UNIVERSITÀ POLITECNICA DELLE MARCHE
SCUOLA DI DOTTORATO DI RICERCA IN INGEGNERIA DELL'INFORMAZIONE
CURRICULUM IN INGEGNERIA INFORMATICA, GESTIONALE E DELL'AUTOMAZIONE

Design of SLAM strategies and technologies in structured and complex environments

Ph.D. Dissertation of:

Silvia Zingaretti

Advisor:

Dott. Ing. David Scaradozzi

Curriculum Supervisor:

Prof. Francesco Piazza

XVI cycle - new series

UNIVERSITà POLITECNICA DELLE MARCHE

DOTTORATO DI RICERCA IN INGEGNERIA DELL'INFORMAIZONE

FACOLTà DI INGEGNERIA

Via Brecce Bianche – 60131 Ancona (AN), Italy

To Matteo

A Matteo

Table of contents

LIST OF FIGURES

LIST OF ABBREVIATIONS

CCD - Charge-Coupled Device

CMOS - Complementary Metal-Oxide Semiconductor

COTS - Commercial Off-The-Shelf

ICP – Iterative Closest Point

IMU – Inertial Measurement Unit

MIBACT - Ministero dei Beni e delle Attività Culturali e del Turismo

PCL – Point Cloud Library

ROI - Region of Interest

SfM - Structure from Motion

SLAM - Simultaneous Localization and Mapping

STMP – Servizi Tecnici Marittimi Portuali

ToF – Time of Flight

ACKNOWLEDGMENTS

At the end of this three-years of doctorate I would like to thank you all the people who, in various ways, have accompanied and supported me during this iter with their suggestions and advices and without which this work thesis would not have been possible.

First of all I would like to thank you my academic supervisor, Prof. David Scaradozzi, for being always available to offer me his methodological and scientific contribution during all the phases of the research work and to let me do stimulating and formative experiences for my growth as a PhD.

I would also like to thank my company tutor, Cristina Cristalli, for giving me the possibility of being part of some big and challenging works and showing me the industrial reality in an innovative company like Loccioni Group.

I would like to say a big thank you to my PhD colleagues and friends, Luca, Laura, Lorenzo and Nicolò, with whom I shared lessons, works, preoccupations and frustrations, but also ideas, satisfactions, apartment sharing and… trips!

Lastly, a special thanks to my family for never making me lack moral support. An immense thanks to Matteo, without whom everything would not have been possible. Thank you for being close to me even when we were physically distant.

ABSTRACT

The research activity has been focused on the study and design of a framework for visual SLAM strategies and technologies, to be applied in robotic systems working in partially structured and complex environments. The framework aims at processing the information coming from the mechatronic subsystems in order to obtain the map of the scene at each instant. In a SLAM problem, in fact, a mobile robot tries to build a map of the environment while simultaneously use it to localize itself. The map can be represented in different ways depending on the sensors used, on the characteristics of the environment and on the estimation algorithm.
Studies about implementation aspects of the SLAM problem in literature have been firstly conducted, in order to properly characterize the entire framework.
Innovative contributions rely on the fact that the proposed framework could have impact on various target groups: it is not designed for a specifically application, but it can be adapted to various scenarios and working environment.
Additional contributions consisted in the intensive adoption of a Computer vision and 3D imaging approach to process and elaborate the sensors information into a useful 3D model. The combination and fusion of many sensors is introduced to exploit the complementarity of each one and increase the accuracy of the 3D model.
The framework subsystems have been customized for its validation in two different environments: two real industrial processes and two real underwater applications have been used as testing situations. In each scenario the number and the type of sensors chosen are different, accordingly to the tasks that the specific robotic system has to perform. Each situation is characterized by a different robotic system and the results have been discussed.

Keywords: SLAM; vision system; underwater applications

SOMMARIO

L'attività di ricerca ha riguardato lo studio e la progettazione di un framework intorno alle strategie e alle tecnologie del Visual SLAM, per essere applicato sia in sistemi robotici in ambienti parzialmente strutturati che in quelli complessi. Il framework ha l'obiettivo di processare le informazioni provenienti dai sottosistemi meccatronici di cui è costituito in modo da ottenere la mappa della scena in ogni istante. In un problema di SLAM, infatti, un robot mobile cerca di costruire la mappa dell'ambiente in cui si muove ed usarla per localizzare se stesso. La mappa può essere rappresentata in diversi modi in base ai sensori utilizzati, alle caratteristiche dell'ambiente e agli algoritmi di stima impiegati.
Sono stati inizialmente condotti studi riguardo gli aspetti implementativi del problema dello SLAM affrontati in letteratura con lo scopo di caratterizzare il framework in maniera appropriata.
Contributi innovativi di questa ricerca hanno riguardato il fatto che il framework proposto può avere un impatto su differenti gruppi di end-users: non è stato progettato per un'applicazione specifica, ma può essere adattato a vari scenari e diversi ambienti di lavoro. Altri contributi sono costituiti da un uso intensivo di tecniche di Computer Vision e 3D Imaging per processare ed elaborare le informazioni provenienti dai sensori in un efficiente modello 3D. La combinazione e la fusione di più sensori sono utilizzate per sfruttare le complementarietà di ognuno e aumentare così l'accuratezza del modello 3D.
I sottosistemi del framework sono stato poi personalizzati per la validazione in due ambienti diversi: due processi industriali reali e due applicazioni subacquee sono state utilizzate come situazioni di test. In ogni scenario il numero e il tipo dei sensori scelti sono differenti in base ai compiti che lo specifico sistema robotico deve compiere. Ogni situazione, infatti, è caratterizzata da un diverso sistema robotico e i risultati ottenuti sono stati discussi.

Chapter 1

INTRODUCTION

1.1 BACKGROUND AND MOTIVATION

Robotics systems share common aspects: the physical world in which they are situated, the use of sensors to comprehend the environment, and its manipulation through actuators. Their most outstanding characteristic is that they operate in increasingly unstructured environments, which are inherently unpredictable, and one of the fundamental requirements is the ability to navigate within it. As a result, robotics is moving into areas where sensor input becomes increasingly important, and where robot software should be robust enough to cope with a range of situations.

Visual inspection and quality control were performed, and in some case still are, by human experts but their performance is susceptible to various factors. In fact, they are slower than the machines and get tired quickly when they perform long and routine tasks. Moreover, they require training and may need time to develop the right skills. In certain applications, precise information must be quickly or repetitively extracted and used (e.g., target tracking and robot guidance) and in some environments (e.g., underwater inspection, nuclear industry, chemical industry etc.) inspection may be difficult or dangerous. Moreover, nowadays most of the robotic applications need to be totally autonomous in their tasks.

Simultaneous Localization and Mapping (SLAM), also known as Concurrent Mapping and Localization (CML), is the process by which a mobile robot can build a map of the environment and at the same time use this map to compute its location. In recent years, the SLAM problem has received considerable attention by the scientific community, and

a flurry of new algorithms and techniques has emerged, which can be categorized in mainly two approaches: on-line and full SLAM.

A robot is usually equipped with internal sensors that are used to perceive its position, pose and motion parameters. The vision sensors made considerably huge progress since its first uses. The progress made has been principally on two separate fronts: vision-based navigation in structured environments, where the geometrical characteristics of the environment are known, and vision-based navigation for complex environments, e.g. outdoor.

Most of vision-based systems are monocular and binocular, although those based on trinocular configurations also exist. For those systems, a sub set of methodology has been studied and presented with the keyword of visual SLAM. Usually, these systems apply algorithms used in the Computer Vision community for Structure from Motion (SfM). SfM has become an emergent methodology in the recent decades; it estimates the camera poses and the scene structure simultaneously without requirement for any previous knowledge, from the information about common points visible in the images. The accuracy of this technique depends on many factors like the resolution of the adopted camera, the working distance, the light conditions, etc., and so sometimes it is too variable, particularly if working in industrial environments.

In the recent years, roboticists are starting to use 3D vision sensors in their applications, and in particular on a new generation of active cameras based on the Time-of-Flight (ToF) principle. The ToF-sensor is a very compact device which allows to acquire data at high frame rates and to obtain 3D point clouds without scanning. Although the use of ToF cameras have several advantages in respect of the stereo systems, the resolution of their depth maps is far below the stereo depth maps and the measurements are greatly corrupted.

1.2 OBJECTIVES

The Ph.D. research activity described in the present dissertation has been funded by a project sponsored by a cooperation between Università Politecnica delle Marche (Dipartimento di Ingegneria dell'Informazione - DII) and Loccioni Group, an Italian company based on Marche region, that develops, measures and tests solutions to improve the quality of products and processes for the manufacturing and service industry

This work arose from the necessity of increasing the capability of scene understanding of a working environment by means of vision sensors. Even if there are other type of sensors that allow a more precise acquisition in many situations, today vision sensors can be implemented in a wide variety of industries and applications, offering cost advantages over standard solutions in difficult sensing environments. To increase the accuracy of data acquisition and the successive elaboration, a combination of different sensors is often introduced to exploit the complementarity of each one, and this concept has been adopted also in the present work.

The aim of this work was to design and develop a framework for visual SLAM strategies and technologies in partially structured and complex environments. This framework is based on mechatronic systems equipped with various high efficient sensors, employed in different robotics applications working in delicate industrial environment or underwater environments. The study, during the three years of activities, considered different optical sensors and the implementation of different techniques to map the environment in which the robot is moving. The information extracted from each sensor at each instant t, t+1, etc., have been merged at high level in order to be employed in SLAM algorithms.

The main objectives of this dissertation can be listed as follows:

1. Design a framework for visual SLAM strategies;
2. Study and develop Computer Vision and 3D imaging techniques to be implemented within the framework, to extract the information from the images acquired and mapping the environment;

3. Implement and validate the framework in different case studies in industrial and underwater environments for inspection or documentation activities.

The application domain, the tasks to be accomplished, the environment, the speed, etc., play an important role in the design and development of a successful vision system. It is essential to define the requirements of the application domain and to understand what kind of information the machine vision system needs to retrieve and how this is translated into measurements or features extracted from images.

The next chapters will present different application fields of SLAM problems in both partially structured and complex environments, for inspection or documentation activities, detailing the development of vision systems for the extraction and the measurement of the information.

1.3 OUTLINE OF THE THESIS

The contents of this thesis can be divided in three parts. The first part overviews the SLAM techniques most implemented in structured and complex environments, with a focus on those which adopt optical sensors (Chapter 2). The second part include the development of the two vision systems in the context of the industrial case studies, with the partnership of Loccioni Group (Chapter 3 and Chapter 4). Lastly, the last part describes the development of mechatronic systems to be applied in not structured environment, specifically for underwater applications (Chapter 5).

A brief description of each chapter is presented in the following:

Chapter 2, “State of the art on SLAM algorithms and techniques” presents the SLAM algorithms and techniques most used in literature from 1977 to 2017, with a focus on Visual SLAM and on the algorithms applied in underwater environments. It then examines the use of 3D cameras in industrial environments, highlighting advantages and disadvantages, for building the map of the scene.

Chapter 3, “First case study: inspection and obstacle detection in industrial environment” presents the first case study: the design and development of a vision system with the use of Time of Flight cameras for an industrial application. The chapter describes the system developed, showing also the results obtained.

Chapter 4, “Second case study: Bayer’s challenge” presents the second industrial application, explaining the context in which it has been implemented and then describing in detail the vision system and the results.

Chapter 5, “Underwater applications” describes two innovative platforms that allow data gathering for documenting and mapping the complex underwater environment, describing their vision systems and the results obtained.

Chapter 2

STATE OF THE ART ON SLAM ALGORITHMS AND TECHNIQUES

Nowadays robots exercise a high degree of autonomy in a huge amount of different applications. One of the fundamental requirements of an autonomous robot is the ability to navigate within a dynamic, unexplored environment. Navigation can be described as the process of determining a suitable and safe path between a starting point and a goal point for the robot travelling between them [1]. For a successful navigation, it is essential to resolve the following three problems:

- Localization: determining where the robot is
- Goal recognition: determining where the robot has to go
- Path planning: determining how the robot should reach the final point.

To address this purpose, a variety of sensors and techniques has been adopted in literature. The success of a path planning and navigation mission of an autonomous vehicle depends on the availability of both a sufficiently reliable estimation of the vehicle location and an accurate representation of the navigation area. The navigation and localization systems can be mainly divided in two groups: those who need a map of the environment before the navigation starts, with different degree of details depending on the study, and those who are able to build the map while moving. Particularly interesting for this work are the second ones.

Simultaneous Localization and Mapping (SLAM), also known as Concurrent Mapping and Localization (CML), is the process by which a mobile robot can build a map of the environment and at the same time use this map to compute its location, as defined by Tim Bailey and Hugh Durrant-Whyte [2]. Initially, both the map and the vehicle position

are not known, the vehicle has a known kinematic model and it is moving through the unknown environment, which is populated with artificial or natural landmarks. SLAM has been formulated and solved as a theoretical problem in a number of different forms. It has also been implemented in many different domains from indoor to outdoor, underwater and airborne systems. A robot starts its navigation in an unknown environment from an unknown location, along a trajectory described by a sequence of random variables, and acquires odometry measurements and observations of the environment while moving. The ability of computing both the map and the robot location is usually due to formulating the statistical correlations between the estimates of the robot position and landmarks, and between those of the landmarks themselves. The overwhelming majority of SLAM algorithms adopts a probabilistic framework that takes advantage of the high degree of correlation between the estimates of individual landmark locations [4] and that tackles the problem of the presence of uncertainty and sensor noise by explicitly modeling different sources of noise and their effects on the measurements [3]. By expressing mapping, so the problem of integrating the information gathered with the robot's sensors into a given representation, and localization, so the problem of estimating the pose of the robot relative to a map, as a joint problem in which the state of the system is composed of both the vehicle pose and every estimated landmark position, it is possible to observe that the correlations grow as the state is updated after each observation.

The problem requires that the probability distribution

$$P(x_k, m \mid Z_{0:k}, U_{0:k}, x_0) \quad (1)$$

where x_k is the state vector describing the location and orientation of the vehicle, m is the map of the environment (the set of all landmarks), $U_{0:k}$ is the history of control inputs, $Z_{0:k}$ the set of all landmarks observations and x_0 the initial state, is computed at each instant time. This probability distribution describes the joint posterior density of the landmark locations and vehicle state (at time k) given the complete history of

landmark observations and control inputs up to and including time k together with the initial state of the vehicle.

In general, a recursive solution to the SLAM problem is desirable. Given both an observation model and a state transition model, this probability function can be computed using Bayes theorem. The **observation model**, that describes the probability of making an observation z_k when the vehicle location and landmark locations are known, is in the form

$$P(z_k \mid x_k, m). \tag{2}$$

The state transition model (**motion model**) is assumed to be a Markov process where the state x_k depends only on the immediately preceding state x_{k-1} and the applied control u_k, and is independent of both the observations and the map. It is in the form:

$$P(x_k \mid x_{k-1}, u_k). \tag{3}$$

For calculating the probability distribution of Eq. 1, the two following equations are recursively applied, as stated in [4].

- Time update (Prediction):

$$P(x_k, m|Z_{0:k-1}, U_{0:k}, x_0) = \int P(x_k| x_{k-1}, U_k) \times P(x_{k-1}, m|Z_{0:k-1}, U_{0:k-1}, x_0) dx_{k-1} \tag{4}$$

- Observation update (Correction):

$$P(x_k, m|Z_{0:k}, U_{0:k}, x_0) = \frac{P(z_k| x_k, m)\, P(x_k, m| Z_{0:k-1}, U_{0:k}, x_0)}{P(z_k|Z_{0:k-1}, U_{0:k})} \tag{5}$$

Therefore, the estimated landmark locations can be obtained by computing the following conditional density, assuming that the vehicle location is known at every instant:

$$P(m_k|X_{0:k}, Z_{0:k}, U_{0:k}) \tag{6}$$

Conversely, assuming that the landmark locations are known, the estimate of the vehicle location can be obtained by computing:

$$P(x_k|X_{0:k}, U_{0:k}, m_k). \quad (7)$$

However, in reality the landmark locations and vehicle position are never known with absolute certainty. What is known is that much of the error in the estimated landmark locations is common to all landmarks and arises from the error in the estimated vehicle position. Therefore, this error is highly correlated, which means that the relative location of landmarks may be known with high accuracy, even though the absolute location of the landmarks is quite uncertain [4].

The poses $x_{1:k}$ and the odometry measurements $u_{1:k}$ are usually represented as 2D or 3D transformations while the map can be represented in different ways. Maps can be parametrized as a set of spatially located landmarks, by dense representations like occupancy grids, surface maps or by raw sensor measurements. The choice of a particular map representation depends on the sensors used, on the characteristics of the environment and on the estimation algorithm. Landmark maps [5,6] are often preferred in environments where locally distinguishable features can be identified and especially when cameras are used. In contrast, dense representations [7-9] are usually used in conjunction with range sensors. Independently of the type of the representation, the map is defined by the measurements and the locations where these measurements have been acquired [10;11].

In recent years, the SLAM problem has received considerable attention by the scientific community, and a flurry of new algorithms and techniques has emerged. There are three main paradigms: Kalman Filters (KF), Particle Filters and Graph-based SLAM. The first two are also referred as filtering techniques, which model the problem as an on-line state estimation where the state of the system consists in the current robot position and the map. The estimate is augmented and refined by incorporating the new measurements when they become available. Due to their incremental nature, these approaches are generally acknowledged as *on-line* SLAM techniques. Conversely, Graph-based SLAM estimate the entire trajectory and the map from the full set of measurements; it is also called *full SLAM* technique.

There is no single best solution to the SLAM problem. The choice of the type of algorithm to use will depend on the peculiarities of the application and on a number of factors, such as the desired map resolution, the update time, the nature of the environment, the type of sensor the robot is equipped with, and so on.

2.1 KALMAN FILTERS

The popularity of the Kalman Filter approaches depends on the fact that it directly provides both a recursive solution to the navigation problem and a means of computing consistent estimates for the uncertainty in vehicle and map landmark locations based on statistical models. In fact, KF relies on the assumption that the next state probability is a linear function in its arguments with added Gaussian noise, the measurement probability must also be linear in its arguments with added Gaussian noise, and the initial belief must be normal distributed. These three properties assure that also posteriors are Gaussian [19]. In practice, the assumptions of linearity are rarely fulfilled. For this reason, two variations of KF are mainly employed in the state-of-the-art SLAM: Extended Kalman Filter (EKF) and Information Filtering (IF), with its extended version.

The EKF overcomes the linearity assumption: the next state probability and the measurement probabilities are described by nonlinear functions [19]. In literature, there exist several examples of the use of the EKF algorithm [13-18,179]. Smith et al. [12] were the first to present the idea of representing the structure of the navigation area in a discrete-time state-space framework, introducing the concept of stochastic map. In [20] authors present an algorithm for the evaluation of the systematic bias error in mathematical model of mobile robot during a 2D SLAM using a combination of recurrent neural network and Extended Kalman Filter. Furthermore, the EKF has been the basis of many recent developments in the field, like in [29].

The Unscented Kalman Filter (UKF) has been developed in recent years to overcome some main problems of the EKF [30]. In the EKF the state distribution is approximated

by a Gaussian Random Variable (GRV), which is then propagated analytically through the first-order linearization of the nonlinear system. These approximations can introduce large errors in the true posterior mean and covariance, which can bring to the divergence of the filter. The UKF addresses this problem by using a deterministic sampling approach. The state distribution is again approximated by a GRV but is represented using a minimal set of carefully chosen sample points, called σ-points. These sample points completely capture the true mean and covariance of the GRV, and when propagated through the true nonlinear system, captures the posterior mean and covariance accurately to the 3rd order of the Taylor series for any nonlinearity [27]. Some examples of the use of UKF for navigation and localization can be found in [31,159,186]. The UKF maintains the same order of the computational complexity of the EKF but it increases with the number of landmarks.

An algorithm that significantly reduces the computational requirement without introducing any penalties in the accuracy of the results is the Compressed Extended Kalman Filter (CEKF) [28]. A CEKF stores and maintains all the information gathered in a local area with a cost proportional to the square of the number of landmarks in the area. This information is then transferred to the rest of the global map with a cost that is similar to full SLAM but in only one iteration.

The dual of the Kalman filter is the Information Filter. It is subject to the same assumptions but the key difference arises in the way the Gaussian belief is represented. In the Kalman filter family of algorithms, Gaussians are represented by their moments (mean, covariance), while in information filters by their canonical representation, which is comprised of an information matrix and an information vector. This difference leads to different update equations [19]. There are several advantages of the IF filter over the KF. Firstly, the data is filtered by simply summing the information matrices and vector, providing more accurate estimates [21]. Secondly, the information filter tends to be numerically more stable than the Kalman filter in many applications [19]. But the prediction step of the IF involves the inversion of two matrices: as the state space grows in dimension, the computational complexity increases. So, in this phase,

the KF is more advantageous because the update step is additive. Anyway, these roles are reversed in the measurement step, illustrating the dual character of Kalman and Information Filters.

The extended version of the Information Filter (EIF) is analog to the EKF, where the nonlinearities are considered and the parameters of the linear model are replaced by nonlinear functions. Thrun et al. [21] developed a variant of the EIF, the Sparse Extended Information Filter (SEIF), that consists in an approximation, which maintains a sparse representation of environmental dependencies, in order to achieve a constant time updating. This algorithm derives from the observation that the normalized information matrix is sparse: it is dominated by a small number of strong links, and possesses many links whose values, when normalized, are near zero. Furthermore, link strength is related to distance of features: the more distant two landmarks are, the weaker their link is. They were inspired by other works on SLAM filters that represent relative distances, such as Newman's geometric projection filter [22] and extensions [23], Csorba's relative filter [24] or Lu and Milion [25] but neither of them are able to perform a constant time updating. Algorithms like SEIF or Ensemble Kalman Filters [160] are well suitable for extension to extremely large-scale settings.
To overcome the difficulties of both EKF and IF, and to be more efficient in terms of computational complexity, a Combined Kalman-Information Filter SLAM algorithm (CF-SLAM) has been adopted in [26]. It is a combination of EKF and EIF that allows executing highly efficient SLAM in large environments.

Although SLAM has been a very active research area in mobile robotics for the last decade, no extensive works on the essential properties of SLAM as an estimation problem have been published. The strongest results concerning the convergence of a SLAM algorithm are available for the Kalman filter based SLAM thanks to Dissanayake et al. [29], who examined a simple linear version of the problem and provided three important properties of SLAM regarding convergence and lower bound on the position uncertainty:

- "the decreasing of the determinant of any submatrix of the map covariance matrix as observations are made successively";
- "the landmark estimates become fully correlated in the limit";
- "the covariance associated with any single landmark pose estimate reaches a lower bound determined only by the initial covariance in the vehicle location estimate at the time of the first sighting of the first landmark" [29].

However, the convergence proofs given are only for a linear case, which do not depict any real SLAM implementation. Many authors showed that the EKF-SLAM produces inconsistent estimates [32-36, 39] and it has been pointed out that linearization is the main cause of it [33]. These problems cannot be eliminated but only reduced, as the basic problem is non-linear.

The second widely investigated aspect in SLAM algorithms is scalability. In general, SLAM algorithms should be capable of computing extensive areas, either using a single or multiple robots. Typically, EKF-based approaches to SLAM suffer from poor scalability and have a limited applicability to large maps, since their update stage has a quadratic dependence with the number of features in the map [29]. The common idea underlying most of some more efficient approaches is to decompose the problem of building one large map into a collection of smaller maps, which can be updated more efficiently [21, 28, 41, 42, 43, 161]. Local maps can be then joined together into a global map that is equivalent to the map obtained by the standard EKF-SLAM approach, except for linearization errors. As most of the mapping process consists in updating local maps, where errors remain small, the consistency of the global map obtained is greatly improved.

2.2 PARTICLE FILTERS

Particle filters [9, 30, 40] comprise a large family of sequential Monte Carlo algorithms [45] for calculating posteriors in partially observable controllable Markov chains with discrete time. The key idea of the particle filter is to represent the posterior by a set of

random state samples, the so-called *particles*, drawn from this posterior [19]. Particle filters are attractive to roboticists for different reasons [50]. First, they can be applied to almost any probabilistic robot model that can be formulated as a Markov chain. Secondly, they do not require a fixed computation time but their accuracy increases with the available computational resource. Finally, they are relatively easy to implement: they do not need to linearize non-linear models and do not worry about closed-form solutions of the conditional probability as in Kalman filters. The main limitation of plain particle filters is their poor performance in higher dimensional spaces. In fact, the number of particles needed to populate a *d*-dimensional space increases exponentially with *d*, so most successful applications have therefore been confined to low-dimensional state spaces. However, many SLAM problems [46,47,48,49] possess structure that can be exploited to develop more efficient particle filters within applications in higher dimensional spaces.

Some of the EKF limitations mentioned in the previous paragraph have been overcome by particle filters. Recent research [9, 51, 52, 53] has led to a family of so called Rao-Blackwellized particle filters that lead to more efficient solutions. These particle filters require time O(M logN) instead of $O(N^2)$, where M is the number of particles. They can also incorporate negative information, hence make better use of measurement data. Furthermore, they provide a better solution to the data association problem. However, these algorithms are susceptible to considerable estimation inconsistencies because they generally underestimate their own error [56]. In large part this is due to degeneracies in the sampling process. Different sampling strategies to improve the consistency of the filter's estimation and the diversity of the trajectory samples have been adopted [55,57,58].

A particular case is FastSLAM [6,9,54]: it denotes a family of algorithms that integrates particle filters and extended Kalman filters. FastSLAM exploits the structural property of the SLAM problem [53] that the features estimates are conditional independent given the observations, the controls, and the robot path. This observation leads to define a factored representation of the posterior over poses and maps: as the individual map

errors are independent, the mapping problem can be factored into separate problems, one for each feature in the map. FastSLAM uses particle filters for estimating the robot path and, for each particle representing a robot path, uses the EKF for estimating feature locations. It offers computational advantages over plain EKF implementations. The filter maintains posteriors over multiple data associations, not just the most likely one, and this feature makes FastSLAM significantly more robust to data association problems than algorithms based on maximum likelihood data association [54]. Furthermore, it can cope with non-linear robot motion models, whereas EKF-style techniques approximate such models via linear functions. However, the presence of a static parameter (the map) in the state space prevents the particle approximation from converging uniformly in time [62]. To overcome this difficulty, [63] introduced a marginal SLAM algorithm: the key difference relies on the nature of the map which is treated as a parameter used to drive a latent data model. This parameter is estimated by a recursive maximum likelihood procedure, solved in practice by a stochastic gradient algorithm [10]. Consequently, this marginal algorithm provides a point estimate of the map and a particle approximation of the marginal posterior distribution of the robot pose at each time.

None of these approaches, however, offer constant time updating while simultaneously maintaining global consistency of the map. To address it, Paskin [59] applied a thin junction trees to the SLAM problem. Currently, researchers are focusing their efforts on Graph-based SLAM algorithms [81], which present better scalability performance and a more compact approach to SLAM (see paragraph 2.4).

2.3 EXPECTATION MAXIMIZATION (EM)

Another popular technique adopted in literature is the Expectation Method (EM) [60]. It is an efficient iterative procedure to compute the Maximum Likelihood (ML) estimate in the presence of missing or hidden data. Each iteration of the EM algorithm consists of two processes: the E-step and the M-step. In the expectation, or E-step, the missing data are estimated given the observed data and current estimate of the model

parameters. In the M-step, the likelihood function is maximized under the assumption that the missing data are known. The estimate of the missing data from the E-step are used in place of the actual missing data. Convergence is assured since the algorithm is guaranteed to increase the likelihood at each iteration. A real-time implementation of this algorithm is described in [61].

Expectation algorithms require the whole data be available at each iteration; for this reason, when processing large data sets or data streams, it becomes impractical. As a solution to this, an online version has been implemented [64], where there is no need to store the data since they are used sequentially. The online E-step is divided into a sequential Monte Carlo step and a stochastic approximation step, which incorporates the information brought by the newly available observation. Therefore, each iteration of the online EM-SLAM provides both a particle approximation of the distribution of the pose and a point estimate of the map. This algorithm has been used also to relax the assumption that the environment in many SLAM problems is static. In fact, at present, most of the methods existing in literature are robust for mapping environments that are static, structured, and limited in size, while mapping unstructured, dynamic, or large-scale environments remains an open research problem. The directions followed in literature are mainly two. The first tries to partition the model into two maps: one map holds only the static landmarks and the other holds the dynamic landmarks [65,66]. The second approach, instead, tries to track moving objects while mapping the static landmarks [67,68]. In [69] authors propose a new SLAM algorithm for non-stationary environments which makes use of a new storing structure, the Histogram Occupancy Grid. The Variational Bayesian Expectation Method (VBEM) algorithm can deal with unknown number of landmarks and can tackle the uncertainties in the data associations efficiently. In [70] a radar map is built assuming known trajectory of the sensor; the VBEM translates an inference problem to an optimization one and it is guaranteed to find a local optimum. Authors in [71] developed a whole SLAM solution by means of that. This method is computationally efficient and easy to implement.

2.4 GRAPH-BASED SLAM

Graph-based SLAM models the SLAM system as a graphical structure. Formulating SLAM in a smoothing context adds the complete trajectory into the estimation problem and postpones the mapping process until the end, thus offering improved performance. While this does not seem intuitive at first, because more variables are added to the estimation problem, the simplification arises from the fact that the smoothing information matrix is naturally sparse, as observed by Golfarelli, Maio and Rizzi [72]. Adopting a graphical formulation is an intuitive way to address the full SLAM problem.

Solving a graph-based SLAM problem involves building a graph whose nodes represent robot poses or landmarks, linked by soft constraints established by sensor measurements [73]. Relaxing these constraints yields the robot's best estimate for the map and the full path. Thus, in graph-based SLAM the problem is decoupled in two tasks. The first refers to the identification of the constraints from the raw measurements and the construction of the graph. Obviously, such constraints can be contradictory since observations are always affected by noise. The graph construction is usually called *front-end* and it is heavily sensor dependent. The second problem is to correct the poses of the robot to obtain a consistent map of the environment given the constraints. This part of the approach is often referred to as the optimizer or the SLAM *back-end*. So, once such a graph is constructed, the crucial problem is to find a configuration of the nodes that is maximally consistent with the measurements. This involves solving a large error minimization problem [90].

These techniques have been firstly introduced by Lu and Milios [25], who represented the SLAM problem as a set of links between robot poses. The map is then refined by globally optimizing the system of equations to reduce the error introduced by constraints. Bosse et al. [74] developed the ATLAS framework, which used multiple connected local maps to integrate global and local mapping, limiting the representation of errors to local regions and adopting topological methods by which local submaps can be managed to provide a global map. Specifically, it consists of a two-level hierarchy of

graphs and employs a Kalman filter to construct the bottom level. Then, a global optimization approach aligns the local maps at the second level. Similarly, Estrada et al. proposed Hierarchical SLAM [75] as a technique for using independent local maps and the work of Nüchter et al. [76] aims at building an integrated SLAM system for 3D mapping.

Gutmann and Konolige [77] proposed an effective way for constructing the network and for detecting loop closures while running an incremental estimation algorithm. Folkesson and Christensen [78,79] exploited the optimization perspective of SLAM by applying gradient descent to the log-likelihood version of the SLAM posterior. Their algorithm reduced the number of variables to the path variables when closing the loop and this reduction significantly accelerated gradient descent. Olson et al. [80] likewise presented an efficient optimization approach which is based on the stochastic gradient descent and can efficiently correct even large pose-graphs. Grisetti et al. proposed an extension of Olson's approach that uses a tree parametrization of the nodes in 2D and 3D. In this way, they increase the convergence speed [81]. Konolige [82] and Montemerlo and Thrun [83] introduced conjugate gradient into the field of SLAM, which is more efficient than gradient descent. Both also reduced the number of variables when closing large cycles. GraphSLAM [84] applies variable elimination techniques to reduce the dimensionality of the optimization problem. In this way, it can handle large number of features and obtains the map and the robot path by resolving the constraints into a globally consistent estimate. The nonlinear constraints are linearized and the resulting least squares problem is solved using standard optimization techniques.

The error minimization in the constraint network has been addressed with many approaches. For example, Howard et al. [85] apply relaxation to localize the robot and build a map. Frese et al. [86] proposed a multi-level relaxation (MLR), which applies it at different resolutions. Dellaert and Kaess [87] were the first to exploit sparse matrix factorizations to solve the linearized problem in full SLAM. Subsequently Kaess [88] presented iSAM, which performs fast incremental updates of the square root

information matrix to compute the full map and trajectory at any time. However, to remain efficient and consistent, iSAM requires periodic batch steps for variable reordering and re-linearization, even it is computational expensive. The Bayes tree proposed in [89] provides higher efficiency while retaining sparseness and full accuracy. The Bayes net is the equivalent of the square root information matrix but it allows a fluid re-linearization of a reduced set of variables.

2.5 VISUAL SLAM

In the SLAM problem, a variety of sensors could be adopted. A robot is usually equipped with internal sensors that are used to perceive its position, pose and motion parameters. In addition to these, other sensors are usually added in order to improve the performances. The progress made in vision-based navigation and localization for mobile robots up to the late 90's was widely surveyed by DeSouza and Kak in [119]. After the late 90's, some authors have hardly surveyed this area [120,121]. Most vision-based systems are monocular and binocular (stereo), although those based on trinocular configurations also exist [91].

In literature, different works that make use of monocular cameras can be found [92,176,177]. Blesser et. al. [93] developed an algorithm for a single self-tracking camera operating in a stable natural environment. The system estimates the motion between frames, which is then used to predict which features should be visible to match them. Then, triangulation is used to compute the 3D position of the features and matching is refined by adding epipolar constraints. They demonstrated that this approach successfully localized the camera in small, feature-rich environments but they suggested that the closure of large loops may not be possible as the feature matching has a strong dependence on the motion estimation, which is susceptible to drift. Overall, this approach is remarkably similar to that of Montiel et. al. [94] except that Montiel used simple image patches, represented feature locations by direction and inverse depth and initialized features as semi-infinite lines.

