

# A MODULAR AND LOW-COST PORTABLE VSLAM SYSTEM FOR REAL-TIME 3D MAPPING: FROM INDOOR AND OUTDOOR SPACES TO UNDERWATER ENVIRONMENTS

F. Menna<sup>1</sup>, A. Torresani<sup>1,2</sup>, R. Battisti<sup>1</sup>, E. Nocerino<sup>3</sup>, F. Remondino<sup>1</sup>

<sup>1</sup> 3D Optical Metrology (3DOM) unit, Bruno Kessler Foundation (FBK), Trento, Italy

Web: <http://3dom.fbk.eu> – Email: <fmenna><atorresani><rbattisti><remondino>@fbk.eu

<sup>2</sup> Department of Information Engineering and Computer Science (DISI), University of Trento, Italy

<sup>3</sup> Dipartimento di Scienze Umanistiche e Sociali, University of Sassari, Sassari, Italy; email:enocerino@uniss.it

## Commission II

**KEY WORDS:** vSLAM, Real Time, Low-Cost, Raspberry, Mobile Mapping, GNSS-denied environments, Indoor, Underwater

### ABSTRACT:

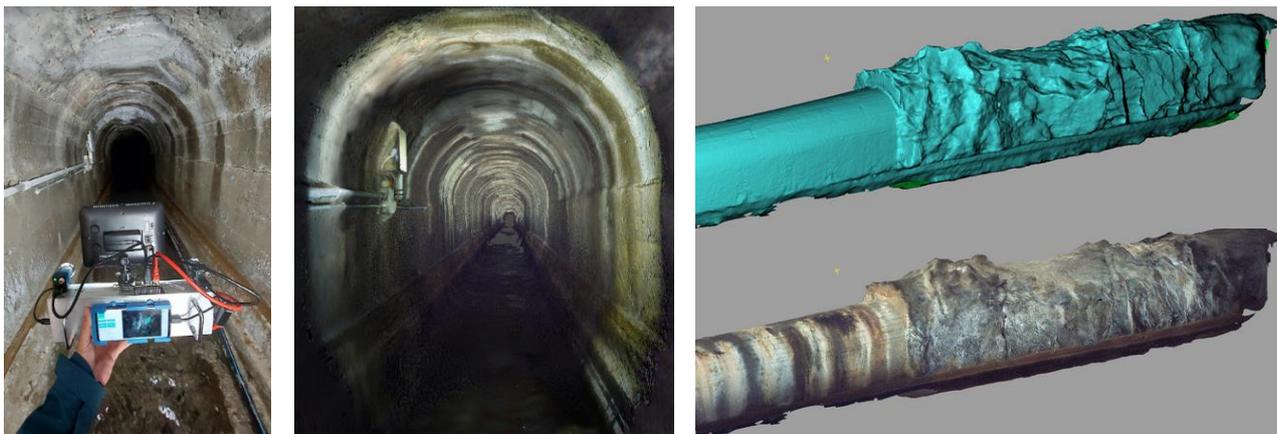
The bond with computer vision and robotics is revolutionizing the traditional surveying approaches. Algorithms such as visual odometry and SLAM are embedded in surveying systems to make on-site and processing operations more efficient both in terms of time and quality of the achieved results. In this paper, we present the latest developments on GuPho, a mobile mapping concept based on photogrammetry that leverages a vSLAM solution to provide innovative and unique features supporting the image acquisition and optimising the processing steps. These include visual feedback on ground sample distance and maximum allowed speed to avoid motion blur. Two efficient image acquisition strategies, based on geometric principles, are implemented to optimise the disk storage, avoiding unnecessary redundancy. Moreover, an innovative automatic exposure control that adjusts the shutter speed or gain based on the tracked object in 3D is part of the system. The paper reports the motivations behind the design choices, details the hardware and software components, discusses several case studies to showcase the potentialities of our low-cost, lightweight, and portable modular prototype system.

## 1. INTRODUCTION

Vision based localization techniques, such as visual odometry (VO) and Simultaneous Localization And Mapping (vSLAM), are getting more and more attention not only in robotics and augmented reality (AR) but also in the geomatic field and in the surveying industry, becoming a key component in many mobile mapping systems, especially portable ones (Otero et al., 2020, Nocerino et al., 2019). VO and vSLAM provide real-time estimation of the position and orientation (also known as the pose or six degrees of freedom-6DoF) of an agent moving in an environment based solely on a sequence of images captured by one or more cameras rigidly mounted on the agent (Cadena et al., 2016). Depending on the vSLAM algorithm used (Taketomi et al., 2017; Cadena et al., 2016), a 3D sparse, semi-dense, or dense reconstruction of the environment is available as a product of the

vSLAM process along with the already mentioned image orientation parameters (camera poses).

vSLAM techniques are beneficial not only for real-time photogrammetry applications (Ortiz-Coder and Sánchez-Ríos, 2020; Menna et al., 2019) but also for mobile mapping systems based on LIDAR technology. Indeed, image orientation is a fundamental step to provide photogrammetric measurements and products like dense point clouds, polygonal meshes, digital elevation models, orthophotos and texturized 3D models (Figure 1). At the same time, for LIDAR-based mobile mapping system, the sensor pose is needed for each laser pulse to produce the point cloud. For this reason, some LIDAR-based mobile mapping systems are integrating, among different sensors, a camera for vSLAM, such as the Leica BLK2GO Handheld Imaging Laser Scanner ([www.leica-geosystems.com](http://www.leica-geosystems.com)) or the KAARTA stencil series ([www.kaarta.com](http://www.kaarta.com)).



**Figure 1.** Example of indoor 3D mapping of an underground tunnel using the developed photogrammetric portable mobile mapping system based on vSLAM for real-time guidance. From left to right: the terrestrial prototype in a tunnel, view of the 3D textured model from the inside, shaded and textured mesh from outside.

vSLAM techniques can support the pose estimation in urban canyons and in GNSS denied environments such as indoor, underground, and underwater. Even terrestrial laser scanners are integrating vSLAM approaches to speed up the co-registration of the point clouds recorded from the different scanning positions. A visual inertial system tracks the motion of the scanner when it is moved from one station to another providing initial approximate parameters for the co-registration (e.g., Leica RTC360, [www.leica-geosystems.com](http://www.leica-geosystems.com)). SLAM is revolutionizing the surveying industry where 3D products have been traditionally obtained in post-processing, in the office, after the data collection on the field. The advantages are several, with the main ones being related to real-time processing: (i) faster response and decision making directly on the field (e.g. in manufacturing, natural disaster assessment, forensics, subsea metrology); (ii) immediate verification of the collected data to guarantee that the survey meets the survey requirements (e.g. resolution, coverage, precision, reliability).

Moved by the above considerations, we developed a low-cost, lightweight, and portable modular prototype system called GuPho (Guided Photogrammetry system) based on stereo vision that uses a customized vSLAM approach to provide real-time guidance to the surveyor during the image capturing phase, ensuring a more reliable, and effective photogrammetric data acquisition and processing. Preliminary results on the algorithmic aspects related to the real-time feedback and image acquisition control were presented in Torresani et al. (2021). In Di Stefano et al (2021) a first metric comparison with other portable mobile systems was presented in the case of underground heritage documentation.

