



**UNIVERSITA' DELLA CALABRIA**

Dipartimento di Ingegneria Meccanica, Energetica e Gestionale

**Scuola di Dottorato**  
**in Scienze Ingegneristiche Pitagora**

**Indirizzo**

**INGEGNERIA MECCANICA**

*Con il contributo di PON Ricerca & Competitività 2007/2013*  
*PON01\_02140 COMAS - CONSERVAZIONE PROGRAMMATA, IN- SITU,*  
*DEI MANUFATTI ARCHEOLOGICI SOMMERSI*

---

**XXVII CICLO**

**OPTICAL SURVEYING TECHNIQUES FOR 3D RECONSTRUCTION  
OF SUBMERGED ARCHAEOLOGICAL ARTEFACTS**

**Settore Scientifico Disciplinare ING-IND/15**

**Coordinatore:** Ch.mo Prof. Leonardo Pagnotta

Firma

**Supervisore/Tutor:** Ch.mo Prof. Fabio Bruno

Firma

**Dottorando:** Dott. Francesco De Filippo

Firma



*To Ida and my family*

*“It is not enough to take steps which may someday lead to a goal;  
each step must be itself a goal and a step likewise.”*

*( Johann Wolfgang von Goethe )*

## ***ABSTRACT***

---

This thesis is concerned with optical surveying techniques for 3D reconstruction in underwater environment. In particular, it focuses on 3D modelling of archaeological areas, lying both in shallow and deep water. Virtual representations of archaeological sites are required by researchers in order to enhance study and preservation methodologies and also to ease promotion and dissemination.

The presented work mainly follows two research lines.

First, it tackles the main problems arisen for standard 3D reconstruction algorithms and procedures when applied on submarine datasets. Multi-View mapping algorithm has been investigated and enhanced in order to achieve high accuracy reconstruction of two submerged *villae*, despite the challenging context of Baia Archaeological Park. Then, the textured and high detailed models can be employed for different purposes; in particular, some examples of virtual exploration, simulation for device design and for intervention planning and monitoring are shown. The second research line deals with the implementation of a Simultaneous Localization and Mapping (SLAM) algorithm for Unmanned Underwater Vehicles (UUVs). Some datasets acquired by an UUV equipped with typical navigation sensors and with a stereo-camera have been processed through the implemented Extended Kalman Filter (EKF), in order to perform sensor fusion and enhance both vehicle tracking and 3D scene reconstruction. Additionally, Augmented State Kalman Filter (ASKF) has been implemented in order to exploit optical data for loop closure detection in order to reduce drift error, typical of dead-reckoning navigation.

## ***SOMMARIO***

---

Il presente lavoro di tesi si occupa di tecniche di rilievo ottico per la ricostruzione 3D in ambiente subacqueo. In particolare, si concentra sulla modellazione 3D di aree archeologiche, site sia a basse che alte profondità. Le rappresentazioni virtuali dei siti archeologici sono uno strumento sempre più utile per i ricercatori, al fine di migliorare le metodologie di studio e conservazione o per facilitare la promozione e la fruizione.

Il presente lavoro segue principalmente due linee di ricerca.

In primo luogo, affronta i principali problemi posti dall'impiego di algoritmi e processi standard di ricostruzione 3D, quando applicati su dataset subacquei. Algoritmo di Multi-View mapping sono stati studiati e migliorati, ottenendo la ricostruzione accurata di due *villae* sommerse, nonostante il difficile contesto del Parco Archeologico di Baia. In seguito, i modelli texturizzati possono essere utilizzati per diversi scopi; in particolare, vengono riportati alcuni esempi di esplorazione virtuale e di simulazioni per la progettazione di dispositivi e per la pianificazione e il monitoraggio degli interventi. La seconda linea di ricerca affronta la realizzazione di un algoritmo di Simultaneous Localization and Mapping (SLAM) per veicoli sottomarini (UUVs). Alcuni dataset acquisiti da un UUV, equipaggiato con i tipici sensori di navigazione e con una stereo-camera, sono stati elaborati dal filtro di Kalman esteso (EKF) sviluppato, al fine di eseguire una sensor fusion e migliorare sia il tracking del veicolo sia la ricostruzione 3D della scena acquisita. Inoltre, è stato implementato un filtro di Kalman a stati aumentati (ASKF) in modo da poter sfruttare i dati ottici per il rilevamento di eventuali chiusure di loop, operazione che permette di ridurre l'errore di deriva, tipico delle applicazioni di navigazione stimata.

## ***ACKNOWLEDGEMENTS***

---

At the end of this experience, I realize what a great opportunity I've had here.

The fact that I reached this point is due to the help and influence of many others.

First, I would like to thank my supervisor, Fabio, whose enthusiasm, understanding and positive attitude encouraged me through these years. In addition, I would like to acknowledge my appreciation for all of the opportunities and experiences he has given me. I am also very grateful to all the people in my Research Group, from all my colleagues to Professors Muzzupappa and Rizzuti, who always have been more than coworkers.

I am also really grateful to ViCOROB as a whole for hosting me in a really edifying research period and for providing some datasets used in this thesis. In particular, I would like to thank Professor Rafael Garcia (Rafa) and Nuno for their unlimited help and willingness, not only in research field. I also thank Dr. Jordi Ferrer (Pio) who developed the code I moved from, and who kindly gave his support whenever needed, even if I have never met him.

An affectionate thought is for all the guys in the lab, coming from all over the world, who shared with me the work but, most of all, dinners and free time.

Additionally I thank the *CoMAS* project that entirely funded this work.

I thank all my friends and housemates for all the moments spent together and for both serious and trivial discussions about research and life.

I cannot finish without thanking my beloved family for all of the love and support they have given me through the years. I always miss you, even if sometimes it could not seem so...

Lastly, but definitely first, thank you to Ida, the most precious discovery that I could have possibly made during my PhD experience. Thank you for being a source of joy and sweetness in my life and for the support that you and your lovely family provided beyond words.

---

# TABLE OF CONTENTS

---

<b>LIST OF FIGURES .....</b>	<b>v</b>
<b>LIST OF TABLES.....</b>	<b>viii</b>
<b>I INTRODUCTION .....</b>	<b>1</b>
1.1 Context.....	2
1.2 Motivation.....	4
1.3 Project Goals.....	5
1.4 Thesis Structure .....	6
<b>II STATE OF ART.....</b>	<b>7</b>
II.1 A Review of Underwater Computer Vision.....	7
II.1.1 Geometry of image formation .....	7
Pinhole Camera Model.....	8
Camera calibration .....	11
II.1.2 Image Registration.....	12
Feature Detection.....	14
Feature Description and Matching.....	17
II.1.3 Image mosaicking.....	20
Temporal Mosaicking.....	20
Global Mosaicking.....	21
II.1.4 3D Reconstruction Techniques.....	21
Epipolar Geometry.....	22
Active and Passive Optical Techniques.....	23
Stereo Triangulation .....	27
Structure from motion.....	29
II.1.5 Imaging constraints in Underwater Environment .....	31
Absorption and Scattering.....	31
Attenuation.....	32
Shallow water.....	33
Registration .....	33
Scene illumination .....	34

---

II.2	<i>A Review of Underwater Technology</i> .....	34
II.2.1	<i>Underwater Vehicles</i> .....	34
	<i>Manned Underwater Vehicles</i> .....	35
	<i>Remotely Operated Vehicles</i> .....	35
	<i>Autonomous Underwater Vehicles</i> .....	36
II.2.2	<i>Underwater Mapping</i> .....	36
	<i>Multi Beam Sonar</i> .....	37
	<i>Sub Bottom profiler</i> .....	37
	<i>Side Scan Sonar</i> .....	37
	<i>Multibeam Imaging Sonar</i> .....	37
II.2.3	<i>Underwater Navigation</i> .....	37
	<i>Long Baseline Navigation</i> .....	38
	<i>Short Baseline (SBL)</i> .....	39
	<i>Ultra Short Baseline (USBL)</i> .....	39
	<i>GPS intelligent buoys (GIB)</i> .....	40
	<i>Doppler Velocity Logs (DVL) and Inertial Navigation System (INS) navigation</i> .....	40
II.2.4	<i>Underwater SLAM</i> .....	43
II.3	<i>Underwater Archaeology Applications: Related works</i> .....	44
<b>III</b>	<b><i>3D MAPPING FOR DOCUMENTATION AND MONITORING OF UNDERWATER ARCHAEOLOGICAL SITES</i></b> ....	<b>46</b>
III.1	<i>Underwater restoration: the CoMAS Project</i> .....	46
III.1.1	<i>CoMAS case study: Ancient Baiae</i> .....	48
III.2	<i>Documentation of the conservation state and interactive visualisation</i> .....	50
III.2.1	<i>Archaeological context</i> .....	51
III.2.2	<i>Experimental setup</i> .....	52
III.2.3	<i>Image acquisition and colour correction</i> .....	52
III.2.4	<i>Multi-view 3D reconstruction</i> .....	55
III.2.5	<i>3D interactive environment</i> .....	60
	<i>Realistic mode</i> .....	61
	<i>Study mode</i> .....	62
	<i>ROV Simulation</i> .....	63
III.3	<i>Monitoring of the experimental cleaning operations</i> .....	64

---

III.3.1	<i>Archaeological context</i> .....	64
III.3.2	<i>The Documentation Process</i> .....	65
	<i>3D Mapping of the experimental site</i> .....	66
	<i>3D Reconstruction of the selected areas</i> .....	67
	<i>Cleaning operations</i> .....	68
	<i>3D Reconstruction of the cleaned areas</i> .....	68
	<i>Results evaluation</i> .....	69
<b>IV</b>	<b><i>KALMAN FILTER THEORY AND MODELLING</i></b> .....	<b>72</b>
IV.1.1	<i>Linear Continuous Kalman Filter</i> .....	72
IV.1.2	<i>Linear Discrete Kalman Filter</i> .....	75
IV.1.3	<i>Discrete Extended Kalman Filter</i> .....	78
IV.1.4	<i>Filter Tuning</i> .....	80
<b>V</b>	<b><i>IMPLEMENTATION OF A 6DOF EKF FOR INCREMENTAL UNDERWATER MAPPING</i></b> .....	<b>82</b>
V.1	<i>Motivations and goals</i> .....	83
V.2	<i>Assumption and Equipment</i> .....	85
	V.2.1 <i>Coordinate frame relationship</i> .....	85
V.3	<i>EKF Formulae for State Prediction and Correction</i> .....	87
	V.3.1 <i>State definition</i> .....	87
	V.3.2 <i>Vehicle model</i> .....	88
	V.3.3 <i>State Prediction</i> .....	89
	V.3.4 <i>State Correction</i> .....	89
V.4	<i>Observation Models</i> .....	90
	V.4.1 <i>Stereo-Tracking Measurements</i> .....	90
	<i>Image Registration and Optical tracking</i> .....	90
	<i>Stereo Tracking EKF Observation Model</i> .....	93
	V.4.2 <i>Navigation Sensor Measurements</i> .....	98
	<i>DVL EKF Observation Model</i> .....	98
	<i>AHRS EKF Observation Model</i> .....	99
	<i>Depth gauge EKF Observation Model</i> .....	100
V.5	<i>Extended Kalman Filter - EKF</i> .....	101
V.6	<i>Augmented State Kalman Filter</i> .....	103

---

V.6.1	<i>Open sequence case</i> .....	103
V.6.2	<i>Loop closure case</i> .....	106
V.7	<i>Algorithm applications</i> .....	108
V.7.1	<i>La Lune Dataset</i> .....	108
	<i>Archaeological context</i> .....	108
	<i>Results</i> .....	109
V.7.2	<i>Cap de Vol Dataset</i> .....	116
V.8	<i>C++ Implementation</i> .....	117
<b>VI</b>	<b><i>CONCLUSIONS AND FUTURE WORK</i></b> .....	<b>119</b>
VI.1	<i>Thesis Summary</i> .....	119
VI.2	<i>Contribution</i> .....	120
VI.3	<i>Future work</i> .....	121
VI.4	<i>Related Publications</i> .....	121
	<b><i>REFERENCES</i></b> .....	<b>123</b>

---

**LIST OF FIGURES**


---

Figure II-1 Pinhole camera model.....	8
Figure II-2 Harris detector: a) flat region; b) edge; c) corner, example taken from [39] .....	15
Figure II-3 Comparison of detectors for test sequence, non-maxima suppression radius 10 pixels: (a) Harris, 271 keypoints. (b) Hessian, 300 keypoints. (c) Laplacian, 287 keypoints. (d) SIFT, 365 keypoints. (e) SURF, 396 keypoints. (f) Preprocessed image actually used for detection, image taken from [38]. .....	16
Figure II-4 Comparison of detectors for test sequence, non-maxima suppression radius 15 pixels, keypoints are displayed at the initial images: (a) Harris, 419 keypoints. (b) Hessian, 439 keypoints. (c) Laplacian, 429 keypoints. (d) SIFT, 509 keypoints. (e) SURF, 484 keypoints. (f) Preprocessed image actually used for detection, image taken from [38]......	16
Figure II-5 In this example, SIFT was used for the matching. About 1000 features have been matched in the couple of images and most of them seem to be correct. ....	19
Figure II-6 Examples of acquisition by passive technique (left) and active technique (right)......	21
Figure II-7 For every pixel $m$ , the corresponding pixel in the second image $m'$ must lie somewhere along a line $l'$ . This property is referred to as the epipolar constraint. See the text for details.....	22
Figure II-8 Triangulation principle in active reconstruction.....	24
Figure II-9 SFM reconstruction example.....	25
Figure II-10 3D reconstruction sequence: point cloud, 3D virtual model (surface), 3D textured model.....	26
Figure II-11 Stereo triangulation example.....	27
Figure II-12 Typical underwater imaging problem. Images taken from [112]......	32
Figure II-13 Shallow water light effects. Images taken from [113] and [114].....	33
Figure II-14 Manned submersibles .....	35
Figure II-15 ROV .....	35
Figure II-16 AUV, Girona 500.....	36
Figure II-17 LBL (a); SBL(b); USBL(c) and GIB systems (d). Images taken from [133]......	38
Figure II-18 DVL sensors.....	41
Figure III-1 Geographical localization of Underwater Archaeological site of Baia (Italy)......	48
Figure III-2 The Baianus Lacus.....	49
Figure III-3 Underwater Archaeological Park: detail of Villa protiro.....	51
Figure III-4 Calibration panel used to manually correct the white balance.....	52
Figure III-5 Image acquisition using a standard aerial photography layout (a) and oblique photographs (b). .....	53
Figure III-6 Original uncorrected image (a), acquired image using pre-measured white balance (b) and final corrected image (c). .....	54
Figure III-7 Sparse point cloud and 533 oriented frames.....	55
Figure III-8 Reconstructed dense point cloud (a), Reconstructed surface(b). .....	56

Figure III-9 Final textured 3D model (a).....	58
Figure III-10 Final textured 3D model (b).....	59
Figure III-11 Scheme of the three applications of 3D interactive environment.....	60
Figure III-12: Screenshot of the application in realistic mode.....	61
Figure III-13 Top: photo of Villa in Baia, bottom: screenshot of the application in realistic mode.....	62
Figure III-14 Left: screen-shot of the application in study mode; right: top-view of study mode.....	63
Figure III-15 Test of reachability for narrow edges and example of measuring.....	63
Figure III-16 Underwater Archaeological Park: detail of Villa dei Pisoni.....	65
Figure III-17 3D surface and textured model of the site used to plan the cleaning intervention.....	66
Figure III-18 The scale bar and a marker over the two areas selected to test the documentation process....	67
Figure III-19 3D reconstruction (with and without texture) of the uncleaned Area 1 and Area 2.....	67
Figure III-20 Cleaning operations performed with nylon and stainless steel brushes.....	68
Figure III-21 3D reconstruction (with and without texture) of the cleaned Area 1 and Area 2.....	69
Figure III-22 Mesh deviations obtained comparing the cleaned and uncleaned surfaces.....	70
Figure III-23 Cross-section analysis used to correlate the amount of removed material with the type of encrustation.....	70
Figure V-1 Representation of world (in black) and local (in green) reference frames.....	86
Figure V-2 Girona 500 equipment (left); Girona 500 AUV with its coordinate frame $\{v\}$ and an example of a sensor frame $\{S\}$ [211].....	86
Figure V-3 Left: Stereo matching. Right: temporal matching.....	91
Figure V-4 Rectified quadruplet example and features matching: each vertex of triangulated structure represents a feature.....	91
Figure V-5 Point $M_i$ is imaged at time $t_0$ in the left and right camera of the stereo-rig at positions $m_{l0i}$ and $m_{r0i}$ respectively. Once the stereo-rig have moved the same world point $M_i$ is imaged in left and right camera coordinates $m_{li}$ and $m_{ri}$ . The stereo rig has a rigid transformation $R_s$ , $t_s$ between left and right cameras. The estimated motion that the stereo-rig suffered is denoted by $R$ , $t$ expressed with respect to the left camera frame.....	95
Figure V-6 Plot of heading values. Considering intervals in which the observed parameter is supposed to be constant, it is possible to evaluate noise and, consequently, a good approximation for sensor covariance.....	100
Figure V-7 EKF scheme.....	101
Figure V-8 Loop closure event.....	106
Figure V-9 La Lune. Left: the only known representation of the 'La Lune' (first ship on the left), from a 1690 painting. (©Musée National de la Marine, S. Dondain, France).Right: bathymetry of the shipwreck (Girona).....	109
Figure V-10 Left: Schema of the Girona 500 AUV, showing the location of principal components and payload area. Right: Deploying of the Girona 500 AUV above the La Lune shipwreck.....	110
Figure V-11 Left: the whole trajectory surveyed by the AUV (two surveys have been performed on the interested area); right: detail of the interested area with a portion of the considered "virtual path".	110

---

<i>Figure V-12 Sparse map estimated by Stereo-Tracker code with standard parameters.....</i>	<i>111</i>
<i>Figure V-13 Sparse map estimated by Stereo-Tracker code with parameters adapted to maximize feature matched. ....</i>	<i>112</i>
<i>Figure V-14 Left: Bathymetry of the area with sensor tracking elaborated by EKF in MATLAB (in red) and by the AUV internal navigation EKF. Right: giga-mosaic of the area with pictures from longitudinal transect highlighted. ....</i>	<i>112</i>
<i>Figure V-15 Trajectory estimated by EKF for the left camera. ....</i>	<i>113</i>
<i>Figure V-16 Trajectory estimated by ASKF for the left camera: blue frames represent loop closures; red frames are non-overlapping subsequent images. ....</i>	<i>114</i>
<i>Figure V-17 Loop closure detail .....</i>	<i>115</i>
<i>Figure V-18 Cap de Vol site in Agisoft Photoscan®: 3D model(above) and camera poses resulting from image registration (bottom). ....</i>	<i>116</i>
<i>Figure V-19 Top: La Lune trajectory estimated by C++ code (in blue) and by MATLAB® EKF (in red) based only on navigation sensors. Bottom: Cap De Vol trajectory estimated by C++ (in blue) and by Agisoft Photoscan® bundle adjustment (in red). ....</i>	<i>118</i>

---

**LIST OF TABLES**

---

<i>Table III-1 Results of the 3D reconstruction related to the cleaned and uncleaned areas .....</i>	<i>68</i>
<i>Table III-2 Results of the analysis conducted comparing the uncleaned and cleaned surfaces.....</i>	<i>70</i>
<i>Table IV-1 Continuous-time Linear Kalman Filter [150].....</i>	<i>75</i>
<i>Table IV-2 Discrete-Time Linear Kalman Filter [150].....</i>	<i>78</i>
<i>Table IV-3 Discrete-Time Extended Kalman Filter [150] .....</i>	<i>80</i>
<i>Table V-1. Sensors employed in our underwater platform.....</i>	<i>85</i>
<i>Table V-2 Generic Pi matrix.....</i>	<i>104</i>
<i>Table V-3 Scheme of a generic Piaugmented matrix.....</i>	<i>105</i>

# ***I INTRODUCTION***

---

Optical 3D reconstruction and mapping techniques have been widely investigated during the last few years. In particular, in the Cultural Heritage field, the use of digital models is becoming popular in application fields such as documentation, digital restoration, visualization, research and promotion [1, 2, 3]. Moreover, the increased Virtual Reality (VR) capabilities allow the employing of digitally modelled archaeological sites in order to promote their exploitation and facilitate the planning of conservation interventions.

This thesis, in particular, addresses the challenging problem of underwater 3D modelling. In effect, many underwater applications require gathering 3D data of flora, fauna, or submerged structures, for different tasks like monitoring, analysis, dissemination, or inspection. In particular, in the Cultural Heritage field, the reconstruction of submerged findings or entire archaeological sites is of great interest to researchers and enthusiasts, especially for inaccessible and hostile environments where data gathering is difficult and expensive. Moreover, according to the guidelines of UNESCO that suggest the *in-situ* preservation of underwater heritages [4], 3D reconstructions techniques and related tools have known a relevant growth in underwater archaeology field during the last decade. Among the various 3D techniques suitable for underwater applications, the photogrammetry represents a valid solution to reconstruct 3D scene from a set of images taken from different viewpoints. The data acquisition devices (both still and video cameras with the proper waterproof housings) are very affordable when compared to devices capable of providing range measurements such as LIDAR and multibeam echo sounders (MBES). Furthermore optical cameras can easily be deployed from underwater robots or handled by scuba divers to document the whole area of interest. When combined with specialized image processing techniques optical images allow the reconstruction of high-resolution textured 3D models.

The data gathering in traditional underwater archaeology is commonly performed via SCUBA diving but it is constrained by the practical depth that a diver can work (normally limited to 50 meters) and the time that can be spent underwater. New technologies, like manned submersibles or Unmanned Underwater Vehicles (UUVs) allow archaeologists to survey in very deep water while reducing operational costs. This increases dramatically the chance to acquire detailed morphometric data in underwater archaeological studies. Although technology plays a significant part in this work, it needs to be combined with the research methodologies used by archaeologists, so that archaeology in deep water conforms to the required standards [5]. Over

the last few years, it has been clearly demonstrated that archaeologists can benefit from new underwater technologies but their requirements pose new and sometimes fundamental problems for engineers [6], such as the need for very accurate navigation and high quality seabed mapping. Furthermore, imaging techniques in underwater suffer from particular problems induced by the medium, that have to be solved in order to achieve performances comparable with on-land activities, in terms of accuracy, resolution and acquisition time.

The combination of vehicle navigation and 3D mapping tasks is commonly known as Simultaneous Localisation And Mapping (SLAM). This is a well-known approach in robotics and it represents a very interesting tool for archaeological interventions and surveys and, in general, for any underwater robot operation. In effect, the possibility to accurately locate the underwater robot in a previously unknown scene can dramatically reduce the time and costs related to the data acquisition task. Previous works on terrestrial and aerial SLAM have produced impacting results [7, 8, 9], whereas in underwater environments there are still several issues to solve in order to achieve satisfying results. The main problem related to the subsea SLAM is that in the water the visible light and, in general, the electromagnetic waves suffer high attenuation and signal alterations. For this reason, different solutions have been investigated ranging from the use of acoustic devices (frequency-dependent range and resolution), near-field (1–5 m) vision, image enhancement and sensor fusion [10]. Thus, in this thesis two studies are presented: the former is intended to evaluate and optimize the performance of dense 3D mapping for documentation and monitoring purposes; the latter is addressed to improve underwater real-time SLAM by combining data from different sensors, usually employed in underwater navigation, with visual algorithms for pre-calibrated stereo-camera.

## ***1.1 Context***

The work of this thesis was developed within the framework of the *CoMAS* Project. *CoMAS* is an acronym for “COnservazione programmata, *in-situ*, dei Manufatti Archeologici Sommersi” (*In-situ* conservation of submerged archaeological artefacts) and its main aim is, indeed, the development of methodologies and tools oriented to the *in-situ* conservation of submerged archaeological artefacts. The most recent UNESCO guidelines for underwater cultural heritage prescribe the promotion, protection, and *in-situ* preservation (where possible) of underwater archaeological and historical heritage. This new orientation encounters a major difficulty related to the lack of knowledge, techniques, and materials suitable for underwater conservation because in the past

few decades, little attention have been devoted to subsea conservation studies, since the recovery of the findings had often been preferred. The *CoMAS* project is focused on the development of new materials, techniques, and methodologies for the conservation and restoration of subsea sites in their natural environment, according to criteria that is normally applied to the existing land Historic monuments.

In this context it clearly appears the need of carefully mapping the intervention scene; a map is a static and accurate representation of a space, mainly used to locate elements within them. Nowadays, 3D maps are a fundamental way to represent an area of interest and are widely used in a large amount of industrial and scientific application fields. In *CoMAS*, together with documentation, virtual restoration and exploration, 3D models of interested areas become essential for planning and monitoring the conservation activities of submerged remains and artefacts.

The *CoMAS* project includes the development of a Remotely Operated Vehicle (ROV) devoted to the routine cleaning and maintenance of the submerged sites. This specific task requires the accurate localization of the vehicle within the operation environment and a clear representation of the submerged scene in presence of low visibility conditions.

The case study for the *CoMAS* project is Baia of Bacoli (Naples, Italy). Its remarkable submerged Archaeological Park represents an ideal test scenario. In particular, it is a critical area because of the high level of turbidity, the heavy presence of marine flora, and the changing of lighting conditions, typical of shallow water, which can seriously affect the acquisition, thus compromising the reliability and the accuracy of the results.

Part of the work on underwater SLAM has been conducted in collaboration with the ViCOROB, a research group in Computer Vision and Robotics of the University of Girona, in Spain. In particular, the Underwater Vision Lab (UVL) team is a leader group in research and development of Unmanned Underwater Vehicles (UUVs) for accurate seafloor mapping and light intervention. For many years, they have dealt with different topics related to underwater vision and robotic for seabed analysis and reconstruction for different range of applications. Research interests encompass robot design and navigation, survey planning and simulation, 2D and 3D mapping of large areas with related underwater image processing requirements [11, 12, 13]. The ViCOROB knowledge in underwater application has gave a remarkable support to research activities related to underwater SLAM discussed in this thesis.

## ***1.2 Motivation***

Water covers more than 70% of our planet and, owing to hostile condition of deep seas, only a little percentage of seafloor is known and explored. The remaining part represents a potential huge source of information for biologists, geologists, archaeologists and many other researchers. Technology progress in recent times provide tools and methodologies to investigate this unknown world. UUVs can perform surveys in areas inaccessible to humans, and deploy a wide range of sensors useful for acquiring relevant data.

Maps are important tools to organize and store spatial information, and are of particular relevance as a display tool when that spatial information has a strong 3D content. Interest towards 3D mapping of the environment is considerably growing because of the possibilities provided by a detailed 3D model or map. Moreover, the latest progress in hardware and software performances allows achieving stunning results even with consumer cameras, so reducing related cost.

Virtual representation of objects and areas can be exploited both for study and dissemination aims. In addition, virtual visualisation and exploration in augmented reality can be very appealing tools.

Optical 3D imaging is even more promising for many disciplines related to the underwater field, because of the already discussed hostility of submarine environment.

The research topics investigated by the present work are ascribable to two very promising research lines:

- *Dense 3D mapping with a single camera performed by scuba divers:*

Scuba divers can perform acquisition of submerged scenes or objects lying at non-prohibitive deep by means of relatively low cost consumer camera. Taking care of underwater imaging peculiar aspects, it is possible to create impressive textured 3D model.

- *Interactive 3D reconstruction for supporting ROV guidance and manipulation:*

As in many applications, when operation scenario becomes hostile for humans, the better solution is to employ robotic platforms. In this case, due to depth and/or extent of interested scene, underwater vehicles become the only available option. In order to reduce risk and cost, UUVs are generally preferred to manned vehicles.

As mention above, the operation of UUVs faces important challenges related to their positioning inside the surveyed location. Consequently, it is necessary to have either a previously built map of the operation scene or to construct one during the surveying. Using optical sensing alone, it is extremely difficult to achieve the high standards for accurate positioning and map quality required by archaeologists. Among several factors, this difficulty is due to water turbidity, changes in natural light conditions in shallow water, and moving shadows when the robot it is fitted with artificial illumination in deeper water. For this reason, positioning data from other sensors will be exploited and used together with optical sensing.

### ***1.3 Project Goals***

Following the research lines highlighted in the previous section, the main goals of the presented work can be organised in two key topics. We try to explain briefly each of those, before giving an accurate analysis in the next Chapters.

- *Experimentation of 3D dense mapping in complex underwater conditions.*

Optical 3D reconstruction can be obtained in different manners and a series of previous studies available in literature or conducted in our Department tested different methodologies both in air and in water [14]. Among the most popular techniques, we tested the most affordable and promising ones to reconstruct some areas from the challenging context of Baia Archaeological Park. Chapter III points out how a certain care have been required during image acquisition and processing phases in order to optimise the results, and shows tangible contributions of accurate 3D mapping to planning and monitoring of conservation and restoration activities.

- *Sensor fusion: Develop an EKF-ASKF for the interactive 3D mapping with UUV*

Even in human reachable areas, UUVs employment can be useful to enhance operation performance surveying the interested area faster and more accurately than scuba divers do. Underwater visual SLAM, or simple 3D mapping, can take big advantages from sophisticate and accurate instrumentation equipping UUVs. It is possible to improve six-degrees-of-freedom (6DOF) tracking and 3D mapping from stereo pair capabilities, exploiting typical dead-reckoning navigation from sensors together with calibrated stereo-rig. A filter well known in navigation applications performs this sensor fusion: the Extended Kalman Filter (EKF), the non-linear counterpart of the Kalman Filter (KF). EKF is the *de facto* standard for navigation, able to give a

good estimate for different variables (the state vector) of a non-linear system, according to different sensor data and to the system model. In particular, we also implemented an Augmented State KF (ASKF), a modified version of the algorithm that stores and relate all state estimations in time, so providing trajectory optimization based on loop-closure detection. The final aim is to test the feasibility of a real time SLAM able to support ROV operator during intervention activities.

## ***I.4 Thesis Structure***

Thesis started with an overview of thesis main topics and objectives, presented in this *Introduction Chapter*. The work encompasses different topics (computer vision, underwater navigation and mapping, etc.).

*Chapter II* tries to give an extensive background of all problems tackled in the thesis, of chosen approaches and relevant possible alternatives.

Then, *Chapter III* treats the 3D mapping for underwater archaeology; it briefly presents the *CoMAS* project, its motivation and its objectives, and the underwater Archaeological Park of Baia as selected case study. Two performed large area 3D reconstructions are shown, together with possible employment of the digital models in documentation, exploitation, intervention planning and monitoring over interested area.

*Chapter IV* discusses different Kalman Filter forms and gives detail on its theory and related formulae.

In *Chapter V*, we illustrate the implementation of an Augmented State Kalman Filter oriented to 6DOF SLAM for underwater vehicle navigation. The implemented filter gathers data coming from on-board navigation sensors and images from a stereo-rig; we analyse the Filter implementation procedure and its characterising models and assumption, referring both to MATLAB and C++ version.

Finally, a *Conclusion Chapter* summarizes topics and results and considers possible further developments and research lines.

## ***II STATE OF ART***

---

The state of art study is a fundamental step for any research activity. It is useful in order to know what has been already done in the interested field, to give new inputs, possible solutions and to support assumed hypothesis. Our research focuses about three main realms: underwater computer vision, underwater technology and their application in underwater archaeology. For the sake of simplicity, we tend to overlook or just briefly describe approaches, tools and methodologies not directly involved in the carried out research work.

### ***II.1 A Review of Underwater Computer Vision***

Computer vision is a widespread research field that covers diverse range of theory and application. Here, the discussion is focused on aspects related with image acquisition and their exploitation for 3D reconstruction, needed to the understanding of the work presented; we analyse principles of image formation and registration, mosaicking, 3D reconstruction techniques and we eventually show the main constraints peculiar to the marine environment.

#### ***II.1.1 Geometry of image formation***

A digital image is a matrix of elements, the pixels (picture element), which contain the radiometric information that can be expressed by a continuous function. The radiometric content can be represented by black and white values, grey levels or RGB (Red, Green, and Blue) values.

Regardless of the method of acquisition of a digital image, it must be considered that an image of a natural scene is not an entity expressible with a closed analytical expression, and therefore it is necessary to search for a discrete function representing it.

The digitization process converts the continuous representation in a discrete representation, by sampling the pixels and quantizing the radiometric values. This is the process that directly occurs in a digital camera.

The imaging sensor collects the information carried by electromagnetic waves and measures the amount of incident light, which is subsequently converted in a proportional tension transformed by an Analogue to Digital converter (A/D) in a digital number.

The light starting from one or more light sources which reflects off one or more surfaces and passes through the camera's lenses, reaches finally the imaging sensor. The photons arriving at this sensor are converted into the digital (R, G, B) values forming a digital image. In following

subsection, we discuss about some fundamental topics of image processing that deal with 3D reconstruction problem.

### Pinhole Camera Model

The pinhole or perspective camera model (shown in Figure II-1) is commonly used to explain the geometry of image formation. Given a camera located at  $C$  (the pinhole or the *camera centre of projection*) and an image plane placed at a distance  $f$  (focal distance) from  $C$ , the pinhole model states that each light ray passes through  $C$  and maps a generic point  $M$  in 3D space to a 2D point  $m$  lying on the image plane.

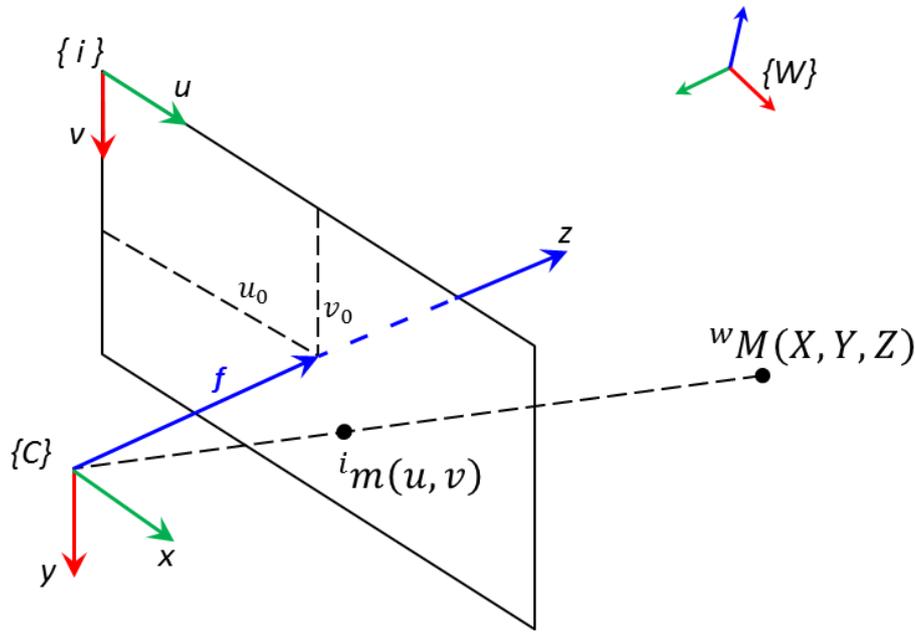


Figure II-1 Pinhole camera model.

With respect to the camera reference system  $\{C\}$  we can refer to  $M$  as  ${}^cM(X, Y, Z)$  and to  $m$ , its projection on image plane, as  ${}^c m(x, y, z)$ . However, 3D points are often represented in a different coordinate system. Defining  ${}^wM$  as the coordinates of  $M$  in reference frame  $\{W\}$ , it results that:

$${}^cM = {}^cR_w {}^wM + {}^c t_w \tag{Eq. II-1}$$

in which  ${}^cR_w$  represents a  $3 \times 3$  rotation matrix and  ${}^c t_w$  represents a translation vector for 3D points. Writing the Eq. II-1 in matrix form, we obtain:

$${}^cM = \begin{bmatrix} {}^cR_w & {}^c t_w \\ 0_{1 \times 3} & 1 \end{bmatrix} {}^wM \text{ i.e. } {}^cM = {}^cT_w {}^wM \tag{Eq. II-2}$$

where  ${}^cT_w$  is the transformation matrix that represents the pose of the world reference frame with respect to the camera one.

Due to triangle similitude, and to the fact that  $z = f$  on image plane,  ${}^cM(X,Y,Z)$  and  ${}^cm(x,y,z)$  are related through the following:

$$x = f \cdot \frac{X}{Z}; y = f \cdot \frac{Y}{Z}$$

Using homogeneous coordinates for  $m$ , this can be written in matrix form:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad \text{Eq. II-3}$$

The line through  $C$ , perpendicular to the image plane is called optical axis and it meets the image plane at the principal point.

Since the image coordinate system is typically not centred at the principal point and the scaling along each image axes can vary, the coordinates  $u, v$  in frame  $\{i\}$  undergoes a similarity transformation, represented by:

$${}^iT_c = \begin{bmatrix} s_x & s_s & p_x \\ 0 & s_y & p_y \\ 0 & 0 & 1 \end{bmatrix}; \quad \text{Eq. II-4}$$

In which,  $s_x$  and  $s_y$  represent the relationship between pixel and world metric units ( $pxs/mm$ ) respectively along  $x$  and  $y$  axes (so giving the possibility to consider rectangular pixels);  $s_s$  is the skew coefficient related to the skew angle between frame vectors  $x$  and  $y$  (it differs from zero only if frame vectors are not orthogonal), and  $(p_x, p_y)$  are coordinate of the principal point, expressed in image reference frame  $\{i\}$ .

Adding this term to the perspective projection equation, we obtain:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = {}^iT_c \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} {}^cT_w {}^wM \quad \text{Eq. II-5}$$

Alternatively, simply:

$${}^im = P {}^wM \quad \text{Eq. II-6}$$

where  $\mathbf{P}$  is a  $3 \times 4$  non-singular matrix called the *camera projection matrix*. This matrix  $\mathbf{P}$  can be decomposed as shown in Eq. II-7, where  $\mathbf{K}$  is the matrix containing the *camera intrinsic parameters*, whereas  $\mathbf{R}$  and  $\mathbf{t}$  together represents the *camera extrinsic parameters*, namely the relative pose of the world frame with respect to the camera.

$$\mathbf{P} = \mathbf{K}[\mathbf{R}|\mathbf{t}] \text{ where } \mathbf{K} = \begin{bmatrix} f s_x & f s_s & p_x \\ 0 & f s_y & p_y \\ 0 & 0 & 1 \end{bmatrix} \quad \text{Eq. II-7}$$

The intrinsic parameters in  $\mathbf{K}$  can be parameterized by  $f$ ,  $s_x$ ,  $s_y$ ,  $s_s$ ,  $p_x$  and  $p_y$  shown in Eq. II-7.

Thus, in the general case, a Euclidean perspective camera can be modelled in matrix form with six intrinsic and six extrinsic parameters (three for the rotation and three for the translation) which correspond to the six degrees of freedom for relative pose representation of world coordinate system with respect to the image coordinate system.

Nevertheless, real cameras deviate from the pinhole model due to various optical effects introduced by the lens system, visible as curvature in the projection of straight lines. Unless this distortion is taken into account, it becomes impossible to create highly accurate reconstructions. For example, panoramas created through image stitching constructed without taking radial distortion into account will often show evidence of blurring due to the misregistration of corresponding features before blending. Radial and tangential distortions are often corrected by warping the image with a non-linear transformation.

Following the distortion model presented in [15], we can consider the distorted image point  $m_d = (x_d, y_d)$  in the real image can be related to its undistorted version  $m_u = (x_u, y_u)$  through radial and tangential terms, by means of the following equations:

$$\begin{bmatrix} x_u \\ y_u \end{bmatrix} = (1 + k_{r1}r^2 + k_{r2}r^4 + k_{r3}r^6 + \dots) \begin{bmatrix} x_d \\ y_d \end{bmatrix} + d_t$$

where:

$$d_t = \begin{bmatrix} 2 \cdot k_{t1} \cdot x \cdot y + k_{t2} \cdot (r^2 + 2 \cdot x_d^2) \\ k_{t1}(r^2 + 2 \cdot y_d^2) + 2 \cdot k_{t2} \cdot x \cdot y \end{bmatrix}; \quad r^2 = x^2 + y^2;$$

Here  $k_r$ ,  $k_t$  are radial and tangential distortion coefficients, and  $r$  represents the distance of interested point from a hypothetical centre of distortion. Practically, tangential distortions are often less pronounced than radial ones and are ignored in some authors; common models consider up to the second or the third radial term and up to the second tangential one.

After evaluating the undistorted coordinates, we can map them to image plane through the intrinsic parameters, obtaining the pixel coordinates of associated point:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{K} \begin{bmatrix} x_u \\ y_u \end{bmatrix}.$$

The explained problem of image formation onto image plane is just the reverse problem of interested one. We need, in effect, to recover the position in the 3D space of a point (or others visible elements) imaged at a certain pixel coordinates in the picture. This goal cannot be achieved from a single view of the scene. Analysing the Figure II-1 it clearly appear that each generic point  $M$ , belonging to the half line through  $C$  and  $m$ , gives the same projection  $m$  on the image plane, so generating ambiguity that can be solved only by at least another view of the same point, as we later discuss.

### ***Camera calibration***

Intrinsic parameter evaluation is performed through a procedure known as camera calibration.

In traditional camera calibration procedures, images of a calibration target represented by an object with a known geometry are first acquired. Then, the correspondences between 3D points on the target and their imaged pixels found in the images are recovered. The next step is represented by the solving of the *camera-resectioning* problem that involves the estimation of the intrinsic and extrinsic parameters of the camera by minimizing the reprojection error of the 3D points on the calibration object.

The camera calibration technique proposed by Tsai [16] requires a three-dimensional calibration object with known 3D coordinates, such as two or three orthogonal planes. This type of object was difficult to build, complicating the overall.

A more flexible method has been proposed by [17] and implemented in a toolbox by [18]. It is based on a planar calibration grid that can be freely moved jointly with the camera. The calibration object is represented by a simple checkerboard pattern that can be fixed on a planar board. It recovers required parameters using Direct Linear Transformation algorithm [19] and a non-linear optimization procedure.

This method can be very tedious when accurate calibration of for large multi-camera systems have to be performed; in effect, the checkerboard can only be seen by a small group of cameras at one time, so only a little part of camera network can be calibrated at the same time. Then it is necessary to merge the results from all calibration session in order to calibrate the whole camera set.

Svoboda et al [20] proposed a new method for multi-camera calibration, based on the use of a single-point calibration object represented by a bright led that is moved around the scene. This led is viewed by all the cameras and can be used to calibrate a large volume.

Although all these methods can produce accurate results, they require an offline pre-calibration stage that is often impractical in some application. Robust structures from motion methods, that allow the recovery of 3D structure from uncalibrated image sequences, have been developed thanks to the significant progress made in automatic feature detection and feature matching across images. Hence, the possibility to concurrently estimate both the camera trajectory and the camera projection matrix parameters (intrinsic and extrinsic values) is available in structure from motion algorithms used with videos [21], or with large unordered image collections [22, 23]. Nevertheless, in order to reduce problem complexity and resource costs and also to increase model quality and robustness, a preliminary calibration step should always be carried out if possible.

### ***II.1.2 Image Registration***

Image registration is the process that aims to join two or more views of the same scene taken from different camera viewpoints but, more generally, also at different times or by different modalities (e.g. optical, infrared, acoustic, etc.).

In our case, as previously mentioned, image registration assumes great importance in order to reconstruct the 3D model of the scene reliably. Matching the same points in different images, in effect, can give information about object shape and also about camera motion in the scene; consequently, it is possible to recover, together with matched points, the relative pose of different views they are visible in, which is crucial information for 3D reconstruction that cannot always be available through other tools.

Image registration is a fundamental task for every activities involving more than one image. This task has application across many different disciplines spanning across scene change detection [24], multichannel image registration [25], image fusion [26], remote sensing [27], multispectral classification [28], environmental monitoring [29], image mosaicking [30], creating a super-resolution image [31], and as well aligning images from different modalities like in medical diagnosis [32].

Authors in [33] classify possible application in four main areas, according to the image acquisition setup:

- *Different viewpoints (multiview analysis)*

Images of the same scene are acquired from different viewpoints. The aim is to gain a larger 2D view or a 3D representation of the scanned scene [34].

- *Different times (multitemporal analysis)*

Images of the same scene are acquired at different times, often on regular basis, and possibly under different conditions. The aim is to find and evaluate changes in the scene, which appeared between the consecutive image acquisitions [35].

- *Different sensors (multimodal analysis)*

Images of the same scene are acquired by different sensors. The aim is to integrate the information obtained from different source streams to gain more complex and detailed scene representation [36].

- *Scene to model registration*

Images of a scene and a model of the scene are registered. The model can be a computer representation of the scene, for instance maps or digital elevation models (DEM) in GIS, another scene with similar content (another patient), “average” specimen, etc. The aim is to localize the acquired image in the scene/model and/or to compare them.

According to the Institute of Scientific Information (ISI), thousands of papers have been published during the last twenty years related to image registration, so confirming that it represent a crucial and on-going topic. Brown [37] did one of the most comprehensive surveys in 1992.

Due to the diversity of the images to be registered, and to various types of issues, it is impossible to design a universal method applicable to all registration tasks. Every method should take into account not only the assumed type of geometric deformation between the images but also radiometric deformations and noise corruption, registration accuracy and application-dependent data characteristics.

Nevertheless, the majority of the registration methods can be summarized in the following four steps:

- *Feature Detection*

Salient and distinctive objects (closed-boundary regions, edges, contours, line intersections, corners, etc.) are manually or (preferably) automatically detected. For further processing, these features can be represented by their point representatives (centres of gravity, line endings, distinctive points), which are called control points (CPs) in literature.

- *Feature Matching*

In this step, the correspondence between the features detected in the sensed image and those detected in the reference image is established. Various feature descriptors and similarity measures, along with spatial relationships among the features, are used for that purpose.

- *Transform model estimation*

The type and parameters of the so-called mapping functions, aligning the sensed image with the reference image, are estimated. The parameters of the mapping functions are computed by means of the established feature correspondence.

- *Image resampling and transformation*

The sensed image is transformed by means of the mapping functions. Image values in non-integer coordinates are computed by the appropriate interpolation technique.

### ***Feature Detection***

A very important requirement for a feature point is that it should be visually dominant and distinctive within the image. If it were not the case, it would not be possible to match it with a corresponding point in the other images. A traditional method for finding features is based on the measure of cornerness illustrated in Figure II-2. Some other detectors find blobs and ridges instead of corners. The main difference between the detectors is the manner to compute the cornerness

and the invariance of the detector to different geometry [38]. The main feature detectors are now summarized.

