IMEKO International Conference on
Metrology for Archaeology and Cultural Heritage
Lecce, Italy, October 23-25, 2017

# Supporting and promoting underwater archeology through computer vision and graphics

Marco Reggiannini and Marco Tampucci

*National Research Council of Italy, Via G. Moruzzi 1, Pisa, e-mail: {name.surname}@isti.cnr.it*

*Abstract* – **The essential mission of an archeologist can be synthesized in two main objectives. The first one concerns the entire system of experimental measurements and subsequent decision making procedures involved in the archaeological fieldwork process. These actions are usually based on the employment of engineering methods and technical tools that enable the implementation of site searching, mapping and documentation collecting tasks. On the other hand, once the fieldwork has been accomplished, the archaeologist main goal becomes the development of specific actions dedicated to the site preservation and prevention from illicit intrusion events and finally, to the dissemination of the gathered knowledge to increase people awareness about cultural heritage. The mentioned issues constitute the primary concerns for every institution involved in the safeguard and promotion of cultural heritage. This is true for on land and for underwater archeology as well. This paper focuses on issues related to the underwater archeology scenario, a circumstance that deserves an in-depth discussion since the intrinsic threats for human safety as well as the complex working conditions pose challenging issues to the archaeological mission. Nevertheless the huge patrimony of wrecks and man made artifacts lying all over the globe sea floors must be preserved in that it represents a collective community asset.**

## I. INTRODUCTION

The survey of an underwater environment, aiming at the discovery of archaeological sites, can be performed by means of unmanned robots, such as Autonomous Underwater Vehicles (AUVs), equipped with payload sensors properly selected in order to perform efficient sensing in the water medium. Given their complementarity features for this specific environment, acoustic and optical sensors (side scan sonar, cameras) turn out to be the most suitable option for the specific purpose of this work. In particular the former sensor provide large-scale maps of the environment, usually at coarse resolution, while the latter returns highly detailed visual representation of small scale sea floor spots, that must be captured at very close range to prevent from relevant signal degradation, due to fast electromagnetic absorption in the water medium. The collected multi-sensor data can be processed adopting a computer vision approach, aiming at providing the robot platform with an autonomous capability to understand the marine environment without human supervision. To pursue these goals, procedures for the automatic analysis of underwater maps have been developed within ARROWS[1] (FP7-Environment-308724), a project that aimed at mapping, diagnosing, cleaning and safeguarding underwater and coastal archaeological sites in the Mediterranean and Baltic seas.

Particular focus has been devoted to the implementation of algorithmic tools in charge of recognizing objects potentially related to man made manufacture, such as geometric shape detectors or textural segmentation and classification procedures. These tools are intended to help the archeological detection process through the extraction of relevant features from the data, followed by their exploitation in order to identify the areas that most likely host interesting objects. The output of these procedures provide a quantitative assessment that may trigger the real time adaptive planning of the survey missions. In a hypothetical circumstance where potentially interesting targets have been detected, this approach enables the on line optimization of the mission plan, for example by launching a dedicated mission to collect further details in the identified sea floor area.

Furthermore the close range captured optical data can be later exploited to generate highly detailed 3D reconstructions of the detected objects through advanced photogrammetry techniques. These reconstructions provide the archaeologist with information regarding the shape of the object and its dimensions and finally, they can also be exploited for the dissemination of the gathered novel information.

The remainder of the paper is structured as follows: the next section II concerns a brief description of the data capture process, its modeling and the preliminary stages of data processing and analysis. Section III focuses on data processing for 3D reconstruction purposes while section IV describes the exploitation of the 3D reconstructed models to set up a virtual environment conceived as a tool for measurement operations and for dissemination purposes.