The Shi and Tomasi algorithm [96] is adopted in [97] to extract the position of the image features which are used as landmarks to guide the navigation process. The use of a wide angle camera improved some aspects: camera motion can be better identified, with particular improvements on rotational and translational movements estimation, the range of movements increase, and large motions or motions with great acceleration are better dealt with, since they appear much less abrupt. Pinies et. al. [99] investigated the benefits of using a low cost IMU to aid an implementation of inverse depth monocular SLAM similar to that of Davison [97]. They demonstrated that the inertial observations constrain the uncertainty of the camera pose which improves the accuracy of the estimated trajectory [99]. Large-scale direct monocular SLAM [100] is an algorithm that uses only RGB images from a monocular camera as information about the environment and sequentially builds topological map. The main property is the ability of reconstructing 3D environments and direct nature process. To improve the accuracy of the features, some works rely on a multi-sensor system, like the one presented by Castellanos and Tardos [41]. Their system consists of a 2D laser scanner and a camera, implementing an EFK-SLAM algorithm. Laser scanners and vision sensors, particularly monocular vision, have quite orthogonal and complementary properties. Hence, it is meaningful to use laser data to estimate the location of the features and to use the vision data to make these features more distinguishable in the data association process [179]. Other examples of the use of EKF in visual SLAM can be found in [95,179,180].
In recent years, an increasing use of Time of Flight (TOF) cameras has been seen, particularly in structured environment as described in paragraph 2.8. Due to the fact that these cameras give directly the 3D coordinates of the image pixels, authors believe that they can help to have a better reconstruction of the environment [101]. Another structure that is gaining popularity because of its advantages is the omnidirectional cameras. Omnidirectional cameras have a 360◦ view of the environment and given that the features stay longer in the field of view, it is easier to find and track them [162,163].

However, the use of a monocular system can lead to failure modes due to non-observability, problems with scale propagation and requires extra computation to

provide depth estimates. To avoid these issues, different works employ a stereo pair. Stereo systems are hugely adopted in different environments, for both landmark detection and motion estimation [102,174,175] in indoor [164-169] and outdoor environments [170-172]. The outstanding work of Sim et al [103,104] implemented a Blackwellised particle filter with the stereo vision system. It implemented a hybrid approach consisting of 3D landmarks extraction for localization and occupancy grid construction for safe navigation. Other examples of the adoption of particle filter algorithms with stereo vision system can be found also in [181,182].

Schleicher et al [105], instead, used a top-down Bayesian method to perform a vision-based mapping process consisting in the identification and localization of natural landmarks from images provided by a wide-angle stereo camera. Simultaneously, a self-localization process is performed by tracking known features of the environment. The authors proved that, thanks to the redundancy of the information extracted from both cameras, robustness and accuracy are increased and the processing time decreased. [106] presents two vision-based SLAM approaches that use 3D points as landmarks: one relies on stereovision, where the landmark positions are fully observed from a single position, and a bearing-only approach that exploits monocular sequences.

The relative SLAM system presented in [173] combines a world representation enabling loop closure in real-time with low level image processing adapted to stereo image pairs. They integrated three key components: a representation of the global environment in terms of a continuous sequence of relative locations; a visual processing front-end that tracks features with sub-pixel accuracy, computes temporal and spatial correspondences, and estimates precise local structure from these features; and a method for loop-closure, which is independent of the map geometry. There have been also many successful approaches to the visual SLAM problem using the RGB-D sensor to exploit the 3D point clouds provided [107,178]. RGB-D cameras are sensing systems that capture RGB images along with per-pixel depth information. They rely on either active stereo or time-of-flight sensing to generate depth estimates at a large number of pixels.

Most of the visual SLAM systems apply algorithms used in the Computer Vision community for Structure from Motion (SfM). Techniques such as Bundle Adjustment (BA) are in fact generating a great deal of interest in the robotics community now due to the availability of high performance computer hardware and the sparse representations of these techniques can in fact improve performance over the EKF [112]. SfM based techniques typically maintain the full trajectory of the camera and use optimization to find the best trajectory and landmark locations; in paragraph 2.8 an insight of this technique is given. The first real time application of BA was the visual odometry work of Mouragon et. al. [113], followed by the ground breaking SLAM work of Klein and Murray [114], known as Parallel Tracking and Mapping (PTAM). SfM algorithms have been extended to work on long image sequences [115,116], but these systems are fundamentally offline in nature, analyzing a complete image sequence to produce a reconstruction of the camera trajectory and scene structure observed. To obtain globally consistent estimates over a sequence, local motion estimates from frame-to-frame feature matching are refined by means of bundle adjustment. But it has been proven that it is indeed possible to achieve real-time localization and mapping with a single freely moving camera as the only data source [117]. In ORB-SLAM [118], thanks to the covisibility graph, tracking and mapping is focused in a local covisible area, independent of global map size. It is an example of the use of ORB features. Usually, the scale invariant feature transform (SIFT) method, developed by Lowe [109], stands out among other image feature or relevant points detection techniques and nowadays has become a method commonly used in landmark detection applications [108]. SIFT-based methods extract features that are invariant to image scaling, rotation, and illumination or camera view-point changes. During the robot navigation process, detected invariant features are observed from different points of view, angles, distances and under different illumination conditions and thus become highly appropriated landmarks to be tracked for navigation, global localization [110] and robust vision-based SLAM performance [111].

A novel visual SLAM approach is illustrated in [122]. It is called Conditional Simultaneous Localization and Mapping (C-SLAM): the camera state transition is derived from image data using optical flow constraints and epipolar geometry in the prediction stage. This not only increases prediction accuracy but also replaces commonly used predefined dynamic models which require additional computation.

2.6 UNDERWATER SLAM

The underwater environment is extremely challenging and a relevant part of this thesis regards applications employed in it. This paragraph presents a summary of SLAM algorithms and techniques adopted in this environment.

In comparison to the works applied in indoor or in land mobile robotics, there has been little work done in the field of underwater SLAM, mostly due to the complexity of the environment and the difficulty in conducting field experiments. Unpredictable currents and surges acting upon the vehicle body can produce motion in any direction; the lighting conditions are highly dynamic, adding complexity to the motion tracking algorithms; the visibility decreases with depth and turbidity of the water. These are only some of the difficulties that marine researchers need to face. For these reasons, usually a variety of onboard sensors (IMU, depth sensor, etc.) and perception sensors are used to estimate an accurate location and mapping of the subsea habitat. The most used perception sensors in literature are Mechanical Scanned Imaging Sonars (MSIS), Forward-Looking Sonars (FLS), Side-Scan Sonars (SSS) and Video cameras.

There are different examples of underwater SLAM where sonars are employed [146-150]. Mallios et al. [123,124] implemented a SLAM algorithm for partially structured environments were DVL and MSIS are used. Due to the distorted images collected from the MSIS while the vehicle is moving, a two-stage EKF algorithm is implemented: a first EKF is used to estimate the trajectory and then correct the MSIS images; the second stage uses motion estimation and undistorted images once data association is performed. While in [125] a MSIS is used through a modified FastSLAM 2.0 method.

Authors in [126] propose a system for mine counter measurement and localization with an AUV where a SSS and a FLS are employed, in addition to the use of the GPS for initial localization. A nonlinear least square optimization is performed to the dead-reckoning (DVL and IMU) sensor and sonar images. In [127] an approach of Graph SLAM is proposed using a set of membership to perform SLAM in a non-linear environment from seamarks located through a SSS.

Vision systems reduce space and cost and increase the resolution, although their range dramatically depends on the water conditions. At present, several solutions for AUVs can be found for many undersea critical applications: undersea infrastructures or installations inspection and maintenance, for any of power, gas or telecommunications transport cases, sea life monitoring, military missions, sea bed reconstruction in deep waters, inspection of sunken ancient ships, etc. Vision has become essential for all these applications, either as a main navigation sensor or as a complement of sonar. Consequently, there exists a good motivation to improve AUVs navigation techniques by expanding their autonomy, capabilities and their usefulness.

Estimation of camera motion in underwater unstructured environments is a tough aspect. If objects like pipes or cables, typical artifacts of interest in underwater operations, are present, the estimation of camera motion is eased. It can be based on edge detectors and Hough transform [128, 129,130] or on different texture groups and segmentation of images in regions with similar textural behavior [131]. In [184] a stereo vision system was integrated in the MARIS intervention AUV to detect cylindrical pipes. Pipe edges are tracked using an alpha-beta filter to achieve robustness and return a reliable pose estimation even in case of partial pipe visibility. Foresti and Gentili [132] implemented a robust neural-based system to recognize underwater objects.

Whereas the environment has no defined references, navigation becomes extremely complicated and challenging. In underwater visual SLAM strategies, features must be found in the overlapping images. There are fundamentally three methods that are used

for this purpose: optical flow, feature tracking or gradient methods. Optical flow-based methods and feature tracking can cause failure in algorithms due to scattering effects, bad image quality or lacking illumination under the sea. Gradient methods use scene properties such as depth, range, shapes or color intensity, that are computationally more efficient and more accurate [133].
Pizarro [136] and Eustice [137] developed an information filter based SLAM algorithm that was exactly sparse and operates completely in the information space, thus obtaining linear complexity in the number of landmarks. It was also demonstrated the ability to produce large-scale 3D reconstructions of underwater environments from monocular video and IMU data logged on several ocean surveys. Monocular camera was also employed in [138,152,154,185], while other researchers opted for stereo images systems [139,153,183].

Sometimes, beyond the estimation of the AUV movement through features identification and correlation-based procedures, the grabbed images are combined producing a photo-mosaic of the traveled area [134,151,140]. Marks et al [141,142] developed a technique to implement real-time mosaics using correlation between on-line images and stored images, matching them by edge detection method while [155] used a Fourier based matching. Previous systems often assumed that the seafloor was plane and static, and that the camera was facing it, making the image plane almost parallel to the seafloor plane. Gracias et al [143] proposed a method for mosaicking and localization that did not make any assumption on the camera motion or its relative position to the sea bottom. The system was based on motion computation by matching areas between pairs of consecutive images of a video sequence. Finally, an interesting contribution to underwater mosaicking and positioning was that by Xu and Negahdaripour in [144]. The vehicle position was computed integrating the motion of the camera from consecutive frames using Taylor series of motion equations, including the second order terms, which in previous research was usually ignored. Due to the shortcomings that both sonars and cameras singularly have, is always more frequent to find hybrid systems in literature to enhance the performances [156,157,158]. Mahon

et. al. [145] developed an EKF visual SLAM algorithm for an AUV fusing the information coming from both a scanning sonar and high resolution stereo vision system to identify and track a sparse set of distinguishable environment landmarks. They identified that in underwater environments where water movement may quickly cause large uncertainty in the vehicle location it is necessary to have many landmarks to ensure they are always visible. Despite the difficulty of combining two modalities that operate at different resolutions, technology innovations and advances in acoustic sensors have progressively allowed the generation of good quality data suitable for integration, which is usually performed at a feature level [156].

2.7 CONSIDERATIONS ON SLAM TECHNIQUES

A total number of 186 works published between 1977 and 2017 have been analyzed to build the state of the art on SLAM. Precisely, the examined bibliography comprises 93 articles from Journals, 86 from Conference, 5 PhD Thesis, as represented in Fig. 1.

From a first analysis on which are the techniques and the algorithms most adopted in SLAM problems, it resulted that the majority of the authors implemented the EKF algorithms in their problems (Fig. 2), followed by the Particle Filter. Even the non-filtering technique of Graph SLAM was adopted in a great extent. It can be noticed also that the EKF was adopted continuously from the '80s to nowadays while other types of techniques were emerging (Fig.3). It can be stated that the main advantage of the Extended Kalman filter is its ability to provide the quality of the estimate (i.e., the variance), and its relatively low complexity, but it can be applied only for Gaussian and linear models or with limited nonlinearity. Unscented Kalman Filter is a more reliable estimator than EKF when the system model is highly nonlinear, since it approximates the probability density function instead of the nonlinear function itself, as previously described, but it does not make any improvement in the high computational load of the EKF. For this reason, it has not overcome the use of the EKF. For non-Gaussian and non-

linear models, particle filtering (PF) is the most appropriate approach, since it is able to provide arbitrarily posterior probability distribution.

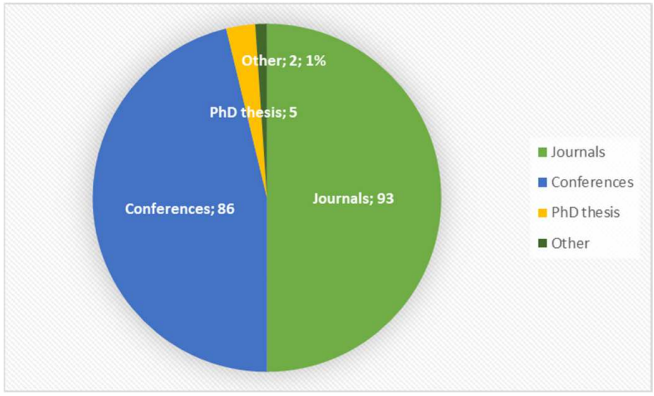

FIGURE 1 EXAMINED BIBLIOGRAPHY

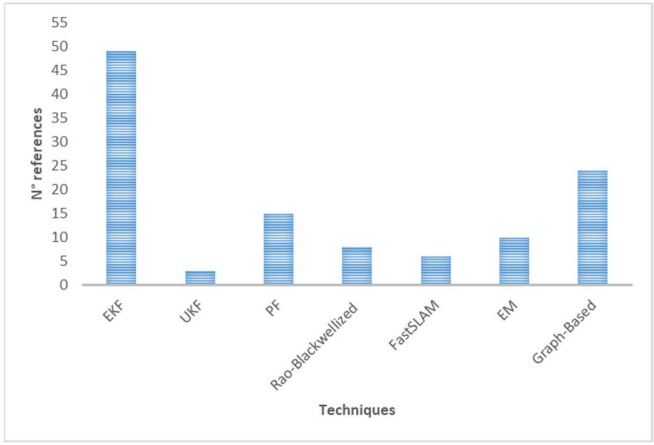

FIGURE 2 MAJOR TECHNIQUES ADOPTED IN SLAM

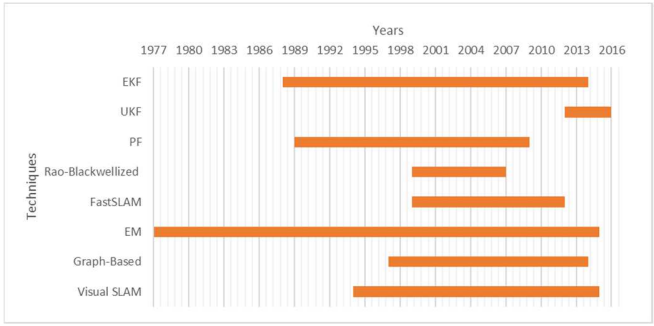

FIGURE 3 TECHNIQUES DURING THE YEARS

The SLAM problems that take into consideration the vision system, the so called Visual SLAM, has been successively analyzed. Within the entire bibliography considered, about 70 articles resulted to use the vision sensor. At first, the analysis on the type of the system adopted was performed. About 50% of the authors employed a single camera, followed by the usage of stereo systems, as illustrated in the graph of Fig.4. Only 7% adopted other types of sensors or composition of the system (i.e. Time of Flight cameras or trinocular systems). Even here the most adopted technique is the EKF, Structure from Motion is also largely used with vision, as it allows to easily detect features in images and so to reconstruct the map of the environment.

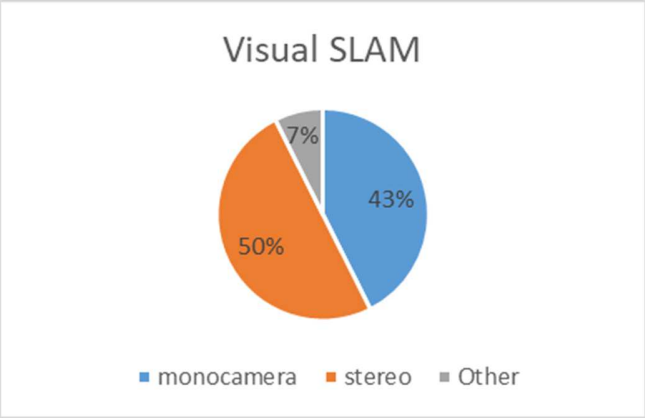

FIGURE 4 VISUAL SLAM SENSOR SYSTEM

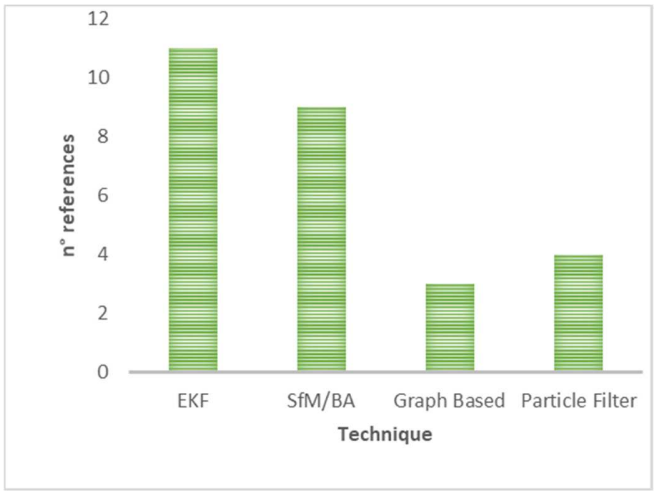

FIGURE 5 VISUAL SLAM TECHNIQUES

The largest number of articles applies SLAM in indoor environments (Fig.6). This is due to errors, their magnitude and the unpredictability that make outdoor navigation so much harder than indoor navigation. It is difficult and time consuming to obtain correct and accurate sensory data in such an environment. In fact, measurement errors cannot be avoided and they are often large and with non-standard distribution. As outlined in paragraph 2.6, a few number of works have been conducted in underwater environment in comparison with those in indoor fields. As results from Fig. 7, in underwater scenarios the most used sensor is the vision, followed by sonar sensors. The latter have several disadvantages:

- they usually only allow limited depth for operation, making applications costly for deep immersions;
- the data observed has limited accuracy, particularly in ocean environments where low-light, strong currents and turbid waters are present;
- the estimated noise ends up causing crucial impact on the tasks of localization and mapping, often leading to a non-convergent system.

Cameras turned out to be the most adopted in literature in the last years. It is required to observe as many as possible distinguishable features in order to reduce the uncertainty caused by vehicle drift, so having distinct landmarks could hugely simplify the data association process. They are low cost but with high and accurate data. It is

also necessary to manage reflections and the dynamic continuous change of underwater natural resources is another factor that makes the recognition of previously visited locations difficult or even impossible. For these reasons, many authors are beginning to adopt a hybrid system composed of both sonar and vision sensors to compensate the disadvantages of employing them singularly.

Regarding the methods implemented in such scenarios, fig. 8 shows that many successful applications are employing an EKF to solve underwater SLAM problems, but it is not the unique. It is easy to implement but the quadratic computational complexity of the EKF makes it difficult to apply to the most diverse type of scenarios. UKF is a more reliable estimator than EKF, as already said before, but it is also slightly slower than the EKF, so it is not frequently adopted. The pose graph formulation can be applicable to many of the forms of underwater SLAM that depend on the sensors available, the operating environment and the autonomous task to be executed. It is faster on large-scale problems because it parameterizes the entire robot trajectory in the information form feature constant-time updates and linear memory requirements. In addition, it takes advantage of re-linearization to better handle nonlinearities. Despite the advantages, this method presents some challenges: firstly, since it operates in the information matrix, it is expensive to recover the joint covariances of the estimated variables; secondly, since it smooths the entire trajectory of the robot, the complexity of the problems grows over time and the performance degrades.

Following this analysis, it results that is not possible to indicate which method is the best to solve SLAM problems in underwater environment because a common benchmark of study is not applicable. The marine environmental conditions in which the robot moves and the equipment of the robots vary situation by situation, based also on the task it has to perform. Nonetheless, the EKF seems to be the most employed in literature for the reasons already explained above.

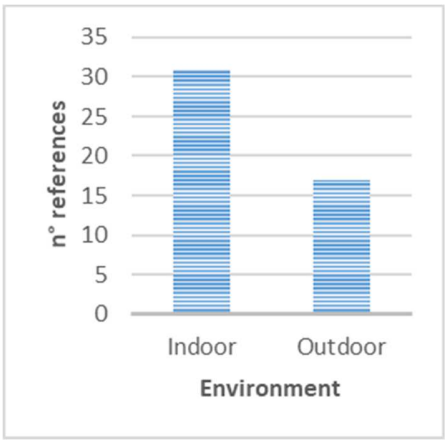

FIGURE 6 INDOOR/OUTDOOR

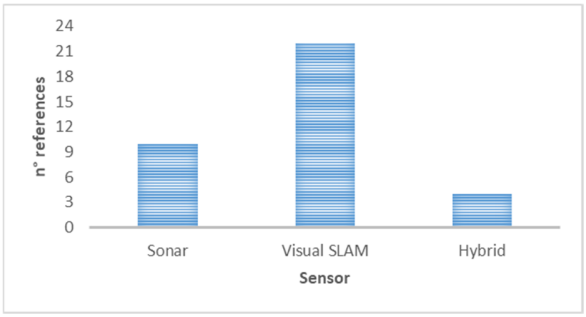

FIGURE 7 UNDERWATER SLAM

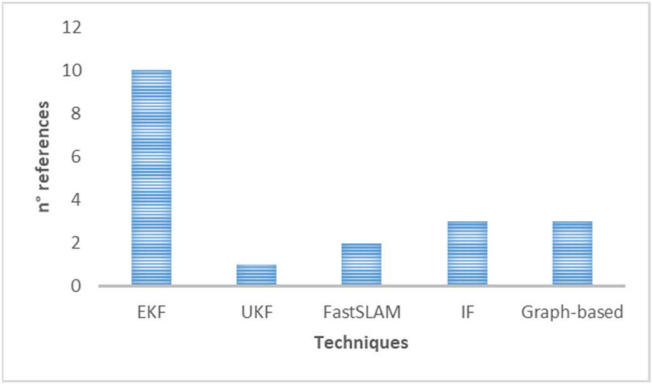

FIGURE 8 TECHNIQUES IN UNDERWATER SLAM

2.8 VISION SYSTEMS AND TECHNIQUES EMPLOYED FOR MAPPING (PARTIALLY) STRUCTURED ENVIRONMENT

The mapping of the working scene and the information extracted at each instant t, t+1, etc., are essential for implementing the SLAM algorithms described in the previous paragraphs, employable in robotic systems with the aim of inspection or documentation.

Considering structured environments, the mapping by means of optical devices has considerably evolved during the recent years. As mentioned before, Structure from Motion has become an emergent methodology in the recent decades, reaching such a state of maturity that some of its algorithms have already climbed to the commercial application level. There are a number of free software options for SfM processing (e.g., Bundler [186], VisualSFM [187,188], and Autodesk ReMake [189]) as well as proprietary software (e.g., Arc3D [190] and Agisoft PhotoScan [191]). It is the line of research that, taking as the only input a set of image correspondences, seeks to infer in a totally automated manner the 3D structure of the scene viewed and the camera locations where the images where captured. The origins of SfM are from the field of computer vision, beginning with Ullman (1979) and evolving into the current iterations of algorithms [186, 192- 195]. The core concept of SfM is the science of photogrammetry, that since the second half of 19th century aimed to extract geometric information from images [196,197]. In traditional photogrammetry, two overlapping images (a stereo pair) are taken with a calibrated (metric) camera(s). By knowing the internal camera geometry, lens distortions, and the distance between photos (parallax), the stereo perspective allows users to calculate the distance from the camera to objects in the photographs using trigonometry. The negative aspect of this type of photogrammetry is that users need a priori knowledge of the exact camera positions, the internal camera geometry, and lens distortions. Stereo photogrammetric datasets also suffer from shadowing, which are areas of missing data from obstructions in the instrument's line of sight.

The research done by the computer vision community has been mostly oriented to achieve the complete automation of the problem and has produced remarkable progress in three aspects: first, the constraints imposed on the motion of the features in two or three images under the assumption of the rigidity of the scene have been formalized, even in the case of degenerate motion and uncalibrated camera [198]; second, intense research on salient feature detection and description with a high degree of invariance has been conducted [199-201]; and third, spurious rejection [202,203]. The result is an automated way of robustly matching point and line features along images and estimate the geometric relations between pairs. SfM estimates the camera poses and the scene structure simultaneously without requiring either to be known a priori, from the information about common points visible in the images. The internal camera geometry and lens distortions, however, should be estimated with simplified models. The multiple views allow for multiple trigonometric measurements that help reducing the error in each 3D point that comes from the simplified camera and lens geometries models. Moreover, by collecting multiple views of all sides of an object or scene the shadowing can be eliminated in the final dataset [204]. Based on these three achievements several methods have been proposed that, from a set of images of a scene, are able to estimate the three-dimensional structure and camera locations up to a projective transformation in the most general case of uncalibrated cameras.
Some of the applications can be found in geomorphology [205-208], volcanology [209,210], forestry and precision agriculture [211,212] and archeology [213].

Independently to the field in which it is applied, conventional SfM systems consist of the following steps:

1. Image features point detection and matching;
2. Relative poses between camera pairs are computed from matched image feature points;
3. all camera poses (including orientations and positions) and scene point coordinates are recovered in a global coordinate system according to these relative poses. If camera intrinsic parameters are unknown, self-calibration

algorithms, e.g. [214], should be applied, to obtain a metric solution up to scale. Some well-known systems, such as [186, 192], compute camera poses in an incremental way, where cameras are added one by one to the global coordinate system. Other successful systems, e.g. [215-217], take a hierarchical approach to gradually merge short sequences or partial reconstructions.

4. Lastly, a global nonlinear optimization algorithm, like bundle adjustment [218], is applied to minimize the reprojection error, which guarantees a maximum likelihood estimation of the result and refines the initial pairwise estimation into a globally consistent one [219].

Fig. 9 shows the typical pipeline of a SfM algorithm.

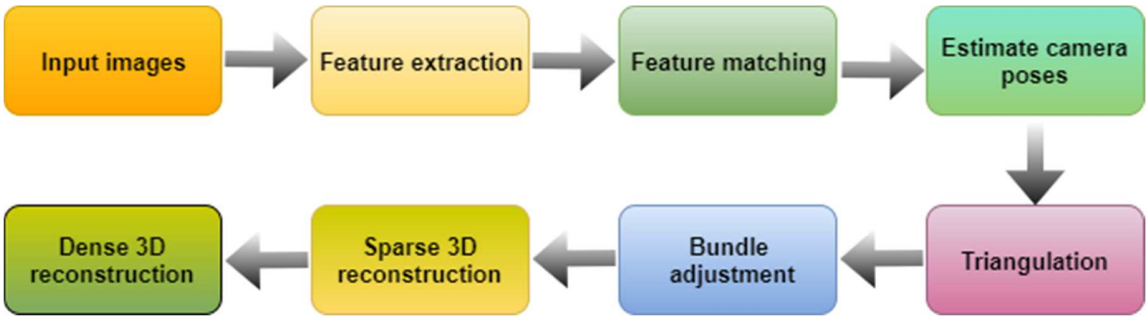

FIGURE 9 SFM PIPELINE

While SfM has dealt with the problem of the estimation of the ego-motion of a mobile platform and its surroundings in its most general form, for any kind of visual input, monocular SLAM has focused in sequential approaches mainly for the processing of video input. This comes because of the specificity of the robotic application, as the sensorial input in a robot comes naturally in the form of a stream. A mobile robot also needs sequential estimation capabilities: at every step the best possible estimation is required in order to insert it in the control loop, and hence batch processing does not make sense. This sequential constraint is not limited to the robotics applications. Augmented reality, for example, also needs a sequential estimation in order to coherently and smoothly insert a virtual character in every frame of a sequence and hence can make use of monocular SLAM algorithms [220].

Although SfM computer vision provides an economic and user-friendly alternative to other 3D scene-capture and modeling tools such as light distance and ranging (LiDAR), the accuracies and resolutions are often not enough satisfactory in industrial applications. In fact, it depends on many factors like the resolution of the adopted camera, the working distance, the light conditions, etc., and so it is too variable. Table 1 shows some experimental results conducted in laboratory with the aim of evaluating the accuracy of the SfM technique in the reconstruction of an object located at about 105 mm distant from the camera. The adopted camera was a Dalsa Genie H1400 (see Table 2 for the characteristics).

TABLE 1

Resolution	1400 x 1024
Total Data Rate	92 MB/s
Max. Frame Rate	75 fps
Pixel Size	7.4 µm
Output Format	GigE Vision
Size	44 mm x 29 mm x 67 mm

TABLE 2

N. photos	**32**	**64**	**135**
Elaboration Time	18.917 min	19.967 min	45.583 min
N. points	4593 pts	8146 pts	12307 pts
GCP transformation - Root of mean squared error	0.070346 (cm)	0.012769 (cm)	0.038887 (cm)
GCP transformation - Mean absolute error	0.061205	0.011775	0.036066

Most mobile robot implementations until today rely on 2D sensors for creating maps, self-localization, and collision avoidance, but multi-view stereo methods [221] frequently fail to properly reconstruct 3D scene geometry. The primary reason for this is the notorious difficulty of finding multi-view correspondence when visible texture is sparse or complex occlusions are present. Although these difficulties could be partially reduced by increasing the set of views or resolution of the images, intrinsic problems still remain [222].

For this reason, in the recent years, roboticists are starting to use 3D vision sensors in their applications. The most popular is the Microsoft Kinect, designed to project a known light pattern into the 3D scene, in the infrared wavelengths for a reconstruction of the illuminated scene. This strategy is only applicable indoors and for a reduced range of a few meters. With the release of the Microsoft Kinect camera, many researchers focused the localization and mapping with hand-held RGB-D cameras. As is always the case when considering image processing, the core issue of the reconstruction of the 3D structure of an object imaged from multiple points of view lies in the quality of the initial images. Not complying with the requirements of image acquisition induces great chances of the failure of the subsequent processing.

In recent years, a new generation of active cameras, based on the Time-of-Flight principle (ToF), has been developed. Being a recent development in imaging hardware, the ToF technology opens new possibilities. Unlike other 3D systems, the ToF-sensor is a very compact device, which already fulfils most of the features desired for real-time distance acquisition. Compared to stereo vision, ToF cameras do not suffer from missing texture in the scene or bad lighting conditions with less computational expensiveness. These devices are usually characterized by no more than a few thousands of tens of pixels, a maximum unambiguous measurement range up to thirty meters and small dimensions. The advantages of ToF cameras over other 3D measurement techniques are the possibility to acquire data at high frame rates, to obtain 3D point clouds without scanning and from just one point of view and the compactness of the sensor. A 3D time-of-flight camera operates by illuminating the

scene with a modulated light source, and observing the reflected light [223]. The reflected signal is received by a combined CCD/CMOS chip. Phase shift between the illumination and the reflection is measured and translated to distance. TOF systems use either pulsed-modulation or continuous wave modulation:

- Pulsed-modulation (Fig. 10): the distance is obtained by measuring the time interval between the transmitted and received light pulses.
    - *Advantages*: direct measurement of time-of-flight; high-energy light pulses limit influence of background illumination; illumination and observation directions are collinear.
    - *Disadvantages*: high-accuracy time measurement is required; measurement of light pulse return is inexact, due to light scattering; difficulty to generate short light pulses with fast rise and fall times; usable light sources (e.g. lasers) suffer low repetition rates for pulses
- Continuous wave modulation (Fig. 11): measures the phase difference between the sent and received signals. Different shapes of signals are possible, e.g., sinusoidal, square waves, etc.
    - *Advantages*: there is a variety of light sources available as no short/strong pulses required; applicable to different modulation techniques (other than frequency); simultaneous range and amplitude images.
    - *Disadvantages*: the integration over time is required to reduce noise and consequently the frame rates are limited; motion blur caused by long integration time.

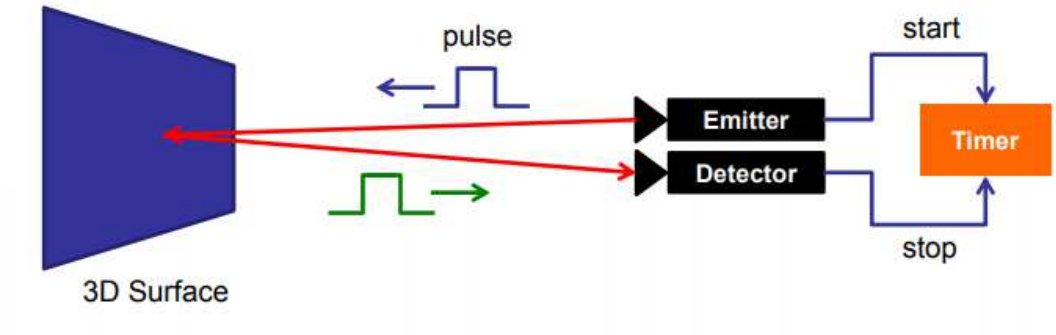

FIGURE 10 PULSED-MODULATION

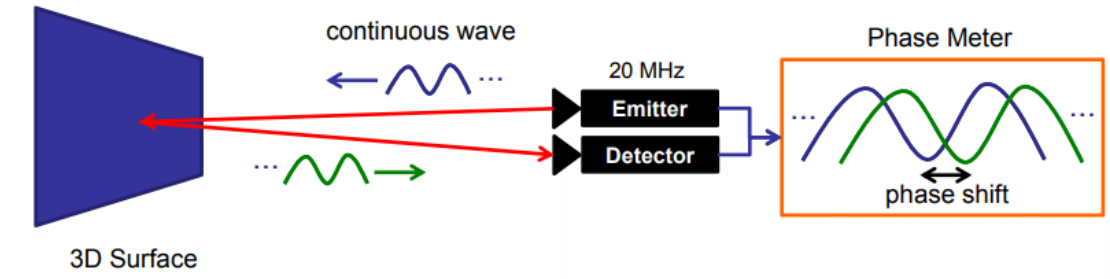

FIGURE 11 CONTINUOUS WAVE MODULATION

ToF range cameras can provide 2D image and 3D data, so it is a better way that the combination image with 3D data is used for solving SLAM problem. ToF cameras usually deliver a range image and an amplitude image: the range image (or depth image) contains for each pixel the radial measured distance between the considered pixel and its projection on the observed object, while the amplitude image contains for each pixel the strength of the reflected signal by the object. In some cases, an intensity image is also delivered, which represents the mean of the total light incident on the sensor (reflected modulated signal and background light of the observed scene) [224].