- *Harris Corner Detection*

The Harris corner methodology [39] underlies the idea of distinguishing the feature from the surrounding area image, because the neighbourhoods of the feature should be different from the neighbourhood obtained after a small displacement. By considering the difference between the patch centred on the pixel and the patches shifted by a small amount in different direction, the algorithm tests each pixel in the image to detect whether a corner is present or not. The Sum of the Square Difference  $SSD$  is computed by taking the sum of each pixel and its neighbourhood in one image and compare it with the  $SSD$  of all of the others pixels in the others images. The Harris corner technique is rotation invariant, invariant to linear changes in illumination, and robust to small amount of image noise.

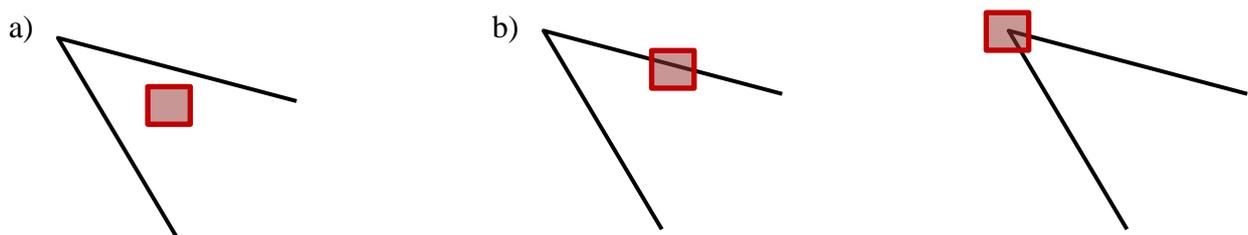


Figure II-2 Harris detector: a) flat region; b) edge; c) corner, example taken from [39]

- *Hessian Blob Detection*

Beaudet was the first to develop this technique in 1978 [40]. He created an operator that is rotationally invariant given a determinant of the Hessian matrix.

The cornerness measured from Beaudet's method is very good for blob and ridges detection. The region detected by the Beaudet method is nearly similar to the region detected by the Laplacian operator (see Figure II-3). It has been shown that the Hessian method is more stable than Harris method [41], even if it is less sensitive to the noise [42].

- *Laplacian of Gaussian (LoG)*

The Laplacian method is one of the most commonly method used blob detectors. This method is very robust for dark blob detection therefore very bad for bright blob detection [41].

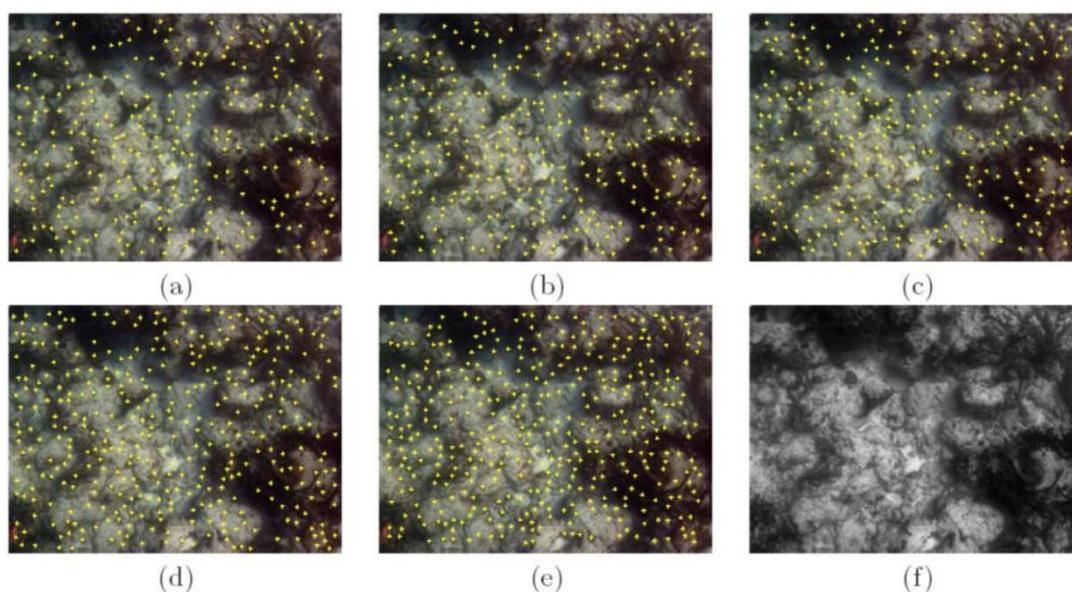


Figure II-3 Comparison of detectors for test sequence, non-maxima suppression radius 10 pixels: (a) Harris, 271 keypoints. (b) Hessian, 300 keypoints. (c) Laplacian, 287 keypoints. (d) SIFT, 365 keypoints. (e) SURF, 396 keypoints. (f) Preprocessed image actually used for detection, image taken from [38].

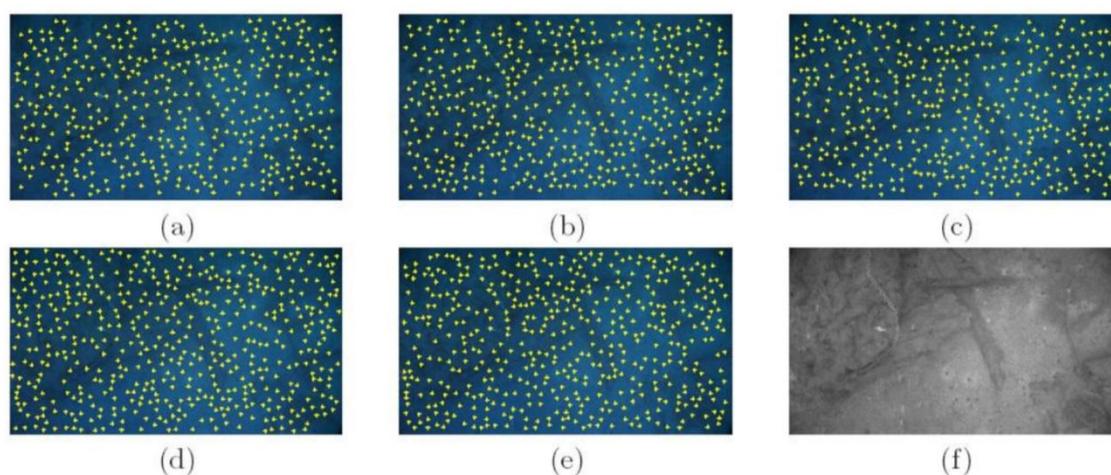


Figure II-4 Comparison of detectors for test sequence, non-maxima suppression radius 15 pixels, keypoints are displayed at the initial images: (a) Harris, 419 keypoints. (b) Hessian, 439 keypoints. (c) Laplacian, 429 keypoints. (d) SIFT, 509 keypoints. (e) SURF, 484 keypoints. (f) Preprocessed image actually used for detection, image taken from [38].

- *Scale Invariant Feature Transform (SIFT) Detector*

The SIFT descriptor published by Lowe [43] uses the Difference of Gaussians (DoG), an approximation of LoG, to extract salient areas. The image  $I$  is convolved with Gaussian kernels  $G$  at different scales  $k$ , taking the difference  $D$  between successive convolved images:

$$\begin{aligned}
 D(x, y, \sigma) &= L(x, y, k_i \sigma) - L(x, y, k_j \sigma); \\
 L(x, y, k\sigma) &= G(x, y, k\sigma) * I(x, y);
 \end{aligned}
 \tag{Eq. 11-8}$$

Convolved images are grouped by octave in which the  $k$  parameter is doubled in subsequent levels of the image stack, called scale space. Local maxima and minima extracted from the this scale space represent the needed keypoints features, that are invariant with respect to image scaling, translation, rotation and partially invariant to illumination changes and affine or 3D projection. Furthermore, sub-pixel accuracy can be achieved using a quadratic Taylor expansion of the scale space.

- *Speeded Up Robust Features (SURF) Detector*

The SURF algorithm [41] is similar to SIFT in the scale space feature extraction using the determinant of the Hessian matrix.

$$H(I(x, y, \sigma)) = \begin{pmatrix} L_{xx}(x, y, \sigma) & L_{xy}(x, y, \sigma) \\ L_{xy}(x, y, \sigma) & L_{yy}(x, y, \sigma) \end{pmatrix};
 \tag{Eq. 11-9}$$

In order to increase performances integral images are employed, together with box filters approximating Gaussian second order derivatives. It is possible, modifying the size of the filter, to have different responses thus obtaining the required scale space in which features are extracted.

- *Maximally Stable Extremal Region (MSER) Detector*

The MSER method developed by [44] is a blob detection algorithm. It uses a series of binarization, with threshold gradually increased, to extract zones with low variance over a large range of considered thresholds, corresponding to the higher or lower intensity values than all surrounding pixels.

### ***Feature Description and Matching***

For accurate image registration, features extracted from an image have to be correctly matched against the ones extracted from the other(s). Using statistics about just one pixel or small regions of pixels to characterize the feature would be highly instable in case of noise, illumination, and geometric changes. For these reasons, it is necessary to appropriately characterize the feature by choosing an adequate feature descriptor. State of art feature characterization methods are widely

discussed in [41, 45]. Here the discussion is focused on SIFT and SURF descriptors, that represent the most reliable ones. It has been shown that SIFT and SURF give better results in several applications and this makes them to be the most used descriptors [42]. Both approaches appears in hundreds of scientific publications because of their superiority with respect to correlation based methods, which can fail more frequently in underwater image registration. Different experiments in literature agree that SURF algorithm is very similar to SIFT in feature robustness, density, and reliability, but is significantly more performing in terms of the computational effort [41].

- *SURF Descriptor*

SURF descriptor is based on the distribution of intensity gradients around the features [41]. The use of gradients provides good feature distinctiveness. The SURF descriptor is extracted as followed:

In the first step, an orientation is obtained by calculating the wavelet response of Haar wavelets in  $x$  and  $y$  directions. The size of each Haar wavelet is four times the scale factor, and the size of circle around each point is six times the scale factor. The final orientation is achieved with reference to the total sum of wavelet responses within the sliding window. This sliding window covers an angle of  $\pi/3$ .

In the second step, in order to obtain the feature descriptor a square region is located at the central point with the reference to the selected orientation. This square is 20 times the scale factor and it is divided into 16 regions, each region is further divided into four smaller regions inside [38]. The final descriptor vector is based on the horizontal and vertical response of the Haar wavelet in each subregion; the descriptor is a four-element vector containing the absolute sum of horizontal and vertical responses of the Haar wavelet, this means that the final feature descriptor will have the length of 64 elements.

- *SIFT Descriptor*

The SIFT DoG detector gives invariance to translation and scale [43], and invariance with respect to rotation, change in illumination and 3D viewpoint can be managed by the descriptor. A local orientation is assigned to the detected keypoint as follows. The smoothed image  $L$  is selected using the keypoint scale in such a way that all computations are performed in a scale-invariant manner. The keypoint's scale can be different from all the ones represented in the scale space

levels. For each image sample,  $L(x, y)$ , at the closest scale the gradient magnitude  $m(x, y)$  and the orientation  $\theta(x, y)$ , are obtained with pixel differences as:

$$m(x, y) = \sqrt{(L(x + 1, y) - L(x - 1, y))^2 + (L(x, y + 1) - L(x, y - 1))^2} \quad \text{Eq. II-10}$$

$$\theta(x, y) = \text{atan2}(L(x, y + 1) - L(x, y - 1), L(x + 1, y) - L(x - 1, y)) \quad \text{Eq. II-11}$$

Using the gradient orientations of points in the neighbourhood of the keypoint, an orientation histogram is obtained. Each point is weighted by its gradient magnitude using a Gaussian-weighted circular window with the standard deviation being 1.5 times the one of the keypoint scale. The highest peak in the histogram corresponds to the dominant direction of the local gradient and is detected, and other peaks within 80% of that value are used to create keypoints with that orientation. Only near 15% of points are assigned multiple orientations, but they increase the matching stability. A parabola is finally added to interpolate the peak position for better accuracy. The resulting SIFT keypoint comprises a vector  $(x, y, \sigma, m, \theta)$  with  $x, y$  representing the coordinates,  $\sigma$  being the scale, and  $m, \theta$  being the gradient magnitude and orientation, respectively. The SIFT descriptor can be also used with Harris, Hessian and Laplacian detectors. In these cases, the orientation is computed based on the keypoints detected by each detector in a single image, without considering scale invariance. The feature descriptors obtained by SIFT contain 128 elements, since the  $4 \times 4$  descriptor array based on 8 bins orientation histogram was used. The set of keypoints given by the detection and the description process is used to find relations between two or more images. In general, depending on the images content and size, SIFT and SURF extract several hundred features, which are characterized by their spatial coordinates  $x$  and  $y$ , orientation  $\theta$ , and scale.

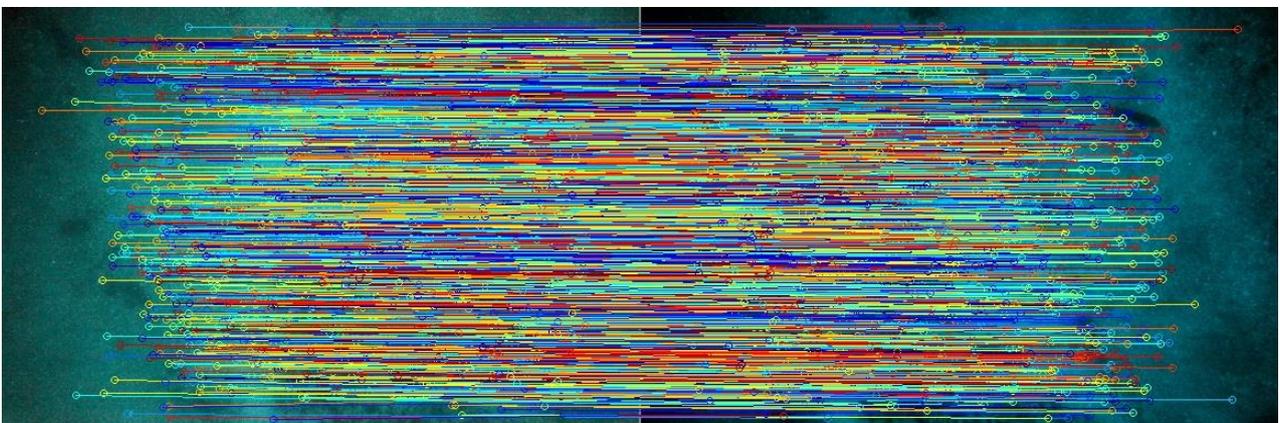


Figure II-5 In this example, SIFT was used for the matching. About 1000 features have been matched in the couple of images and most of them seem to be correct.

The basic criterion to associate features from two images is to find the nearest descriptors in terms of inner product between the descriptor vectors. In order to filter out associations between features with weakly discriminating descriptions (because they are too often present in the baseline, such as those that are extracted from background patterns often repeated), Lowe's method [43] evaluates the nearest neighbour and second nearest neighbour of each descriptor. When the ratio of distances is smaller than 0.8, the match is rejected because it will be considered as ambiguous. By this criterion, Lowe managed to eliminate 90% of false matches, while losing less than five of correct matches, in the test data reported in [38].

### ***II.1.3 Image mosaicking***

Mosaicking is the task of combining two or more images such that the resulting composite image has an increased effective FOV. The problem has been extensively studied [46, 47, 48, 49, 50], with early roots in aerial and satellite imaging where the planar parametric motion-model is well approximated due to the large separation between camera and scene. Planar parametric motion-models yield a composite image that is theoretically exact under only two conditions [51]:

- *the scene structure is arbitrary and the camera undergoes rotation about its optical centre*
- *the camera motion is arbitrary, but the scene being viewed is planar*

Both of these conditions are equivalent to no observed parallax in the input images.

#### ***Temporal Mosaicking***

Early methods in mosaicking by the computer graphics community approached the problem in a temporally causal manner [46, 47, 48, 49, 50]. These approaches processed the imagery in a sequential manner to determine the pairwise homographies relating the temporal sequence, and constructed a composite view by concatenation (thus, warping all images to a common reference frame). While the pairwise homographies accurately describe the local registration, the small residual local alignment errors, coupled with errors in the applied motion-model, lead to an amplified global error when simply concatenated over long sequences. Since the image to reference frame homographies calculated by compounding do not attempt to achieve global consistency, images that are not temporal neighbours, but are spatial neighbours, may not be co-registered in the resulting mosaic.

### ***Global Mosaicking***

More recently, efforts have focused on imposing the available non-temporal spatial constraints to produce a globally consistent mosaic [52, 53, 54, 55]. These methods formulate the problem as the optimization of a global cost function parameterized by all of the image to mosaic frame homography parameters. The mosaic topology may initially be derived in a coarse manner assuming simplified motion-model parameters between temporally connected neighbours. From this roughly estimated topology, new spatial neighbours are hypothesized and then tested. This process is iterated until a stable image topology emerges. The optimization of the cost function incorporates these spatial constraints to produce a globally consistent mosaic with enhanced quality and robustness as compared to simpler mosaicking methods.

#### ***II.1.4 3D Reconstruction Techniques***

During last few decades 3D reconstruction techniques have been widely investigated and improved. Our study focused on reconstruction techniques commonly employed, moving on from different reviews available in literature [56, 57, 58, 59] and from previous works conducted by our Department [14].

Optical reconstruction techniques allow obtaining a 3D digital model of real objects using pictures or videos of it. Image-based 3D modelling techniques (also called imaging techniques) are usually classified into passive, if the lightning source, natural or artificial, is needed only to light the object, and active, if lightning source takes part in the 3D reconstruction process [60].



*Figure II-6 Examples of acquisition by passive technique (left) and active technique (right).*

Photogrammetric techniques have known an important development during the last decade. Comparison between passive and active techniques have been studied both in air [60] and in

water [61] and the influence of water on calibration procedures has been investigated, too [62, 63].

Hereafter, we only touch on active techniques whereas we analyse more deeply some peculiar topics of passive techniques that, according to many tests and authors, appear to be more practical and promising for underwater applications.

### ***Epipolar Geometry***

The epipolar geometry captures the geometric relation between two images of the same scene. When a 3D point  $M$  projects to pixels  $m$  and  $m'$  onto two images,  $m$  and  $m'$  are said to be in correspondence. For every point in the first image, the corresponding point in the second image is constrained to lie along a specific line called the *epipolar line*.

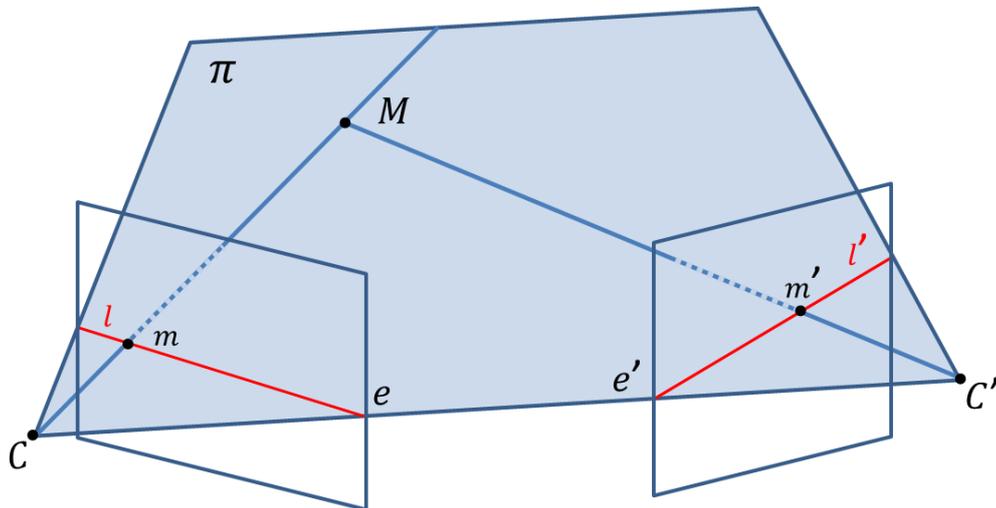


Figure II-7 For every pixel  $m$ , the corresponding pixel in the second image  $m'$  must lie somewhere along a line  $l'$ . This property is referred to as the epipolar constraint. See the text for details.

Every plane such as  $\pi$  that contains the *baseline* (the line joining the two camera centres), must intersect the two image planes in corresponding epipolar lines, such as  $l$  and  $l'$ , respectively. All epipolar lines within an image intersect at a special point called the *epipole*. Algebraically:

$$m'^T \mathbf{F} m = 0 \quad \text{Eq. II-12}$$

where  $\mathbf{F}$  is called the *fundamental matrix* and has rank two.

Points  $m$  and  $m'$  can be transferred to the corresponding epipolar lines in the other image, using the following relations.

$$l = F^T m' \quad l' = Fm \quad \text{Eq. II-13}$$

The epipoles are also the left and right null-vector of the fundamental matrix:

$$Fe = 0 \quad F^T e' = 0 \quad \text{Eq. II-14}$$

Since  $F$  is a  $3 \times 3$  matrix unique up to scale, it can be linearly computed from 8 pair of corresponding points in the two views using Eq. II-12, which is often called the epipolar constraint. This is known as the *8-point algorithm*. However, when the rank constraint is enforced,  $F$  can be computed from seven pairs of correspondences, using the non-linear *7-point algorithm*. Refer to [19] for the details. Any pair of cameras denoted by camera matrices  $P$  and  $P'$  results in a unique fundamental matrix. Given a fundamental matrix  $F$ , the camera pairs are determined up to a projective ambiguity.

### ***Active and Passive Optical Techniques***

3D active techniques involve the object illumination in the shape recover process. Particular kinds of artificial light sources illuminate the scene and are exploited to evaluate the position in space of image points. It is possible to achieve this aim through different ways, but all uses either time-delay or triangulation principles.

In particular, in the first category, relying on time delay, 3D reconstruction is based on the measurement of the time of flight (TOF) of a light pulse sent towards the object. Assuming a constant light speed, it is possible to calculate the distance from the device to the object.

Time delay active sensors are employed, instead, for large objects, i.e. architectures, rooms, walls and archaeological sites [64]. These sensors work on longer distances (i.e. 2 - 1000 m), but are less accurate than triangulation systems.

In the latter techniques, the geometry of the object is reconstructed thanks to the particular configuration of the measuring devices, which made possible the use of triangulation principles.

The system has to be calibrated in order to know the extrinsic parameters and so the relative pose of source and sensor. In effect, looking at Figure II-8, being known the baseline  $d$  (the distance

between sensor and source), and the angles  $\alpha$  e  $\beta$ ,  $L_o$  and  $L_s$  are derivable from easy geometric relations.

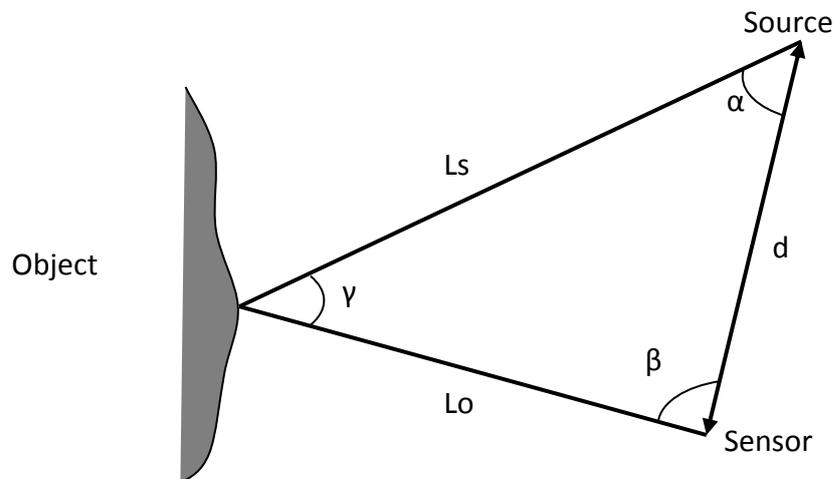


Figure II-8 Triangulation principle in active reconstruction.

The direction of emission patterns and the location of source and sensor have to be known through previous calibration. Both laser sources and sources of white light (digital projectors) can be used as sources of emission patterns; projecting an easy distinguishable pattern (coloured, monochromatic stripes, grey-scale or RGB sinusoidal fringes, etc.), it is possible to codify small areas in the images, so allowing triangulation and 3D reconstruction [59].

Triangulation-based laser scanners provide high accuracy in the 3D measurements, essential for 3D documentation of very detailed object as sculptures, bas-reliefs and other complex small objects. For example, experimented a laser scanner and a structured-light scanner to realize a digital model of the Minerva of Arezzo statue [65], whereas the renowned “Mona Lisa” has been studied by [66] exploiting shadow moiré method and fringe projection profilometry, based on the pattern projection technique.

Recently, [67, 68] achieved promising results using the Microsoft *Kinect* device, which exploit speckle pattern of infrared laser light, as source of data useful for 3D reconstruction.

Amongst the possible projected patterns, we analyzed those that seem more suitable for underwater application, considering errors and acquisition time. *Fringe projection* techniques, for example, employ a few number of lightened scene images, so reducing working time with respect, for example, to Grey-Code, which can require more than fifty patterns. In particular, *Phase Shifting Profilometry* can be implemented with three or four patterns, whereas the *Wavelet Transform Profilometry* can extract the 3D shape employing just a single shot [69]. Other promising

improvements to this realm, like astigmatic patterns [70], one-shot monochromatic pattern [71], multiple pattern for *phase shifting* [72], and many others have been developed.

These techniques produce good results in controlled scenes, but there are some issues when using them in real environment.

Passive methods recover all the information only from the object's texture. Possibly, illumination source are used just to lightening the scene either in case of low ambient light or when controlled environment is required.

These techniques are based on different shape-from-x approaches: shape from texture [73], shape from silhouettes [74], shape from shading [75], shape from focus [76], shape from stereo, etc. All of them are characteristics that contribute in human vision perception of object 3D shapes. When some information about the scene is known, it is possible to recover coarse shape and fairly inaccurate 3D information from a single image relative to an unknown scene [77, 78, 79].

It is possible to recover the 3D structure using geometric techniques such as multi-view triangulation by computing dense pixel-to-pixel correspondences between multiple calibrated images. The 3D reconstruction problem is quite challenging due to image noise, reflecting surfaces, bad illumination, moving objects, texture-less areas and occlusion. Often the reconstruction conduct to a computational expensive optimization problem, involving different mathematical formulations based on different criteria [80, 81, 82].

These methods can be classified into *local* and *global* methods. While global methods usually ensure higher accuracy, their computational complexity is usually much higher than the local methods [83], and cannot be used for real time applications.



Figure II-9 SFM reconstruction example.

There exist both methods employing imaging systems with two or more cameras (i.e. photogrammetry) and techniques requiring just one moving camera (i.e. Structure From Motion, SFM).

Even if good results have been reached using active techniques [84], passive techniques allow for a fast and accurate 3D survey by using a simple and low-cost hardware, also suitable for submerged environments. One of their most important developments is related to the combination with algorithms for SFM; this allows for reconstructing a 3D model of the scene through the application of photogrammetric techniques on pictures captured from different points of view by a single moving camera; each pose can be estimated exploiting triangulation principles and feature-detection algorithms (SIFT and SURF above all). After that, through photogrammetric constraints, it is possible to locate the point in the in the 3D space, so creating a point cloud of the scene. This point cloud can later be densified and used to create three-dimensional surfaces or solid structures, on which it is possible to warp even the texture recovered from pictures.

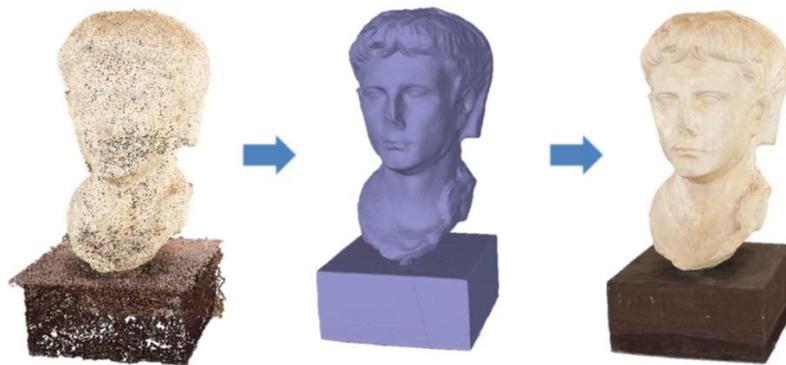


Figure II-10 3D reconstruction sequence: point cloud, 3D virtual model (surface), 3D textured model

The increasing effectiveness of multi-view techniques makes them ideal for 3D underwater surveys. In effect, in underwater environment, the acquisition by stereo-systems needs complex hardware configuration and handling. The possibility to use just a simple camera to take pictures, without following an *a priori* defined path, is a great advantage for scuba divers. Nevertheless, the employment of calibrated camera systems eases the reconstruction process and enhances the achieved results.

Obviously, due to different lighting conditions of underwater environment [85], image-based techniques employed in air could not give satisfying results if the right attention is not paid to avoid and/or correct typical problems of underwater vision.

### Stereo Triangulation

Stereo vision and stereo camera systems have been extensively used in latest years, not only by restricted research realm but also by the popular cinematographic and videogames industries.

The easiest way to obtain the depth of a scene is the stereopsis (or binocular vision), that is commonly known as *visual perception* process which gives the sensation of depth from two slightly different views of the scene. Generally, at least two images are required. This technique operates on the same principle as the human vision system. When a person looks at a point, the distance to the point is determined by comparing its apparent shift in position between the two eyes. Then, these techniques use 2D image measurements to recover 3D object information through a mathematical model. The observation of this principle has led to the development of photogrammetric techniques [86] based on the triangulation of corresponding points on the two views, by applying the epipolar constraints and computing the fundamental matrix to estimate the relative camera positions [87]. This requires a solution to the so-called correspondence problem [83], i.e. matching common points extracted, for example, by a SIFT operator.

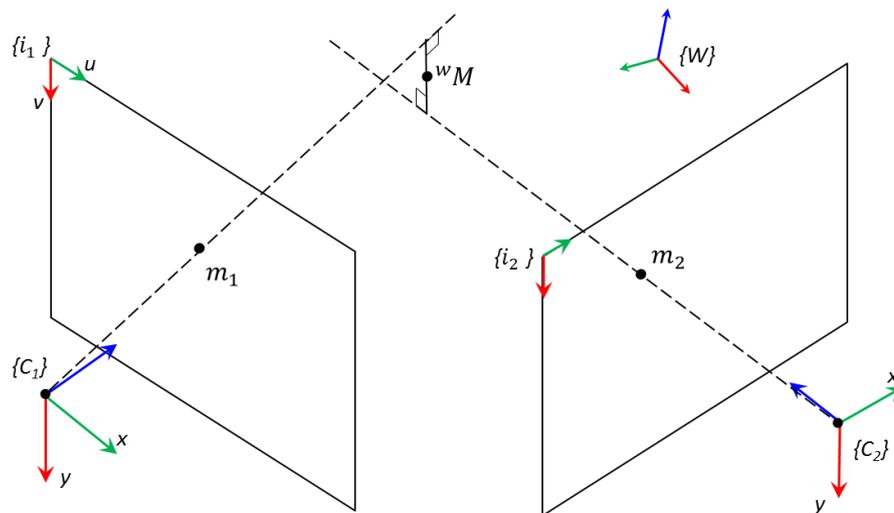


Figure II-11 Stereo triangulation example

Stereo triangulation problem can be reduced to evaluation of line intersection under certain condition. Imagine having a couple of cameras, of which both intrinsic and extrinsic parameters are known: we can recognize the same point  $M$  mapped in the two pictures at  $m_1$  and  $m_2$  and consequently evaluate its pixel position in each image; the 3D position  ${}^w p$  of the interested point can be retrieved by triangulation process on the lines-of-sight. These lines are ray passing through both the centre of camera and an imaged point (see Figure II-11).

In a perfect system, the 3D point would be exactly at intersection of line-of-sight but in real environment, due to errors, these rays do not intersect each other. Hence, the point  ${}^wM$  will be placed in the middle of the shortest segment connecting the two lines.

Recalling that from Eq. II-6:

$$m = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{P} \cdot {}^wM;$$

For each imaged point  $m_1, m_2$  we can impose:

$$m_1 \times \mathbf{P}_1 \cdot {}^wM = 0;$$

$$m_2 \times \mathbf{P}_2 \cdot {}^wM = 0;$$

Each of the previous constraints generates a system of three equations with three variables, corresponding to the 3D space coordinates of  ${}^wM$ , but only two of them are linear independent. Thus, as expected, in order to evaluate the three unknown variables, we need to impose an equation system, choosing at least three of the four linear independent equations generated. Usually, due to noise that affects measurements, the overdetermined system is solved through approximate method such as least squares formula.

The multiple-camera system, with overlapping FOVs, allows for a much wider FOV for panoramic views, photo-mosaics, 3D motion estimation and positioning [88, 89]. The configuration with three synchronized still cameras is a good compromise between accuracy and encumbrance, so it can be a good solution for underwater surveys conducted by a scuba diver [90]. With respect to a standard stereo system, the use of the third camera can improve the quality of the acquisition, reducing the errors due to the mismatching and the occlusions, because it makes the epipolar constraint more robust by means of the trifocal tensor [19]. Quadranocular stereo (two pairs) videos are used for 3D reconstruction of reefs and benthic structures [91].

Using pre-calibrated camera system, the exact relative position between cameras is known *a priori* and the obtained 3D point cloud is already in real scale.

When the camera parameters are unknown, there is a higher degree of ambiguity in the reconstruction – it can only be determined up to a projective transformation of 3D space. By recovering the whole projective structure starting just from point correspondences in multiple views, one is able to compute a *projective reconstruction* of the cameras and the scene. Note that this can be done without any knowledge of the camera intrinsics, and it makes possible to reconstruct 3D scenes from uncalibrated sequences.

This projective cameras and scene differs from the actual cameras and scene (often referred to as a Euclidean or *metric reconstruction*) by a projective transformation. There exists classical techniques to transform a projective reconstruction to a metric one by computing this unknown homography – this is called *auto-calibration* or *self-calibration* [92]. Please refer to [19] and [21] for more details.

### ***Structure from motion***

Most of the structure from motion techniques used for sequence of unknowns cameras start by estimating the fundamental matrix if the correspondence is performed in two views or the trifocal tensor in the case of three view correspondences. The trifocal tensor and the fundamental matrix present the same role in the three-view and two views case respectively [19].

Most of the approaches used for large-scale structure from motion compute in an incremental way the reconstruction of the cameras and the scene, though in the state of the art are present various approaches that perform the reconstruction simultaneously [93, 94].

The SFM process developed by [21] starts from an initial reconstruction recovered from two views, then other cameras are added making possible to reconstruct the scene incrementally. A projective bundle adjustment [95] refines all the camera parameters and the 3D point previously computed, by minimizing the reprojection error.

The obtained reconstruction is determined up to an arbitrary projective transformation. Therefore, it is necessary a method that can lead to a metric reconstruction determined up to an arbitrary Euclidean transformation and a scale factor. This can be done though the imposition of some constraints on the intrinsic camera parameters (self-calibration) followed by a *Euclidean bundle adjustment* to determine the optimal camera parameters.

#### Multi view stereo mapping

In multi-view stereo algorithms [96] the object's colour or texture is used to compute dense correspondence between pixels in different calibrated views, then the depth of a 3D point in the scene can be recovered by triangulating corresponding pixels. Image matching algorithms often fail to find robust correspondences because of ambiguities due to texture-less regions, problems due to occlusions or non-Lambertian properties of the surface.

Stereo matching algorithms represent the object's shape using a disparity map, which represents the distance  $d$  (disparity) between a pixel  $p_1(i, j)$  and its correspondent  $p_2(i+d, j)$  along an

epipolar line. This is a 2D search problem because the images can be previously rectified in order to ensure that corresponding pixels lie on the same scanline (epipolar line).

In the multi-view case, the correspondence between matched pixels is represented by a depth-map that is the depth from a particular viewpoint. The disparity map computation can be viewed as a pixel labelling problem, which can be solved by energy minimization methods such as graph cuts [82] and loopy belief propagation [97].

The (*dense*) *correspondence problem* in computer vision involves finding, for every pixel in one image, the corresponding pixel in the other image. An individual pixel value is subjected to the presence of noise and alone is not distinctive enough; thus, the similarity is often computed using a support window around a pixel. Typically it is an  $n \times n$  square window centred in the current pixel, but some methods make use of adaptive windows [98, 99]. Different windows-based similarity functions based on difference and correlation measures are used for this task.

One of the most common pixel-based matching methods, is the Sum of Absolute Differences (SAD) [100] given by the expression:

$$\sum_{(u,v) \in W} |I_1(x_1 + u, y_1 + v) - I_2(x_2 + u, y_2 + v)| \quad \text{Eq. II-15}$$

Another common metric is the Sum of Squared Differences (SSD) [101] given by the expression:

$$\sum_{(u,v) \in W} (I_1(x_1 + u, y_1 + v) - I_2(x_2 + u, y_2 + v))^2 \quad \text{Eq. II-16}$$

The Normalized Cross Correlation (NCC) [101, 102, 103] is given by the expression:

$$\frac{\sum_{(u,v) \in W} I_1(x_1 + u, y_1 + v) \cdot I_2(x_2 + u, y_2 + v)}{\sqrt{\sum_{(u,v) \in W} I_1^2(x_1 + u, y_1 + v) \cdot \sum_{(u,v) \in W} I_2^2(x_2 + u, y_2 + v)}} \quad \text{Eq. II-17}$$

The Zero-mean Normalized Cross Correlation (ZNCC) [104] is given by the expression:

$$\frac{\sum_{(u,v) \in W} (I_1(x_1 + u, y_1 + v) - \bar{I}_1) \cdot (I_2(x_2 + u, y_2 + v) - \bar{I}_2)}{\sqrt{\sum_{(u,v) \in W} (I_1(x_1 + u, y_1 + v) - \bar{I}_1)^2 \cdot \sum_{(u,v) \in W} (I_2(x_2 + u, y_2 + v) - \bar{I}_2)^2}} \quad \text{Eq. II-18}$$

SAD and SSD produce a value of zero for identical support windows, but are not normalized because the similarity measure depends on the window extension.

Although it is the most expensive in terms of computational time, contrary to SAD, SSD and NCC, the ZNCC tolerates linear intensity changes providing better robustness and should be used when

brightness differences are present in the images. When a relevant noise is present in the images, similar *texture-less* areas provide high similarity scores using SAD and SSD, when the ZNCC instead produce a low similarity score because it tries to correlate two random signals. For more details about similarity measures in the context of stereo matching, please refer to [105].

### ***II.1.5 Imaging constraints in Underwater Environment***

Underwater environment can seriously compromise imaging results due to a series of problems that arises because of water properties and operative conditions. Scattering and absorption in the medium decrease image contrast and attenuate light intensity with consequent loss of details and colour alteration. Numerous studies and tests have been carried out about water influence on lights and colour [106, 107, 108, 12] and on registration and reconstruction techniques.

#### ***Absorption and Scattering***

Absorption and scattering are phenomena characterizing the light propagation in aquatic environment (Figure II-12). The former is the light disappearing from image-forming process, especially the low frequencies of the spectrum; the latter consists in different effects due to the direction change of photons colliding with particles suspended in water column:

- *Backscattering:*
- *particles reflecting light back to the camera appears as bright spots in the image.*
- *Forward scattering:*
- *particles deflect lighting rays so creating blurring and cluttering effects.*

The absorption and scattering of light through the medium of water was first understood in a physics-based context with the pioneering work by Duntley [109]. Duntley showed that the propagation of light underwater suffers from a wavelength-dependent exponential attenuation. In more recent years, McGlamery [110] investigated the fundamentals of the image formation process by computer modeling the absorption coupled with the direct, forward, and backscatter light components. Jaffe [111] later extended McGlamery's work to determine the idealized vehicle lighting configuration for minimal backscatter and good scene illumination. His results for standard lighting configurations confirmed that large horizontal camera-to-light separations were desirable to reduce backscatter — the principle cause being the reduction of common volume between the camera FOV and volume of projected light. However, Singh [10] recently showed that there are

theoretical limits to the benefits of large camera-to-light separation as applied to practical vehicle configurations.

a) Backscattering



b) Forward scattering



c) Absorption



d) Attenuation



Figure II-12 Typical underwater imaging problem. Images taken from [112].

### ***Attenuation***

In conjunction with the constraint of minimizing backscatter, the rapid attenuation of light through water imposes additional challenges when collecting underwater imagery (Figure II-12). Light attenuation limits the altitude at which a vehicle can fly from the seafloor and collect imagery. As deep-sea vehicles are required to carry their own light sources, this constraint has implications in both minimizing terrain parallax effects and in generating large-area imagery since the constraining altitude is typically 3–10 m [10]. In addition, the light source moves with the vehicle, leading to non-uniform illumination and moving shadows — both of which pose additional challenges during image registration. Vehicles are forced to fly close to the seafloor where terrain

relief may be comparable to the camera to seafloor separation, inducing gross perspective changes. Also, each image encompasses a small area of the seafloor, reducing the overall FOV. For mosaicking, this implies that many images must be registered to increase the effective FOV, and that terrain distortion.

### ***Shallow water***

Acquisition in shallow water is another crucial aspect of submarine surveys. Natural illumination in shallow water can vary continuously during the operation and poses challenging aspects like the sunflicker effect, created from refracted sunlight casting fast moving patterns on the seafloor and high reflectance of some particular seafloors. Furthermore meadows of moving plants or schools of fish are more frequent in warmer and brighter waters than in deep seas.



*Figure II-13 Shallow water light effects. Images taken from [113] and [114]*

### ***Registration***

Image registration can also be more difficult with underwater imagery than with land-based acquired imagery. Unstructured surveys by vehicles with low-resolution navigation and heading inaccuracies are common. This results in imagery with gross motions between temporal frames, often with minimum overlap due to strobed lighting [115]. In addition, the types of imaged scenery can be vastly different ranging from highly 3D coral reefs [116] to featureless muddy bottoms [117]. Man-made features such as edges, corners, and parallel lines, prominent in land-based images, cannot be reliably expected to occur in underwater imagery. Researchers have also compared the effectiveness of different image enhancement methods [118] and key-point detectors [119] applied to underwater images.

### ***Scene illumination***

The illumination of the scene is an important aspect in underwater surveys. In effect, natural sunlight disappears within a very short depth beneath the surface of the water and the need for artificial lighting appears evident. Power budget limitations of AUVs are hence an important consideration in the design of imaging systems. The amount of energy expended in illuminating the scene has a direct negative effect on the endurance of these battery-limited vehicles [115]. Typically, AUVs cannot afford to put out the continuous lighting needed for video frame rates because it would come at the sacrifice of precious bottom-time. Rather, strobed lighting is often used to conserve power [120]. Additionally, the low amount of image overlap afforded by this illumination scheme precludes optical-flow image registration methods such as [121, 122]. Therefore, the unique energy constraints of AUVs are a major driver to be able to handle low overlap imagery (i.e., 15–35% temporal overlap).

## ***II.2 A Review of Underwater Technology***

In this chapter a review of commonly employed underwater technologies is given. The attention will be focused on three fundamental aspects in underwater operation: vehicles, navigation, and acoustic mapping. Finally, a last subsection treats about the SLAM topic that tackle the problem of concurrent navigation and mapping.

### ***II.2.1 Underwater Vehicles***

Underwater vehicles can represent a powerful tool for every kind of underwater activities. In many circumstances, their employment is indispensable because of hostile time and condition operation. Scuba diver operations are, indeed limited up to fifty meters. Moreover, the operation time is inversely proportional to the depth.

For these reasons, underwater robotic vehicles are widely employed in different discipline, especially those requiring high accuracy standards, like archaeology.

Even if they can noticeably differ from each other in size, costs, and capabilities, a general classification divides underwater vehicles in manned and unmanned (UUVs); the latter can be differentiated in Remotely Operated Vehicles (ROVs) and Autonomous Underwater Vehicles (AUVs).

### ***Manned Underwater Vehicles***

Manned submersible can dive up to known depths, directly driven by an operator, and sometime, hosting more than one human per mission. They can recover artifacts, manipulate the environment having scientist *in-situ*. Big disadvantages are the extremely high costs, both of vehicle and of deployment operation, and the limited underwater time.

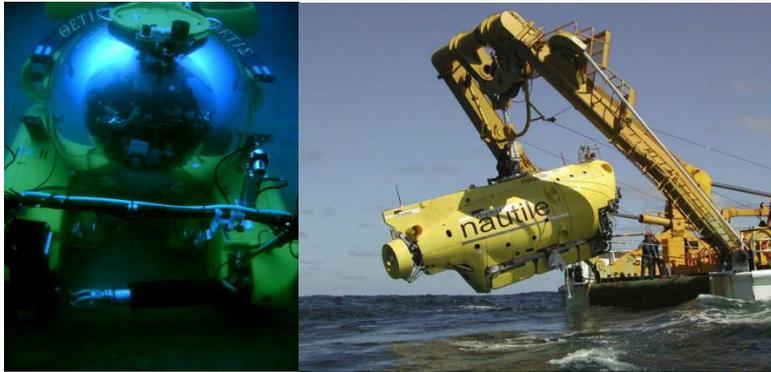


Figure II-14 Manned submersibles

### ***Remotely Operated Vehicles***

At the present day, ROVs are probably the most used platforms, able to operate with guidance from outside while doing a survey. There is a link between the surface and the underwater vehicle made up by a so-called umbilical cable, which is used for the vehicle control, energy supply and data exchange. They are characterized by having their own means of propulsion. Generally, ROVs can perform different operations uninterruptedly and are able to recover artefacts. Apart from operator fatigue, their main constraints are due to the umbilical that limits depths, range and mobility; moreover, for large depths the umbilical winch increases and consequently the size of the ship that can support it.

