Automatic scale estimation of structure from motion based 3D models using laser scalers in underwater scenarios

Klemen Istenic, Nuno Gracias, Aurélien Arnaubec, Javier Escartin, Rafael Garcia

ISPRS Journal of Photogrammetry and Remote Sensing 159 (2020) 13–25

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Automatic scale estimation of structure from motion based 3D models using laser scalers in underwater scenarios

Klemen Istenič[a,b,*], Nuno Gracias[a], Aurélien Arnaubec[c], Javier Escartín[d], Rafael Garcia[a]

[a] Underwater Robotics Research Center (CIRS), Computer Vision and Robotics Institute (VICOROB), University of Girona, Edifici P-IV, Campus de Montilivi, 17071 Girona, Spain
[b] Coronis Computing, S.L., Science and Technological Park of UdG, Carrer Pic de Peguera, 15, 17003 Girona, Spain
[c] IFREMER, Ctr Mediterranee, Unité Syst. Marins, CS 20330, F-83507 La Seyne Sur Mer, France
[d] Université de Paris, Institut de Physique du Globe de Paris, CNRS, F-75005 Paris, France

ARTICLE INFO

Keywords:
Structure-from-motion
Underwater 3D reconstruction
Photogrammetry
Laser scalers

ABSTRACT

Improvements in structure-from-motion techniques are enabling many scientific fields to benefit from the routine creation of detailed 3D models. However, for a large number of applications, only a single camera is available for the image acquisition, due to cost or space constraints in the survey platforms. Monocular structure-from-motion raises the issue of properly estimating the scale of the 3D models, in order to later use those models for metrology. The scale can be determined from the presence of visible objects of known dimensions, or from information on the magnitude of the camera motion provided by other sensors, such as GPS.

This paper addresses the problem of accurately scaling 3D models created from monocular cameras in GPS-denied environments, such as in underwater applications. Motivated by the common availability of underwater laser scalers, we present two novel approaches which are suitable for different laser scaler configurations. A fully unconstrained method enables the use of arbitrary laser setups, while a partially constrained method reduces the need for calibration by only assuming parallelism on the laser beams and equidistance with the camera. The proposed methods have several advantages with respect to existing methods. By using the known geometry of the scene represented by the 3D model, along with some parameters of the laser scaler geometry, the need for laser alignment with the optical axis of the camera is eliminated. Furthermore, the extremely error-prone manual identification of image points on the 3D model, currently required in image-scaling methods, is dispensed with.

The performance of the methods and their applicability was evaluated both on data generated from a realistic 3D model and on data collected during an oceanographic cruise in 2017. Three separate laser configurations have been tested, encompassing nearly all possible laser setups, to evaluate the effects of terrain roughness, noise, camera perspective angle and camera-scene distance on the final estimates of scale. In the real scenario, the computation of 6 independent model scale estimates using our fully unconstrained approach, produced values with a standard deviation of 0.3%. By comparing the values to the only other possible method currently usable for this dataset, we showed that the consistency of scales obtained for individual lasers is much higher for our approach (0.6% compared to 4%).

1. Introduction

In increasing number of remote sensing applications photogrammetry is used to obtain reliable geometric information about the environment. These optical-based reconstruction procedures, generally based on the Structure from Motion (SfM) approach, have gained significant popularity due to multiple factors. The improvements in both speed and robustness of many image processing techniques (Snavely et al., 2008; Remondino et al., 2008; Agarwal et al., 2009; Triggs et al., 1999) together with the increased computational capabilities of commonly available processing hardware, enable nowadays nearly black-box type of data processing, where there is little to no need for user intervention. The abundance of low cost cameras that can easily be mounted on a variety of vehicles, or used hand-held, has further

* Corresponding author at: Underwater Robotics Research Center (CIRS), Computer Vision and Robotics Institute (VICOROB), University of Girona, Edifici P-IV, Campus de Montilivi, 17071 Girona, Spain.
E-mail addresses: klemen.istenic@gmail.com (K. Istenič), ngracias@silver.udg.edu (N. Gracias), aurelien.arnaubec@ifremer.fr (A. Arnaubec), escartin@ipgp.fr (J. Escartín), rafael.garcia@udg.edu (R. Garcia).

https://doi.org/10.1016/j.isprsjprs.2019.10.007
Received 19 June 2019; Received in revised form 14 October 2019; Accepted 15 October 2019
Available online 14 November 2019

Fig. 1. ROV VICTOR 6000 (IFREMER), used among other, in the SUBSAINTES 2017 cruise (doi:10.17600/17001000).

spearheaded the widespread application of these techniques in a variety of fields (e.g., Wallace et al., 2016; Javernick et al., 2014; Anderson and Gaston, 2013).