ToF cameras have been used to sense relatively large depth values for mapping or obstacle avoidance in mobile robotics, and also for human detection and interaction. Weingarten et. al. [225] used a CSEM ToF camera prototype for basic obstacle avoidance and local path planning. They evaluated and compared the results to a trajectory from 2D laser range-finders. Their experiments showed that path planning

and obstacle avoidance based on the ToF camera data could prevent the robot from colliding with an obstacle that was not detected by the 2D laser range finder [226]. At closer distances, ToF cameras have been applied to object modelling [227,228], precise surface reconstruction [229] and to grasp known [230] and unknown [231,232] objects. As shown in many previous works [233-239], the distance measurements of ToF cameras are influenced by some systematic errors. In order to model these errors, previous ToF camera calibration approaches performed a linear mapping, adjusting range measurement errors using look-up-tables [235] or splines [236]: in both cases, they determined the distance errors using high precision measurement racks. In [224] authors attempted to obtain a distance error model of a ToF camera with a low-cost custom-made system.

The use of ToF cameras have several advantages in respect of the stereo systems: only one camera is required, the depth is not necessary to be computed manually, the acquisition of 3D scene geometry is performed in real-time, it has a reduced dependence on scene illumination and has almost no dependence on surface texturing. But, despite the fact that ToF sensors can provide dense depth maps even where stereo setups typically fail, they have two main challenges:

1. the resolution of ToF depth maps is far below the resolution of stereo depth maps from color images, and
2. measurements are greatly corrupted by non-trivial systematic measurement bias and random noise [222].

In fact, reconstruction is challenging when dealing with complex 3D shapes, especially if fine details are of interest or surfaces have specular reflection characteristics. For the localization of such objects sensors need to be selected with respect to the desired working range or the physical principle. In many cases, a combination of different sensors is beneficial. Sensor fusion can help to reconstruct the environment. The challenge of multi-sensor data fusion and mapping lies in the variety of measurement characteristics. Differences in resolution, frame rate, range, accuracy or sensor noise makes need for specific mathematical sensor models [240]. With the fusion of a stereo

camera and a ToF camera drawbacks of both principles can be compensated. In [241], a PMD stereo combination has been introduced to exploit the complementarity of both sensors. In [242], it has been shown that ToF-stereo combination can significantly speed up the stereo algorithm and can help to manage texture-less regions. The approach in [243] fuses stereo and ToF estimates of very different resolutions to estimate local surface patches including surface normals.

Most fusion systems differ mainly in how the data is merged, once it has been brought into the same reference frame. In [244] authors implemented the fusion step through a combination of the following:

- the ToF depth and the output of a stereo algorithm are computed individually and then fused;
- the ToF data is used as an initial guess and/or to reduce the search space for subsequent stereo refinements;
- the depth reconstruction algorithm uses both stereo and ToF costs as data terms.

The fusion methods can be grouped based on the optimization strategy that is employed. *Local* methods [245 - 252] tend to be faster and parallelizable but cannot cope with locally erroneous data. They are often based on a line search that is guided by the ToF data. The basic optimization employed only takes into account a local sets of pixel values. *Global* methods [222, 253 - 261] add the ToF information as an additional data term in a global energy functional that is then jointly optimized. While the depth maps obtained are smoother due to the usage of prior information, this is at the cost of additional computational resources. The global techniques can be further grouped depending the framework that was chosen for optimization:

- *Graphical* models: Here the problem of correspondence estimation is treated as a labeling problem, where each discrete label corresponds to a disparity value. The energy is interpreted as the negative logarithm of a joint probability distribution defined on a graph, where each node corresponds to an observed

(data term) or latent (depth) random variables the probabilities are defined on cliques of these graphs. Examples can be found in [254-257,261].

- *Variational* fusion approaches [251,258] consider both a continuous image domain and continuous variables (functions), indicated by a dependency on the image coordinates.
- The last sub-group of the global methods [222,253,259,260] contains those which use other non-local optimization strategies. Kim et al. [222] propose a volumetric approach: the initial surface is given by the ToF depth, that is further refined by optimizing an energy function including ToF, stereo, silhouette terms and a Laplacian prior. Based on locally consistent stereo [262] the technique adopted in [260] uses a segmentation of the RGB image to guide a bilateral filter for ToF data upsampling. The algorithm takes two depth hypotheses from a stereo algorithm (semi global matching) and ToF respectively, which are calculated independently. Each pixel then propagates both depth hypothesis independently to surrounding pixel based on color similarity, spatial proximity and photo consistency. Every pixel then has a number of `votes' casted by neighbouring pixels. From these hypotheses, the one with the highest plausibility is finally chosen. In Gandhi et al. [259] the basic idea is to combine the reprojection and interpolation step of the ToF depth map on the reference frame with a stereo matching procedure.

Sometimes a combination of local and global methods is preferred. Nair et al. used a local fusion, based on stereo block matching and subsequently a variational fusion based on total variation to increase smoothness of data [251].

A suitable application in industrial environments for sorting of parts with focus on high flexibility in pick-and-place tasks is bin picking: via a sensor, mounted on the robot itself or from an external view point (see Figure 12), data from the environment is processed by a computer unit in order to classify objects and determine their pose. With known size, position and orientation grasping and manipulation of objects can be performed. In [240] they present a representation for 3D multisensory data fusion with focus on

object localization. Multiple measurements are fused in a generic truncated signed distance representation, from which smooth point clouds with minimal noise can be extracted. In [263] and [264] a Tof and a Microsoft Kinect, respectively, are adopted for bin picking problem.

In [222] authors present a novel approach for the fusion of multiple ToF sensors with stereo yielding 3D reconstructions superior to the ones obtainable with individual sensing modalities alone. They propose an integrated multi-view method that:

- utilizes the multi-view setup to compensate complex systematic measurement bias,
- fuses multi-view ToF measurements into a single complete initial geometry estimate that drastically reduces the random noise by incorporating the directional noise characteristics of ToF sensors,
- integrates the ToF sensor noise characteristics and stereo cues via probabilistic framework that refines the initial geometry estimate and achieves the accuracy of stereo when reliable constraints exist.

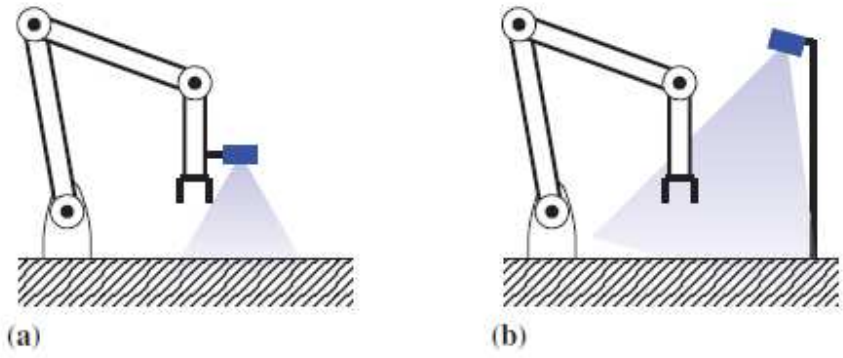

FIGURE 12 BIN PICKING APPLICATIONS

The statistics of the natural environment are such that a higher resolution is required for color than for depth information. Therefore, another typical combination present in literature is a union of 2D RGB with a ToF-sensor. It can be used a binocular

combination of a ToF camera with one [265-269] or with several conventional cameras [270], thereby enhancing the low resolution ToF data with high resolution color information. Such fixed camera combinations enable the computation of the rigid 3D transformation between the optical centers of both cameras (external calibration) as well as the intrinsic camera parameters of each camera. By utilizing this transformation, the 3D points provided by the ToF camera are co-registered with the 2D image, thus color information can be assigned to each 3D point.
In all cases, an external calibration of the two sensors is required. In many approaches, a rather simple data fusion is used by mapping the ToF pixel as 3D point onto the 2D image plane resulting in a single-color value per ToF pixel [266,267]. A more sophisticated approach presented in [268] projects the portion of the RGB image corresponding to a representative 3D ToF pixel geometry, e.g. a quad, using texture mapping techniques. Furthermore, occlusion arteacts in the near range of the binocular camera rig are detected.

2.9 FRAMEWORK DESIGN

Typically, vision applications are challenging because each application is often unique. In developing the most efficient solution for a given task, in fact, there is the necessity of answering to some initial questions that can provide valuable information in the selection of appropriate system components. However, there are some expected functions from a vision machine that allow to design a general scheme which can then be adapted to various situations. For example, the exploitation and imposition of the environmental constraint of a scene, the capturing of the images, the analysis of those captured images, the recognition of certain features within each image and the initiation of subsequent actions are general operations that must be present in the framework; the specific hardware or computer vision techniques in each component may change in specific applications.

The framework studied and designed in this dissertation is shown in Figure 13. It includes systems and sub-systems, which depend on the type of applications and required tasks.

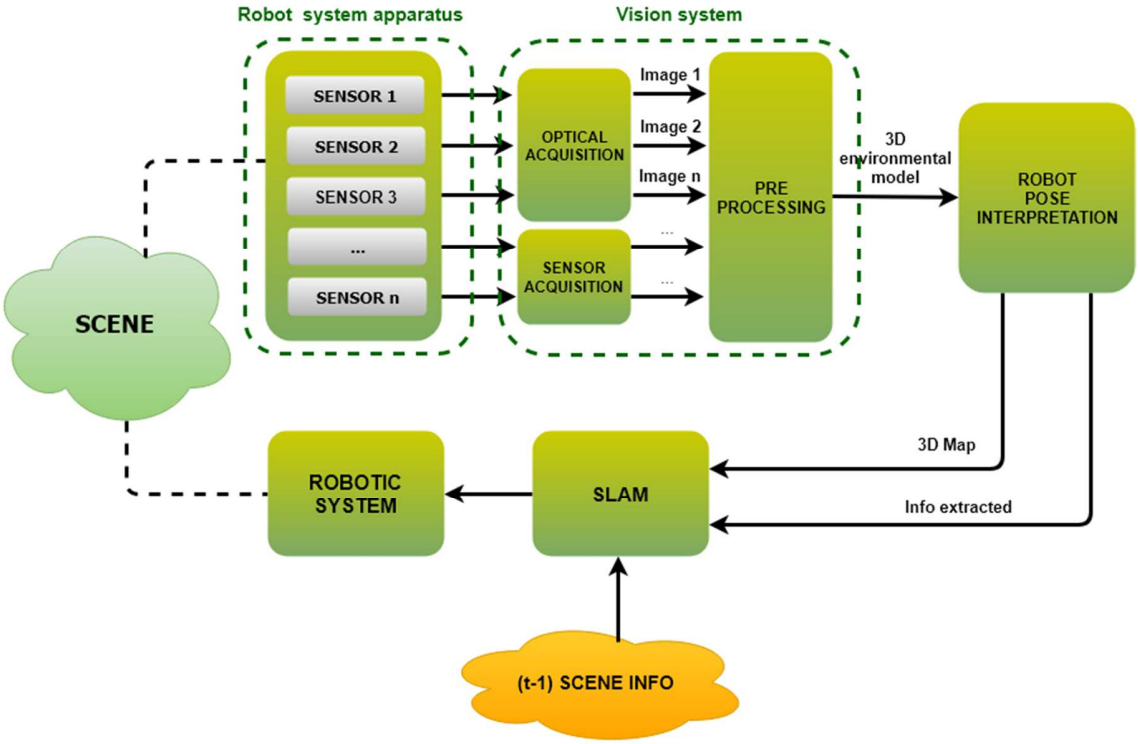

FIGURE 13 FRAMEWORK SCHEME

Scene constraint is the first consideration for the entire system. In fact, the scene in which the robot is moving could vary in each application (indoor, outdoor, etc.), but the situation must be recognized by the machine vision designer and the robotic system apparatus needs to be chosen accordingly to the scene specific characteristics,.

The robotic system apparatus can be composed of a variable number of sensors: optical, proprioceptive or exteroceptive. They acquire information from the scene in order to be manipulated and transformed by the next blocks in the most suitable format for the application. Image acquisition is an important process for the performance of the entire framework, because a high-quality image would surely ease the following processing

and analysis. Both hardware and software are involved in this process and the selection of proper components is crucial to image acquisition process.

The Pre-Processing block includes computer vision techniques to elaborate the single input images, that could be 2D or 3D depending on the sensors chosen in the Robotic System Apparatus, in order to obtain a unified 3D environmental model. The next block duty is to enrich the 3D model with the robot position information coming from the sensors, in order to recover a complete map of the environment. It is then applied in SLAM algorithms that, exploiting also the (t-1) scene data, allows the navigation of the robot in the scene.
This research focused mainly on the design and development of the upper side subsystems of Figure 13: the process that goes from the scene capturing to the output of the Robot Pose Interpretation.

Chapter 3 FIRST CASE STUDY: INSPECTION AND OBSTACLE DETECTION IN INDUSTRIAL ENVIRONMENT

In the recent years, there has been a lot of progress in the field of computer vision applied to industrial robotics. The study of vision systems for industrial robots has been incentivized by various factors like, for example, the implementation of algorithms for image elaboration, the availability of valuable and optimized three-dimensional data , the decreasing of the cost of high performance hardware and the availability of low cost but accurate vision sensors.

Industrial robotics systems lacking a vision system do not have the perception of the environment and so they need a structured and well-defined context to correctly operate. This constraint influences the entire production chain because it is necessary the study of ad hoc solutions to manipulate the products between a working phase and the next one.

In the field of computer vision, the object recognition consists of localizing the current object in the scene, that could be bi-dimensional or three-dimensional. The human being is able to recognize big quantities of objects in few seconds, also with different scale or point of view; even when the object is partially hidden, humans have no difficulty. The same job performed by a machine becomes complex and computationally expensive, especially when the scene in which the objects have to be recognized is three-dimensional. Today, a robust, accurate and speed algorithm for object recognition in computer vision is fundamental for many applications, like augmented reality, or for the detection of objects related to robotics manipulation, like in this work.

In the present activity, an algorithm for object detection in three-dimensional industrial environment has been implemented. The specific field of application will not be mentioned because a Non-Disclosure Agreement between the committee and Loccioni Group is active. For the same reason, there will not be shown images depicting the real object to inspect, but just the initial environment employed for preliminary testing.

The vision system, which consists of Time of Flight cameras, performs the recognition taking advantage of the PCL (Point Cloud Library) 1.7.2 library [1]. The Point Cloud Library is a large-scale, open project [185] for 2D/3D image and point cloud processing that contains numerous state-of-the art algorithms. The developed algorithm compares two point clouds, one considered as reference, and it is based on the Iterative Closest Point algorithm.
In the present chapter, after an introduction on Loccioni Group, the company in partnership during the three years of activities, the problem is then defined in paragraph 3.2 and the results are analyzed in paragraph 3.3.

3.1 LOCCIONI GROUP

The Loccioni Group was established in 1968 by Enrico Loccioni with the aim of creating in his territory – and delivering to the world – an entrepreneurial model for the work and knowledge development. By integrating ideas, people and technologies, the company develops, measures and tests solutions to improve the quality of products and processes for the manufacturing and service industry. Each project is customized on the customers' requirements integrating internal and external competences and technologies and building with customers and partners long-term relations for mutual development. The commitment is measuring for improving, thus helping all those who realize products or offer services to do it in the best way, saving time, money and respecting the environment.

[1] http://pointclouds.org/

The company creates very high-level networks where researchers' communities have the common goal of improving the life quality through the technologies development and integration. The main sectors on which Loccioni Group operates are:

- Energy: Integrated energy efficiency solutions, solutions for energy production from renewable sources and green IT.
- Environment: Integrated solutions for environmental monitoring.
- Industry: Measure, assembly and quality control for industrial processes, products and buildings.
- Humancare: Automating and quality control solutions for health care.
- Mobility: Assembly, testing and quality control systems for automotive components.
- Train & Transport: Integrated solutions for transport and railways network.
- Aerospace: Measure, automation and quality control solutions for aeronautic and aerospace processes, systems and components

The unit involved in the present PhD is the Loccioni *Research for Innovation*. The team is dedicated to the development of long-term research projects, following transversal strategic trends, and a network of Research for Development Laboratories, in which innovation becomes applied research developed according to the specific markets' and clients' requirements. To be always on the frontiers of innovation, to be competitive in terms of value proposition, to respond to constantly changing challengers, the main goal of Loccioni Research activity is transforming data into value.

3.2 PROBLEM DEFINITION

The objective of the present application is to detect and notify the presence of small obstacles, of the order of maximum 5 cm, that could be present in the scene of work before a robot starts moving and executing the planned tasks. This job relies on a 3D vision system positioned at the base of the robot, at about 80 cm distant from the area of inspection.

3.2.1 THE VISION SYSTEM

Three IFM O3D303 ToF cameras have been chosen for this application. Table 3 shows the specifications of this sensor. The relative position between the cameras was defined in order to maximize the capturing area. The three cameras are identified in the following as "left camera", "right camera" and "front camera". The final system will not have overlapping areas between the cameras, but in the tests executed and described in this work there will be some parts overlapping.

TABLE 3

Camera Model	**IFM O3D303**
Technology	ToF
Image size	176 x 132 pixels
Frame rate	
Lens	
Range	300...8000 mm
Field of view	
Focus	
Illumination	
Images	
Interface	Digital I/O; Ethernet

The first step executed was the calibration of the entire system. Because of the specific position of the cameras, a tailored calibration panel was designed, defining a particular pattern that allows to easily match the known points in the panel with the ones in the images. The panel has a size of an A0 (841 × 1189 mm) and its pattern scheme is shown in Figure 14, where there are also drawn in red the areas captured from each camera. Figure 15, instead, represents the real images captured from the cameras for the calibration. Each "X" of the pattern is equally distant from its neighbors, so all the patterns have a known position in the panel.

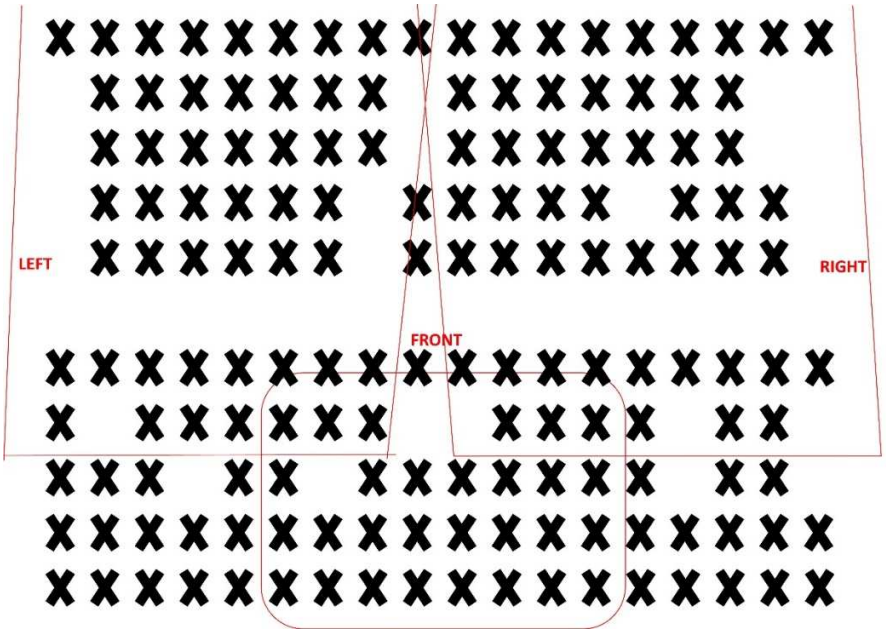

FIGURE 14 CALIBRATION PANEL

FIGURE 15 CALIBRATION PANEL CAPTURED FROM LEFT, FRONT AND RIGHT CAMERAS RESPECTIVELY

The calibration software has been developed in LabVIEW, as almost the rest of the application. The main screen (Figure 16) consists of three steps:

1. First of all, it is necessary to select the path of the image to load and the path where saving the file with Roto-translation matrix of each camera. The user

needs to select which camera (Left, Front or Right) is working with before executing to the next step.

2. The user has then to pick the Selection Tool or the Point tool from the menu box "Current ROI tool" and then he can start selecting the points in the image. As the "X" are equally distant from each other, to have a better calibration the user should select the centers of the "X" in the image. The user can select one point each time and press the *Save Point* button to memorize it. It is then necessary to insert in *Data set REF* the known coordinates related to the points selected in the image, otherwise the calibration is not possible.
3. When the user considers sufficient the amount of points chosen (at least 10 points), he should press *SAVE ALL and CALCULATE R T* button. At this moment, all the points are saved in Global variables, in order to be available to other .vi, and the Roto-translation matrices, one for each camera, are saved in the respective files.

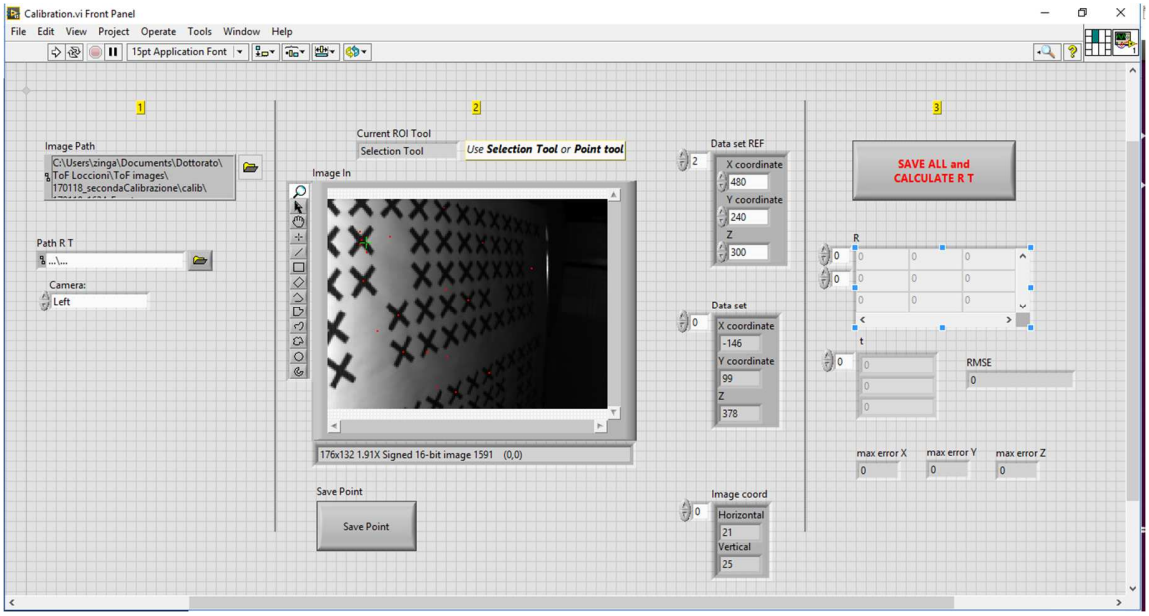

FIGURE 16 CALIBRATION SOFTWARE MAIN SCREEN

The algorithm implemented for the 3D correspondence, so for obtaining the Rotation and Translation matrices, is the one described by Arun [271].

3.2.2 POINT CLOUD ACQUISITION

The algorithm is developed using the PCL 1.7.2 library and LabVIEW. The PCL functions are written in C++, so in order to be used in LabVIEW the code that uses the PCL functions has been written in C++ and then imported in LabVIEW as a DLL.

The input data to start a comprehension of the scene are essentially two:

- the model, that is given from the CAD of the item that needs to be recognized,
- the scene, that is given from the point cloud obtained from the cameras.

After the calibration is performed, the three cameras are on the same reference system. When the three cameras acquire a frame, each roto-translation matrix is applied to the point cloud obtained, in order to get a calibrated one trough

$$P' = RP + T$$

where P' is the new point coordinates, R a 3x3 rotation matrix, P the input coordinates and T a 3x1 translation vector.
Through the calibration, the roto-translation matrices of the three cameras obtained are:

$$R_{front} = \begin{bmatrix} 0.998856 & 0.012797 & 0.046085 \\ -0.011932 & 0.999748 & -0.019007 \\ -0.046317 & 0.018435 & 0.998757 \end{bmatrix}$$

$$t_{front} = [476.762491 \quad 568.411944 \quad 219.981145]$$

RMSE: 8.174187 mm

$$R_{left} = \begin{bmatrix} -0.994309 & -0.023044 & 0.104010 \\ -0.030963 & -0.871666 & -0.489122 \\ 0.101933 & -0.489559 & 0.865992 \end{bmatrix}$$

$$t_{left} = [97.158399 \quad 458.188907 \quad 19.274511]$$

RMSE: 18.468957 mm

$$R_{right} = \begin{bmatrix} -0.994045 & 0.071434 & -0.082285 \\ -0.028111 & -0.897697 & -0.439716 \\ -0.105278 & -0.434785 & 0.894359 \end{bmatrix}$$

$$t_{right} = [891.616812 \quad 454.404261 \quad -14.833234\ ]$$

RMSE: 7.741729 mm

Each point cloud is then exposed to a double filtering. Firstly, a filter on the z coordinate of each point of the point cloud is applied. A minimum and a maximum threshold value is defined and if the z value is less or higher than the thresholds, the point is removed. Secondly, a Radius Outlier Removal filter is applied to clean a little more the cloud. It filters points in a cloud based on the number of neighbors they have. It Iterates through the entire input once, and for each point, retrieves the number of neighbors within a certain radius. The user specifies a number of neighbors, which every indice must have within a specified radius to remain in the point cloud.

When all the point clouds are calibrated and cleaned, they are unified in order to obtain a unique point cloud that will be then elaborated to detect the presence of obstacles in the scene.

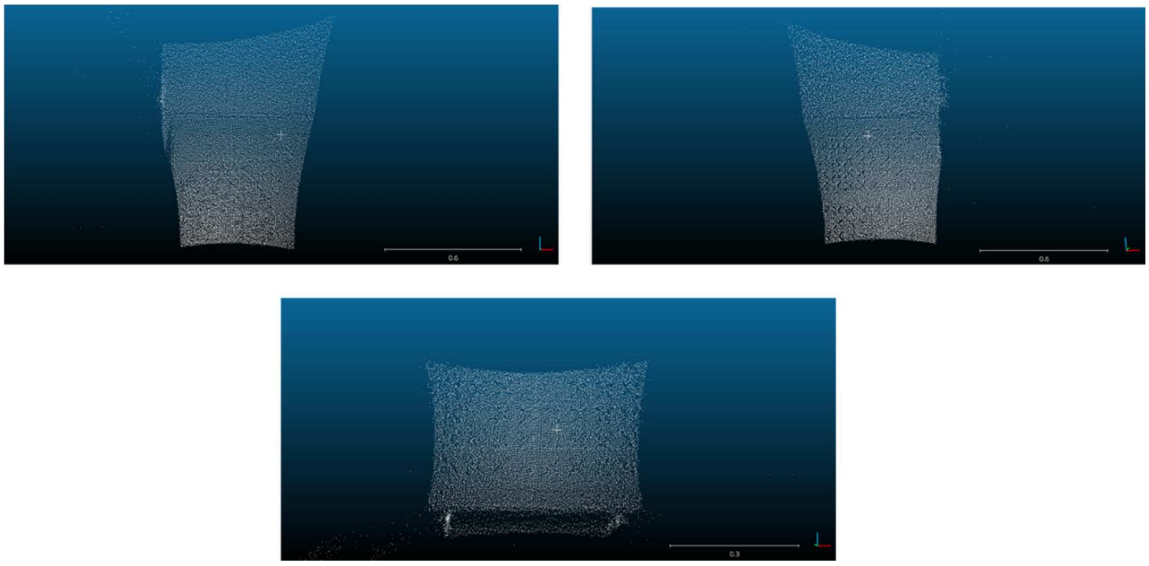

FIGURE 17 POINT CLOUD CALIBRATION

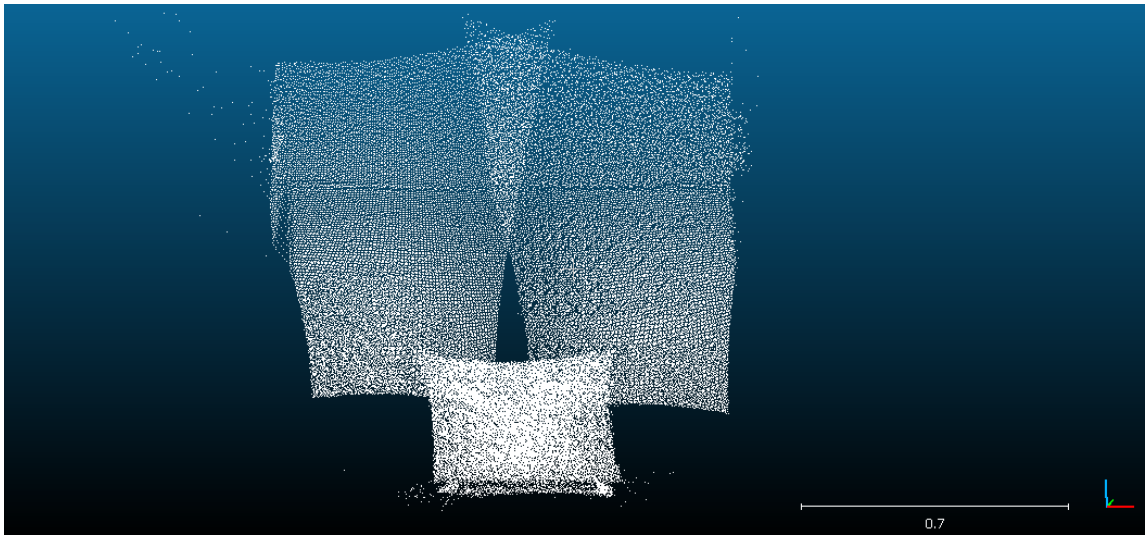

FIGURE 18 POINT CLOUD UNIFIED

3.2.3 POINT CLOUD REGISTRATION

Given two sets of points in different coordinate systems, or equivalently in the same coordinate system with different poses, the goal is to find the transformation that best aligns one of the point-sets to the other. The process that aligns two point clouds is often

called registration. Registration algorithms are able to estimate the ego-motion of the robot by calculating the transformation that optimally maps two point clouds, each of which is subject to camera noise.

The registration algorithms can be classified into rigid and non-rigid approaches. A rigid transformation is defined as a transformation that does not change the distance between any two points. A non-rigid registration yields a non-rigid transformation which maps one point set to the other. In the context of point set registration, non-rigid registration typically involves nonlinear transformation; these methods are able to cope with articulated objects or soft bodies that change shape over time.

Registration algorithms are used in different fields and applications, such as 3D object scanning, 3D mapping, 3D localization and ego-motion estimation, human body detection. Most of these state-of-the-art applications employ either a simple Singular Value Decomposition (SVD) [272] or Principal Component Analysis (PCA) based registration, or use a more advance iterative scheme based on the Iterative Closest Point (ICP) algorithm [273]. Among all the numerous registration methods proposed in literature, the Iterative Closest Point (ICP) algorithm [273-275], introduced in the early 1990s, is the most well-known algorithm. Its concept is simple and intuitive: ICP iteratively estimates the transformation between two point clouds, the *model point set* and the *scene point set*. In every iteration, the point correspondences between model and scene are determined by a nearest neighbor search and the transformation between the point correspondences is estimated by a least squares minimization. The mean squared error of the estimated transformation applied to the scene is determined in every iteration. The algorithm iterates until the error converges or a maximum number of iterations is reached. Given an initial transformation (rotation and translation), it alternates between building closest-point correspondences under the current transformation and estimating the transformation with these correspondences, until convergence. Appealingly, point-to-point ICP is able to work directly on the raw point-sets, regardless of their intrinsic properties (such as distribution, density and noise level). Due to its conceptual simplicity, high usability and good performance in

practice, ICP and its variants are very popular and have been successfully applied in numerous real world tasks [276,277]. The application of the ICP to ToF camera data has also been studied [278].

Recently, many variants on the original ICP approach have been proposed, the most important of which are non-linear ICP [279], generalized ICP [280], and non-rigid ICP [281]. The choice for one of these algorithms generally depends on several important characteristics such as accuracy, computational complexity, and convergence rate, each of which depends on the application of interest. Moreover, the characteristics of most registration algorithms heavily depend on the data used, and thus on the environment itself.

However, ICP is also known for its susceptibility to the problem of local minima, due to the non-convexity of the problem as well as the local iterative procedure it adopts [282]. Being an iterative method, it requires a good initialization, without which the algorithm may easily become trapped in a local minimum. If this occurs, the solution may be far from the true (optimal) solution, resulting in erroneous estimation. More critically, there is no reliable way to tell whether or not it is trapped in a local minimum.

To deal with the issue of local minima, previous efforts have been devoted to widening the basin of convergence [283,284] performing heuristic and nondeterministic global search [285,286] and utilizing other methods for coarse initial alignment [287,288] etc. However, global optimality cannot be guaranteed with these approaches. Furthermore, some methods, such as those based on feature matching, are not always reliable or even applicable when the point-sets are not sampled densely from smooth surfaces.

As mentioned above, the final aim of the present work is to detect the presence of obstacles in the area where the robot has to perform some planned actions. To reach this objective, it was decided to use the ICP as registration algorithm. In fact, there was the necessity to align two point clouds: the one obtained from the three ToF cameras, after the three single point clouds unified in order to have one point cloud at last, and the CAD model of the area to inspect.