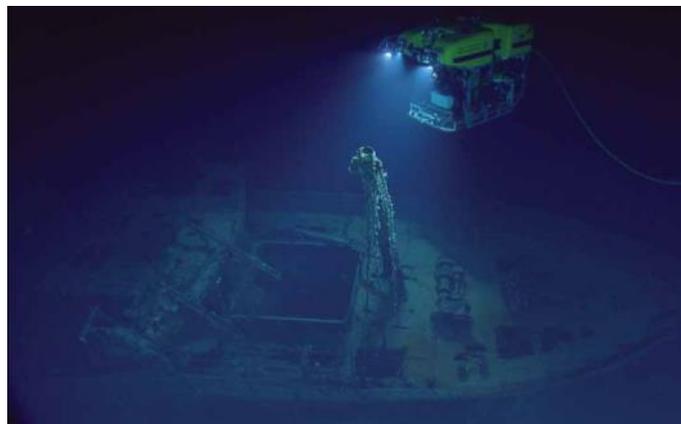


Figure II-15 ROV

### ***Autonomous Underwater Vehicles***

AUVs have a high degree of autonomy in the sense that they can navigate and locate itself using only their on-board sensors, and without permanent communication. AUVs have unmanned and untethered design, which makes them well suited to extended exploratory surveys requiring minimal user intervention and support [123]. Meanwhile, their autonomous free-swimming capability has added a new paradigm of ocean sampling to the scientific user community as demonstrated. They complement the capabilities of tethered remotely operated vehicles (ROVs). They have relatively low operational costs, also because of fast surveying capabilities. Main drawbacks are the limited battery autonomy and the self-position estimation inside surveyed scene that can drift in large areas. Moreover, capability of recovering and interventions are nowadays limited and object of studies [124].

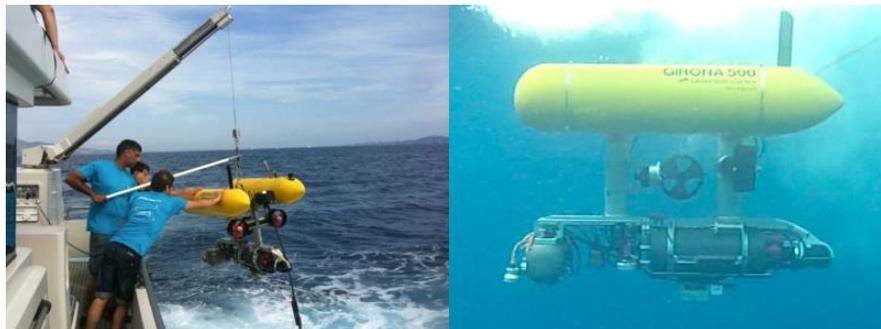


Figure II-16 AUV, Girona 500

### ***II.2.2 Underwater Mapping***

Underwater mapping problem has been tackled for years employing acoustic sensors. Sound wave can propagate in water for thousands of meters without suffering for attenuation and being practically not influenced by water turbidity. Thus, sonar system can survey huge areas with appreciable reliability, also because of typical underwater open spaces that limit undesired echoes and sound reflections. Sonars are technologically advanced devices: sophisticated designs of the transducer heads and the use of beam forming techniques make obtaining narrow beams that can produce remarkably precise measurements.

They can be classified in two categories depending on whether they produce only a set of range and bearing measurements or an acoustic image of the scene. Among those in the first category, the most commonly used are the Multibeam sonar and the Sub Bottom Profiler, whereas Sidescan sonar and Multibeam imaging sonar belong to the latter one.

### ***Multi Beam Sonar***

This sensor is specifically designed to produce bathymetric maps of large areas of the seabed. It is composed of an array of hydrophones which can emit fan shaped beams towards the bottom and measure the range of a strip of points placed perpendicularly to the direction of the vehicle's movement. These measurements can be produced at a high rate and resolution.

### ***Sub Bottom profiler***

Sub Bottom profiler is a particular kind of echo sounder, operating by emitting a pulse from its transducer. When this pulse reflects off a surface, it returns to the sensor head and the time of flight can be measured and therefore, the distance estimated. These kinds of devices are usually mounted in a down-looking position to find the altitude of the vehicle with respect to the seabed. Sub Bottom Profiler pulse generally contains different frequencies. The lowest ones can penetrate the seabed, so giving a characterization of the layers of sediment or rock under the seafloor with high resolution and penetration.

### ***Side Scan Sonar***

This sonar is designed for imaging large seabed areas. Its mode of operation is analogous to that of multibeam echo sounders, but oriented to imaging tasks. While the sonar is moved along a survey path (either mounted on a vehicle or towed by a ship), it emits fan shaped pulses down toward the seabed across a wide angle perpendicular to the direction of the movement, producing a strip of echo intensity measurements.

### ***Multibeam Imaging Sonar***

Also known as *Electronically scanned imaging* and *forward-looking imaging sonar*, this sonar is equipped with an array of hydrophones which allows, with the emission of a single pulse, producing the complete acoustic image of the insonified area. This area is usually limited to a small sector in the front of the sensor, but can be scanned at very high rates. Its main drawback is the cost which can be around ten times the price of a mechanically scanned unit.

## ***II.2.3 Underwater Navigation***

The development of UUVs has offered numerous advantages, but has also presented new challenges. One of the crucial aspects of a mission performed with Unmanned Underwater

Vehicles (UUV) is their localisation within the surveying area. Whereas gyroscopes and depth gauges directly provide drift-free measurements of attitude and depth, only complex acoustic positioning systems, such as Long, Short or Ultra-Short Baseline (LBL, SBL, USBL) or GPS intelligent buoys (GIB) transponder units can obtain the absolute measurements of longitude and latitude, with high associated mission costs.

Dead-Reckoning (DR) techniques can be used with less expensive sensors like Inertial Navigation systems (INS) and Doppler Velocity Logs (DVL) to estimate the motion of the vehicle with respect to the previous position. They are becoming more and more popular, affordable, and decreasing in size. The main drawback is that the estimation of absolute positioning from incremental measurements is affected by drift, due to error accumulation. For more accurate survey on this topic please refer to [125, 126, 127, 128, 129, 130, 131, 132].

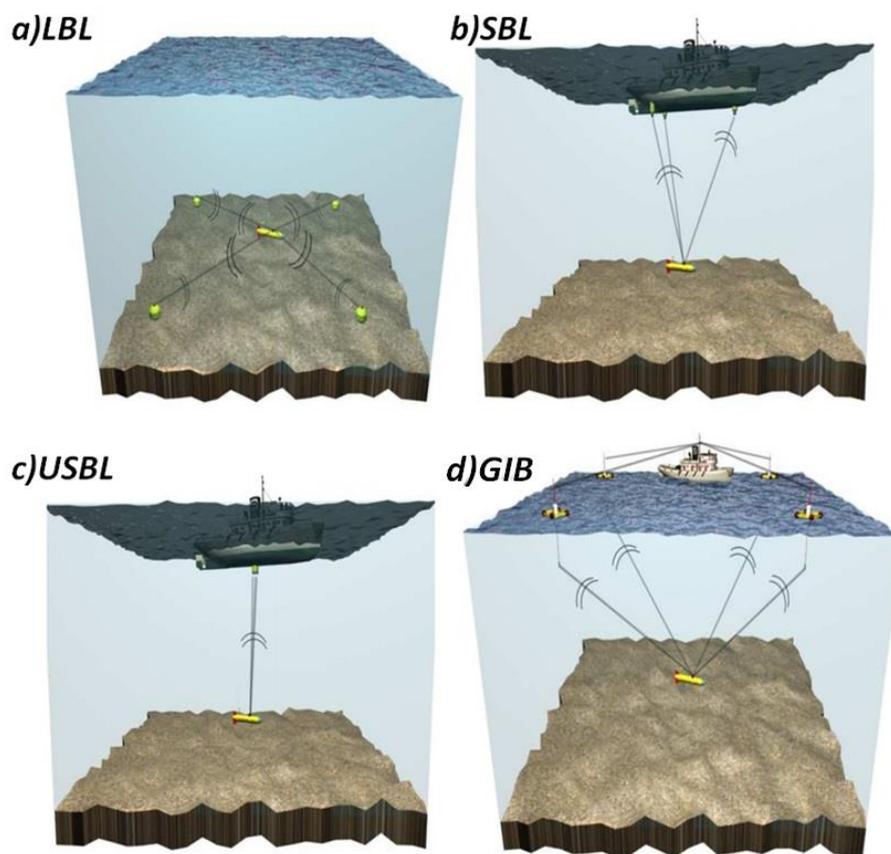


Figure II-17 LBL (a); SBL(b); USBL(c) and GIB systems (d). Images taken from [133]

### ***Long Baseline Navigation***

LBL system consists of a transponder network moored on the ocean seafloor. Transducers are mounted on the ROV or AUV, which sends acoustic signals to the underwater transponders. Each transponder of the network replies and a computer can determine the distances between each

transponder and the underwater vehicle. Then, the position can be calculated by triangulation of at least 3 of the distances measured from the beacons. If the ship receives and processes Global Positioning System (GPS) signals, the position can be georeferenced [134]. LBL systems typically operate at frequencies of 8 – 15 kHz covering ranges up to 5 – 10 km.

The main drawback of LBL systems is that the transponders network must be placed and calibrated before using them and recovered after the mission finishes making them suitable only for large-scale or long-term surveys (e.g. deep seafloor surveys in the Mid Atlantic Ridge). Although this sensor architecture provides a global position and bounded error within the network, the measurement frequency depends on the transponder distances since it uses sonar technology. Therefore, the longer the distance between transponders, the lower measurement frequency and accuracy (typically in the order of a few meters).

### ***Short Baseline (SBL)***

Short baseline systems use baselines approximately 1-100 meters and are usually composed of a set of receiver hydrophones mounted on the hull of a support vessel or in a rigid structure [135]. Usually, SBL systems are used to track an underwater vehicle from the surface, so that its position is computed and known at a surface ship but not at the vehicle (Figure II-17). In this case, the system provides only relative position estimates between the SBL hydrophones and the vehicle. If absolute position estimates are required, it is necessary to accurately know the absolute position and orientation of the structure (usually a support vessel) where the hydrophones are mounted [133].

### ***Ultra Short Baseline (USBL)***

In USBL networks, a sonar array is employed to determine the range and bearing to the vehicle [126]. This device consists of a transceiver, which is usually placed on the surface, on a pole under the ship, and a transponder/responder mounted on the UUV (Figure II-17). The device determines the position of the vehicle by calculating the range and angles obtained after the transmission and reply of an acoustic pulse between the transceiver and the transponder/responder. USBL are commonly used to track underwater vehicles, but cannot be directly used for vehicle navigation, as the position measurements are only known on the support ship carrying the USBL transducer. Some commercial units incorporate an acoustic modem capability so that position information can then be sent back to the vehicle through the acoustic channel. There is an alternative

configuration called inverted USBL that allows for vehicle navigation. In this case, the vehicle carries the USBL transducer head, and navigates by using an acoustic pinger with known position. The modest infrastructure required for USBL navigation (i.e., a hull mounted transducer) has resulted in its widespread utilization in a variety of scientific, industrial, and military underwater vehicles (e.g., [136, 137]). USBL systems require alignment calibration of the transponder and ship's positioning system (typically GPS), although the recent development of USBL transponders with integrated GPS systems could minimize this error [138]. Supplementing the vehicle range and bearing measurements with range and bearing measurements from a fixed sea floor transponder has been shown to improve the precision of USBL navigation [139, 140]. In addition to vehicle tracking, USBL navigation systems have been employed for the task of docking a vehicle to a transponder-equipped docking station [141, 135] .

### ***GPS intelligent buoys (GIB)***

The GIB system consists of four surface buoys equipped with Differential Global Positioning System (DGPS) receivers and submerged hydrophones. Each of the hydrophones receives the acoustic impulses emitted periodically by a synchronized pinger installed on-board the underwater platform and records their times of arrival (TOA). The buoys communicate via radio with a central station (typically on-board a support vessel) where the position of the underwater target is computed and displayed (see Figure II-17). The depth of the target is also available from the GIB system by coding that info in the acoustic emission pattern. The pinger emits two successive acoustic pulses during each emission cycle, the time delay between the two pulses being proportional to the pinger depth [133]

### ***Doppler Velocity Logs (DVL) and Inertial Navigation System (INS) navigation***

Recent advances have been made in the area of DR vehicle navigation with the advent of the bottom-lock Doppler velocity log (DVL). The DVL provides a measurement of seafloor-referenced vehicle velocity, which can be integrated over time to provide XYZ positional information [126].



Figure II-18 DVL sensors

The basic working principle behind these bottom-referenced velocity measurements is the acoustic Doppler Effect, which states that a change in the observed sound pitch results from relative motion. This change in sound pitch is directly proportional to the relative radial velocity between the source and receiver and can be used to recover seafloor-referenced vehicle velocity. Additionally, a DVL can also be used to measure water-referenced velocities. Unlike previous sensor networks, in order to track the vehicle position, DVL data has to be further elaborate, so giving, measurement drift that can pose serious problem in large areas and/or long surveys.

Most off-the-shelf DVLs use a Janus transducer configuration [142], which consists of four downward-looking acoustic transducers each oriented at  $30^\circ$  from the vertical [143](see Figure II-18). In this configuration, each transducer measures the sensor's velocity with respect to the seafloor as projected onto the centerline of its acoustic beam axis, resulting in four measurements of beam-component velocity. The beam component velocity measurements can be mapped to a standard Cartesian fixed instrument frame by the static  $4 \times 4$  instrument transformation matrix  $M$  parameterized by the transducer geometry [144], giving:

$$v_{sensor}(t) = \begin{bmatrix} v_{sx}(t) \\ v_{sy}(t) \\ v_{sz}(t) \\ e(t) \end{bmatrix};$$

Where  $e(t)$  is the estimated measurement error, whereas the other three terms are the velocities along  $x$ ,  $y$  and  $z$  in the local sensor reference frame. Through  ${}^wR_s$ , the  $3 \times 3$  rotation matrix between the sensor and the world frame arranged using information from orientation sensors, and ignoring the error term, it is possible to obtain the velocities expressed in a local frame aligned with world frame:

$$v_{world}(t) = {}^wR_s v'_{sensor}(t); \quad \text{Eq. II-19}$$

Integrating the Eq. II-19 navigation frame velocities it is possible to obtain a dead-reckoned bottom track DVL position.

$$x(t) = x(t_0) + \int_{t_0}^t v_{world}(\tau) d\tau;$$

Even if the dead-reckoning integration can be performed internal to the DVL using its on-board tilt and magnetic flux gate compass for orientation, it is typically computed exploiting measurements from vehicle's orientation sensors for better precision. For these setups, the error dependence in the integrated vehicle position can be less than 1% of total distance travelled [142]. Obviously, great attention has to be given to sensor positioning and calibration. In effect, the main part of accumulated error can be ascribed to small errors occurred in evaluating sensor relative poses; little deviation from designed position and orientation can dramatically affect the final result, especially in case long time and/or long track performed.

There are several types of sensors and methodologies to perform DR navigation, varying in complexity and their long time accuracy. Inertial Navigation Systems (INS) are one of the most widely employed. They are typically can provide position and attitude information with good short time accuracy and high update rates. INS systems are usually made up by different sensors (accelerometers, rate gyros, and magnetometers) and by a computer that process data giving the estimated pose of the vehicle. Their price is often prohibitive for low cost applications. However, it is possible to rely either on single sensors or on more economic sensor units like IMU (Inertial Measurement Unit) and AHRS (Attitude Heading Reference System) and implement dedicated DR algorithm. Both IMU and AHRS provide velocity, orientation, and gravitational forces, using a combination of accelerometers and gyroscopes, sometimes magnetometers. AHRS, thanks to its integrated processing system, is able to directly estimate and provide vehicle attitude and heading, whereas IMU just sends data to an external processing unit that solves the problem.

The most employed algorithm to incorporate multiple sensors for navigation is definitely the Kalman Filter (KF) originally published in 1960 [145]. Since then, the advances in digital computing made the usage of this filter practical and its applications has remarkably growth in number and complexity [146, 147]. Due to non-linearity of more engineering systems, the most influential development is certainly represented by the formalisation of its non-linear version known as Extended KF (EKF), firstly attempted by NASA researchers [148, 149]. EKF gives a good estimation results for the inherently more difficult nonlinear systems, where the first order Taylor series linearization sufficiently approximates the nonlinear motion characteristics [150]. Therefore, the

standard EKF has stayed the most popular filter for nonlinear estimation for long time. Nowadays, to overcome instability and divergence problems in high order systems, other design approaches are investigated for high performance applications [150]. For instance, amongst a series of possible variations, it is important to cite the Unscented Kalman Filter (UKF), developed by Julier, Uhlman and Durrant-White [151], which offers higher accuracy but slightly higher computational costs.

### ***II.2.4 Underwater SLAM***

SLAM is a fundamental task of robotic, aiming to concurrently map the working environment and positioning the vehicle inside of it. This issue, a typical chicken-or-the-egg problem, represents the aim for fully autonomous robot operations, expecting to give them the capability of performing intervention without *a priori* knowledge of the scene. The ideal robot behaviour should not differ from the human one: visiting a totally new environment, a series of recognisable landmarks are saved, in order to be used to reconstruct and link different portion of the scene, and also to recover self-position after got lost. The main problem is the sensor inaccuracy that brings to the erroneous robot positioning and, consequently, a distorted landmark and scene evaluation.

Works by Moutarlier and Chatila [152] and Smith, Self, and Cheeseman [153, 154] can be considered the first examples of SLAM. Since that time, significant advances have been made in dealing with several fundamental issues such as environmental scalability [7, 155, 156, 157, 158, 8] (i.e., how many landmarks can the robot maintain in its map), data association [159, 160, 161] (i.e., the problem of establishing land mark correspondence to measurements), and map representation [162, 163, 153] (i.e., how to model the environment or landmarks within it). These advances have led to the demonstration of impressive large-scale autonomous map making under challenging circumstances including large cyclic environments and poor odometry [7, 8, 9], and represent a significant fundamental achievement in our collective understanding of navigation with mobile robotics.

Hence, algorithms SLAM purposes have been widely investigated and tested in air, but the possibilities of visual techniques, are finding growing interest amongst researchers for underwater environments, too, also because the possibility of loop closure allows to significantly reduce the final error and uncertainty [164]. Unfortunately, most of on-land methodologies are impracticable in water due to medium characteristics that remarkably limit the employment of electromagnetic

waves. Moreover, terrestrial scenes mapped are usually rich of simple human made structures, that ease the landmark recognition process.

For this reason, SLAM in underwater environment has incited new research lines even if this remains a not extensively explored field. All information gathered by the whole UUVs' sensor equipment and by visual systems can be exploited and combined for 3D scene reconstruction and position tracking. Different authors made promising integration of optical acquisition tools with positioning sensors [165] or acoustic mapping devices [166, 167]. It is possible to find examples of tasks like seabed analysis, mapping, and 3D reconstruction performed exploiting different versions of EKF to merge data from cameras [12] or echo-sounder [11] with information from different navigation sensors carried by UUVs.

The Sparse Extended Information Filters (SEIF), previously implemented by Thrun et al. [168] moving from the information form of the EKF, has been used by Eustice et al. in [164, 169, 170, 171] for SLAM purpose on a underwater dataset, merging data from cameras and navigation sensors.

### ***II.3 Underwater Archaeology Applications: Related works***

Underwater environment represent an attractive realm for various knowledge field. In literature, it is possible to find interesting applications of underwater archaeology [172, 173], biology [174], geology and many others.

Underwater archaeology researchers need to survey submerged areas in order to obtain maps and 3D models of findings, artefacts and also entire sites [175], both through optical and acoustic technologies. In [176] an automatic, but very slow pre-processing methodology is proposed, intended to detect the best combination of image corrections and feature recognition algorithms. Passive photogrammetry and Structure from Motion (SFM) have been already applied in marine environment for seabed reconstructions, sometimes related to archaeological findings. Some example of seabed reconstruction from a video taken by a camera mounted on a Remotely Operated Vehicle (ROV) are shown in [177] and in [178].

In [179], researchers propose a methodology based on open-source algorithms and software tools to reconstruct the Mazotos area, trying to satisfy different requirements of speed, accuracy, cost and easiness of use for documentation purposes. Using a digital camera the captured images have been elaborated by using open source softwares (PMVS and Blundler), widely and effectively used in air for multi-view 3D reconstruction, thus without take into account the optical refraction

model. The site of Mazotos lies in deep water, so factors of disturbance like turbidity, variation light condition, and presence of marine flora have not been considered in the experimentation.

In [172] an Hellenistic shipwreck consisting mainly of a cargo of amphorae has been surveyed; after having conducted the camera calibration procedure, a photogrammetric technique has been applied on a set of images captured by a scuba diver, corrected and scaled thanks to a previous trilateration adjustment with tape measures. The obtained Digital Surface Model have been used to properly place the 3D theoretical model of the amphora.

New methodologies and underwater technologies have also been investigated in different works. Using a submarine robot, a technique similar to [172] has been previously employed in [180]; in particular, an underwater photogrammetry approach has been used to survey the archaeological site of Pianosa, where the Grand Ribaud Etruscan Wreck lies in deep water [181, 182], in almost impossible conditions for a diver. The digital photogrammetry method employed was based on a non-metric digital camera, mounted in a waterproof housing attached to a bar on the submarine robot. The huge amount of collected data has been used above all for cataloguing and remote visualization purposes. Authors in [183] tested some novel methodologies and tools applied to archaeological surveys on the shipwreck of Dolia in the Tuscan Archipelago, whereas [184] illustrates a project intended to provide methodologies and tools for the mapping, preservation and virtual exploitation of underwater and coastal heritage.

Some example of seabed reconstruction from a video taken by a camera mounted on a Remotely Operated Vehicle (ROV) are shown in [177] and in [185].

In [186] a new type of relief was carried out to document the phases of the restoration of a room paved with opus sectile in the baths of Punta Epitaffio (-5 m) using *Naumacos L1* scanner laser.

### ***III 3D MAPPING FOR DOCUMENTATION AND MONITORING OF UNDERWATER ARCHAEOLOGICAL SITES***

---

As previously mentioned, 3D models and maps are widely employed in various application. The ever-increasing interest in 3D vision and reconstruction, and the related technological development, remarkably influenced all research fields and not only Computer Science or Engineering.

3D models are useful for monitoring the current state of an object or area and for related intervention planning. More relatively to archaeological purposes, it is a very appealing and practical tool for conservation intervention and data storing, reference and reuse, but also for ease enjoyment by general non expert public, in museum or even on the web.

In particular, the underwater archaeology represents a new challenging field. Submerged Cultural Heritage realm would extremely benefit from these kinds of technologies. However, due to underwater environment intrinsic constraints, it requires the implementation of dedicated tools and the development of *ad-hoc* methodologies and software.

This section starts with an overview about the *CoMAS* project, which directly tackle these issues, and about the specific project test bed of Baia Underwater Archaeological Park.

Some tests performed in the area gave interesting results in 3D mapping. In particular, we illustrate the 3D reconstruction of two *Villae* from the archaeological site, and focus the attention on some particular applications of virtual models as means of Cultural Heritage dissemination and valorisation, and as conservation, monitoring and planning tool for archaeological purposes.

#### ***III.1 Underwater restoration: the CoMAS Project***

During the last few years, the scientific community and international Cultural Heritage safeguarding Offices have shown an ever-growing interest towards submerged Heritage and towards the need for promoting, protection, and conservation, possibly *in-situ*, of underwater findings. The *CoMAS* Project (“*CO*nservazione programmata, *in-situ*, dei Manufatti Archeologici Sommersi”) grows out of these needs, and aims to bridge the gaps that currently hinder the conservation planning of submerged archaeological artefact. In particular, the lack of knowledge and above all of techniques and materials specific for underwater interventions moves the research activity towards the implementation of new dedicated methodologies and tools for

restoration and conservation of underwater archaeological areas. The project test-bed is the Underwater Archaeological Park of Baia, a valuable archaeological site lying in shallow water, not far from Naples, and with a series of characteristics that make it challenging for underwater mapping, as described in the next subsection.

The CoMAS Project started in July 2011 and extensively worked on both materials and tools for archaeological interventions, so gathering researchers from archaeology, mechatronic, and material science in order to achieve excellent results combining different proficiencies.

Putting aside the details of interest in new materials suitable for artefacts protection that does not pertain to the objectives of this thesis, we focus the attention on innovative intervention tools and methodology involved in the project.

Nowadays, operations on underwater findings are performed almost exclusively by qualified scuba divers, which seem not to take advantage from technology progress that involved the great part of knowledge fields. In effect, the available tools are quite rudimentary and negatively affect costs and time of operations.

In CoMAS, a series of instruments have been designed *ad-hoc* for underwater studying and operation purpose. These new instruments are oriented to both human and machine use. In effect, the wide-spreading employment of underwater vehicles (UV), both manned and unmanned, also involved the archaeological field, urged by hostility of subsea conditions.

Apart from machine devices, also the 3D reconstruction of structures and objects represent another important tool for the underwater realm. In effect, the impressive high detailed models, achievable with off-line 3D reconstruction techniques, represent a very appealing tool for both researcher and lovers. In particular, in Underwater Archaeology, virtual models are easy to catalogue, study and disseminate. A virtual representation can be comfortably studied and enriched with metadata; it can be used for intervention planning and simulation and also for virtual restoration. Lovers and museum visitors particularly appreciate to access multimedia contents and, above all, really enjoy virtual exploration (or visualisation) of virtual sites (or smaller findings) and the possibility to compare the current conservation state with the digitally restored artefact version. Unfortunately, underwater imaging is a challenging task and requires different precautions in order to overcome the typical problems of the subsea environment related, above all, to light propagation in water, turbidity and operational constraints.

We tested that 3D dense stereo mapping can give satisfying results also in underwater applications. Examples of how an accurate virtual model can document the conservation state and

the results of the restoration process are given in the following sections, after the presentation of Baia archaeological site.

### III.1.1 CoMAS case study: Ancient Baiae.

The underwater archaeological site of Baia is located few kilometers North of Naples (Italy) (Figure III-1) and belongs to the imposing volcanic complex of the Phlegrean Fields.



Figure III-1 Geographical localization of Underwater Archaeological site of Baia (Italy).

Some explorations and underwater researches conducted in the site of *Punta Epitaffio* and in the Gulf of Baia have identified the ancient site of *Baiae*, which is spread into the sea up to 400 m. approximately from the coast, whereas 60% of the site is on land. In Roman times, the current gulf was occupied by the *Baianus Lacus*, or *Portus Baiarum* ("Lake" and "Harbour", respectively), as reported by Seneca, Tacitus, Martial, Pliny the Elder and Florus, which later disappeared due to bradyseism [187] (Figure III-2).

*Baiae* was a famous seaside town much prized in Antiquity for its temperate climate, beautiful setting and the properties of its mineral waters which have been exploited since the second century B.C. . It was the most popular resort of Roman Aristocracy and the Imperial Family up to the end of the third century A.D., when the bradyseism caused the submersion of the city. Now the remains of villas lie underwater along the seaside to a distance of 400-500 meters. Next to luxurious maritime villas and imperial buildings it is possible to find more modest residences, private *thermae*, the so-called *tabernae* and all those structures that characterize the Roman age cities.

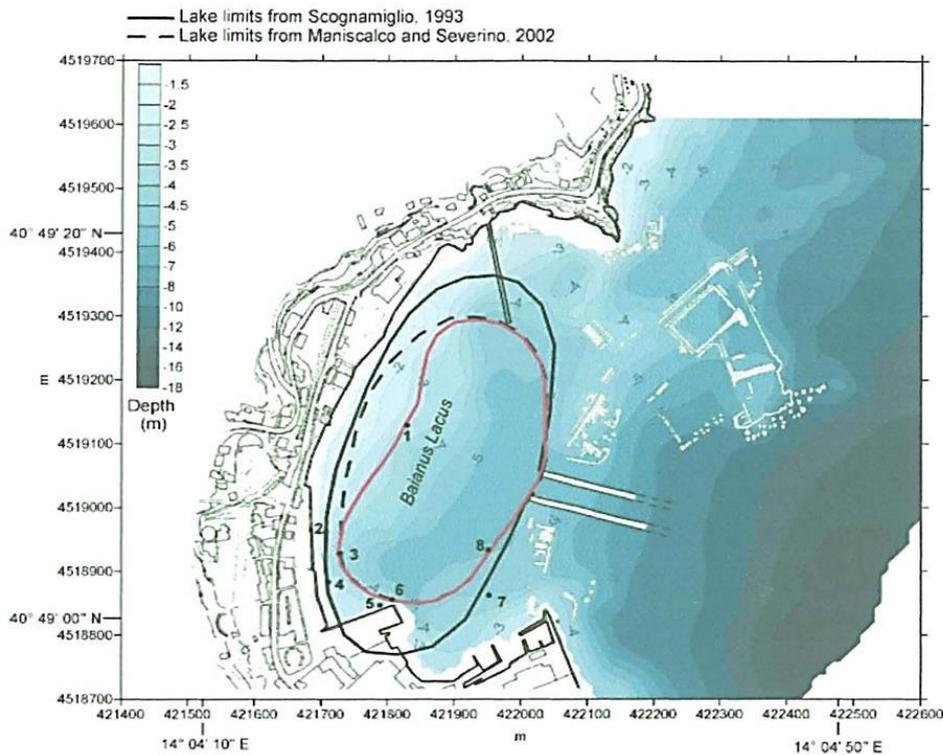


Figure III-2 The Baianus Lacus.

The site is recognizable from the ancient ruins located on the banks and the access channel at about -6/8 m., in the centre of the roadstead, cutting the Isthmus that separated it from the sea. To the east of the *nymphaeum* of Claudius (I<sup>st</sup> c. A.D., exc. 1981/82) and the stone-paved road toward the foothills (I<sup>st</sup>-IV<sup>th</sup> c. A.D., exc. and rest. 2008), there were some *thermae* (I<sup>st</sup>-III<sup>rd</sup> c. A.D.) and a villa dating back to the age of Hadrian (117-138 AD) located south-east of *Punta Epitaffio*, built on the remains of an older villa (late I<sup>st</sup> B.C.-early I<sup>st</sup> c. A.D.), attributed to the *Pisones*.

The vast complex featured *thermae*, gardens, a maritime quarter, with residential rooms, cisterns, fish ponds and two landing sites, protected by breakwaters, partly similar to *Villa Adriana* (Tivoli), and it has probably been built under the direct intervention of Hadrian. An element of great interest could be a large peristyle, *viridarium*, fish pond or thermal pool (still waiting for archaeological confirmations), like the one found in Bath (UK). Other remains have been found beyond the channel and on the seabed in front of the Castle of Baia.

As it is well known, the ancient submerged remains are subjected to erosion phenomena, manmade damage, and colonization by biological agents. All these factors are problematic for restoration and preservation, and involve high maintenance costs, as reported during the ISCR experimental restorations (2003-2012) and the P.O.R. of *Baia Sommersa* (2008) [188]; the restored paths are currently in need of maintenance.

In the past years the ISCR with the project 'Restoring Underwater' has restored some sectors of ancient villas and structures situated in the Park of Baia [189, 190]: the *Villa con ingresso a Protiro*, the *Villa dei Pisoni*, the *Via Herculanea* and the *Edificio con cortile porticato at Portus Iulius*.

This case study is particularly critical because of the high level of turbidity, the heavy presence of marine flora and the lighting conditions changing typical of shallow water. All these disturbances cause a decreasing of the contrast and an alteration of the colours in the acquired images.

All these problems can seriously affect the 3D reconstruction and we were not sure about the possibility to use a multi-view photogrammetric technique for obtaining a sufficiently detailed model of the structures present in the site.

### ***III.2 Documentation of the conservation state and interactive visualisation***

In this section we show the test activity for a photogrammetric technique based on multiple captures for 3D model reconstruction of submerged structures.

As mentioned in the previous sections, the archaeological site of *Baiae* presents several peculiarities compared to other underwater sites. In effect, the above-mentioned *Baiae* environment conditions, in terms of water turbidity and presence of flora, can dramatically affect the use of a multi-view photogrammetric technique.

For this reason, a first feasibility study has been performed to test a 3D reconstruction technique available in literature to acquire a portion of the site elected as a case-study. Thus, the main goal of this study was to verify if a well-known 3D reconstruction technique based on multi-view stereo may be employed in the critical conditions of the *Baia* site.

Furthermore, the obtained model has been loaded into a software application that allows the visualization and the exploration of the room in order to study and simulate the possibility to adopt a ROV or an AUV for operating on the submerged structures.

The room selected as a case-study is a portion of *Villa con ingresso a protiro* (henceforth called *Villa protiro*) sized 5 x 8 meters. The room is delimited by a wall with an elevation ranging around 0.5 meters. There is one aperture on the west side and it presents several blocks of bricks spread over the floor. This particular room has been selected because it presents various materials (bricks, mortar, tile floors, etc.) and, at acquisition time, it was affected by a strong colonization of algae and other kind of bio-fouling, so constituting a very hard test case for the 3D acquisition techniques.

### III.2.1 Archaeological context

The test has been performed on the thermal room (*caldarium*) of *Villa protiro*. This *caldarium*, together with another room paved with a white mosaic, was restored in the 2003 and in 2009. The *Villa* lies at 5 meters depth; the rooms that composed the *Villa* extend for 40 meters on the road flanked by *thermae*, *tabernae* and other *villas*; however, its real size must have been larger. Two red-plastered masonry benches delimit the threshold, framed in stucco pilasters.

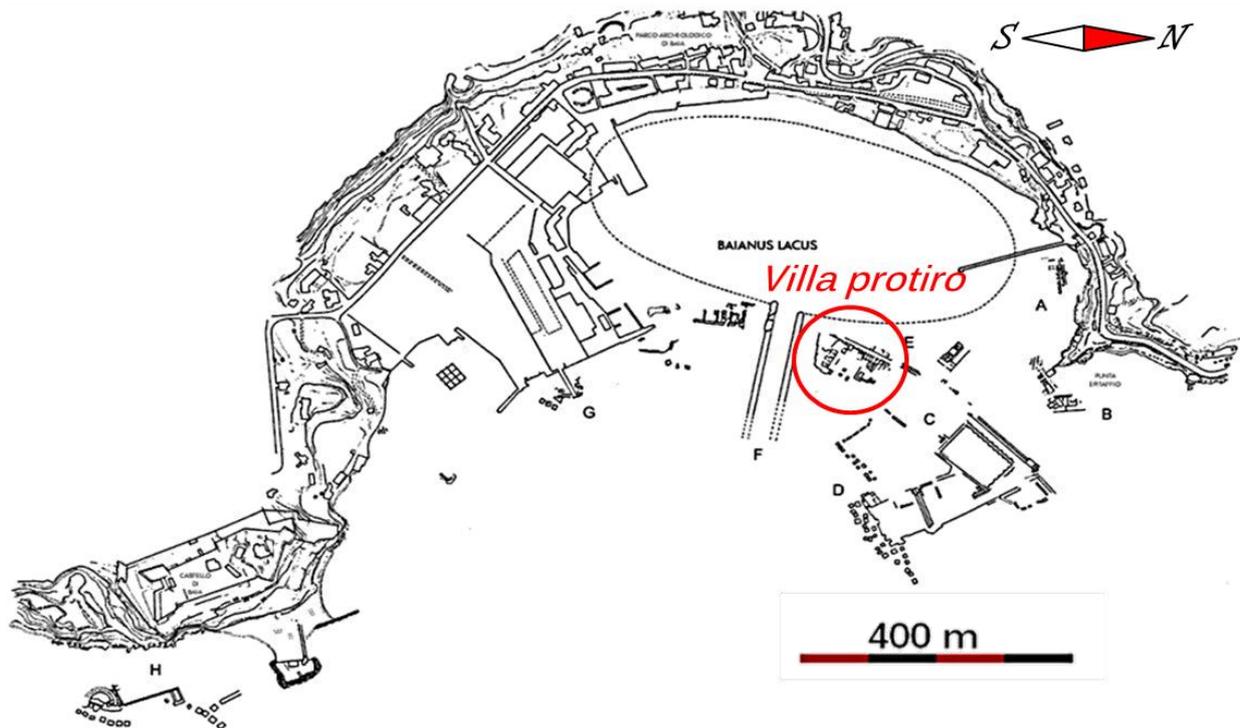


Figure III-3 Underwater Archaeological Park: detail of *Villa protiro*

The name *prothyrum* comes from the presence of two stuccoed columns shafts, no longer existing, that were placed on two short parting walls built in front of the threshold. The villa's floors and walls are largely decorated with marbles; almost all the surveyed rooms have marble-covered walls, and mosaic floors mostly made of white tesserae (e.g.: room no restored by ISCR in 2003). A white and black mosaic patterned with hexagons, rosette diamonds and plants, 2nd century ad, decorated a room on the Northeast of the entry hall; this entry hall had a sheets wainscot in red marble of *Tenario* with "Lunense" marble mouldings. The villa underwent a second construction phase represented by the building of a wide apsidal room south of the hall. The room had two floor levels, built from large white marble sheets, which were used as covers for the walls as well. The villa is also composed of other service rooms such as an area identified as a kitchen in the northwestern section, and a rectangular courtyard. A garden was located behind these areas, it

contained the *Alkamenes Afrodithe in kepois statue (l'Afrodite dei giardini)* recovered in 1993 and now displayed in the garden of *Campi Flegrei Archaeological Miuseum*.

### ***III.2.2 Experimental setup***

The equipment used to acquire the underwater pictures consists of a Nikon D7000 reflex camera with a 20 mm fixed lens. The camera is mounted in an underwater case manufactured by Ikelite equipped with a spherical housing. In addition, two underwater strobes have been used to acquire pictures of small portion of the site that, for the presence of shadows, are in very low lighting conditions. Another diver used a compact camera, precisely a Canon S100, mounted in a proprietary underwater housing, to record a series of full-HD videos in order to have redundancy of data reducing the risk of uncovered areas.

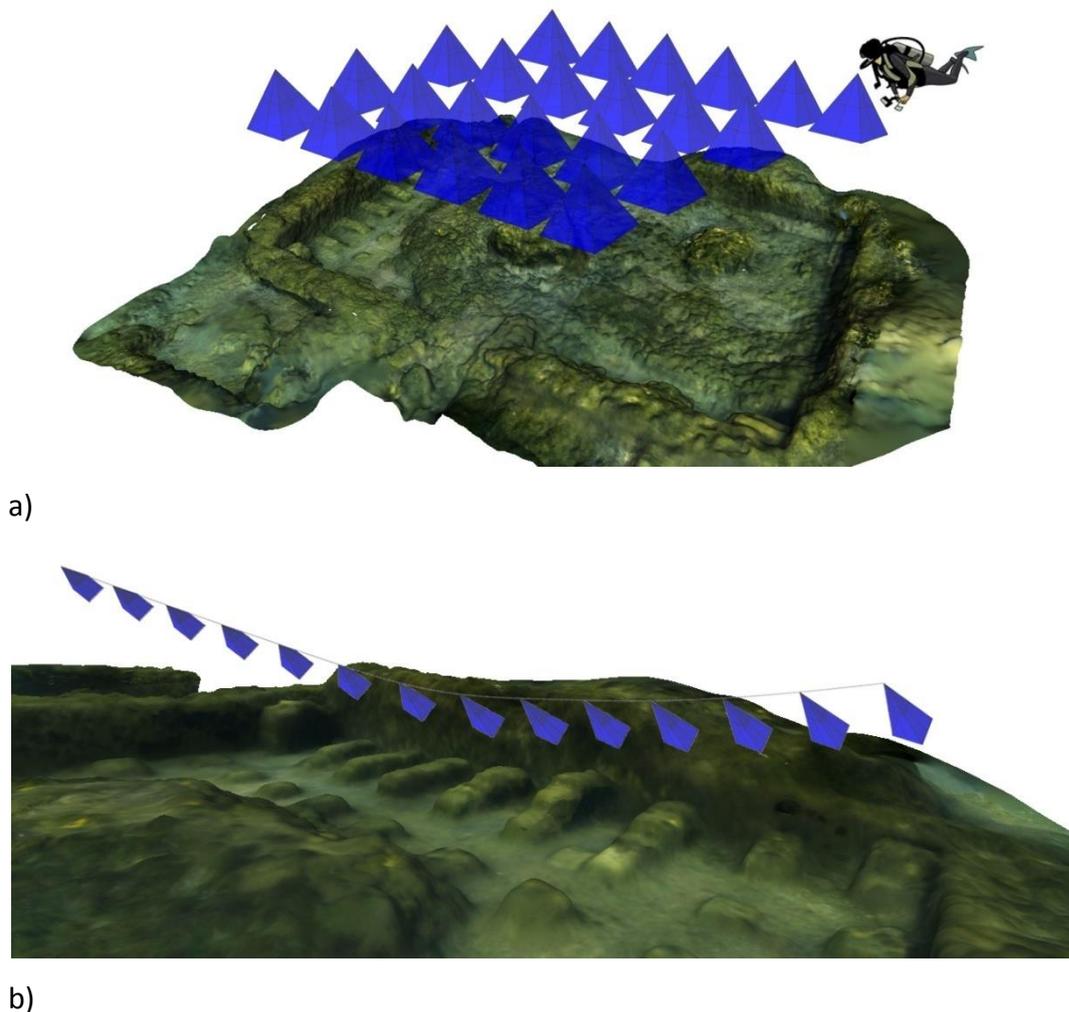
### ***III.2.3 Image acquisition and colour correction***

The acquisition has been carried out in two different dives, related to the North part and to the South part of the submerged area. A third dive has been planned in order to complete the missing areas. The manual mode has been used during the acquisition, setting values between 1/250 and 1/320 of a second for the exposure time and an aperture between f-5.6 and f-8 to ensure a sufficient depth of field. The automatic ISO sensitivity function has been used in order to ensure that exposure adapts to variable lighting conditions, taking advantage of the D7000 performance, capable of producing images with no noise even at high ISO. In order to test the effect on the 3D reconstruction, a set of images has been acquired using the pre-measured value obtained acquiring an underwater white balance panel *Lastolite* (Figure III-4).



*Figure III-4 Calibration panel used to manually correct the white balance.*

Following a standard aerial photography layout, the diver swims at a certain height/depth above the site (about 2 meter in this particular experimentation) taking overlapping pictures ensuring about a 75% of common coverage between two consecutive photos, following a straight line with constant speed in North-South direction (Figure III-5-a). Several overlapping stripes have been acquired covering the whole area. Another set of stripes have been acquired in East-West direction. The occluded areas like walls and all the features not visible in vertical pictures, have been acquired a series of oblique photographs (Figure III-5-b). In total about 700 images have been gathered.



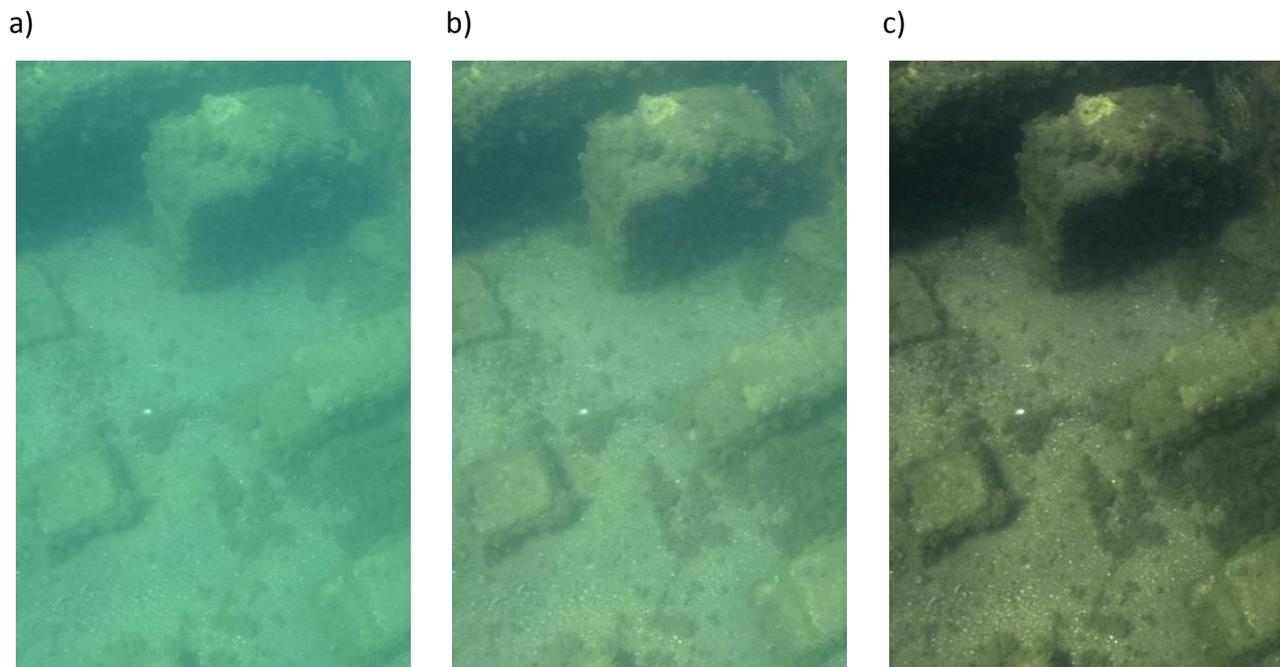
*Figure III-5 Image acquisition using a standard aerial photography layout (a) and oblique photographs (b).*

As previously discussed, underwater photography is not an easy task, because the environmental conditions lead to several problems related to visibility, refraction and changing of lighting. Underwater pictures generally suffer of light absorption, which causes some defects mostly on the

red channel that is the first component of the light spectrum that is absorbed, and this effect is notable already from few meters. The pre-processing of underwater images can be conducted with two different approaches: image restoration techniques or image enhancement methods [191, 118]. Image restoration techniques need some known environment parameters such as scattering and attenuation coefficients, while image enhancement methods do not require an *a priori* knowledge of the environment. In this study we implemented a time-saving methodology based on enhancement, using batch actions in Adobe Photoshop® to rescue the maximum amount of information from a set of defective and noisy pictures.