[1] `http://www.arrowsproject.eu/`

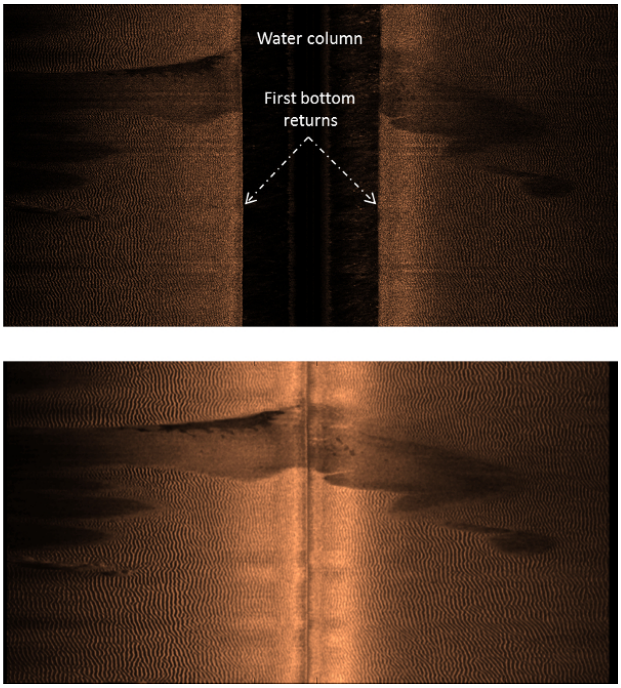


*Fig. 1. Side scan sonar map represented by Slant (up) and Ground (down) corrected coordinates.*

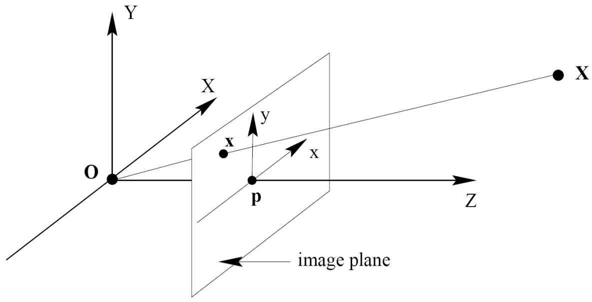


*Fig. 2. Pinhole Modeling of the optical data capture.*

## II. ATTENTIVE DATA CAPTURE AND PROCESSING

The image intensity in a sonograph is proportional to the strength of the returned signal, and the horizontal coordinate measured from the middle of a line scan corresponds to the time elapsed since the pulse was sent out. Equivalently the horizontal coordinate may be viewed as range from the side scan sonar, assuming that the sound pulse travels at a constant speed. The vertical coordinate of the sonograph corresponds to distance traveled by the side scan support (AUV). The dark stripe in the middle of the image corresponds to the interval of time between the emission of the sound pulse and the instant it first reaches the seabed, during which there are no reflection except those from fishes or other objects in the water column.

One type of geometric distortion is inherent to side-scan sonar: the so-called slant range distortion, a consequence of the cross-track coordinate of sonographs being range to the sensor rather than horizontal distances on the bottom. This is why sonograph maps require preliminary processing in order to correct geometrical artifacts due to the peculiar perspective of the backscatter capture process. A slant to ground (figure 1) transform is required to obtain a restored 2D map, on which correctly perform distance measurements.

For what concerns optical sensor the main issues emerging when capturing optical data concern the geometrical distortions caused by the propagation of light through the optical system. On the other hand it is of paramount relevance to process 2D images in order to extract metric information. Performing calibration on the optical cameras may solve this issue. Calibration takes place by performing camera observations of a planar pattern with known features, captured from a few different directions. The pattern to be used may be a chessboard with known squares dimensions. The calibration allows to estimate the intrinsic parameters of the camera (image center position, focal length, image scaling factors) by taking measurements directly on the captured images. The estimated parameters are then used to correct the captured data in order to remove the distortions introduced by the lens.