In this paper we provide a broader overview of the motivations driving the development behind such a system and its lightweight architecture. New details are presented on different surveying scenarios enabled by its modularity that allows it to be handheld by a person, when dexterity is key, or remotely controlled on a ground robot vehicle for confined or hazardous places. Section 2 introduces the context of portable mobile mapping systems developed within research and commercial domains that are more akin to GuPho's concept. Section 3 discusses in detail the requirements and motivations that lead to the specific design choices. The low-cost hardware components and overall modular architecture are described.

In section 4 different case studies will be showcased to explain the different features and modularity of the system: from indoor and underground to underwater environments. We report novel unpublished results including its underwater version and future developments to improve the reliability of the system in challenging scenarios, e.g. in presence of poorly textured and dynamic scene such as crowded public spaces or underwater (fish, seaweeds, caustics, benthic species).

## 2. PORTABLE MOBILE MAPPING SYSTEMS

Over the last few years, 3D mapping of natural and man-made structures has been revolutionised by the introduction of portable mobile mapping systems, such as backpack and handheld devices (Nocerino et al., 2019). Such a technology has led to significant changes over the planning, execution, and processing of the data, driven by a general requirement for efficiency, speed and effectiveness of the 3D survey techniques. However, this change of surveying paradigm has come with some downsides when compared to classic surveying techniques such as, for example, terrestrial laser scanning (TLS), with the main ones being an overall reduction of the achievable accuracy and resolution as well as limited range (only a few meters for handheld devices).

Portable mobile mapping technology enables faster surveys with little or any need for control over the acquisition parameters during the survey, especially for those based on LIDAR. Indeed, most of the research and commercial systems (Torresani et al., 2022, Nocerino et al., 2019;) only require the surveyor to walk the environment for it to be mapped in 3D. When high resolution colour information, or contextual images are needed along with the geometry acquired by the LIDAR sensors, one or more cameras may be integrated as optional sensors, depending on the system used. For systems that do not rely on vSLAM as main technique for the trajectory estimation (for example the GeoSLAM Zeb Horizon - <https://geoslam.com/>), the need for colour information may significantly reduce the overall speed of acquisition and thus efficiency of the system. Indeed, indoor, underground, or in any environment with low-light conditions, the walking speed must be significantly reduced and adapted so that, for a given exposure time of the used camera, the motion blur does not affect the sharpness of the image.

On the other hand, till recently, portable mobile mapping systems based on vision have received limited attention if compared to LIDAR based systems. Photogrammetry is known to be among the most flexible surveying techniques, although it requires specific knowledge and expertise to plan and correctly carry out a photogrammetric survey that meets predefined metric requirements (Menna et al., 2016a; 2020; Nocerino et al., 2013; Fraser, 1984). Moreover, over the last decade, structure from motion (SfM) and multi view stereo algorithms have significantly automatised and speeded up the photogrammetric data processing (Remondino et al., 2017, 2014), yet full resolution and highest accuracy potential performances are not reachable in real-time in portable handheld devices.

For this reason, most of the commercial and research systems have focused on devices aimed at supporting the image capturing in the field with the full processing performed post-survey, in the office. For example, a commercial solution integrating a GNSS receiver and a multi-camera system for panoramic imaging was commercialised by Trimble ([www.trimble.com](http://www.trimble.com)) with its V10 Imaging Rover system almost ten years ago (Baiocchi et al., 2017). A concept reintroduced more recently by Hexagon with the Leica GS18 I Rover, which uses visual positioning and GNSS-IMU data fusion ([www.leica-geosystems.com](http://www.leica-geosystems.com)) to provide interactive photogrammetric measurement of unreachable or under overhang points, in the field, using the provided tablet and stylus. Nevertheless, this system is conceived to work with the main device in open sky to allow RTK positioning and can perform only a limited number of manual measurements in real-time. Solutions able to work in real-time, in GNSS denied environments, without any positioning system were proposed in conjunction with depth sensors, such as using the popular Kinect device (Menna et al., 2011) and fusion libraries (Pagliari et al., 2014) or based on the project Tango Development Kit Tablet (Schöps et al., 2017), leveraging GPU acceleration (Nvidia Tegra K1 chipset), a visual-inertial odometry algorithm from the tablet monochrome fisheye camera, and a truncated signed distance function (TSDF) volumetric approach to compute in real-time dense 3D models of large scale outdoor scenes.

Mobile phone-based solutions were proposed using, for example, a collaborative approach over the internet where the images collected by a smart phone were sent to a server for an online photogrammetric processing while a 3D sparse reconstruction and camera poses were visible with low-latency on the mobile phone (Nocerino et al., 2017). Following the same idea, in 2020 the company Pix4D released Pix4DCatch ([www.pix4d.com](http://www.pix4d.com)), which takes advantage of mobile augmented reality frameworks

to assist the user during the image acquisition. The app allows the user to see in real-time the locations of the acquired images and the sparse point cloud of the acquired scene. Underwater, where satellite-based global positioning systems are not available due to electromagnetic absorption of water, visual odometry and visual SLAM techniques have shown great potential in a number of different applications ranging from archaeology to subsea metrology (Menna et al., 2019; Nawaf et al., 2018; Drap et al. 2015). In these works, the images, captured by a stereo-camera, were transferred to a computer for real-time processing to support navigation while performing the photogrammetric survey.

These systems rely on powerful computational resources, often with GPU capabilities and delocalised on a remote computer wirelessly over the Internet or via a cable. For a more comprehensive review of visual based handheld portable mobile mapping systems the reader is addressed to Torresani (2022). To the best of our knowledge, none of the available solutions provide real-time guidance and control of the system, such as automatic image acquisition, distance to the object, maximum allowed speed to avoid motion blur, exposure control.

While in open sky real time kinematic (RTK) can provide precise positioning for attaining, automatically, a specific camera network (Menna et al., 2020), in GNSS denied environments, such as indoor, underground, and underwater, current photogrammetric surveys still rely on the photogrammetrist expertise for a careful plan and execution of the survey mostly with limited control over the image acquisition parameters in the field (Menna et al 2016a, 2013). The GuPho concept presented in this paper, with the two developed modular prototypes, aims at providing a series of tools supporting the photogrammetrist in the field, whether it be for mapping unexplored areas, documenting a heritage site or for monitoring purposes.

### 3. GUPHO - THE GUIDED PHOTOGRAMMETRY CONCEPT

#### 3.1 Hardware and software requirements

We gathered a list of features deemed key in a portable device system for photogrammetric measurements, based upon experiences and lessons learned over the years in the implementation of different 3D documentation and monitoring projects, in particular indoor and in GNSS denied environments, from cultural heritage, to underwater subsea metrology.

*Lightweight:* to fully exploit human dexterity and reduce the fatigue in handling the system, we aimed at a weight ideally less than a professional DSLR camera (e.g. less than 2 kilograms, lens and battery included). A lightweight system can be used as payload for small robots and UAVs and can be easily transported even in remote areas reachable only after long hikes such as in the case of the 3D documentation of WWI tunnels in the alpine regions (Nocerino et al., 2018; Rodríguez-González et al., 2015).

*Modularity:* we aimed at both hardware and software modularity. The cameras setup, such as baseline length and configuration need to be easily changeable, depending on the application of interest, for example from parallel stereo to convergent axes or from rectilinear to fisheye lenses. Focus needs to be adjustable for very close range surveys. Visualization of the real-time processed data should be possible on site on a display but also remotely via cable or wirelessly for example to allow the system to be operated by robots, UAVs, ROVs, etc.