The main screen of the software is shown in Figure 19.

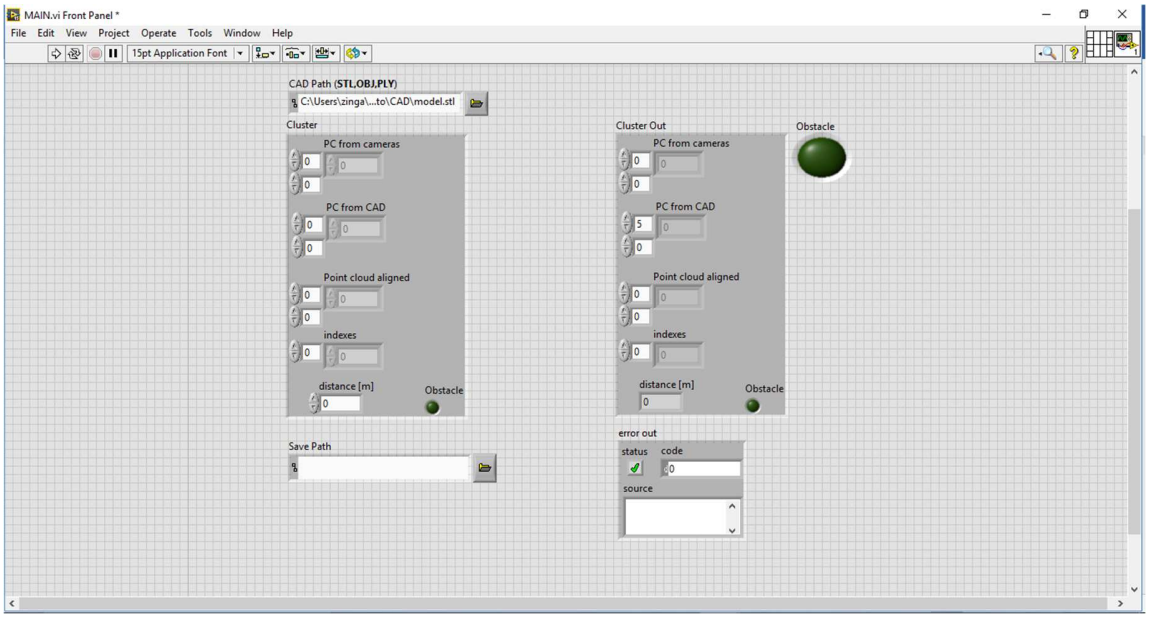

FIGURE 19 MAIN SCREEN

First of all, the CAD model, that must be in .stl, .obj or .ply format, is acquired and converted into a point cloud manageable by the PCL functions. Then the two point clouds are given as input of the ICP algorithm.

PCL contains a set of powerful algorithms that allow the estimation of multiple sets of correspondences, as well as methods for rejecting bad correspondences, and estimating transformations in a robust manner from them.

The output of the problem of registering a pair of point cloud datasets together as pairwise registration, is usually a rigid transformation matrix (4x4) representing the rotation and translation that would have to be applied on one of the datasets (let's call it source) in order for it to be perfectly aligned with the other dataset (let's call it target, or model).

The steps performed in a pairwise registration are shown in the diagram below (Figure 20), which represents a single iteration of the algorithm.

The computational steps for two datasets are straightforward:

- from a set of points, identify interest points (i.e., keypoints) that best represent the scene in both datasets;
- at each keypoint, compute a feature descriptor;
- from the set of feature descriptors together with their XYZ positions in the two datasets, estimate a set of correspondences, based on the similarities between features and positions;
- given that the data is assumed to be noisy, not all correspondences are valid, so reject those bad correspondences that contribute negatively to the registration process;
- from the remaining set of good correspondences, estimate a motion transformation.

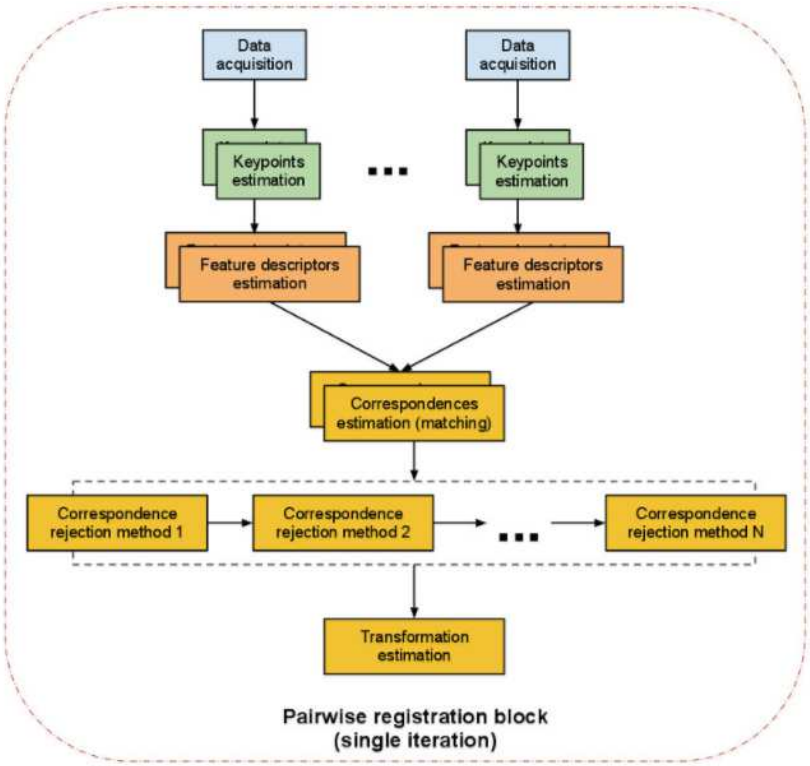

FIGURE 20 POINT CLOUD LIBRARY REGISTRATION API

The transformation is estimated based on Singular Value Decomposition (SVD). The algorithm has several termination criteria:

1. the number of iterations has reached the maximum user-imposed number of iterations;
2. The epsilon (difference) between the previous transformation and the current estimated transformation is smaller than an user imposed value (TransformationEpsilon)
3. The sum of Euclidean squared errors is smaller than a user defined threshold (EuclideanFitnessEpsilon).

The method that implements the ICP algorithm has many input parameters. The essential ones are:

- Number of columns of source cloud
- Number of rows of source cloud
- X,Y,Z values of source cloud
- Number of columns of target cloud
- Number of rows of target cloud
- X,Y,Z values of target cloud
- Max number of iterations
- Maximum distance threshold between two correspondent points in source and target. If the distance is larger than this threshold, the points will be ignored in the alignment process.
- Euclidean Fitness Epsilon
- Transformation Epsilon

The important output parameters are:

- A Boolean to indicate if the algorithm converged or not
- the Euclidean fitness score (the sum of squared distances from the source to the target) from two sets of correspondence distances (distances between source and target points)
- the final transformation matrix estimated by the registration method
- the mean Euclidean distance between the aligned and target points

- the vector with the correspondences between input and target cloud.

In the following figures, some results of the alignment are shown. Different input parameters, in particular regarding the number of iterations, the Euclidean Fitness Epsilon and the Transformation Epsilon, have been tested in order to find the best values combination. Figure 21 presents the initial position of the source and the target point clouds, while Figure 22 illustrates the point cloud obtained after the registration algorithm, with the approximated distances histogram.

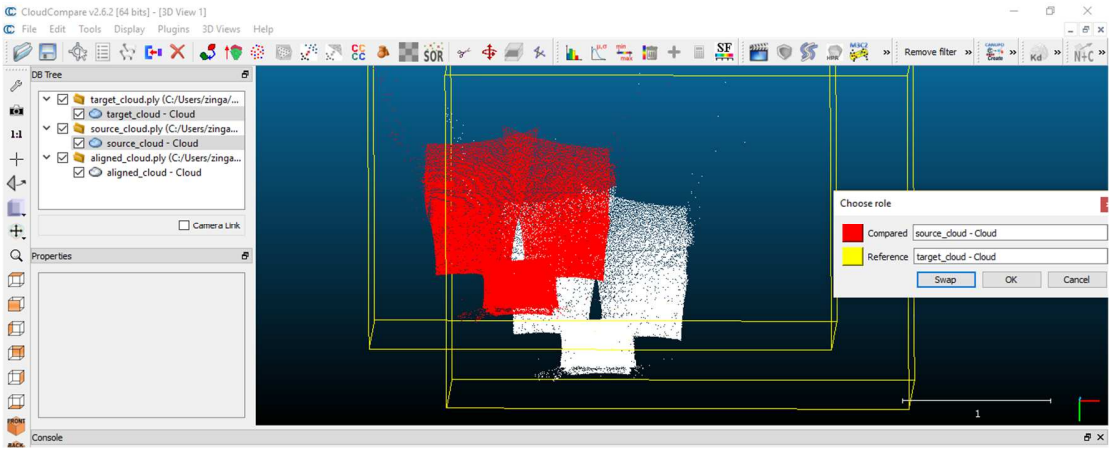

FIGURE 21 TARGET AND SOURCE POINT CLOUDS

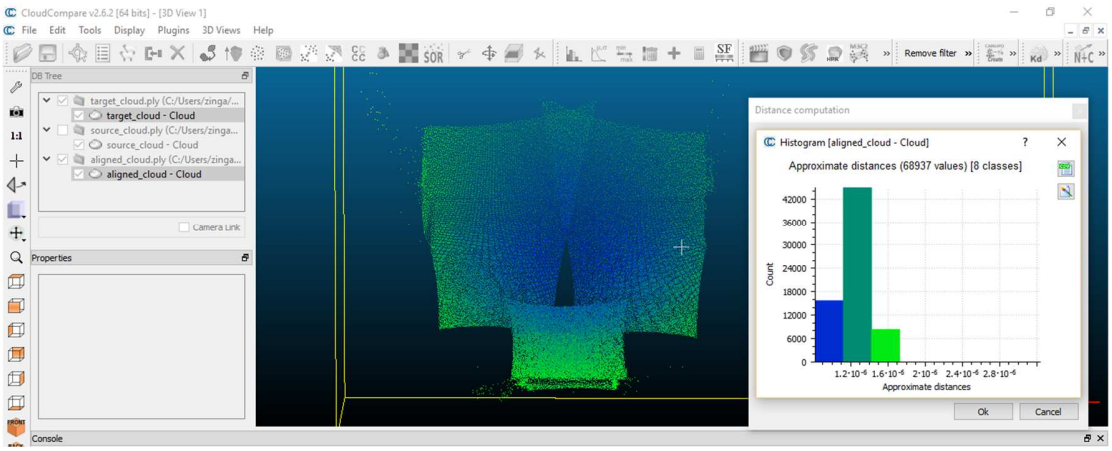

FIGURE 22 ALIGNED POINT CLOUD

3.2.4 OBSTACLE DETECTION AND RESULTS

One of the output of the alignment process is the vector containing the aligned cloud indexes of the points that have the correspondence in the target cloud. This vector, together with the aligned and target point clouds and with the mean distance calculated, is then transferred to the obstacle detection .vi.

The algorithm extracts from the original CAD point cloud the points that have the correspondences with the final cloud. The distance between each aligned point and the correspondent one of the CAD model is then calculated. Subsequently, it is compared with the mean Euclidean distance value obtained by the registration process: if the first is bigger than the mean value plus an offset (a 50% is considered in the preliminary studies), the relative point is considered to be part of the possible obstacle point set. After all the points are processed and a set of points constituting the obstacle has been defined, this vector is given as input to a Radius Outlier Removal filter in order to clean it. The final output is the obstacle detected.

To test the algorithm, an extraneous object has been put in the working scene. Figures 23-27 shows some results obtained from the entire process.

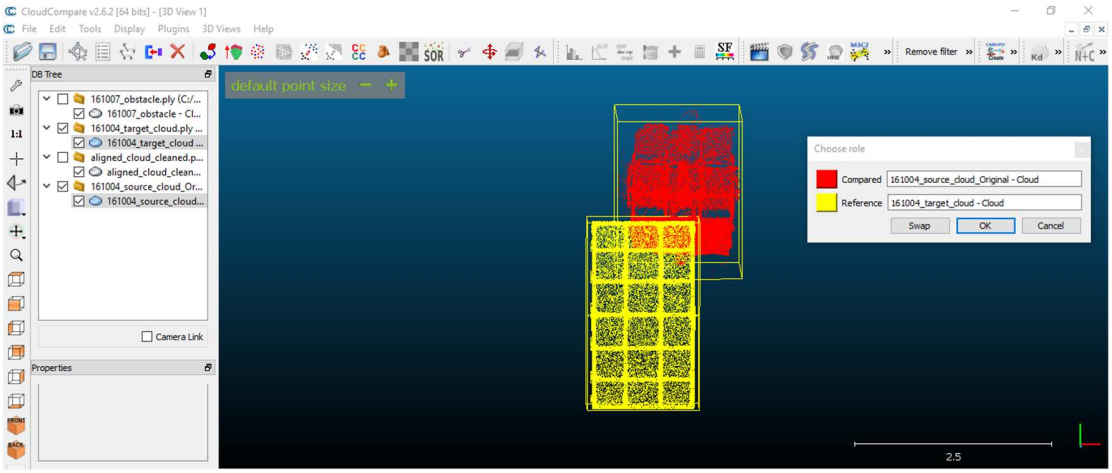

FIGURE 23 TARGET AND SOURCE POINT CLOUDS. IN RED: SOURCE, IN YELLOW: TARGET

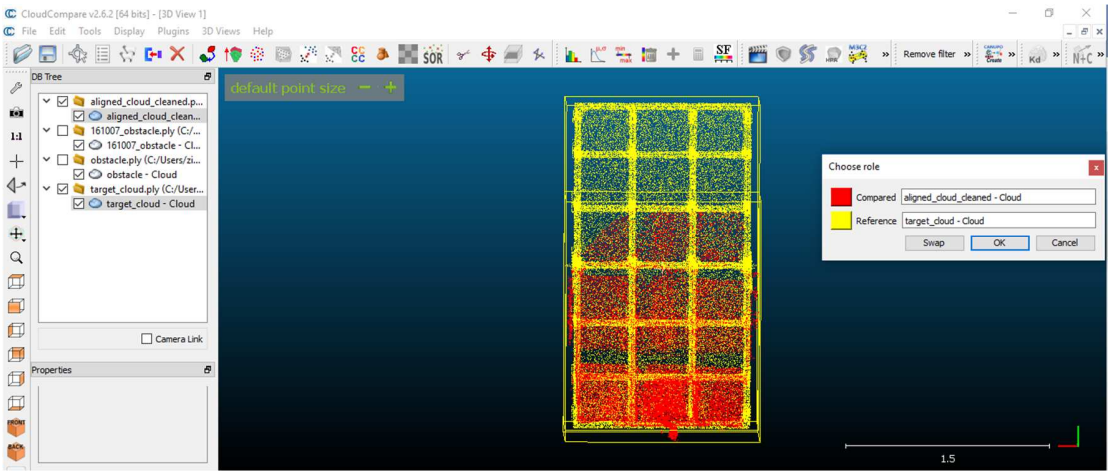

FIGURE 24 TARGET AND ALIGNED POINT CLOUDS. IN RED: ALIGNED, IN YELLOW: TARGET

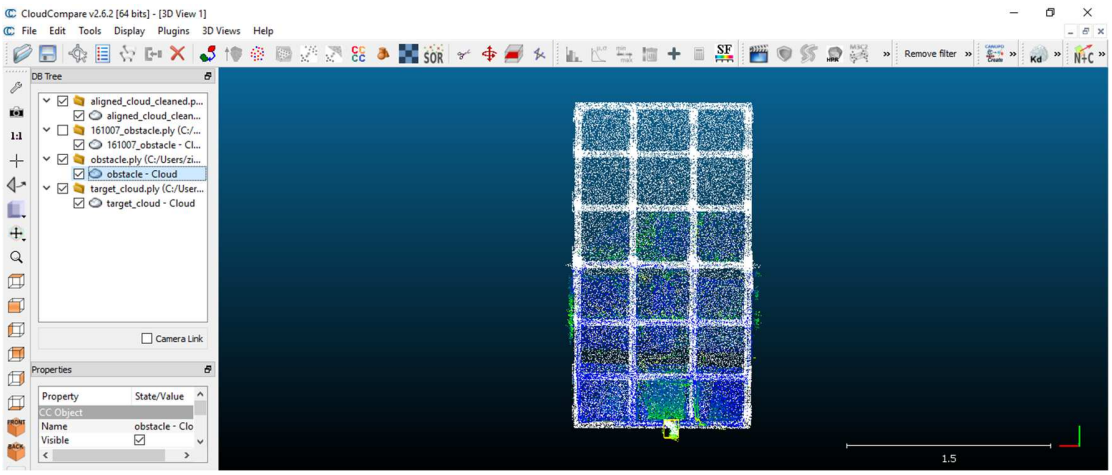

FIGURE 25 OBSTACLE DETECTED (YELLOW RECTANGLE)

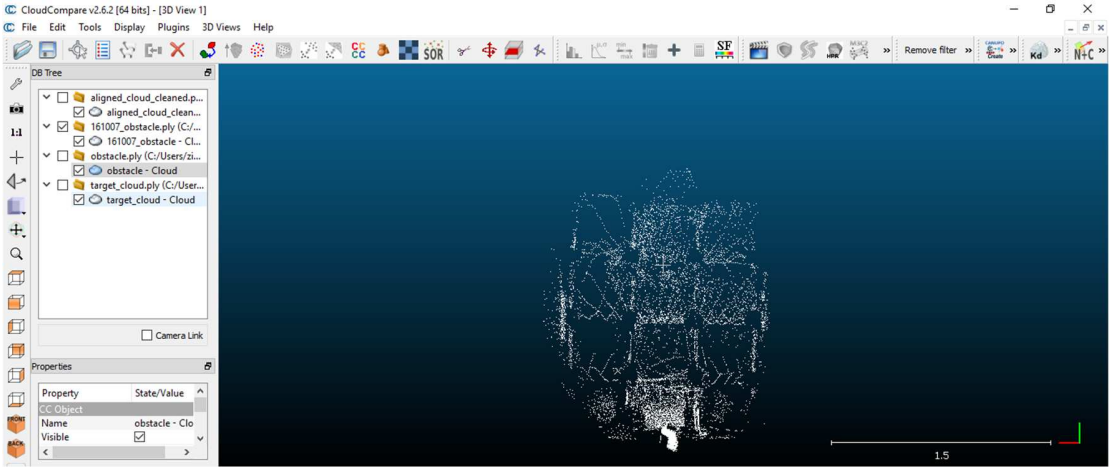

FIGURE 26 OBSTACLE POINTS SET BEFORE FILTERING

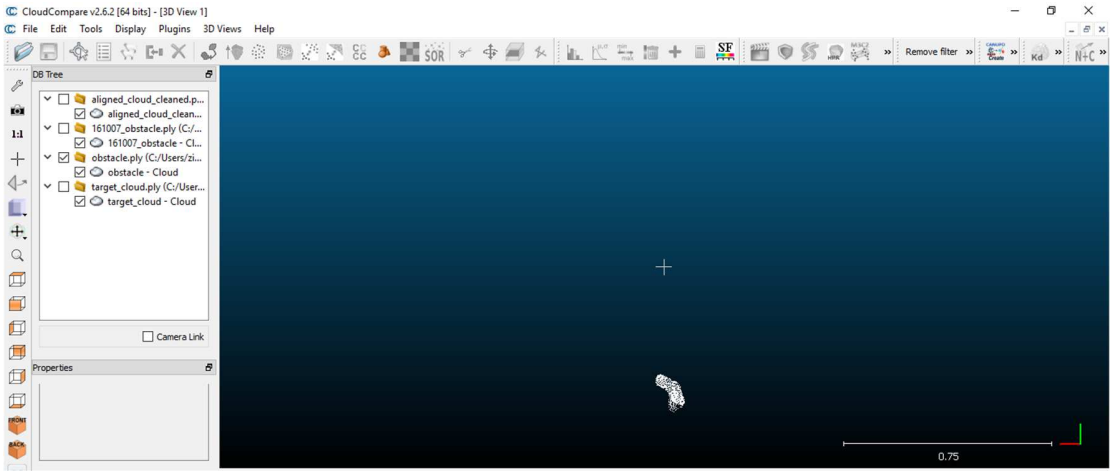

FIGURE 27 OBSTACLE AFTER RADIUS OUTLIER REMOVAL

The rotation and translation matrix obtained from the ICP alignment are:

$$R = \begin{bmatrix} -0.999588 & -0.00236201 & 0.0286864 \\ 0.000345411 & -0.997544 & -0.070085 \\ 0.0287802 & -0.0700434 & 0.997137 \end{bmatrix}$$

$$t = \begin{bmatrix} 0.478804 \\ -1.78924 \\ -0.796692 \end{bmatrix}$$

Lastly, the algorithm has been compiled for the sbRIO-9651 provided by National Instruments in order to test it in the real working environment. sbRIO-9651 includes NI Linux® Real-Time, which combines the performances of a real-time operating system with the Linux environment. As the PCL library is not available for such platform, there was the need to cross-compile it with all its dependencies.

Chapter 4

SECOND CASE STUDY: BAYER'S CHALLENGE

Bayer is an innovation company with a more than 150-year history and core competencies in the fields of health care and agriculture. It develops new molecules for use in innovative products and solutions to improve the health of humans, animals and plants. Its research and development activities are based on a profound understanding of the biochemical processes in living organisms. Manufacturing of its products requires cutting edge technology and the highest standards of quality. However, for certain tasks automation of unstructured activities, it creates a challenge. The 2017 Grants4Tech competition aims to strengthen the bond between the Lifescience industry and academic robotic research communities. It shall give insights into the types of unstructured automation challenges they face while advancing new solutions in the area of robotic automation. The challenge will ask participants to build the robotic hardware and develop the software (a "Robot") that will attempt to master a simplified version of taking samples for quality control from a drum in goods receipt. The Robot will be tasked to open the drum, open the inner bag, tag the sample and finally close the bag and drum. This task is done hundreds of times every day at Bayer's multiple sites worldwide. The challenge combines image recognition, movement planning, precise handling of different materials and tools and error recognition in a fixed amount of time. The Robots will be scored by fulfilling a sub step successfully.

4.1 THE CHALLENGE

With the Grants4Tech robotic challenge, Bayer aims to strengthen the bond between the Lifescience industry and academic robotic research communities and enthusiasts. They aim for a complete new class of applications based on modern robotics.

Sampling of powder for quality control from a drum of incoming raw materials is done hundreds of times a day at each of our multiple sites worldwide. Solving this challenge will have a disruptive impact to future operations and robotic applications. The challenge expects to use robotic hardware and develop the software to perform sequentially several tasks. Some of them are mandatory in order to success; others are not core but optional.

4.1.1 STEPS TO FULFILL

In the following, all the steps that the robot has to fulfill are described.

1. Open drum

Bayer provides two different drum sizes. During the challenge presentation, all steps will be fulfilled with one drum picked by Bayer. In sequence, the steps are:

a. Break the seal
b. Open the clamping ring
c. Securely place the clamping ring aside
d. Securely place the lid aside

b, c, d steps are mandatory.

2. Open inner bag

This step includes cutting the cable strap.

a. Cut the cable strap
b. Place it aside
c. Open the bag

All these steps are mandatory.

3. Take sample and place into sample holder

This task is optional.

   a. Take a spoon
   b. Take a sample (minimum 20g)
   c. Put the sample into the sampling container
   d. Securely place the lid on the sampling container
   e. Put the spoon and sample away

4. Close inner bag
   a. Grab and close the bag
   b. Attach a new cable strap to inner bag
   c. Push down the closed inner bag

5. Close drum and attach new seal
   a. Take lid
   b. Put the lid on the drum
   c. Replace the clamping ring on the drum
   d. Apply a new seal

4.2 THE PROPOSED SYSTEM

In order to accomplish the tasks presented in the Bayer challenge, an ad hoc team has been created. Loccioni Group and Istituto Italiano di Tecnologia (IIT) formed a consortium with the purpose of maximizing experience and know-how, with the aim of presenting a convincing solution. This team, called *Laial*, in fact combines the knowledge of IIT about the design and development of new robotics components, with the expertise of Loccioni in integrating and deploying advances robotics technologies in industrial environments.

The technical concept is based on the use of two robotic arms supported by several sensors and tools, together with a software layer that will make possible the coordination and execution of the tasks to perform. The robotic system for powder sampling is composed of two collaborative robots (Universal Robots) equipped with electric grippers (Schunk). The robots are then equipped with customized end-effectors and appropriate sensors, in order to manipulate the drums and to perform all the operations requested by the challenge. The fingers of each gripper are designed by IIT in order to engage different types of tools to cut, take and move. The movements of the two robots are coordinated by an external PC and guided by a vision system composed of a 2D and a 3D camera (IDS). The two cameras provide information about both the colors and the distances in the acquired scene.

Loccioni integrates all these technologies in an innovative robotic solution able to:

- Cut the seal
- Open the drum
- Open the inner bag
- Sample the powder
- Close the drum
- Put a new seal

To fulfill them, the team decided to operate as follows.

The drums are placed between the two robots, without any requirement on its angular orientation in the vertical direction. The 3D vision system is positioned on top of the working area. In this way, it is able to detect the clamping device as well as the seal that has to be cut to open the ring. The output of the 3D vision system will be the X-Y-Z coordinates of the detected object, that will be provided to the software application in order to proceed with a coordinated motion of the robots. In this way, lid and ring removal can be carried out with the manipulators' grippers.

The closure of the drum will be addressed similarly, exploiting the information coming from the 3D camera.

Regarding the cable strap in the inner bag, a solution for both the cutting and the replacement has been conceived: firstly, the 3D vision system detects the cable-strap thanks to the information of the 2D color camera. Once approached it, the cut is done using proper end-effectors, without piercing or damaging the inner bag.

Concerning the sampling process, one of the two robots inserts vertically the sampling tool in the inner bag.

Figure 28 shows the whole system developed and presented to Bayer.

FIGURE 28 SYSTEM PROPOSED

The main activity that is going to be described in this chapter is the one related to the vision system. In fact, the student focused mainly on the vision algorithms implementation and their integration with the rest of the system.

4.2.1 VISION SYSTEM

The vision system is composed of a 2D and a 3D camera provided by IDS. The chosen 3D camera is an Ensenso 3D N35, which specification are detailed in Table 4. It has two monochrome CMOS sensors (Global Shutter, 1280 x 1024 pixels) and a projector.

TABLE 4 ENSENSO N35 SPECIFICATIONS

1280 x 1024 px 1/1.8″ global shutter CMOS sensors
FlexView pattern projection module (N35 only)
Interface: Gigabit Ethernet with Power over Ethernet
Water and dust proof according to IP65, IP67
Optional 12-24V external power supply
Available focal lengths: 6 – 16 mm
Light sensitive optics with F=1.6 apertures
Available with blue (465nm) and infrared (850nm) wavelengths
12-24V GPIO, trigger input and flash output
All connectors lockable
Dimensions: 175 x 50 x 52 mm
Factory calibrated

FIGURE 29 ENSENSO N35

The EnsensoSDK provides the interface to the stereo cameras. Its main components are the NxView demo program and the NxLib library. NxView demonstrates the main functions of the NxLib library. It allows to open one or multiple stereo or color cameras and displays the captured depth and texture data. It is possible to interactively adjust parameters like exposure, binning, AOI and stereo matching volume. NxLib is the core library of the SDK. It contains the stereo matching algorithm, calibration and processing functionality.

The 2D camera is a USB uEye color camera, provided by IDS. They are positioned in a fixed position to always focus the scene. They can be combined in order to have both the distances and color information of the scene. This information is fundamental for the present challenge because the color recognition is an essential aspect for fulfilling different steps of the challenge.

4.2.1.1 CAMERAS CALIBRATION

NxLib provides a number of functions to simplify the usage of multi camera setups. Before cameras can be combined into a single capture system, they have to be calibrated together. This is currently done pairwise using the NxView program. One can simply

select two cameras and calibrate them with a calibration pattern. After the calibration, the relative position and orientation of the two cameras is known and saved into one of the camera's EEPROM. Each camera can store one link (relative position and orientation) to another coordinate system. The linked coordinate system can either be another camera, or the workspace system.

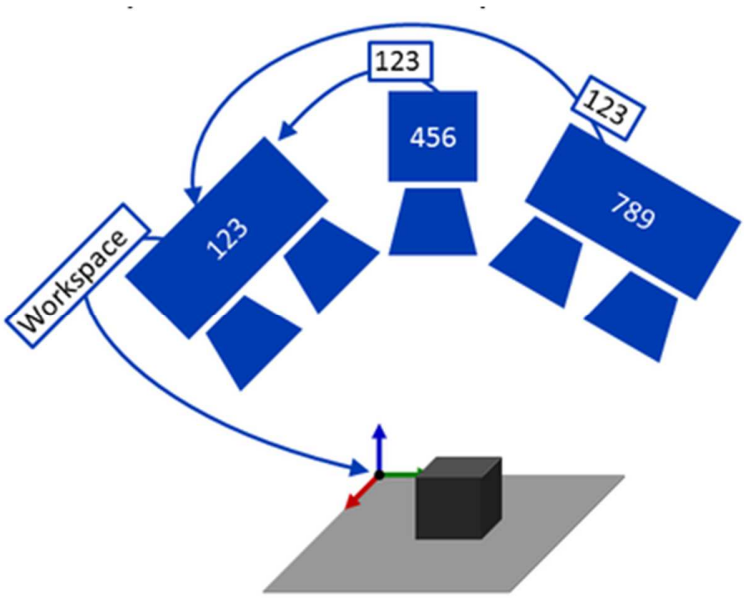

FIGURE 30 MULTI CAMERA SETUP

As each camera stores exactly one transformation into a destination coordinate system, a multi camera setup can always be seen as a tree of linked cameras. The link is stored as a transformation, together with the name of the destination coordinate system, which can be either a camera serial number string, or the special name "Workspace", to refer to the workspace system. Whenever coordinates are output by a camera, this is done relative to the top-most coordinate system that can be reached via the camera's outgoing link. This might either be the workspace, or the camera at the tree root, when no workspace system is calibrated.

It is possible to calibrate a capture system consisting of two or more stereo cameras and one or more color cameras; we can assume to have two stereo cameras ("123" and

"789") and one color camera ("456"). After mounting the cameras, firstly the two stereo cameras are calibrated together. It is possible to decide where the resulting link should be stored (for example, in the stereo camera with serial number "789"). Then it is possible to add the color camera "456", by calibrating it against the stereo camera "123"; the calibrated link is automatically stored in the color camera. When opening all three cameras together, it is additionally allowed to calibrate the workspace system. NxView then automatically decides to store the workspace link in the camera "123", as it is the only camera with an empty link. The resulting link tree looks like Figure 31.

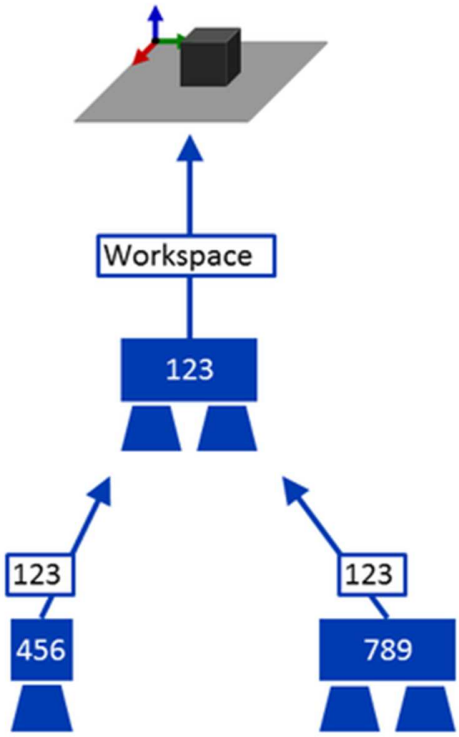

FIGURE 31 LINKING TREE

When capturing point clouds with the stereo cameras, each point cloud is then transformed into the common reference system. In other words, each camera follows its links along the tree towards the root and transforms its point cloud into the found root system by applying the transformations along the path to the root. Thus, the point clouds all live in the same coordinate systems and can be combined by simple merging.

In the present application, the 3D stereo camera has been calibrated together with the 2D color camera; the workspace reference system has been taken as the center of the drum. In the following, the resulting calibration matrices are reported.

$$K_{2D} = \begin{bmatrix} 2311.620849609375 & 0 & 1243.04638671875 \\ 0 & 2311.620849609375 & 1019.9951782226562 \\ 0 & 0 & 1 \end{bmatrix}$$

is the calibration matrix of the color camera, while

$$K_{left} = \begin{bmatrix} 1475.4223243059644 & 0 & 539.962122752121 56 \\ 0 & 1475.4223243059644 & 510.78653198826697 \\ 0 & 0 & 1 \end{bmatrix}$$

$$K_{right} = \begin{bmatrix} 1475.4223243059644 & 0 & 677.58857543798172 \\ 0 & 1475.4223243059644 & 510.78653198826697 \\ 0 & 0 & 1 \end{bmatrix}$$

are the calibration matrices of the left and right cameras of the stereo system of the Ensenso N35. The link between the Ensenso and the monocamera is represented by an angle and axis, which represent the rotation of the color camera in respect to the other, and a translation vector:

$$Angle = 0.051236202639722013$$

$$Axis = \begin{bmatrix} 0.83165780884521079 \\ 0.19789968403360292 \\ 0.51882656451475484 \end{bmatrix}$$

$$Translation = \begin{bmatrix} -56.381912231445312 \\ 78.608634948730469 \\ -12.772295951843262 \end{bmatrix}$$

The vision system is then calibrated with the robotic system. The rototranslation matrices between this and each robotic arm have been calculated so that the two systems rely on the same reference system. In this way, the vision system could give to

the robots the correct 3D data information. The fact that the two cameras are calibrated together means also that the 3D data is combined with the color info. An example of the result of the calibration is illustrated in Figures 32 and 33.