Camera calibration is typically based on the pinhole camera modeling [1] of the image formation process. Under this hypothesis a point in the world is projected onto the image plane according to the geometry illustrated in figure 2. The analytic relation between a 3D world point, identified by vector $\mathbf{X}$, and the corresponding projection, identified by vector $\mathbf{x}$, on the camera sensor plane is given by

$$s\mathbf{x} = \mathbf{A}\,[\mathbf{R}\;\mathbf{t}]\,\mathbf{X} \tag{1}$$

where $s$ is a generic scale factor and matrix $\mathbf{A}$ is defined as:

$$\mathbf{A} = \begin{pmatrix} \alpha & c & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2}$$

with $\alpha, \beta$ scaling factors in the image formation, $c$ relates to the skewness of the axis and $u_0, v_0$ are the coordinates of the principal point expressed in the image pixels units. $[\mathbf{R}\;\mathbf{t}]$ represents the extrinsic term in the pinhole model, taking into account the roto-translational misalignment between the camera reference frame and the world coordinate system.

The estimation of the camera projection model can be exploited to rectify the images and restore the correct geometrical properties of the optical data.

### A. *Primitive curves detector*

Searching for man made artifacts within a complex environment such as the underwater scene represents a demanding task to be assigned to a robot. Indeed the efficient human capabilities in terms of acquiring knowledge about the surroundings and take suitable decisions, are not easily transferred to a machine framework. Interesting objects may exhibit typical features that may refer to different attributes categories (peculiar morphology traits of surface appearance). Therefore a preliminary step concerning the identification of the feature typologies that the machine will have to look for is a paramount requirement for a mid-level machine processing stage. This starting point allows to endow the computer with proper skills, aimed at building a robust awareness about the explored scenario [2].

The silhouettes of man made artifacts can be thought of combinations of segments of regular geometrical shapes. This is typically the case of archaeological objects like amphoras, vases or generic pottery. Regular shapes can also be observed in the external structures of ships and wrecks. It is therefore interesting to implement a procedure to process the maps captured by the on board sensors, in order to detect and classify those pixels that belong to some primitive geometrical curve and classify the curve typology (line, circle or ellipse). The curve detection process can be exploited to assess the presence of regularity in the environment and define an interest measure (as in [3]) of a specific area on the basis of the number of detected curves, their length and their persistence in the chronological sequence.

The curve detection problem is an old computer vision issue that has been tackled and solved by means of popular mathematical tools such as the Hough transform and its variations [4]. With this approach the typical computational time required to detect the curves represented in an image grows proportionally with the number of parameters of the sought curve. This means that the detection of an ellipse requires asymptotically $O(n^5)$ operations to be performed, where $n$ is the number of pixels. In order to make faster the execution of the detection procedure a statistical approach has been adopted during the algorithm implementation phase. This allows to reduce the computation time aiming at a real time development of the detection process. We describe in the following the implementation of an efficient parameterless line and ellipse segment detector. Its efficiency relies on the statistical framework on which the procedure is based, that allows to describe each considered candidate in terms of a probabilistic measure. The procedure goes through three main steps: i) a first candidate selection stage concerning the grouping of those image pixels that fulfill specific collinearity constraints, ii) a candidate validation stage where the candidates determined in the former stage are considered again to be labeled with a rank value indicating how much every candidate is likely to be perceived as a good observation in an unstructured "noisy" image and iii) a final third stage concerning the association between the survived candidates and the most suitable model.

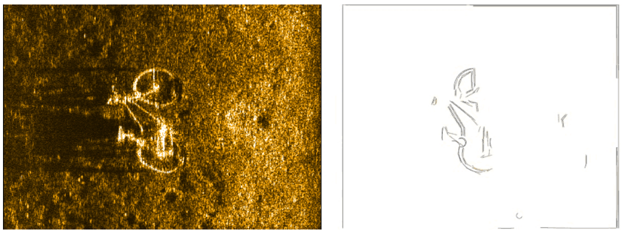

*Fig. 3. Geometry detector applied to a sidescan map. Original side scan map available on http://www.jfishers.com/.*

In a nutshell this method is based on the so called *Helmholtz perception principle* which formally states that in an unstructured image, only a very small number of detections (false alarms) should take place. The decision about a meaningful candidate curve is based on the probability of observing candidates as structured as the considered one: the smaller this probability value is, the more meaningful the candidate curve is to be considered.