For this purpose, a representative image of the underwater scenario has been chosen as sample in order to perform retouching and equalization, applying a four-step procedure:

- a) preliminary histogram stretching to improve the contrast;*
- b) mixing of the colour channels to balance the missing information on the red one;*
- c) creation of a set of adjustment layers on the sample image, ranging from saturation enhancement for some missing hues, contrast masks, colour balancing and equalizing;*
- d) export of the statistics of the sample image on the other ones, with different mixing percentages (Figure III-6).*



*Figure III-6 Original uncorrected image (a), acquired image using pre-measured white balance (b) and final corrected image (c).*

### **III.2.4 Multi-view 3D reconstruction**

The 3D reconstruction pipeline starts from the orientation of the 722 pictures composing the whole dataset by using Bundler [23]. The first phase uses the SIFT to extract a set of descriptors across the collection of images that are subsequently matched. The output of Bundler consists of the intrinsic and extrinsic parameters for each camera pose and in a sparse reconstruction of the scene. We find that the colour correction step improves considerably the matching results. In effect, the full set of images has not been successfully oriented, and it has been necessary to divide the dataset into two groups, for the North and South part, that have been reconstructed separately. In particular, 384 images have been oriented for the North dataset and 116 for the South dataset. This is mainly caused by the sandy seabed present in the central part of the room, which makes difficult the extraction and matching of feature due to the low contrast. Furthermore, the lack of overlapping areas in the reconstructed surface has made it impossible to align the two blocks.

Moreover, the use of uncorrected pictures produces a low-quality texture making the model unusable for interactive visualization purposes.

Instead, the use of enhanced images, allowed orienting a subset of 533 pictures relative to the whole area, making possible to generate a complete 3D point cloud without aligning different meshes (Figure III-7).



*Figure III-7 Sparse point cloud and 533 oriented frames.*

The Bundler's outputs and the undistorted images have been processed with PMVS2 (Patch-based Multi-View Stereo) [192]. This algorithm estimates the surface orientation while enforcing the local photometric consistency, which is important to obtain accurate models for low textured objects or with images affected by blur due to turbidity in underwater environment. Furthermore, PMVS2 automatically reject moving objects, such as fishes and algae. The output is a dense point cloud (about 10 millions of points) with RGB information for each 3D point to create a dense point cloud of the whole site (about 10 millions of points) with RGB information for each 3D point (Figure III-8(a)).

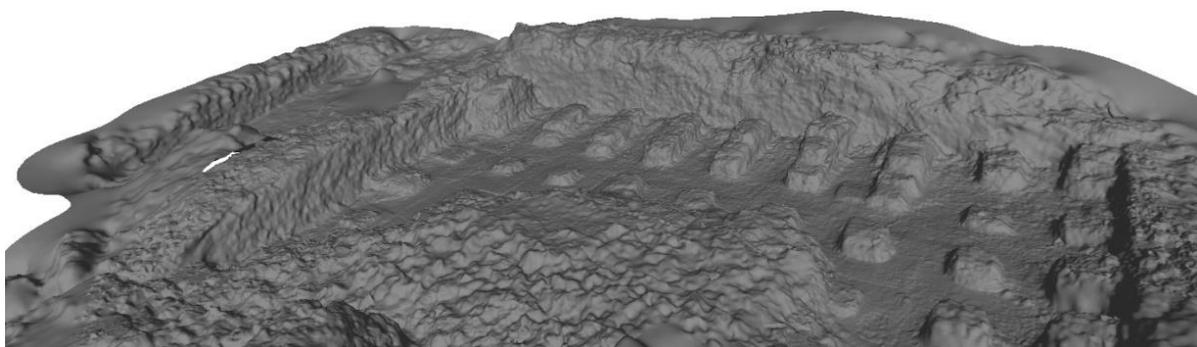
The point cloud has been elaborated by using Meshlab tools [193]. The first operation was a manual selection and deletion of unwanted areas. Then, a watertight surface with about 25 millions of triangles (Figure III-8(b)) has been obtained through a Poisson Surface Reconstruction algorithm.

The resulting surface has been subsequently decimated in a mesh of 6.5 million of triangles and 3 million of points, in order to be handled more effectively without losing details. Since the camera orientation procedure has been carried out with an unknown scale factor, it is necessary to scale the model by selecting two points with a known distance. In this experimentation a scale bar has been placed in the scene and reconstructed in order to evaluate the scale factor.

**a)**



**b)**

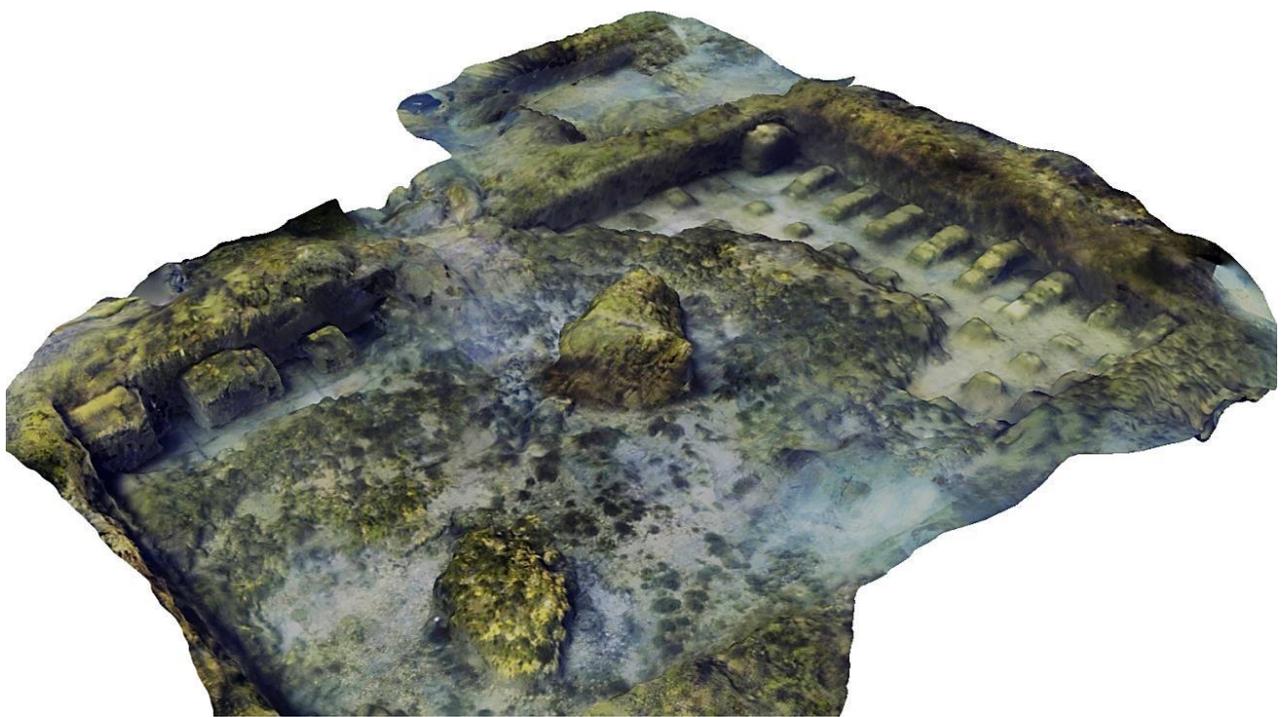
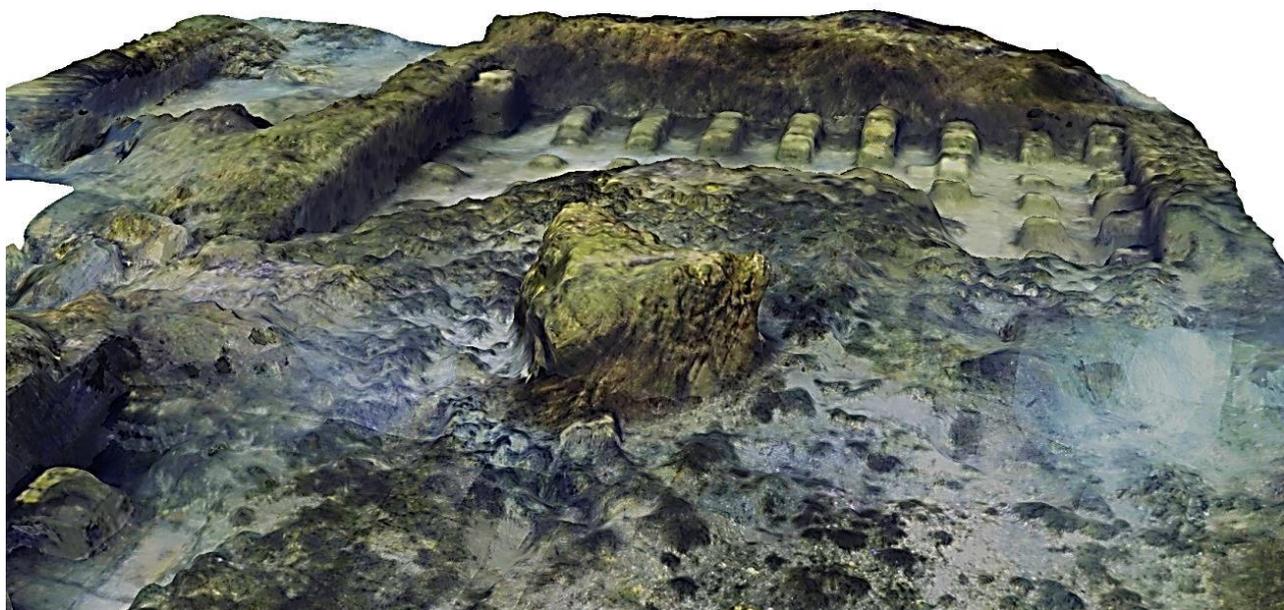


*Figure III-8 Reconstructed dense point cloud (a), Reconstructed surface(b).*

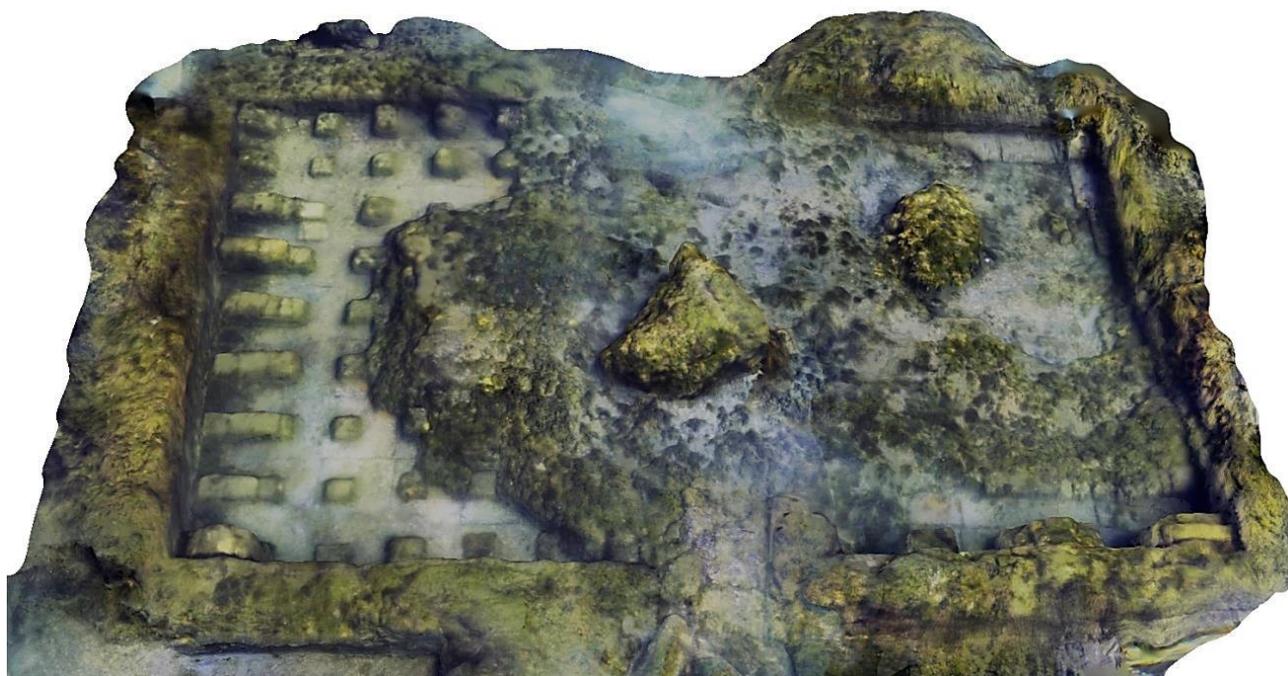
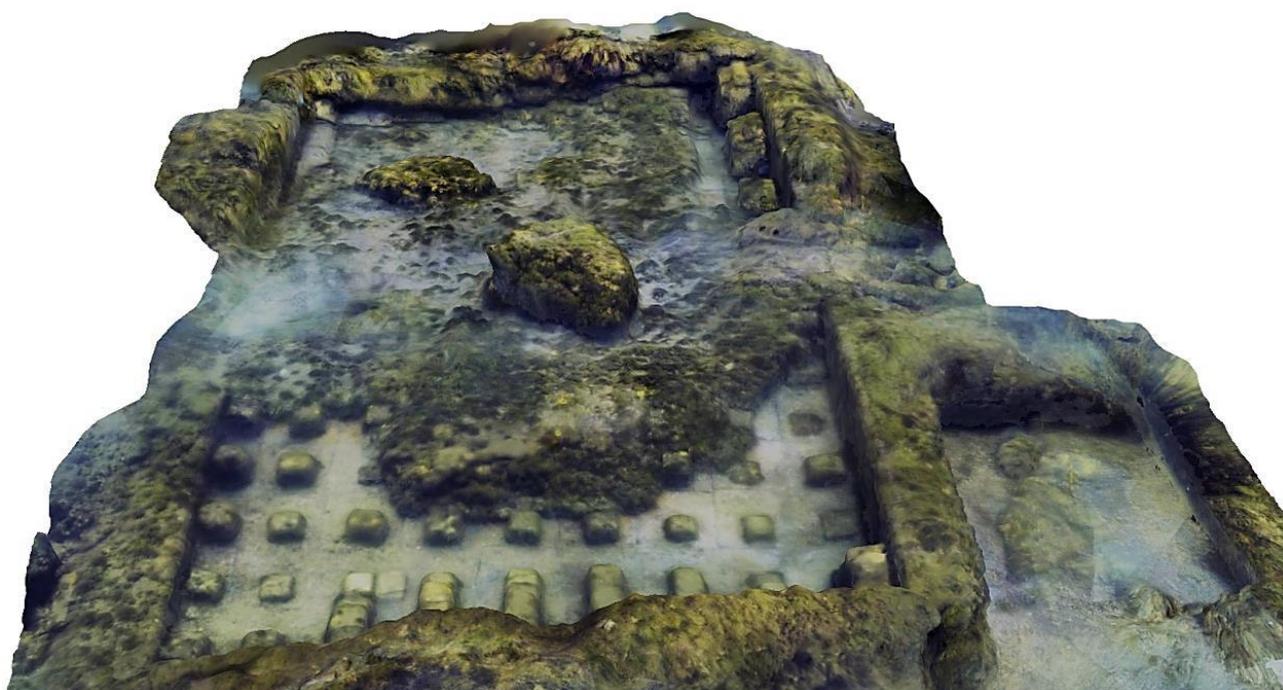
The last step consists in the application of the texture on the 3D surface. Colour information can be extracted directly from the coloured point cloud, but this method does not allow the creation of a high quality texture, because its resolution depends on the point cloud density. Moreover, since the enhancement is often performed to favour the feature extraction process by increasing the contrast without taking into account the fidelity of the colours (usually a single component or grey-scale images are used), the colour information stored into the pixels cannot be used.

Since the camera positions are known, the method chosen for texture mapping consists in the projection and blending of the high resolution images directly on the 3D surface. In particular, an image subset has been selected because the averaging among neighbourhood values during the blending on the images works better if a small overlapping area is present (blur effects are reduced).

The result of this procedure is a texture with a resolution comparable to the original images and characterized by a high quality level (Figure III-9 and Figure III-10).



*Figure III-9 Final textured 3D model (a).*



*Figure III-10 Final textured 3D model (b).*

### III.2.5 3D interactive environment

The 3D model of the Villa Protiro room has been loaded into a developed interactive application that tries to reach three different goals (Figure III-11):

1. support the dissemination of the results of the *CoMAS* project by giving the possibility to a large audience to explore a 3D underwater site reconstructed in an appealing and realistic way;
2. support the study of the archaeological site by giving the possibility to navigate in the 3D reconstruction of the room;
3. support the requirement analysis for the development of a robotic arm to be mounted on a ROV specifically addressed to cleaning and maintenance operations on the underwater site.

Since the 3D environment has three different goals, the application that has been developed offers three different modes (realistic, study and simulation) characterized by different interfaces and different functionalities. The realistic mode is addressed to a general public, the study mode is intended for archaeologists and the simulation mode has to support the collaboration between engineers and archaeologists in defining the requirements of the ROV arm. The 3D application has been developed in Virtools Dassault®. The first step in the development of the application is to import the textured 3D model. The file format chosen for transferring the data between Meshlab and Virtools was the *.3ds*, because of its high portability.

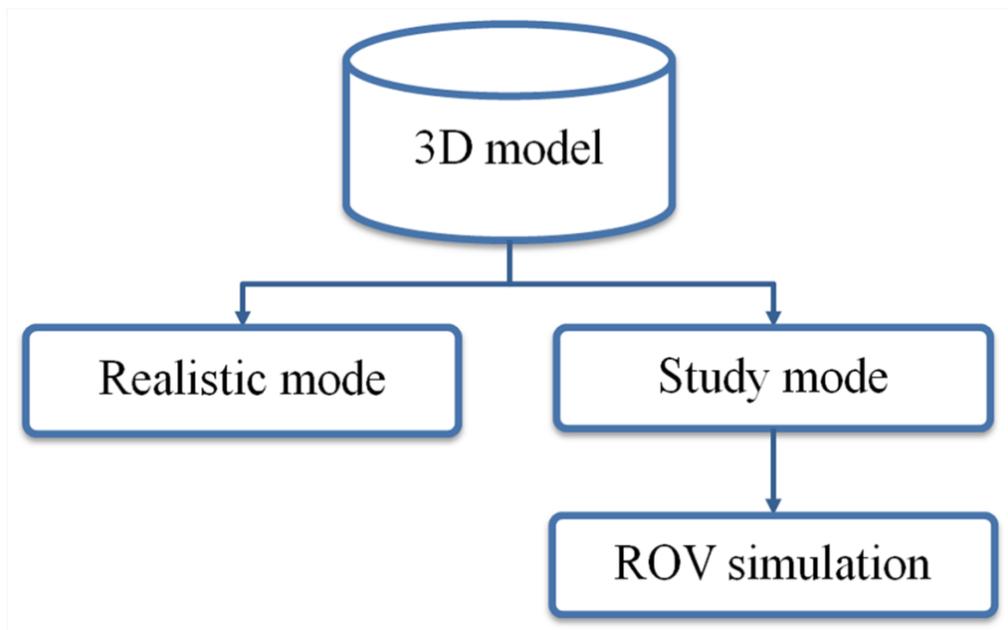


Figure III-11 Scheme of the three applications of 3D interactive environment.

### ***Realistic mode***

The realistic mode makes use of some computer graphics techniques to reproduce typical effects of an actual underwater scene. As visible in Figure III-12, reconstructed environment presents no ambient light and different types of artificial spotlights, caustics effects and a strong light attenuation simulating turbidity. In the scene, it is also possible to see fish swimming, moving fins and tail.

In real underwater environment, the density of the water causes the light to be absorbed with increasing distance, preventing gradually the vision of the scene. In addition, the absorption of the red component of sunlight increases with the depth, flattening the shade of the surrounding colours toward green. To simulate these effects we used a virtual greenish-blue fog. Caustic effects result from light rays reflecting or refracting from a curved surface and hence focusing only in certain areas of the receiving surface. In order to simulate this effect we combined and mapped two textures on the seabed material. The first texture is the one obtained by the 3D reconstruction process; the second one is an image of caustics on black background.



*Figure III-12: Screenshot of the application in realistic mode.*

The caustics motion on the seabed, due to wave motion of sea surface, has been simulated using a shader (a program used primarily to calculate rendering effects on graphics hardware) that implements a sine equation. The effects of brightness, fog and caustics were manually calibrated to obtain a view as close as possible to the actual one. Figure III-13 illustrates a comparison

between a detail of Villa Protiro room (a) and the same detail in a screenshot from the virtual application (b). The use of computer graphics techniques like fog, shaders, transparency effects, increases the resource consumptions, so the user can disable them in order to speed-up the application. The realistic mode has been also converted into a web application in order to promote the Baia site and ease its exploitation. The web version, obtained through the Virtools web player, requires a strong decimation of the 3D model in order to reduce the file size and the hardware requirements. The final version has a size of 62 Mb and requires a minimum of 3 Gb RAM and a Graphic Card whit 512 Mb dedicated memory. The decimation process has reduced the number of triangles by 90% of the original 3D model, passing from 3 million of vertex and 6.5 millions of triangles to 496.997 vertex and 613.272 triangles.

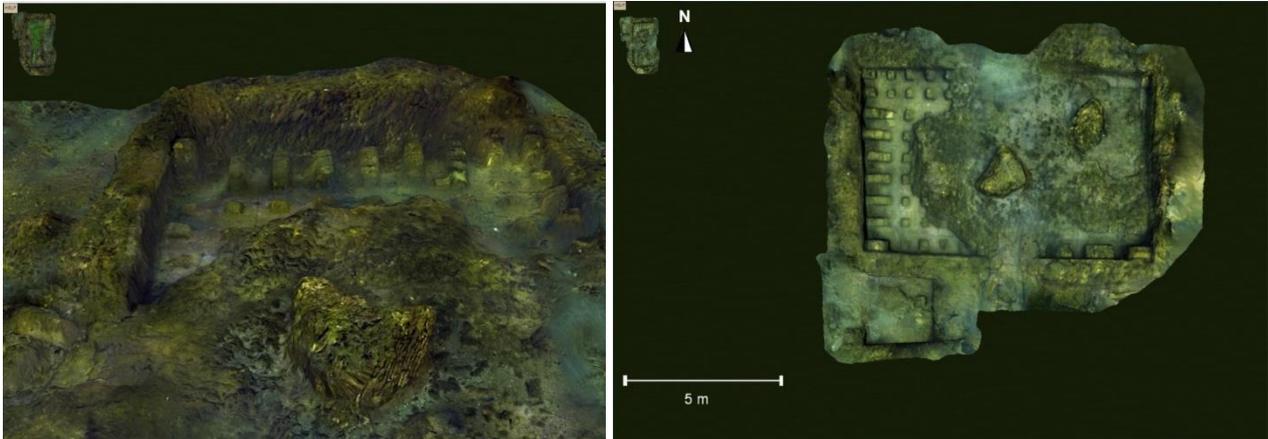


*Figure III-13 Top: photo of Villa in Baia, bottom: screenshot of the application in realistic mode.*

### ***Study mode***

For the study mode (Figure III-14) we used the model with the highest quality without any decimation. The rendering does not use any visual effect for turbidity or caustics simulation and only a strong ambient light illuminates the scene. This kind of visualization is obviously not realistic but it is intended to allow researchers to clearly observe and study the structures. The user can explore the site by choosing among keyboard, gamepad, or graphical user interface. A site

navigation map that shows the entire exploration area and the current viewpoint position is available whereas a compass clarifies the orientation of the map. Figure III-14 (right) shows the site in Top View default orientation, in which only pan and zoom actions are enabled.

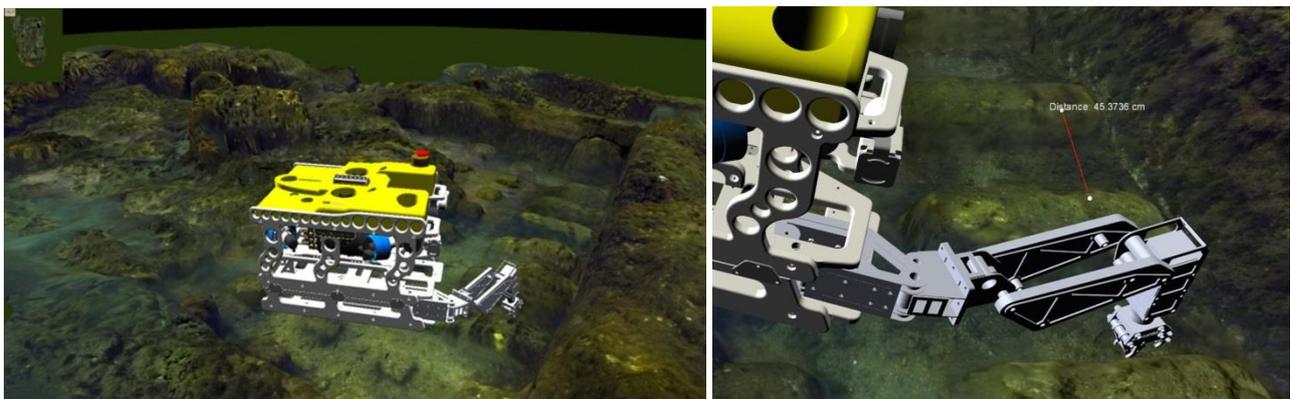


*Figure III-14 Left: screen-shot of the application in study mode; right: top-view of study mode.*

This default orientation includes a graduated bar as visual reference system for distances: the bar size changes according to zoom level so allowing incremental steps for the reference unit.

### ***ROV Simulation***

The ROV simulation mode keeps all functionalities of study mode visualisation. Furthermore, the ROV 3D model with its own arm has been imported in the application; a set of additional commands is provided to move and orient the ROV and the robotic arm and to provide distance measurement (Figure III-15).



*Figure III-15 Test of reachability for narrow edges and example of measuring.*

The robotic arm has three degrees of freedom and it is controlled by direct kinematics so user can set pitch and yaw for shoulder and pitch for elbow. Kinematics constraints are imposed by

inserting the range of rotation for each beam. Collision detection between the ROV and seabed model have been implemented. Since the collision detection is a time consuming task, we implemented the possibility to set it on the whole ROV or only on a particular portion (e.g. the body, the end effector etc.).

Moreover, it is also possible to choose whether the contact consequences (sliding effects) have to be simulated or not. We used a Box-Face collision detection by considering only the bounding box of the ROV, whereas seabed collision surface is simulated with its own mesh.

### ***III.3 Monitoring of the experimental cleaning operations***

In this study we try to highlight and exploit a new possible and interesting aspect related to the use of 3D models; in effect, to the best of our knowledge, this powerful tool has been used above all for surveying and virtual exploitation [194]. In this work, we aim to show how a comparison between 3D models of the same area, surveyed at different times (for example during a cleaning intervention) can provide important information both about the performances of intervention tools and about characteristics and conditions of flora and fauna colonization.

In this case-study, the multi-view stereo algorithms implemented in Agisoft Photoscan® [195] have been used to reconstruct the underwater scenario, making it possible to orient more than 1500 pictures and to compute the camera calibration parameters of the un-calibrated camera being used.

#### ***III.3.1 Archaeological context***

This process has been tested during an experimentation carried out in the “*Villa dei Pisoni*”, an Early Imperial Age residence.

*Villa dei Pisoni* is located at about 150 m southeastward from the Punta Epitaffio Cape and is also visible from aerial and satellite images. It is composed by several rooms with a global rectangular plan structure elongated 160×120 m (Figure III-16). Most of the internal sector was occupied by gardens (about 95×65 m) and surrounded from apses and columns. Thanks to the findings of a “*fistula aquaria*” containing the inscription “*L. Pisonis*”, the Villa property was attributed to the important “*Pisonis*” family and dated first century A.C.. The Pisonis family took place to the conspiracy against the Emperor Nerone, and the Villa was confiscated and became imperial property in the second century B.C. South-eastward, the Villa structure is linked to an area where there are remains of a thermal bath complex, piers and fish ponds.

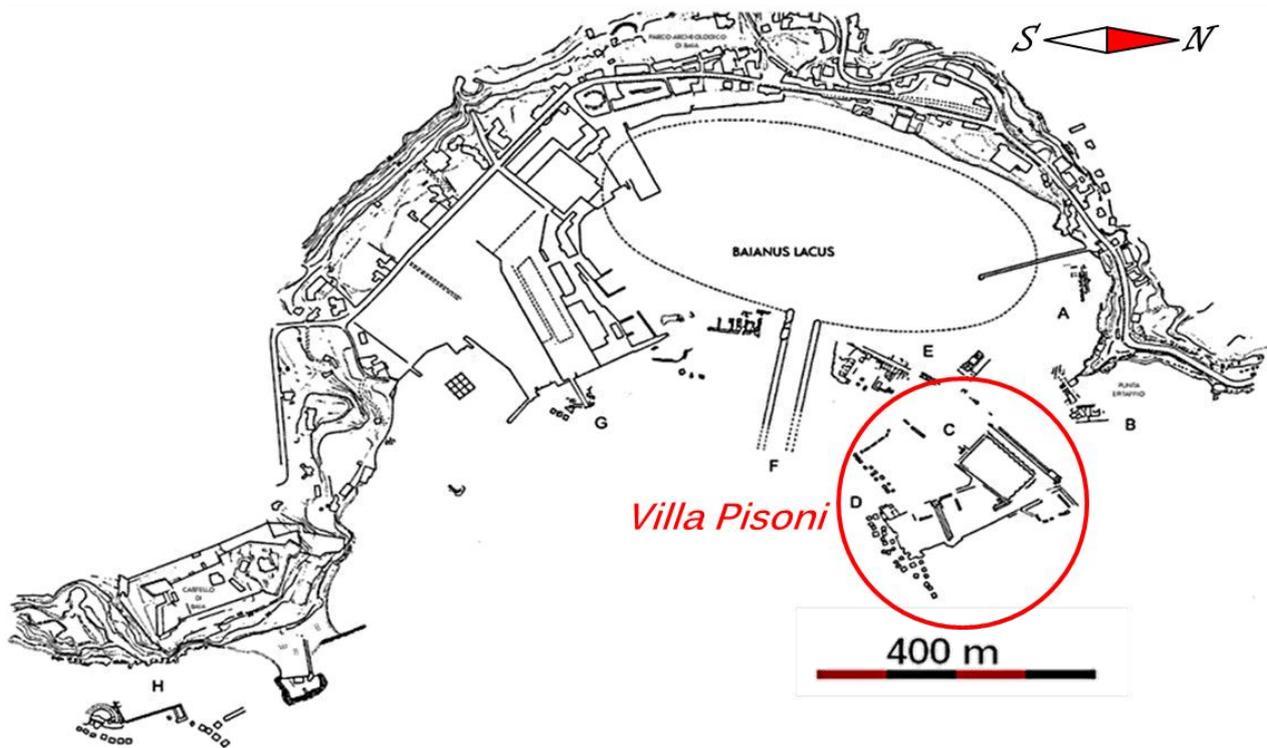


Figure III-16 Underwater Archaeological Park: detail of Villa dei Pisoni

### III.3.2 The Documentation Process

The documentation process, defined to support the experimentation of the cleaning operations, consists of the five subsequent phases:

- a) 3D mapping of the experimental site
- b) 3D Reconstruction of the selected areas
- c) Cleaning operations
- d) 3D Reconstruction of the cleaned areas
- e) Result evaluation

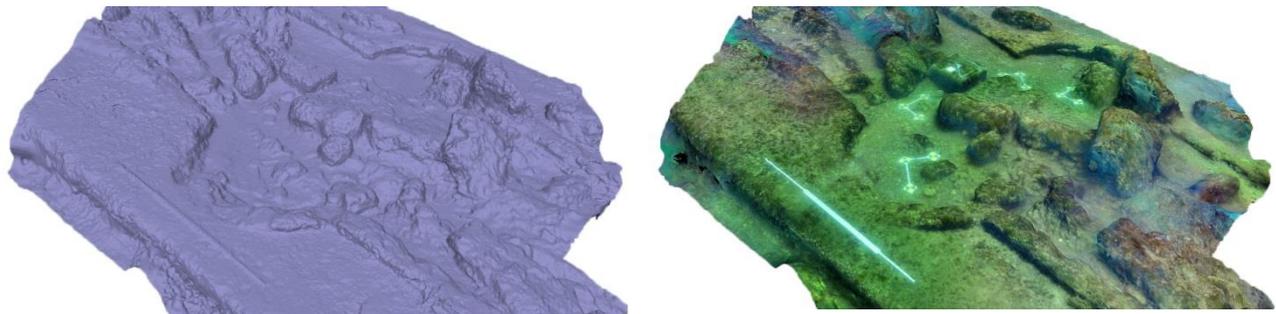
Hereafter, we describe each phase implemented during the intervention survey performed for testing of some devices developed within the CoMAS project.

### ***3D Mapping of the experimental site***

This phase aims to obtain a virtual model of the entire area to be able to plan following operation and defining the best target zone. This phase is equitable with the previous 3D reconstruction test case.

Again, the equipment used to acquire the underwater pictures consists of a Nikon D7000 reflex camera with a 20 mm fixed lens mounted in an underwater housing. In this case, two underwater strobes were necessary to acquire pictures of the portion of the site that was characterized by low lighting conditions.

The acquisition has been carried out in a single diving session, in which the scuba diver swam at a certain distance from seabed (about 2 meters) following a standard aerial photography layout. The diver followed straight lines with constant speed in North-South and East-West directions, taking more than 1500 pictures trying to ensure an overlapping rate of about 75% between two consecutive photos. A set of targets with known dimensions and a 4 meters long scale bar have been placed on the seabed in order to scale the model.



*Figure III-17 3D surface and textured model of the site used to plan the cleaning intervention.*

A preliminary 3D reconstruction of the whole site has been performed immediately after the diving session. In particular, a set of 1560 pictures, down-sampled to 800x600 pixels to save some computational time and pre-processed in order to increase the accuracy of the matching process. The images have been processed with the multi-view 3D reconstruction software Phototscan® by Agisoft, which automates the pipeline illustrated in the previous section and offers some interesting tools and settings.

The processing has been conducted on a workstation based on a CPU Intel core i7, Geforce GTX 680 2Gb and 48 Gb of RAM, and took about 5 hours. The resulting 3D model (Figure III-17) contains about 3.3 million points and 6.5 million triangles.

Since the 3D reconstruction is carried out up to an unknown scale factor, the model has been scaled using the known distances between each couple of targets (targets visible in Figure III-18). In particular, the mean value of the measurements has been used to increase the accuracy of the reconstruction, in order to make a 3D map of the site suitable to plan the cleaning operations. The accuracy of the whole 3D model can be estimated through the average error measured on targets after the scaling process, which is about 4.5 millimetres.



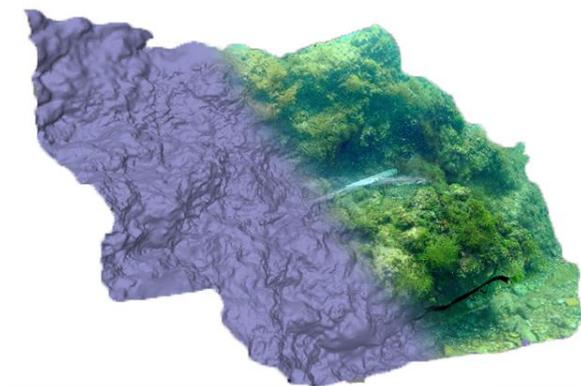
*Figure III-18 The scale bar and a marker over the two areas selected to test the documentation process.*

### ***3D Reconstruction of the selected areas***

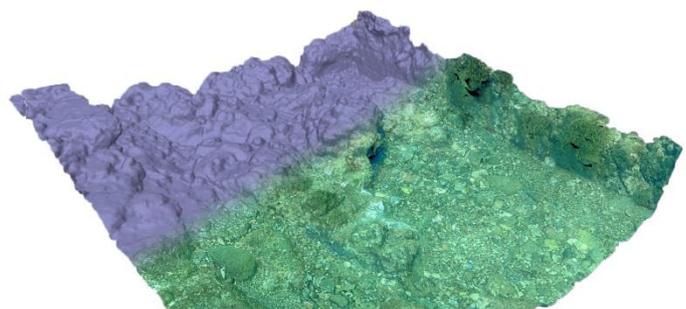
The two selected areas have been acquired at the beginning of the diving session, prior to any cleaning operation. To help the alignment and scaling process, a set of markers and scale bars have been placed on the seabed, surrounding the cleaning intervention area (Figure III-18). About 50 pictures have been acquired for each area, in order to include the surrounding zones to be used as references in the subsequent alignment process between the 3D models, obtained both before and after the cleaning intervention.

The two areas cover a surface of 3 and 1.5 m<sup>2</sup>, respectively, and have been acquired from a distance of about 2 meters.

#### ***Area 1***



#### ***Area 2***



*Figure III-19 3D reconstruction (with and without texture) of the uncleaned Area 1 and Area 2.*

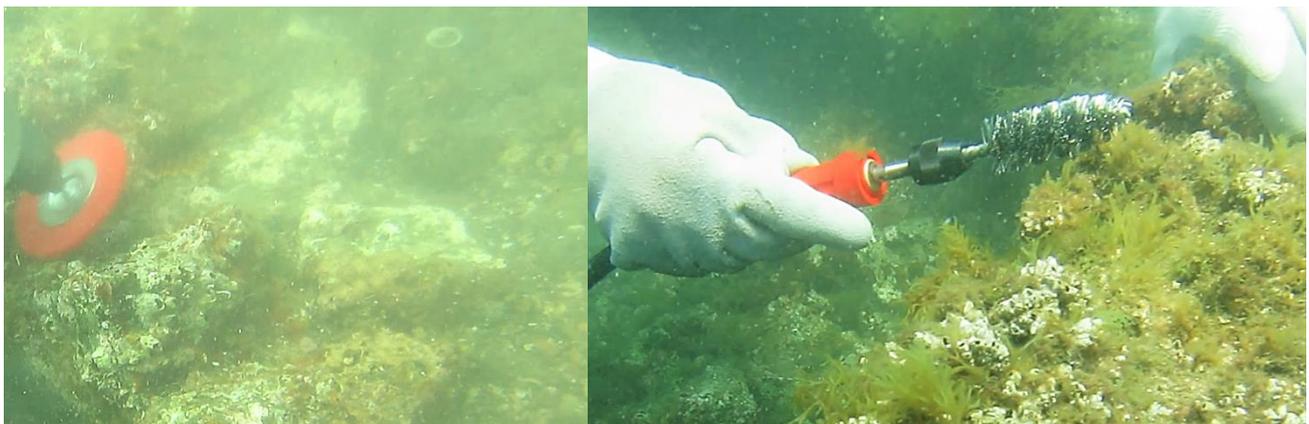
Despite the presence of marine flora and fauna and the bad visibility conditions, more than 1 million points have been reconstructed for both areas (see Figure III-19 and Table III-1); in this case, the average error drops below 1 millimetre.

*TABLE III-1 RESULTS OF THE 3D RECONSTRUCTION RELATED TO THE CLEANED AND UNCLEARED AREAS*

<b>Area</b>	<b>Images</b>		<b>Points (x 1000)</b>		<b>Triangles (x 1000)</b>	
	<i>Unclean</i>	<i>Clean</i>	<i>Unclean</i>	<i>Clean</i>	<i>Unclean</i>	<i>Clean</i>
1	50	100	1100	1500	2200	3500
2	56	50	1500	3000	2900	6000

### ***Cleaning operations***

In this step, the selected areas have been cleaned using the devices developed in the CoMAS project equipped with various tools suitable for the different types of encrustation. In particular nylon and stainless steel brushes, of variable shapes and sizes, have been used (Figure III-20).



*Figure III-20 Cleaning operations performed with nylon and stainless steel brushes.*

### ***3D Reconstruction of the cleaned areas***

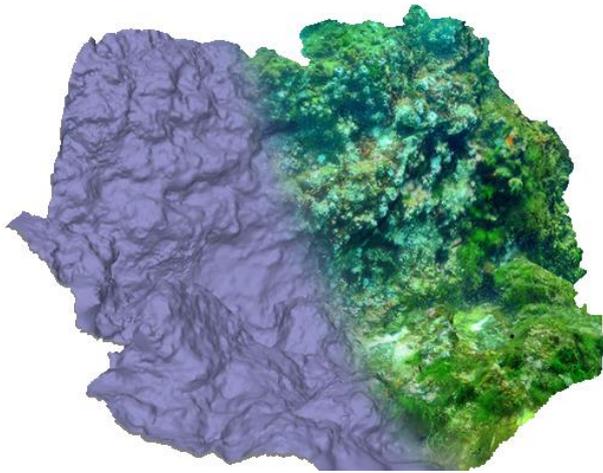
After the cleaning intervention, an additional set of pictures has been acquired for both areas. In particular, a sequence of about 100 pictures has been taken for the first area, whereas about 50 pictures regarded the second one. The pictures have been taken from a distance of about 1 meter, including the surrounding areas as a reference for the alignment process. The results of the 3D reconstruction are summarized in Table III-1.

The cleaned surface presents a higher number of points, since the presence of marine organisms that cause problems in the reconstructed surface is highly reduced.

The reconstructed surfaces are shown in Figure III-21. The uncleaned surface presents more defects, due to the presence of algae and other types of bio-fouling, whereas on the cleaned

surface it is possible to distinguish the various types of encrustations that remained on the surface.

**Area 1**



**Area 2**

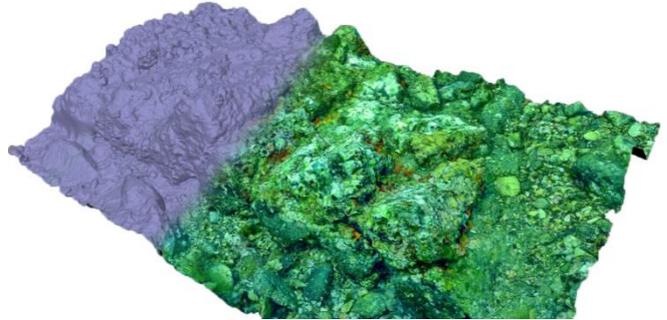
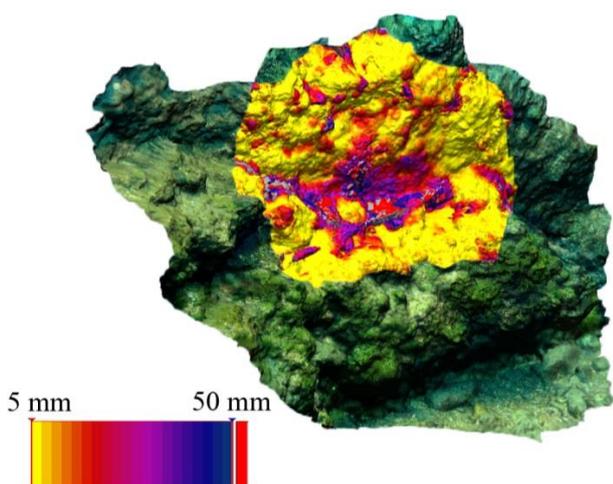


Figure III-21 3D reconstruction (with and without texture) of the cleaned Area 1 and Area 2.

**Results evaluation**

The 3D models of both the cleaned and uncleaned surfaces have been aligned in Rapidform® [196] by identifying corresponding points in the overlapping areas surrounding the cleaned surface. The use of targets and scale bars was helpful for the alignment process. Once the alignment is done, it is possible to analyse the effect of the cleaning operations on the area of interest by comparing the 3D models obtained before and after the cleaning. Figure III-22 shows the cleaned models, highlighting the misalignment with respect to the uncleaned one: in each particular, the cleaned area is coloured from yellowish to bluish according to its distance from the original surface.

**Area 1**



**Area 2**

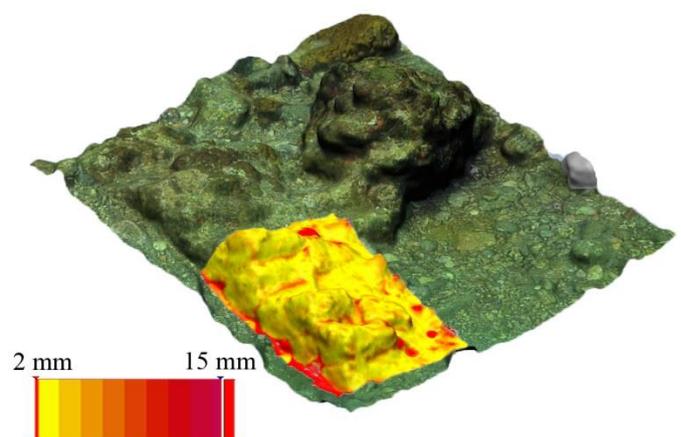


Figure III-22 Mesh deviations obtained comparing the cleaned and uncleaned surfaces.

Different measurements have been conducted on the cleaned areas, summarized in Table III-2. The two areas have an extension of 3 and 1.5 m<sup>2</sup> respectively, and two surfaces of 0.6 and 0.4 m<sup>2</sup> have been used to perform the measurements. The mean distance is related to the deviation between the reconstructed surfaces before and after the cleaning intervention, while the removed volume has been evaluated by measuring the volume enclosed between the uncleaned and cleaned surfaces, using the mesh processing and analysis tools included in Rapidform®.

TABLE III-2 RESULTS OF THE ANALYSIS CONDUCTED COMPARING THE UNCLEANNED AND CLEANED SURFACES.