Concurrently, the field of underwater photogrammetry has also grown considerably with the availability of underwater vehicles. Whereas traditional aerial and terrestrial vehicles are increasingly equipped with single or multi-camera set-ups (e.g., stereo cameras, multi-camera systems), most underwater remotely operated vehicles (ROVs) and autonomous underwater vehicles (AUVs) that are nowadays used in scientific missions (e.g., VICTOR 6000 from IFREMER depicted in Fig. 1) have limited optical sensing capabilities. Common optical systems consist of a single main camera used by the ROV-pilot or, in the case of the larger workclass ROVs, also of additional cameras for maneuvering. As these are typically unsynchronized and have non-overlapping fields-of-view, they are not suited for stereo image processing. While multi-camera underwater metrology systems are starting to appear as commercially available products and services (Rovco, 2019; Comex, 2019), such systems are still too large and expensive for most ROV science applications. Nonetheless, the ability to produce accurate 3-dimensional (3D) models from monocular cameras despite the unfavorable properties of the water medium (i.e., light attenuation and scattering, among other effects) has given scientists unprecedented access to the underwater environment and its ecosystems, from shallow waters (Pizarro et al., 2017; Storlazzi et al., 2016; Rossi et al., 2019) to the deep ocean (Bingham et al., 2010; Escartín et al., 2016; Bodenmann et al., 2017).

Performing SfM based reconstruction using single camera imagery has an important limitation as it precludes obtaining a metric scale of the resulting model. The image formation process of projecting the 3D world onto 2-dimensional (2D) image planes obviously causes the loss of a dimension. When performing the reconstruction, this results in scale ambiguity, i.e. the estimated parameters of 3D structure and camera trajectory can be multiplied by an arbitrary factor and still give rise to the same image observations (Lourakis and Zabulis, 2013; Hartley and Zisserman, 2003). This also precludes or at least limits the possibility of conducting quantitative measurements based on geometric parameters (e.g., distances, areas, volumes, etc.) obtained from the models. To resolve the ambiguity, a general trend in aerial and terrestrial problems is to fuse the image measurements with other sensors (e.g., inertial navigation system (INS) (Spaenlehauer et al., 2017; Zhang and Singh, 2015) and Global Navigation Satellite System (GNSS) (SSoloviev and Venableoloviev and Venable, 2010; Mian et al., 2016)) or using ground control points (GCPs) (Eltner and Schneider, 2015; Mertes et al., 2017). These control points are extremely hard, if not impossible, to establish underwater, while the absorption of electromagnetic waves in water prevents the use of GPS. Hence the scale is normally disambiguated either using INS (Sedlazeck et al., 2009; Pizarro et al., 2009; Campos et al., 2016) or through the introduction of known distances between points in the scene (Garcia et al., 2011). It is worth noting that reliable displacement information may not be available in smaller ROVs, since this normally requires a dedicated INS complemented with a Doppler Velocity Log (DVL) (Ribas et al., 2011) and there are rarely any known measurements readily available in real underwater scenarios. The scale is therefore often determined by placing objects with known dimensions (e.g., scaling cube (Cocito et al., 2003), locknuts (Kalacska et al., 2018), graduated bars (Neyer et al., 2018), etc.) into the scene. While such an approach does not require any additional equipment (with the exception of auxiliary objects), it does however involve their transport and placement, which can be challenging in deep-sea environments.

Alternatively, the distance between known points on the model can also be established from the projections of laser beams with known geometry (Robert et al., 2017; Bergmann et al., 2011; Tusting and Davis, 1992). The use of laser scalers to provide an absolute size reference in photographs is one of their most widespread applications (Tusting and Davis, 1992, 1993). Their initial use dates back to the late 1980s (Tusting and Davis, 1986; Caimi and Tusting, 1987). To compensate for the lack of knowledge about the scene and camera-scene distance, these methods require a perfect alignment of parallel lasers with the camera, planarity of the scene surface and perpendicularity between the camera and the scene. Comparing the spacing between two laser spots on the image and the known beam spacing, any measurement in the plane of the lasers, regardless of the camera-to-scene range, should be correctly estimated.