The discussed geometry detection technique (an example of its application is illustrated in figure 3) can also be exploited to perform attentive analysis of the collected optical and acoustic maps. By applying the curve recognition algorithm to the new maps, as soon as they become available during the mission, an instantaneous label of interest is assigned to the surveyed regions. This enables to perform a real-time assessment about the presence of geometrical curves in the data. By evaluating the real time evolution of the geometrical presence in the images the system can be programmed in order to autonomously and promptly decide whether a specific area is worth of more detailed inspection or not.

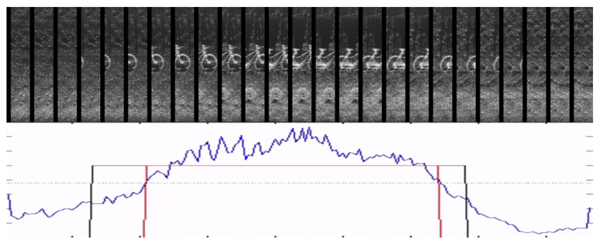

*Fig. 4. Attentive analysis procedure based on geometry detection applied to the side scan sonar map shown in the previous sections. The blue curve growth is proportional to the number of detected curves, the red line represents the threshold over which the detected number of curves is to be considered relevant and the black line represents a ground truth reference.*

Consider as an example the acoustic signal flow depicted in figure 4, corresponding to a set of $n$ frames. The primitive curves have been detected for a first set of $n_A$ frames and exploited to compute a weighted average ($A_w$) of the detected curves. Hence two discovery thresholds have been defined as $\sigma_o = 1.2 \times A_w$ for optical maps or $\sigma_a = 1.1 \times A_w$ for acoustic maps. Later, the geometry detector is applied to the remaining set of $n_B$ frames. The computed thresholds $\sigma_o$, $\sigma_a$ are compared to the geometry presence index evaluated for each frame. Finally a warning is produced whenever a value of the geometry index overshoots the threshold. In the example illustrated in figure 4 a substantial correlation between detections and ground truth can be observed.

### B. *Texture based classifier*

The observation of the environment by means of optical or acoustic sensing devices entails the mapping of the 3D world scenario onto the 2D image domain. This transformation usually preserves interesting scenario features that are crucial because they hold relevant information about the peculiar nature of the objects located in the environment. Image segmentation consists in splitting an input image into piecewise connected subregions and in assigning each region a specific class label corresponding to each region typology. Segmentation is based on the analysis of the texture attributes of the image, that is the spatial arrangement of gray values of pixels and their neighbors, and the their variation properties.

Specific regularity patterns observed in the pixel intensity values can be interpreted as identifying traits of the class of objects featuring that pattern. These features can be exploited to discern and classify image areas belonging to various sea floor categories (e.g. sand, rock or vegetation). This is possible since every texture class typically exhibits one or more dominant component in its spatial frequency content, and that can be exploited as a discriminative signature of the texture.

Psychophysical and psycho physiological experiments fostered the thesis that the human visual perception is based on the decomposition of the retinal image into multiple images (*channels*), each of which can be interpreted as the result of primary processing the retinal image by a suitable filter. Each filtered output features a spectrum that corresponds to a subset of the original image spectrum, drawn from a sufficiently narrow frequency window. The procedure discussed in this work took inspiration from the above mentioned conceptual framework.

Based on the representation of the captured maps in the spatial frequency domain a multichannel approach has been implemented by considering each channel as a filtering kernel. The overall set of channels may also be seen as a set of basis vectors for the 2D real valued functions. Therefore this basis functions can be exploited to represent the image as a linear combination of the mentioned vectors. A suitable and popular choice for the basis functions is given by the *Gabor wavelet* functions. These functions were proved to be appropriate filtering kernels for the mentioned segmentation purposes, since they realistically mimic the human vision apparatus.