*Smart image capturing:* to maximise the efficiency in data storage and post-processing, we sought a simple, yet effective solution that skips unnecessarily redundant image data, such as acquiring and storing new images only when these are beneficial for the camera network geometry and object coverage, differentiating it from a “blind” timed acquisition at a specific frame rate where if the operator remains still during the mobile acquisition unnecessary images are stored.

*Memory efficient:* the system should be able to track and map unexplored areas for several minutes without memory issues due to the growth of the map.

*Relocalisation:* in order to map large areas in tiles or monitoring a same scene over time, a relocalisation of the device is sought for the operator to understand the agent’s pose and continue the image acquisition of unexplored/unmapped regions or remeasure a previously surveyed area.

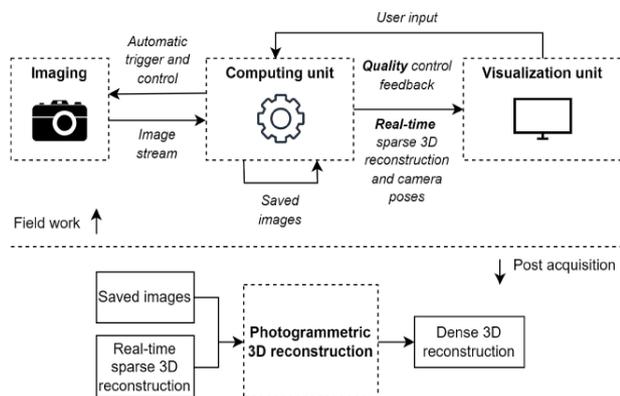
*Low-cost:* cameras are less expensive than active sensors and a low price system is attractive, especially considering the limited budget available today for many non-industrial projects, such as ecology and cultural heritage.

*Energy saver:* a non-energy-hungry device requires less power, which simplifies the cooling of the system (for example using a passive heatsink), and reduces the battery capacity and weight, an important aspect to consider when traveling as airline restrictions allow only a limited battery capacity to carry onboard a plane (e.g. 100Wh).

*Robustness to dynamic scene:* the survey of indoor spaces, especially public ones may introduce significant disturbances and thus outliers in the scene that may lead to several different issues, such as loss of tracking and bundle adjustment instability. A typical application is in real estate surveys. Similarly, underwater, wildlife like fish or water caustics in shallow water may represent a problem for vSLAM algorithms due to the continuously changing appearance of the scene.

#### 3.2 vSLAM framework

Among the different available vSLAM frameworks (Torresani, 2022), the OpenVSLAM project (Sumikura et al., 2019), based on ORB-SLAM2 (Mur-Artal and Tardós, 2017), was chosen as building block for our developments, meeting it several of the requirements listed in the previous subsection. ORB-SLAM2 is an open-source a feature-based SLAM system for monocular, stereo and RGB-D cameras, an evolution of a vSLAM method originally developed for monocular systems by the same authors (Mur-Artal et al., 2015). The choice to use OpenVSLAM project (Sumikura et al., 2019) was motivated by important features such as: i) the possibility to use both rectilinear and fisheye camera models, and ii) the real-time visualisation relies on a web-based viewer, which reduces the computation load by keeping separate the vSLAM computation and the 3D visualisation. To help in the image acquisition, we use the camera pose estimation and 3D sparse reconstruction performed in real-time to introduce visual feedbacks shown on a display. Moreover, additional features were introduced such as, for example: (i) camera-to-object distance warnings to guarantee the expected ground sample distance (GSD), (ii) automatic baseline acquisition to meet the planned overlap, (iii) actual speed warnings to avoid motion blur, (iv) automatic image exposure computed on the 3D surveyed scene and not on the entire image.



**Figure 2.** High level architecture and workflow of GuPho's concept.

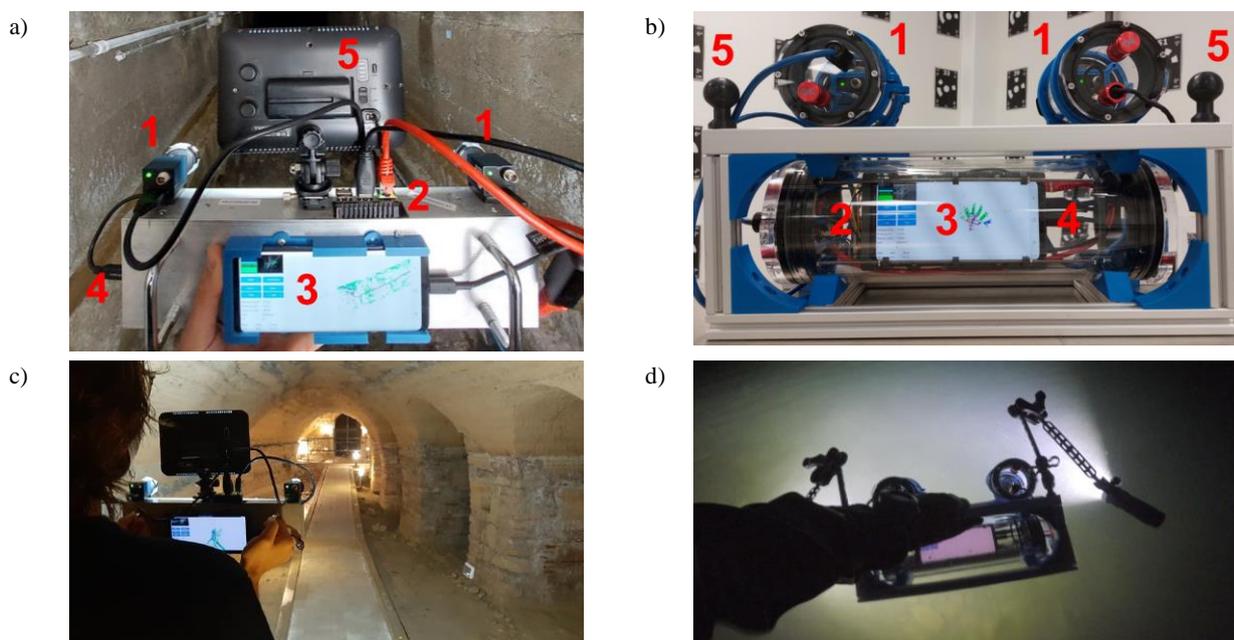
Additionally, the system can support multi-temporal monitoring tasks thanks to its capabilities to store the map and extracted visual features in a local database that can be reutilized for future revisitation of the same area. In a typical scenario, a camera network previously surveyed or designed can be loaded in the system and used to automatically trigger the image acquisition when the exterior orientation of the camera in real-time is found to be within a given threshold (e.g. Euclidean distance or attitude) from the same picture taken in a previous survey or from a synthetically planned camera network (subsection 4.4). Among the portable mobile mapping systems today available, in the authors knowledge, this is the first example of a visual handheld mobile mapping device that provides real-time feedback and assistance during the acquisition of the images, which cover both geometric and radiometric aspects of the imagery.