Seen as the most restrictive requirement, the necessity of perpendicularity between the optical axis of the camera and the plane of the scene has been addressed in various approaches with the introduction of additional lasers and sensors. Wakefield and Genin (1987) first introduced the idea of perspective grids to enable oblique camera views. Although being an improvement, the method imposed additional constraints on the camera-scene distance (altitude) and fixed inclination angle. To provide additional information about the camera-scene relationship, more lasers have also been added to the systems. A configuration consisting of three lasers, two aligned with the optical axis of the camera and a third laser oriented at an angle, has been described by Davis and Tusting (1991). It enables the estimation of range and size of objects from direct scaling of the position of the light spots on the image. An underwater photogrammetric system using several sensors to provide precision navigation for benthic surveys is described in Kocak et al. (2002, 2004). One of them, the ring laser gyroscope, made for measuring pitch/roll motions is integrated into a custom software package which establishes the scale reference from the projections of the three beam laser system. To enable the measurement of distance between any two points on the image, Pilgrim et al. (2000) presented a multi-laser approach. It gains information about the camera's inclination angle and distance to the scene by using four parallel lasers positioned equidistant from the camera center together with either a fifth laser set at an angle parallel to the bottom or a side pair, similar to the three-beam approach. The method works under the assumption of scene flatness and the restraint of the camera in either pan or tilt planes with respect to the sea bottom. A more versatile method capable of determining an arbitrary tilt of a surface was presented by Davis and Tusting (1991) which requires four parallel lasers aligned with the optical axis of the camera.

Due to the lack of a better approach, image-scaling methods are still commonly used for scaling 3D models, and therefore require not only that the image containing the projections of lasers be acquired in flat areas of the scene, but also complex laser alignment with the optical axis of the camera. In real scientific cruises these strict rigidity constraints can be nearly impossible to maintain, due to the frequent need to mount and dismount equipment, especially when the camera is not

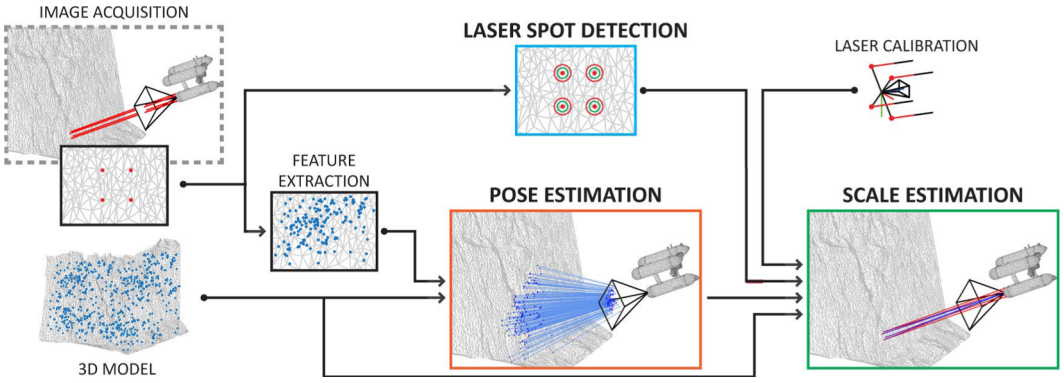

Fig. 2. Flowchart of the scale estimation process depicting three crucial steps in scale estimation: laser spot detection, pose estimation, and scale estimation.

rigidly attached to the laser scalers. As obtaining accurate geometrical information would entail repetitive calibration procedures, it significantly limits its usability. Furthermore, given that the image-scaling techniques only provide the estimated distance between points on an image, this information is not directly related to the model itself. In order to scale any model, a separate identification of these laser points has to be carried out on the model itself. As the identification of image points on the model is done manually, it is extremely error-prone and time consuming.

The main goal of this paper is to present two novel automated approaches to solve the scaling problem for SfM based 3D models, using commonly available laser scalers. The image information is exploited beyond the automatic location of laser spots, compensating for the known geometry of the laser scalers. The need for laser alignment with the optical axis, scene geometry or camera position are thus abolished together with the error-prone manual identification of 3D points on the model.

Each of the two proposed methods (i.e., fully unconstrained and partially constrained) is suitable for a different laser scaler configuration. While the fully unconstrained approach enables an arbitrary laser setup, the required rigidity between the lasers and the camera can be extremely limiting in real scenarios. To overcome this, we also present an alternative approach in which the required relation of the lasers to the camera is significantly relaxed at the cost of requiring the lasers to be parallel among themselves (but not necessarily with the optical axis). As the fully unconstrained method utilizes a fully-determined laser geometry, it is able to estimate the scale using a single laser, while the partial method requires a laser pair. Any additional laser measurements are used to further reduce the potential effect of noisy laser spot detections. These methods are considered universal, as they can be applied to standard imagery acquisitions, and are not linked to data acquired with specific sensors or hardware (e.g., stereo cameras). Hence, it is possible to process legacy data from previous missions acquired using different vehicles and imaging systems. However, as the novel methods require a description of a scene in a form of a 3D model, they cannot be utilized on moving objects (e.g., fish, benthonic species, soft corals, etc.) or in highly dynamic environments.

