we adjusted the ELSD detector in order to perform on-board data analysis during the AUV survey, and to transmit alert signals whenever a high density of geometrical shapes is detected. We consider the amount of regular curves and geometric patterns contained in a scene as an estimator of the probability of object discovery. We hence define a *finding event* as the circumstance in which the total number of the detected geometrical shapes exceeds a properly determined threshold. In order to overcome the typical randomness of environmental conditions, we propose the following processing pipeline that can be applied to both optical and acoustic data:

1. Raw data image acquisition and pre-processing;
2. ELSD for a sequence of N consecutive frames; fine tuning of internal parameters (e.g. ρ, τ, D, computation of the *discovery threshold* based on the weighted sum of curve detections);
3. ELSD for a sequence of M adjacent frames; reporting supra-threshold detection.

The *discovery threshold* is updated at every cycle restart since new surveyed sites may show new properties in terms of environmental conditions, specifically for what concerns the background noise level. We set {3; 2; 1} as weights for each type of sought curve {ellipse, circle, line}. This choice is motivated by the strong belief that elliptical and circular arcs are more meaningful (and rare) than line segments in the archaeological object we look for (e.g. amphorae, and plates, commonly found in the cargoes of ancient vessels). Then we compute the weighted average w_a over the detected curves for the test set of N frames and set the discovery threshold to w_a incremented of a 20%; this threshold is then applied to each frame (included the first N frames on which it is computed) and a warning is produced for each frame containing a supra-threshold number of detections. Even without data pre-processing, we experienced quite good performances of ELSD on underwater natural images (see Figure 3).

Our detection method has been tested on acoustic data too. Both optical and acoustic data have been processed in a similar way by extracting a subset of frames from the available data and by applying the algorithm according to the above described outline. Generally speaking, the ELSD performance can be badly affected by *granular noise* (like *speckle* or *salt-and-pepper*), which produces over-fragmentation and clutter in the images. Hence, before applying ELSD to sonograms, we should carefully smooth the image without losing detail. The image in Figure 4 has been retrieved from the internet and it has been pre-processed by a Gaussian blur in order to filter out the granular noise content. Then a sequence of frames has been created by selecting small consecutive and partially overlapping sections of the original sonar image. After having suitably set the internal parameters (ρ, τ, D) in order to increase the detection sensitivity, we finally applied ELSD to the obtained sequence of frames. As it can be seen in Figure 4, the algorithm was able to detect the salient features appearing in the sonar frames and extract the geometrical content carried by the acoustic data. Plots in Figure 5 show a good correlation between these three sets of data: detections, discovery threshold and ground-truth (which has been evaluated by manually labeling every frame with a degree of interest on a scale of four arbitrary levels).

4. **3D PROCESSING**

The captured data essentially result from a projection transform of 3D objects' profiles onto a 2D plane. Depending on the environmental conditions, sensor settings and morphology features univocal 2D maps are produced by the optical and acoustic devices. The primitive tridimensional properties of the objects can be estimated by processing the raw data with appropriate computer vision algorithms.

Figure 6: Virtual environment recreated by means of the estimated disparity map

<u>3D from stereo:</u> The optical data can be exploited to infer tridimensional properties of the scene by means of *stereo photogrammetry* reconstruction methods. We started from a simple geometrical modeling of the light projection on the camera system (*pinhole camera model*).

As known the stereo configuration of the cameras allows to recover the disparity map from every synchronized and rectified pair of frames. From the disparity map we can build a 3D model of what is visible in both frames.

In recent years, some improved methods have been developed to compute the 3D model from a stereoscopic pair of images, and more generally, from a possible large set of images of the same object or scene (see [9]). The main goal of our work is to speed up and combine existing methods to produce a real time efficient method able to analyze a pair of video streams and to produce a textured 3D scene. Nevertheless, for the purpose of image mosaicking and data fusion, it is sufficient to use as input selected pairs of frames from the video streams and apply known methods for the computation of the 3D structure. The resulting output can be further processed and rendered into a virtual environment, suitably designed to reproduce the underwater site exploration (see Figure 6).

3D from *acoustic shape from shading:* The 3D bathymetric profile of the seabed can be inferred from sonar maps using a *shape from shading* approach [10]. The intensity of a sonogram pixel is related to the backscattered echo intensity, which is in turn dependent from various environmental factors, e.g. the seafloor bathymetry *Z*, the transducer directivity pattern Φ and the reflectivity *R*. If we assume a particular model for the acoustic wave scattering (for example *Lambertian*) and the quantities involved in the model are known, then we can calculate the intensity of a specific pixel in the sonar map. In the actual experimental setting the physical parameters (*Z, R,* Φ) are not known; we only know the pixel intensity from the measurements. The *Shape From Shading* algorithm consists in inverting the problem of sonographic map formation starting from the above cited hypothesis on the acoustic back-scattering law and minimizing an appropriate cost function.