		<b>Area 1</b>	<b>Area 2</b>
<b>Selected area</b>	(m <sup>2</sup> )	3	1.5
<b>Analysed area</b>	(m <sup>2</sup> )	0.6	0.4
<b>Mean distance</b>	(mm)	9	3.6
<b>Max distance</b>	(mm)	47	12
<b>Removed volume</b>	(dm <sup>3</sup> )	5.4	1.5

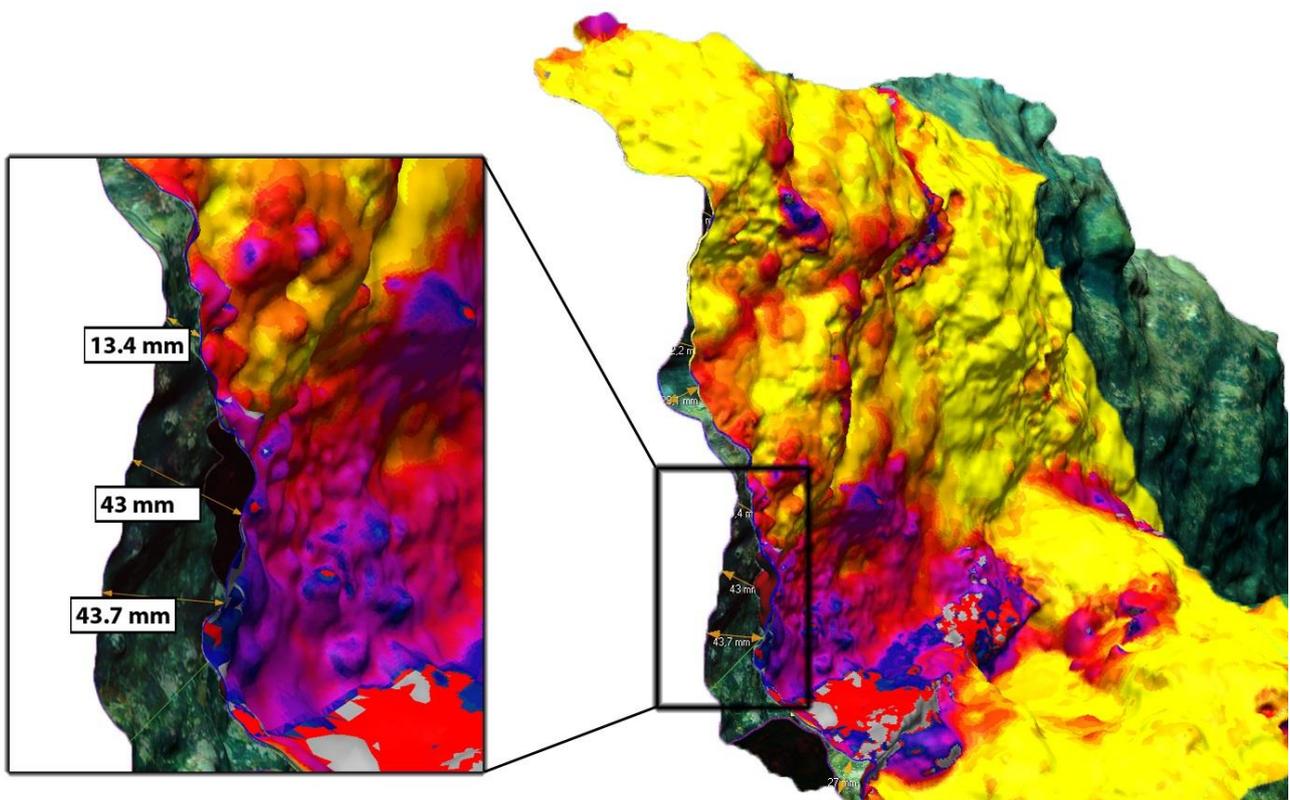


Figure III-23 Cross-section analysis used to correlate the amount of removed material with the type of encrustation.

Additional measurements have been conducted on sections obtained by using different cutting planes. Figure III-23 represents the superposition between the cleaned surface and the initial one, coloured according to the distance; the detailed focuses on possible direct measurements of the removed material. These kinds of information are useful to correlate the amount of removed material with the type of bio-fouling. Higher values have been obtained, for example, on the materials characterized by the presence of algae and plants, lower values are related to the presence of strong encrustation, such as *Thoracica*, so suggesting different strategies and tools and /or intervention time.

## IV KALMAN FILTER THEORY AND MODELLING

This chapter focuses on Kalman Filter (KF) algorithm. We present the equations for the linear and non-linear formulations of the KF. A more detailed description on this topic can be found in [145, 197].

The KF is a recursive data processing algorithm that addresses the general problem of estimating the state of a stochastic system using a model of the system and a set of sensor measurements that are functions of the state. KF contains different structures depending on the linearity and nonlinearity of the system and measurements. In the following subsections the explanations about the Linear Kalman filter, Linear Discrete Kalman filter, and Extended Kalman Filter (EKF) is given in details.

### IV.1.1 Linear Continuous Kalman Filter

Kalman Filter can be derived using continuous-time models and measurements and this derivation approach provides some unique perspectives that are especially useful for small sampling intervals. However, due to the extensive use of digital computers in today's time, the continuous-time Kalman filter is not widely used in practice [150].

Consider the standard linear system model and linear measurement model defined in continuous time:

$$\frac{dx(t)}{dt} = \dot{x}(t) = A(t)x(t) + B(t)u(t) + W(t)w(t); \quad \text{Eq. IV-1a}$$

$$\tilde{z}(t) = H(t)x(t) + v(t); \quad \text{Eq. IV-1b}$$

In the Eq. IV-1a, referred to the system state model, we have:

- $x(t)$ , the  $n \times 1$  state vector
- $u(t)$ , the  $s \times 1$  input or control vector,
- $w(t)$ , a  $q \times 1$  vector of zero-mean Gaussian white noise associated to process uncertainty
- $A(t)$ , the  $n \times n$  state (or model) matrix of the system
- $B(t)$ , the  $n \times s$  input matrix of the system
- $W(t)$ , the  $n \times q$  noise matrix of the system.

Instead, the Eq. IV-1b is the system observation (or measurement) model; here we find:

- $\bar{z}(t)$ , the  $m \times 1$  observation (or measurement) vector
- $v(t)$ , a  $r \times 1$  vector of zero-mean Gaussian white noise associated to observation uncertainty
- $H(t)$ , the  $m \times n$  observation(or measurement) matrix of the system

The hypothesis on noise vectors are the following:

$$E\{w(t)w^T(\tau)\} = Q(t)\delta(1 - \tau); \quad \text{Eq. IV-2a}$$

$$E\{v(t)v^T(\tau)\} = R(t)\delta(1 - \tau); \quad \text{Eq. IV-2b}$$

$$\text{where, } \delta(1 - \tau) = \begin{cases} 1 & ; t = \tau \\ 0 & ; t \neq \tau \end{cases} \text{ is the Kronecker expression} \quad \text{Eq. IV-2c}$$

$$E\{v(t)w^T(\tau)\} = 0 \quad \text{Eq. IV-2d}$$

The last equation implies that  $w(t)$  and  $v(t)$  are uncorrelated.

The Kalman Filter structure for  $\hat{x}(t)$  and  $\hat{z}(t)$ , respectively the estimation of the state and of output vectors, is given by the following equations:

$$\dot{\hat{x}}(t) = A(t)\hat{x}(t) + B(t)u(t) + K(t)[\bar{z}(t) - H(t)\hat{x}(t)]; \quad \text{Eq. IV-3a}$$

$$\hat{z}(t) = H(t)\hat{x}(t); \quad \text{Eq. IV-3b}$$

In which  $K(t)$  is the Kalman gain. The dynamics of the state vector estimation error parameters, defined as  $\tilde{x}(t) = \hat{x}(t) - x(t)$ , can be reformulated by using the Eq. IV-1 and Eq. IV-3 as follows:

$$\dot{\tilde{x}}(t) = L(t)\tilde{x}(t) + m(t); \quad \text{Eq. IV-4}$$

where:

$$L(t) = A(t) - K(t)H(t); \quad \text{Eq. IV-5}$$

$$m(t) = K(t)v(t) - W(t)w(t); \quad \text{Eq. IV-6}$$

When using the matrix exponential solution for the Eq. IV-6, the following expression is obtained for the state error propagation:

$$\tilde{x}(t) = \Phi(t, t_0)\tilde{x}(t_0) + \int_{t_0}^t \Phi(t, \tau) m(\tau) d\tau; \quad \text{Eq. IV-7}$$

Here, it is possible to note that  $u(t)$  does not appear in the error state. Since the system error  $w(t)$  and the measurement error  $v(t)$  are uncorrelated, the following expression represents the covariance matrix of the measurement error:

$$E\{m(t)m^T(\tau)\} = [W(t)Q(t)W^T(t) + K(t)R(t)K^T(t)]\delta(1 - \tau); \quad \text{Eq. IV-8}$$

The state error covariance is defined by:

$$P(t) \equiv E\{\tilde{x}(t)\tilde{x}^T(\tau)\}; \quad \text{Eq. IV-9}$$

As a result, using the Eq. IV-7 and Eq. IV-8, the time derivative expression of the covariance matrix is obtained. The simplified form of this expression is given as follows:

$$\begin{aligned} \dot{P}(t) = & [A(t) - K(t)H(t)]P(t) + P(t)[A(t) - K(t)H(t)]^T \\ & + W(t)Q(t)W^T(t) + K(t)R(t)K^T(t); \end{aligned} \quad \text{Eq. IV-10}$$

In order to determine the Kalman gain  $K(t)$ , the minimization on Eq. IV-10 with respect to  $K(t)$  is done. The necessary conditions lead to:

$$\frac{\partial J}{\partial K(t)} = 0 = 2K(t)R(t) - 2P(t)H^T(t); \quad \text{Eq. IV-11}$$

Solving the Eq. IV-11 for  $K(t)$ , the following expression is obtained:

$$K(t) = P(t)H^T(t)R^{-1}(t); \quad \text{Eq. IV-12}$$

By substituting this Kalman gain expression into Eq. IV-10 the following expression, known as *continuous Riccati Equation*, is obtained:

$$\dot{P}(t) = A(t)P(t) + P(t)A^T(t) - P(t)H^T(t)R^{-1}(t)H(t)P(t) + W(t)Q(t)W^T(t); \quad \text{Eq. IV-13}$$

TABLE IV-1 CONTINUOUS-TIME LINEAR KALMAN FILTER [150]

<b>Initialize</b>	$\hat{x}(t_0) = \hat{x}_0$ $P_0 = E\{\tilde{x}(t_0)\tilde{x}^T(t_0)\}$
<b>Model</b>	$\dot{x}(t) = A(t)x(t) + B(t)u(t) + W(t)w(t), \quad w(t) \sim N(0, Q(t))$ $\tilde{z}(t) = H(t)x(t) + v(t), \quad v(t) \sim N(0, R(t))$
<b>Gain</b>	$K(t) = P(t)H^T(t)R^{-1}(t)$
<b>Covariance</b>	$\dot{P}(t) = A(t)P(t) + P(t)A^T(t) - P(t)H^T(t)R^{-1}(t)H(t)P(t) + W(t)Q(t)W^T(t)$
<b>Estimate</b>	$\hat{\dot{x}}(t) = A(t)\hat{x}(t) + B(t)u(t) + K(t)[\tilde{z}(t) - H(t)\hat{x}(t)]$

Table IV-1 summarizes the continuous-time Kalman filter. Firstly, state error and covariance are initialized. Then, the Kalman gain is calculated with the initial covariance value. Next, the covariance and estimated states are numerically integrated using the continuous-time measurement. The integration of the estimated state and covariance continues until the final measurement time is reached.

### IV.1.2 Linear Discrete Kalman Filter

In many engineering applications both the model and the measurements are available in discrete-time form. The truth linear system model and measurement model for this discrete-time case is given by [150]:

$$x_{i+1} = A_i x_i + B_i u_i + W_i w_i; \quad \text{Eq. IV-14a}$$

$$\tilde{z}_i = H_i x_i + v_i; \quad \text{Eq. IV-14b}$$

Where  $w_i$  and  $v_i$  are assumed zero-mean Gaussian white noise processes with covariances given by

$$E\{w_i w_j^T\} = R_i \delta_{ij}; \quad \text{Eq. IV-15a}$$

$$E\{v_i v_j^T\} = Q_i \delta_{ij}; \quad \text{Eq. IV-15b}$$

where,  $\delta_{ij} = \begin{cases} 1 & ; i = j \\ 0 & ; i \neq j \end{cases}$  is the Kronecker expression Eq. IV-15c

$$E\{v_i w_j^T\} = 0 \quad \text{Eq. IV-15d}$$

This Kronecker delta requirement preserves the block diagonal structure of the covariance and weight matrices, and it is also assumed that  $v_i$  and  $w_i$  are uncorrelated.

The propagation of the current estimate and the update of the state by using the Kalman gain and measurements are done by using the equations given below:

$$\hat{x}_{i+1}^- = A_i \hat{x}_i + B_i u_i; \quad \text{Eq. IV-16a}$$

$$\hat{x}_{i+1} = \hat{x}_{i+1}^- + K_i [\tilde{z}_i - H_i \hat{x}_{i+1}^-]; \quad \text{Eq. IV-16b}$$

In which  $\hat{x}_{i+1}^-$  and  $\hat{x}_{i+1}$  are respectively the *a-priori* and *a-posteriori* (or corrected) estimation of the state vector for the current  $i + 1$  step. The state error covariance matrix is defined using the following expression:

$$\begin{aligned} P_i^- &\equiv E\{\tilde{x}_i^- \tilde{x}_i^{-T}\}; & P_{i+1}^- &\equiv E\{\tilde{x}_{i+1}^- \tilde{x}_{i+1}^{-T}\}; \\ P_i &\equiv E\{\tilde{x}_i \tilde{x}_i^T\}; & P_{i+1} &\equiv E\{\tilde{x}_{i+1} \tilde{x}_{i+1}^T\}; \end{aligned} \quad \text{Eq. IV-17a}$$

$$\begin{aligned} \tilde{x}_i^- &\equiv \hat{x}_i^- - x_i; & \tilde{x}_{i+1}^- &\equiv \hat{x}_{i+1}^- - x_{i+1}; \\ \tilde{x}_i &\equiv \hat{x}_i - x_i; & \tilde{x}_{i+1} &\equiv \hat{x}_{i+1} - x_{i+1}; \end{aligned} \quad \text{Eq. IV-17b}$$

The expressions given in Eq. IV-17b are the state errors using in the state prediction and state update. Here, the aim is to derive an expression for both  $P_{i+1}^-$  and  $P_{i+1}$ , and also an optimal expression for the Kalman gain  $K_i$ . Since Eq. IV-16a is not a direct function of the gain  $K_i$ , it is fairly straightforward to derive the expression  $P_{i+1}^-$ . After substituting the Eq. IV-14a and Eq. IV-16a into Eq. IV-17b, the following expression for  $P_{i+1}^-$  is obtained:

$$\begin{aligned} P_{i+1}^- &\equiv E\{\tilde{x}_{i+1}^- \tilde{x}_{i+1}^{-T}\} \\ &= E\{A_i \tilde{x}_{i+1} \tilde{x}_{i+1}^T A_i^T\} - E\{A_i \tilde{x}_{i+1} w_i^T W_i^T\} \\ &\quad - E\{W_i w_i \tilde{x}_{i+1}^T A_i^T\} + E\{W_i w_i w_i^T W_i^T\}; \end{aligned} \quad \text{Eq. IV-18}$$

After carrying out certain simplifications, finally the following expression is obtained for the state covariance propagation:

$$P_{i+1}^- = H_i P_i H_i^T + W_i Q_i W_i^T; \quad \text{Eq. IV-19}$$

The next step is to obtain an expression for the state covariance update. For this purpose the state propagation and the measurement update equations are used. This derivation for the expression for  $P_i$  is listed in the following set of equations:

$$\tilde{x}_i = (I - K_i H_i)\hat{x}_i^- + K_i H_i x_i + K_i v_i - x_i; \quad \text{Eq. IV-20}$$

$$\begin{aligned} P_i &\equiv E\{\tilde{x}_i \tilde{x}_i^T\} \\ &= E\{(I - K_i H_i)\tilde{x}_i^- \tilde{x}_i^{-T}(I - K_i H_i)^T\} \\ &\quad + E\{(I - K_i H_i)\tilde{x}_i^- v_i^T K_i^T\} + E\{K_i v_i \tilde{x}_i^{-T}(I - K_i H_i)^T\} \\ &\quad + E\{K_i v_i v_i^T K_i^T\} \quad ; \end{aligned} \quad \text{Eq. IV-21a}$$

$$\text{Since } v_i \text{ and } w_i \text{ are uncorrelated} \Rightarrow E\{\tilde{x}_i^- v_i^T\} = E\{v_i \tilde{x}_i^{-T}\} = 0 \quad \text{Eq. IV-21b}$$

$$P_i = [I - K_i H_i]P_i^- [I - K_i H_i]^T + K_i R_i K_i^T \quad \text{Eq. IV-22}$$

In order to determine the Kalman gain  $K_i$ , the minimization on Eq. IV-22 with respect to  $K_i$  is done. The necessary conditions lead to:

$$\frac{\partial J}{\partial K_i} = 0 = -2(I - K_i H_i)P_i^- H_i^T + 2K_i R_i; \quad \text{Eq. IV-23}$$

$$K_i = P_i^- H_i^T [H_i P_i^- H_i^T + R_i]^{-1}; \quad \text{Eq. IV-24}$$

Substituting Eq. IV-24 into Eq. IV-22, the simplified expression for the state covariance update is obtained:

$$P_i = [I - K_i H_i]P_i^-; \quad \text{Eq. IV-25}$$

The above Eq. IV-25 infers that while the propagation step in Eq. IV-19 *increases* the covariance; in the opposite side, the update stage of the discrete-time Kalman filter *decreases* the covariance. This observation is instinctively consistent since in general more measurements improve the state estimate [150]. A further expression for the state update can be derived by using Kalman gain, real measurements and expected measurements as follows:

$$\begin{aligned} \hat{x}_i &= [I - K_i H_i]\hat{x}_i^- + K_i \tilde{z}_i \\ &= \hat{x}_i^- - K_i (\tilde{z}_i - H_i \hat{x}_i^-); \end{aligned} \quad \text{Eq. IV-26}$$

A summary for the algorithms of the discrete-time Kalman filter is given in Table IV-2. First, the initializations for the state and covariance are done. If a measurement is available, the state and covariance are updated by using Kalman gain values and the propagation of the state estimate and

covariance are calculated for the next step. If a measurement is not available, state and covariance are propagated and this process is repeated until the new measurement is available.

*TABLE IV-2 DISCRETE-TIME LINEAR KALMAN FILTER [150]*

<b>Model</b>	$x_{i+1} = A_i x_i + B_i u_i + W_i w_i, \quad w_i \sim N(0, Q_i)$ $\tilde{z}_i = H_i x_i + v_i, \quad v_i \sim N(0, R_i)$
<b>Initialize</b>	$\hat{x}(t_0) = \hat{x}_0$ $P_0 = E\{\tilde{x}(t_0)\tilde{x}^T(t_0)\}$
<b>Gain</b>	$K_i = P_i^- H_i^T [H_i P_i^- H_i^T + R_i]^{-1}$
<b>Update</b>	$\hat{x}_i = \hat{x}_i^- - K_i [\tilde{z}_i - H_i \hat{x}_i^-];$ $P_i = [I - K_i H_i] P_i^-;$
<b>Propagation</b>	$\hat{x}_{i+1}^- = A_i \hat{x}_i + B_i u_i;$ $P_{i+1}^- = A_i P_i A_i^T + W_i Q_i W_i^T;$

### ***IV.1.3 Discrete Extended Kalman Filter***

As described above, the Kalman filter addresses the general problem of trying to estimate the states of a linear discrete-time process; however, a large class of estimation problems involves nonlinear models. A vast majority of such nonlinear applications uses a Kalman filter that linearizes about the current mean and covariance, and this type of filter is called as an Extended Kalman filter (EKF) [198].

In order to derive the algorithm for EKF, first the discrete nonlinear system and measurement model are expressed as follows:

$$x_{i+1} = f(x_i, u_{i+1}, w_i, i); \tag{Eq. IV-27a}$$

$$\tilde{z}_i = h(x_i, v_i, i); \tag{Eq. IV-27b}$$

where random variables  $w_i$  and  $v_i$ , again, represent a zeros-mean Gaussian process and measurement noise. In practice, the values of the noise  $w_i$  and  $v_i$  are unknown at each time-step. However, the state and measurement vectors can be defined without considering these values:

$$\tilde{x}_{i+1} = f(\hat{x}_i, u_{i+1}, 0, i); \tag{Eq. IV-28a}$$

$$\tilde{z}_i = h(\tilde{x}_i, 0, i); \tag{Eq. IV-28b}$$

The main difference between the linear and extended Kalman filter is in use of the linearized system and measurement models on the steps of the filtering algorithm. The linearization is performed by using first-order Taylor series expansion. The first-order expansion of the system and measurement model is given as follows:

$$f(x_i, u_{i+1}, i) \cong f(\bar{x}_i, u_{i+1}, i) + \left. \frac{\partial f}{\partial x} \right|_{\bar{x}_i} [x_i - \bar{x}_i]; \quad \text{Eq. IV-29a}$$

$$h(x_i, i) \cong h(\bar{x}_i, i) + \left. \frac{\partial h}{\partial x} \right|_{\bar{x}_i} [x_i - \bar{x}_i]; \quad \text{Eq. IV-29b}$$

The propagations of the current estimate and covariance matrix are done by using the following equations:

$$\hat{x}_{i+1}^- = \hat{x}_i + \int_i^{i+1} f_k(\hat{x}_i, i) dt = (I - A_i dt) \hat{x}_i; \quad \text{Eq. IV-30}$$

$$P_{i+1}^- = (I + A_i dt) P_i (I + A_i dt)^T + Q; \quad \text{Eq. IV-31}$$

Where  $A_i$  is the Jacobian matrix of first derivatives of  $f$  with respect to the state variables:

$$A_i = \frac{\partial f_i(\hat{x}_i, i)}{\partial \hat{x}_i}; \quad \text{Eq. IV-32}$$

Then Kalman gain  $K_i$  is computed when the measurement vector is available. In this computation, the linearized measurement model is used:

$$K_i = P_i^- H_i^T [H_i P_i^- H_i^T + R_i]^{-1}; \quad \text{Eq. IV-33}$$

$$H_i = \frac{\partial h_i(\hat{x}_{i+1}^-, i)}{\partial \hat{x}_{i+1}^-}; \quad \text{Eq. IV-34}$$

In order to determine the error between the actual measurement and expected measurement the innovation step is executed:

$$e_i = \tilde{z}_i - H \hat{x}_{i+1}^-; \quad \text{Eq. IV-35}$$

Finally, the expressions that provide the state update and covariance update is given as follows:

$$\hat{x}_{i+1} = \hat{x}_{i+1}^- - K_i e_k; \quad \text{Eq. IV-36}$$

$$P_{i+1} = [I - K_i H_i]P_{i+1}^-; \quad \text{Eq. IV-37}$$

The following Table IV-3 summarizes the procedures of the EKF. First the initialization is done for both state and covariance matrix. Then, Kalman gain is computed when the measurement is available. After computing the error between the real measurements and expected measurements, the state update and covariance update steps are implemented.

TABLE IV-3 DISCRETE-TIME EXTENDED KALMAN FILTER [150]

<b>Model</b>	$x_{i+1} = A_i x_i + B_i u_i + W_i w_i, \quad w_i \sim N(0, Q_i)$ $\tilde{z}_i = H_i x_i + v_i, \quad v_i \sim N(0, R_i)$
<b>Initialize</b>	$\hat{x}(t_0) = \hat{x}_0$ $P_0 = E\{\tilde{x}(t_0)\tilde{x}^T(t_0)\}$
<b>Gain</b>	$K_i = P_i^- H_i^T [H_i P_i^- H_i^T + R_i]^{-1}$
<b>Update</b>	$\hat{x}_i = \hat{x}_i^- - K_i [\tilde{z}_i - H_i \hat{x}_i^-];$ $P_i = [I - K_i H_i]P_i^-;$
<b>Propagation</b>	$\hat{x}_{i+1}^- = A_i \hat{x}_i + B_i u_i;$ $P_{i+1}^- = A_i P_i A_i^T + W_i Q_i W_i^T;$

It is clear that Extended Kalman filter works well only in the region where the first-order Taylor series linearization adequately approximates the nonlinear behaviour of the system. For this reason, when the estimated initial state is far from the true state, instabilities may occur in estimation process. To overcome these instabilities EKF can be reconfigured by adding the second-order terms in Taylor series, but in that case, the computational burden becomes an important factor. Therefore, in practice, this standard EKF has remained the most popular method for nonlinear estimation problems. Other filters (like Unscented and Particle Kalman filters) are investigated only when the performance of the standard EKF is not sufficient, and the quantity or the performance of the sensors is limited [150].

In this thesis EKF is used for navigation algorithms and several sensor measurements are fused through EKF. The details about this system are given in the following chapter.

#### IV.1.4 Filter Tuning

The performance of the Kalman filter depends on the filter tuning: the selection of the covariance matrices  $P_0$ ,  $Q$ , and  $R$ . Filter tuning has two main objectives: to obtain maximum estimation

accuracy and to converge to an accurate estimate timely [147]. The  $P_0$  matrix determines the rapidity of the initial convergence and this matrix has no contribution on the steady-state performance of the filter. A  $P_0$  matrix with large entries, compared to the  $Q$  and  $R$ , provides a rapid initial convergence on estimation. In filter, the trade-off between the rapid tracking of disturbance noise induced state variations is determined by the selection of the covariance matrices  $Q$  and  $R$  [147]. These matrices filter also the measurement noises. The  $Q$  and  $R$  also determine the filter stability in steady-state condition. Generally the values of  $Q$  is selected as the one hundredth or one thousandth of the  $P_0$  ; and the measurement noise level determines the value of  $R$  matrix.

## ***V IMPLEMENTATION OF A 6DOF EKF FOR INCREMENTAL UNDERWATER MAPPING***

---

The possibilities given by the use of underwater vehicles are amongst the underwater archaeology main interests. The advantages are evident in terms of costs, time, and repeatability. Besides, the use of such vehicles is mandatory in many occasions, due to impossibility for human researchers or operators to reach deeper sites or to work underwater for long time. For these reasons, the *CoMAS* project is also oriented towards capabilities of underwater vehicles. In particular, ROVs interventions are matter of study, because of less prohibitive costs and safer working conditions with respect to manned vehicles. All the main archaeological operations of interest that that are feasible *in-situ* are analysed in order to find new solutions, especially regarding automation possibilities, to reduce costs and working time and/or enhance safety and results.

Test campaign oriented towards environment variables and parameters have been conducted. Indeed, it is crucial to characterize new environment and working condition to guarantee the reliability of methodologies and tools and to ease the control of UV and mechanical devices employed, for example, in cleaning, micro-probing or carriage operations.

Nonetheless, it stands to the reason that operative conditions are strictly prescribed for all activities on archaeological findings, and UUVs need a robust design oriented to their intervention capabilities in underwater context. The manipulation of valuable artefacts requires a certain intervention precision and, therefore, great stability. In effect, in addition to inertia and sea current effects, it is necessary to counterbalance arbitrary moments and forces caused by interactions between tool and artefact (Station keeping problem). This is usually achieved through force sensing used directly in control loops [199, 200, 201], even if several attempts of vision based station keeping have been made [202, 203, 204, 205].

In this context, it clearly appears the need for careful vehicle navigation and, concurrently, accurate site mapping. To this end, real-time or almost-real-time SLAM capabilities would be of high importance, in order to either avoid or, at least, to limit preliminary survey missions and consequently reduce the total time dedicated to a unique intervention. Real time visual SLAM can suffer from different problems, even more in underwater condition because of the phenomena described in section II.1.5. Furthermore, the turbidity due to intervention itself can seriously impair visibility. Thus, a valid solution could be the integration of data from different sensors; two ways have been considered in the *CoMAS* project: acquire the scene also by means of an acoustic

camera and exploit information from navigation sensors in order to increase positioning reliability. The former solution consists in merging the point cloud models recovered from the optical and acoustic camera, if both available, or use at least the one available to track the motion and continue the mapping. The optical point cloud could be very accurate but only available in favourable visibility and relatively near field condition, whereas the acoustic one, is not affected by turbidity or bad illumination but generally has lower resolution and does not give colour information.

Another promising solution is the exploitation of UUVs standard navigation sensors (i.e. AHRS, DVLs, depth gauges etc.) in order to mutual enhance 3D mapping performances and vehicle navigation. The navigation sensors, even if dead-reckoned and, thus, affect by drift in long surveys, can be managed to provide a reliable pose estimation that can be exploited to noticeably ease the 3D scene mapping. Furthermore, optical device have been employed in the so-called Vision-Based Navigation (VBN) techniques [206], to reduce errors, especially during station keeping and in case of trajectory loop closure. This approach has been the one addressed by the presented work.

## ***V.1 Motivations and goals***

The *CoMAS* Project involves the development of a ROV suited for surveys and restoration interventions. For this purpose, the ROV is equipped with common navigation sensors like AHRS and DVL, and with a 3D opto-acoustic camera. This device is a prototype dedicated to real time 3D mapping in underwater environment, which merges point clouds coming from a 3D imaging sonar and from a stereo-camera. The system has to be pre-calibrated to give the aligned 3D point cloud, and works on the two different sensing modalities (optical and acoustic) in order to exploit their respective advantages and overtake drawbacks. The stereo-camera can give a detailed 3D vision of the imaged area but the acoustic relief is necessary for vision in turbidity condition, typical during intervention operations like cleaning and probing.

Moreover, AHRS and DVL can support navigation and, consequently, alignment of point clouds over time. As previously said, concurrent navigation and mapping estimation, represent the so-called SLAM problem, a well-known task in robotics. This task can be accomplished through sensor fusion techniques using an Extended Kalman Filter that enhances 3D mapping performance.

Both the ROV and the opto-acoustic camera have been under development during the research period covered by this thesis. For this reason, the implementation and testing of the SLAM

algorithm have been carried out at UVL of ViCoRob in Girona, which also made available some dataset acquired during previous campaigns.

The EKF algorithm has been implemented starting from the available Stereo-Tracking code, a particular implementation for optical SLAM, developed by ViCoRob researchers, able to perform an on-line incremental 3D reconstruction, even if at quite low frame rates. In the presented work, it has been adapted for the specific purposes and used as an input for the Extended Kalman Filter. In particular, the camera trajectory is estimated using an EKF, which also integrates the measurements provided by the other navigation sensors. The use of an EKF contributes to attain two main enhancements, which are strictly related due to the mutual contribution of vision and navigation sensors:

- *Correct alignment of successively estimated point clouds:*

The integration of different sensors in the pose estimation of the vehicle (and camera) allows to better estimate the incremental pose, so granting a good alignment of successive 3D sparse maps. These would otherwise be significantly difficult or totally impossible to be aligned in case of low overlapping among subsequent 3D maps

- *Better global estimation of the camera (and the vehicle) trajectory:*

Reducing the error in the estimation of subsequent poses, the whole trajectory and map can be estimated much more accurately. The trajectory can be further enhanced at each loop-closure event, thanks to the employment of vision navigation with a slightly modified EKF, the Augmented State Kalman Filter (ASKF) [207, 12, 164]. This capability corrects the drift error that affects DR sensors in long surveys and, consequently, allows to adjust the reconstructed 3D map.

Estimating the correct trajectory during on-line application can have other positive effects on survey operation. By having the estimated camera poses allows to avoid image registration process required for dense stereo mapping. The available accurate sparse cloud and trajectory can be exploited to speed up the process for the 3D reconstruction of the detailed model, starting the process directly from the point cloud densification.

Hereafter, we describe the implementation steps of the EKF and ASKF algorithms in MATLAB® environment. An overview about C++ porting, which should remarkably increase performances, is also given. The final section of the chapter shows some achieved results.

## V.2 Assumption and Equipment

The presented application is tested with an AUV equipped with a downward-looking digital-still stereo-rig of calibrated cameras. The measurements of velocities, absolute orientation, and depth, performed by vehicle navigation sensors, are processed inside the filter together with optical data to estimate the trajectory covered during the survey. The particular AUV platform is extensively described in [208, 209]. In Table V-1 we summarize the employed measurements, the related sensors, and the associated errors (estimated or assumed):

TABLE V-1. SENSORS EMPLOYED IN OUR UNDERWATER PLATFORM.

Measurement	Sensor	Precision
Roll/Pitch/Heading	Attitude and Heading Reference System (AHRS)	$\pm 1e-2$ deg
Body Frame Velocities	Acoustic Doppler (DVL)	$\pm 1e-3$ m/sec
Depth	Pressure Sensor	$\pm 1e-2$ m

Our stereo-rig has been calibrated in order to compensate for geometric distortions of the image. In the implemented method, we assume that all coordinate transformation between every sensor and the vehicle reference frame are known.

### V.2.1 Coordinate frame relationship

In this section, we describe the reference frames adopted in vehicle navigation and their 6-DOF coordinate frame relationships as illustrated in Figure V-1 and Figure V-2.

Throughout this article, we adopt the Smith, Self, and Cheeseman coordinate frame convention [153].

The pose of a reference frame  $\{j\}$  with respect to a reference frame  $\{i\}$  can be described by a vector  $x_{ij}$ , consisting of three translation components ( $x$ ,  $y$  and  $z$ ) and three rotation angles ( $\varphi$ ,  $\vartheta$  and  $\psi$  that denote respectively roll, pitch and yaw angles):

$$x_{ij} = [\varphi \quad \vartheta \quad \psi \quad x \quad y \quad z]^T = [\Theta_{ij} \quad t_{ij}]^T ;$$

where  $\Theta_{ij}$  is the orientation vector defined by  $\varphi$ ,  $\vartheta$  and  $\psi$  Euler angles around  $x$ ,  $y$ , and  $z$  axis respectively, and  $t_{ij}$  is the translation vector. The orientation part can also be described by a rotation matrix  ${}^i_jR$  defined as:

$${}^i_jR = rot_{xyz}(\Theta_{ij}).$$

It is possible to arrange  ${}^i_jR$  matrix and  $t_{ij}$  vector in order to obtain the standard homogeneous coordinates transformation matrix that describes both position and orientation of  $\{j\}$  with respect to  $\{i\}$ :

$${}^i_jT = \begin{bmatrix} {}^i_jR & t_{ij} \\ 0 & 1 \end{bmatrix}.$$

Standard coordinate transformation operations are the compounding and inverse coordinate frame relationships, which are denoted as  $x_{ik} = x_{ij} \oplus x_{jk}$ , and  $x_{ji} = \ominus x_{ij}$ , and corresponds to  ${}^i_kT = {}^i_jT * {}^j_kT$  and  ${}^j_iT = {}^i_jT^{-1}$ , respectively.

Following the standard SNAME (The Society of Naval Architects and Marine Engineers) convention [210], we define the right-handed vehicle coordinate frame  $\{v\}$ , to be coincident with a fixed point on the vehicle (corresponding to the centre of gravity in our case) and oriented such that the positive  $x_v$ -axis is aligned with the bow, positive  $y_v$ -axis to starboard, and  $z_v$ -axis down.

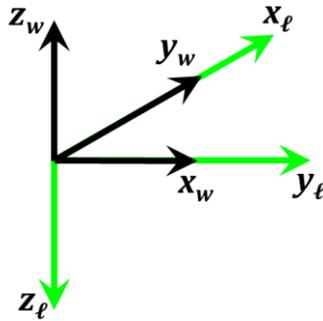


Figure V-1 Representation of world (in black) and local (in green) reference frames.

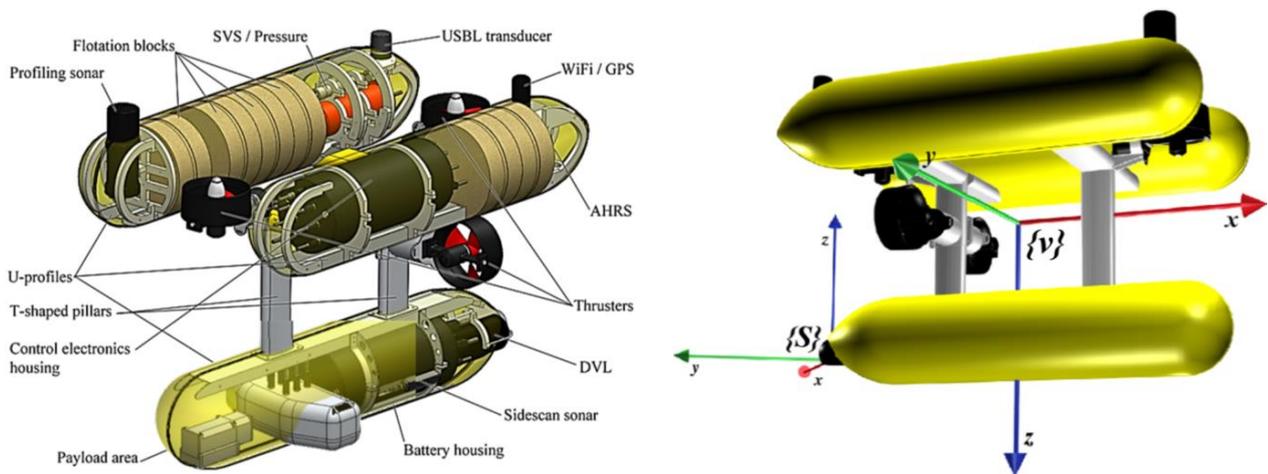


Figure V-2 Girona 500 equipment (left); Girona 500 AUV with its coordinate frame  $\{v\}$  and an example of a sensor frame  $\{S\}$  [211].

Each sensor's internal coordinate frame (in which measurements are expressed) is assumed static and known with respect to vehicle frame, and obtained in a previous calibration process. Sensor pose in vehicle coordinate frame relationship is denoted by the vector  $x_{vs}$ .

In order to express the vehicle pose, other two navigation frames are defined: the world frame  $\{w\}$  and the local-level frame  $\{\ell\}$ . The former follows the standard map convention, so allowing to easily display results; it is a static reference frame located at the water surface oriented with  $x_w$ -East,  $y_w$ -North, and  $z_w$ -Up. It is denoted by subscript  $w$  whereas the vehicle position with respect to this frame is denoted by  $x_{wv}$ . The latter frame, identified by the subscript  $\ell$ , is the local-level frame, coincident with the world frame, but oriented with  $x_\ell$ -North,  $y_\ell$ -East, and  $z_\ell$ -Down. It represents a zero-orientation version of the vehicle frame. Subscript  $x_{\ell v}$  denotes vehicle position in this frame.

### ***V.3 EKF Formulae for State Prediction and Correction***

The Extended Kalman Filter (EKF) is a standard tool both to fuse data coming from different sensors, and to estimate the motion.

The implemented EKF merges data from DVL, AHRS, depth gauge and optical tracking to estimate the vehicle pose with respect to the fixed world reference frame; each sensor contribution is weighted according to its accuracy, through the provided measurement covariance. The obtained pose is exploited to correct the optical tracking and the associated point cloud registration previously performed, so allowing a more correct incremental 3D reconstruction of the scene.

The specific EKF formulation, adapted to our problem, is presented in the following.

#### ***V.3.1 State definition***

Our model uses a 6-degrees-of-freedom (DOF) representation of the vehicle pose, consisting of three translation components ( $x$ ,  $y$  and  $z$ ) and three rotation angles ( $\varphi$ ,  $\vartheta$  and  $\psi$  that denote respectively roll, pitch and yaw angles), represented in the local-level frame.

The pose vector and the kinematic vector at a generic step can be defined, respectively, as:

$$x_p = [\varphi \ \vartheta \ \psi \ x \ y \ z]^T; \quad x_k = [\omega\varphi \ \omega\vartheta \ \omega\psi \ vx \ vy \ vz]^T;$$

Where  $\omega$  and  $v$  denotes angular and linear velocities; note that our  $x_p$  consists of vehicle pose parameters represented in local-level frame; thus, it can be assumed as the better estimate for actual pose  $x_{\ell v}$ : ideally,  $x_{\ell v}$  cannot be known exactly but just estimated with some errors.

Apart from these twelve variables, the state vector  $x$  at  $i$ -th iteration stores also the six pose variables relative to the last step in which optical data has been used (denoted by subscript  $last\_opt$ ). It assumes the following form:

$$x_i = \begin{bmatrix} x_{p_i} \\ x_{k_i} \\ x_{p_{last\_opt}} \end{bmatrix};$$

In addition, a matrix  $P_i$  is defined as the covariance matrix of the state vector  $x_i$  at the  $i$ -th iteration.

### **V.3.2 Vehicle model**

Our EKF uses a simplified constant velocity model with no control inputs to determine how the state evolves from step  $i-1$  to step  $i$ . The prediction step is evaluated using only the previous position and speed. While being an approximation, it provides acceptably good results in case of fixed frame representation and small time intervals.

The system model at  $i$ -th step can be represented as a function of  $x_{i-1}$ ,  $dt$  and  $w$ .

$$x_i = \begin{bmatrix} x_{p_i} \\ x_{k_i} \\ x_{p_{last\_opt}} \end{bmatrix} = f(x_{i-1}, dt_i, w);$$

In particular, the model that describes the state vector portions  $x_{p_i}$  and  $x_{k_i}$  is assumed as:

$$x_{p_i} = x_{p_{i-1}} + dt_i * x_{k_{i-1}} + w_p; \quad x_{k_i} = x_{k_{i-1}} + w_k;$$

Where:

- $dt_i$  is the time interval between iterations;
- $w_p$  and  $w_k$  are 6-elements vectors of zero-mean white Gaussian process noise that can be used to take in account all the model approximations (and simplification of the linearized dynamic model).

The pose part  $x_{p_i}$  is evaluated from the previous estimated values whereas the kinematic part  $x_{k_i}$  is supposed to be constant.

Associated with the process noise vector  $w$  there is the process noise covariance  $Q_i$  that has been set empirically assuming the pose variables covariance matrix  $\sigma^2$  [ref]:

$$Q_i = \begin{bmatrix} \frac{1}{4} dt_i^4 \sigma^4 & \frac{1}{2} dt_i^3 \sigma^2 & 0_6 \\ \frac{1}{2} dt_i^3 \sigma^2 & dt_i^2 \sigma^2 & 0_6 \\ 0_6 & 0_6 & 0_6 \end{bmatrix}; \quad \text{where } \sigma^2 = \begin{bmatrix} \sigma_\theta^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_\phi^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_\psi^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_x^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_y^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_z^2 \end{bmatrix};$$

### V.3.3 State Prediction

Being defined the vehicle model, we can write the equation for the *a priori* state vector  $\hat{x}_i^-$  and *a priori* error covariance matrix  $P_i^-$ , evaluated from  $\hat{x}_{i-1}^-$  and  $P_{i-1}^-$ , that represent the *a posteriori* estimations of the previous step:

$$\hat{x}_i^- = \begin{bmatrix} \hat{x}_{p_i}^- \\ \hat{x}_{k_i}^- \\ \hat{x}_{p_{last\_opt}}^- \end{bmatrix} = A_i \hat{x}_{i-1}^-;$$

$$P_i^- = A_i P_{i-1}^- A_i^T + W_i Q_{i-1} W_i^T;$$

In which the  $A_i$  is the Jacobian matrix of partial derivatives of  $f$  with respect to the state variables and  $W_i$  is the Jacobian matrix of partial derivatives of  $f$  with respect to the process noise  $w$ .

$$A_i = \begin{bmatrix} I_6 & dt_i \cdot I_6 & 0_6 \\ 0_6 & I_6 & 0_6 \\ 0_6 & 0_6 & I_6 \end{bmatrix}; \quad W_i = [I_{18}];$$

in which  $I_n$  and  $0_n$  are, respectively, an identity and a zero matrix of  $n \times n$  elements.

### V.3.4 State Correction

The Kalman filter considers a measurement vector  $z_i$  that can be linked to the state vector through the observation function  $h$  and a vector of zero-mean white Gaussian noise  $v$ :

$$z_i = h(x_i) + v_i;$$

The  $z_i$  vector contains every sensor measurements available at  $i$ -th iteration, so its size and the Jacobian matrix  $H$  associated to  $h$ , are variable; consequently, also the noise vector and the associated Jacobian matrix  $R$  changes in size according to available measurements. We will later explain, in following sections, how  $H$  and  $R$  matrices are evaluated and arranged. The linearized model for the *a priori* estimation of the observation vector is defined as:

$$\tilde{z}_i = H \hat{x}_i^-;$$

The correction of the *a priori* state and covariance by means of information provided by sensors available in  $z_i$  can be made through the equations:

$$K_i = \frac{P_i^- H_i^T}{(H_i P_i^- H_i^T + R_i)};$$
$$\hat{x}_i = \hat{x}_i^- + K_i (z_i - \tilde{z}_i);$$
$$P_i = (I - K_i H_i) P_i^-;$$

Where  $K_i$  is the Kalman gain and  $I$  is an identity matrix.

Thus, the EKF implemented gives the best estimation for the vehicle pose according to all the available sensors and to their associated errors. We further use this information to correct the optical tracking and, consequently, to orient in space the point clouds for achieving the incremental 3D mapping of the imaged scene.

## ***V.4 Observation Models***

In this section, we specify the kind of measurement performed by each sensor and the provided information, in order to define the  $h$  function that links the state vector variables to the physical quantities measured by sensors. We discuss about the nature of the sensor output in order to establish relation with defined state vector.

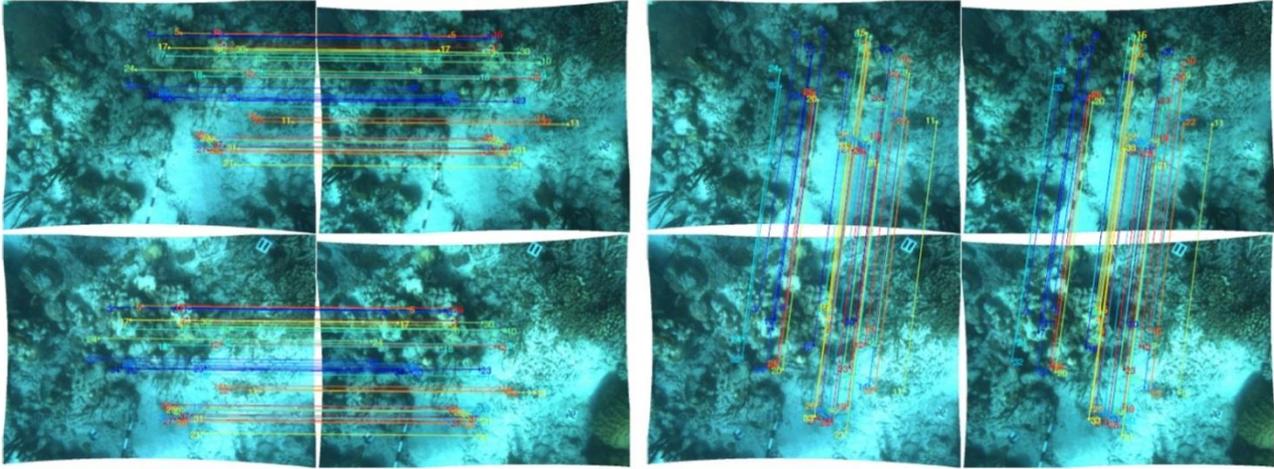
### ***V.4.1 Stereo-Tracking Measurements***

In AUV applications, continuous video acquisition is often avoided because of the power consumption required by the artificial illumination; for this reason, optical data is usually acquired by interval shooting of still images, where the frame rate is adjusted to provide enough overlapping to ensure adequate estimation of the full trajectory. When this condition is met, then a calibrated stereo-rig can ideally be used to create a 3D point cloud of the scene and track the camera movement in that space, through the use of trigonometry principles and scale invariant feature matching in the acquired stereo-pairs. These aspects are detailed next.