### 3.3 System architecture and workflow overview

A high-level overview of the proposed system is shown in Figure 2, which depicts its three main components: imaging, computing and visualization. It is composed of a pair of global shutter 1.3 Mpx cameras, a Raspberry Pi4 microcomputer, a mobile phone used for 3D visualization and a power bank. A detailed overview

of the hardware is visible in Table 1. The first prototypes of the system used for terrestrial and underwater applications are visible respectively in figure 3a,c and 3b,d. The underwater version uses the same hardware enclosed in waterproof housings made of polycarbonate, featuring dome ports and is rated to 100m depth. There were different aspects considered in the choice of a monocular versus synchronized stereo camera configuration. A stereo camera requires to double the costs of the imaging hardware as well as to halve the number of processed frames per second in real-time. However, the benefits are several considering the low-cost hardware employed with the current achievable frame rate of 5Hz, fitting well the most typical walking/ swimming speed scenario of the target applications. In fact, indoor, the agent speed is mainly limited by the exposure time of the cameras rather than by the processing of the data in real-time.

The stereo configuration provides metric 3D measurements of the scene for each epoch without the need to use scalebars placed on the object. Although monocular non-contact techniques exist using for example IMU sensors, the use of a stereo camera provides better accuracy than monocular based vSLAM (Campos et al., 2021). The known relative orientation between the two cameras or even simply the baseline determined with a geometric calibration of the system can be inputted as known constraints in the bundle adjustment. The imaging components are controlled by the computing unit which receives synchronized stereo pairs at every triggering command. Although hardware synchronisation is possible, with the currently implemented software trigger we observed a maximum synchronisation error of less than a ms, verified with an own developed led stopwatch. For walking and swimming speeds characterising the application of interest, this synchronisation error can be neglected. The stereo images are first pre-processed (i.e. undistorted and stereo-rectified) and then used by the computing unit to continuously estimate in real-time the pose of the system and a three-dimensional reconstruction of the environment in the form of a sparse point cloud.



**Figure 3.** GuPho's prototypes used respectively for terrestrial (a,c) and underwater (b,d) photogrammetric surveys. The main hardware components are shown: 1) cameras; 2) computing unit; 3) visualisation unit; 4) battery; 5) light panel//supports

IMAGING			
Model	Sensor	Lenses	
Daheng Imaging MER-131- 210U3C USB3.0 Cameras (x2)	1/2" Global shutter CMOS1280x1024 (pixel size 4.8µm)	4.0 mm rectilinear wide angle 1.85 mm fisheye lens	
COMPUTING UNIT			
Model	CPU	RAM	Disk
Raspberry Pi 4 model B	Cortex-A72 1.5 GHz	8GB	128 GB SD
VISUALISATION AND CONTROL			
Model	CPU	RAM	GPU
Samsung S9 (6.2 inches screen)	Qualcomm 845	6GB	Adreno 630
BATTERY	WEIGHT		
5V/3A 10400mAh power bank	1.4 kg without light (terrestrial version) 5.3 kg above the water without lights (underwater version) 0 Kg (neutral buoyancy) with lights once immersed in fresh water (underwater version)		

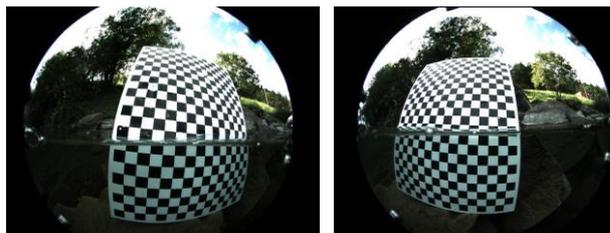
**Table 1.** main hardware components and characteristics of GuPho prototypes

These estimated poses are then exploited by four different modules to control the saving of the images, provide feedback on the acquisition distance and speed, and update the camera acquisition parameters such as the camera exposure that is optimized on the reconstructed 3D scene, excluding parts of the scene not of interest (e.g. sky, foreground, background). The sparse three-dimensional reconstruction, the poses of the saved images, and the quality control feedback are instantly and incrementally made available to the user through the visualization unit and a custom viewer. After the acquisition, the saved images can be further processed with SfM and MVS pipelines commonly available today in many photogrammetric software applications. The full resolution metric capabilities of the proposed procedure are thus obtained post-survey, in the office, but leveraging the vSLAM output provided in real-time by the system to speed up the process such as skipping the image orientation phase or using the image poses as initial approximations for an off-line bundle adjustment.

### 3.4 System calibration

The system calibration is performed by acquiring a set of images of the testfield available in the 3DOM laboratory for the terrestrial prototype. For the underwater prototype a portable testfield, or targets and scale bars are placed at the operative depth of the survey. The images are acquired according to regular self-calibration protocols including rolled and convergent images of a three-dimensional testfield. The interior as well exterior relative orientation parameters are computed through a joint bundle block adjustment of the two cameras using baseline constraints. The baseline distance between the two cameras is considered fixed and unknown. The relative orientation is thus extracted from the bundle using the procedure described in Nocerino et al. (2021). Once computed, the calibration parameters are kept fixed during the vSLAM process. For the underwater prototype, called “Frog”, the spherical dome surface were centred with respect to the entrance pupil position (for paraxial rays) using the optical alignment procedure described in Menna et al. (2016b) or adjusted after computing the decentring offsets as proposed in Rofallski et al. (2022). Figure 4 shows a split view of a partially immersed checkerboard respectively for the left and right cameras showing no visible refraction effects (except for the very borders of the image format). This design makes it possible to use the device both above and under the water with the same calibration when relaxed accuracy requirements exist. Indeed, in our tests the vSLAM algorithm could track, relocalise the agent, and close the

loops above the water even using an underwater calibration providing length measurement errors of less than 2% on 2m scalebars. Nevertheless, the best accuracy is always achieved with the calibration performed in the respective medium, right before or after the survey where we obtained sub-millimetre length measurement errors over the calibrated scale bars, both above and under the water at very close range (GSD better than 0.5mm). Table 2 shows the asymmetric relative orientation of the right camera with respect to the left one for different configurations.



**Figure 4.** Split views of left and right fisheye images with both the domes and checkerboard partially immersed in water to show the absence of significant refractive effects and thus a proper centering procedure of the domes.

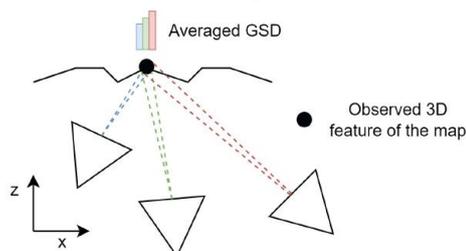
### RIGHT CAMERA ASYMMETRIC RELATIVE ORIENTATION

[mm]	[deg]		
$b \pm \sigma_b$	$\omega \pm \sigma_\omega$	$\phi \pm \sigma_\phi$	$\kappa \pm \sigma_\kappa$
Rectilinear - terrestrial			
$245.32 \pm 0.11$	$-0.094 \pm 0.039$	$-22.112 \pm 0.012$	$-0.024 \pm 0.034$
Fisheye - terrestrial			
$305.93 \pm 0.17$	$0.405 \pm 0.013$	$1.970 \pm 0.016$	$0.042 \pm 0.007$
Fisheye - underwater			
$244.09 \pm 0.24$	$-0.175 \pm 0.048$	$-0.929 \pm 0.025$	$2.754 \pm 0.042$

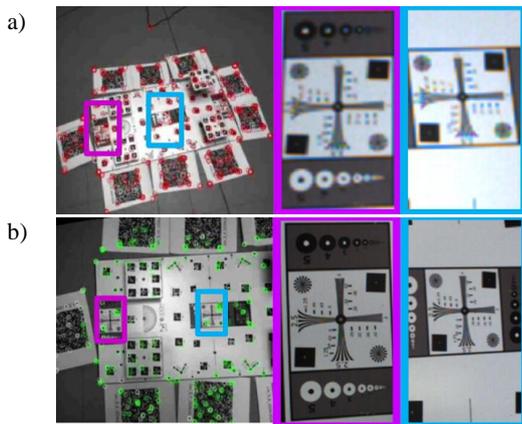
**Table 2.** Asymmetric relative orientation of the right camera with respect to the left one for different configurations.