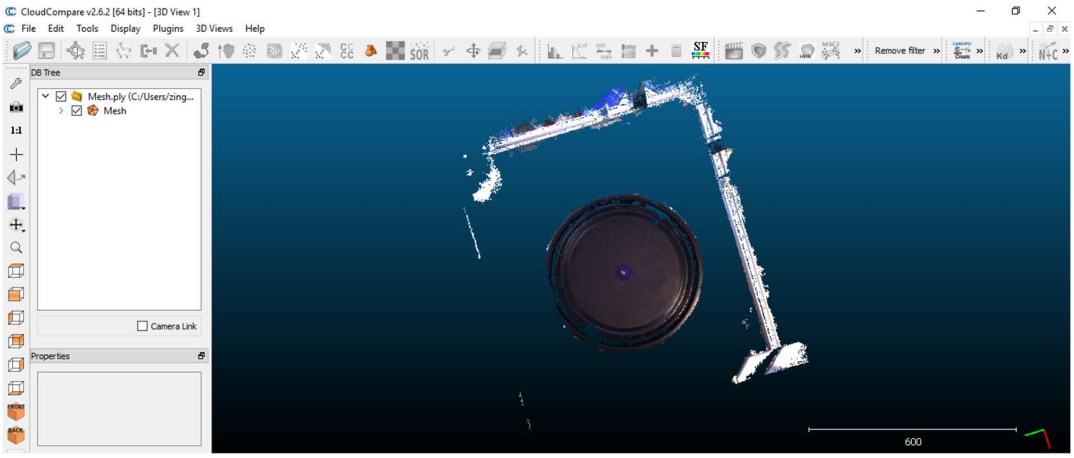

FIGURE 32 CAMERAS CALIBRATION RESULT /1

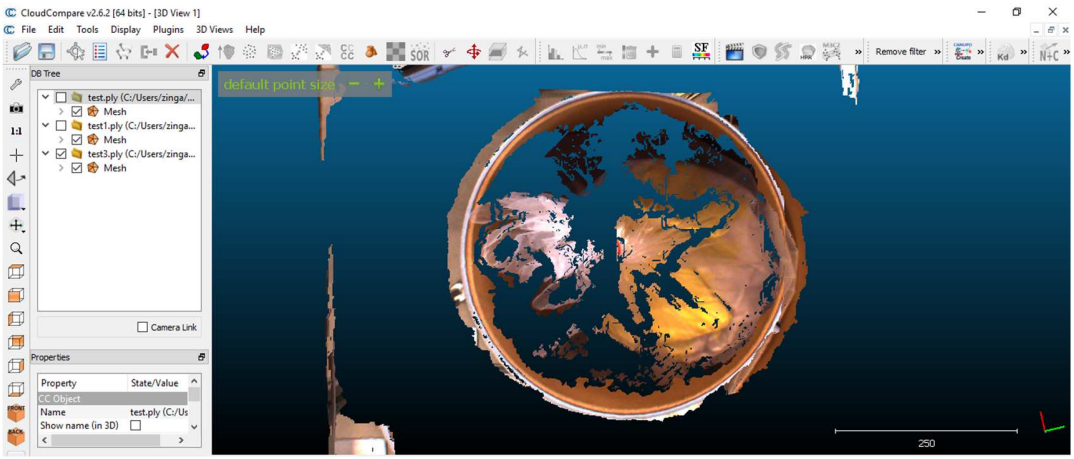

FIGURE 33 CAMERAS CALIBRATION RESULT /2

4.3 ALGORITHM IMPLEMENTATION

The activity carried out in the context of the presented challenge regards the design and development of the vision algorithm in order to give information to the robots. In fact, for some of the mandatory tasks, the robotic system needs information about the

localization of specific objects in the scene. Specifically, the first task of Step 1 (Open drum) asks the robot to break a yellow seal, while Step 2 (Open the inner bag) requires cutting a red cable strap. Without the data from the vision system, these tasks are even more difficult to fulfil.

The algorithm has been implemented in LabVIEW because of full compatibility with the software implemented for the robotic systems. The EnsensoSDK makes available the .net libraries through which the connection of the cameras with LabVIEW has been established.

The next paragraphs focus on the description of the algorithm implemented to reach the two objectives of finding the yellow seal and to localize in the three-dimensional space the red cable strap.

After both the cameras are started, the parameters of the Ensenso have been set. The available drums could be of two sizes, a smaller one and a bigger one. On the base of which type of drum Bayer should have chosen to use during the challenge, the 3D camera has different parameters in order to have the best quality of images.

When an image is grabbed, from the libraries it is possible to recover both the single 2D images (the monocular from the stereo camera and the color one) and 3D data.

4.3.1 SEAL RECOGNITION

The first step that the vision system is required to perform is to identify a yellow seal in the lid of the drum. The robots have to break it in order to open the drum for Step 1. The arms need its position in the reference system. It has been decided to operate through color recognition.

FIGURE 34 YELLOW SEAL

First of all, a brightness, contrast, and gamma correction to each color plane separately is carried out. Then, the software applies a threshold to the three planes of an RGB. The threshold is based on the yellow RGB values determined off-line. Some morphological transformations are then applied in order to find a unique particle in the image corresponding to the seal and its position is then retrieved and passed to the robots. The rot-otranslations of the arms are applied so that they are able to pick and break the seal.

4.3.2 RED CABLE STRAP LOCALIZATION

Step 2, open the inner bag, requires cutting the cable strap that keeps closed the bag which contains the powder to analyze. The cable strap is red and so is fundamental to take this color information into consideration for identifying it. The length of the cable is not fixed and depends on how much tight the bag is closed. So, there could be a long part of cable that, if cut, does not allow to really open the bag (Figure 35). Another situation that could occur is when part of the cable is hidden under the bag and in the image there are two pieces of the cable, like in Figure 36. The difficulty relies on the fact that the robot needs a point on the cable that allows to open the bag, so a point that is in the piece of cable that practically closes the bag.

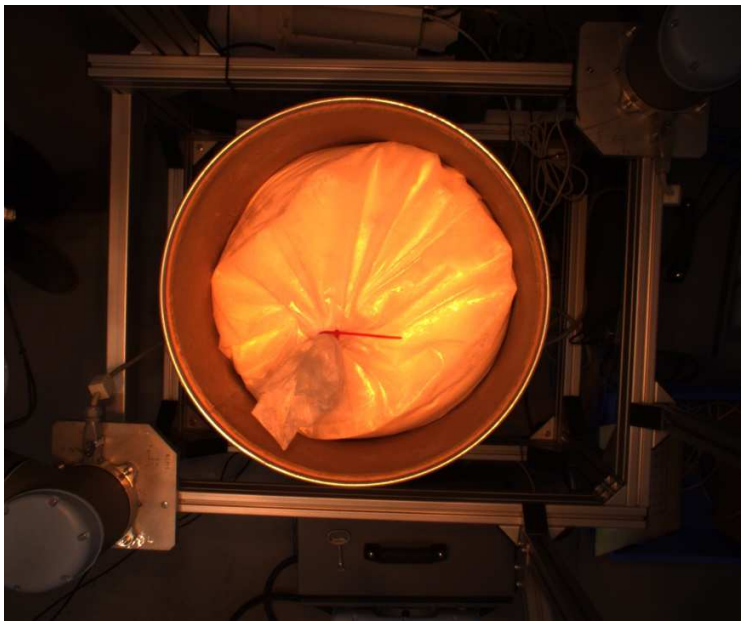

FIGURE 35 LONG CABLE STRAP

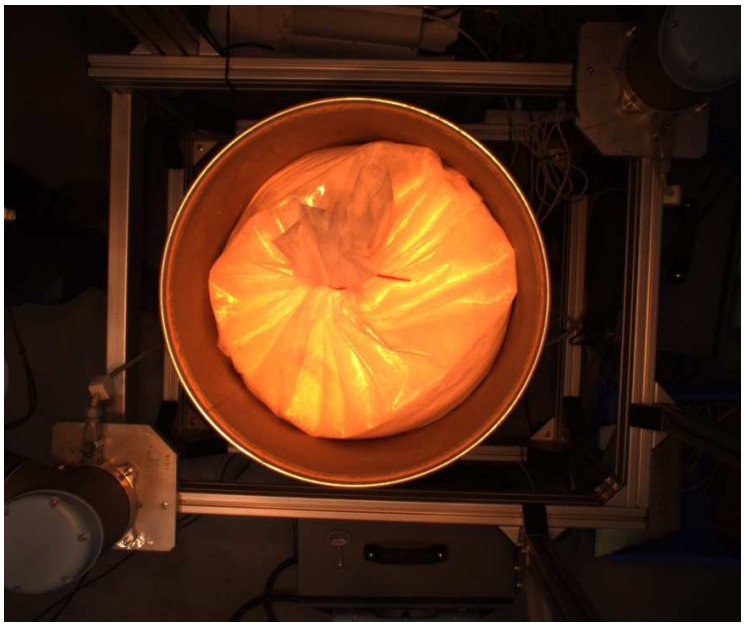

FIGURE 36 TWO PIECES OF CABLE STRAP

To address these difficulties, different methods have been implemented and tested. The one with the best capacities of recognizing the most suitable cutting point was finally chosen.

First of all, after the images are acquired, a ROI is created. The circumference of the drum is first detected; to limit the area of searching and to maximize the colors contrast, a subtraction of 250 pixel from the radius is performed.

The second step consists on finding the red cable strap in the image. After the application of the ROI, a threshold on the RGB values is computed; in this way, it is possible to keep only the pixels having RGB values contained in a range around the red value of the cable. Some morphological transformations are computed and the number of particles remained in the image are calculated. If the number of particles is more than one, just one (the biggest) will be selected. If it is equally to zero it means that no cables are found; otherwise, when it is equal to one, the X, Y of the center of mass and the orientation are extracted. Furthermore, the slope and the intercept of the cable are calculated and saved because they will be useful in the following steps. An example of the result of the elaboration is illustrated in Figure 37: in the right side of the picture the cable strap has been isolated.

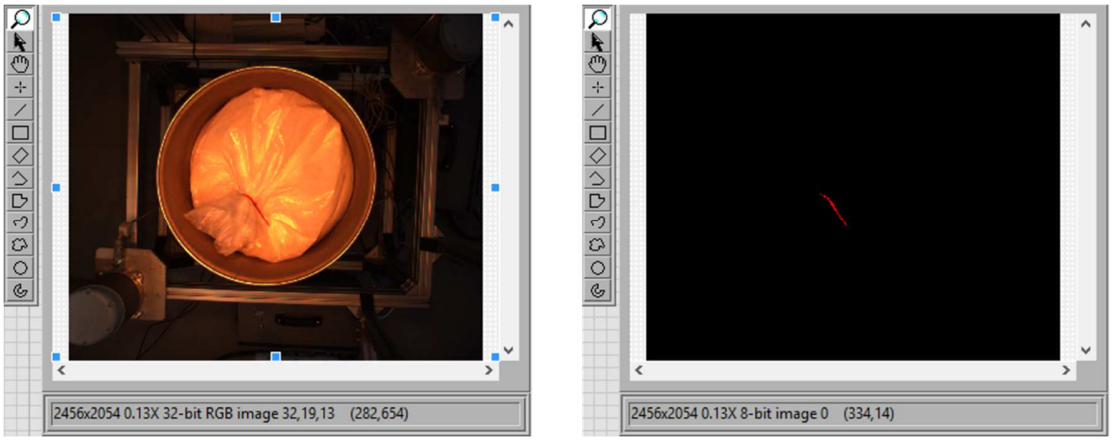

FIGURE 37 RED CABLE DETECTION

Having only the orientation and the position of the center of mass of the cable strap is not sufficient to find the proper point to be given to the robot. In fact, if we take the

mean as the point to cut, we cannot assure that it will be sufficient to effectively cut the strap. If the cable is long, the point could result in the piece outside the grip.

For this reason, it was necessary also to take into consideration other aspects of the scene. The tuft of the bag is, often, the highest part inside the drum, so the one nearer to the camera and its identification could ease the recognition of the cable strap. To individuate the part of the tuft beyond the cable strap and then extract the exterior contours of it, namely the perimeter, a sequence of steps have been performed. From the 3D point cloud, a grayscale image is obtained (see in the right side of Figure 38)., where the value of each pixel is the information on the Z axis, so the distance from the camera –more the point is distant, higher is the pixel value Working with the grayscale image, some morphological transformations are applied.. First of all, a ROI (Region of Interest) is created with the information about the drum diameter calculated during the seal recognition phase, in order to take the unnecessary elements in the image out. Then, 10 iterations of Erosion, to eliminate isolated background pixels, followed by Dilation are applied. The process continues with the application of a threshold. The minimum and the maximum pixel values of the grayscale image, that correspond to the minimum and maximum distances from the cameras, are calculated. The interval considered for the threshold is [minimum distance; minimum distance + 300mm]. A Particle Filter algorithm is then applied to remove small or isolated parts. If only one area remains in the image, the perimeter is extracted from the exterior contours of the particle; if there are more than one areas, the perimeter is drawn out the biggest one.

After that, two perpendicular lines (number 1 – Fig. 39) are marked on the cable strap, equally distant from the center point. On these lines, a line profile analysis is conducted and the perpendicular that has the point with the minimum Z value is taken to calculate the new "mean point" of the cable strap (point A – Fig.39): it is obtained from the average point between the center of the cable strap and the intersection with the perpendicular chosen. To find the point to cut with more precision, the minimum distance between each point of the tuft's perimeter and point A is searched; the intersection point between the two lines is then calculated.

4.3.3 RESULTS

The method described above was the result of many trials of different modalities. A statistical analysis on the success rate of each method, tested with about 40 images, has been done and the one with the best rate was chosen. Figure 38 shows the main window where the X and Y pixels of the point to cut is highlighted; the correspondent 3D coordinated, calibrated with the robotic system, are exchanged with the robots. Figures 39 and 40 illustrates a couple of good identification of the point, where in red is drawn the cable strap line, in yellow the perpendiculars and in blue the point found. Figure 41, instead, represents a quite good approximation of the point but the algorithm did not completely succeed in the individuation of the proper one.

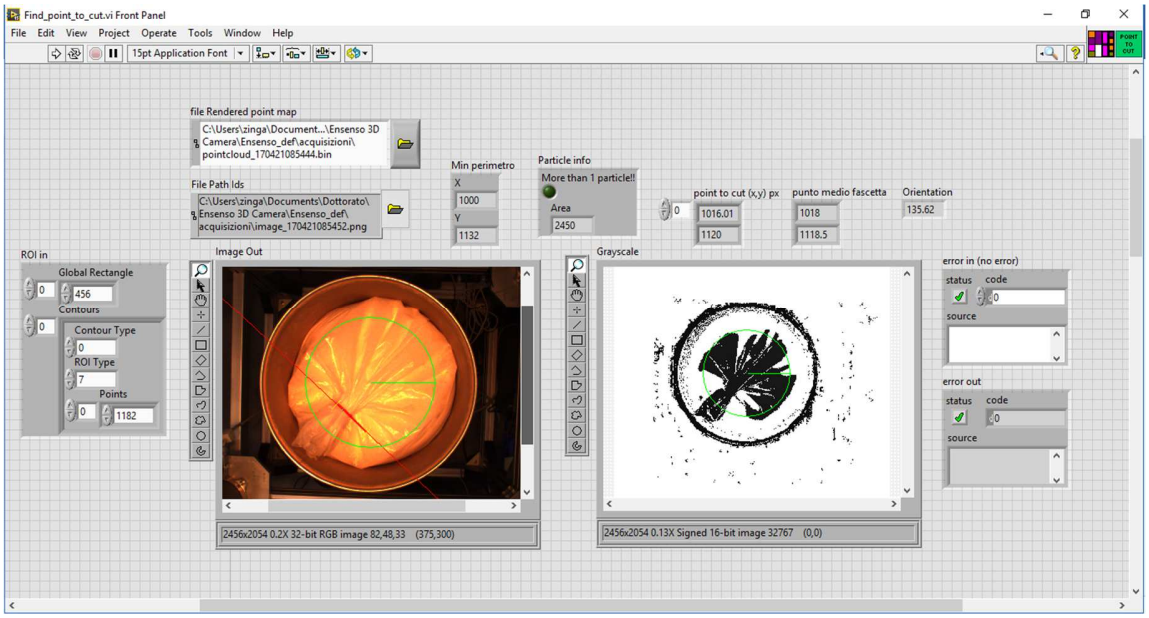

FIGURE 38 MAIN WINDOW OF RED CABLE STRAP PIPELINE

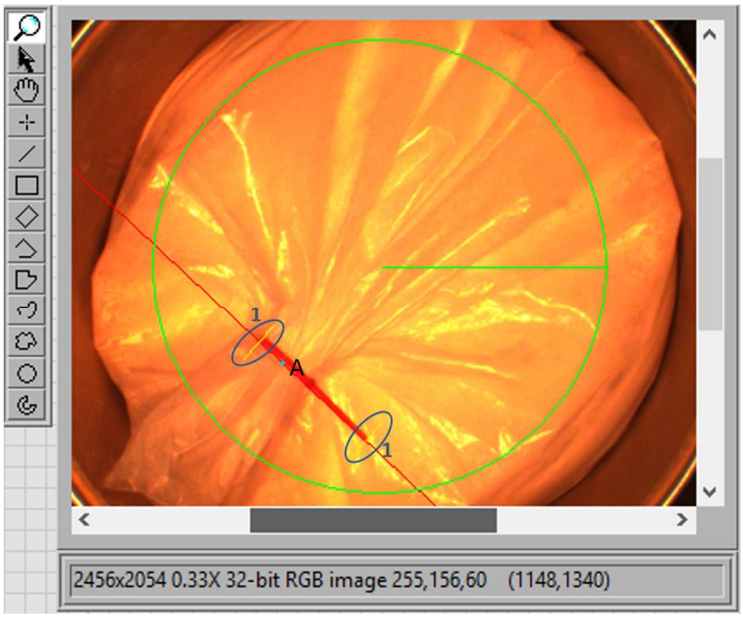

FIGURE 39 POINT TO CUT IDENTIFIED

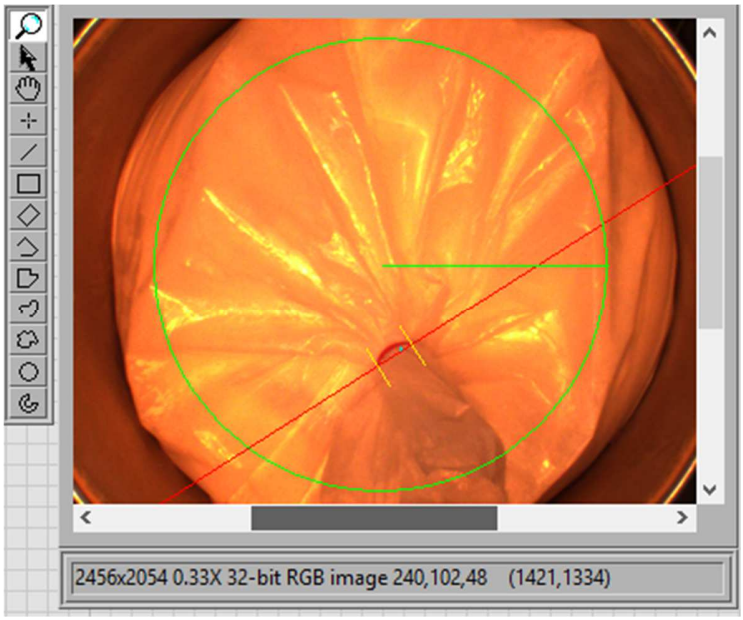

FIGURE 40 POINT TO CUT IDENTIFIED /2

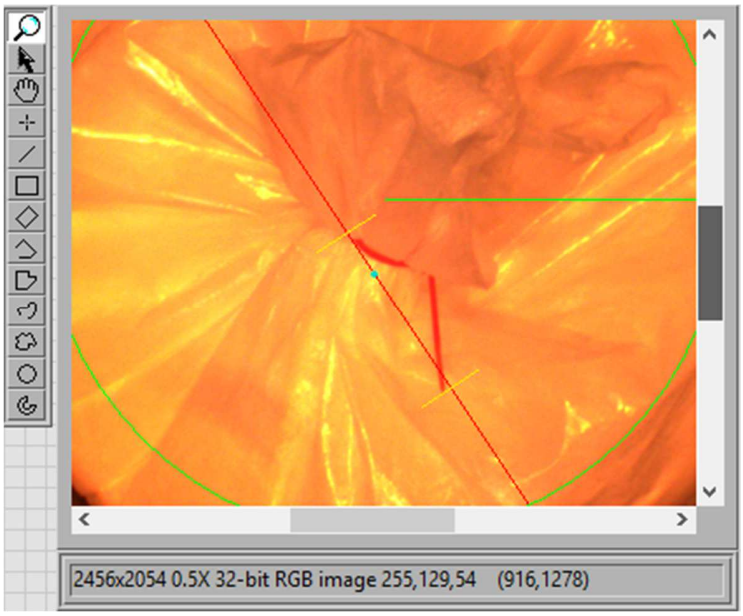

FIGURE 41 POINT TO CUT IDENTIFIED /3

Chapter 5

UNDERWATER APPLICATIONS

The work presented in the previous parts focused on the vision systems applied to structured environments, specifically the industrial one, where the robotic systems navigate and perform actions in the scene using the information coming from its vision system.

Target detection and localization in complex environments like the underwater one are more challenging. As mentioned in Chapter 2, many types of sensors are used for these purposes but vision systems are always more employed since they reduce space and cost and increase the resolution, although their range significantly depends on the water conditions.

The common need for this kind of applications is to obtain enriched maps of the underwater environment or precise representation of specific details in an efficient, economic and non-invasive way. Some examples are monitoring of underwater building status, biological marine environment monitoring and analysis or survey archaeological sites.

In order to address these aspects, the last part of this thesis concentrates on the design, development and test of the framework presented in paragraph 2.7 applied to underwater devices or vehicles. The final aim is producing geo-referenced 3D maps of the underwater environment and the localization of the robot or the diver. These maps, extracted at subsequent instants, should be used in SLAM algorithms, as described in Chapter 2.
In both the applications presented, the Structure from Motion technique has been adopted.

5.1 DOCUSCOOTER

5.1.1 GREEN BUBBLES PROJECT

Green Bubbles project (http://www.greenbubbles.eu/) has received funding from the European Commission's Horizon 2020 research and innovation programme under the Marie Sklodowska- Curie grant agreement No 643712 and it involves 4 academic and 5 non-academic participants (4 SMEs) from 6 countries across 3 continents. Its central objective is to maximize the benefits associated with diving, whilst minimizing its negative impacts, thus achieving the environmental, economic and social sustainability of the system.

Recreational SCUBA diving has become a mass leisure activity engaging millions of divers worldwide and the diving industry generates large direct and indirect revenues for local communities and Marine Protected Areas (MPAs). Other benefits linked to diving include the promotion of environmental and ocean stewardship, contribution to scientific research, fostering social inclusion and personal development. Yet, diving has also negative impacts, due to damage or disturbance of habitats and organisms and to conflicts with local communities for the access to/use of the same resources, equity issues, or cultural clashes. These aspects clearly relate to the three pillars of sustainability, covering environmental, economic and social dimensions and can only be addressed by a systemic approach.

This will be done by:

- Carefully assessing and modelling the system itself;
- Developing innovative products based on the issues and needs highlighted by assessment and modelling;
- Promoting the uptake of such products by the system designing tailored business models and marketing plans. Direct engagement with selected stakeholders (divers, professionals, operators, certification agencies - CAs,

MPAs, NGOs) will ensure relevant feedback throughout the project's lifetime, as well as effective uptake of results at the end of the project.

5.1.1.1 SPECIFIC OBJECTIVES

Green Bubbles is a project about diving and its positive as well as negative potential impacts. It was conceived to support the evolution of the diving system towards sustainability and to be implemented together with stakeholders. The project intends to address the issues mentioned above pursuing the following objectives:

TABLE 5 GREEN BUBBLES SPECIFIC OBJECTIVES

1 Put the European diving system in focus, establishing a dialogue with all stakeholders.
Dedicated research with, about and for the European diving industry, parallel comparative work on (much more studied) coral reef diving industry
2 Thorough segmentation of 3 key components of the system (divers, professionals and operators)
Baseline assessment of internal and external determinants of behaviour. This will inform all the subsequent research work.
3 Enhance the traditional offer by the diving industry
Assessment of products, business models, marketing approaches. Case studies to design new products that include cultural components or strengthen the nature-based component. Training modules for operators and professionals that offer the new products. Specific marketing plans. Dialogue with stakeholders throughout the work
4 Bring innovation to the diving industry by introducing diving and Ocean Literacy principles in school curricula.
Content analysis of diving & school teaching material in the light of Ocean Literacy principles. Creation of school teaching plans linking Ocean Literacy + diving to national curricula, and to diving programmes. Dedicated marketing/business plans.
5Bring innovation to the diving industry by transforming Citizen Science in an opportunity for operators and MPAs.
Analysis of existing Citizen Science projects. (Re)design of projects with clear scientific, managerial, educational, economic goals. Software and hardware, business models and co-management approaches. Design of dedicated marketing plan.
6 Bring innovation to the diving industry by creating a quality label that considers environmental, social and safety aspects
Analysis of environmental & safety-related behaviour of divers, professionals, operators (in the water and outside). Definition of environmental/safety-related targets and guidelines, labelling system, business models, co-management approaches. Dedicated marketing plan

7 Bring innovation to the diving industry by developing IT tools supporting the other objectives
For enhanced offers: gamification and portals. For Ocean Literacy: gamification. For Citizen Science: u/w geo-data tools; novel sensors; one-point-shop portal. For quality label: u/w dive-team control system; boat count system
8 Design tailored business models and promotional strategies for the adoption of innovation by the system.
Marketing and business models analysis, market segmentation, tailored strategies. Direct engagement with major stakeholders to elicit meaningful dialogue and feedback, promote participation and awareness, foster the adoption of the project's outcomes

5.1.2 Background and Motivation

Marine and coastal environments are extremely complex systems, characterised by strong links between their physical and chemical processes and biological population. The scientific interest in biodiversity losses and global change of the environment has increased dramatically in recent decades. The preservation of the underwater habitats provides benefits to the whole society and it constitutes a basic need for the mankind. It is therefore essential to monitor factors and indicators in order to evaluate the effectiveness of protected area establishment and management, and to put in place adaptive measures to address emerging challenges such as climate change [289]. A consistent part of the work carried out by the scientific community consists in the acquisition of robust and measurable data. The resulting analysis and the identification and quantification of threats could then reveal the necessity of taking effective measures and improve the monitoring and management system of a particular area. But tracking, understanding and enhancing biodiversity losses, and more in general the monitoring of the underwater environment, require repeated and widespread surveys, which have traditionally been done by research groups of professional divers [290,291]. The number of sites surveyed by researchers is very tiny in comparison to the several thousands that should ideally be surveyed annually such that good and fully interpretations could be made [289].

As a partial solution to these issues, the last years have seen an increasing exploitation of the citizen science in marine monitoring projects [289,292-294]. Citizen science is defined as scientific research that enlists members of the general public in the

collection, categorization or analysis of scientific data across a wide range of disciplines, in collaboration with or under the direction of professional scientists and scientific institutions.

The engagement of such non-specialist volunteers brings to many advantages: the provision of manpower sufficient to conduct extensive surveys, large financial savings and an increase in the level of public awareness of ecological problems through active participation in ecological survey work [292,293]. In literature, there is a debate on the usefulness of the citizen science, specifically on the quality of the data collected by non-experts [295-297]. Despite the advantages previously mentioned, the use of volunteers is often criticised on the grounds that the information collected are unreliable as a result of either insufficient training or a lack of consistency due to the large numbers of different observers. Several research teams tried to answer to these doubts organizing training days before the missions and finding out that the data collected by the general public are consistent with those collected by the scientists [296, 298-301]. To determine the accuracy of the data acquired, they are compared with those collected by marine biologists, who dived simultaneously with the volunteers but not interacting with them. A fundamental aspect that strongly contributed to reduction in costs, time, equipment and logistics for tasks that could have been otherwise impossible to perform has been the establishment of diving as a mass leisure activity. It has broadened the scope of citizen science to the underwater world and it opened the involvement of diving centers in research projects. At a higher level of cooperation and engagement, recreational and professional divers can team up with researchers and managers to identify issues that need to be addressed, and design strategies to do so [302].

There are also major educational and social benefits from the involvement of volunteers in scientific projects. By providing hands-on and engaging experiences in natural settings, the direct involvement of citizens in marine monitoring foster environmental education, stewardship and sense of ownership, that enhance awareness of environmental issues. Citizen science engagement gives people an opportunity to get involved in environmental conservation issues in a novel way,

balancing the need to collect good quality data with public education, and it could encourage an attitude and behavioral change around these thematic [296, 298, 302].

Volunteer monitoring programs can increase the amount of information about the ecological condition of the environment supporting data sharing for community education and providing networks for volunteers and schools [295]. Shared data access and databases, shared involvement in planning and project design, as well as high levels of communication about processes and outcomes between stakeholders, are perceived as central to grow the trust of individuals, community groups and broader society with institutional science, policy and management [297]. In order to achieve these objectives, a suitable suite of scientific workflow products is needed to facilitate the archiving, assimilation, visualization, modelling and interpretation of the information contained in the acquired data. The adoption and development of new technology can significantly facilitate this action. In the recent years an increasingly interest in the realization of user-friendly technologies and tools for carrying out efficiently activities in the marine environment has been seen by the robotic community [303-306]. Many scientific and technical tools contributed to achieve an appropriate and comprehensive understanding of the functioning of the marine environment at different geographic and temporal scales, and the monitoring of marine and maritime activities to ensure their sustainable development [307,308]. What the marine community is lacking is a system where all the processes of acquisition, processing, analysis and output representation are integrated. In literature there are different examples about improving the collection or the analysis steps. Numerous studies, for example, have demonstrated the value of using photographic or video techniques to evaluate benthic habitat conditions rather than employing other high costly equipment [304,309-311]. Recent advances in image processing algorithms and computing power have seen a proliferation of techniques to estimate structural complexity for large reef areas based on digital photogrammetry and application of structure from motion software [310, 312-316] or using annotated underwater imagery [317,318]. But all of these advances do not take in consideration the whole process from data acquisition to output visualization. For this reason, a low-cost and flexible IT tool that allows professionals

and/or citizen science volunteers to collect, analyze and archive simultaneously different types of information from the marine environment during a unique diving session could be very helpful and useful to the marine community.

Within the GreenBubbles project, an innovative integrated device that allows scuba divers to automatically collect, analyze and visualize data from the ocean during their leisure activity, called DocuScooter, has been designed and developed. The user can upload all the information acquired on an appropriate web server as soon as an Internet connection is available and launch a 3D reconstruction process with the uploaded photos and other materials and the results can be stored in the most common three-dimensional formats. The system can provide ad-hoc outputs for different types of applications, like data mining, gamification and 3D printing. With the use of this system it would be also possible to create a wide photo database of the seabed worldwide oriented to different types of end users.

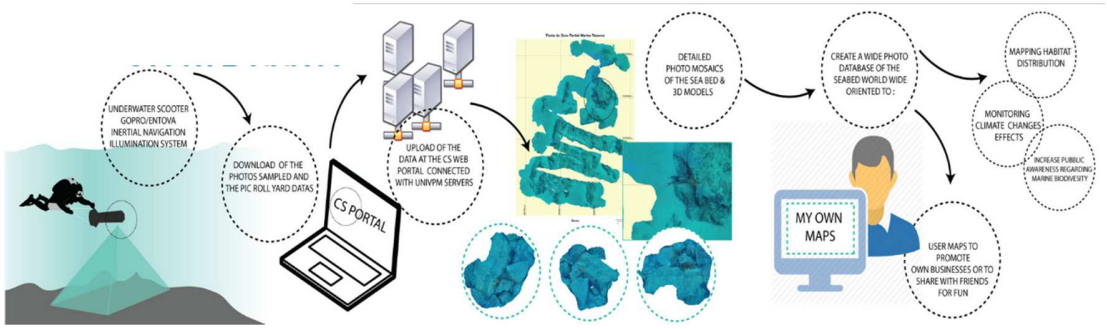

FIGURE 42 CITIZEN SCIENCE PROJECT

5.1.3 DocuScooter architecture

DocuScooter integrates different hardware solutions with the purpose of allowing recreational divers to capture images and data during a diving session, and use them to perform a 3D reconstruction of the marine environment. The acquired data are then available for successive analysis in order to make it useful and exploitable for environmental monitoring. With this kind of product, users can acquire new information and increase their environmental awareness.

The platform consists of a modular robotics system connected with a variable number of payloads such as underwater cameras, lights and other types of sensors, e.g. turbidity or pH measurement sensors, to equip an underwater scooter for scuba divers' activities. The diver can directly manage and control each device with the use of an underwater Android tablet, where a proper application is installed. The entire system is built to work underwater, so all the connectors and the housing are carefully chosen or designed to be suitable for this kind of applications. In the following, the system architecture design and the choice of the hardware components are going to be detailed.

Figure 43 shows the entire system architecture. It is possible to recognize different modules and this modularity is one of the essential aspects to be noticed. The figure shows five peripherals (two underwater cameras, a generic sensor, one underwater light and a battery module) but others can be added or removed on the fly at any time without compromising the entire system.