#### ***Image Registration and Optical tracking***

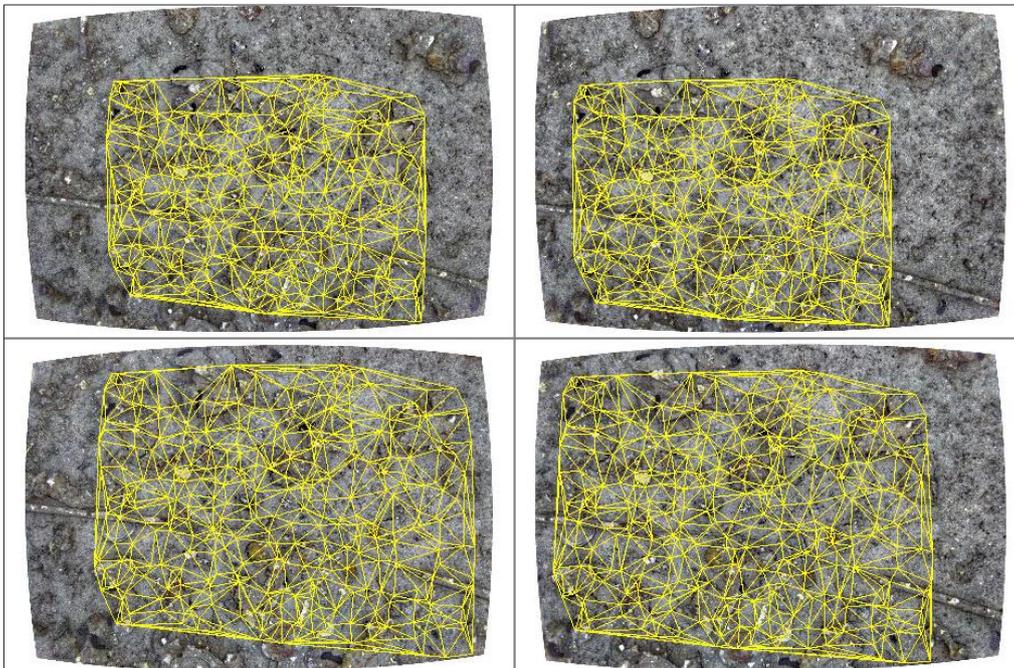
The optical tracking code had been previously developed by ViCoRob as a first step towards the incremental on-line 3D reconstruction. Having a SLAM purpose, it performs the camera tracking and simultaneously creates the incremental 3D model of the acquired scene. The optical algorithm requires subsequent overlapping stereo-pairs in which it recognizes and matches scale-invariant

descriptors (SURF or SIFT) features. Each feature is seen at least in four images. In effect, after image undistortions, rectifications and enhancement, features are extracted and matched between left and right images inside the same stereo-pair Figure V-3(left); matched features are then compared with those matched in previous stereo-frame (Figure V-3(right)), in order to keep only features visible in a quadruplet (at least) so increasing algorithm robustness (Figure V-4).



*Figure V-3 Left: Stereo matching. Right: temporal matching.*

*For visualization clarity, only few features are showed*



*Figure V-4 Rectified quadruplet example and features matching: each vertex of triangulated structure represents a feature.*

This phase can be controlled through different settings. Starting from original stereo-pairs, they can be resized, undistorted and/or rectified according to previously evaluated calibration parameters; then it is possible to choose the technique for grayscale conversion (red, green and yellow single channel, Y component of YIQ model or Principal Component Analysis - PCA) and the algorithm for contrast and brightness adjustment (Normalization, Equalization to 255 or CLAHE [212]). In many cases the PCA – CLAHE combination is preferred, but some tests demonstrated that for our purpose Green channel – CLAHE represents the best solution in terms of quality and time costs.

Other parameters control image sequence and overlapping area in different ways:

- limiting number and/or density of matched features in subsequent pairs
- requiring a minimum number of features matched in time to perform tracking
- requiring and a minimum number of subsequent step in which a feature has to be visible to be taken into account.

These settings can have a relevant impact on the algorithm speed and robustness. In our case, due to speed requirements and low overlapping of subsequent pairs, we just match features from  $t$  step with those from  $t-1$  step.

After stereo and temporal matching, a real scale 3D point cloud for each stereo-pair can be obtained from the remaining set of points, using the calibration data and triangulation principles. The two point clouds obviously represent the same scene portion seen from subsequent positions of the stereo-rig in consecutive instants. Assuming that the scene remains rigid over time, these point structures can be registered using the RANSAC algorithm (RANDOMSample Consensus [213]) applied on an Absolute Orientation algorithm [214]. This process also provides estimates of the relative poses of the cameras in the considered time steps.

A high degree of robustness of the obtained 3D point cloud to outliers and noisy image points, is ensured by the geometric constraints used, and by filtering the data through different adjustable thresholds. Each point is visible in no fewer than four images, and matched features must satisfy epipolar geometry in stereo matching; RANSAC filtering algorithm ensures homography constraints and motion coherence of the rigid 3D structure; furthermore, it is possible to set the minimum distance of points from camera, below which they are rejected. Eventually, both the estimated motion and the 3D structure are refined through a local Sparse Bundle Adjustment procedure [215].

The output of the tracking algorithm is the 6-DOF incremental pose represented with a six-element vector of estimated relative pose, arranged with three Euler angles and three-translation component. Together with the estimated relative pose the algorithm, it provides the uncertainty of measurement, and the whole covariance matrix, essential for combining data from different sensors. The incremental pose and the associated covariance are fed to the EKF, which merges this data with measurements provided from the others sensors, calculating a better estimation of the vehicle pose with respect to the fixed world reference frame. Consequently, this new pose is exploited to correct the previous optical estimation and to enhance subsequent point cloud registration, so allowing a more correct incremental 3D reconstruction of the scene.

### ***Stereo Tracking EKF Observation Model***

We can assume the reference frame  $\{c\}$  placed in the optical centre of the left camera and  $x_{vc}$  the vector that describes the pose of  $\{c\}$  with respect to the vehicle reference frame  $\{v\}$ .  $x_{vc}$  is static and known from vehicle configuration and calibration. Assuming known the vehicle position at  $i$ -th step with respect to the local frame that corresponds to the static part of Kalman state vector, the camera pose in  $\{\ell\}$  can be expressed as:

$$x_{\ell c_i} = x_{\ell v_i} \oplus x_{vc};$$

The six elements of stereo-tracking output vector is expressed with respect to a previous step reference frame and can be denoted as  $x_{c_i c_k}$ , that represents the relative pose between the position at a generic step  $k$  and the current position at step  $i$  (with  $k < i$ ). The measurement vector  $z_i$  related to stereo-tracking can be linked to the state vector elements through the following:

$$\begin{aligned} z_i = x_{c_i c_k} &= x_{c \ell_i} \oplus x_{\ell c_k} = \ominus x_{\ell c_i} \oplus x_{\ell c_k} = x_{cv} \oplus x_{v \ell_i} \oplus x_{\ell v_k} \oplus x_{vc}; \\ z_i &= \ominus (x_{\ell v_i} \oplus x_{vc}) \oplus x_{\ell v_k} \oplus x_{vc}; \end{aligned} \quad \text{Eq. V-1}$$

As we previously said, the better estimation for  $x_{\ell v}$  is  $x_p$ , the static part of the current state vector  $x$ , so we can define the observation model of the optical part, affected by the error  $v_c$ , a vector of zero-mean white Gaussian noise, as:

$$z_{c_i} = h_c(x_i) + v_{c_i} = \ominus (x_{p_i} \oplus x_{vc}) \oplus x_{p_{last\_opt}} \oplus x_{vc} + v_{c_i}; \quad \text{Eq. V-2}$$

Note that this is true also for any other previous step in which optical data has been used.

The defined  $h_c$  is derived with respect to the 18 state variables in order to obtain the Jacobian matrix  $H_c$  concerning the camera measurements.

$$H_{c_{m,n}} = \frac{\partial h_{c_m}}{\partial x_n} = \partial(h_{c_m}, x_n);$$

Where  $x_n$  is the  $n$ -th variables of eighteen-elements state vector  $x$ ,  $h_{c_m}$  is the  $m$ -th equation of the system that links the six-element vector  $z_c$  to the vector  $x$ , and  $H_{c_{m,n}}$  means  $(m, n)$  element of the 6x18 matrix  $H_c$ . Moreover, we define  $\partial(a, b)$  as the partial derivation operation of  $a$  with respect to  $b$ . In particular:

$$H_{c_i} = \begin{bmatrix} \partial(h_c(\theta), x_{p_i}) & 0_{3 \times 6} & \partial(h_c(\theta), x_{p_{last\_opt}}) \\ \partial(h_c(t), x_{p_i}) & 0_{3 \times 6} & \partial(h_c(t), x_{p_{last\_opt}}) \end{bmatrix};$$

Where:

- $h_c(\theta)_i$  and  $h_c(t)_i$  represent two parts of  $h_c$  equation system, respectively related to the orientation and to the translation observation variables
- $x_{p_i}$  and  $x_{p_{last\_opt}}$ , as previously defined, are the pose information about the current and previous “optical” step, both stored in the state vector of the current iteration.

Obviously, at evaluation time, the best estimation of the real state  $x_i$  available to evaluate  $H_c$  matrix is its *a priori* estimation  $\hat{x}_i^-$ , whereas we can use the *a posteriori* estimation  $\hat{x}_{last\_opt}$ , stored in the state vector.

In order to include data from optical tracking in the EKF and ASKF [216] to perform SLAM, the covariance of the estimated parameters must be calculated.

Figure V-5 shows a single step from time  $t_0$  to time  $t$  done by the stereo-rig. Let us denote  $R_s, t_s$  the fixed transformation between left and right cameras of the stereo-rig and  $\hat{R}(\varphi, \vartheta, \psi)$ ,  $\hat{t} = (t_x, t_y, t_z)$  denote the estimated motion performed by the stereo-rig. World points  $M_i$  are imaged in all camera frames (i.e. in the left and right cameras at time  $t_0$  and in the left and right cameras at time  $t$ ) at positions  $ml_0^i, mr_0^i, ml^i, mr^i$ , respectively.

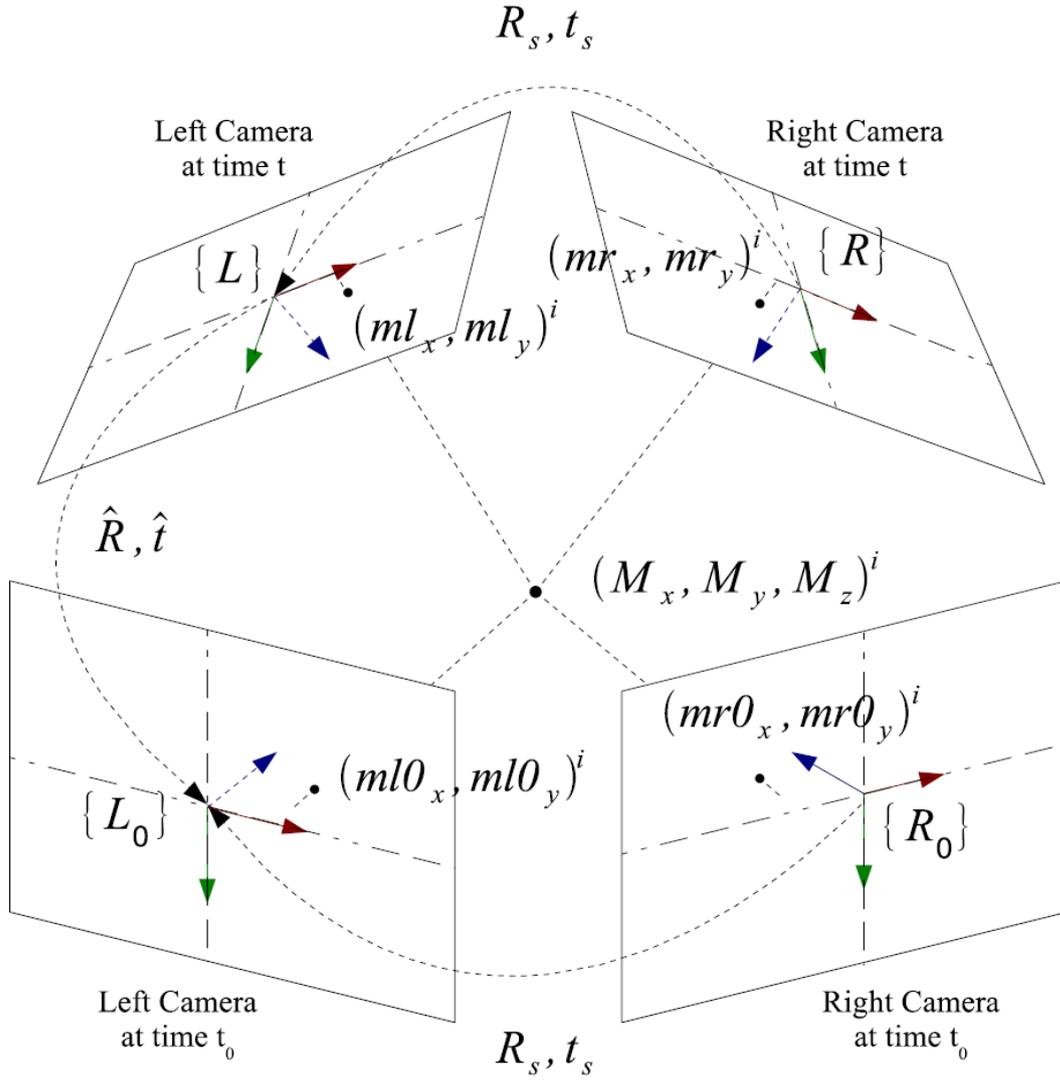


Figure V-5 Point  $M_i$  is imaged at time  $t_0$  in the left and right camera of the stereo-rig at positions  $ml0^i$  and  $mr0^i$  respectively. Once the stereo-rig have moved the same world point  $M_i$  is imaged in left and right camera coordinates  $ml^i$  and  $mr^i$ . The stereo rig has a rigid transformation  $R_s, t_s$  between left and right cameras. The estimated motion that the stereo-rig suffered is denoted by  $\hat{R}, \hat{t}$  expressed with respect to the left camera frame.

Let us denote  $\Psi$  the vector of all estimated parameters during a Stereo Tracker step,

$$\Psi = (\varphi, \vartheta, \psi, t_x, t_y, t_z, M_x^1, M_y^1, M_z^1, \dots, M_x^n, M_y^n, M_z^n);$$

Where  $\varphi, \vartheta, \psi$ , and  $t_x, t_y, t_z$ , define, respectively, the Euler angles and the translation between left camera at time  $t_0$  and time  $t$ . Therefore the left camera  $L$  is expressed as  $L = \hat{R}L_0 + \hat{t}$ , where  $\hat{R}$  is the rotation represented in mobile axes convention [217],

Let  $X$  be the vector of observed values containing the projections of all world points  $M_i$  in all the cameras

$$\begin{aligned}
 X = & (ml0_x^1, ml0_y^1, mr0_x^1, mr0_y^1, ml_x^1, ml_y^1, mr_x^1, mr_y^1, \\
 & ml0_x^2, ml0_y^2, mr0_x^2, mr0_y^2, ml_x^2, ml_y^2, mr_x^2, mr_y^2, \\
 & \dots, \\
 & ml0_x^n, ml0_y^n, mr0_x^n, mr0_y^n, ml_x^n, ml_y^n, mr_x^n, mr_y^n);
 \end{aligned}$$

where,  $ml0^i$  and  $mr0^i$  are the projections of  $M_i$  in the left and right cameras at time  $t_0$ , respectively, and  $mli$  and  $ml^i$  and  $mr^i$  are the projections of  $M_i$  in the left and right cameras at time  $t$ .

The method described by Haralick [218] propagates the covariance of an input variable  $X$  perturbed by additive random noise through any kind of calculation that extremizes a function  $F(X, \Psi)$ . The method determines the covariance of the  $\Psi$  parameters without the knowledge of how the minimization was computed. It is only assumed that the function  $F(X, \Psi)$  has finite second partial derivatives and that the perturbations are small enough to be correctly approximated by a first order Taylor series expansion.

In the case of the stereo rig, the function  $F(X, \Psi)$  that is being minimized is the reprojection error for all the world points projected onto all the camera frames defined in the following equation:

$$F(X, \Psi) = \underset{\hat{P}_k, \hat{M}_i}{\operatorname{argmin}} \sum_{k=1}^m \sum_{i=1}^n d(m_{ki}, \hat{P}_k \hat{M}_i)^2; \quad \text{Eq. V-3}$$

where  $P_k$  are the camera projection matrices  $k \in (L_0, R_0, L, R)$ ,  $M_i$  is the set of  $i$  3D world points and  $m_{ki}$  their respective projections in the camera frames. Operator  $d(\cdot, \cdot)$  is defining the Euclidean distance. Expanding the Eq. V-3 we obtain the function detailed for every camera frame:

$$\begin{aligned}
 F(X, \Psi) = \underset{\hat{P}_k, \hat{M}_i}{\operatorname{argmin}} \sum_{k=1}^m \sum_{i=1}^n & d(ml0_i, K_l \cdot \hat{M}_i)^2 + \\
 & d(mr0_i, K_r \cdot (R_s \cdot \hat{M}_i + t_s))^2 + \\
 & d(ml_i, K_l \cdot (\hat{R} \cdot \hat{M}_i + \hat{t}))^2 + \\
 & d(mr_i, K_r \cdot (R_s \cdot (\hat{R} \cdot \hat{M}_i + \hat{t}) + t_s))^2;
 \end{aligned}$$

where  $K_l$  and  $K_r$  are the intrinsic parameters matrices for the left and the right camera, respectively. Finally, the following equation defines the parameter covariance estimator  $\Sigma_\Psi$  in function of the partial derivatives of  $F(X, \Psi)$  and the observations covariance  $\Sigma_X$ :

$$\Sigma_\Psi \approx \frac{\partial g^{-1}}{\partial \Psi}(X, \Psi) \cdot \frac{\partial g^T}{\partial X}(X, \Psi) \cdot \Sigma_X \cdot \frac{\partial g}{\partial X}(X, \Psi) \cdot \frac{\partial g^{-1}}{\partial \Psi}(X, \Psi)$$

Where  $g(X, \Psi) = \frac{\partial F}{\partial \Psi}$  and  $\Sigma_X$  is a diagonal matrix containing the noise variance of the observed projections  $ml0_i$ ,  $mr0_i$ ,  $mr0_i$  and  $mr_i$ . A naive implementation of this method can solve the problem for a fixed number of 3D world points  $M_i$  by constructing the respective matrices of partial derivatives of  $F(X, \Psi)$  (i.e.  $\frac{\partial^2 F}{\partial \Psi \partial \Psi} = \frac{\partial g}{\partial \Psi}$  and  $\frac{\partial^2 F}{\partial X \partial \Psi} = \frac{\partial g}{\partial X}$ ). However, the number of 3D points depends on the stereo-rig step: at every consecutive pair of stereo rig positions, a different number of 3D points can be estimated and is highly dependent on the scene content. Nevertheless, for what concerns the covariance needed in filter, only data about the first six parameters of  $\Psi$  is needed. It comes from  $B$ , the upper left corner of  $\frac{\partial F}{\partial \Psi \partial \Psi}$  :

	$\varphi$	$\vartheta$	$\psi$	$t_x$	$t_y$	$t_z$	$M_x^1$	$M_y^1$	$M_z^1$	...	$M_x^n$	$M_y^n$	$M_z^n$
$\varphi$	$\frac{\partial^2 F}{\partial^2 \varphi}$	$\frac{\partial^2 F}{\partial \varphi \partial \vartheta}$	$\frac{\partial^2 F}{\partial \varphi \partial \psi}$										
$\vartheta$	$\frac{\partial^2 F}{\partial \vartheta \partial \varphi}$	$\frac{\partial^2 F}{\partial^2 \vartheta}$			...								
$\psi$	$\frac{\partial^2 F}{\partial \psi \partial \varphi}$	$F$	$\frac{\partial^2 F}{\partial^2 \psi}$					$A_1^1$		$A_1^i$		$A_1^n$	
$t_x$				$B$									
$t_y$		$\vdots$				$\ddots$							
$t_z$													
$M_x^1$													
$M_y^1$			$A_2^1$					$A_3^1$		0		0	
$M_z^1$													
...			$A_2^i$					0		$A_3^i$		0	
$M_x^n$													
$M_y^n$			$A_2^n$					0		0		$A_3^n$	
$M_z^n$													

Other three set of equation ( $A_p, p \in [1, 2, 3]$ ) have been identified, that can be particularized for the correspondent 3D point  $M^i$  and the respective projections in all the camera frames.

The whole  $\frac{\partial^2 F}{\partial \Psi \partial \Psi}$  matrix and also the  $\frac{\partial^2 F}{\partial X \partial \Psi}$  matrix are evaluated in order to determine the covariance of 3D points but only the  $B$  portion is feed to the EKF as:

$$R_c = B;$$

For the sake of clarity,  $R_c$  is the covariance matrix associated to camera observation used in the filter.

### **V.4.2 Navigation Sensor Measurements**

In addition to the vector of incremental pose coming from optical data, other navigation data is incorporated. In particular, the other exploited sensors are:

- *a DVL that provides velocities in the sensor frame;*
- *an Attitude and Heading Reference System (AHRS), which tracks the orientation and gives the absolute angular pose with respect to the fixed  $\{\ell\}$  frame ;*
- *a pressure gauge with the depth as output.*

In the following subsections, we analyse each sensor output in order to describe its relation with our state vector and the resulting observation model.

#### **DVL EKF Observation Model**

The DVL output  $z_d$  consists of a vector of three instant velocities expressed in the DVL mobile frame  $\{d\}$ ; its pose is known and described by  $x_{vd}$  with respect to  $\{v\}$ . DVL measurements are obviously related both with linear and angular velocities of the vehicle. In particular, if we express  $x_{vd}$  by means of a rotation matrix  ${}^v_dR$  and a translation vector  $t_{vd}$ , we can write the following:

$$\begin{aligned} z_d = \begin{bmatrix} vx \\ vy \\ vz \end{bmatrix}_d &= {}^v_dR^{-1} \left( \begin{bmatrix} vx \\ vy \\ vz \end{bmatrix}_v + \begin{bmatrix} \omega\varphi \\ \omega\vartheta \\ \omega\psi \end{bmatrix}_v \times t_{vd} \right) + v_{di} \\ &= {}^v_dR^{-1} \left( {}^v_wR \begin{bmatrix} vx \\ vy \\ vz \end{bmatrix}_w + \left( {}^v_wR \begin{bmatrix} \omega\varphi \\ \omega\vartheta \\ \omega\psi \end{bmatrix}_w \right) \times t_{vd} \right) + v_{di}; \end{aligned}$$

In which  ${}^v_wR$  is the rotation matrix that links  $\{v\}$  frame with the fixed world frame  $\{w\}$ .

Linear and angular velocities are estimated by our EKF and stored in  $x_k$ , the kinematic part of the defined state vector. In particular, we can remember that:

$$\begin{bmatrix} \omega\varphi \\ \omega\vartheta \\ \omega\psi \end{bmatrix}_w = x_k(1, 2, 3) = x_k(\omega); \quad \begin{bmatrix} vx \\ vy \\ vz \end{bmatrix}_w = x_k(4, 5, 6) = x_k(v);$$

In which  $x_k(\omega)$  and  $x_k(v)$  are, respectively, the angular and linear velocities of vehicle expressed in  $\{w\}$  frame, stored in kinematic part of state vector.

Hence, we can define the observation function  $h_d$  for the DVL sensor at  $i$ -th step and write the following observation model, affected by the error  $v_d$ , a vector of zero-mean white Gaussian noise:

$$z_{d_i} = h_d(x_i) + v_{d_i} = {}^v_d R^{-1} \left( {}^v_w R \cdot x_k(v) + ({}^v_w R \cdot x_k(\omega)) \times t_{vd} \right) + v_{d_i};$$

The linearized model requires  $H_d$ , the Jacobian matrix of partial derivatives of  $h_d$  with respect to the state variables:

$$H_{d_{m,n}} = \frac{\partial h_{d_m}}{\partial x_n} = \partial(h_{d_m}, x_n);$$

Where  $x_n$  is the  $n$ -th variables of eighteen-elements state vector  $x$ ,  $h_{d_m}$  is the  $m$ -th equation of the system that links the three-element vector  $z_d$  to the vector  $x$ , and  $H_{d_{m,n}}$  means  $(m, n)$  element of the 3x18 matrix  $H_d$ :

$$H_{d_i} = \begin{bmatrix} 0_{3 \times 6} & \partial(h_{d_i}, x_k) & 0_{3 \times 6} \end{bmatrix};$$

where  $x_{p_i}$ , as previously defined, is the pose part of the state vector of the  $i$ -th iteration.

DVL covariance has not been analytically evaluated; in effect, the employed sensor provides an estimate of the error for each measurement. This error  $e_d$  can be used to arrange the covariance matrix of DVL observation,  $R_d$ , assuming the measurement uncorrelations:

$$R_{d_i} = \begin{bmatrix} e_{d_i}^2 & 0 & 0 \\ 0 & e_{d_i}^2 & 0 \\ 0 & 0 & e_{d_i}^2 \end{bmatrix};$$

### **AHRS EKF Observation Model**

The Attitude and Heading Reference System output  $z_f$  consists of a vector of three Euler angles expressed in the AHRS fixed frame  $\{a\}$ ; its pose is known and described by  $x_{va}$  with respect to  $\{v\}$ . The sensor pre-processes mainly data from fibre optic gyroscope, accelerometer and compass, providing the three absolute Euler angles that describes the sensor (and vehicle) orientation in  $\{a\}$ . The simple relation that links measurements and state vector orientation variables only requires the relative pose between  $\{a\}$  and  $\{w\}$ :

$$z_{a_i} = h_a(x_i) + v_{a_i} = {}^a_w R \cdot x_{p_i}(\theta) + v_{a_i};$$

In which  $x_{p_i}(\theta) = x_p \quad (1, 2, 3) = \begin{bmatrix} \varphi \\ \vartheta \\ \psi \end{bmatrix}_w$ ; consequently, the  $H_a$  matrix can be defined as:

$$H_{a_{m,n}} = \frac{\partial h_{a_m}}{\partial x_n} = \partial(h_{a_m}, x_n);$$

Where  $h_{a,m}$  is the  $m$ -th equation of the system that links the three-element vector  $z_a$  to the vector  $x$ , and  $H_{a,m,n}$  means  $(m, n)$  element of the  $3 \times 18$  matrix  $H_a$ :

$$H_{a_i} = \begin{bmatrix} \partial \left( h_a, x_{p_i}(\theta) \right) & 0_{1 \times 15} \end{bmatrix};$$

Also the covariance  $R_a$  associated with AHRS has not been analytically evaluated; the employed sensor does not provide any estimation of the error. Hence, in this case the error  $e_a$  has been evaluated exploiting acquired measurements.

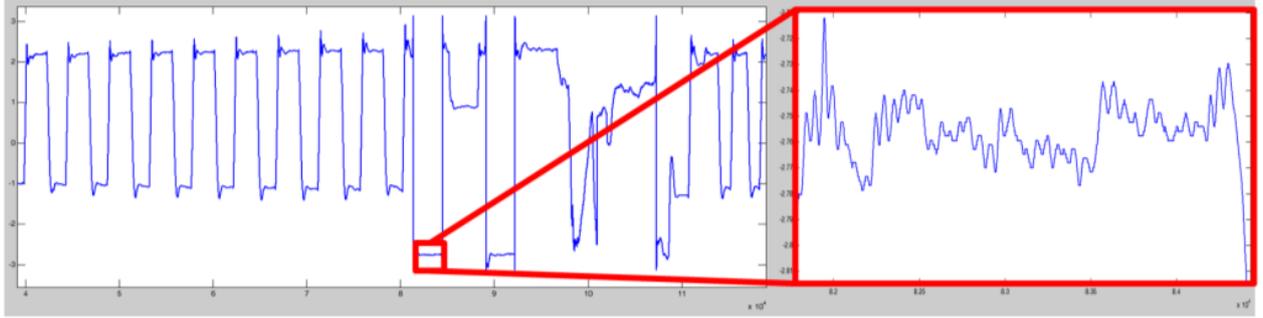


Figure V-6 Plot of heading values. Considering intervals in which the observed parameter is supposed to be constant, it is possible to evaluate noise and, consequently, a good approximation for sensor covariance.

In effect, considering intervals of acquisition relative to constant values of interested variable, it is possible to calculate the mean value and the standard deviation  $\sigma_a$ , used to arrange the covariance matrix  $R_a$ , always assuming the uncorrelation of measurements:

$$R_{a_i} = \begin{bmatrix} \sigma_{a_i}^2 & 0 & 0 \\ 0 & \sigma_{a_i}^2 & 0 \\ 0 & 0 & \sigma_{a_i}^2 \end{bmatrix};$$

### Depth gauge EKF Observation Model

Finally, we consider the Depth Gauge sensor. Its output  $z_g$  is a scalar representing the depth according to sensor reference frame  $\{g\}$ . Assuming small values for roll and pitch angles, it can be directly linked to the  $z$ -coordinate of world reference system. The only required adjustment is the offset between  $\{v\}$ , the vehicle reference system, and  $\{g\}$ :

$$z_{g_i} = h_g(x_i) + v_{g_i} = x_{p_i}(z) + t_{gv}(z) + v_{g_i};$$

In which  $x_{p_i}(z) = [z]_w$  and  $t_{gv}(z)$  is the  $z$ -coordinate of the translation vector between  $\{g\}$  and  $\{v\}$ ; consequently, the  $H_g$  matrix can be simply defined as:

$$H_{g_i} = \frac{\partial h_g}{\partial x_n};$$

and  $H_{a_n}$  means  $(1, n)$  element of the  $1 \times 18$  matrix  $H_g$ :

$$H_{g_i} = [0_{1 \times 5} \quad \partial(h_g, x_{p_i}(z_i)) \quad 0_{1 \times 12}] = [0_{1 \times 5} \quad 1 \quad 0_{1 \times 12}];$$

The covariance  $R_g$  associated with Depth Gauge has been evaluated exploiting the acquired measurements, as previously described for AHRS.

$$R_{g_i} = [\sigma_{g_i}^2];$$

## V.5 Extended Kalman Filter - EKF

The implemented EKF incorporates the 6DOF incremental pose representation coming from optical data and other navigation data from different sensors. If the number of matched features in the subsequent pairs is lower than a minimum threshold, the filter only considers the observations from the other sensors. In particular, the absolute angular pose is measured by an AHRS, the depth by a pressure gauge, and velocities, in sensor local frame, by a Doppler Velocity Log (DVL). The described mathematical model describes the evolution of the system state in order to predict its values (in our case the vehicle pose). Evaluated observation models are required to relate the different sensor measurements with the desired variables. The filter merges all the information coming from sensors and system model to estimate the interested state variables, according to their associated errors and covariances.

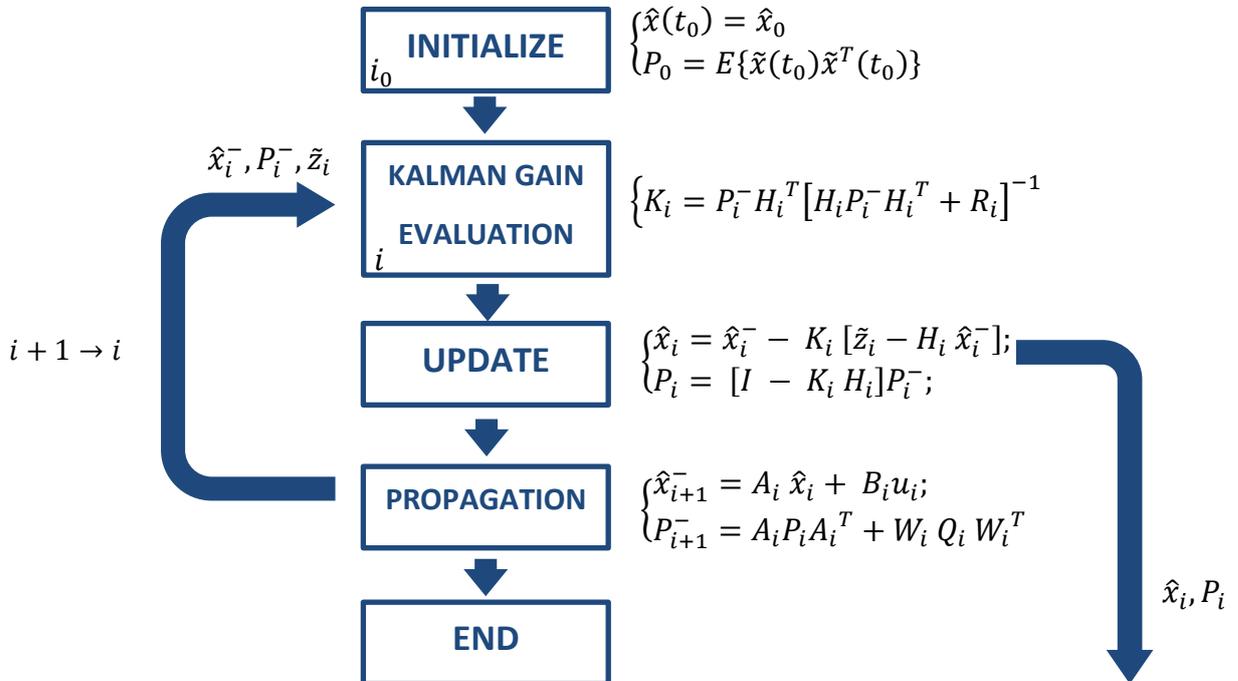


Figure V-7 EKF scheme

The algorithm follows the classical EKF implementation, schematized in Figure V-7.

State and covariance are initialized with a proper value. After that the Kalman filter is iterated until needed. In the loop, first, the available measurements and associated errors and the *a priori* state and error covariance, estimated in the previous step, are exploited to evaluate the Kalman gain; then the predictions are corrected in the Update phase; finally, in the propagation step, prediction for state and error covariance of the successive iteration are estimated. Each iteration provides the *a posteriori* estimation of state and error covariance.

In our case, the algorithm is tested with an off-line dataset. It assumes a given starting pose for the first iteration, associated with the first desired stereo-pair. Then the code reads DVL measurements, which have been acquired with a sample time of about 300 milliseconds, and, based on synchronised time-stamps, it search for other concurrent sensor data.

Cameras are noticeably slower than other sensors; new images are normally available every two seconds and optical tracking algorithm provides data relative to the previous optical observation. For these reasons, optical data cannot be considered at each iteration and it is necessary to store in the state vector and in the error covariance matrix information about the last pose corresponding to a camera acquisition. This means that  $x_{p_{last\_opt}}$  part of the state vector is exploited (and obviously updated), only when a new stereo-pair is available to be processed by the Stereo Tracking algorithm. In case of low overlapping subsequent stereo pairs, it is not possible to evaluate the relative camera pose needed by EKF, and optical tracking should terminate. Nevertheless, in this eventuality, the absolute pose, evaluated by the filter only considering the sensor observations, is stored in  $x_{p_{last\_opt}}$  in order to provide a new consistent starting point for further optical tracking.

The EKF is written to be flexible with respect to the available sensors and, in theory, is able to work also without any input, just following the system model. In effect, the current observation vector, matrix, and covariance are obtained by properly assembling the entities related to each single sensor; hence, they can be different for each step both in values and in size. When information from all sensors are available they configures as following:

$$\tilde{z}_i = \begin{bmatrix} \tilde{z}_{c_i} \\ \tilde{z}_{a_i} \\ \tilde{z}_{d_i} \\ \tilde{z}_{g_i} \end{bmatrix}_{13 \times 1} ; \quad H_i = \begin{bmatrix} H_{c_i} \\ H_{a_i} \\ H_{d_i} \\ H_{g_i} \end{bmatrix}_{13 \times 18} ; \quad R_i = \begin{bmatrix} R_{c_i} & 0_{6 \times 3} & 0_{3 \times 3} & 0_{3 \times 1} \\ & R_{a_i} & 0_3 & 0_{3 \times 1} \\ & & R_{d_i} & 0_{3 \times 1} \\ Symm. & & & R_{g_i} \end{bmatrix}_{13 \times 13} ;$$

If any of the sensor is missing, then the associated vector and matrices are empty, erasing respective row and/or column. For example if DVL is not available for  $k$ -th iteration, then we have:

$$\tilde{z}_k = \begin{bmatrix} \tilde{z}_{c_k} \\ \tilde{z}_{a_k} \\ \tilde{z}_{g_k} \end{bmatrix}_{10 \times 1} ; \quad H_k = \begin{bmatrix} H_{c_k} \\ H_{a_k} \\ H_{g_k} \end{bmatrix}_{10 \times 18} ; \quad R_k = \begin{bmatrix} R_{c_k} & 0_{6 \times 3} & 0_{3 \times 1} \\ & R_{a_k} & 0_{3 \times 1} \\ \text{Symm.} & & R_{g_k} \end{bmatrix}_{10 \times 10} ;$$

After the estimation of the pose, the 3D point structure and the associated uncertainty, both previously obtained from vision data only, are corrected according to the EKF estimated pose.

## **V.6 Augmented State Kalman Filter**

The Augmented-State EKF is a particular implementation that allows to store the whole set of previous poses and to add constraints related to loop closure events exploiting optical data. In the presented test case, the filter state is augmented each time a new image is available, storing each pose associated with a stereo-pair acquisition. The augmentation also regards the error covariance matrix, which grows fast in size, together with the other matrices involved (i.e.  $H, K, A, W$ ).

In the following sections, we explain how the augmentation is performed in detail, and how the ASKF works both in case of normal open sequence and in case of loop closure event.

### **V.6.1 Open sequence case**

The open sequence case is obviously the most frequent one occurring in the estimation. In this case, the filter works simply solving the described equation for the state and for the error covariance prediction and update.

When a new optical data is available, related information about relative pose is added to the current measurement vector. After each step corresponding to a camera acquisition pose, both the model state and covariance are augmented. The augmentation results in increasing their size storing information about current step and its relation with the previous ones.

We define as  $x_{pc_j}$  the vehicle pose vector relative to  $j$ -th optical acquisition. If optical data is available at current  $i$ -th iteration, the evaluated  $\hat{x}_{p_i}$  is assumed as current  $x_{pc_j}$  and stored in the state vector. As previously said, in the EKF algorithm this operation is also needed and performed

using the  $x_{p_{last\_opt}}$  part of the state vector. The difference in ASKF implementation consists in not substituting old values but adding new ones, so increasing state vector size.

$$\begin{aligned}
 \hat{x}_0 &= [\hat{x}_{p_0} \quad \hat{x}_{k_0} \quad \hat{x}_{pc_0}]^T & ; & \quad \hat{x}_{pc_0} = \hat{x}_{p_0} & \leftarrow \text{Start Optical Data} \\
 \hat{x}_1 &= [\hat{x}_{p_1} \quad \hat{x}_{k_1} \quad \hat{x}_{pc_0}]^T \\
 \hat{x}_2 &= [\hat{x}_{p_2} \quad \hat{x}_{k_2} \quad \hat{x}_{pc_1} \quad \hat{x}_{pc_0}]^T & ; & \quad \hat{x}_{pc_1} = \hat{x}_{p_2} & \leftarrow \text{New Optical Data} \\
 & \vdots \\
 \hat{x}_i &= [\hat{x}_{p_i} \quad \hat{x}_{k_i} \quad \hat{x}_{pc_j} \quad \dots \quad \hat{x}_{pc_0}]^T & ; & \quad \begin{cases} \hat{x}_{pc_j} = \hat{x}_{p_i}; & \text{New Optical Data} \\ \hat{x}_{pc_j} \neq \hat{x}_{p_i}; & \text{No New Optical Data} \end{cases}
 \end{aligned}$$

In our case, 6 elements, corresponding to the current pose, are added to the state vector; in addition, information about current error covariance (6x6 matrix) is integrated in the covariance matrix.

A generic  $P_i$  matrix is schematised in Table V-2. The various blocks are moved and duplicated in order to obtain the  $P_{i_{aug}}$  augmented matrix, represented in

Table V-3.

Table V-2 Generic  $P_i$  matrix

	$x_{p_i}$	$x_{k_i}$	$x_{pc_{j-1}}$	$x_{pc_{j-2}}$	$\dots$	$x_{pc_1}$	$x_{pc_0}$
$x_{p_i}$	$P_{x_p x_p}$	$P_{x_p x_k}$	$P_{x_p x_{pc}}$				
$x_{k_i}$	$P_{x_k x_p}$	$P_{x_k x_k}$	$P_{x_k x_{pc}}$				
$x_{pc_{j-1}}$	$P_{x_{pc} x_p}$	$P_{x_{pc} x_k}$	$P_{x_{pc} x_{pc}}$				
$x_{pc_{j-2}}$							
$\vdots$							
$x_{pc_1}$							
$x_{pc_0}$							

Table V-3 Scheme of a generic  $P_i$  augmented matrix

	$x_{p_i}$	$x_{k_i}$	$x_{pc_j} \equiv x_{p_i}$	$x_{pc_{j-1}}$ $x_{pc_{j-2}}$ $\dots$ $x_{pc_1}$ $x_{pc_0}$
$x_{p_i}$	$P_{x_p x_p}$	$P_{x_p x_k}$	$P_{x_p x_{pc_j}}$	$P_{x_p x_{pc}}$
$x_{k_i}$	$P_{x_k x_p}$	$P_{x_k x_k}$	$P_{x_k x_{pc_j}}$	$P_{x_k x_{pc}}$
$x_{pc_j}$	$P_{x_{pc_j} x_p}$	$P_{x_{pc_j} x_k}$	$P_{x_{pc_j} x_{pc_j}}$	$P_{x_{pc_{j+1}} x_{pc}}$
$x_{pc_{j-1}}$				
$x_{pc_{j-2}}$				
$\vdots$	$P_{x_{pc} x_p}$	$P_{x_{pc} x_k}$	$P_{x_{pc} x_{pc_j}}$	$P_{x_{pc} x_{pc}}$
$x_{pc_1}$				
$x_{pc_0}$				

Notice that, as previously said, when a new camera observation is available,  $x_{pc_j} \equiv x_{p_i}$ . The notation  $P_{a|b}$  refers to the covariance of  $a$  with respect to  $b$ , where  $a$  and  $b$  are related to a variable or to a group of variables. In particular:

$x_p \equiv x_{pc_j}$  = state variables related to the pose part of the state vector (6 elements).

$x_k$  = state variables related to the kinematic part of the state vector (6 elements).

$x_{pc}$  = state variables related to the part of the state vector that stores the previous vehicle poses corresponding to a camera acquisition occurrence ( $n-12$  elements). We highlight also that the dimension of the state vector  $n$ , at  $i$ -th iteration, respects the relation:

$$n = 6 \cdot n_{frame} + 12$$

In which  $n_{frame}$  is the number of stereo pairs included in the ASKF at  $i$ -th step.

Obviously,  $A$  and  $W$  matrices also need to be augmented in order to correctly match the size of the current state vector, defined as  $n$ . After each step, the augmentation of  $A$  and  $W$  is performed as follows:

$$A_{aug} = \begin{bmatrix} A & 0_6 \\ 0_6 & I_6 \end{bmatrix}; \quad W_{aug} = \begin{bmatrix} W \\ 0_6 \end{bmatrix};$$

The filter requires that  $H$  matrix incorporate information about relations between current measurement vector and all state vector variables. Hence, in general, the  $H$  matrix evaluated has to be augmented with a series of zeros in order to adjust its dimension:

$$H_{aug} = [H \quad 0_{m \times (n-18)}];$$

Where  $m$  is the size of the current measurement vector.

This is valid only in case of open sequence; in fact, when a loop closure event occurs, current optical measurement is related to more previous states and not only to the last one.

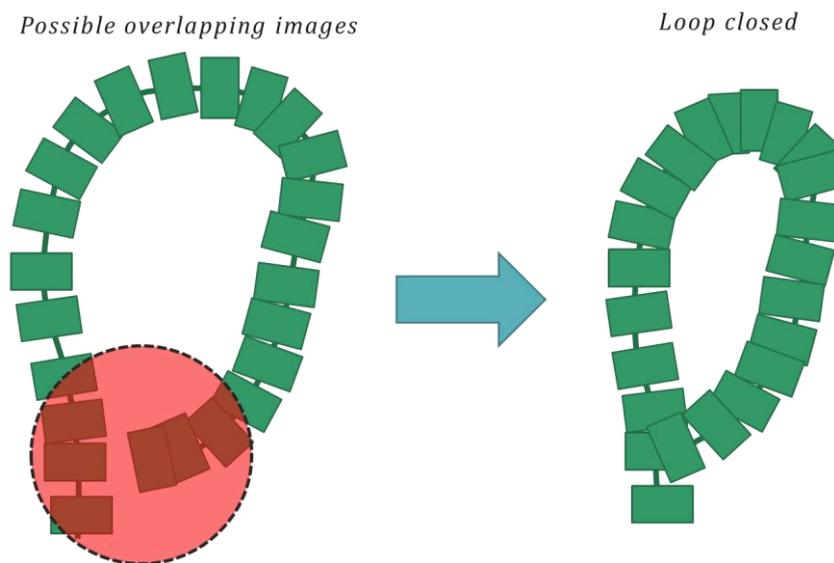
The augmented vector and matrices are then used in the propagation step, and the new iteration can be normally performed with Kalman gain evaluation and with updating step.

### ***V.6.2 Loop closure case***

At each step, after estimating the current pose, candidates for loop closure are searched for.

Closing loops allows partially recovering the drift error that affect dead-reckoning navigation, exploiting optical data to recognize already surveyed areas to adjust the current pose and the whole trajectory to fulfil the new constraint(s) (Figure V-8).

When a crossing path happens, the implemented algorithm also correct the 3D structure and the covariance associated to each 3D point of it.