### 3.5 Ground sample distance real-time feedback

In a typical survey, the GuPho system is set to reach a target ground sample distance range (GSD) for example between 1 and 2 mm. As soon as the survey is started, a sparse point cloud is triangulated by the system over stereo matched ORB features and shown on the display both in 3D and on the images. Each point is coloured according to whether it is within (green) or not (red) the proper distance to meet the planned GSD range. Points too close are visualised in blue (Figure 5). As the agent starts to move, new points are matched across the stereo pairs and with the previous frames. The average GSD is then displayed. Red areas (Figure 6a) of the scene can be thus “fixed” by getting closer until the average GSD for those points has not changed to green (Figure 6b). Moreover, the average distance of the scene seen by the stereo camera is shown on the screen coloured according to the same criteria described above. When the distance is red the object is too far and the agent needs to get closer until the distance becomes green (blue viceversa).



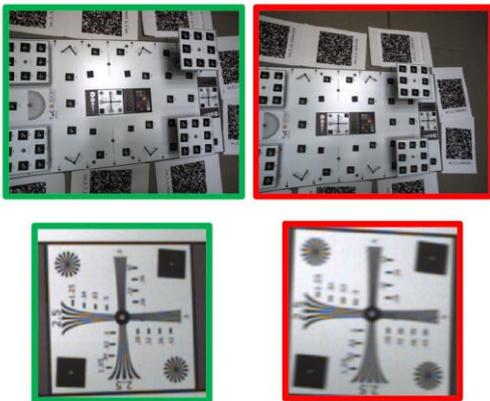
**Figure 5.** Principle of the average GSD used in GuPho and colour schema with red points too far, green within the proper distance range and blue too close (Torresani, 2022).



**Figure 6.** Example of real-time GSD feedback where red tie points (a) become green (b) once the agent gets to the proper distance to meet the planned GSD (in this example about 2 mm) as verified using the resolution wedges of the testfield (Torresani et al., 2021).

### 3.6 Motion blur real-time feedback

Given a specific exposure time of the cameras, the agent can move up to a maximum speed allowed to avoid that motion blur affects the image sharpness (Torresani et al., 2021). We display the speed of the agent computed over the last two successive stereo frame with the same colour schema described in 3.5. Figure 7 shows a demonstration of how an uncontrolled speed introduces motion blur effects. These features allow to optimise the survey in real-time by using always the maximum speed allowed by the current exposure time. If necessary, the user can increase the gain of the cameras to reduce the exposure time and speed up the survey.

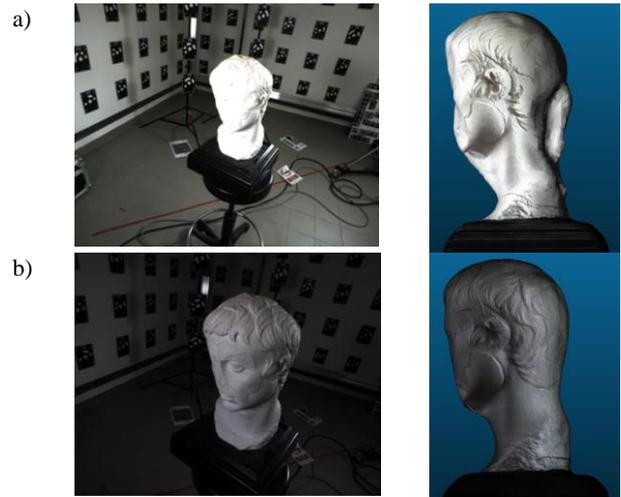


**Figure 7.** Resolution patterns extracted from two images captured while the speed was within the allowed range (left in green) and while a red warning was shown on the display (right in red). In this case the 2mm target GSD cannot be reached because of the introduced motion blur.

### 3.7 Image capturing strategies

In GuPho there are two strategies to save the images according to a specified overlap set by the user (for example 80%). The first is based on standard rules of aerial photogrammetry. Assuming a flat object surveyed at a constant flying height, the next image is saved only when the distance from the previously saved image is equal to the expected baseline. The other rule is similar to the keyframe selection criterium used by OpenVSLAM with an image stored if the number of newly triangulated points, not matching the existing 3D sparse point cloud (map), are above a

specific threshold. The proposed method provides an efficient image storage where, if the surveyor remains still during the tracking, no images are stored in memory. In an outdoor test presented in Torresani et al (2021), we measured a storage saving in terms of number of images with respect to a timed image acquisition of 1Hz of at least 5 times, with only 271 against 1080 images that would have been recorded during the 18 minutes acquisition.



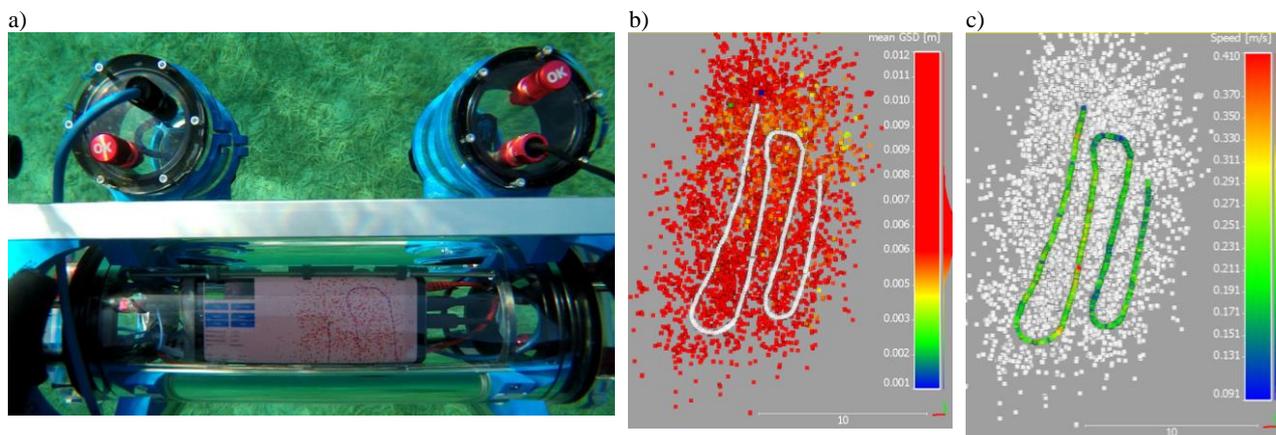
**Figure 8.** Comparison between in-camera (a) and proposed (b) exposure control algorithm showing one of the recorded images and the corresponding reconstructed 3D mesh. The proposed method (b) is independent of foreground and background scene illumination and thus much more reliable and robust.