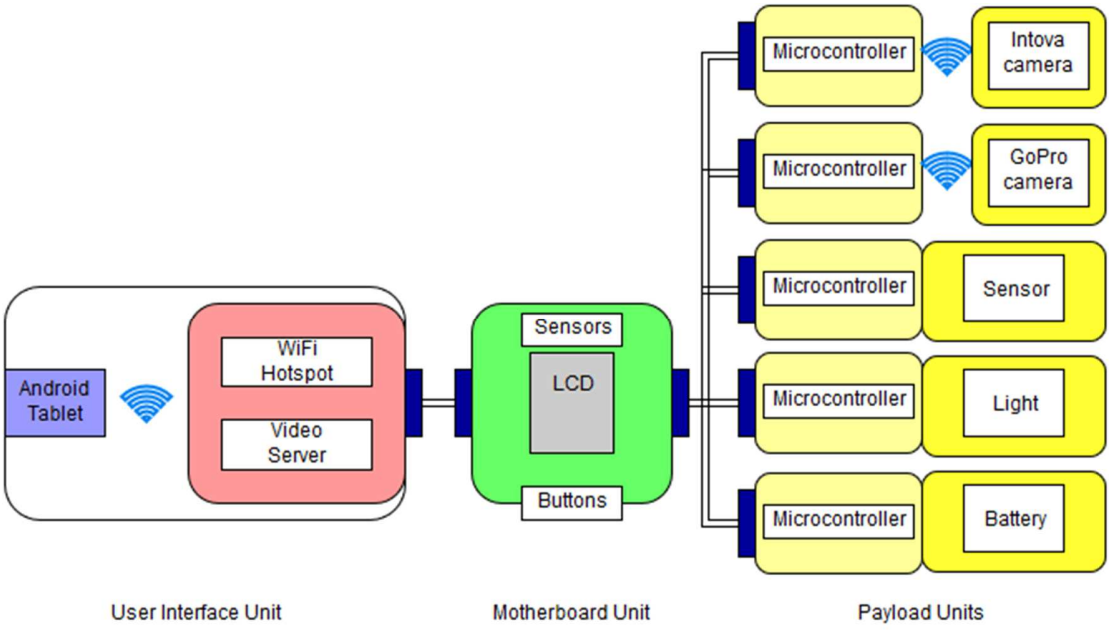

FIGURE 43 DOCUSCOOTER ARCHITECTURE

The first module is the *User Interface Unit*. It consists of two parts: an Android Pad and a subsystem. The underwater tablet employed for this purpose is the Alleco AllTab 2, which contains a Samsung Galaxy Tab S2 8", but the system is adaptable to be employed with different tablet brands; the only constraint is the Android operating system and the necessity of a WiFi connectivity. Through an application specifically developed for the DocuScooter, the scuba diver is able to perform some basic or advanced tasks like following the mission, capturing photos, recording videos and so on. The underwater pad housing is equipped, in the exterior part, with a subsystem that contains a COTS and programmable Wi-Fi Access Point, through which the connection with the tablet is established, in order to exchange information with the other two system units, equipped with UART and Ethernet ports.
For a basic functioning of the system like general exploration and photo documentation, the User Interface Unit it is not necessarily required; the rest of the system is capable of working also without it. This unit communicates with the next one with cables and a Wet Connector.

The *Motherboard Unit* represents the second module. It is composed of:

- a board with a Texas Instruments' MSP430;
- an LCD screen capable of showing important information about the system status and operation (battery charge, mission time, depth, etc…);
- some buttons, in order to interact with the system (shoot photos, turn on/off lights, etc…);
- a series of embedded sensors, that are exclusive to this unit.

The main duties of this layer are to forward messages coming from the User Interface Unit, when it is present in the system, which are directed to the system peripherals and to provide back some specific information from the on-board embedded sensors, like depth. This unit has also discovery capabilities of the connected payloads. If the system is not provided with the User Interface Unit, it is possible to operate directly by means of the interface provided on the LCD screen of the Motherboard Unit.

The last layer shown in the scheme of Figure 43 is composed by one or more *Payload Units*. A "payload" is any compatible device or sensor that the end user wants to use and control in order to acquire data. It comprehends a variable number of peripherals, that can be hot added or removed at any time. In fact, the platform is developed in a flexible way: to attach a new kind of device is mandatory only to respect the transmission protocol and the cable connection. The peripherals are a sort of intelligent devices, in fact as shown in Figure 43, the single component or sensor is not directly connected to the bus but through a microcontroller. The microcrontroller adopted for each unit is a Flyport Pro provided by OpenPicus, that was programmed using C language.

As low-cost action cameras have recently become highly popular in the marine research due to their relative ease of use and high resolution/low-cost ratio, it has been decided to consider them as optical sensors in the Robot System Apparatus of Fig. 13 (Chapter 2). There are three action cameras: an Intova Edge X, a GoPro Hero3 Black™ and a Kodak PixPro SP360. They have been identified as suitable for the platform because all of them are highly configurable through Wi-Fi. Furthermore, the Kodak PixPro Sp360 allows to capture 360˚ of HD video without the need of multiple cameras. Each module is composed of a microcontroller that communicates with the camera using the wireless link; a specific communication protocol has also been defined to exchange data with the rest of the system. Each camera has its own battery, so it does not need a power source. Exploiting the Wi-Fi communication allows all the final users to use their personal Wi-Fi camera within this infrastructure, without limiting customers only to one camera brand. In this way, the flexibility of the system is maximised.

The rest of the preliminary set of payloads is composed of:

- *Light Unit*: this unit is a single, pilotable light source of the system.
- *Battery Unit*: this unit is a single, pilotable power source of the system.
- *Sensor unit:* it represents any specialized sensor the user wants to employ in the mission.

With the option to add as many devices as there are needs, in addition to the fact that the peripherals are independently controlled, it is possible to change each part with an updated one, for example, in the future, without compromising the entire system. It is not limited to the specific equipment presented in this work and it represents a considerable advantage.

5.1.3.1 ANDROID APP

A custom android application has been developed in order to have the full management and control of the payloads and, more in general, of the entire platform. Due to the fact that the application should be easily employed by divers in harsh conditions, the principal goal is to provide a user-friendly interface, for data management and acquisition of information during an underwater mission. Beyond the abilities of displaying data regarding the ongoing and past missions and providing the functionality of exporting them, shooting photos and recording videos, the application is able also to acquire the georeferenced underwater positions provided by an ad-hoc algorithm. Moreover, a worthy and fundamental feature is the detection and communication with all the units in the system. In fact, the application is capable of configuring properly any type of known payload unit attached. Each type of unit has a proper set of commands that can receive. For example, when a light with variable light intensity capability is connected, it will be possible to vary the level of emitted light directly from the application.
Going into detail of the Graphical User Interface, there are three main screens:

- *Mission Database*: here it is possible to visualize all data of previous missions, export them and upload them to the external web server. From this screen, it is possible to continue a selected mission.
- *New Mission Configuration*: in this screen it is possible to set all parameters and information regarding the new mission that is going to be performed.
- *Device Manager*: here it is possible to visualize, customize, configure and test all the devices attached to the system.

From both *Mission Database* and *New Mission Configuration* screens, the user is redirected to *Mission Cockpit* window. This is the main screen regarding the ongoing mission where all the information related to that are shown, like real time position, depth, roll, pitch and yaw angles and other useful information for the user. In order to let the diver interact with the payloads attached, buttons are provided. Specifically regarding the cameras, there are two ways of use: if more cameras are connected, it is possible to ask an action from any single camera (e.g. shoot photo from camera one) or ask the same action from all the cameras (e.g. shoot photo from all cameras).

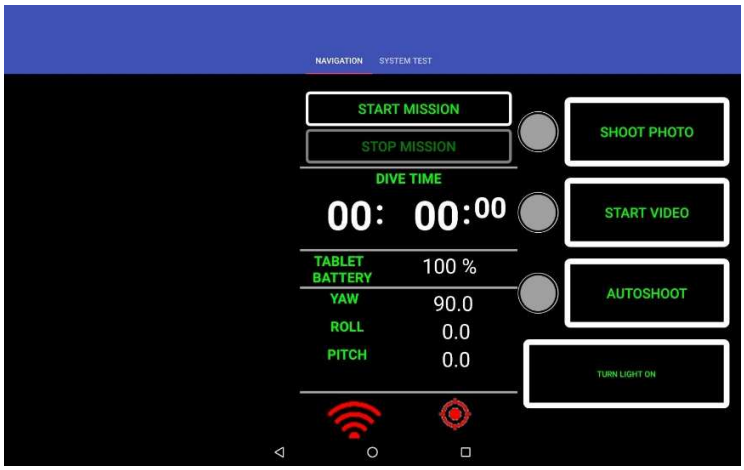

FIGURE 44 DOCUSCOOTER APP

5.1.4 VALIDATION AND RESULTS

The first prototype of the DocuScooter has been tested different times by marine biologists during the mobilities done at Hopkins Marine Station of Stanford University in the course of the three years of activities.

The electronics of the User Interface Unit, the Motherboard unit and the Payload unit have been inserted in a unique watertight box in order to perform the preliminary tests. After the system has been verified to correctly work out of water, a series of test have been conducted in water. The aim of the missions was to evaluate the functionality of the system and the inspection of a semi-submerged pipe located at the bottom of the seabed. The diver was instructed to survey the pipe different times in order to have

images redundancy for the 3D reconstruction process of an area of about 2 x 14 meters. Figure 45 presents the prototype of DocuScooter.

FIGURE 45 DOCUSCOOTER PROTOTYPE

The time delay between the pad and the cameras are quite satisfactory: the capture time for one photo is about 700-800ms. The 85% of this is due to the answer time of the adopted camera, a parameter that we are not able to manage. In fact, those commands that not need to pass through the WiFi connection of the cameras (e.g. the request of the connection status of the cameras) have an average response time of 90ms.

The time for obtaining the 3D model depends mainly on two factors: the number of photos and the quality requested. A low quality model can be available with few hours of elaboration and this means that the researchers can better organize another survey for the same day or the day after analysing the site models obtained with the feedback. The high quality model is instead available in a longer time, e.g. 1 month, together with the possibility of having also its 3D print. The output can be produced in different formats, so manageable with different kind of software, and it can be also imported in virtual environments like common game engines or common GIS which makes its fruition easier. Figure 46 shows the low quality model resulted from the elaboration of about 200 photos in four hours.

In parallel to the reconstruction process, from the cameras alignment and correspondence steps, is possible to recover also the trajectory followed by the diver. An example of this output related to the pipe inspection is illustrated in Figure 47. It can be noticed that the diver started the survey from the middle of the pipe and the continued along its length.

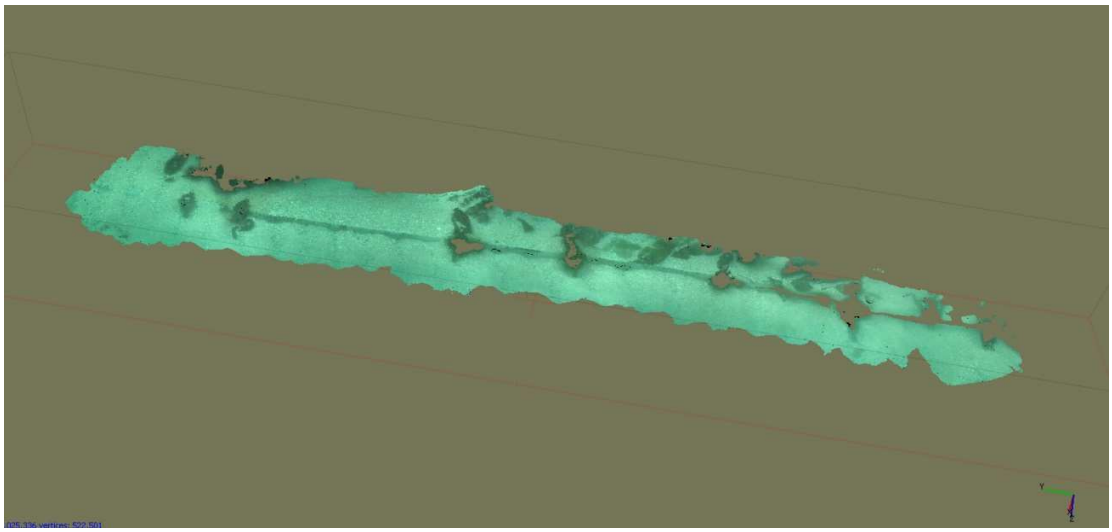

FIGURE 46 RECONSTRUCTION OF A PIPE FROM DOCUSCOOTER TEST

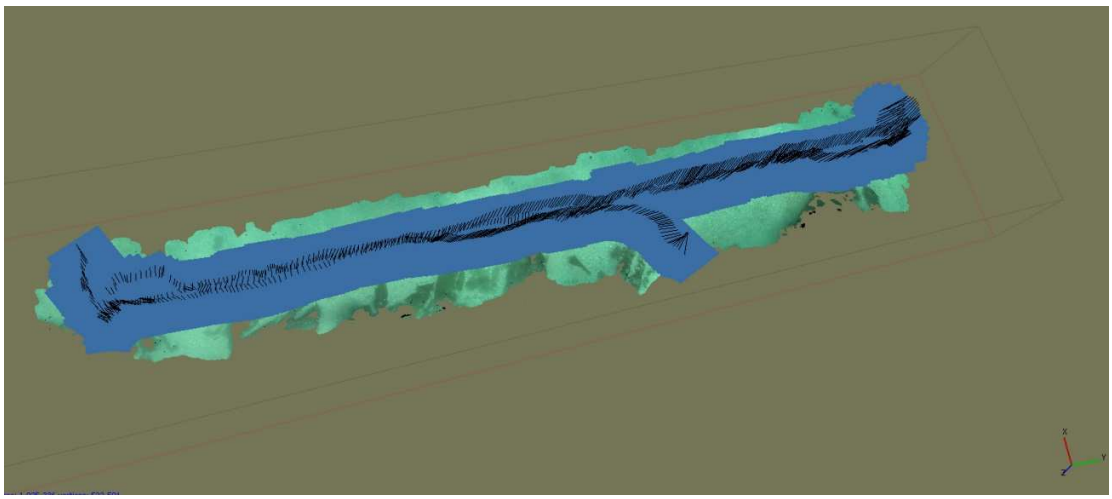

FIGURE 47 RECONSTRUCTION WITH TRAJECTORY FOLLOWED

5.2 BRAVE

5.2.1 BRAVE ARCHITECTURE

The second case study addressed in the complex underwater environment regards the design, development and test of the visual framework for a novel, partially-biomimetic vehicle, called BRAVe, with hybrid propulsion system for documentation purposes.

Biomimetic applied to underwater robotics has given rise to a number of innovative solutions in recent years and the study of fish-like robots has become one of the hot topics in the field of AUVs. In fact, biological inspirations from various species of fishes can help to complement disadvantages of the conventional screw propellers in terms of manoeuvrability and energy efficiency of the thrust.

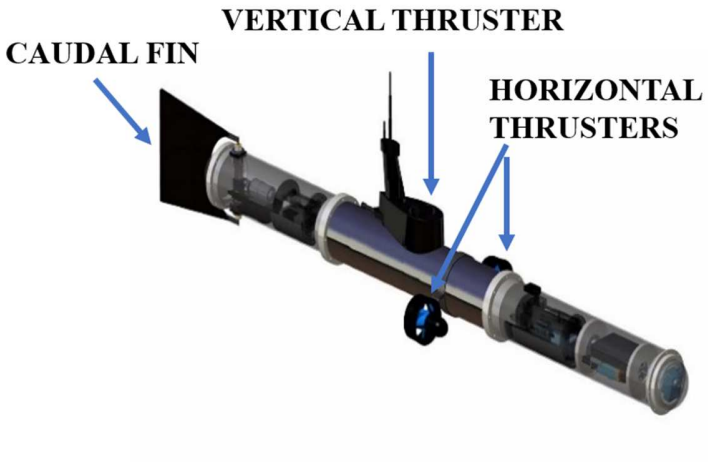

FIGURE 48 THRUSTERS CONFIGURATION

The hybrid propulsive system consists of two different type of thrusters (see Figure 48) and is designed to compensate the reciprocal flows in order to achieve a more efficient system overall: specifically, while the high-efficient biomimetic thruster, shaped like a caudal fin, is employed to navigate over large survey areas saving battery power and maximizing autonomy, the powerful horizontal screw propellers allow the vehicle to stop and accelerate quickly from and up to cruise speed. Moreover, they create a

propulsive redundancy that improves the vehicle movement capabilities allowing, for example, the agile yaw rotation required to adjust its heading and thus consent a complete and precise documentation of underwater remains within short distances, also thanks to the presence of a vertical thruster which allows the vehicle depth control.

BRAVe has a modular and fully reconfigurable architecture composed by an assembly of three compartments sharing some common features: a waterproofing case, a battery pack and a wired communication device designed to connect the various modules of the robot. As can be seen in Figure 49, going from prow to stern, there are:

1. the Vision Compartment (VC),
2. the Core Compartment (CC),
3. and the Propulsive Compartment (PC).

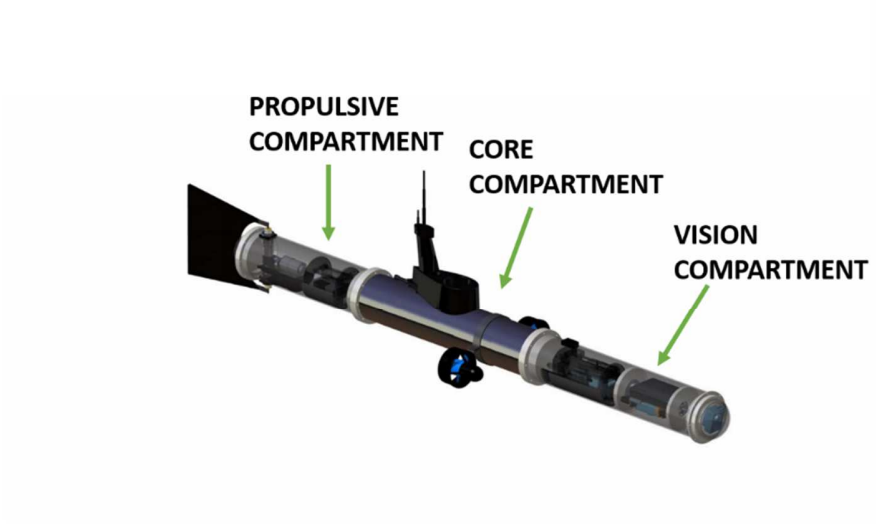

FIGURE 49 COMPARTMENTS ASSEMBLY

The Vision Compartment houses the vehicle sensory payload, namely a set of cameras arranged in a custom layout, the control devices required to run the acquisition-processing software and a dedicated battery pack capable of powering the electronics of the whole module during a mission. Since the VC function is to acquire pictures and

video from the outside environment, the compartment pressure hull has been made of perspex.

The Core Compartment houses the main controller of the vehicle, the vertical thruster, two BlueRobotics T200 horizontal thrusters and a battery pack fully capable to power the electronics and save data even in case of a major failure of the other modules. Since the weight and the buoyancy due to the Vision and Propulsive modules must be supported by the CC during out-of-water and underwater operations, its hull is made of aluminium. Two myRIO boards provided by National Instruments manage the vehicle during its mission. Specifically, one is exclusively dedicated to the Vision module in order to allocate more resources to the video and image processing algorithms, while the other manages almost all of the rest. For some specific aspects, also some microcontrollers are employed, for example to manage the switching on/off and data record of the cameras or to manage the pilot and the power of the Propulsive Compartment.

Finally, the Propulsive Compartment houses the thrust-generating and steering device: an oscillating foil, shaped like a caudal fin, moved by a custom, cam-like mechanism actuated by a gear motor. This solution complies with the propulsive hydrodynamic principles exploited by biological systems during swimming. Steering is possible thanks to a second mechanism capable of moving the neutral plane of oscillation outside the vertical plane of symmetry of the vehicle in order to create a centripetal thrust component. The PC pressure hull is made of perspex to allow a close inspection of the thrust generating device.

5.2.2 VISION MODULE

The Robot system apparatus (Figure 13), also in this case, is mainly composed of low cost action cameras. There are two main blocks: the frontal block, with the functions of navigation and detection of objects and obstacles, and the central one. The first consists of one GoPro Hero 3+ mounted with an inclination of 15° in order to have a wider view of the marine seabed; the camera has a wide-angle lens with an angle of view of 170°,

which shrinks to about 155° in water. The central block is composed of two GoPro Hero 3+: the adoption of a wide-angle lens also on these devices allowed to assemble the cameras at an angle of 30°, thus creating a stereo vision system capable to perform depth calculation during navigation. Figure 50 shows details of the assembly.

A third camera is included in the system. While the above vision blocks are mainly employed for documentation purposes, off-line mosaicking and 3D reconstruction, a USB camera "Genius wide cam f100", positioned at the centre of the compartment, is used to collect data for navigation purposes.

Calibration is an essential step to ensure high accuracy for the adopted 3D reconstruction technique. The stereo calibration has been carried out in water using a grid of 20 x 20 dots, mounted on an aluminium frame. The calibration has been performed at about 30 cm of distance, as shown in Figure 51. The refractive indices of air, the dome port and water, are not considered in this linear calibration model. A microprocessor controls all the cameras through the UART interface, with the duty of turning them on and off. Both the frontal and the central block are managed by a NI myRIO, which acquires also the analog videos from the video server Ip9100a through the TCP protocol.

Figure 52 shows how the areas covered by the cameras overlap on the seabed to collect images of the surveyed area.

FIGURE 50 VISION COMPARTMENT ASSEMBLY

FIGURE 51 LEFT AND RIGHT CAMERA CALIBRATION

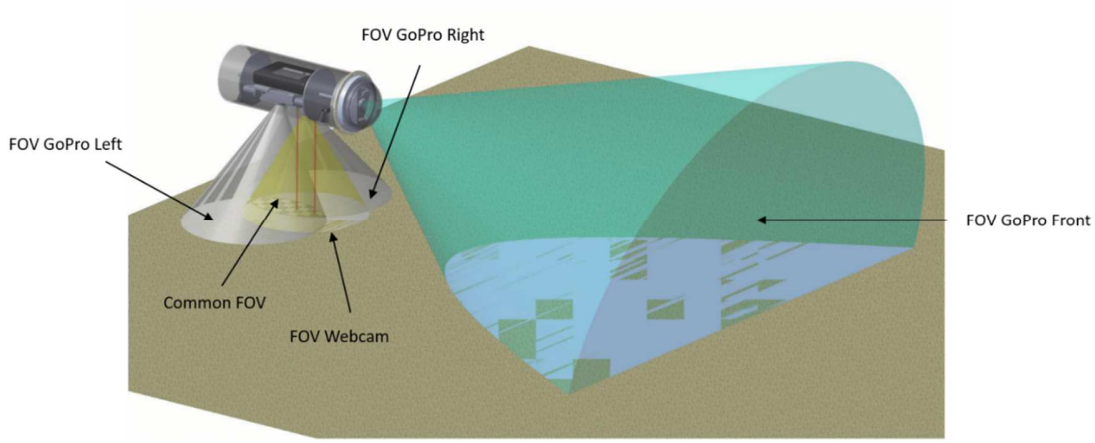

FIGURE 52 AREAS COVERED BY THE VISION DEVICES ABOARD

The Vision Module is also equipped with two laser diodes: one of them is mounted above the cameras while the other one is under them. During the Pre-Processing phase of the framework scheme (Figure 13) the red component of the RGB image of the seabed is analysed to detect and track the laser spots in order to compute the scene depth. The distance of the framed object is obtained depending on the distance in pixels of the lasers pointers into the image. In fact, as the relation between the distance in pixel and the depth is non-linear, a series of offline measurements from exactly-known positions have been carried out to obtain a calibration table. When the laser pointers are identified in the image, a linear interpolation between the founded measure and the nearest values in the calibration table is performed. Finally, in addition to the lasers, also an IMU and a depth-gauge are included in the Robot system apparatus. The information coming from these sensors are then fused to other data in the Robot Pose Interpretation phase (Figure 13) to improve the navigation algorithm.

The software for the Vision Module comprises of two parallel threads, one for the frontal block and the other for the central block. Each of them executes the following common phases:

- Camera initialization
- Frame acquisition
- Image filtering and object detecting
- Update of shared variables.

The image processing algorithms have been implemented through the use of NI Vision Development Module, an extension of LabVIEW.

5.2.3 EXPERIMENTAL TESTS

The visual framework applied in BRAVe was validated at a first instance during the archaeological mission at the Gulf of Baratti (Tuscany, Italy) on the site of Punta Tonnarelle with the cooperation of MIBACT (Italian Ministry of Heritage and Cultural Activities) and STMP (Servizi Tecnici Marittimi Portuali). This consists of an intact *dolium*, a broken one and numerous fragments scattered on the seabed. The site and the relative documentation is very important because the way Romans used to build these ships had a short life, not exceeding a century, so it well characterizes an historical period and the relative Romans' commercial routes of transportation. The experience provided by the archaeologists and the professional divers of STMP was essential to examine the optimal components position to achieve the best image resolution and to define the final design of the module.

The planned tests consisted on surveying the interested archaeological site following a straight line right over it, from north to south, until an interest artefact was found and later, a circular trajectory around the area closer to the sighted object. Figures 53-55 show the 3D reconstructions obtained with about a hundred images in different motions, using a SfM technique, considering the two cameras of the stereo system and the frontal camera independently for the reconstruction. In the figures, it is also possible to notice the trajectory followed.

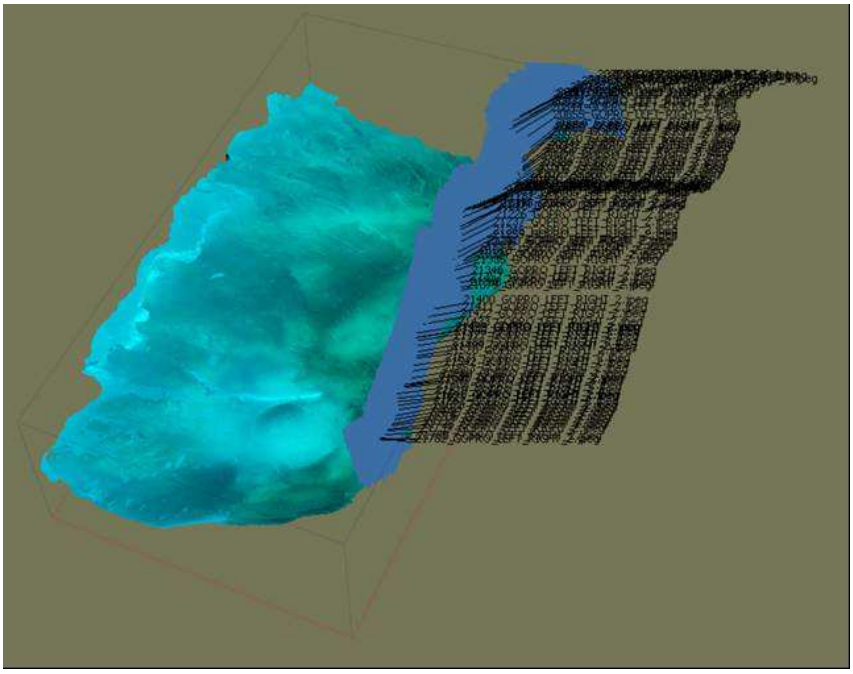

FIGURE 53 RIGHT CAMERA NORTH-SOUTH 3D RECUNSTRUCTION OBTAINED WITH 112 IMAGES

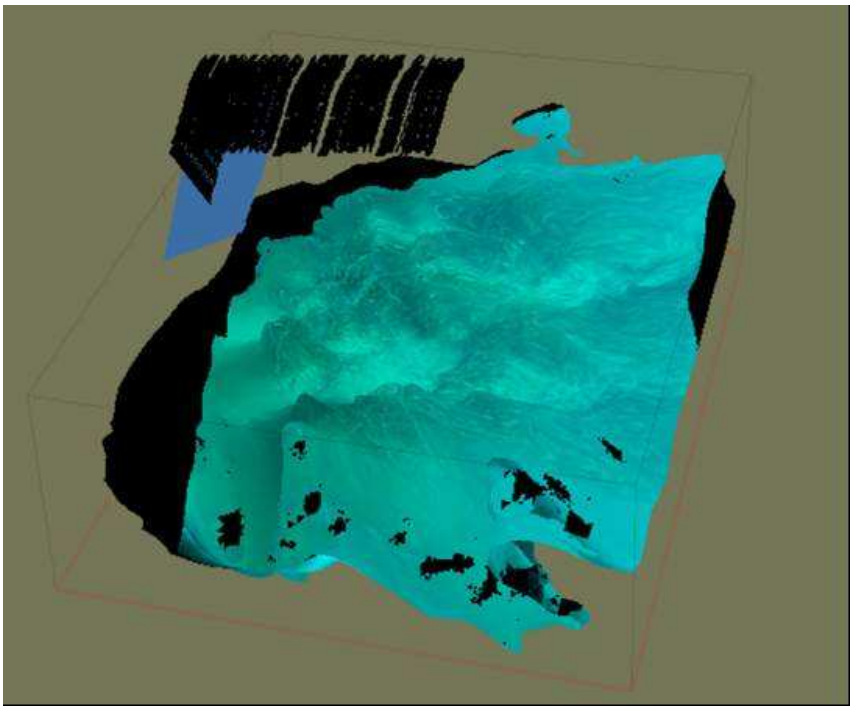

FIGURE 54 LEFT CAMERA NORTH-SOUTH 3D RECONSTRUCTION OBTAINED WITH 84 IMAGES

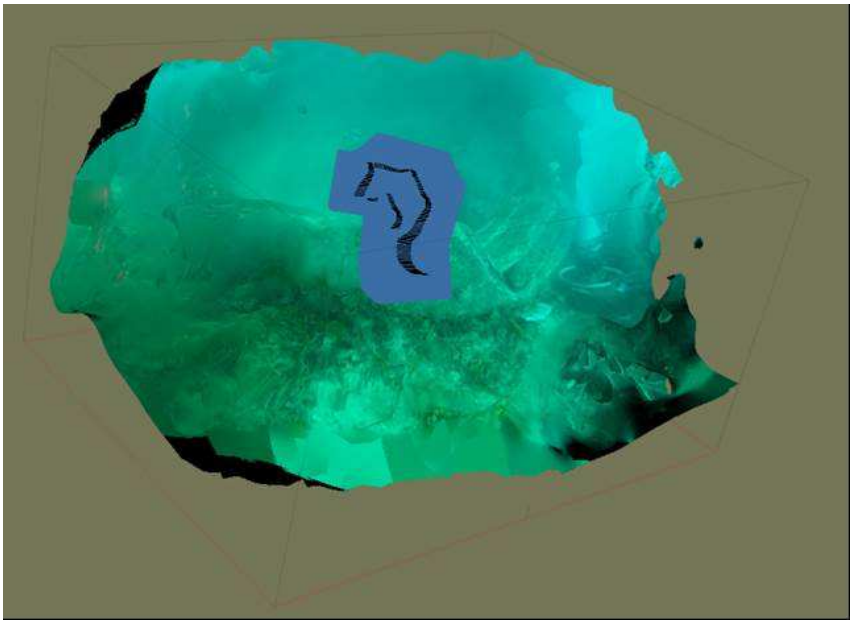

FIGURE 55 FRONTAL CAMERA CIRCULAR TRAJECTORY 3D RECONSTRUCTION OBTAINED WITH 156 IMAGES

Chapter 6

CONCLUSIONS AND FUTURE WORKS

This chapter concludes the work presented throughout this dissertation. It first summarizes the thesis by reviewing the contents described in each chapter. All aspects which have not been accomplished as well as some interesting future research issues are commented in this future work section.

6.1 REVIEW OF CHAPTERS

The characteristics of the map is one of the aspect to take into consideration in the choice of the SLAM algorithm to adopt in applications for delicate environments, partially structured or complex ones. The sensors used and the characteristics of the environment surely contribute to the computation of the map.

Chapter 2 presented the SLAM state of the art, analyzing 186 works published between 1977 and 2017. It was highlighted that it is not possible to indicate which is best algorithm to solve SLAM problems, especially in underwater environment where the conditions are even more challenging. The Chapter continued with a focus on the optical technologies most used for mapping industrial environments in the recent years, describing their advantages and disadvantages, with an excursus to those which have been adopted in the case studies presented.

Chapter 3 proposed the first case study developed for an industrial application. It described the vision system applied on an industrial robot, which is moving in the working scene and needs to individuate and avoid obstacles. The vision system consisted of a combination of three Time of Flight cameras, where a tailored algorithm was implemented in order to build the map of the environment in front of the robot and

allowing it to avoid obstacles while moving. Results confirmed a good behavior of the algorithm.

Chapter 4 described the second industrial case study. Here the extraction and the measurement of the scene information was obtained from a combination of a 3D stereo camera with a 2D RGB camera, thus merging information from different type of sensors in order to exploit the potentiality of each one and increasing the quality of the mapping. The system was applied to a robotic system composed of two collaborative robots. Results showed that the proposed system is able to extract and measure information at each instant allowing the robots to move and perform the predefined tasks.

Chapter 5 focused on applications in the complex underwater environment where SLAM algorithms could be applied, as described in Chapter 2. This chapter presented two technological devices for data gathering, whose vision systems were developed within the context of this thesis to map the navigation area. The 3D reconstructions computed from images gathered during real field missions by means of the SfM technique were shown.

6.2 REALIZATION OF OBJECTIVES AND MAIN CONTRIBUTIONS

The main objectives of this dissertation have been fulfilled during this three-years project. In this dissertation, we showed that with the proposed visual framework, composed of a sequence of computational blocks it is possible to solve the SLAM problem in different types of environment.

Specifically, the main contributions that this research offers can be listed in the following:

1. a framework for visual SLAM strategies that could have impact on various target groups, including marine robotics engineers, designer of industrial robotic systems, marine biologists, fishery authorities, natural reserve authorities, marine observatories, etc.

2. A Computer vision and 3D imaging approach to map a generic environment in order to employ the map in SLAM algorithms.
3. The combination of different sensors (optical and not), chosen accordingly to the environment in which the robot is moving, to extract the relevant data useful for SLAM techniques.
4. The validation of the framework in two delicate and challenging scenarios with very different characteristics: industrial and underwater.

6.3 FUTURE WORKS

The development of a research project always pushes the discovery of new problems as well as new interesting research projects. Future works will consider the adoption in complete SLAM algorithms, described in Chapter 2, of the information and the measures extracted at each instant t, t+1, etc., with the systems and methods implemented and described in this work concerning the mapping phase. The combination of the information coming from the vision systems could considerably help a robot to be fully autonomous. The application of these techniques could really improve the navigation and localization of the robot in delicate partially structured and complex environments and so they are interesting to be studied and tested in other situations and applications.