*Figure V-8 Loop closure event*

Candidates for matching operation are chosen amongst all previous stored positions that are within a certain range around the current pose, considering, for simplicity, only the 2D position on the survey plane. Among them, using the same algorithm above mentioned for subsequent pairs,

we match the features in the quadruplet and if their number overcomes a threshold that ensures a satisfying overlapping, we estimate the relative pose of the current frame with respect to all the matched ones. This information is incorporated in ASKF by means of the observation vector and matrix, properly arranged. In particular for each matched image, the six variables describing the relative transformation between current and matched vehicle pose are attached to the measurement vector  $\tilde{z}_c$ ; supposing to match  $p$  previous frames he can define:

$$\tilde{z}_c = [\tilde{z}_c^1 \quad \tilde{z}_c^2 \quad \dots \quad \tilde{z}_c^p]^T;$$

Usually the first one of them is the last frame before the current (even if it is possible to have a current stereo-pair matching with other previous frames, but the last one). The ASKF receives the  $\tilde{z}_c$  vector and also information about matched frames and their position inside the  $\tilde{z}_c$  vector, in order to arrange the  $H_c$  matrix correctly. The sub-matrix  $H_{c_m}$ , relative to the  $m$ -th observation, remains very similar to the one defined for simple EKF, with the first part always related to the current predicted pose, and the second part relative to the interested previous pose and not to always to the last one:

$$H_{c_m} = \begin{bmatrix} \partial (h_c(\theta), x_{p_i}) & \dots 0 \dots & \partial (h_c(\theta), x_{pc_m}) & \dots 0 \\ \partial (h_c(t), x_{p_i}) & \dots 0 \dots & \partial (h_c(t), x_{pc_m}) & \dots 0 \end{bmatrix};$$

Supposing that  $\tilde{z}_c^1$  refers to  $x_{pc_{j-1}}$ ,  $\tilde{z}_c^2$  refers to  $x_{pc_l}$ , and so on until ,  $\tilde{z}_c^p$  that refers to  $x_{pc_p}$ , we can write a generic  $H_{c_i}$  matrix as:

	$x_{p_i}$	$x_{k_i}$	$x_{pc_{j-1}}$	...	$x_{pc_l}$	...	$x_{pc_p}$	...	$x_{pc_0}$
$\tilde{z}_c^1$	$\begin{matrix} \partial (h_c(\theta), x_{p_i}) \\ \partial (h_c(t), x_{p_i}) \end{matrix}$	$0$	$\begin{matrix} \partial (h_c(\theta), x_{pc_j}) \\ \partial (h_c(t), x_{pc_j}) \end{matrix}$	$0$	$0$	$0$	$0$	$0$	$0$
$\tilde{z}_c^2$	$\begin{matrix} \partial (h_c(\theta), x_{p_i}) \\ \partial (h_c(t), x_{p_i}) \end{matrix}$	$0$	$0$	$0$	$\begin{matrix} \partial (h_c(\theta), x_{pc_l}) \\ \partial (h_c(t), x_{pc_l}) \end{matrix}$	$0$	$0$	$0$	$0$
$\vdots$	$\vdots$	$0$	$0$	$0$	$0$	$\ddots$	$0$	$0$	$0$
$\tilde{z}_c^p$	$\begin{matrix} \partial (h_c(\theta), x_{p_i}) \\ \partial (h_c(t), x_{p_i}) \end{matrix}$	$0$	$0$	$0$	$0$	$0$	$\begin{matrix} \partial (h_c(\theta), x_{pc_m}) \\ \partial (h_c(t), x_{pc_m}) \end{matrix}$	$0$	$0$

Notice that each matrix cell is a 6x6 sub-matrix.

The described  $\tilde{z}_c$  and  $H_c$  are exploited to re-evaluate the current state vector with another ASKF iteration that corrects the current pose (used as starting point for loop detection) and all previous poses inside the loop. Finally, all the new poses and associated error covariances are used to correct the related 3D point structures.

In order to speed-up the process, features descriptors from each stereo-pair is stored in properly arranged matrices, so avoiding to perform feature extraction and the stereo-matching step more than once. Nevertheless, the considerably bigger size of involved matrices and the required matching and correction operations noticeably decrease performances for big image sets.

## ***V.7 Algorithm applications***

In this section, we illustrate test results obtained on two datasets acquired by Girona 500, the UUV unmanned vehicle, during two surveys: *La Lune* and *Cap de Vol* dataset.

### ***V.7.1 La Lune Dataset***

#### ***Archaeological context***

The first dataset is relative to *La Lune* shipwreck. *La Lune* was a three-masted, 38m long vessel, and was part of the fleet of King Louis XIVs (Figure V-9). It was lost soon after returning from the North of Africa, on November 6<sup>th</sup> 1664, with 350 sailors and at least 450 soldiers aboard. The ship had returned already in poor condition from a campaign in Djidjelli, nowadays part of Algeria [11]. Upon arriving to France, it was hastily refitted, so that it could be promptly used to carry a large amount of people from Toulon to the nearby island of Porquerolles for quarantine. The ship started taking in water shortly after leaving to port of Toulon, and eventually broke in two pieces, only a few miles from coast. The breaking and sinking happened in quick succession, which contributed to the large number of casualties [11]. More than 700 people perished in the accident. The wreck was discovered accidentally in the spring of 1993 during a test trial of the IFREMER submarine Nautilus [12]. Soon after its discovery, the wreck site was assessed by the French Departement des Recherches Archeologiques Subaquatiques et Sous-Marines (DRASSM), using a ROV [13]. Upon conclusion of that initial archaeological survey, the wreck was classified as one of the best preserved of its kind in the world. The site is considered as an experimental laboratory for the development and testing of new extraction methods for deep-water archaeology [14].

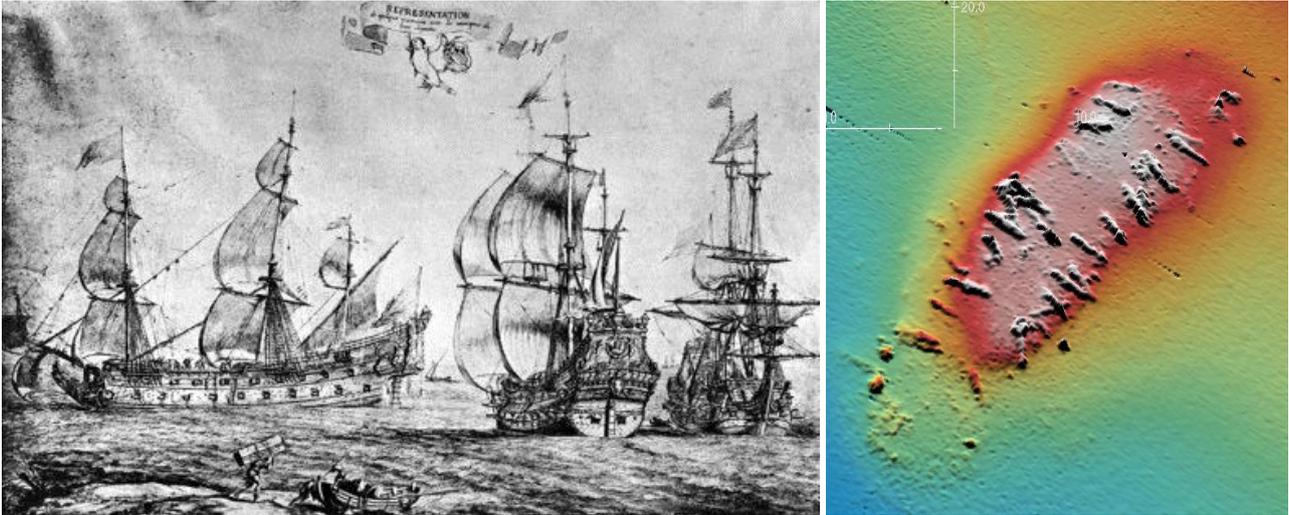


Figure V-9 *La Lune*. Left: the only known representation of the 'La Lune' (first ship on the left), from a 1690 painting. (@Musée National de la Marine, S. Dondain, France). Right: bathymetry of the shipwreck (*Girona*)

In early 2012, the site was mapped using a Kongsberg Maritime EM2040 multibeam echosounder, deployed from the AsterX AUV [15]. These data allowed for obtaining a detailed bathymetry of the site, with a spatial cell resolution of 10 cm. Subsequently, the site was optically surveyed in August 2012 with the *Girona 500* AUV [cit IEEE].

Due to hostile acquisition condition, the acquired dataset is particularly challenging, not only for the homogeneous sandy seabed that can produce wrong matched features; the hostile weather conditions caused navigation problems and the area extent required relatively fast survey that not always allowed to guarantee the correct overlapping area between subsequent stereo-pairs.

### **Results**

The presented work has been tested with the real dataset acquired by UVL team during the survey performed on the *La Lune* shipwreck with the *Girona 500* AUV. It is a high-reconfigurable, compact-size AUV developed by the University of Girona (Figure V-10).

Apart from cited navigation sensors, the AUV was equipped with a high-resolution stereo-rig. The stereo imaging system consists of two Canon EOS 5D Mark II cameras electronically synchronised, installed in the free payload area of the AUV together with the illumination system. They were in a down-looking configuration, pre-calibrated and they shot at 0.5Hz. The AUV surveyed at a nominal altitude of 3.68 m and at a nominal forward speed of 0.5 m/s. The total bottom time was 103 minutes, split into two surveys in which 2,757 stereo images were acquired.

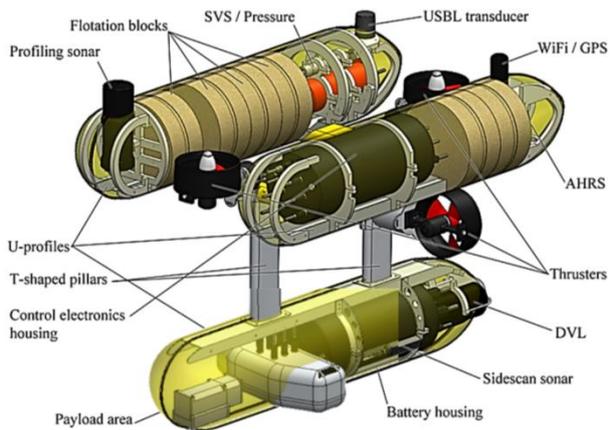


Figure V-10 Left: Schema of the Girona 500 AUV, showing the location of principal components and payload area. Right: Deploying of the Girona 500 AUV above the La Lune shipwreck.

Acquired images have been resized at 1404x936 pixels, a quarter of the original dimension, so decreasing processing time and also enhancing results, thanks to the implicit de-noising effect of rescaling operation. In these conditions, extracted features are limited at 15,000 and a value of 5 pixels is chosen for Non Maximum suppression Radius parameter.

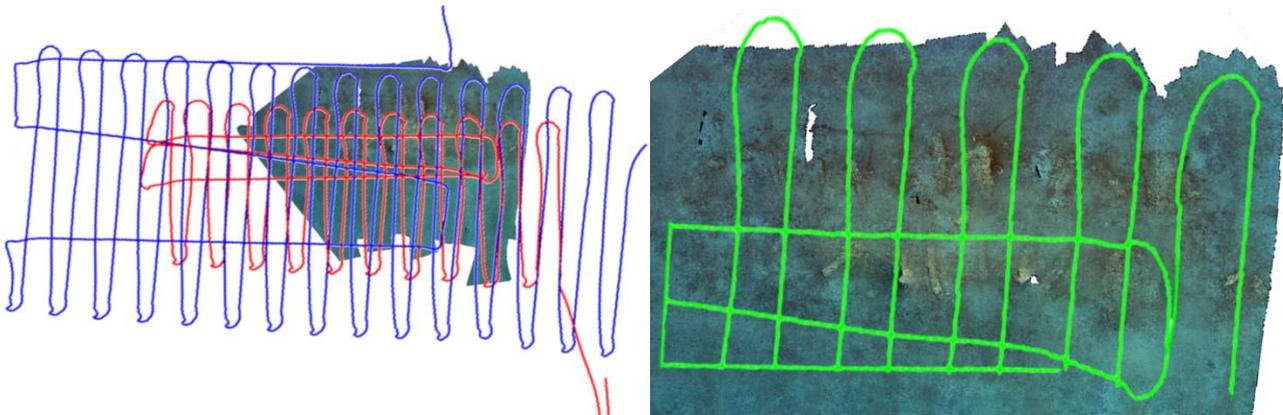
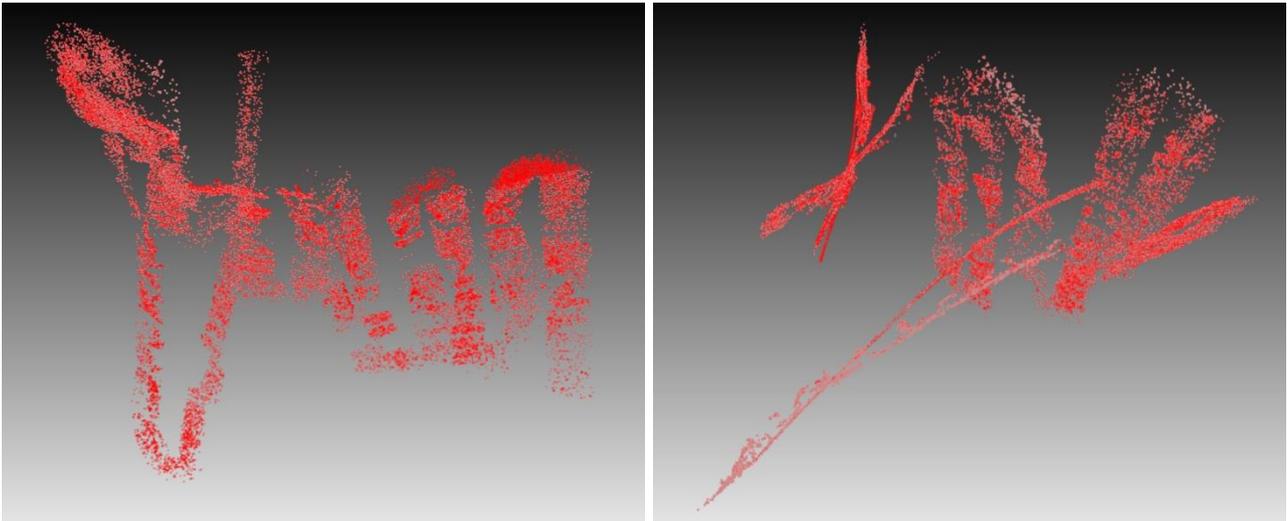


Figure V-11 Left: the whole trajectory surveyed by the AUV (two surveys have been performed on the interested area); right: detail of the interested area with a portion of the considered “virtual path”.

For testing purposes, a subset of 310 stereo images has been extracted and processed following the pipeline described above. The chosen images have been reordered thanks to the already available giga-mosaic obtained from the same images, in order to create a “virtual path” (Figure V-11). This choice is the consequence of the bad survey condition that obliged to perform two different surveys and acquire an area bigger than the interested one because of localisation low accuracy.

The *La Lune* site is basically flat and reliefs are limited to little structures like cannons; thus, in a sparse map they will not be appreciable; nevertheless, point cloud will be clearly show the “optical swath” imaged by the stereo rig.

The first attempt regarded the optical part. The full estimation of the trajectory (and map) was not possible due to low overlapping of subsequent stereo-pairs; in certain cases, there has been a complete gap between acquisitions, probably because of missing shots, and the algorithm totally failed. “Tricking” the algorithm (that requires subsequent images) by adding some shots taken from other swaths it has been possible to complete the trajectory using just the Stereo-Tracker modified algorithm. Resulting trajectory is estimated in reasonable time (about 2 sec/image, more or less equal to the surveying time), shown in Figure V-12, is heavily affected by errors mainly due to the lack of features matched in some low overlapping subsequent pairs.



*Figure V-12 Sparse map estimated by Stereo-Tracker code with standard parameters.*

Increasing the feature limit and density parameters, better trajectory estimation is obtainable, visible in Figure V-13. Here the seabed imaged is much more flat and trajectory followed by the vehicle becomes to be visible. Anyway, the trajectory and point cloud are absolutely far from being acceptable and useful. Moreover, time spent to produce this result is extremely high (more than 500 sec/images), and also required a final global Bundle Adjustment.

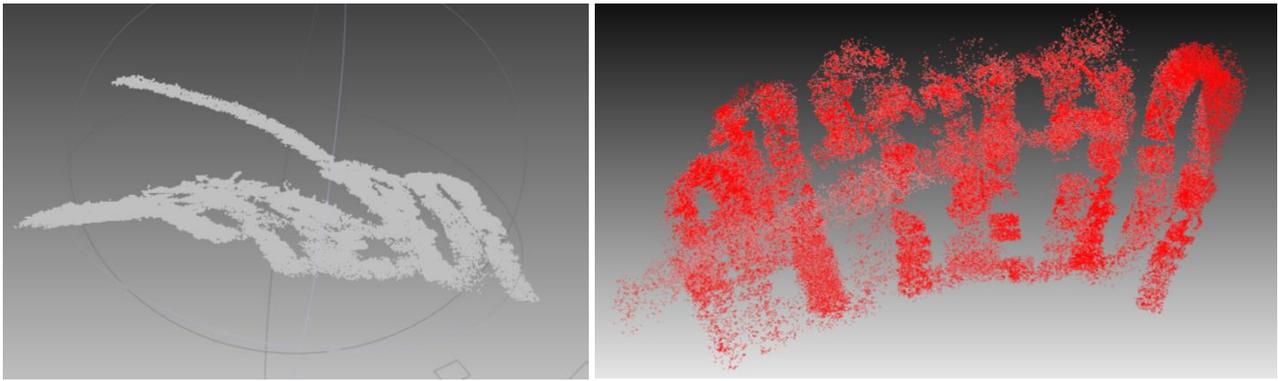


Figure V-13 Sparse map estimated by Stereo-Tracker code with parameters adapted to maximize feature matched.

Bad results obtained relying only on the optical data highlighted the need of an alternative way, justifying the use of the EKF.

Some tests have been preliminary carried out feeding only navigation sensor measurements to the EKF. Results are visible in Figure V-14: on the left, the estimated path (in red) is directly juxtaposed with navigation data provided by AUV internal EKF (in blue) and with the bathymetry of the shipwreck; on the right, the trajectory recoverable from the giga-mosaic is highlighted for indirect comparison.

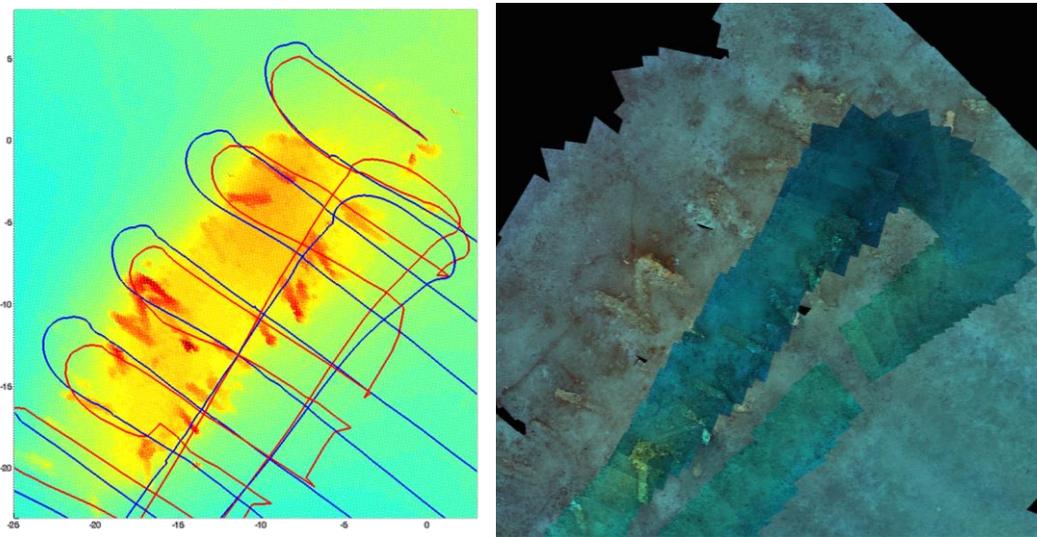


Figure V-14 Left: Bathymetry of the area with sensor tracking elaborated by EKF in MATLAB (in red) and by the AUV internal navigation EKF. Right: giga-mosaic of the area with pictures from longitudinal transect highlighted.

Simulating an on-line application, the filter loads each new stereo-pair and evaluates map and relative pose. Estimated pose is used as input for the EKF. Stereo system is the slowest among



The following step has been the ASKF test in order to evaluate loop closure performances, visible in Figure V-16. Blue frames show where a loop is closed, whereas red frames are subsequent non-overlapping images.

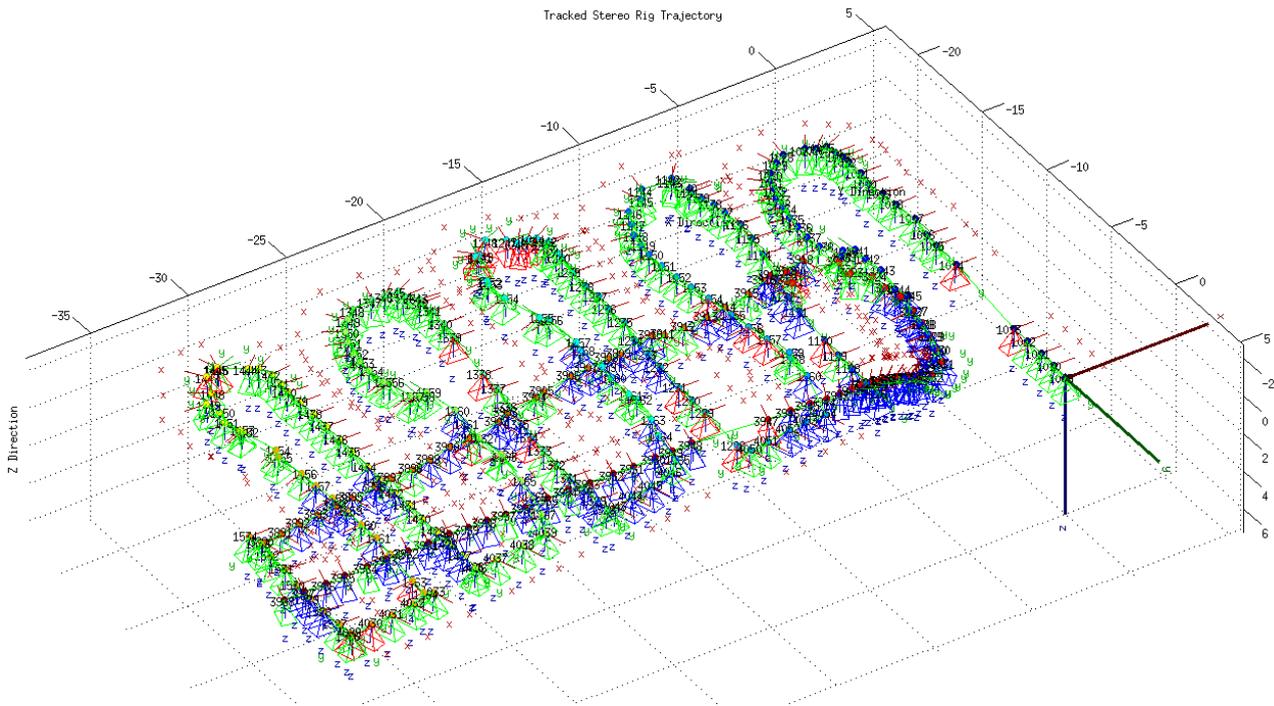


Figure V-16 Trajectory estimated by ASKF for the left camera: blue frames represent loop closures; red frames are non-overlapping subsequent images.

To better illustrate advantages of loop closure we put in evidence a particular part of the trajectory. The area highlighted in Figure V-17 evidences a particular “virtual link”. The second and the third vertical swaths are virtually linked thanks to images from the turn of horizontal one.

We avoided, when possible, to use the same information more than once, but, due to lack of images in that area, we decided to use twice that particular series of images: as said, once to connect vertical transects, and once in its original temporal sequence, at the turn of the transversal swath. Thus, being exactly the same pictures, the corrected estimation of relative pose must be the same both in the first and in the second evaluation.

It clearly appears how in the simple EKF this condition is not respected, whereas the constraints added in the ASKF after the loop closure detection allow to correct the trajectory, so recovering drift error. Figure V-16 shows that a lot of loop are closed, adding constraints to the trajectory that remains smooth because of intrinsic correction capabilities of ASKF. It is crucial to correct even the smallest error to avoid that a little orientation inaccuracy become a big displacement error after some meters.

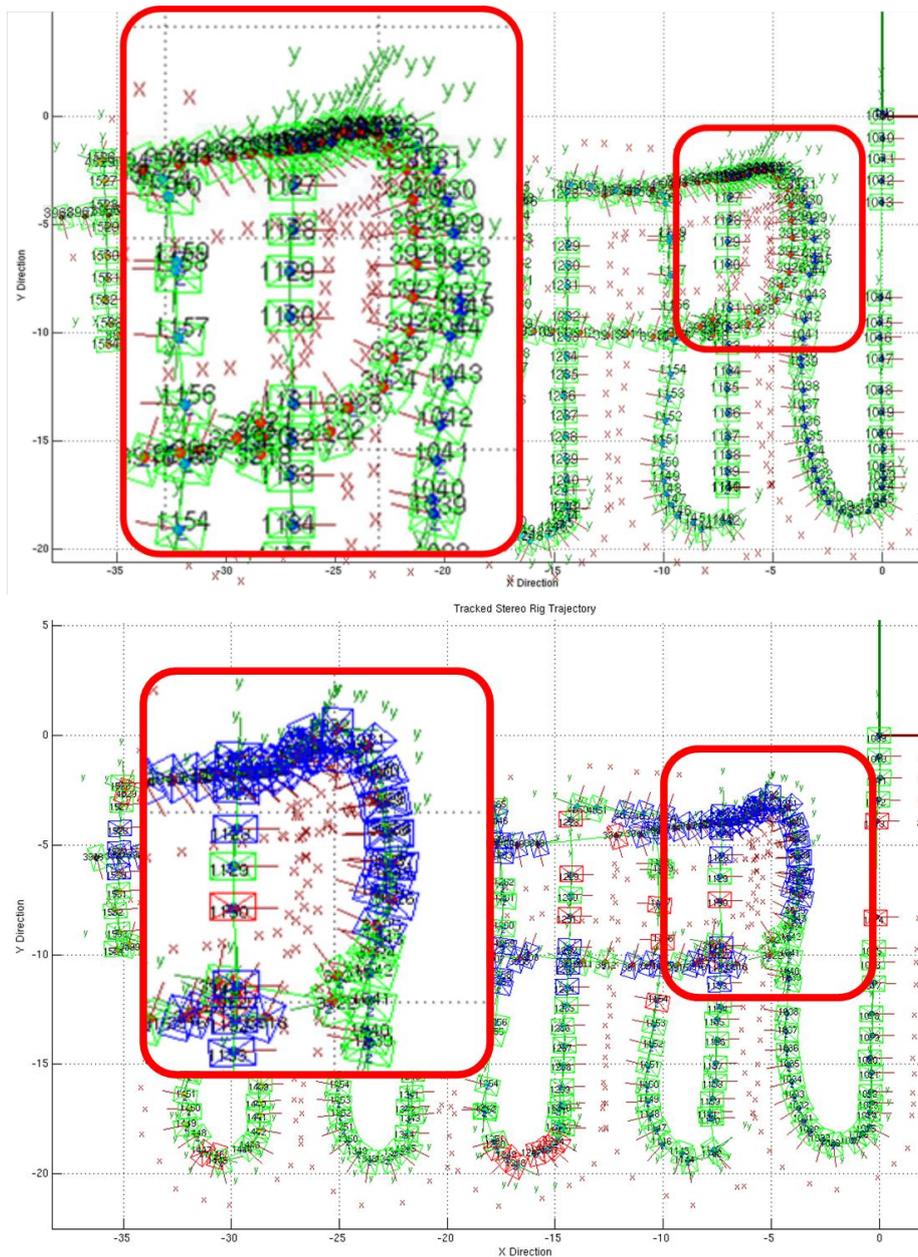


Figure V-17 Loop closure detail

Unfortunately, ASKF has an important drawback related to time performances. In effect, the size of the state vector and of the model covariance increase at each augmentation and with it also the process time per image grows. This is a noticeable limit for dataset with more than 100 images. The big problem is given by the inversion of a matrix needed for Kalman gain calculation. That matrix can be represented as a sparse matrix to save resources, but the best solution could be the C++ implementation.

### V.7.2 *Cap de Vol Dataset*

*Cap de Vol* dataset refers to a roman shipwreck lying at about 20 metres deep off the coasts of Puerto de la Selva (Girona, Spain). This dataset is less problematic than *La Lune* one, thanks to good overlapping between subsequent images and low depth that ease navigation.

Firstly, a 3D model of the site has been reconstructed in order to have an idea of surveyed area and of covered path.

*Cap de Vol* dataset has been fed to the MATLAB® algorithm and tested also with the C++ version. Results are shown in the following subsection.

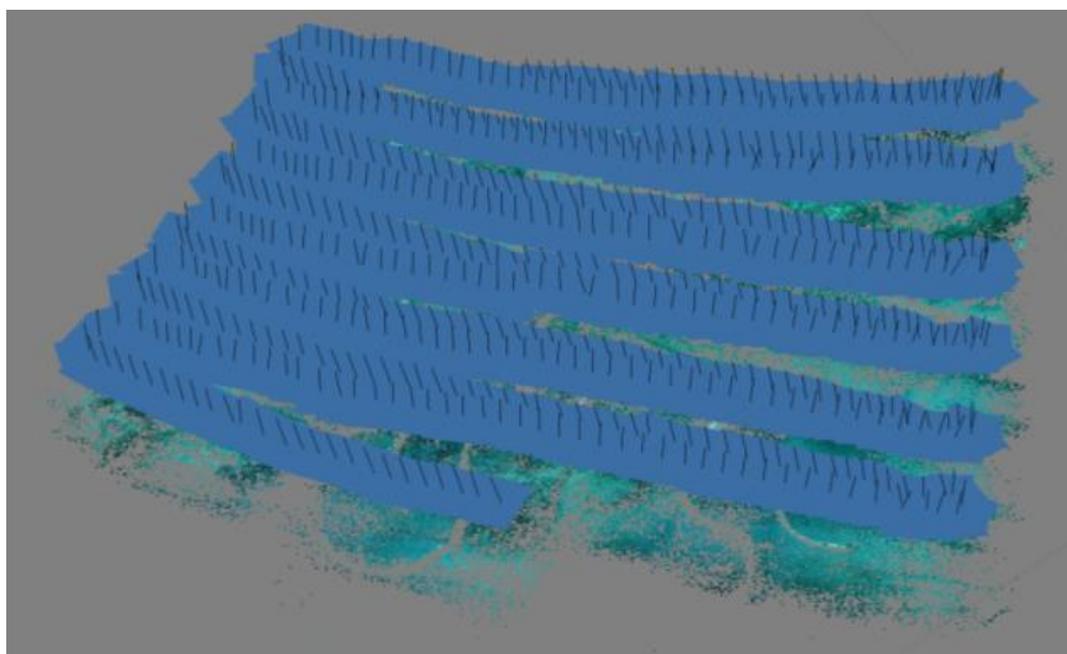
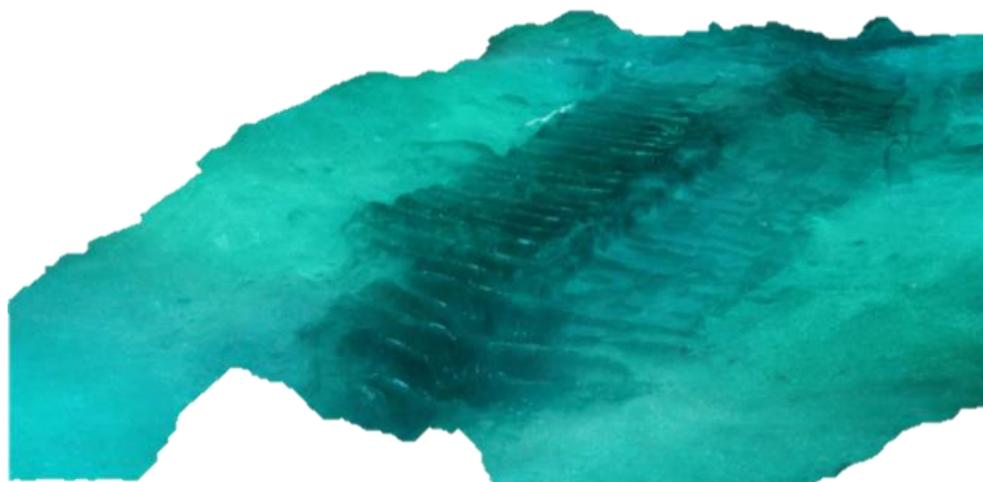


Figure V-18 *Cap de Vol* site in Agisoft Photoscan®: 3D model(above) and camera poses resulting from image registration (bottom).

## V.8 C++ Implementation

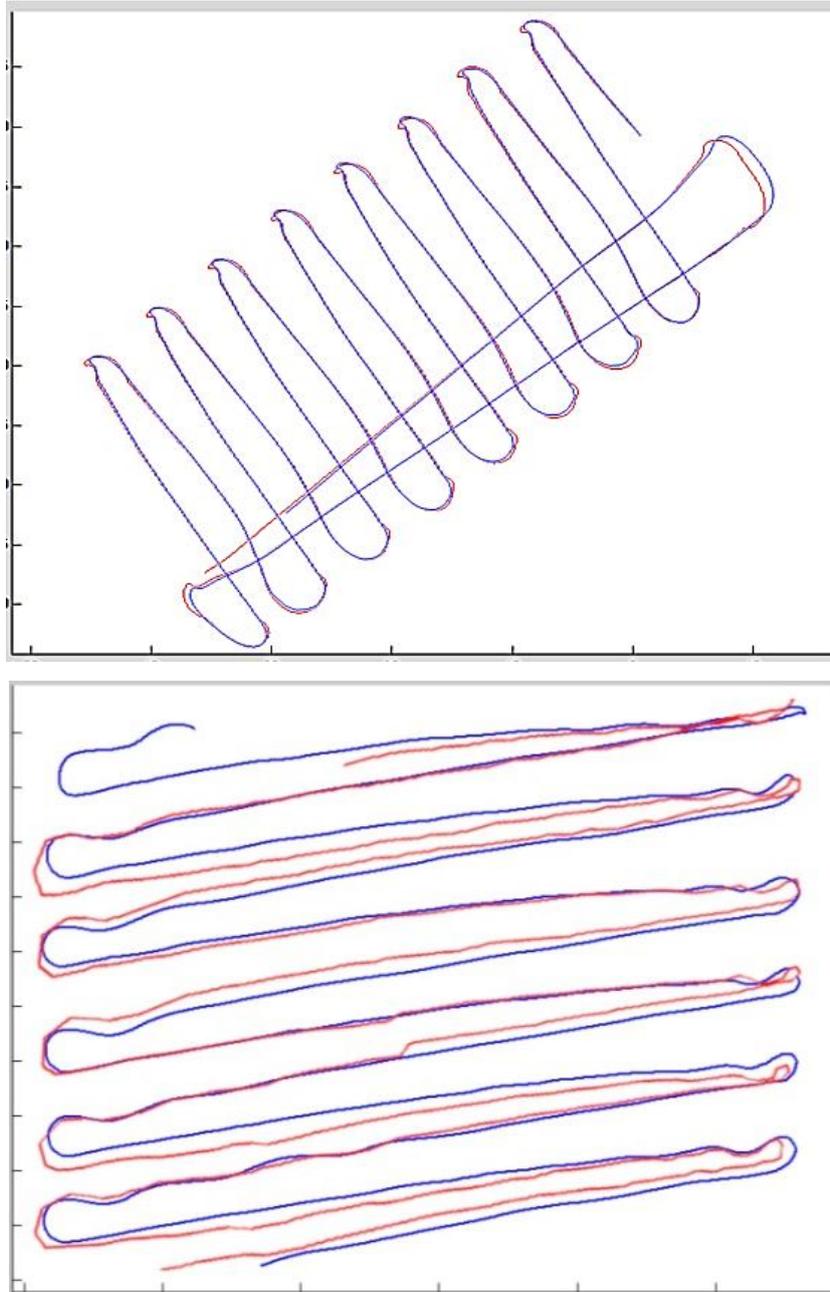
In order to increase both portability and performances, the whole code has been gradually implemented in C++. The testing of some available implementations was not satisfying at all, mostly due to lack of flexibility, which is an essential fulfilment of required Kalman Filter.

After the relatively simple development of propagation and updating phases, the definition of specific observation models for the sensors required the implementation of very complex matrices in order to provide the required flexibility. This has been achieved through parsing in C++ language of symbolic matrices already evaluated in MATLAB®.

The idea, already pursued in MATLAB® implementation, is to find symbolic representations of various matrices just once and then use the evaluated expressions after the variable numeric definition. In particular, after defining symbolic variables and writing symbolic equation system relative to each observation function  $h$ , the associated Jacobian matrix is evaluated through symbolic calculation. The obtained values can be copied-pasted into another MATLAB® script (or function) in which the independent variables keep the same name but are defined as numerical variables, changing their own values at each iteration. Consequently, the independent variables, precisely the predicted measurements and the observation matrix elements, are evaluated through the found relations, and used by the EKF (or ASKF).

Unfortunately, this step is not so direct in C++. Tested libraries for symbolic calculation usually does not provide an output directly “paste-ble” as input expression; hence we decided to convert MATLAB® expressions, mainly because of possibilities to directly test them after the evaluation and before the parsing operation.

*Eigen* library is used to ease coding and enhance performances. The developed C++ code is yet under testing and optimisation, above all for what concerns optical data, but first encouraging results are already available. Figure V-19 represents comparison between trajectory estimated by C++ and other trajectories differently estimated. In particular, Figure V-19 (left) compares *La Lune* trajectory estimated by C++ code (in blue) and by MATLAB® EKF (in red). Figure V-19 (right) is referred to *Cap De Vol* dataset and shows the C++ estimated trajectory (in blue) and the camera trajectory resulting from Agisoft Photoscan® bundle adjustment (in red).



*Figure V-19 Top: La Lune trajectory estimated by C++ code (in blue) and by MATLAB® EKF (in red) based only on navigation sensors. Bottom: Cap De Vol trajectory estimated by C++ (in blue) and by Agisoft Photoscan® bundle adjustment (in red).*

## ***VI CONCLUSIONS AND FUTURE WORK***

---

This final chapter outlines the work presented throughout the thesis. After a summary of treated topics, it points out the main contributions of the work, the main unsolved issues, and possible further developments. Finally, it lists the publications related with the carried out research.

### ***VI.1 Thesis Summary***

The research topics discussed in the present work are mainly referred to the 3D reconstruction issue. We dealt with 3D modelling in underwater environment. A main objective of the work has been the testing of 3D dense mapping of underwater artefacts and the related possible applications oriented towards Underwater Archaeology realm. Another aim pursued regards the employment of underwater vehicles in Underwater Archaeology. We investigated the possibility of combining information from classical navigation instruments and optical sensor, by means of Augmented State Kalman Filter, in order to mutual enhance 3D mapping performances and vehicle navigation.

Chapter II discusses about related works on underwater archaeology and about computer vision aspects essential to the understanding of the thesis. Moreover, an overview of main underwater technology is given.

In Chapter III, we show two studies performed on the *CoMAS* test-bed of *Baia* that illustrate the adopted pipeline for 3D dense stereo-mapping of challenging underwater scenes, with relative encountered problems and tested solution. We also present how a detailed virtual model of reconstructed area can support many purposes of underwater researchers, such as exploitation and dissemination, studying and cataloguing of findings, planning and documentation of intervention operations.

After a discussion about Kalman Filter theory and its main variations in Chapter IV, the implemented EKF-ASKF version is illustrated in Chapter V, which explains the particularization of the general form for the specific aim. The implemented algorithm and the resulting test, aiming the feasibility evaluation of interactive 3D mapping with UUV, are analysed in detail, showing the positive feedback of test performed on off-line datasets.

## **VI.2 Contribution**

This thesis made some contributions in 3D vision for underwater environment application. In particular, the carried out work accomplished the proposed goals defined according to the research context urgencies:

*Experimentation of 3D reconstruction in complex underwater conditions for documentation purposes.*

Among the most popular 3D reconstruction techniques, we tested Multi-View mapping to reconstruct some areas from the challenging context of Baia Archaeological Park. A certain care have been required during image acquisition and processing phases in order to optimise the results. The implemented pipeline allows to reconstruct large areas even in the hostile condition of shallow water lightning, turbidity and biological colonisation. The final 3D model is textured and high detailed and has been used for different purposes. This high accuracy achieved represents the most remarkable contributions for underwater application and archaeology in particular. In effect, apart from visualisation and virtual exploration, it has been showed how an accurate 3D model can be exploited for operation planning, accurate device simulation and design, and for monitoring and documentation of helpful operation results and parameters.

*Feasibility test of 3D mapping and navigation with UUV for ROV guidance support during interventions*

We tested the feasibility of an architecture for underwater visual SLAM with six degrees-of-freedom pose estimation. The implemented EKF-ASKF algorithm exploits typical DR navigation from sensors together with calibrated stereo-rig data, improving 6DOF tracking, and concurrently reconstructing an incremental 3D map of the explored scenario. The filter gives the better estimation of the vehicle (and camera) poses using data coming from DVL, AHRS, depth gauge, and optical tracking. This pose estimation is used, at each step, to enhance the registration of the subsequent point clouds and incrementally reconstruct a 3D sparse map of the surveyed area. Storing all previous position and constraints, the ASKF allows to further enhance the trajectory tracking, thanks to Visual-Based implemented algorithm. Exploiting current position and feature-matching, it is possible to look for already surveyed area and, in case of loop closure detection, to

correct the whole trajectory surveyed in between, reducing the drift error, main drawback of dead-reckoning navigation.

### **VI.3 Future work**

The presented work highlighted new issues to solve and new possible research lines.

The main problem that affects 3D reconstruction is related to performances in terms of time.

Even if satisfying for on-line mapping, the SLAM application requires a significantly higher frame-rate to be considered as a valid visualisation tool for remote operation. This can be achieved completing the porting in C++ and optimising the implemented algorithms, above all the one related to feature detection and matching. Moreover, the employment of ASKF seems to raise serious problems for real-time capabilities. Optimisation should be performed according to matrix sparseness, statistic managing of stored paths, and code parallelisation possibilities.

For what concerns 3D dense mapping, information on camera poses coming from SLAM could be exploited to speed-up the reconstruction problem overlooking the image registration part.

The techniques for image enhancement and colour correction can further be refined and optimised through on-site pre-calibration operations. Furthermore, even if accuracy test have been performed thanks to calibrated bars on scene, it would be useful to test the encouraging 3D reconstructions with *ad-hoc* specimen and compare it with other measurement and reconstruction methodologies employed in underwater environment. This will better motivate the decision to extensively use this process in the archaeological experimentation and intervention operations.

### **VI.4 Related Publications**

- *F. Bruno, A. Gallo, M. Muzzupappa, L. Barbieri, F. De Filippo, A. Angilica, R. Petriaggi, B. Davide, M. Ciabattini, R. Mancinelli, F. Lucci, G. G. de Ayala e P. Caputo, "New technologies and tools for the documentation and restoration of submerged archaeological sites", in ISUR 8 - 8th INTERNATIONAL SYMPOSIUM ON UNDERWATER RESEARCH, Procida (NA) - Italy, 26-29 March 2014.*

- *F. Bruno, A. Gallo, F. De Filippo, M. Muzzupappa, B. Davide Petriaggi e P. Caputo, "3D documentation and monitoring of the experimental cleaning operations in the underwater archaeological site of Baia (Italy)", in Digital Heritage, International Congress 2013, Marseille, France, 28 Oct - 01 Nov 2013.*
- *A. Tonazzini, E. Salerno, V. Palleschi, G. Bianco e F. De Filippo, "Extracting Information from Multimodal Images of Documents and Artworks", in 6th International Congress "Science and Technology for the Safeguard of Cultural Heritage in the Mediterranean Basin", Athens, Greece, 22 – 25 October 2013.*
- *F. De Filippo, N. Gracias, R. Garcia, J. Ferrer e F. Bruno, "Incremental underwater mapping in 6DOF with stereo tracking", in MARTECH workshop 2013, Girona, Spain, 9 - 11 October 2013.*
- *N. Gracias, P. Ridaó, R. Garcia, J. Escartin, M. L'Hour, F. Cibecchini, R. Campos, M. Carreras, D. Ribas, N. Palomeras, L. Magi, A. Palomer, T. Nicosevici, R. Prados, R. Hegedus, L. Neumann, F. De Filippo e A. Mallios, "Mapping the Moon: Using a lightweight AUV to survey the site of the 17th Century ship 'La Lune'", in Proceedings of the Oceans MTS/IEEE OCEANS conference 2013, Bergen, Norway, June 2013.*
- *A. Gallo, A. Angilica, G. Bianco, F. De Filippo, M. Muzzupappa, B. Davide e F. Bruno, "3D Reconstruction and Virtual Exploration of Submerged Structures: a Case Study in the Underwater Archaeological Site of Baia (Italy)", in VAST: International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage, Brighton, UK, 19-21 November 2012.*

---

**REFERENCES**

---

- [1] T. L. Evans and P. T. Daly, *Digital Archaeology: Bridging Method and Theory*, T. L. Evans and P. T. Daly, Eds., London: Routledge, 2006.
- [2] F. Stanco, S. Battiato and G. Gallo, *Digital Imaging for Cultural Heritage Preservation Analysis, Restoration, and Reconstruction of Ancient Artworks*, CRC Press, July 28th 2011.
- [3] P. Reilly, "Towards a virtual archaeology.", in *Computer Applications and Quantitative Methods in Archaeology 1990 (BAR International Series 565)*, S. Rahtz and K. Lockyear, Eds., Oxford, CAA90, 1991, pp. 132-139.
- [4] UNESCO, 2001. [Online]. Available: <http://www.unesco.org/new/en/culture/themes/underwater-cultural-heritage/>.
- [5] P. Holt, "An assessment of quality in underwater archaeological surveys using tape measurements.", *International Journal of Nautical Archaeology*, vol. 32, no. 2, pp. 246-251, October 2003.
- [6] B. Bingham, B. Foley, H. Singh, R. Camilli, K. Delaporta, R. Eustice, A. Mallios, D. Mindell, C. Roman and D. Sakellariou, "Robotic tools for deep water archaeology: Surveying an ancient shipwreck with an autonomous underwater vehicle.", *Journal of Field Robotics*, vol. 27, no. 6, pp. 702-717, 2010.
- [7] M. N. P. Bosse, J. Leonard, M. Soika, W. Feiten and S. Teller, "An Atlas framework for scalable mapping.", in *Robotics and Automation, 2003. Proceedings. ICRA '03. IEEE International Conference on*, 14-19 Sept. 2003.
- [8] S. Thrun, Y. Liu, D. Koller, A. Y. Ng, Z. Ghahramani and H. Durrant-Whyte, "Simultaneous Localization and Mapping with Sparse Extended Information Filters.", *The International Journal of Robotics Research*, vol. 23, pp. 693-716, August 2004.
- [9] S. Thrun, M. Montemerlo, D. Koller, B. Wegbreit, J. Nieto and E. Nebot, "FastSLAM: An efficient solution to the simultaneous localization and mapping problem with unknown data association.", *Journal of Machine Learning Research*, 2004.
- [10] H. Singh, J. Howland and O. Pizarro, "Advances in large-area photomosaicking underwater.", *Oceanic Engineering, IEEE Journal of*, vol. 29, no. 3, pp. 872-886, July 2004.
- [11] D. Ribas, P. Ridao and J. Neira, "Underwater SLAM for structured environments using an imaging sonar.", 2010.
- [12] A. Elibol, N. Gracias and R. Garcia, "Augmented State-Extended Kalman filter combined framework for topology estimation in large area underwater mapping.", *Journal of Field Robotics*, vol. 27, no. 5, pp. 656-674, 2010.
- [13] R. Campos, J. Ferrer, M. Villanueva, L. Magí and R. Garcia, "Trinocular System for 3D motion and Dense Structure Estimation.", *Instrumentation Viewpoint*, vol. 45, 2009.
- [14] G. Bianco, A. Gallo, F. Bruno and M. Muzzupappa, "A comparison between active and passive techniques for underwater 3D applications.", in *3D-ARCH 2011*, Trento, Italy,, 2-4 March 2011.
- [15] D. C. Brown, "Close-range camera calibration.", *Photogrammetric Engineering*, vol. 37, no. 8, pp. 855-866, 1971.