The results of our methods are validated using a 3D model constructed using real underwater data and comparing them to the results which would have been obtained using an image-scaling method supporting an arbitrary tilt of the surface (Davis and Tusting, 1991). The effects of noise, camera perspective angle and camera-scene distance on our process and final estimates of scale are further analyzed. Finally, the results of using our method to scale a model reconstructed from data acquired during the SUBSAINTES 2017 cruise (doi:10.17600/17001000) (Escartín et al., 2017) are presented.

2. Scaling of SfM-based 3D models

Optical-based 3D models are produced using a set of images through a sequential series of steps. A sparse set of 3D points representing the general 3D geometry of the scene can be obtained by exploiting multiple projections of the same 3D point in overlapping images through the equations of projective geometry (Hartley and Zisserman, 2003). By extracting salient features and matching them across the image set, the 3D locations of these points (the structure) are estimated together with the camera parameters (the motion) through a technique called Structure from Motion (SfM). An accurate and highly detailed description of the scene is subsequently obtained through a multi-view stereo densification process followed by an estimation of a surface from the noisy point cloud. The final photo-realistic 3D model representation is achieved by finding a consistent high-quality texture through seamlessly mapping input images to a high-resolution triangle representation of the surface. If the imagery used in the process was acquired using one or more unsynchronized cameras, and no other auxiliary data is used, it is impossible to determine the correct scale of the model. Such a result can be visually pleasing but cannot be used for further scientific purposes where knowledge of the distances, areas and volumes is required. Therefore, a scale estimation step is vital in the reconstruction for scientific purposes.

Nowadays, the most common uses of laser scalers are based on the image-scaling multi-laser approaches introduced by Pilgrim et al. (2000) and Davis and Tusting (1991). The requirements associated with these methods, i.e. laser alignment with the optical axis and manual identification of the image points on the 3D models, while originally reasonable, are becoming constricting in an increasing number of situations in which data for photogrammetry can be collected.

In this section, we present two novel methods for scale estimation, namely the fully unconstrained method (FUM) and the partially constrained method (PCM), suitable for different laser-scaler configurations and scenarios. Both methods, based on computer vision techniques of image localization and ray casting, exploit the information acquired with an optical image in which the intersection of laser beams with the scene (laser spots) are visible. Both methods consist of three main steps, as depicted in Fig. 2. The two initial steps are identical in both methods. First, a laser detection method is required to determine the locations of laser spots on an image. Secondly, the pose of the camera (wrt. the 3D model), at the moment at which the image was acquired, is estimated through a feature-based localization process. These estimations are used in the third step, which differs between the two methods and depends on available laser configuration information. The scale of the model is finally computed after determining the 3D position of laser beams intersecting with the scene.

It is worth noting that our approaches are independent of the method used for detecting laser spots on the image. Laser spots can be selected either manually, through a simple method (e.g., color

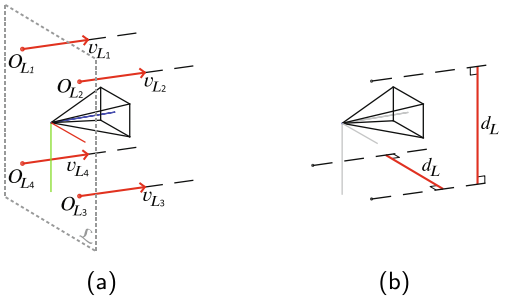

Fig. 3. (a) Fully- and (b) partially-calibrated setups of the measuring device (optical camera and separate lasers) with the required information marked in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

thresholding) or even with a more complex approach such as machine learning (Rzhanov et al., 2005; Schoening et al., 2015).

2.1. Measuring device

The measuring setup required consists of two devices commonly used in underwater surveying using ROVs and AUVs: A laser scaler, which can contain a variable number of lasers, and a monocular optical camera. If the laser geometry (origins O_L and directions v_L) with respect to the optical axis of the camera are known, the setup is considered fully calibrated (Fig. 3a). The origins are defined as points on a plane $\mathcal{L}$, which is perpendicular to the optical axis of the camera and contains the optical center, while the directions are unit vectors expressed wrt. the camera's optical axis.