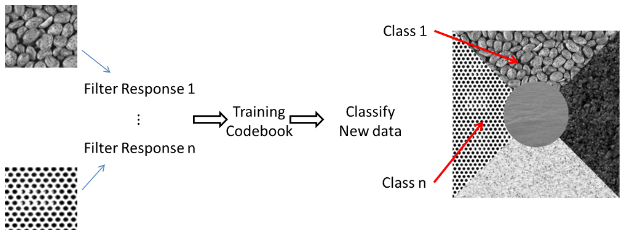


*Fig. 5. Preliminary training for the texture classification algorithm (original image taken from http://note.sonots.com/)*

The segmentation algorithm can be executed as an unsupervised procedure where the cluster centroids are estimated iteratively from the data, by applying K-means from scratch every time a new mission takes place. An advanced implementation is based on a preliminary training stage in which a set of known classified patterns are processed with different Gabor wavelets in order to provide the main spectral features for various classes, such as rock, sand, mud, posidonia, etc. The results obtained this way are then used to train the algorithm, which is later employed to classify areas in new captured data, based on the gathered a priori knowledge. A conceptual sketch of this procedure is represented in figure 5 and a result of the segmentation process applied on real side scan sonar data is represented in figure 6.

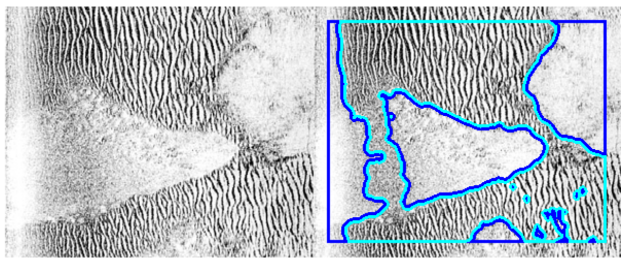

*Fig. 6. Texture classification of side scan sonar maps by means of Gabor filtering (original image taken from http://www.ise.bc.ca/)*

This way it is expected that the resulting process may be employed also for online purposes, aiming at a fast preliminary classification of the environment to quickly identify the interesting spots.

## III. RECONSTRUCTION CHAIN

Exploration of the underwater environment by means of a robotic system usually mandates the acquisition of the 3D bottom profile, as bathymetric information about the sea floor details, which is a fundamental step to automatically understand the surrounding scenario. The sea floor relief can be estimated by exploiting 3D reconstruction techniques, such as photogrammetry methods. These are mainly based on sampling the environment by capturing multiple images from many different perspectives. Then a feature detection (for example SIFT features) stage is applied, followed by a feature matching procedure which allows to identify those 3D points that occur repeatedly in the images sequence. Finally a 3D coordinates estimation of the salient keypoints is performed, for example by applying a Structure from Motion method [5].

In detail, the reconstruction process is composed by five steps. This process can be completely automatic or semi-automatic since each step produces an output which can be accessed and modified by the user. In the first step, the frames of the video acquired by the AUVs are aligned and, for each point of view, the corresponding pose is estimated. Secondly a first sparse point cloud is generated, exploiting the previous frame alignment and pose estimation results. The results of these two preliminary steps are jointly exploited in the third step, consisting of the dense point cloud generation, i.e. a refined version of the first one with a larger number of reconstructed points. In the following fourth step the dense point cloud is processed to generate a mesh surface, that is then refined in the final step, concerning the generation of a textured mesh. The texture information exploited in the latter stage is extracted from the starting acquired frames.

As mentioned before, the output of each step can be manually modified to provide better results. For instance, the frames alignment can be improved by adding markers to the frames. Also, sparse cloud, dense cloud and the mesh surface can be manually corrected in case of identified irregularities. In this circumstance the mesh processing tool Meshlab[2] has been employed aiming at the refinement of the meshes generated by Photoscan[3]. In particular the dense meshes provided by Photoscan have been manipulated by performing simplification and adaptation operations, such as mesh decimation, scaling and roto-translation transformations. The mesh decimation is particularly desirable for our purposes since typical full resolution reconstructions consist of several millions of points. This causes management issues in the subsequent processing steps, including the management of such big models during the virtual environment generation step.