- 
- [16] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses.", in *Radiometry*, USA, Jones and Bartlett Publishers, Inc., 1992, pp. 221-244 .
- [17] Z. Zhang, "Flexible Camera Calibration by Viewing a Plane from Unknown Orientations.", in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, 20 Sep 1999-27 Sep 1999.
- [18] J.-Y. Bouguet. [Online]. Available: <http://www.vision.caltech.edu/bouguetj/>. [Accessed 29 November 2014].
- [19] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge, UK: Cambridge University Press, 2004, Second Edition.
- [20] T. Svoboda, D. Martinec and T. Pajdla, "A convenient multicamera self-calibration for virtual environments.", *Presence: Teleoperators and Virtual Environments*, vol. 14, no. 4, pp. 407-422 , August 2005.
- [21] M. Pollefeys, L. Van Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops and R. Koch, "Visual modeling with a hand-held camera.", *International Journal of Computer Vision*, vol. 59, no. 3, pp. 207-232, 2004.
- [22] M. Brown and D. Lowe, "Unsupervised 3D object recognition and reconstruction in unordered datasets.", in *3-D Digital Imaging and Modeling, 2005. 3DIM 2005. Fifth International Conference on*, 3-16 June 2005.
- [23] N. Snavely, S. M. Seitz and R. Szeliski, "Photo tourism: Exploring photo collections in 3D.", *SIGGRAPH Conference Proceedings*, pp. 835-846, 2006.
- [24] R. Kumar, H. Sawhney, S. Samarasekera, S. Hsu, H. Tao, Y. Guo, K. Hanna, A. Pope, R. Wildes, D. Hirvonen, M. Hansen and P. Burt, "Aerial video surveillance and exploitation.", *Proceedings of the IEEE*, vol. 89, no. 10, pp. 1518-1539, October 2001.
- [25] Y. Li and R. Verma, "Multichannel Image Registration by Feature-Based Information Fusion.", *Medical Imaging, IEEE Transactions on* , vol. 30, no. 3, pp. 707 - 720, November 2010.
- [26] Z. Zhang and R. Blum, "A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application.", *Proceedings of the IEEE* , vol. 87, no. 8, pp. 1315 - 1326, August 1999.
- [27] Y. Na, M. Ehlers, W. Yang and Y. Li, "Imaging mechanism based remote sensing image fusion with multi-wavelet transform.", *Signal Processing, 2004. Proceedings. ICSP '04. 7th International Conference on* , vol. 3, pp. 2190 - 2193 , 31 Aug.-4 Sept 2004 .
- [28] I. Villalon-Turrubiates and M. Llovera-Torres, "Multispectral classification of remote sensing imagery for archaeological land use analysis: Prospective study.", in *Geoscience and Remote Sensing Symposium (IGARSS), IEEE International*, Honolulu, HI, 25-30 July 2010.
- [29] P. Weibring, H. Edner and S. Svanberg, "Versatile mobile lidar system for environmental monitoring.", *Applied Optics*, vol. 42, no. 18, pp. 3583-3594 , 2003.
- [30] R. Prados, R. Garcia, N. Gracias, J. Escartin and L. Neumann, "A Novel Blending Technique for Underwater Gigamosaicing.", *Oceanic Engineering, IEEE Journal of*, vol. 37, no. 4, pp. 626-644, July 2012.
- [31] S. C. Park, M. K. Park and M. G. Kang, "Super-resolution image reconstruction: a technical overview.",

- Signal Processing Magazine, IEEE*, vol. 20, no. 3, pp. 21-36, May 2003.
- [32] J. Liu, B. Vemuri and J. Marroquin, "Local frequency representations for robust multimodal image registration.", *Medical Imaging, IEEE Transactions on*, vol. 21, no. 5, pp. 462-469, May 2002.
- [33] B. Zitová and J. Flusser, "Image registration methods: a survey.", *Image and Vision Computing*, vol. 21, no. 11, pp. 977-1000, October 2003.
- [34] G. Blais and M. Levine, "Registering multiview range data to create 3D computer objects.", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 17, no. 8, pp. 820 - 824, August 1995.
- [35] X. Dai and S. Khorram, "The effects of image misregistration on the accuracy of remotely sensed change detection.", *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 36, no. 5, pp. 1566 - 1577, August 2002.
- [36] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal and P. Suetens, "Multimodality image registration by maximization of mutual information.", *Medical Imaging, IEEE Transactions on*, vol. 16, no. 2, pp. 187 - 198, April 1997.
- [37] L. G. Brown, "A survey of image registration techniques.", *ACM Computing Surveys (CSUR)*, vol. 24, no. 4, pp. 325-376, December 1992.
- [38] M. Kudzinava, R. Garcia and J. Marti, "Feature-based Matching of Underwater Images.", *Instrumentation Viewpoint Journal*, 2007.
- [39] C. Harris and M. Stephens, "A Combined Corner and Edge Detector.", *Proceedings of the 4th Alvey Vision Conference*, vol. 15, pp. 147-151, 1988.
- [40] P. Beaudet, "Rotationally invariant image operators.", *4th Int. Joint Conference on Pattern Recognition*, pp. 579-583, 1978.
- [41] H. Bay, A. Ess, T. Tuytelaars and L. Van Gool, "SURF: Speeded Up Robust Features.", *Computer Vision and Image Understanding (CVIU)*, vol. 110, no. 3, pp. 346-359, 2008.
- [42] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir and L. Van Gool, "A Comparison of Affine Region Detectors.", *International Journal of Computer Vision*, vol. 65, no. 1-2, pp. 43-72, November 2005.
- [43] D. Lowe, "Object recognition from local scale-invariant features.", in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, Kerkyra, Corfu, Greece, 1999.
- [44] J. Matas, O. Chum, M. Urban and T. Pajdla, "Robust Wide Baseline Stereo from Maximally Stable Extremal Regions.", *Proceedings of the British Machine Conference*, pp. 36.1-36.10, September 2002.
- [45] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors.", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 10, pp. 1615 - 1630, October 2005.
- [46] M. Irani and P. Anandan, "Robust multi-sensor image alignment.", *Proceedings of the International Conference on Computer Vision*, pp. 959-966, January 1996.
- [47] R. Kumar, P. Anandan, M. Irani, J. Bergen and K. Hanna, "Representation of scenes from collections of images.", *Representation of Visual Scenes, 1995. (In Conjunction with ICCV'95), Proceedings IEEE Workshop on*, pp. 10 - 17, 24 June 1995.
- [48] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, San Mateo, CA: Morgan Kaufmann Publishers, , 1988.

- 
- [49] H. Sawhney, S. Ayer and M. Gorkani, "Model-based 2D & 3D dominant motion estimation for mosaicing and video representation.", in *Proceedings of the International Conference on Computer Vision*, Cambridge, MA, 1995.
- [50] R. Szeliski, "Image mosaicing for tele-reality applications.", May 1994.
- [51] S. Kang, "A survey of image-based rendering techniques.", August 1997.
- [52] A. Can, C. Stewart, B. Roysam and H. Tanenbaum, "A feature-based technique for joint, linear estimation of high-order image-to-mosaic transformations: Application to mosaicing the curved human retina.", in *In Proceedings of the IEEE Conference on Computer Vision*, Hilton Head Island, SC, USA, June 2000.
- [53] O. Pizarro and H. Singh, "Towards large area mosaicing for underwater scientific applications.", *IEEE Journal of Oceanic Engineering*, vol. 28, no. 4, pp. 651-672, October 2003.
- [54] H. Sawhney, S. Hsu and R. Kumar, "Robust video mosaicing through topology inference and local to global alignment.", *Proceedings of the European Conference on Computer Vision — ECCV'98*, vol. 2, pp. 103-119, 1998.
- [55] H. Sawhney, R. Kumar, G. Gendel, J. Bergen, D. Dixon and V. Paragano, "Videobrush: Experiences with consumer video mosaicing.", in *Proceedings of the Fourth IEEE Workshop on Applications of Computer Vision*, Princeton, NJ, USA, October 1998, pp. 56-62.
- [56] F. Chen, G. M. Brown and M. Song, "Overview of threedimensional shape measurement using optical methods.", *Optical Engineering*, vol. 39, no. 1, pp. 10-22, 2000.
- [57] F. Blais, "A review of 20 years of range sensor development.", *Journal of Electronic Imaging*, vol. 13, no. 1, pp. 231-243, 2004.
- [58] S. G. Narasimhan and K. N. Shree, "Structured Light Methods for Underwater Imaging: Light Stripe Scanning and Photometric Stereo.", in *Proceedings of 2005 MTS/IEEE OCEANS*, Washington, September, 2005.
- [59] J. Salvi, S. Fernandez, T. Pribanic and X. Lladó, "A state of the art in structured light patterns for surface profilometry.", *Pattern Recognition*, no. 43, pp. 2666 - 2680, 2010.
- [60] F. Remondino and S. El-Hakim, "Image-based 3D modelling: A review.", *Photogrammetric Record*, vol. 21, no. 115, p. 269–291, 2006.
- [61] M. Muzzupappa, F. Bruno, S. Barone, A. Razionale and G. Bianco, "Experimentation of structured light and stereo vision for underwater 3D reconstruction.", *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, pp. 508-518, 2011.
- [62] A. Gallo, G. Bianco, F. Bruno and M. Muzzupappa, "Sperimentazione di tecniche fotogrammetriche per il rilievo 3D di reperti archeologici sommersi.", in *AIAS 40° convegno nazionale*, Palermo, 7-10 settembre 2011.
- [63] A. Piaia Silvatti, F. A. Salve Dias, P. Cerveri and R. M. L. Barr, "Comparison of different camera calibration approaches for underwater applications.", *Journal of Biomechanics*, vol. 45, no. 6, pp. 1112 - 1116, 2012.
- [64] R. Fontana, M. C. Gambino, G. Gianfrate, M. Greco, E. Pampaloni and L. Pezzati, "A 3D scanning device for architectural survey based on time-of-flight technology.", *Proc. SPIE 5457, Optical Metrology in Production Engineering*, vol. 5457, pp. 393-400, 10 September 2004.

- 
- [65] R. Fontana, M. Greco, M. Materazzi, E. Pampaloni, L. Pezzati, C. Rocchini and R. Scopigno, "Three-dimensional modelling of statues: the Minerva of Arezzo.", *Journal of Cultural Heritage*, vol. 3, no. 4, p. 325–331, 2002.
- [66] a. Brémand, P. Doumalin, J.-C. Dupré, F. Hesser and V. Valle, "Optical Techniques for Relief Study of Mona Lisas Wooden Support.", in *Proceedings of the XIth International Congress and Exposition*, Orlando, Florida, USA, June 2-5, 2008.
- [67] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison and al, "KinectFusion: Real-Time Dense Surface Mapping and Tracking.", *IEEE ISMAR 2011*, 2011.
- [68] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe and al, "KinectFusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera.", in *UIST '11*, Santa Barbara, CA, USA., October 16-19, 2011.
- [69] J. P. J. Salvi and J. Batlle, "Pattern codification strategies in structured light systems.", *Pattern Recognition*, vol. 37, no. 4, pp. 827-849, 2004.
- [70] G. Birch, J. S. Tyo and J. Schwiegerling, "3D imaging made possible through astigmatic projected pattern.", *Optics Express*, vol. 20, no. 6561-6574, p. 65, 2012.
- [71] J. Xu, N. Xi, C. Zhang, Q. Shi and J. Gregory, "Real-time 3D shape inspection system of automotive parts based on structured light pattern.", *Optics & Laser Technology*, vol. 43, no. 1, pp. 1-8, February 2011.
- [72] T. Pribanić, S. Mrvoš and J. Salvi, "Efficient multiple phase shift patterns for dense 3D acquisition in structured light scanning.", *Image and Vision Computing*, vol. 28, no. 8, pp. 1255-1266, August 2010.
- [73] J. R. Kender, *Shape from texture*, Pittsburgh, PA, USA: Carnegie Mellon University , 1981.
- [74] D. Hirano, Y. Funayama and T. Maekawa, "3D Shape Reconstruction from 2D Images.", *Computer-Aided Design and Applications*, vol. 6, no. 5, pp. 701-710, 2009.
- [75] B. K. P. Horn and M. J. Brooks, *Shape from shading*, Cambridge, MA, USA : MIT Press , 1989.
- [76] S. Nayar and Y. Nakagawa, "Shape from focus.", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 16, no. 8, pp. 824 - 831, August 1994.
- [77] A. Criminisi, I. Reid and A. Zisserman, "Single View Metrology.", *International Journal of Computer Vision*, vol. 40, no. 2, pp. 123-148, 2000.
- [78] D. Hoiem, A. ., E. Efros and M. Hebert, "Automatic photo pop-up.", *ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH 2005*, vol. 54, no. 3, pp. 577-584, July 2005.
- [79] A. Saxena, S. H. Chung and A. Y. Ng, "3-D Depth Reconstruction from a Single Still Image.", *International Journal of Computer Vision*, vol. 76, no. 1, pp. 53-69, 2008.
- [80] O. Faugeras and R. Keriven, "Complete dense stereovision using level set methods.", *Computer Vision — ECCV'98*, vol. 1406, pp. 379-393, 1998.
- [81] Y. Boykov and V. Kolmogorov, "Computing geodesics and minimal surfaces via graph cuts.", *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, vol. 1, pp. 26-33, 13-16 October 2003.
- [82] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max- flow algorithms for energy minimization in vision.", *Pattern Analysis and Machine Intelligence, IEEE Transactions on* , vol.

- 26, no. 9, pp. 1124-1137, September 2004.
- [83] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms.", *International Journal of Computer Vision*, vol. 47, no. 1-3, pp. 7-42, April 2002.
- [84] L. Bartolini, L. De Dominicis, M. Ferri de Collibus, G. Fornetti, M. Guarneri, E. Paglia, C. Poggi and R. Ricci, "Underwater three-dimensional imaging with an amplitude-modulated laser radar at a 405 nm wavelength.", *Applied Optics*, vol. 44, no. 33, pp. 7130-7135, 2005.
- [85] Y. Schechner and N. Karpel, "Clear underwater vision.", in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition.*, 2004.
- [86] E. M. Mikhail, J. S. Bethel and J. C. McGlone, "Introduction to modern photogrammetry. John Wiley and Sons, 2001. 16.", *John Wiley and Sons*, vol. 1, 2001.
- [87] Y. Ma, S. Soatto, J. Kosecka and S. Sastry., *An invitation to 3d vision: From images to geometric models.*, Springer Verlag, 2003.
- [88] P. Firoozfam and S. Negahdaripour, "A multi-camera conical imaging system for robust 3D motion estimation, positioning and mapping from UAVs.", *Advanced Video and Signal Based Surveillance, 2003. Proceedings. IEEE Conference on*, pp. 99-106, 21-22 July 2003.
- [89] S. Negahdaripour, H. Zhang, P. Firoozfam and J. Oles, "Utilizing panoramic views for visually guided tasks in underwater robotics applications.", *OCEANS, 2001. MTS/IEEE Conference and Exhibition*, vol. 4, pp. 2593-2600, 5-8 November 2001.
- [90] Menci, "mencisoftware.", [Online]. Available: <http://www.menci.com/>. [Accessed 31 November 2014].
- [91] H. Zhang and S. Negahdaripour, "On reconstruction of 3D volumetric models of reefs and benthic structures from image sequences of a stereo-rig.", *In Oceans Conference Record (IEEE)*, vol. 5, pp. 2553-2559, 2003.
- [92] C. Fraser, "Digital camera self-calibration.", *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 52, no. 4, pp. 149-159, August 1997.
- [93] P. Sturm and B. Triggs, "A factorization based algorithm for multi-image projective structure and motion.", *Computer Vision — ECCV '96 — Lecture Notes in Computer Science*, vol. 1065, pp. 709-720, 1996.
- [94] B. Triggs, "Factorization methods for projective structure and motion.", in *Computer Vision and Pattern Recognition, 1996. Proceedings CVPR '96, 1996 IEEE Computer Society Conference on*, San Francisco, CA, 18-20 June 1996.
- [95] B. Triggs, P. F. McLauchlan, R. I. Hartley and A. W. Fitzgibbon, "Bundle adjustment—a modern synthesis.", *Vision Algorithms: Theory and Practice*, vol. 1883, pp. 298-372, 2000.
- [96] S. Seitz, B. Curless, J. Diebel, D. Scharstein and R. Szeliski, "A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms.", *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 1, pp. 519-528, 17-22 June 2006.
- [97] J. S. Yedidia, W. T. Freeman and Y. Weiss, "Generalized Belief Propagation.", *Neural Information Processing Systems (NIPS)*, vol. 13, pp. 689-695, June 2000.
- [98] F. Tombari, S. Mattoccia, L. D. Stefano and E. Addimanda, "Classification and evaluation of cost

- aggregation methods for stereo correspondence.”, *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2008), Anchorage, AK*, pp. 1-8, 23-28 June 2008.
- [99] T. Kanade and M. Okutomi, “A stereo matching algorithm with an adaptive window: Theory and experiment.”, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 16, no. 9, p. 920–932, September 1994.
- [100] T. Kanade, “Development of a video-rate stereo machine. In Image Understanding Workshop.”, *Image Understanding Workshop, In Proceedings of*, p. 549–557, 1994.
- [101] M. J. Hannah, Computer Matching of Areas in Stereo Images, Ph.D. thesis, Stanford University. , (1974).
- [102] R. C. Bolles, H. H. Baker and M. J. Hannah, “ The JISCT stereo evaluation.”, *Image Understanding Workshop*, p. 263–274, 1993.
- [103] G. D. Evangelidis and E. Z. .. P. i. a. u. e. c. c. m. I. T. o. P. A. a. M. I. 3. Psarakis, “Parametric image alignment using enhanced correlation coefficient maximization. IEEE Transactions on Pattern Analysis and Machine Intelligence.”, vol. 30, no. 10, p. 1858–1865, 2008.
- [104] O. Faugeras, B. Hotz, H. Mathieu, T. Viéville, Z. Zhang, P. Fua, E. Théron, L. Moll, G. Berry, J. Vuillemin, P. Bertin and C. Proy, “Real time correlation-based stereo: algorithm, implementations and applications.”, INRIA, 1993.
- [105] D. S. Heiko Hirschmuller, “Evaluation of cost functions for stereo matching.”, *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pp. 1 - 8, 17-22 June 2007.
- [106] M. Jenkin, B. Verzijlenberg and A. Hogue, “Progress towards underwater 3D scene recovery.”, in *Canadian Conference on Computer Science & Software Engineering* , 2010.
- [107] G. D. Gilbert and J. C. Pernicka, “Improvement of Underwater Visibility by Reduction of Backscatter with a Circular Polarization Technique.”, *Applied Optics*, vol. 6, no. 4, pp. 741 - 746.
- [108] “Detection of Interest Points in Turbid Underwater Images.”, in *IEEE OCEANS Conference 2011*, Santander, Spain, 2011.
- [109] S. Duntley, “Light in the sea.”, *Journal of the Optical Society of America*, vol. 53, no. 2, pp. 214-233, 1963.
- [110] B. McGlamery, “Computer analysis and simulation of underwater camera system performance.”, Technical Report SIO Ref. 75-2, Scripps Institution of Oceanography, January 1975.
- [111] J. Jaffe, “Computer modeling and the design of optimal underwater imaging systems.”, *Oceanic Engineering, IEEE Journal of*, vol. 15, no. 2, pp. 101 - 111, April 1990.
- [112] J. Ferrer, “Optical Seafloor Mapping, Master Thesis.”, [Online]. Available: <http://eia.udg.edu/~jferrerp/papers/OpticalSeafloorMapping.pdf>. [Accessed 1 December 2014].
- [113] Y. Swirski, Y. Schechner, B. Herzberg and S. Negahdaripour, “Stereo from flickering caustics.”, in *Computer Vision, 2009 IEEE 12th International Conference on*, Kyoto, 2009.
- [114] A. Shihavuddin, N. R. E. Grácias and R. García Campos, “Online Sunflicker Removal using Dynamic Texture Prediction.”, *VISAPP 2012*, vol. 1, pp. 161-167, 2012.
- [115] A. Bradley, M. Feezor, H. Singh and F. Sorrell, “Power systems for autonomous.”, *IEEE Journal of Oceanic Engineering*, vol. 26, no. 4, pp. 526-538, 2001.

- 
- [116] H. Singh, R. Eustice, C. Roman and O. Pizarro., "The SeaBED AUV - a platform for high resolution imaging.", in *Unmanned Underwater Vehicle Showcase*, Southampton, Oceanography Centre, UK, 2002.
- [117] H. Singh and J. Howland, "A forensic analysis of the remains of flight EA990. Poster.", in *Woods Hole Oceanographic Institution*, 1999.
- [118] R. Schettini and S. Corchs, "Underwater Image Processing: State of the Art of Restoration and Image Enhancement Methods.", *EURASIP Journal on Advances in Signal Processing*, vol. 2010, pp. 1-14, 2010.
- [119] R. Garcia and N. Gracias, "Detection of Interest Points in Turbid Underwater Images.", in *IEEE OCEANS Conference 2011*, Santander, Spain, 2011.
- [120] H. Singh, R. Armstrong, F. Gilbes, R. Eustice, C. Roman, O. Pizarro and J. Torres., "Imaging coral I: Imaging coral habitats with the SeaBED AUV.", *Subsurface Sensing Technologies and Applications*, vol. 5, no. 1, pp. 25-42, 2004.
- [121] S. Negahdaripour, X. Xu and L. Jin, "Direct estimation of motion from sea.", *Oceanic Engineering, IEEE Journal of*, vol. 24, no. 3, pp. 370-382, 1999.
- [122] S. Negahdaripour and X. Xu, "Mosaic-based positioning and improved motion-estimation methods for automatic navigation of submersible vehicles.", *Oceanic Engineering, IEEE Journal of*, vol. 27, no. 1, pp. 79 - 99, 2002.
- [123] D. P. Connelly, B. J. Bett, H. A. Ruhl, K. J. Morris, J. Peakall, D. R. Parsons, E. J. Sumner, S. E. Darby, R. M. Dorrell and J. E. Hunt, "Autonomous Underwater Vehicles (AUVs): Their past, present and future contributions to the advancement of marine geoscience.", *Marine Geology*, vol. 352, pp. 451-468, 2014.
- [124] M. Prats, D. Ribas, N. Palomeras, J. C. García, V. Nannen, S. Wirth, J. J. Fernández, J. P. Beltrán, R. Campos, P. Ridao, P. J. Sanz, G. Oliver, M. Carreras, N. Gracias, R. Marín and A. Ortiz, "Reconfigurable AUV for intervention missions: a case study on underwater object recovery.", *Intelligent Service Robotics*, vol. 5, no. 1, pp. 19-31, 2012.
- [125] J. Leonard, A. Bennett, C. Smith and H. Fed, "Autonomous underwater vehicle navigation.", *MIT Marine Robotics Laboratory Technical Memorandum 98-1*, 1998.
- [126] J. Kinsey, R. Eustice and L. Whitcomb, "A Survey of Underwater Vehicle Navigation: Recent Advances and New Challenges.", in *Proc. of MCMC'2006 - 7th Conference on Manoeuvring and Control of Marine Craft*, Lisbon, Portugal, 2006.
- [127] K. Vickery, "Acoustic positioning systems. A practical overview of current systems.", in *Autonomous Underwater Vehicles, 1998. AUV'98. Proceedings of the 1998 Workshop on*, Cambridge, MA, 20-21 Aug 1998.
- [128] K. Vickery, "Acoustic positioning systems. New concepts-the future.", in *Autonomous Underwater Vehicles, 1998. AUV'98. Proceedings of the 1998 Workshop on*, Cambridge, MA, 20-21 Aug 1998.
- [129] P. Milne, Underwater acoustic positioning systems, 1983.
- [130] M. M. Hunt, W. M. Marquet, D. A. Moller, K. R. Peal and W. K. Smith, "An Acoustic Navigation System, Technical Report WHOI-74-6.", Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, USA, 1974.
- [131] E. Geyer, P. Creamer, J. D'Appolito and R. Rains, "Characteristics and capabilities of navigation systems for unmanned untethered submersibles.", in *Unmanned Untethered Submersible*

- Technology, Proceedings of the 1987 5th International Symposium on Vol.(5)*, Durham, NH, USA, 1987.
- [132] A. Matos and N. Cruz, "AUV navigation and guidance in a moving acoustic network.", *Oceans 2005 - Europe*, vol. 1, pp. 680 - 685, 20-23 June 2005.
- [133] A. Alcocer, P. Oliveira and A. Pascoal, "Underwater Acoustic Positioning Systems Based on Buoys with GPS.", in *Proceedings of the Eighth European Conference on Underwater Acoustics, 8th ECUA*, Carvoeiro, Portugal, June 2006.
- [134] J. Escartín, R. García, O. Delaunoy, J. Ferrer, N. Gracias, A. Elibol, X. Cufi, L. Neumann, D. J. Fornari, S. E. Humphris and J. Renard, "Globally aligned photomosaic of the Lucky Strike hydrothermal vent field (Mid-Atlantic Ridge, 37°18.5'N): Release of georeferenced data, mosaic construction, and viewing software.", *Geochemistry, Geophysics, Geosystems*, vol. 9, no. 12, 2008.
- [135] S. Smith and D. Kronen, "Experimental results of an inexpensive short baseline acoustic positioning system for AUV navigation.", in *OCEANS '97. MTS/IEEE Conference Proceedings*, Halifax, NS, October 1997.
- [136] J.-P. Peyronnet, R. Person and F. Rybicki, "POSIDONIA 6000: a new long range highly accurate ultra short base line positioning system.", in *OCEANS '98 Conference Proceedings*, Nice, 28 Sep-1 Oct 1998.
- [137] B. Jalving, N. Kjeller, K. Gade, O. Hagen and K. Vestgard, "A toolbox of aiding techniques for the HUGIN AUV integrated inertial navigation system.", in *OCEANS 2003. Proceedings*, San Diego, CA, USA, 2003.
- [138] M. Audric, "GAPS, a new concept for USBL [Global Acoustic Positioning System for Ultra Short Base Line positioning].", *OCEANS '04. MTT/IEEE TECHNO-OCEAN '04*, vol. 2, pp. 786 - 788, 9-12 Nov. 2004.
- [139] F. Parthiot and J.-F. Denis, "A better way to navigate on deep sea floors.", in *OCEANS '93. Engineering in Harmony with Ocean. Proceedings*, Victoria, BC, 18-21 Oct 1993.
- [140] J. Opderbecke, "At-sea calibration of a USBL underwater vehicle positioning system.", in *OCEANS '97. MTS/IEEE Conference Proceedings*, Halifax, NS, 6-9 Oct 1997.
- [141] H. Singh, J. Catipovic, R. Eastwood, L. Freitag, H. Henriksen, F. Hover, D. Yoerger, J. Bellingham and B. Moran, "An integrated approach to multiple AUV communications, navigation and docking.", in *OCEANS '96. MTS/IEEE. Prospects for the 21st Century. Conference Proceedings*, Fort Lauderdale, FL, 23-26 Sep 1996.
- [142] N. Brokloff, "Matrix algorithm for Doppler sonar navigation.", in *OCEANS '94. 'Oceans Engineering for Today's Technology and Tomorrow's Preservation.' Proceedings*, Brest, 1994.
- [143] R. L. Gordon, "Acoustic Doppler Current Profiler: Principles of operation a practical primer.", RD Instruments, San Diego, CA, USA,, 1996.
- [144] R. Instruments, "ADCP coordinate transformation: Formulas and calculations.", RDI Manual, San Diego, CA, USA, 1998.
- [145] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems.", *Journal of Fluids Engineering*, vol. 82, no. 1, pp. 35-45, 1960.
- [146] W. Steyn, "A Multi-Mode Attitude Determination and Control System for SUNSAT.", in *3rd International Symposium on Small Satellites Systems and Services*, Annecy, France, December 1995.
- [147] M. Psiaki, "Three-Axis Attitude Determination via Kalman Filtering of Magnetometer Data.", AIAA,

---

vol. 13, no. 3, pp. 506-514, May-June 1990.

- [148] B. A. McElhoe, "An Assessment of the Navigation and Course Corrections for a Manned Flyby of Mars or Venus.", *Aerospace and Electronic Systems, IEEE Transactions on*, Vols. AES-2, no. 4, pp. 613 - 623, July 1966.
- [149] L. A. Mc Gee, S. F. Schmidt and G. L. Smith, "APPLICATION OF STATISTICAL FILTER THEORY TO THE OPTIMAL ESTIMATION OF POSITION AND VELOCITY ON BOARD A CIRCUMLUNAR VEHICLE.", NASA Technical Documents, 1962.
- [150] J. L. Crassidis and J. L. Junkins, *Optimal Estimation of Dynamic Systems*, Second Edition, CRC Press, 2011.
- [151] S. Julier, J. Uhlmann and H. Durrant-Whyte, "A new approach for filtering nonlinear systems.", in *American Control Conference, Proceedings of the 1995*, Seattle, WA, 21-23 Jun 1995.
- [152] P. Moutarlier and R. Chatila, "An experimental system for incremental environment modeling by an autonomous mobile robot.", in *Proceedings of the 1st International Symposium on Experimental Robotics*, Montreal, Canada, June 1989.
- [153] R. Smith, M. Self and P. Cheeseman, "Estimating Uncertain Spatial Relationships in Robotics.", in *Autonomous robot vehicles*, New York, NY, USA, Springer-Verlag, 1990, pp. 167-193.
- [154] R. Smith, M. Self and P. Cheeseman, "A stochastic map for uncertain spatial relationships.", in *Autonomous Mobile Robots*, S.S. Lyengar and A. Elfes, editors. IEEE Computer Society Press, 1991, pp. 323-330.
- [155] J. Guivant and E. Nebot, "Optimization of the simultaneous localization and map building algorithm for real time implementation.", *IEEE Transactions on Robotics and Automation*, vol. 17, no. 3, pp. 242-257, June 2001.
- [156] J. Leonard and H. Feder., "Decoupled stochastic mapping.", *IEEE Journal of Oceanic Engineering*, vol. 26, no. 4, pp. 561-571, 2001.
- [157] J. Leonard and P. Newman, "Consistent, convergent, and constant-time SLAM.", in *Proceedings of the International Joint Conference on Artificial Intelligence*, Acapulco, Mexico, Aug. 2003.
- [158] M. Montemerlo, S. Thrun, D. Koller and B. Wegbreit, "FastSLAM: A Factored Solution to the Simultaneous Localization and Mapping Problem.", in *Proceedings of the AAAI National Conference on Artificial Intelligence*, Edmonton, Canada, 2002.
- [159] J. Leonard and R. Rikoski, "Incorporation of delayed decision making into stochastic mapping.", in *Experimental Robotics VII, volume 271 of Lecture Notes in Control*, Springer-Verlag, 2001, pp. 533-542.
- [160] J. Leonard, R. Rikoski, P. Newman and M. Bosse, "Mapping partially observable features from multiple uncertain vantage points.", *International Journal of Robotics Research*, vol. 21, no. 10, pp. 943-975, October 2002.
- [161] J. Neira and J. Tardos, "Data association in stochastic mapping using the joint compatibility test.", *Robotics and Automation, IEEE Transactions on*, vol. 17, no. 6, pp. 890 - 897, Dec 2001.
- [162] F. Lu and E. Milios, "Globally Consistent Range Scan Alignment for Environment Mapping.", *Autonomous Robots*, vol. 4, no. 4, pp. 333-349, 1997.
- [163] H. P. Moravec, "Sensor Fusion in Certainty Grids for Mobile Robots.", *AI Magazine*, vol. 9, no. 2, pp.

61-74, 1988.

- [164] R. M. Eustice, O. Pizarro and H. Singh, "Visually augmented navigation for autonomous underwater vehicles.", *IEEE Journal of Oceanic Engineering*, vol. 33, no. 2, pp. 103-122, April 2008.
- [165] H. Singh, G. Salgian, R. Eustice and R. Mandelbaum, "Sensor fusion of structure-from-motion, bathymetric 3D, and beacon-based navigation modalities.", in *IEEE International Conference on Robotics and Automation, Proceedings. ICRA '02.*, Washington, D.C., 2002.
- [166] N. Farr, A. Bowen, J. Ware, C. Pontbriand and M. Tivey, "An integrated, underwater optical/acoustic communications system.", in *OCEANS 2010 IEEE*, Sydney, 2010.
- [167] S. Negahdaripour, H. Sekkati and H. Pirsiavash, "Opti-Acoustic Stereo Imaging: On System Calibration and 3-D Target Reconstruction.", *Image Processing, IEEE Transactions on*, vol. 18, no. 6, pp. 12030-1214, June 2009.
- [168] S. Thrun, D. Koller, Z. Ghahramani, H. Durrant-Whyte and A. Y. Ng, "Simultaneous Mapping and Localization with Sparse Extended Information Filters: Theory and Initial Results.", *Algorithmic Foundations of Robotics V*, vol. 7, pp. 363-380, 2004.
- [169] R. M. Eustice, H. Singh, J. J. Leonard and M. R. Walter, "Visually Mapping the RMS Titanic: Conservative Covariance Estimates for SLAM Information Filters.", *The International Journal of Robotics Research*, vol. 25, no. 12, pp. 1223-1242, December 2006.
- [170] R. Eustice, H. Singh and J. Leonard, "Exactly Sparse Delayed-State Filters for View-Based SLAM.", *Robotics, IEEE Transactions on*, vol. 22, no. 6, pp. 1100 - 1114, Dec. 2006.
- [171] R. Eustice, H. Singh, J. Leonard, M. Walter and R. Ballard, "Visually Navigating the RMS Titanic with SLAM Information Filters.", in *Proceedings of Robotics - Science and Systems*, Cambridge, USA, 2005.
- [172] E. Diamanti, A. Georgopoulos and F. Vlachaki, "Geometric documentation of underwater archaeological sites.", in *XXIII CIPA Symposium, Proceedings*, 12/16 September 2011.
- [173] R. Ricci, L. De Dominicis, M. F. De Collibus, G. Fornetti, M. Guarneri, M. Nuvoli, M. Francucci and , "A 3D laser scanner prototype aimed at underwater archaeology applications.", *Lasers in the Conservation of Artworks VIII*, pp. 221 - 226, 2010.
- [174] S. Cocitoa, S. Sgorbini, A. Peirano and M. Valle, "3-D reconstruction of biological objects using underwater video technique and image processing.", *Journal of Experimental Marine Biology and Ecology*, vol. 297, no. 1, pp. 57 - 70, December 2003.
- [175] H. Singh, J. Adams, D. Mindell and B. Foley, "Imaging Underwater for Archaeology.", *Journal of Field Archaeology*, vol. 27, no. 3, pp. 319-328(10), 2000.
- [176] A. Mahiddine, J. Seinturier, D. Peloso, J.-M. Boï, P. Drap and D. Merad, "Underwater Image Preprocessing for Automated Photogrammetry in High Turbidity Water.", in *Proceedings of VSMM2012*, Milano, 2012.
- [177] A. Sedlazeck, K. Köser and R. Koch, "3D Reconstruction Based on Underwater Video from ROV Kiel 6000 Considering Underwater Imaging Conditions.", in *OCEANS 2009 - EUROPE*, Bremen, Germany, 11-14 May 2009.
- [178] T. Nicosevici, N. Gracias, S. Negahdaripour and R. Garcia, "Efficient 3D Scene Modeling and Mosaicing.", *Journal of Field Robotics.*, vol. 26, no. 10, p. 759-788, 2009.
- [179] D. Skarlatos, S. Demestiha and S. Kiparissi, "An 'Open' Method for 3D Modelling and Mapping in

- Underwater Archaeological Sites.”, *International Journal of Heritage in the Digital Era*, vol. 1, no. 1, 2012.
- [180] P. Drap, L. Long, A. Durand and P. Grussenmeyer, “From Underwater photogrammetry to a web integrated documentation system: the case of the 'Grand Ribaud F' etruscan wreck.”, in *CIPA 2001 International Symposium.*, Germany, 2001.
- [181] P. Drap, E. Bruno, L. Long, A. Durand and P. Grussenmeyer, “Underwater photogrammetry and xml based documentation system: The case of the 'Grand Ribaud F' etruscan wreck.”, in *ISPRS Comm. V Symposium*, CORFOU, Greece, 2002.
- [182] P. Drap, D. Scaradozzi, Gambogi and G. F. P., “Underwater photogrammetry for archaeology- the venus project framework.”, in *GRAPP'08*, Funchal, Madeira - Portugal , 2008.
- [183] L. Sorbi, D. Scaradozzi, G. Conte and P. Gambogi, “Geoposition data aided 3D documentation of archaeological and biological sites.”, in *ECUA 2012 11th European Conference on Underwater Acoustics*, July, 2012.
- [184] D. J. Gregory, “Development of Tools and Techniques to Survey, Assess, Stabilise, Monitor and Preserve Underwater Archaeological Sites: Sasmap.”, *International Journal of Heritage in the Digital Era*, vol. 1, no. 1, pp. 367-372, October, 2012.
- [185] T. Kwasnitschka, T. H. Hansteen, C. W. Devey and S. Kutterolf, “Doing fieldwork on the seafloor: Photogrammetric techniques to yield 3D visual models from ROV video.”, *Computers & Geosciences*, vol. 52, pp. 218 - 226, 2013.
- [186] B. Davidde, R. Petriaggi and G. Gomez de Ayala, “3D Documentation for the Assessment of Underwater Archaeological Remains.”, in *Computer Applications and Quantitative Methods in Archaeology (CAA) 2012*, Southampton, 26-30 March, 2012, [in press].
- [187] P. Caputo, “La topografia di Baia sommersa.”, in *P. Miniero e F. Zevi, Museo archeologico dei Campi Flegrei. Catalogo generale. Litternum, Baia, Miseno*, Napoli, Electa, 2008, pp. 142-143.
- [188] P. Caputo and N. Severino, “Parco sommerso di Baia: il restauro paesaggistico-ambientale di Punta Epitaffio, il recupero della strada Erculanea e i servizi per la fruizione delle aree archeologiche sommerse.”, in *F. Escalona, R. Ruggiero, Il progetto Integrato Campi Flegrei*, Napoli, Electa, 2008, p. 56.
- [189] R. Petriaggi and R. Mancinelli, “An experimental conservation treatment on the mosaic floor and perimeter walls of room n. 1 of the so-called «Villa con ingresso a protiro» in the underwater archaeological park of Baia (Naples).”, *Archaeologia Maritima Mediterranea*, pp. 109-26, 1 2004.
- [190] R. Petriaggi and B. Davidde, “The ISCR Project "Restoring Underwater": an evaluation of the results after ten years.”, in *4th Conference on Preserving Archaeological Remains in Situ PARIS4, vol. 14, 2012, pp.192-199*, 2012.
- [191] A. Mahiddine, J. Seinturier, D. Peloso, J.-M. Boï, P. Drap and D. Merad, “Underwater image preprocess- ing for automated photogrammetry in high turbidity water.”, in *VSMM2012*, Milan Italy, 2012.
- [192] Y. Furukawa and J. Ponce, “Accurate, Dense, and Robust Multi-View Stereopsis.”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 32, no. 8, pp. 1362-1376, 2010.
- [193] “MESH LAB.”, [Online]. Available: <http://meshlab.sourceforge.net/>. [Accessed 29 11 2014].
- [194] J. Henderson, O. Pizarro, M. Johnson-Roberson and I. Mahon, “Mapping Submerged Archaeological

- Sites using Stereo-Vision Photogrammetry.”, *International Journal of Nautical Archaeology*, 2013.
- [195] “Agisoft PhotoScan®.”, [Online]. Available: <http://www.agisoft.ru/>. [Accessed 4th July 2013].
- [196] “Rapidform®.”, [Online]. Available: <http://www.rapidform.com/home/>. [Accessed 29th November 2014].
- [197] P. S. Maybeck, *Stochastic models, estimation and control*. Vol. 1, Academic Press, 1982.
- [198] G. Welch and G. Bishop, “An Introduction to the Kalman Filter.”, University of North Carolina at Chapel Hill, Department of Computer Science. TR95-041, Chapel Hill, NC, USA, 2011. First edition 1995.
- [199] L. Whitcomb and D. Yoerger, “Preliminary experiments in model-based thruster control for underwater vehicle positioning.”, *Oceanic Engineering, IEEE Journal of*, vol. 24, no. 4, pp. 495 - 506, Oct 1999.
- [200] T. McLain, S. Rock and M. Lee, “Experiments in the coordination of underwater manipulator and vehicle control.”, in *OCEANS '95. MTS/IEEE. Challenges of Our Changing Global Environment. Conference Proceedings.*, San Diego, CA, 9-12 Oct 1995.
- [201] J. Yuh, “Design and Control of Autonomous Underwater Robots: A Survey.”, *Autonomous Robots*, vol. 8, no. 1, pp. 7-24, 2000.
- [202] R. Marks, H. Wang, M. Lee and S. Rock, “Automatic visual station keeping of an underwater robot.”, in *OCEANS '94. 'Oceans Engineering for Today's Technology and Tomorrow's Preservation.' Proceedings*, Brest, 1994.
- [203] X. Cufi, R. Garcia and P. Ridaó, “An approach to vision-based station keeping for an unmanned underwater vehicle.”, *Intelligent Robots and Systems, 2002. IEEE/RSJ International Conference on*, vol. 1, pp. 799 - 804, 2002.
- [204] L. Jin, X. Xu, S. Negahdaripour, C. Tsukamoto and J. Yuh, “A real-time vision-based stationkeeping system for underwater robotics applications.”, in *OCEANS '96. MTS/IEEE. Prospects for the 21st Century. Conference Proceedings*, Fort Lauderdale, FL, 23-26 Sep 1996.
- [205] S. van der Zwaan and J. Santos-Victor, “Real-time vision-based station keeping for underwater robots.”, in *OCEANS, 2001. MTS/IEEE Conference and Exhibition*, Honolulu, HI, 2001.
- [206] A. Ohya, A. Kosaka and A. Kak, “Vision-based navigation by a mobile robot with obstacle avoidance using single-camera vision and ultrasonic sensing.”, *Robotics and Automation, IEEE Transactions on*, vol. 14, no. 6, pp. 969 - 978, December 1998.
- [207] R. García, J. Puig, P. Ridaó and X. Cufi, “Augmented State Kalman Filtering for AUV Navigation.”, *International Conference on Robotics and Automation - ICRA*, pp. 4010-4015, 2002.
- [208] D. Ribas, N. Palomeras, P. Ridaó, M. Carreras and A. Mallios, “Girona 500 AUV: From survey to intervention.”, *Mechatronics, IEEE/ASME Transactions on*, vol. 17, no. 1, pp. 46-53, 2012.
- [209] D. Ribas, P. Ridaó, L. Magi, N. Palomeras and M. Carreras, “The Girona 500, a multipurpose autonomous underwater vehicle.”, in *OCEANS, 2011 IEEE - Spain*, Santander, 6-9 June 2011.
- [210] T. Fossen, *Guidance and Control of Ocean Vehicles*, New York: John Wiley and Sons Ltd., 1994.
- [211] N. Palomeras, S. Nagappa, D. Ribas and N. Gracias, “Vision-based localization and mapping system for AUV intervention.”, in *OCEANS - Bergen, 2013 MTS/IEEE*, Bergen, 10-14 June 2013.

- [212] P. S. Heckbert, in *Graphics Gems IV*, Boston: Academic Press, p. 474–485.
- [213] M. A. Fischler and R. C. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography.”, *Communications of the ACM*, vol. 24, no. 6, p. 381–395, 1981..
- [214] B. K. P. Horn, “ Closed–Form Solution of Absolute Orientation Using Unit Quaternion.”, *Journal of the Optical Society of America*, vol. 4, p. 629–642, 1987.
- [215] L. M. A. and A. Argyros, “SBA: A Software Package for Generic Sparse Bundle Adjustment.”, *ACM Trans. Math. Software*, vol. 36, no. 1, pp. 1-30, 2009.
- [216] H. Zhang and S. Negahdaripour, “EKF-Based Recursive Dual Estimation of Structure and Motion From Stereo Data.”, *Oceanic Engineering, IEEE Journal of*, vol. 35, no. 2, pp. 424 - 437, 26 aprile 2010.
- [217] J. Craig, *Introduction to Robotics: Mechanics and Control*, New York: Addison-Wesley Longman Publishing, 1989.
- [218] R. M. Haralick, “Propagating Covariance In Computer Vision.”, Jerusalem, Israel, Oct. 1994..