Depending on the circumstances (e.g., multiple dives involving mounting and dismounting of equipment with associated misalignments), the strict rigidity constraints between the lasers and the camera are very difficult to maintain, especially if the camera and the laser scaler are not rigidly coupled. As any change would thus entail a new calibration procedure, which may not be feasible in certain situations, we also present an alternative approach, in which laser pairs have to be parallel, along with the sole condition that the camera be equidistant to their origins (Fig. 3b). As there is no requirement of parallelism between the laser beams and the optical axis of the camera, this partially constrained approach permits alterations in the orientation between the camera and laser scaler, making it more suitable for scenarios with multiple mounting and dismounting operations, or situations in which an accurate calibration procedure is not possible or unavailable. These relaxed constraints render the system more usable in practice.

2.2. Pose estimation

The scale estimation process requires the knowledge of the camera pose $\boldsymbol{P} = [\boldsymbol{R}^T | -\boldsymbol{R}^T\boldsymbol{t}] \in \mathbf{SE}(3)$ defined as the projection from the world to the camera frame at the moment the image was taken. As these images contain laser spots, they do not reflect the real appearance of the environment and are, as such, considered undesirable in the 3D reconstruction process, specially in the densification and texture mapping steps. To estimate the poses, images can potentially still be included in the SfM step (and excluded from rest) or, alternatively, the pose can be computed through a separate feature-based image localization method presented here.

Salient 2D features extracted from the image, are matched with a full set of features associated with the model's sparse set of 3D points. Feature detection and matching procedures can be adjusted for each specific dataset, and do not influence the scale estimation process, as long as it is possible to produce successful pairs of 3D-2D observations ($\mathcal{F} = \{X_k, x_j\}$). Such matches are then exploited to obtain an initial estimate of camera extrinsic parameters $\boldsymbol{P}$ (and possible camera intrinsics $\boldsymbol{K}$). In cases in which the camera has been calibrated, the solution is obtained by solving a minimal case ($n = 3$) of the Perspectiven-Point (PnP) problem (Ke and Roumeliotis, 2017), while alternatively a Direct Linear Transform (DLT) (Hartley and Zisserman, 2003) algorithm can be used. As feature observations are noisy and might contain outliers, the process must be carried out in conjunction with a robust estimation method A Contrario Ransac (AC-RANSAC) (Moisan et al., 2012). Initial parameter values are subsequently refined through a non-linear optimization. Using Bundle Adjustment (BA) the re-projection error of known (and fixed) 3D points and their 2D observation is minimized:

$$\min_{P,K} \sum_{\mathcal{F}} \|x_j - \text{proj}(\boldsymbol{K}, \boldsymbol{P}, \mathbf{X}_k)\|^2. \tag{1}$$

2.3. Scale estimation

The scale of a 3D model is obtained as the ratio between a known quantity m and its model based estimate $\hat{m}$:

$$s = \frac{m}{\hat{m}}. \tag{2}$$

Using the location of recorded and detected laser spots x_L and previously estimated parameters of the camera $\{\boldsymbol{K}, \boldsymbol{P}\}$, it is possible to predict the geometry of the laser scaler which produced the recorded results. Given that the prediction is based on the 3D model, it is directly affected by the scale of the model and can therefore be used to determine it. Depending on the availability of information about the geometry of the lasers and the camera, we can either use the distance between the laser origins and camera's optical center (FUM) or the perpendicular distance between the two parallel beams (PCM).

2.3.1. Fully unconstrained method

When the complete laser geometry (origins O_L and directions v_L) are known, the position from where the lasers must be emitted $\widehat{O}_L$ in order to produce the observed result can be determined regardless of potential non-parallelism between the lasers (Fig. 4). The position of origin of each laser can be estimated independently by exploiting the known direction of the laser beam and the determined position of the laser beam intersection with the scene X_L. As this point is seen on the image, the actual 3D point X_L had to be in the line-of-sight of the camera and can therefore be deduced using a ray casting procedure. The location is computed by finding the first surface of the 3D model which is intersected by a ray originating in the camera center and passing through the location of the detected laser spot on the image. Subsequently, to obtain the location of the origin, the point X_L expressed in the camera frame is back-projected according to a known direction of the beam v_L onto the plane $\mathcal{L}$ (Eq. (3)). Once known, the scale can be determined by comparing the displacement $\hat{m}_L = \|\widehat{O}_L\|$ with its *a priori* known value m_L.