Examples of real data processing are provided in the following, concerning a data set captured during an ARROWS campaign, at the artificially flooded quarry of the Rummu lake, Tallinn (Estonia). In that circumstance the distance separating two detected machineries lying on the lake floor, has been estimated by means of the Sonar data captured by a Side Scan Sonar installed on an AUV. This data has been georeferenced by means of the positioning sensors aboard the vehicle, hence providing true ground reference. The optical captured data has been off line processed to provide an overall view (mosaic) and a textured mesh of the surveyed area by means of the previously discussed methods.

Figure 7 shows a mosaic obtained by processing the frames acquired during experiment and a view of the generated 3D mesh.

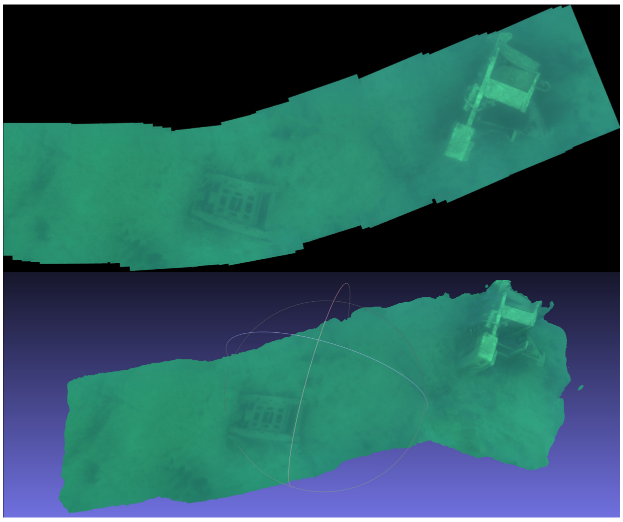

*Fig. 7. Reconstruction of machineries observed during Rummu Quarry experiment.*

## IV. MARINE VIRTUAL ENVIRONMENT

The 3D models reconstructed by processing the data captured during the ARROWS missions, have been employed for the design and the implementation of a virtual environment, called the Marine Virtual Environment (MVE), specifically conceived to provide a tool for the realistic and immersive fruition of the archaeological exploration experience.

The 3D reconstructed models and the eventual virtual environment have been designed and realized aiming at a double purpose functionality. Indeed the MVE simultaneously represents an useful tool for the archaeologist concerned with the exploration and the understanding of the underwater site while at the same time it can be exploited as a system for results dissemination, aiming at efficiently engaging the user and enabling the general public to access to the underwater site, otherwise out of reach to the majority of people. In order to accomplish with this twofold intention, metric photo realistic renderings are needed so

[2] http://www.meshlab.net/
[3] http://www.agisoft.com/

as to provide a reliable tool for the archaeologist analysis and to make the environment immersive and captivating.

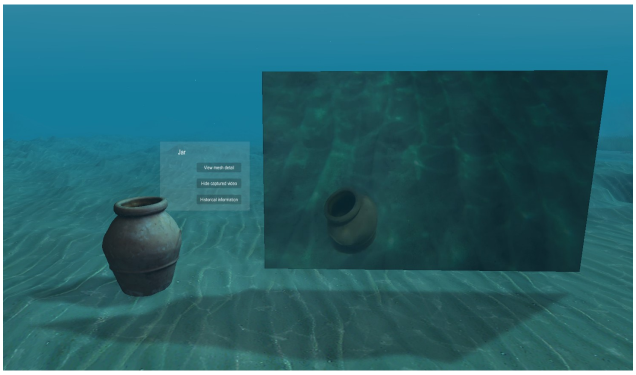

*Fig. 8. Marine Virtual Environment.*

The MVE has been realized exploiting Unity[4] game engine, since it provides a set of functionalities to efficiently set up the scene and manage the interactions with the scene itself and the included objects. Marine Virtual Environment is also connected with an informative system which contains information about the objects in the scene. Indeed, through this connection, the system is able to provide the users with historical information about the observed objects.