5. DATA FUSION

A detection procedure based on multi-modal data processing (optical and acoustic) can enforce the analysis of surveyed areas and improve the reliability of the archaeological mission. There exist many information levels at which we can fuse: Data, Feature, Decision. Respectively these are related to the different levels of representation, from low to high ones: Signal/Pixel, Feature, Symbol. Signal/pixel level fusion is the combination of the raw (or preprocessed) data from multiple source images into a single image; feature level fusion requires the extraction of different features from each source data; and decision level fusion combines the results from multiple algorithms to yield a final fused decision.
The most common fusion approaches used in the past were of two types: either at data level or at features level. In other words: first fusion then feature extraction, or first feature extraction then fusion. Especially in the application to underwater navigation, it seems to be preferred the first fusion method, because the seabed feature extraction (such as for points or lines) is less robust in the underwater environment. In our case instead, we have two data layers, in which each spatial point carries its optical and acoustic information, and both layers are referred to the same reference system. For this reason, we decide to combine the layers and project all sensor information into a common multidimensional state-space map. In [11] we propose the integration between many layers, differently processed, of the multidimensional state-space map. The purpose of this integration is to obtain a data set on which one can apply specific algorithms, for instance for the detection of a relevant finding. More in detail, we assign to each point $p \in \mathbb{R}^2$ in a mosaic a vector $\boldsymbol{v}(p)$ whose values are grouped (and preliminary limited to) as follows:

1. INTENSITY: RGB values, and acoustic echo intensity
2. ALTITUDE: elevation (*z* coordinate) estimated by sensor measurements, or inferred from acoustic *shape from shading* or the *depth map* resulting from the stereo image analysis
3. SURFACE: optical and acoustic texture classification
4. GEOMETRY: affinity to a specific curve family (counted with multiplicity)

Once the investigated underwater scene has been represented as above described, specific algorithms can be implemented. For instance:

a) geometric pattern detection on both optical and acoustic maps followed by a comparison procedure to recognize common characteristics.

b) components selection from the vector $\boldsymbol{v}(p)$ and processing of the unique fusion map to identify all other points revealing a strong similarity to it.

6. CONCLUSIONS

In this paper we have analyzed a set of methods to process multi-sensor data related to the underwater environment. A descriptor based on the detected number of circular shapes has been produced and tested. We aim at developing a data fusion model in which the information provided by the multi-sensor platform can be exploited for a higher level interpretation of the underwater scene. Starting from that, we implemented a procedure for automatic recognition of interesting objects. In the framework of our research work we aim to produce new technologies for underwater archaeological survey, and we believe that the joint exploitation of standard survey sensors, such as acoustic and optical sensors, can be promising.

7. ACKNOWLEDGEMENTS

This work has been partially supported by THESAURUS project (*Techniques for Underwater Exploration and Archaeology through Swarms of Autonomous Vehicles* - PAR FAS Tuscany Project) and ARROWS project (*ARchaeological RObot systems for the World's Seas* - FP7 EU Project).

BIBILIOGRAPHY

[1] O. Pizarro, R. Eustice and H. Singh, "Large area 3D reconstructions from underwater optical surveys," *IEEE Journal of Oceanic Engineering,* vol. 34(2), pp. 150-169, 2009.

[2] F. Bonin, A. Burguera and G. Oliver, "Imaging systems for advanced underwater vehicles," *Journal of maritime research,* vol. VII(1), pp. 65-86, 2011.

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

[4] 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*, 2000.

[5] A. Martin, G. Sevellec and I. Leblond, "Characteristics vs. decision fusion for sea-bottom caracterization," in *Colloque Caractérisation in-situ des fonds marins*, Brest, 2004.

[6] L. Lengrand-Lambert, A. Martin and R. Courtis, "Sonar images segmentation and classification," in *ECUA 2010*, 2010.

[7] D. Moroni, M. A. Pascali, M. Reggiannini and O. Salvetti, "Curve recognition for underwater wrecks and handmade artefacts," in *4th International Workshop on Image Mining. Theory and Applications. VISIGRAPP 2013*, 2013.

[8] V. Patraucean, P. Gurdjos and R. G. Von Gioi, "A parameterless line segment and elliptical arc detector with enhanced ellipse fitting," in *ECCV 2012 - Lecture notes in computer science*, 2012.

[9] J. Ponce and Y. Furukawa, "Accurate, dense and robust multi-view stereopsis," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* pp. 1362-1376, 2010.

[10] E. Coiras, Y. R. Petillot and D. M. Lane, "Multiresolution 3D reconstruction from side-scan sonar images," *IEEE Transactions on Image Processing,* vol. 16 (2), pp. 382-390, 2007.

[11] D. Moroni, M. A. Pascali, M. Reggiannini and O. Salvetti, "Underwater scene understanding by optical and acoustic data integration," *Proceedings Of Meetings on Acoustics,* vol. 17, 2013.