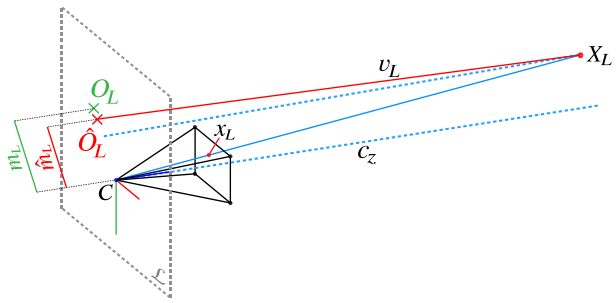

Fig. 4. Scale estimation procedure using the fully unconstrained approach, based on the 3D model and optical image depicting the laser beam projection of the laser intersection with the scene. The displacement of the predicted laser origin $\hat{m}_L$, obtained by projecting the 3D point X_L onto the plane $\mathcal{L}$ according to the known direction of the laser beam v_L, is compared to its known value m_L.

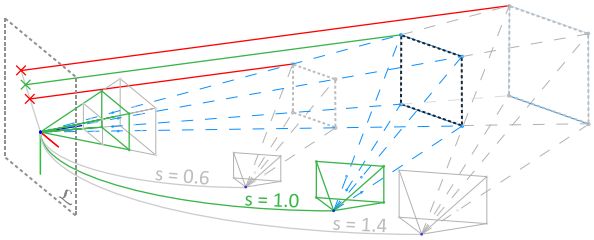

Fig. 5. The effect of scale ambiguity on a 3D reconstruction and its projection on images. The structure and motion can be multiplied by an arbitrary factor (incorrect – light gray, correct – black) and still produce projections in the same positions on the images. The error in scale can be determined by comparing the predicted locations of the laser origin (incorrect – red, correct – green cross). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$\widehat{O}_L = \boldsymbol{P}X_L - \frac{\boldsymbol{P}X_L \cdot c_z}{v_L \cdot c_z} v_L, \tag{3}$$

where c_z represents the optical axis of the camera.

Fig. 5 depicts the effect different scaling factors (affecting the 3D model) have on the displacement of the predicted laser origin (represented as cross on plane $\mathcal{L}$). Due to the scale ambiguity, all variations of the model (light gray dotted line), are valid solutions of the 3D reconstruction process. As shown, only the laser intersection point (blue) obtained using the model with the true scale (dark gray dotted line) produces the correct prediction of the laser origin on the plane $\mathcal{L}$ (green cross) by using known directions of the laser beams.

2.3.2. Partially constrained method

While fully unconstrained method enables the use of an arbitrary laser setup, the required rigidity between the lasers and the camera can be extremely limiting in certain real scenarios. To alleviate this, we present an alternative approach, in which the required relation between the camera and the lasers is significantly less rigid. This approach only requires the two lasers to be parallel and equidistant from the camera. As opposed to the existing image-scaling methods, the lasers do not have to be aligned with the optical axis of the camera. The scale of the model is therefore estimated by comparing a known perpendicular distance between the two parallel beams to the one estimated from the image and the model $\widehat{d}_L$ (Fig. 6). To overcome the fact that the direction of the parallel beams wrt. the camera is not known, we exploit the knowledge that the lasers are equidistant to the camera and approximate the direction with the direction of the vector connecting the camera center and the midpoint between the two points of laser beam intersections with the model X_{L_1} and X_{L_2}. Since it is reasonable to expect the camera-scene distance to be significantly greater than its difference measured at the two points, this approximation leads to a negligible error. Similarly to the FUM, the location of laser beam intersections with the scene X_{L_1} and X_{L_2} are determined through a ray casting procedure and are affected by the same scale as the model and therefore affect the final estimated distance $\widehat{d}_L$ by the same factor:

$$\cos\alpha = \frac{v_{1,2} \cdot v_{\mathrm{CM}}}{|v_{1,2}||v_{\mathrm{CM}}|}, \tag{4}$$

$$\widehat{d}_L = \sin\alpha \cdot |v_{1,2}|, \tag{5}$$

where $v_{1,2}$ represents the vector between scene points X_{L_1} and X_{L_2} and v_{CM} the vector connecting camera center with the middle point X_M.

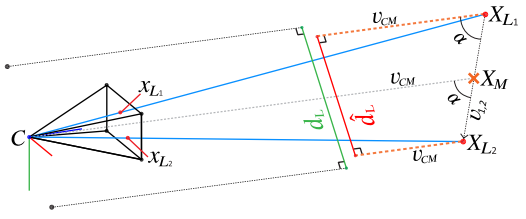

Fig. 6. Scale estimation procedure using the partially constrained method, based on the 3D model and optical image depicting the laser beam projections of the laser intersections with the scene. The direction of the parallel laser beams (wrt. the camera) is approximated with the vector v_{CM} and used to compute the perpendicular distance between the predicted laser beams originating at the point of laser intersenctions with the scene (X_{L_1} and X_{L_2}).