REFERENCES

[1] H. Choset, K. M. Lynch, S. Hutchinson, G. Kantor, W. Burgard, L. E. Kavraki and S. Thrun, "Principles of robot motion: Theory, Algorithms, and Implementations", *MIT Press*, Boston, 2005

[2] T. Bailey, H. Durrant-Whyte, "Simultaneous Localisation and Mapping (SLAM): Part II", *Robotics and Automation Magazine*, September, 2006.

[3] S. Thrun, "Robotic mapping: a survey", In *Exploring artificial intelligence in the new millennium*, Gerhard Lakemeyer and Bernhard Nebel (Eds.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2003, pp.1-35.

[4] H. Durrant-Whyte, T. Bailey, "Simultaneous Localisation and Mapping (SLAM): Part I", *IEEE ROBOTICS AND AUTOMATION MAGAZINE*, 2006

[5] R. Smith, M. Self, P. Cheeseman, "Estimating uncertain spatial relationships in robotics", in *Autonomous Robot Vehicles*, I. Cox and G. Wilfong,] editors, Springer Verlag, 1990, pp 167–193.

[6] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit, "FastSLAM: A factored solution to simultaneous localization and mapping", in *Proc. of the National Conference on Artificial Intelligence (AAAI)*, Edmonton, Canada, 2002, pp. 593–598.

[7] R. Triebel, P. Pfaff, W. Burgard, "Multi-level surface maps for outdoor terrain mapping and loop closing", in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2006, pp. 2276-2282.

[8] G. Grisetti, C. Stachniss, W. Burgard, "Improved techniques for grid mapping with raoblackwellized particle filters", *IEEE Trans. on Robotics*, Vol. 23, no.1, pp. 34–46, Feb. 2007.

[9] D. Hahnel, W. Burgard, D. Fox, S. Thrun, “An efficient FastSLAM algorithm for generating maps of large-scale cyclic environments from raw laser range measurements”, in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, Las Vegas, NV, USA, 2003, pp. 206–211.

[10] K. Konolige, “A gradient method for realtime robot control”, in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2000, Vol.1, pp. 639-646.

[11] K. Konolige, J. Bowman, J. D. Chen, P. Mihelich, M. Calonder, V. Lepetit, P. Fua, “View-based maps”, in *International Journal of Robotics Research (IJRR)*, Vol. 29, no. 10, 2010.

[12] S. Thrun, W. Burgard, D. Fox, “Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)”, *The MIT Press*, 2005.

[13] P.M. Newman, J.J. Leonard, “Consistent, convergent, and constant-time SLAM”, *International Joint Conference on Artificial Intelligence (IJCAI)*, Acapulco, Mexico, 2003, pp. 1143-1150.

[14] J.A. Castellanos, “Mobile robot localization and map-building: a multisensor fusion approach”, Ph.D. dissertation, Universidad de Zaragoza, Spain, 1998.

[15] K.S. Chong, L. Kleeman, “Feature-based mapping in real, large scale environments using an ultrasonic array”, in *International Journal of Robotics Research*, Vol. 18, no. 2, 1999, pp. 3-19.

[16] A.J. Davison, D.W. Murray, “Mobile robot localisation using active vision”, in *Proceedings of the Fifth European Conference on Computer Vision*, Freiburg, Germany, 1998, pp. 809–825.

[17] H.F. Durrant-Whyte, M.W.M.G. Dissanayake, P.W. Gibbens, “Toward deployments of large scale simultaneous localisation and map building (slam) systems”, in *Proceedings of the Ninth International Symposium of Robotics Research*, Snowbird, UT, 1999, pp. 121–127.

[18] Y.D. Kwon, J.S. Lee, “A stochastic map-building method for mobile robot using 2-D laser range finder”, in *Autonomous Robots*, vol. 7, 1999, pp. 187-200.

[19] P. Skrzypczyski, “Simultaneous localization and mapping: A feature based probabilistic approach”, in *International Journal of Applied Mathematics and Computer Science*, vol. 19, no. 4, December 2009, pp. 575-588.

[20] R. Smith, M. Self, P. Cheeseman, “A stochastic map for uncertain spatial relationships”, in *The Fourth International Symposium of Robotics Research*, R. Bolles and B. Roth editors, The MIT Press, Cambridge, 1988, pp. 467-474.

[21] K. Konstantin, A. Maltsev, M. Sobolev, “Recurrent neural network and extended Kalman filter in SLAM problem”, *3rd IFAC International Conference on Intelligent Control and Automation Science*, September 2-4, 2013, Chengdu, China, pp. 23-26.

[22] G. Dissanayake, P. Newman, S. Clark, H.F. Durrant-Whyte, M. Csorba, “A solution to the simultaneous localisation and map building (SLAM) problem”, in *IEEE Transactions of Robotics and Automation*, vol. 17, 2001, pp.229-241.

[23] S. J. Julier, J. K. Uhlmann, “New extension of the Kalman filter to nonlinear systems”, in *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 3068, 1997, pp. 182-193.

[24] E.A. Wan, R. van der Merwe, “The Unscented Kalman Filter”, in *Kalman Filtering and Neural Networks* (ed S. Haykin), John Wiley & Sons, Inc., New York, USA, 2001.

[25] M. Barisic, A. Vasilijevic, D. Nad, “Sigma-point unscented kalman filter used for auv navigation”, in *Proceedings of the 20th Mediterranean conference on control and automation*, Barcelona, Spain, 2012, pp. 1365-1372.

[26] M. Karimi, M. Bozorg, A., "A comparison of DVL/INS fusion by UKF and EKF to localize an autonomous underwater vehicle", in *Proceedings of the first RSI/ISM international conference on robotics and mechatronics (ICRoM),* Tehran, Iran, 2013.

[27] B. Allotta, A. Caiti, R. Costanzi, F. Fanelli, D. Fenucci, E. Meli, A. Ridolfi, "A new AUV navigation system exploiting unscented Kalman filter", in *Ocean Eng*, vol. 113, 2016, pp. 121-132.

[28] J.E. Guivant, E.M. Nebot, "Optimization of the Simultaneous Localization and Map-Building Algorithm for Real-Time Implementation", in *IEEE Transactions on Robotics and Automation*, vol. 17, no. 3, 2001.

[29] S.Thrun, D. Koller, Z. Ghahramani, H. Durrant-Whyte, A. Y. Ng, "Simultaneous Mapping and Localization With Sparse Extended Information Filters: Theory and Initial Results", in *Proceedings of the Fifth International Workshop on Algorithmic Foundations of Robotics*, Nice, France, 2002.

[30] P. Newman, "On the Structure and Solution of the Simultaneous Localisation and Map Building Problem", PhD dissertation, University of Sydney, 2000.

[31] M. Deans, M. Hebert, "Invariant filtering for simultaneous localization and mapping", *IEEE International Conference on Robotics and Automation*, 2000, pp. 1042-1047.

[32] M. Csorba, "Simultaneous Localization and Map Building", PhD dissertation, Univ. of Oxford, 1997.

[33] F. Lu, E. Milios, "Globally consistent range scan alignment for environment mapping", in *Autonomous Robots*, vol. 4, pp. 333–349, 1997.

[34] G. Evensen, "The ensemble Kalman filter: Theoretical formulation and practical implementation", in *Ocean Dynamics*, vol. 53, pp. 343-367, 2003.

[35] C. Cadena, J. Neira, “SLAM in O(logn) with the Combined Kalman-Information Filter”, in *Robotics and Autonomous Systems*, Vol. 58, no. 11, pp. 1207-1219, 2010.

[36] S.J. Julier, J.K. Uhlmann, “A counter example to the theory of simultaneous localization and map building”, in *Proc. of the Int. Conference on Robotics and Automation ICRA*, Seoul, Korea, 2001, pp. 4238-4243.

[37] J.A. Castellanos, J. Neira, J.D. Tardos, “Limits to the consistency of EKF based SLAM”, in *Proc. of the 5th IFAC/EURON Symposium on Intelligent and Autonomous Vehicles IAV,* Lisboa, Portugal, 2004.

[38] U. Frese, “A discussion of simultaneous localization and mapping”, in *Autonomous Robots,* vol. 20, no. 1, pp.25-42, 2006.

[39] A. Martinell, N. Tomatis, R. Siegwart, “Some results on SLAM and the closing of the loop problem”, in *Proc. of the IEEE/RSJ Int. Conference on Intelligent Robots Systems IROS*, Edmonton, AB, Canada, 2005, pp. 334-339.

[40] M.W.M.G. Dissanayake, S. Huang, “Convergence and consistency analysis for extended Kalman filter based SLAM”, in *IEEE Transactions on Robotics and Automation*, vol. 23, no. 5, pp. 1036-1049, 2007.

[41] T. Bailey, J. Nieto, J. Guivant, M. Stevens, E. Nebot, “Consistency of EKF SLAM algorithm”, in *Proc. of the IEEE Int. Conf. on Intelligent Robot Systems IROS*, Beijing, China, 2006, pp. 3562-3568.

[42] J.A. Castellanos, J.M.M. Montiel, J. Neira, J.D. Tardòs, “The SPmap: A probabilistic framework for simultaneous localization and map building”, in *IEEE Transactions on Robotics and Automation*, vol. 15, no. 5, pp. 948–953, 1999.

[43] J.J. Leonard, H.J.S. Feder, "A computationally efficient method for large-scale concurrent mapping and localization", in *Proceedings of the Ninth International Symposium on Robotics Research*, J. Hollerbach and D. Koditschek, editors, Salt Lake City, Utah, 1999.

[44] M.A. Paskin, "Thin junction tree filters for simultaneous localization and mapping, in *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI)*, Acapulco, Mexico, 2003.

[45] J. D. Tardós, J. Neira, P. Newman, J. Leonard, "Robust mapping and localization in indoor environments using sonar data", in *International Journal of Robotics Research*, vol. 21, pp. 311-330, 2002.

[46] L. Moreno, S. Garrido, D. Blanco, M. L. Muñoz, "Differential evolution solution to the SLAM problem", in *Robotics and Autonomous Systems*, vol. 57, pp. 441-450, 2009.

[47] J. Liu, R. Chen, "Sequential monte carlo methods for dynamic systems", in *Journal of the American Statistical Association*, vol.93, pp. 1032-1044, 1998.

[48] A. Doucet, J.F.G. de Freitas, N.J. Gordon, editors, "Sequential Monte Carlo Methods In Practice", Springer, New York, 2001.

[49] S. Thrun, "Particle Filters in Robotics", in *Proceedings of the 17th Annual Conference on Uncertainty in AI (UAI),* 2002.

[50] G. Dissanayake, P. Newman, S. Clark, H.F. Durrant-Whyte, M. Csorba, "An experimental and theoretical investigation into simultaneous localisation and map building (SLAM)", in Lecture Notes in Control and Information Sciences, *Experimental Robotics VI*, Springer, pp. 265-2742000.

[51] J.J. Leonard, H.F. Durrant-Whyte, I.J. Cox, "Dynamic map building for an autonomous mobile robot", in *International Journal of Robotics Research*, vol. 11, no. 4, 1992.

[52] P. Moutarlier, R. Chatila, “An experimental system for incremental environment modeling by an autonomous mobile robot”, in *Experimental Robotics I: The First International Symposium Montreal*, pp. 327-346, 1990

[53] R. Smith, M. Self, P. Cheeseman, “Estimating uncertain spatial relationships in robotics”, in *Autonomous Robot Vehicles*, Springer, pp.167-193, 1990.

[54] A Doucet, N. de Freitas, K. Murphy, S. Russell, “Rao-Blackwellised particle filtering for dynamic Bayesian networks”, in *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence,* Craig Boutilier and Moisés Goldszmidt (Eds.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 176-183, 2000.

[55] M. Montemerlo, W. Whittaker, S. Thrun, “Conditional particle filters for simultaneous mobile robot localization and people-tracking”, in *IEEE International Conference on Robotics and Automation (ICRA)*, 2002, pp. 695-701.

[56] K. Murphy, “Bayesian map learning in dynamic environments”, in *Proceedings of the 12th International Conference on Neural Information Processing Systems*, S. A. Solla, T. K. Leen, and K. Müller (Eds.), MIT Press, Cambridge, MA, USA, pp. 1015-1021, 1999.

[57] T. Bailey, J. Nieto, E. Nebot, “Consistency of the FastSLAM algorithm”, in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2006, pp. 424–429.

[58] K. R. Beevers, W. H. Huang, “Fixed-lag Sampling Strategies for Particle Filtering SLAM”, *Proceedings of the IEEE International Conference on Robotics and Automation,* 2007, pp. 2433-2438.

[59] A. Doucet, M. Briers, S. Sénécal, “Efficient block sampling strategies for sequential Monte Carlo methods”, in *Journal of Computational and Graphical Statistics*, vol. 15, no. 3, pp. 693-711, 2006.

[60] W.R. Gilks, C. Berzuini, “Following a moving target - Monte Carlo inference for dynamic Bayesian models”, in *Journal of the Royal Statistical Society B*, vol. 63, no. 1, pp. 127-146, 2001.

[61] M. Montemerlo, S. Thrun, “Simultaneous localization and mapping with unknown data association using FastSLAM”, in *Proc. of the Int. Conference on Intelligent Robotics and Automation*, Taipei, Taiwan, vol. 2, 2003, pp. 1985-1991.

[62] D. Crisan, A. Doucet, “A survey of convergence results on particle filtering methods for practitioners”, in *IEEE Trans. Signal Process*, vol. 50, no. 3, pp. 736-746, 2002.

[63] R. Martinez-Cantin, “Active map learning for robots: insights into statistical consistency”, PhD dissertation, University of Zaragoza, 2008.

[64] M.A. Paskin, “Thin junction tree filters for simultaneous localization and mapping”, in *Proceedings of the 18th international joint conference on Artificial intelligence*, Acapulco, Mexico, 2003, pp. 1157-1164.

[65] G. Grisetti, C. Stachniss, W. Burgard, “Non-linear constraint network optimization for efficient map learning”, in *IEEE Transactions on Intelligent Transportation Systems*, Vol. 10, no. 3, pp. 428-439, 2009.

[66] A. P. Dempster, N. M. Laird, D. B. Rubin, “Maximum likelihood from incomplete data via the em algorithm”, in *Journal of the Royal Statistical Society: Series B*, vol. 39, no. 1, pp. 1-38, 1977.

[67] S. Thrun, C. Martin, Y. Liu, D. Hahnel, R. E. Montemerlo, D. Chakrabarti, W. Burgard, “A Real-Time Expectation Maximization Algorithm for Acquiring Multi-Planar Maps of Indoor Environments with Mobile Robots”, in *IEEE Transactions on Robotics and Automation*, vol. 20, no. 3, 2004.

[68] S. Le Corff, G. Fort, E. Moulines, “Online Expectation Maximization algorithm to solve the SLAM problem”, in *IEEE Statistical Signal Processing Workshop (SSP)*, 2011, pp. 225-228

[69] D. Hahnel, D. Schulz, W. Burgard, “Map building with mobile robots in populated environments”, in *International Conference on Intelligent Robots and Systems*, 2002, pp. 496–501.

[70] D. H¨ahnel, R. Triebel, W. Burgard, S. Thrun, “Map building with mobile robots in dynamic environments”, in *International Conference on Robotics and Automation*, 2003, pages 1557–1563.

[71] C. Bibby, I. Reid, “Simultaneous localisation and mapping in dynamic environments (SLAMIDE) with reversible data association”, in *Proceedings of Robotics: Science and Systems Conference,* Atlanta, 2007.

[72] J.G. Rogers III, A. J. B. Trevor, C. Nieto-Granda, H. I. Christensen, “SLAM with Expectation Maximization for Moveable Object Tracking”, in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Taipei, Taiwan, 2010.

[73] N. Mitsou, C.Tzafestas, “Maximum Likelihood SLAM in Dynamic Environments Conference”, in *Proceedings of the 19th IEEE International Conference on Tools with Artificial Intelligence*, Vol. 01, Pp.152-156, 2007.

[74] M. Lundgren, “Bayesian filtering for automotive applications,” Ph.D. dissertation, Chalmers University of Technology, 2015.

[75] M. Fatemi, L. Svensson, L. Hammarstrand, et al., "Variational Bayesian EM for SLAM", in *IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Cancun, Mexico, 2015, pp. 501-504.

[76] M. Golfarelli, D. Maio, S. Rizzi, “Elastic correction of dead-reckoning errors in map building”, in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Victoria, Canada, 1998, pp. 905-911.

[77] K. Konolige, J. Bowman, J. D. Chen, P. Mihelich, M. Calonder, V. Lepetit, P. Fua, “View-based maps”, in *International Journal of Robotics Research (IJRR)*, vol. 29, no. 8, pp. 941-957, 2010.

[78] G. Grisetti, R. Kummerle, C. Stachniss, W. Burgard, “A tutorial on graph-based slam”, IEEE Intelligent Transportation Systems Magazine Volume: 2, Issue: 4, 2010

[79] M. Bosse, P.M. Newman, J.J. Leonard, S. Teller, “An ATLAS framework for scalable mapping”, in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, pages 1899-1906, Taipei, Taiwan, 2003.

[80] C. Estrada, J. Neira, J.D. Tardos, “Hierachical slam: Real-time accurate mapping of large environments”, in *IEEE Transactions on Robotics*, vol. 21, no. 4, pp. 588-596, 2005.

[81] A. Nuchter, K. Lingemann, J. Hertzberg, H. Surann, “6d SLAM with approximate data association”, in *Proc. of the 12th Int. Conference on Advanced Robotics (ICAR)*, pp. 242-249, 2005.

[82] J.S. Gutmann, K. Konolige, “Incremental mapping of large cyclic environments”, in *Proc. of the IEEE Int. Symposium on Computational Intelligence in Robotics and Automation (CIRA)*, 1999, pp. 318-325.

[83] J. Folkesson, H.I. Christensen, “Graphical SLAM: A self-correcting map”, in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, New Orleans, USA, 2004, pp. 383-390.

[84] J. Folkesson, H.I. Christensen, “Robust SLAM”, in *the 5th Symposium on Intelligent Autonomous Vehicles*, Elsevier, 2004.

[85] E. Olson, J. Leonard, S. Teller, “Fast iterative optimization of pose graphs with poor initial estimates”, in *Proc. of the IEEE Int. Conf. On Robotics & Automation (ICRA)*, pp. 2262–2269, 2006.

[86] K. Konolige, “Large-scale map-making”, in *Proceedings of the AAAI National Conference on Artificial Intelligence*, pp. 457–463, San Jose, CA. 2004.

[87] M. Montemerlo, S. Thrun, “Large-scale robotic 3-d mapping of urban structures”, in *Experimental Robotics IX: The 9th International Symposium on Experimental Robotics,* Springer, pp. 141-150, 2006.

[88] S. Thrun, M. Montemerlo, “The graph SLAM algorithm with applications to large-scale mapping of urban structures”, in *Int. Journal of Robotics Research,* vol. 25, pp. 403-429, 2006.

[89] A. Howard, M.J. Mataric, G. Sukhatme, “Relaxation on a mesh: a formalism for generalized localization”, in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2001

[90] U. Frese, P. Larsson, T. Duckett, “A multilevel relaxation algorithm for simultaneous localisation and mapping”, in *IEEE Transactions on Robotics*, vol. 21, no. 2, pp.1-12, 2005.

[91] F. Dellaert M. Kaess, “Square root SAM: Simultaneous location and mapping via square root information smoothing”, in *Int. Journal of Robotics Research*, pp. 1181-1203, 2006.

[92] M. Kaess, A. Ranganathan, F. Dellaert, “iSAM: Fast incremental smoothing and mapping with efficient data association”, in *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, Rome, Italy, 2007.

[93] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, F. Dellaert, “iSAM2: Incremental Smoothing and Mapping Using the Bayes Tree”, in *the International Journal of Robotics Research*, Vol 31, no. 2, pp. 216-235, 2011.

[94] G.N. DeSouza, A.C. Kak, “Vision for Mobile Robot Navigation: A Survey”, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 2, pp. 237-267, 2002.

[95] A.C. Kak, G.N. de Souza, “Robotic Vision: What Happened to the Visions of Yesterday?”, in *Proc. Of 16th International Conference on Pattern Recognition*, pp. 839-847, 2002.

[96] J. Abascal, E. Lazkano, B. Sierra, “Behavior-Based indoor navigation”, in *Ambient Intelligence for Scientific Discovery*, Yang Cai (Ed.). Springer-Verlag, Berlin, Heidelberg, pp. 263-285, 2005

[97] S. Se, D. G. Lowe, J. J. Little, “Vision-based global localization and mapping for mobile robots,” in *IEEE transactions on robotics*, vol. 21, no. 3, pp. 364–375, 2005.

[98] F. Bonin-Font, A. Ortiz, G. Oliver, “Visual Navigation for Mobile Robots: a Survey”, in *Journal of Intelligent and Robotic Systems*, vol. 53, no. 3, pp. 263-296, 2008.

[99] D. Burschka, G. D. Hager, “V-gps(slam): Vision-based inertial system for mobile robots,” in *IEEE International Conference on Robotics and Automation*, New Orleans, LA, USA, April 2004.

[100] P. Lothe, S. Bourgeois, F. Dekeyser, E. Royer, M. Dhome, “Monocular SLAM Reconstructions and 3D City Models: Towards a Deep Consistency”, in *Computer Vision, Imaging and Computer Graphics. Theory and Applications: International Joint Conference, VISIGRAPP 2009, Lisboa, Portugal,* Springer, pp. 201-214, 2010.

[101] A. Cumani, S. Denasi, A. Guiducci, G. Quaglia, “Integrating monocular vision and odometry for SLAM”, in *WSEAS Transactions on Computers*, vol. 3, no. 3, pp. 625-630, 2004.

[102] G. Bleser, M. Becker, D. Stricker, “Real-time vision-based tracking and reconstruction,” in *Journal of real-time image processing*, vol. 2, pp. 161-175, 2007.

[103] J. Montiel, J. Civera, A. J. Davison, "Unified inverse depth parametrization for monocular slam," in *In Proceedings of Robotics: Science and Systems,* Philadelphia, USA, August 2006.

[104] J. Shi, C. Tomasi, "Good Features to Track", in *Proc. of IEEE Intl Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 593-600, 1994.

[105] A. J. Davison, I. D. Reid, N. D. Molton, O. Stasse, "Monoslam: Real-time single camera slam," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, 2007.

[106] P. Pinies, T. Lupton, S. Sukkarieh, J. D. Tardos, "Inertial aiding of inverse depth slam using a monocular camera," in *IEEE International Conference on Robotics and Automation*, April 2007, pp. 2797–2802.

[107] J. Engel, T. Schöps, D. Cremers, "LSD-SLAM: Large-Scale Direct Monocular SLAM", in: Fleet D., Pajdla T., Schiele B., Tuytelaars T. (eds) *Computer Vision – ECCV 2014*. ECCV 2014. Lecture Notes in Computer Science, vol 8690. Springer, Cham, 2014.

[108] A. Davison, "Real-time simultaneous localisation and mapping with a single camera," in *Int. Conf. on Computer Vision*, 2003, pp. 1403-1416.

[109] A. Nemra N. Aouf, "Robust Airborne 3D Visual Simultaneous Localization and Mapping with Observability and Consistency Analysis", in *Journal of Intelligent and Robotic Systems*, vol. 55, pp. 345-376, January 2009.

[110] X.Hai-Xia, Z. Wei, Z. Jiang, "3D Visual SLAM with a Time-of-Flight Camera", in *IEEE Workshop on Signal Processing Systems (SiPS)*, 2015

[111] D. Caruso, J. Engel, D. Cremers, "Large-Scale Direct SLAM for Omnidirectional Cameras", in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015.

[112] D. Gutierrez, A. Rituerto, J. Montiel, and J. J. Guerrero, "Adapting a real-time monocular visual slam from conventional to omnidirectional cameras," in *IEEE International Conference on Computer Vision Workshops (ICCV Workshops),* 2011, pp. 343-350.

[113] S. Lacroix, A. Mallet, I.-K. Jung, T. Lemaire, J. Sola, "Vision-based slam," SLAM Summer School 2006, Oxford, 2006.

[114] D. Schleicher, L. M. Bergasa, M. Ocana, R. Barea, E. Lopez, "Realtime hierarchical stereo Visual SLAM in largescale environments", in *Robotics and Autonomous Systems*, vol. 58, no. 8, pp.991-1002, August 2010.

[115] A. Cumani, S. Denasi, A. Guiducci, and G Quaglia, "Robot localisation and mapping with stereo vision", in *WSEAS Transactions on Circuits and Systems*, vol. 3, no. 10, pp. 2116-2121, 2004.

[116] S. Atiya, G.D. Hager, "Real-Time Vision-Based Robot Localization," in *IEEE Trans. Robotics and Automation*, vol. 9, no. 6, pp. 785-800, Dec. 1993.

[117] H.I. Christensen, N.O. Kirkeby, S. Kristensen, L. Knudsen, "Model-Driven Vision for In-Door Navigation," in *Robotics and Autonomous Systems*, vol. 12, pp. 199-207, 1994.

[118] M. Hashima, F. Hasegawa, S. Kanda, T. Maruyama, T. Uchiyama, "Localization and Obstacle Detection for a Robot for Carrying Food Trays," *Proc. IEEE Int'l Conf. Intelligent Robots and Systems*, pp. 345-351, Sept. 1997.

[119] A.C. Kak, K.M. Andress, C. Lopez-Abadia, M. S. Carroll, J.R. Lewis, "Hierarchical Evidence Accumulation in the PSEIKI System and Experiments in Model-Driven Mobile Robot Navigation," *Uncertainty in Artificial Intelligence*, M. Henrion, R. Shachter, L.N. Kanal, and J. Lemmer, eds. pp. 353-369, Elsevier, 1990.

[120] D.J. Kriegman, E. Triendl, T.O. Binford, "Stereo Vision and Navigation in Buildings for Mobile Robots," in *IEEE Trans. Robotics and Automation*, vol. 5, no. 6, pp. 792-803, 1989.

[121] L. Matthies, S.A. Shafer, "Error Modeling in Stereo Vision," in *IEEE J. Robotics and Automation*, vol. 3, no. 3, pp. 239-248, 1987.

[122] K. Sutherland, W. Thompson, "Localizing in Unstructured Environments: Dealing with the Errors," in *IEEE Trans. Robotics and Automation*, vol. 10, no. 6, pp. 740-754, Dec. 1994.

[123] L. Matthies, T. Balch, B. Wilcox, "Fast Optical Hazard Detection for Planetary Rovers Using Multiple Spot Laser Triangulation," in *Proc. IEEE Int'l Conf. Robotics and Automation*, vol. 1 pp. 859-866, Apr. 1997.

[124] R. Simmons, E. Krotkov, L. Chrisman, F. Cozman, R. Goodwin, M. Hebert, L. Katragadda, S. Koenig, G. Krishnaswamy, Y. Shinoda, W. Whittaker, P. Klader, "Experience with Rover Navigation for Lunar-Like Terrains," in *Proc. IEEE Int'l Conf. Intelligent Robots and Systems*, pp. 441-446, Aug. 1995.

[125] R. Sim, P. Elinas, M. Griffin, A. Shyr, J. J. Little, "Design and Analysis of a Framework for Real-Time Vision-Based SLAM Using Rao-Blackwellised Particle Filters", in *Proc. of the 3rd Canadian Conference on Computer and Robotic Vision*, 2006.

[126] R. Sim, J. J. Little, "Autonomous Vision-Based Exploration and Mapping Using Hybrid Maps and Rao-Blackwellised Particle Filters", in *Proc. of IEEE Int'l Conf. on Intelligent Robots and Systems (IROS),* 2006.

[127] A. Gil, O. Reinoso, M. Ballesta, M. Julia, "Multirobot visual SLAM using a Rao-Blackwellized particle filter," in *Robotics and Autonomous Systems*, vol. 58, no. 1, pp.68-80, January 2010.

[128] V. Pradeep, G. Medioni, J. Weiland, "Visual loop closing using multi-resolution SIFT grids in metric-topological SLAM," in *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1438-1445, June 2009.

[129] D. Schleicher, L. M. Bergasa, R. Barea, E. Lopez, M. Ocaa, “Real-Time Simultaneous Localization and Mapping Using a Wide-Angle Stereo Camera”, in *Proc. of IEEE Workshop on Distributed Intelligent Systems: Collective Intelligence and Its Applications* (DIS’06), pp. 55–60, 2006.

[130] T. Lemaire, C. Berger, IK. Jung, et al., “Vision-Based SLAM: Stereo and Monocular Approaches”, in *International Journal of Computer Vision*, vol. 74, no.3, pp. 343-364, 2007.

[131] C. Mei, G. Sibley, M. Cummins, P.Newman, I. Reid, “RSLAM: A System for Large-Scale Mapping in Constant-Time Using Stereo”, in *International Journal of Computer Vision*, vol. 94, no. 2, pp. 198-214, June 2010.

[132] Doaa M. A.-Latif, Mohammed A.-Megeed Salem, H. Ramadan, Mohamed I. Roushdy, “3D Graph-based Vision-SLAM Registration and Optimization”, in International Journal of circuits, systems and signal processing, vol.8, 2014.

[133] N. Engelhard, F. Endres, J. Hess, J. Sturm, W. Burgard, “Realtime 3D visual SLAM with a hand-held RGB-D camera”, in *Proceedings of the RGB-D workshop on 3D perception in robotics* at the European robotics forum, Vasteras, Sweden, number c, 2011.

[134] R. I. Hartley, B. Triggs, P.F. McLauchlan, A. W. Fitzgibbon, “Bundle adjustment - a modern synthesis”, Lecture Notes in Computer Science, pp. 298-372, 1999.

[135] E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser, P. Sayd, “Real time localization and 3d reconstruction,” in *Computer Vision and Pattern Recognition*, IEEE Computer Society Conference on, vol. 1, 2006, pp. 363-370.

[136] G. Klein, D. Murray, “Parallel tracking and mapping for small AR workspaces,” in *IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, Nara, Japan, November 2007, pp. 225–234.

[137] A.W. Fitzgibbon, A. Zisserman, “Automatic Camera Recovery for Closed or Open Image Sequences,” in *Proc. European Conf. Computer Vision*, pp. 311-326, June 1998.

[138] M. Pollefeys, R. Koch, L.V. Gool, “Self-Calibration and Metric Reconstruction in Spite of Varying and Unknown Internal Camera Parameters,” *Proc. Sixth International Conf. Computer Vision*, pp. 90-96, 1998.

[139] A. J. Davison, I. D. Reid, N. D. Molton, O. Stasse, “MonoSLAM: Real-Time Single Camera SLAM”, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.29, no. 6, June 2007.

[140] R.Mur-Artal, J. M. M. Montiel, J. D. Tardòs, “ORB-SLAM: A Versatile and Accurate Monocular SLAM System”, in *IEEE Transactions on Robotics*, vol. 31, no. 5, October 2015.

[141] D.G. Lowe, “Object Recognition From Local Scale Invariant Features”, in *Proc. of International Conference on Computer Vision (ICCV)*, pp. 1150-1157, 1999.

[142] P. Henry, M. Krainin, E.Herbst, X. Ren, D. Fox, “RGBD mapping: Using depth cameras for dense 3D modeling of indoor environments”, in the *12th International Symposium on Experimental Robotics (ISER)*, 2010.

[143] S. Se, D.G. Lowe, J. Little, “Vision-Based Global Localization and Mapping for Mobile Robots”, in *IEEE Transactions on Robotics*, vol. 21, no. 3, pp. 364-375, 2005.

[144] S. Se, D.G. Lowe, J. Little, “Mobile Robot Localization and Mapping with Uncertainty Using Scale- Invariant Visual Landmarks”, in the *International Journal of Robotics Research*, vol. 21, no. 8, pp. 735-758, 2002.

[145] L. Jigang, L. Dongquan, C. Jun, et al., “Conditional simultaneous localization and mapping: A robust visual SLAM system”, Neurocomputing, Vol. 145, pp. 269-284, 2014.

[146] A. Burguera, Y. Gonzalez, G. Oliver, Y. González, "Underwater SLAM with robocentric trajectory using a mechanically scanned imaging sonar," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2011, pp. 3577-3582.

[147] A. Mallios, P. Ridao, D. Ribas, E. Hern, and E. Hernandez, "Probabilistic sonar scan matching SLAM for underwater environment," in *OCEANS'10 IEEE SYDNEY*, 2010, pp. 1-8

[148] B. He, "Exploration with loop-closing in depth-fixed navigation for autonomous underwater vehicle," in *IEEE Intell. Veh. Symp*., pp. 459-463, Jun. 2009.

[149] I. T. Ruiz, S. De Raucourt, Y. Petillot, D. M. D. M. Lane, I. Tena Ruiz, S. de Raucourt, "Concurrent mapping and localization using sidescan sonar," in *IEEE Journal of Ocean Engineering*, vol. 29, no. 2, pp. 442- 456, Apr. 2004.

[150] M. Walter, F. Hover, J. Leonard, "SLAM for ship hull inspection using exactly sparse extended information filters," in *IEEE International Conference on Robotics and Automation*, Pasadena, CA, USA, May 19-23, 2008, pp. 1463-1470.

[151] A. Mallios, P. Ridao, D. Ribas, E. Hernandez, "Scan matching SLAM in underwater environments", in *Autonomous Robots*, vol. 36, pp. 181-198, 2014.

[152] D. Ribas, P. Ridao, J. D. Tard, J. Domingo Tardos, J. Neira, "Underwater SLAM in a marina environment," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2007, pp. 1455-1460.

[153] B. He, Y. Liang, X. Feng, R. Nian, T. Yan, M. Li, S. Zhang, "AUV SLAM and experiments using a mechanical scanning forward looking sonar," *Sensors* (Basel)., vol. 12, no. 7, pp. 9386–410, Jan. 2012.