### 3.8 Exposure control

In GuPho we implement a new exposure control algorithm that adjusts the shutter speed (the aperture value remains fixed during the survey) with a proportional controller using only the parts of the image corresponding to the object of interest. Instead of using the entire image we select 5x5 pixel image patches around each stereo matched ORB feature discarding those for which the distance is not within the range that gives the desired GSD. In Figure 8 we show an example of exposure control for a small white statue standing against a darker background. When using the automatic exposure implemented in the camera firmware, overall, the images are exposed correctly but at a closer look the object of interest is oversaturated and has lost texture and details. With the novel exposure control proposed by the authors, the exposure is adjusted only around the ORB features belonging to the statue, no matter where the statue appears in the image (middle or in the corner), as long as it is within the target distance. This simple, yet effective exposure method allows to filter any background or foreground interference providing better results both in real-time tracking, and in the high resolution photogrammetric post-processing, as shown for the computed mesh in Figure 8.

## 4. PHOTOGRAMMETRIC PROCESSING CASE STUDIES

In the following subsections we show some key features of the photogrammetric processing carried out post-survey by leveraging the real-time outputs of the vSLAM algorithm. We present some relevant case studies to qualitatively and quantitatively show the improvements of the implemented procedures.



**Figure 9.** Outputs of the VSLAM process in real time useful for quality check. Trajectory of an underwater survey over seagrass meadow (*Posidonia Oceanica*) as seen by the diver while being recorded in real-time (a) and corresponding sparse point cloud coloured according to the average GSD (b) or according to the swimming speed (c) for motion blur analysis.

#### 4.1 vSLAM outputs and visual quality control

GuPho organises the images collected in real-time in folders for the left and right cameras and named according to the system timestamp. The parent folder contains: camera calibration, trajectory, sparse point cloud (the map), a database containing the keyframe poses, ORB features, bag-of-words (BoW) representation of keyframes, and the co-visibility graph for a successive reuse of the map.

The image trajectory file lists all the camera poses at the maximum frequency (up to 5Hz for our system) each containing:

- the timestamp;
- the left camera exterior orientation (position and orientation as quaternion);
- additional metadata useful for real time visualisation as well as post-survey checks, such as median distance of stereo-matched tie points, speed of the agent, exposure time, gain, etc.

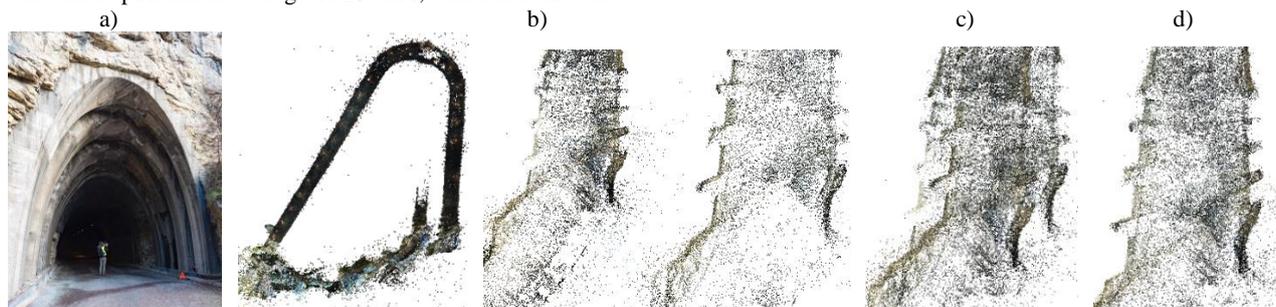
Similarly, the sparse point cloud contains the 3D coordinates for each point along with the mean GSD and multiplicity, i.e. the number of stereo pairs in which the point was observed.

Figure 9a shows an example of the display of the underwater prototype with the trajectory of the parallel transects while being recorded in real-time over seagrass meadow (*Posidonia Oceanica*) in Sardinia, Italy in very shallow water (maximum depth 3m). Figure 9b shows the corresponding sparse point cloud coloured according to the average GSD for each point providing a check on the achieved GSD over the object (in the example planned to be 3 mm). Figure 9c shows the swimming speed for each camera position in average of 20 cm/s, which is below the

motion blur limit for the maximum shutter speed recorded of 1ms.

#### 4.2 Leveraging the pre-calibration in a stereo camera system

The post-survey workflow uses the system calibration to improve the accuracy of the photogrammetric products. Not only are these already scaled but they are also more accurate than if obtained with a monocular system. Indeed, Figure 10 shows an example related to a dataset collected on the Dòs Trent: an old tunnel in Trento, Italy considered a heritage site. The figure shows the main entrance where the loop closure is analysed. The dataset, a 450m long loop was processed using different strategies after importing the images in Agisoft Metashape ([www.agisoft.com](http://www.agisoft.com)). Figure 10b shows the sparse point cloud after a regular SfM procedure using the recorded images without leveraging the relative orientation of the stereo camera and thus performing only an average scale a posteriori, after the bundle adjustment. This represents an unconstrained sequential image orientation and bundle adjustment without loop closure, with the error accumulation that produces an evident scale drift after returning at the entrance of the tunnel up to about 40% difference in scale, with the entrance split in two sections for a loop error of about 17m. Using the baseline constraint (Figure 10c) in the bundle adjustment provides a substantial improvement with little scale drift (about 1.5%) when comparing the distance between the same target measured, respectively, only in the images belonging to the beginning and to the end of the trajectory the loop closure error reduces to about 3.8m.



**Figure 10.** Photogrammetric processing of the Dòs Trent tunnel in Trento, Italy showing the main entrance (a) where the loop closure is analysed. In (b) the sequential image orientation (no loop closure) is done without stereo baseline constraints in the bundle adjustment, in (c) baseline constraints are added and in (d) the poses of (c) are used as initial approximation for a more exhaustive matching (loop closure). For more details refer to subsection 4.2.



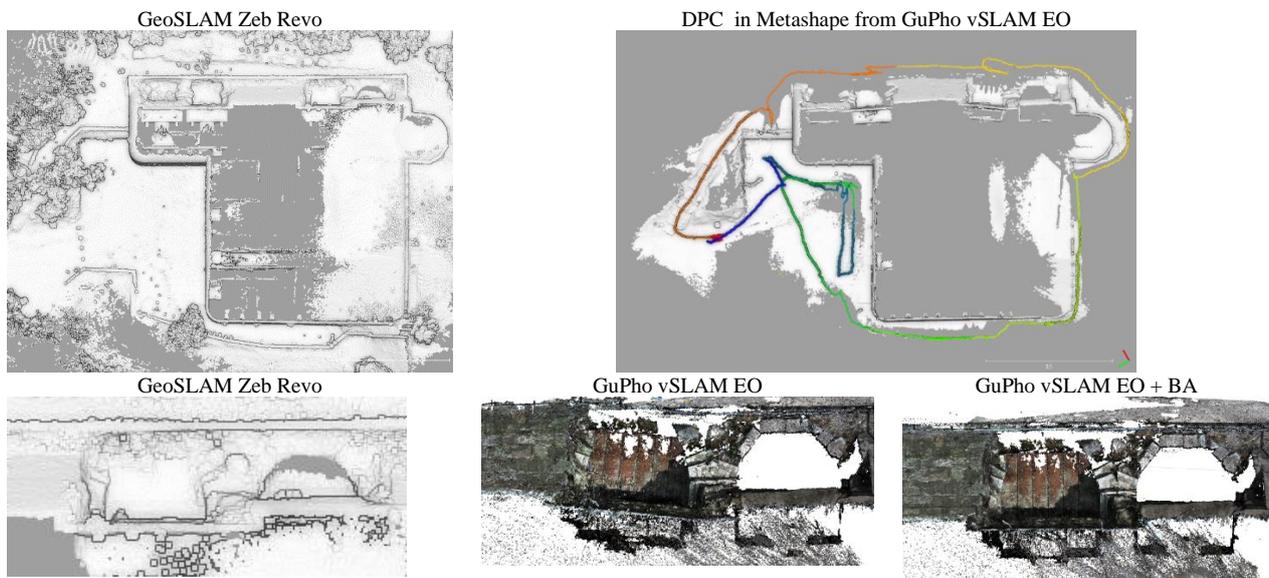
**Figure 11.** A close up view of the main entrance of Batteria di Mezzo fort in Riva del Garda, Italy (left) and the two point clouds of its facade, surveyed with the GeoSLAM Zeb Revo (middle), and the proposed prototype (right), respectively.