2.4. Camera – laser calibration

Depending on the method used for estimating the scale, several parameters describing the geometric arrangement between the camera and the laser scaler need to be estimated. To obtain a fully calibrated system, both origins and orientations of the laser beams with respect to the optical camera have to be known. Normally, the calibration procedure consists of the acquisition of images with clearly visible laser-surface intersections for which the distances between the camera and surface are known or can be easily computed (e.g., using a checkerboard pattern). A set of points lying on the laser beam is thus obtained by expressing the 3D positions of these intersections in the camera coordinate system. Given that the spread of distances at which the data is collected is sufficient, the direction of the laser beam can be confidently estimated through a line fitting procedure minimizing the sum of squared perpendicular distances between the 3D points and the laser beam line. To avoid any potentially erroneous 3D points affecting the final calibration, a robust estimation method such as RANSAC can be utilized. Finally, the origin of the laser is determined by computing the intersection between the now estimated direction of the laser beam and plane $\mathcal{L}$. It is important to note that, in underwater scenarios, significant refraction can occur at the air-acrylic-water interface of the laser housing. This effect has to be considered in the calibration procedure, either mathematically or by performing the data acquisition underwater.

Alternatively, the partially constrained method requires the knowledge of the distance between the parallel pair of lasers and that the camera center is equidistant to the laser origins (without the need for the knowledge of its value). For most cases, commercially available laser scalers have laser beams that are adequately parallel. On the contrary, ensuring that the camera center is equidistant to the beams is more challenging, unless one is using a purposely designed mounting bracket for the camera/laser system. Having this in mind, it is important to be aware of a potential error in the accuracy of the scale estimation caused by not having the camera equidistant to the laser origins.

The direct consequence of such error is the inaccuracy induced in the estimation of the vector v_{CM} used for approximating the direction of the parallel lasers. Given that the magnitude of the misalignment error will always be disproportionately small compared to the camera-scene distance (mm/cm vs. m), the effect on the estimated direction and subsequently on the final scale estimation will be negligible small as seen in Fig. 7. The analysis of the induced error due to the camera center misalignment up to 5 cm showed that the error increases with

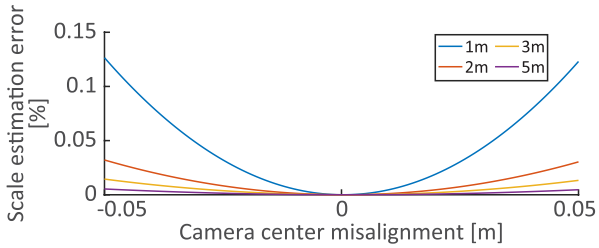

Fig. 7. Error in the estimation of scale using partially constrained method due to the misalignment of camera center with the laser pair.

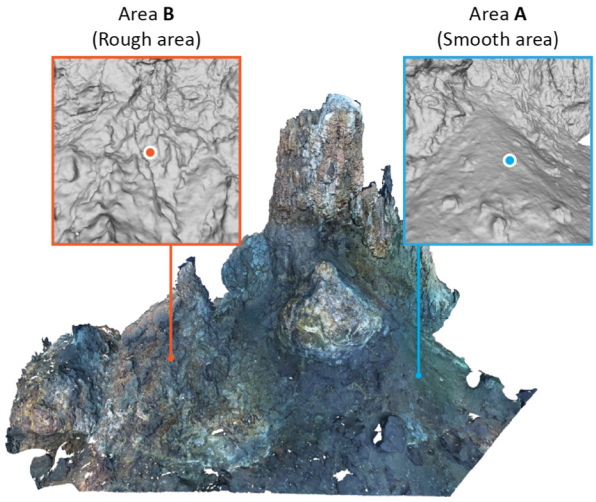

Fig. 8. 3D model of an underwater hydrothermal vent (Eiffel Tower at Lucky Strike vent field, Mid-Atlantic Ridge) used for model reconstruction evaluation at two marked areas. Data acquired during the 2015 MOMARSAT cruise (doi:10.17600/15000200).

decreased camera-scene distance as well as the error is small (less than 0.15% in the worst case). To ensure the error is in fact due to the misalignment, the camera was positioned parallel to a flat surface.