[154] M. F. Fallon, J. Folkesson, H. McClelland, J. J. Leonard, "Relocating Underwater Features Autonomously Using Sonar-Based SLAM," in *IEEE Journal of Ocean Engineering*, vol. 38, no. 3, pp. 500-513, Jul. 2013.

[155] L. Jaulin, “A Nonlinear Set Membership Approach for the Localization and Map Building of Underwater Robots,” in *IEEE Trans. Robot*., vol. 25, no. 1, pp. 88-98, Feb. 2009.

[156] J. Hallset, “Testing the Robustness of an Underwater Vision System” in Laplante P, Stoyenko A (eds), *Real-time imaging: theory, techniques, and applications.* IEEE Press, pp. 225-260, 1996.

[157] S. Matsumoto, Y. ito, “Real-time Based Tracking of Submarine Cables for AUV/ROV”, in *Proc. Of IEEE Oceans*, vol.3, pp.1997-2002, 1995.

[158] P. Rives, J.J Borelly, “Visual Servoing Techniques Applied to an Underwater Vehicle”, in *Proc. of IEEE Int’l Conf. on Robotics and Automation (ICRA)*, pp. 20-25, 1997.

[159] A. Grau, J. Climent, J. Aranda, “Real-time Architecture for Cable Tracking Using Texture Descriptors”, in *Proc. of IEEE Oceans Conference*, vol. 3, pp. 1496-1500, 1998.

[160] D. Lodi Rizzini, F. Kallasi, J. Aleotti, F. Oleari, S. Caselli,” Integration of a stereo vision system into an autonomous underwater vehicle for pipe manipulation tasks,” in *Computers & Electrical Engineering*, Vol. 58, February 2017, pp. 560-571.

[161] G.L. Foresti, S. Gentili, “A Hierarchical Classification System for Object Recognition in Underwater Environments”, in *IEEE Journal of Oceanic Engineering*, vol. 1, no. 27, pp. 66-78, 2002.

[162] F.R. Dalgleish, S.W. Tetlow, R.L. Allwood, “Vision-Based Navigation of Unmanned Underwater Vehicles: a Survey. Part II: Vision Based Station Keeping and Positioning”, in *Proc. of the Institute of Marine Engineering*, *Science and Technology. Part B, Journal of Marine Design and Operations*, vol. 8, pp. 13-19, 2004.

[163] O. Pizarro, R. Eustice, H. Singh, “Large area 3d reconstructions from underwater surveys,” in *OCEANS 2004 MTS/IEEE Conference and Exhibition*, vol. 2, November 2004, pp. 678-687.

[164] R. Eustice, H. Singh, J. Leonard, M. Walter, R. Ballard, "Visually navigating the RMS titanic with SLAM information filters," in *Proceedings of Robotics: Science and Systems (RSS)*, June 2005.

[165] A. Kim, R. M. R. Eustice, "Real-Time Visual SLAM for Autonomous Underwater Hull Inspection Using Visual Saliency," in *IEEE Trans. Robot*., vol. 29, no. 3, pp. 719-733, June 2013.

[166] J. Aulinas, M. Carreras, X. Llado, J. Salvi, R. Garcia, R. Prados, Y.R. Petillot, "Feature extraction for underwater visual SLAM", in *Oceans, 2011 IEEE*, Spain, 2011.

[167] M. Meireles, R. Lourenco, A. Dias, J. M. Almeida, H. Silva, A. Martins, "Real time visual SLAM for underwater robotic inspection", in *Oceans*, St. John's, 2014, pp. 1-5.

[168] F. Ferreira, G. Veruggio, M. Caccia, G. Bruzzone, "Speeded Up Robust Features for vision-based underwater motion estimation and SLAM: comparison with correlation-based techniques", in *Proceedings of the 8th IFAC International Conference on Manoeuvring and Control of Marine Craft*, September 16-18, 2009, Guarujá (SP), Brazil

[169] J. M. Saez, A. Hogue, F. Escolano, M. Jenkin, "Underwater 3d slam through entropy minimization," in *IEEE International Conference on Robotics and Automation*, May 2006, pp. 3562-3567.

[170] I. Mahon, S. B. Williams, O. Pizarro, and M. Johnson-Roberson, "Efficient View-Based SLAM Using Visual Loop Closures," IEEE Trans. Robot., vol. 24, no. 5, pp. 1002–1014, Oct. 2008

[171] S. Pi, B. He, S. Zhang, R. Nian, Y. Shen, T. Yan, "Stereo visual SLAM system in underwater environment", in *Oceans* 2014, TAIPEI.

[172] P. L. N. Carrasco, F. Bonin-Font, M. Massot Campos, G. O. Codina, "Stereo-Vision Graph-SLAM for Robust Navigation of the AUV SPARUS II", in *IFAC-PapersOnLine*, Vol. 48, no. 2, 2015, pp. 200-205.

[173] R. Garcia, X. Cufi, P. Ridao and M. Carreras, "Constructing photo-mosaics to assist UUV navigation and station-keeping", in *Robotics and Automation in the Maritime Industries*; 6th AUTOMAR meeting. Editores: J. Aranda, P. Gonzalez de Santos, J.M. de la Cruz. (ESP), 2006.

[174] A. Elibol, N. Gracis, R. Gracias, "Augmented state-extended Kalman filter combined framework for topology estimation in large-area underwater mapping", in *Journal of field Robotics*, vol. 27, no. 5, pp. 656-674, 2010.

[175] R. Haywood, "Acquisition of a Micro Scale Photographic Survey Using an Autonomous Submersible", in *Proc. of IEEE Oceans*, vol.18, pp. 1423-1426, 1986.

[176] R.L. Marks, S. Rocks, M.J. Lee, "Real-Time Video Mosaicing of the Ocean Floor", in *Proc. Of Symposium on Autonomous Underwater Vehicle Technology*, pp. 21-27, 1994.

[177] S.D. Fleischer, R.L. Marks, S.M. Rock, "Improving Real-Time Video Mosaicing of the Ocean Floor", in *Proc. of IEEE Oceans*, 1995.

[178] Y. Rzhanov, L.M. Linnet, R. Forbes, "Underwater video mosaicking for seabed mapping", in *Int. Conf. Image Process*, 2000, pp.224-227.

[179] N. Gracias, J. Santos-Victor, "Underwater Video Mosaics as Visual Navigation Maps", in *Computer Vision and Image Understanding*, vol. 79, no. 1, pp. 66-91, 2000.

[180] X. Xu, S. Negahdaripour, "Mosaic-based positioning and improved motion estimation methods for automatic navigation of submersible vehicles", in *IEEE Journal of Oceanic Engineering*, vol. 27, no. 1, pp. 79-99, 2002.

[181] N. Hurtos, X. Cufi, J. Salvi, "Integration of optical and acoustic sensors for 3D underwater scene reconstruction", in *3th International Workshop on Marine Technology(MARTECH'09*), Vilanova i la Geltrú (ESP), 2009.

[182] S. Negahdaripour, H. Sekkati, H. Pirsiavash, "Opti-Acoustic stereo imaging: on system calibration and 3D target reconstruction", in *IEEE Transactions on Image Processing*, Vol.18, no. 6, June 2009.

[183] M. Bayat, A.P. Aguiar, "SLAM for an AUV using vision and an acoustic beacon", in *7th IFAC Symposium on Intelligent Autonomous Vehicle*, Vol. 43, no.16, 2010, pp. 503-508.

[184] I. Mahon, S. Williams, "Slam using natural features in an underwater environment," in *IEEE Control, Automation, Robotics and Vision Conference*, vol. 3, December 2004.

[185] R. B. Rusu e S. Cousins, "3D is here: Point Cloud Library (PCL)" IEEE International Conference on Robotics and Automation (ICRA), 2011.

[186] N. Snavely, S. Seitz e R. Szeliski, "Photo Tourism: Exploring Image Collections in 3D", ACM Transactions on Graphics, New York, p. 1–12, 2006.

[187] C. Wu, S. Agarwal, B. Curless e S. M. Seitz, "Multicore Bundle Adjustment" in CVPR, 2011.

[188] C. Wu, "Towards Linear-time Incremental Structure From Motion" in 3DV, 2013.

[189] Autodesk, Autodesk, 2016, 2016.

[190] D. V. M. V. G. L. Tingdahl, "ARC3D: A public Web service that turns photos into 3D models", Digital Imaging for Cultural Heritage Preservation: Analysis, Restoration, and Reconstruction of Ancient Artworks, 2011.

[191] Agisoft, Agisoft PhotoScan Professional Edition, 2016.

[192] S. Agarwal, N. Snavely, I. Simon, S. M. Seitz e R. Szeliski, "Building Rome in a Day", 2009.

[193] Y. Furukawa e J. Ponce, "Accurate Camera Calibration from Multi-View Stereo and Bundle Adjustment", International Journal of Computer Vision, vol. 84, n. 3, 2009.

[194] N. Snavely, S. Seitz and R. Szeliski, "Modeling the World from Internet Photo Collections," International Journal of Computer Vision, vol. 80, 2007.

[195] N. Snavely, S. Seitz e R. Szeliski, “Skeletal graphs for efficient structure from motion”, in CVPR, 2008.

[196] J. McGlone and G. Lee, "Manual of photogrammetry," American Society for Photogrammetry and remote Sensing, 2013.

[197] H. Sharp, “Practical Photogrammetry”, 1951.

[198] R. Hartley e A. Zisserman, “Multiple View Geometry”, 2004.

[199] J. Canny, “A computational approach to edge detection”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 8, n. 6, 1986.

[200] C. Harris e M. Stephens, “A combined corner and edge detector”, in Proceedings of the 4th Alvey Vision Conference, 1988.

[201] D. Lowe, “Distinctive image features from scale-invariant keypoints”, International Journal of Computer Vision, vol. 60, n. 2, 2004.

[202] M. Fischler e R. Bolles, “Random Sample Consensus, a Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography”, Communications of the ACM, vol. 24, n. 6, 1981.

[203] P. Rousseeuw and A. Leroy, "Robust regression and outlier detection," Wiley-Interscience, 1987.

[204] J. T. Dietrich, “Applications of Structure-from-Motion Photogrammetry to Fluvial Geomorphology”, 2014.

[205] M. Fonstad, J. Dietrich, B. Courville e P. Carbonneau, «Topographic structure from motion: a new development in photogrammetric measurements,» Earth Surface Processes and Landforms, Vol. %1 di %220: 817-827, 2013.

[206] L. Javernick, J. Brasington e B. Caruso, «Modeling the topography of shallow braided rivers using Structure-from-Motion photogrammetry,» Geomorphology , Vol. %1 di %2213: 166- 182, 2014.

[207] M. Westoby , J. Brasington, N. Glasser , M. Hambrey e M. Reyonds, «Structure from Motion photogrammetry: a low-cost, effective tool for geoscience applications,» Geomorphology, Vol. %1 di %2179: 300-314, 2012.

[208] N. Micheletti, J. Chandler e S. Lane, «Investigating the geomorphological potential of freely available and accessible structurefrom-motion photogrammetry using a smartphone,» Earth Surface Processes and Landforms, 2014.

[209] M. James e S. Robson, «Straightforward reconstruction of 3D surfaces and topography with a camera: accuracy and geoscience application,» Journal of Geophysical Research: Erath Surface, vol. 117, 2012.

[210] M. James e N. Varley, «Identification of structural controls in an active lava dome with high resolution DEMs: Volcán de Colima, Mexico,» Geophysical Research Letters, vol. 39, 2012.

[211] J. Dandois e E. Ellis, «High spatial resolution three-dimensional mapping of vegetation spectral dynamics using computer vision,» Remote Sensing of Environment, Vol. %1 di %2136: 259-276, 2013.

[212] A. Mathews e J. Jensen, «Visualizing and Quantifying Vineyard Canopy LAI Using an Unmanned Aerial Vehicle (UAV) Collected High Density Structure from Motion Point Cloud,» Remote Sensing, vol. 5, n. 5: 2164-2183, 2013.

[213] A. Koutsoudis , B. Vidmar, G. Ioannakis, F. Arnaoutoglou , G. Pavlidis e C. Chamzas, «Multi-image 3D reconstruction data evaluation,» Journal of Cultural Heritage, Vol. 15 n. 1: 73-79, 2014.

[214] M. Pollefeys, R. Koch e L. Gool, «Self-calibration and metric reconstruction inspite of varying and unknown intrinsic camera parameters,» IJCV, vol. 32, n. 1, 1999.

[215] A. Fitzgibbon e A. Zisserman, «Automatic camera recovery for closed or open image sequences,» Proc. ECCV, 1998.

[216] M. Lhuillier e L. Quan, «A quasi-dense approach to surface reconstruction from uncalibrated images,» IEEE Trans. PAMI, vol. 27, n. 3, 2005.

[217] M. Havlena, A. Torii, J. Knopp e T. Pajdla, «Randomized structure from motion based on atomic 3d models from camera triplets,» Proc. CVPR, 2009.

[218] B. Triggs , P. Mclauchlan, R. Hartley e A. Fitzg, «Bundle adjustment - a modern synthesis,» Lecture Notes in Computer Science, 2000.

[219] J. Sancho, «Real-Time EKF-Based Structure from Motion,» University of Saragoza, PhD Thesis, 2009.

[220] G. Klein e D. Murray, «Parallel tracking and mapping on a camera phone,» Proceedings of the 8th IEEE and ACM International Symposium on Mixed and Augmented Reality, 2009.

[221] Y. Furukawa e J. Ponce, «Accurate, Dense, and Robust Multiview Stereopsis,» Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 32, n. 8, 2010.

[222] Y. Kim, C. Theobalt, J. Diebel, J. Kosecka, B. Miscusik and S. Thrun, "Multi-view Image and ToF Sensor Fusion for Dense 3D Reconstruction," Computer Vision Workshops (ICCV Workshops), 2009.

[223] Texas Instruments, «Technical White Paper,» SLOA190, January 2014.

[224] F. Chiabrando, R. Chiabrando, D. Piatti e F. Rinaudo, «Sensors for 3D Imaging: Metric Evaluation and Calibration of a CCD/CMOS Time-of-Flight Camera,» Sensors, vol. 9, n. 12: 10080-10096, 2009.

[225] J. Weingarten, G. Gruener e R. Siegwart, «A state-of-the-art 3d sensor for robot navigation,» Vol. %1 di %23: 2155-2160, 2004.

[226] D. Droeschel, S. May, D. Holz, P. Ploger e S. Behnke, «Robust Ego-Motion Estimation with ToF Cameras,» Conference: Proceedings of the 4th European Conference on Mobile Robots, 2009.

[227] B. Dellen, G. Alenyà , S. Foix e C. Torras, «3D object reconstruction from Swissranger sensors data using a spring-mass model,» Proc. 4th Int. Conf. Comput. Vis. Theory Appl., vol. 2:, pp. 368-372, 2009.

[228] S. Foix, G. Alenyà, J. Andrade-Cetto e C. Torras, «Object modeling using a ToF camera under an uncertainty reduction approach,» Proc. IEEE Int. Conf. Robotics Autom., pp. 1306-1312, 2010.

[229] S. Fuchs e S. May, «Calibration and registration for precise surface reconstruction with time of flight cameras,» Int. J. Int. Syst. Tech. App., vol. 5, n. 3-4, pp. 274-284, 2008.

[230] J. Kuehnle, Z. Xue, M. Stotz, J. Zoellner, A. Verl e R. Dillmann, «Grasping in depth maps of time-of-flight cameras,» Proc. Int. Workshop Robot. Sensors Environ, pp. 132-137, 2008.

[231] A. Saxena, L. Wong e A. Ng, «Learning grasp strategies with partial shape information,",» Proc. 23th AAAI, 2008.

[232] Z. Xue, S. Ruehl, A. Hermann, T. Kerscher e R. Dillmann, «Autonomous grasp and manipulation planning using a ToF camera,» Robotics and Autonomous Systems, vol. 60, n. 3, pp. 387-395, 2012.

[233] D. Anderson, H. Herman e A. Kelly, «Experimental characterization of commercial flash ladar devices».Proceedings of International OCnference on Sensing technologies.

[234] D. Falie e V. Buzuloiu, «Noise characteristics of 3D Time-of-Flight cameras,» In Proceedings of IEEE Symposium on Signal Circuits & Systems (ISSCS), pp. 229-232.

[235] T. Kahlmann, F. Remondino e H. Ingensand, «Calibration for increased accuracy of the range imaging camera SwissRanger,» Int. Soc. Photogramm. Remote Sens., vol. XXXVI, pp. 136-141, 2006.

[236] M. Lindner e A. Kolb, «Lateral and depth calibration of PMD-distance sensors,» Proceedings of ISVC, pp. 524-533, 2006.

[237] H. Rapp, M. Frank, F. Hamprecht e B. Jähne, «A theoretical and experimental investigation of the systematic errors and statistical uncertainties of TIme-of-Flight cameras,» IJISTA, vol. 5, pp. 402-413, 2008.

[238] D. Lichti, «Self-Calibration of a 3D Range Camera,» In Proceedings of ISPRS Archives, vol. XXXVII, n. B5, pp. 927-932, 2008.

[239] C. Weyer, K. Bae, K. Lim e D. Lichti, «Extensive metric performance evaluation of a 3D range camera,» Int. Soc. Photogramm. Remote Sens., vol. XXXVII, pp. 939-944, 2008.

[240] C. Pfitzner, W. Antal, P. Hess, S. May, C. Merkl, P. Koch, R. Koch e M. Wagner, «3D MIlti-sensor data fusion for object localization in idustrial applications,» 41st International Symposium on Robotics, 2014.

[241] K. Kuhnert e M. Stommel, «Fusion of stereo-camera and PMD-camera data for real-time suited precise 3D environment reconstruction,» Intelligent Robots and Systems (IROS), pp. 4780-4785, 2006.

[242] S. Gudmundsson , H. Aanæs e R. Larsen, «Fusion of stereo vision and time-of-flight imaging for improved 3D estimation. Int. J. on,» Intell. Systems Techn. And App., Issue on Dynamic 3D Imaging, 2008.

[243] C. Beder, B. Bartczak and R. Koch, "A combined approach for estimating patchlets from PMD depth images and stero intensity images," F. Hamprecht, C, Schnorr, B Jahne editors, Proc. of the DAGM, LNCS,, pp. 11-20, 2007.

[244] R. Nair, K. Ruhl, F. Lenzen, S. Meister, H. Schafer, C. Garbe, M. Eisemann, M. Magnor e D. Kondermann, «A survey on Time-of-Flight Stereo Vision,» Sensors, Algorithms, and Applications, 2013.

[245] K. Kuhnert e M. Stommel, «Fusion of stereo-camera and pmd-camera data for real-time suited precise 3d environment reconstruction,» Int. Conf. on Intelligent Robots and Systems, pp. 4780-4785, 2006.

[246] S. Gudmundsson, H. Aanaes and R. Larsen, "Fusion of stereo vision and time-of-ight imaging for improved 3d estimation.," IJISTA, 5(3):425{433, 2008., vol. 5, no. 3, pp. 425-433, 2008.

[247] B. Bartczak e R. Koch, «Dense depth maps from low resolution time-of-ight depth and high resolution color views,» Advances in Visual Computing, pp. 228-239, 2009.

[248] U. Hahne e M. Alexa, «Depth imaging by combining time-of-ight and on-demand stereo.,» Dynamic 3D Imaging, pp. 70-83, 2009.

[249] C. Dal Mutto , P. Zanuttigh e G. Cortelazzo, «A probabilistic approach to tof and stereo data fusion,» 3DPVT, 2010.

[250] Q. Yang, K. Tan, B. Culbertson e J. Apostolo, «Fusion of active and passive sensors for fast 3d capture.,» MMSP, 2010.

[251] R. Nair, F. Lenzen, S. Meister, H. Schafer, C. Garbe e D. Kondermann, «High accuracy tof and stereo sensor fusion at interactive rates,» Computer Vision ECCV Workshops and Demonstrations, pp. 1-11, 2012.

[252] C. Beder, B. Bartczak e R. Koch, «A combined approach for estimating patchlets from pmd depth images and stereo intensity images,» Pattern Recognition, pp. 11-20, 2007.

[253] J. Fischer, G. Arbeiter e A. Verl, «Combination of time-of-flight depth and stereo using semiglobal optimization,» Int. Conf. on Robotics and Automation (ICRA), pp. 3548-3553, 2011.

[254] U. Hahne e M. Alexa, «Combining time-of-flight depth and stereo images without accurate extrinsic calibration,» IJISTA, vol. 5, n. 3, pp. 325-333, 2008.

[255] J. Zhu, L. Wang, R. Yang e J. Davis, «Fusion of time-of-flight depth and stereo for high accuracy depth maps,» IEEE Proceedings of CVPR, pp. 1-8, 2008.

[256] J. Zhu, L. Wang, R. Yang, J. Davis e e. al., «Reliability fusion of time-of-flight depth and stero for high quality depth maps,» TPAMI, vol. 99, 2011.

[257] J. Zhu, L. Wang, J. Gao e R. Yang, «Spatial-temporal fusion for high accuracy depth mpas using dynamic mrfs,» Pattern Analysis and Machine Intelligence, IEEE TRansactions on, vol. 32, n. 5, pp. 899-909, 2010.

[258] K. Ruhl, F. Klose, C. Lipski e M. Magnor, «Integrating approximate depth data into dense image correspondence estimation,» Proc. European Conference on Visual Media Production (CVMP), 2012.

[259] V. Gandhi, J. Cech e R. Horaud, «High resolution depth mpas based on tof-stero fusion,» Robotics and Automation (ICRA), IEEE International Conference on, pp. 4742-4749, 2012.

[260] C. Dal Mutto, P. Zanuttigh, S. Mattoccia e G. Cortelazzo, «Locally consistent tof and stereo data fusion,» Computer Vision ECCV, Workshops and Demonstrations, pp. 598-607, 2012.

[261] Y. Song, C. Glasbey, G. Van der Heijden, G. Polder e J. Dieleman, «Combining stereo and time-of-flight images with application to automatic plant phenotyping,» Imsge Analysis, pp. 467-478, 2011.

[262] S. Mattoccia, «A locally global approach to stereo correspondence,» Computer Vision Workshops (ICCV Workshops), IEEE 12th International Conference on, pp. 1763-1770, 2009.

[263] S. Fuchs, S. Haddadin, M. Keller, S. Parusel, A. Kolb e M. Suppa, «Cooperative bin-picking with time-of-flight camera and impedance controlled dlr lightweight robot iii,» IROS, IEEE, pp. 4862-4867, 2010.

[264] M. Nieuwenhuisen, D. Droeschel, D. Holz, J. Stuckler, A. Berner, J. Li, R. Klein e S. Behnke, «Mobile bin picking with an anthropomorphic service robot,» ICRA, IEEE, pp. 2327-2334, 2013.

[265] A. Linarth, J. Penne, B. Liu, O. Jesorky e R. Kompe, «Fast fusion of range and video sensor data,» Advanced Microsystems for Automotive Applications (2007), pp. 119-134, 2007.

[266] B. Huhle , P. Jenke e W. Strasser, «On-the-fly scene acquisition with a handy multisensor-system,» Int. J. on Intell. Systems Techn. and App., Issue on Dynamic 3D Imaging 5, 3/4, pp. 255-263, 2008.

[267] P. Jenke, B. Huhle e W. Strasser, «Selflocalization in scanned 3DTV sets,» 3DTV CON - The True Vision, 2007.

[268] M. Lindner, A. Kolb e K. Hartmann, «Datafusion of PMD-based distance-information and highresolution RGB-images,» IEEE Sym. on Signals Circuits & Systems (ISSCS), session on Alg. for 3D ToF cameras, pp. 121-124, 2007.

[269] B. Streckel, B. Bartczak, R. Koch e A. Kolb, «Supporting structure from motion with a 3Drange-camera,» Scandinavian Conf. Image Analysis (SCIA), LNCS, pp. 233-242, 2007.

[270] S. Gudmundsson, R. Larsen, A. H e C. J. R. PARDÁS M., «TOF imaging in smart room environments towards improved people tracking,» In IEEE Conf. on Computer Vision & Pattern Recogn.; Workshop on ToF-Camera based Computer Vision, 2008.

[271] K. Arun, T. Huang e S. Blostein, «Least-Squares Fitting of Two 3-D Point Sets,» IEEE Trans. Patt. Anal. Machine Intell., Pami, vol. 9, pp. 698-700, 1987.

[272] S. Marden e J. Guivant, «Improving the Performance of ICP for Real- Time Applications using an Approximate Nearest Neighbour Search,» Australasian Conference on Robotics and Automation, ACRA, New Zealand, 2012.

[273] P. Besl e N. McKay, «Method for registration of 3-D shapes,» IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 14, n. 2, p. 586–606, 1992.

[274] Y. Chen e G. Medioni, «Object modeling by registration of multiple range images,» Proc. IEEE Int'l Conf. Robotics and Automation, vol. 3, p. 2724–2729, 1991.

[275] Z. Zhang, «Iterative point matching for registration of freeform curves and surfaces,» Int'l J. Computer Vision, vol. 13, n. 2, pp. 119-152, 1994.

[276] R. Newcombe, A. Davison, S. Izadi, P. Kohli, O. Hiliges, J. Shotton, D. Molyneaux, S. Hodges, D. Kim e A. Fitzgibbon, «Kinectfusion: Real-time dense surface mapping and tracking,» IEEE Int'l Symp. Mixed and Augmented Reality, pp. 127-136, 2011.

[277] S. Seitz, B. Curless, J. Diebel, D. Scharstein e R. Szeliski, «A comparison and evaluation of multi-view stereo reconstruction algorithms,» Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. 1, pp. 519-528, 2006.

[278] S. May, «3D Time-of-Flight Ranging for Robotic Perception in Dynamic Environments,» PhD thesis, Universitt Osnabrck, 2009.

[279] S. Fantoni, U. Castellani e A. Fusiello, «Accurate and Automatic Alignment of Range Surfaces,» 2012 Second International Conference on 3D Imaging,Modeling, Processing, Visualization & Transmission IEEE, pp. 73-80, 2012.

[280] A. Segal, D. Haehnel e S. Thrun, «Generalized-ICP,» Proceedings of Robotics: Science and Systems, Seattle, 2009.

[281] D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach e D. Hawkes, «Nonrigid registration using free-form deformations: application to breast MR images,» IEEE transactions on medical imaging, vol. 18, n. 8, pp. 712-21, 1999.

[282] J. Yang, H. Li, D. Campbell e Y. Jia, «Go-ICP: A Globally Optimal Solution to 3D ICP Point-Set Registration,» IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016.

[283] A. Fitzgibbon, «Robust registration of 2D and 3D point sets,» Image and Vision Computing, vol. 21, n. 13, pp. 1145-1153, 2003.

[284] Y. Tsin e T. Kanade, «A correlation-based approach to robust point set registration,» Proc. European Conf. Computer Vision, pp. 558-569, 2004.

[285] R. Sandhu, S. Dambreville e A. Tannenbaum, «Point set registration via particle filtering and stochastic dynamics,» IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 32, n. 8, pp. 1459-1473, 2010.

[286] L. Silva, O. Bellon e K. Boyer, «Precision range image registration using a robust surface interpenetration measure and enhanced genetic algorithms,» IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 5, 2005.

[287] R. Rusu, N. Blodow e M. Beetz, «Fast point feature histograms (FPFH) for 3D registration,» Proc. IEEE Int'l Conf. Robotics and Automation, pp. 3212-3217, 2009.

[288] A. Makadia, A. Patterson e K. Daniilidis, «Fully automatic registration of 3D point clouds,» Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. 1, pp. 1297-1304, 2006.

[289] Hodgson, G (1999). "A Global Assessment of Human Effects on Coral Reefs", Marine Pollution Bulletin Vol. 38, No. 5, pp. 345-355.

[290] Hochachka, WM, Fink, D, Hutchinson, RA, Sheldon, D, Wong, W, and Kelling, S (2012). "Data-intensive science applied to broad-scale citizen science", Trends in Ecology and Evolution, Vol. 27, No. 2.

[291] Danielsen, F, Burgess, ND, Balmford, A (2005). "Monitoring Matters: Examining the Potential of Locally-based Approaches", Biodiversity and Conservation 14:2507–2542.

[292] Brightsmith, DJ, Stronza, A, Holle, K (2008). "Ecotourism, conservation biology, and volunteer tourism: A mutually beneficial triumvirate", Biological Conservation 141, 2832-2842

[293] Branchini, S, Pensa, F, Neri, P, Tonucci, BM., Mattielli, L, Collavo, A, Sillingardi, M.E, Piccinetti, C, Zaccanti, F, Goffredo, S (2015). "Using a citizen science program to monitor coral reef biodiversity through space and time", Biodiversity and conservation 24, pp. 319 – 336.

[294] Bonney, R, Shirk, JL, Phillips, TB, Wiggins, A, Ballard, HL, Miller-Rushing, AJ, Parrish, JK (2014). "Next Steps for Citizen Science", Science, Vol. 343, Issue 6178, pp. 1436-1437.

[295] Fore, LS, Paulsen, K, O' Laughlin, K (2001). "Assessing the performance of volunteers in monitoring streams", Freshwater Biology 46, 109–123.

[296] Van der Velde, T, Milton, DA, Lawson, TJ, Wilcox, C, Lansdell, M, Davis, G, Perkins, G, Hardesty, BD (2016). "Comparison of marine debris data collected by researchers and citizen scientists: Is citizen science data worth the effort?", Biological Conservation.

[297] Vann-Sander, S, Clifton, J, Harvey, E (2016). "Can citizen science work? Perceptions of the role and utility of citizen science in a marine policy and management context", Marine Policy 72, 82–93.

[298] Foster-Smith, J, Evans, SM (2003). "The value of marine ecological data collected by volunteers", Biological Conservation 113, 199–213.

[299] Delaney, DG, Sperling, CD, Adams, CS, Leung, B (2008). "Marine invasive species: validation of citizen science and implications for national monitoring networks", Biological Invasions 10, 117–128.

[300] Theobald, EJ, Ettinger, AK, Burgess, HK, DeBey, LB, Schmidt, NR, Froehlich, HE, Wagner, C, HilleRisLambers, J, Tewksbury, J, Harsch, MA, Parrish, JK (2015). "Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research", Biological Conservation 181, 236–244.

[301] Halusky, JG, Seaman, W, Strawbridge, EW (1994). "Effectiveness of trained volunteer divers in scientific documentation of artificial aquatic habitats", Bulletin of marine science, 55(2-3): 939-959.

[302] Cerrano, C, Milanese, M, Ponti, M (2016). "Diving for science - science for diving: volunteer scuba divers support science and conservation in the Mediterranean Sea: Citizen Science for Mediterranean MPAs", Aquatic conservation marine and freshwater ecosystems

[303] Scaradozzi, D, Conte, G, De Capua, GP, Sorbi, L, Luciani, C, De Cecco, PG, Sorci, A (2009). “Innovative technology for studying growth areas of Posidonia oceanica”, IEEE Workshop on Environmental, Energy, and Structural Monitoring Systems (EESMS).

[304] Zapata-Ramírez, PA, Scaradozzi, D, Sorbi, L, Palma, M, Pantaleo, U, Ponti, M, Cerrano, C (2013). “Innovative study methods for the Mediterranean coralligenous habitats”, Advances in Oceanography and Limnology 4:102-119.

[305] Williams, Stefan B., et al. (2012). "Monitoring of benthic reference sites: using an autonomous underwater vehicle", IEEE Robotics & Automation Magazine 19.1,73-84.

[306] Puillat, I, Prevosto, M, Mercier, H, Thomas, S (2014). “Time series analysis of marine data: A key knowledge at the crossroads of marine sciences”, Journal of Marine Systems, Vol. 130, Pages 1–3.

[307] Hedley, John D. et al. (2016). "Remote sensing of coral reefs for monitoring and management: A review", Remote Sensing 8.2:118.

[308] Rende, SF, Irving, AD, Bacci, T, Parlagreco, L, Bruno, F, De Filippo, F, Montefalcone, M, Penna, M, Trabucco, B, Di Mento, R, Cicero, AM (2015). “Advances in micro-cartography: A two-dimensional photo mosaicing technique for seagrass monitoring”, Estuarine, Coastal and Shelf Science, Volume 167, Part B, Pages 475–486.

[309] Muzzupappa, M, Gallo, A, Spadafora, F, Manfredi, F, Bruno, F, Lamarca, A (2013). “3D reconstruction of an outdoor archaeological site through a multi-view stereo technique”, Digital Heritage International Congress (DigitalHeritage).

[310] Westoby, M, et al. (2012). "‘Structure-from-Motion’ photogrammetry: A low-cost, effective tool for geoscience applications", Geomorphology 179: 300-314.

[311] Zhang, Caiyun, et al. (2013). "Object-based benthic habitat mapping in the Florida Keys from hyperspectral imagery", Estuarine, Coastal and Shelf Science 134: 88-97.

[312] Ferrari, Renata, et al. (2016). "Quantifying the response of structural complexity and community composition to environmental change in marine communities", Global change biology.

[313] Figueira, W, et al. (2015). "Accuracy and precision of habitat structural complexity metrics derived from underwater photogrammetry", Remote Sensing 7.12: 16883-16900.

[314] McCarthy, J (2014). "Multi-image photogrammetry as a practical tool for cultural heritage survey and community engagement", Journal of Archaeological Science 43: 175-185

[315] Storlazzi, CD, Dartnell, P, Hatcher, GA, Gibbs, AE (2016). "End of the chain? Rugosity and fine-scale bathymetry from existing underwater digital imagery using structure-from-motion (SfM) technology", Coral Reefs: 1-6.

[316] Westoby, M, et al. (2016). "From an 'ice-see' perspective: The current use, potential and limitations of Structure-from-Motion photogrammetry for cryospheric applications", EGU General Assembly Conference Abstracts Vol. 18.

[317] Beijbom, O, et al. (2012). "Automated annotation of coral reef survey images", IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[318] González-Rivero, M, et al. (2016). "Scaling up Ecological Measurements of Coral Reefs Using Semi-Automated Field Image Collection and Analysis", Remote Sensing 8.1: 30.