Using the trajectory obtained in Figure 10c as initial guess for the image orientation, with an exhaustive matching of the tie points, bundle adjustment, and thus “loop closure”, the scaling error is reduced to 0.3% and the loop closure is 3.6 cm (Figure 10d).

#### 4.3 Exploiting the vSLAM outputs to speed up the photogrammetric process

The vSLAM sensor poses (exterior orientation – EO) can be directly used to generate a dense point cloud (DPC) through multi view stereo. In Torresani et al. (2021) the authors showed that in a single loop survey, the image orientation obtained in real-time through the vSLAM algorithm, provides results not much different from those obtained by orienting the images with a SfM process. We further investigated the possibility to directly use the vSLAM poses without any additional adjustment and found out that, in well textured scenes and not challenging illumination conditions, the results previously found are confirmed. Nevertheless, the benefits of a full bundle adjustment translates in an improved reliability of the products at little extra cost in terms of computational time. Figure 11 shows some results of a survey campaign carried out using different mobile mapping devices such as GuPho and the GeoSLAM Zeb Revo for 3D documentation of the fort called Batteria di Mezzo, a military heritage site in Riva del Garda, Italy. Here the GuPho prototype was equipped with rectilinear lenses and despite the low resolution of the cameras, the level of detail of the dense cloud

generated in Metashape is comparable if not better than the GeoSLAM Zeb Revo. However, due to the reduced field of view of the rectilinear lenses, some passages around the structure were particular challenging, causing orientation problems visible in the dense cloud shown in figure 12. By running the bundle adjustment (BA) with the baseline constraint the orientation issue was fixed and the dense point cloud produced became comparable to the one surveyed with the GeoSLAM Zeb Revo with differences between the two clouds of only few centimetres. Table 3 compares the computation time of different post survey workflows on a Thinkpad 51 laptop with an Intel® Core™ i7-7820HQ CPU 2.90Ghz and Nvidia Quadro M1200 GPU. When importing the GuPHo EO in Agisoft Metashape, the tie point extraction and matching (column D) is necessary to have 3D seeds for any successive product such as DPC generation through MVS. For the Batteria di Mezzo dataset, the BA requires only an additional minute (column C), while the image orientation leveraging the approximate values (column B) triples the minimum time. A standard workflow in Metashape (column A) would require, for the same number of images, about 14 times more than the minimum necessary using the vSLAM approximate EO. We did not find significant differences in accuracy between the processing A and B in table 3, still the approximate EO provides a computational time reduction of a factor almost 5.



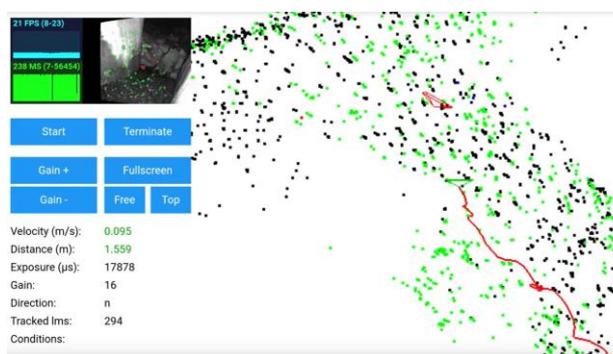
**Figure 12.** Top view of the two point clouds of the Batteria di Mezzo fort in Riva del Garda, Italy surveyed respectively with the GeoSLAM Zeb Revo (left), and the proposed prototype (right). See subsection 4.3 for further details.

	A	B	C	D
	Standard Workflow Image orientation + bundle	Tie point extraction + Image orientation from approximate values	Tie point extraction + BA	Tie point extraction
Matching Time	1h 51m	0h 09m	0h 9 m	0h 9m
Alignment Time (+BA)	0h 12m	0h 17m	-	-
Optimization Time (BA)		-	0h 1m	-
<b>Total time</b>	<b>2h 03m</b>	<b>0h 26m</b>	<b>0h 10m</b>	<b>0h 9m</b>

**Table 3.** Computation times for the different processing strategies in the case study shown in Figure 12. See subsection 4.3 for further details.

#### 4.4 Map reuse for monitoring and re-photographing

As described in 4.1, once a survey is completed, the map can be stored to be reused in a successive survey. Figure 13 shows an example of relocation and mapping for rephotographing and monitoring a vault collapse in a tunnel. The figure shows the screen display where a previously taken image is represented by a red pyramid. The current camera pose is displayed in green together with the newly mapped 3D points and the trajectory in red. New images are stored and labelled when the new pose gets closer enough to the red one.



**Figure 13.** Example of reuse of the map for 3D monitoring purposes.

## 5. CONCLUSIONS AND FUTURE WORKS

We reported the latest developments about GuPho, a new mobile mapping concept based on photogrammetry that leverages the vSLAM technique to support and ease the image acquisition in the field. The visual feedback on GSD and motion blur, and the two implemented image capturing criteria relieve the photogrammetrist from time consuming operations in the field to verify the proper imaging network geometry. These features proved particularly useful and critical in GNSS denied environment, like in close range photogrammetry indoor, and especially underwater, where the unstructured environment, the freedom of motion in any direction, and the absence of well recognisable landmarks make it difficult for the surveyor to navigate and position her/himself with respect to the area of interest. The automatic exposure algorithm tracks the parts of the object at the proper distance to compute the exposure making it robust in high contrast or dynamic scenes. The calibration procedures, and the chosen design for the underwater prototype allow to use the system without particular modifications to the implemented algorithms, observing similar precisions for the relative orientation parameters and length measurement errors. Further tests are necessary to validate the (mechanical) stability of these calibrations, especially underwater, where pressure and temperature variations put the device under extreme mechanical stresses. The implemented image acquisition strategies make the system more efficient with respect to a simple 1Hz timed acquisition, reporting a storage save of up to 5 times during

prototype testing. In the future we will improve the graphical feedback and user interface for a better real-time experience. Further accuracy tests above and under the water will be planned in the future in particular to evaluate the system robustness to dynamic environments such as in presence of caustics in shallow water or over moving seagrass, such as the *Posidonia Oceanica*. Preliminary tests were encouraging against moderate caustics on a quasi static seagrass meadow. Also, stress tests regarding maximum acquisition time will be carried out. In a recent test of the system done underwater in the Garda Lake, Italy, we continuously tracked the lake bottom for 74 minutes down to 32m depth in a 300 m long loop. In the future we will extend the motion blur to rotations as at the moment we only consider translational motion blur, more relevant in standard photogrammetric transects.