### A. *Interactions & Gestural devices*

The virtual environment has been enriched with a set of dedicated functionalities. In order to guarantee the system's appealing features, the MVE provides the users with several data regarding historical information and the exploration of the site. By interacting with the object placed in the scene the user can access to videos captured during the ARROWS missions (figure 8) or already available from pre-existing resources, to the raw data captured by different sensors (e.g. sonograms) but also to any kind of supplementary information such as the reconstructed 3D meshes of the objects, displayed separately from the scene and available for observations from multiple points of view. Moreover, as mentioned before, the MVE is connected with an informative system which contains historical and general information (such as the objects' purpose, the material, etc.) about the explored sites.

Finally, to ensure the immersive features of the system, alternative scene fruition modalities have been designed and developed. Indeed, the scene can be explored through the most modern head mounted visors such as Oculus Rift[5] or using gestural devices such as Leap Motion[6] and Kinect[7].

[4] `https://unity3d.com/`
[5] `https://www.oculus.com/rift/`
[6] `https://www.leapmotion.com/`
[7] `https://en.wikipedia.org/wiki/Kinect`

## V. CONCLUSIONS

The developed Marine Virtual Environment is accessible by all kind of users without any restriction. Indeed, MVE is Operative System independent and can be used both on Windows, Linux or MacOS; also mobile and browser versions are in development. The system has been developed able to provide support both for professional (such as archaeologist) and general user.

MVE offers a set of functionalities to analyze the scene objects and perform analysis of the site directly while exploring the scene. Moreover, by accessing raw data, users can also confirm the deductions obtained exploring the scene. The accurate positioning of the object inside the scene helps the archaeologists to discover the causes behind the sink and the generation of the site itself.

On the other hand, exploiting a captivating graphics and advanced gestural and vision devices, the Marine Virtual Environment is attractive for the general user interest. The virtual diving in underwater scenarios, thoroughly reconstructed and modeled by processing the data taken by the AUV multi-sensor platforms, features a large series of possible choices in terms of exploration actions, thus recreating in a strongly realistic way the survey of underwater locations and the discovery of interesting archaeological sites.

Most of the presented results, including the collection of the data, its processing using the reported methods, the 3D reconstructions and the virtual scenarios developed with the aim of replicating the experience of wreck exploration and survey, have been made possible in the framework of the European FP7 project ARROWS.

## REFERENCES


[1] P. Drap, J. Seinturier, D. Scaradozzi, P. Gambogi, L. Long and F. Gauch, "*Photogrammetry for virtual exploration of underwater archaeological sites*", Proceedings of the 21st CIPA symposium, Athens 2007.

[2] M. Reggiannini and O. Salvetti, "*Seafloor analysis and understanding for underwater archeology*", Journal of Cultural Heritage, vol. 24, pp. 147 - 156. Elsevier Masson SAS, 2017, doi = 10.1016/j.culher.2016.10.012.

[3] D. Moroni, M.A. Pascali, M. Reggiannini and O. Salvetti, "*Underwater Man-made and Archaeological Object Detection in Optical and Acoustic Data*", Pattern Recognition and Image Analysis. Advance in Mathematical Theory and Applications, vol. 25, no.2, pp. 310-317, Nauka/Interperiodica 2014, doi = 10.1134/S1054661814020138.

[4] P.V.C., Hough, "*Machine Analysis of Bubble Chamber Pictures*", International Conference on High Energy Accelerators and Instrumentation, 1959.

[5] R.I. Hartley A. and Zisserman, "*Multiple View Geometry in Computer Vision*", 2nd ed., 2004, Cambridge University Press.