## ACKNOWLEDGEMENTS

This work is partly supported by EIT RawMaterials GmbH under Framework Partnership Agreement No. 19018 (AMICOS - Autonomous Monitoring and Control System for Mining Plants). The authors are thankful to Fabio Mosna, Sara Messina, Andrea Marconi, Carlo Longin, Andrea Montagner and Emiliano Trentinaglia from Rane Nere Sub Trento for the support during diving operations. Authors also thank Matteo Rapanà for giving access to the Fort in Riva del Garda.

## REFERENCES

- Baiocchi, V., Barbarella, M., Del Pizzo, S., Giannone, F., Troisi, S., Piccaro, C. and Marcantonio, D., 2017. AUGUSTO'S Sundial: Image-Based Modeling for Reverse Engineering Purposes. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, p.63.
- Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., Reid, I. and Leonard, J.J., 2016. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on robotics*, 32(6), pp.1309-1332.
- Campos, C., Elvira, R., Rodríguez, J.J.G., Montiel, J.M. and Tardós, J.D., 2021. Orb-slam3: An accurate open-source library for visual, visual-inertial, and multimap slam. *IEEE Transactions on Robotics*, 37(6), pp.1874-1890.
- Di Stefano, F., Torresani, A., Farella, E.M., Pierdicca, R., Menna, F., Remondino, F., 2021. 3D Surveying of Underground Built Heritage: Opportunities and Challenges of Mobile Technologies. *Sustainability*, Vol.13, 13289
- Drap, P., Merad, D., Hijazi, B., Gaoua, L., Nawaf, M.M., Saccone, M., Chemisky, B., Seinturier, J., Sourisseau, J.C., Gambin, T. and Castro, F., 2015. Underwater photogrammetry and object modeling: a case study of Xlendi Wreck in Malta. *Sensors*, 15(12), pp.30351-30384.

- Fraser, C.S., 1984. Network design considerations for non-topographic photogrammetry. *Photogrammetric Engineering and Remote Sensing*, 50(8), pp.1115-1126.
- Menna, F., Nocerino, E., Remondino, F., Saladino, L. and Berri, L., 2020. Towards online UAS-based photogrammetric measurements for 3D metrology inspection. *The Photogrammetric Record*, 35(172), pp.467-486.
- Menna, F., Nocerino, E., Nawaf, M.M., Seinturier, J., Torresani, A., Drap, P., Remondino, F. and Chemisky, B., 2019, June. Towards real-time underwater photogrammetry for subsea metrology applications. In OCEANS 2019-Marseille (pp. 1-10). IEEE.
- Menna, F., Nocerino, E., Remondino, F., Dellepiane, M., Callieri, M. and Scopigno, R., 2016a. 3D digitization of an heritage masterpiece-a critical analysis on quality assessment. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 41.
- Menna, F., Nocerino, E., Fassi, F. and Remondino, F., 2016b. Geometric and optic characterization of a hemispherical dome port for underwater photogrammetry. *Sensors*, 16(1), p.48.
- Menna, F., Nocerino, E., Troisi, S. and Remondino, F., 2013, May. A photogrammetric approach to survey floating and semi-submerged objects. In Videometrics, Range Imaging, and Applications XII; and Automated Visual Inspection (Vol. 8791, pp. 117-131). SPIE.
- Menna, F., Remondino, F., Battisti, R. and Nocerino, E., 2011, June. Geometric investigation of a gaming active device. In Videometrics, Range Imaging, and Applications XI (Vol. 8085, pp. 173-187). SPIE.
- Mur-Artal, R. and Tardós, J.D., 2017. Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE transactions on robotics*, 33(5), pp.1255-1262.
- Mur-Artal, R., Montiel, J.M.M. and Tardos, J.D., 2015. ORB-SLAM: a versatile and accurate monocular SLAM system. *IEEE transactions on robotics*, 31(5), pp.1147-1163.
- Nawaf, M.M., Merad, D., Royer, J.P., Boi, J.M., Saccone, M., Ben Ellef, M. and Drap, P., 2018. Fast visual odometry for a low-cost underwater embedded stereo system. *Sensors*, 18(7), p.2313.
- Nocerino, E., Rodríguez-González, P. and Menna, F., 2019. Introduction to mobile mapping with portable systems. In *Laser Scanning* (pp. 37-52). CRC Press.
- Nocerino, E., Menna, F., Toschi, I., Morabito, D., Remondino, F. and Rodríguez-González, P., 2018. Valorisation of history and landscape for promoting the memory of WWI. *Journal of Cultural Heritage*, 29, pp.113-122.
- Nocerino, E., Poiesi, F., Locher, A., Tefera, Y.T., Remondino, F., Chippendale, P. and Van Gool, L., 2017. 3D reconstruction with a collaborative approach based on smartphones and a cloud-based server. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42(W8), pp.187-194.
- Nocerino, E., Menna, F., Remondino, F. and Saleri, R., 2013. Accuracy and block deformation analysis in automatic UAV and terrestrial photogrammetry-Lesson learnt. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2(5/W1), pp.203-208.
- Ortiz-Coder, P. and Sánchez-Ríos, A., 2020. An integrated solution for 3D heritage modeling based on videogrammetry and V-SLAM technology. *Remote Sensing*, 12(9), p.1529.
- Otero, R., Lagüela, S., Garrido, I. and Arias, P., 2020. Mobile indoor mapping technologies: A review. *Automation in Construction*, 120, p.103399.
- Pagliari, D., Menna, F., Roncella, R., Remondino, F. and Pinto, L., 2014. KINECT FUSION IMPROVEMENT USING DEPTH CAMERA CALIBRATION. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 45.
- Remondino, F., Nocerino, E., Toschi, I. and Menna, F., 2017. A critical review of automated photogrammetric processing of large datasets. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42, pp.591-599.
- Remondino, F., Spera, M.G., Nocerino, E., Menna, F. and Nex, F., 2014. State of the art in high density image matching. *The photogrammetric record*, 29(146), pp.144-166.
- Rodríguez-González, P., Nocerino, E., Menna, F., Minto, S. and Remondino, F., 2015. 3D surveying & modeling of underground passages in WWI fortifications. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, 40, p.17.
- Rofallski, R., Menna, F., Nocerino, E. and Luhmann, T., 2022. An efficient solution to ray tracing problems for hemispherical refractive interfaces. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2, pp.333-342.
- Schöps, T., Sattler, T., Häne, C. and Pollefeys, M., 2017. Large-scale outdoor 3D reconstruction on a mobile device. *Computer Vision and Image Understanding*, 157, pp.151-166.
- Sumikura, S., Shibuya, M. and Sakurada, K., 2019, October. OpenVSLAM: A versatile visual SLAM framework. In Proceedings of the 27th ACM International Conference on Multimedia (pp. 2292-2295).
- Taketomi, T., Uchiyama, H. and Ikeda, S., 2017. Visual SLAM algorithms: A survey from 2010 to 2016. *IPSP Transactions on Computer Vision and Applications*, 9(1), pp.1-11.
- Torresani A., 2022. A portable V-SLAM based solution for advanced visual 3D mobile mapping. PhD dissertation, Trento, Italy.
- Torresani, A., Menna, F., Battisti, R., Remondino, F., 2021. A V-SLAM Guided and Portable System for Photogrammetric Applications. *Remote Sensing*, Vol.13(12), 2351.