3. Results

To assess the applicability and theoretical accuracy of the two proposed approaches (PCM and FUM), tests were performed on both real and simulated scenario datasets. To validate the performance using different laser configurations and acquisition conditions, we have used a real 3D model built using underwater imagery, as depicted in Fig. 8. Various laser measurements were generated as they would have been captured during an ROV survey. As the absolute scale of the model is not precisely known, for the purpose of this evaluation, it was assumed that the model and its scale are correct. Therefore, the performance can be evaluated by comparing the deviations of the estimated scales with the assumed (imposed) correct value of the scale of the model ($s = 1$). This allowed us to confirm the correctness of our approaches, as well as analyze the effects of various types and levels of noise have on the estimation.

Given our goal of developing methods usable in real world scenarios, three distinct laser configurations were devised (Fig. 9)) to test the performance:

(A) Lasers are parallel and aligned with the optical axis of the camera;
(B) Lasers are parallel and positioned equidistant from the camera center, but not aligned with the optical axis;
(C) Lasers have arbitrary positions and directions.

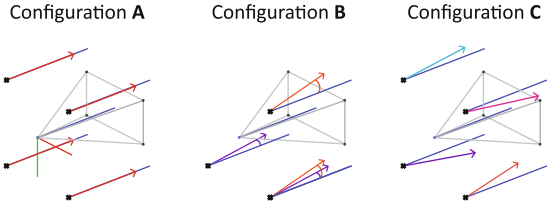

Fig. 9. Various laser configurations used in evaluation: (*A*) Optical axis aligned laser beams; (*B*) Pair-wise parallel laser pairs; (*C*) Lasers with arbitrary origins and orientations. Blue lines represent the optical axis, and the remaining lines depict lasers which are parallel among themselves. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

To illustrate the advantages of our proposed methods in comparison to commonly used image-scaling approaches, the approach by Davis and Tusting (1991) was additionally evaluated, as one of the most versatile methods. The procedure requires four parallel lasers aligned with the optical axis of the camera as well as assuming scene flatness. By exploiting the known spacing between the laser spots on the image and the known displacement of laser origins from the optical center of the camera, distances between various points on the image can be computed for an arbitrary tilt and pan of the camera. As only laser configuration *A* meets the requirements of their method, and other configurations cause dramatic and unpredictable errors, we limit the reporting of the results for Davis and Tusting's approach to laser configuration *A*. Another commonly used method presented by Pilgrim et al. (2000) was not evaluated, as this method requires the restriction of the pose of the camera in either pan or tilt with respect to the scene, which can only be a reasonable restriction if the scene is flat (e.g., the sea bottom), which is almost never the case in models reconstructed using SfM.

3.1. Data

The generation of image and laser data as they would have been recorded in real scenarios enabled us to simulate different perspective angles and camera-scene distances, and analyze their effects on the resulting estimations of scales. The measuring system consisted of an ideal pinhole camera and laser scalers. Both the intrinsic parameters of the camera and the refraction of the lasers occurring at the air-housing-water interface were correctly modelled in the calibration. A real 3D model depicted in Fig. 8 was used in this simulation. The 3D chimney was reconstructed from 908 images of an underwater vent field at the deep-sea Lucky Strike area, collected during the MOMARSAT 2015 cruise (doi:10.17600/15000200). The model covers an area of approximately 200 m^2 with a height range of ~13 m. Assuming the 3D model has a correct scale, we can compute the location of laser spots and feature points as they would appear on the images taken from different poses and according to the pre-determined laser configurations. The number of feature points has been selected to reflect an average number of successfully matched features per image in underwater scenarios ($n = 1500$). To mimic the various perspective angles of the camera, we generate views for which the image plane is not only perpendicular to the surface normal (at the point viewed by the principal point of the camera), but also at a wide range of angles. In total 289 different views were created from different combinations of pitch and roll angles deviating from between $-40°$ and $40°$ in $5°$ steps (Fig. 10). If not specified differently, the camera-scene distance (i.e., distance between the camera center and the point of interest on the surface) has been kept constant at 3 m; based on our experience, this is a reasonable assumption for a typical ROV survey of the scene in this type of environments.

The lasers have been positioned according to the configurations envisioned in different scenarios (Fig. 9). In configuration *A*, the lasers

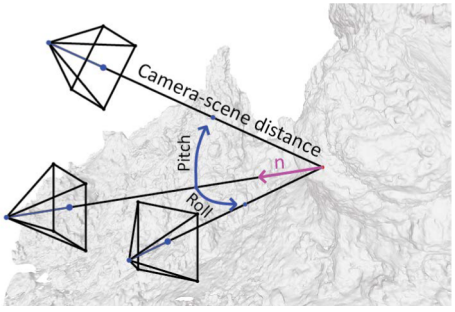

Fig. 10. Definition of perspective angles and camera-scene distance used in the generation of the evaluation data.

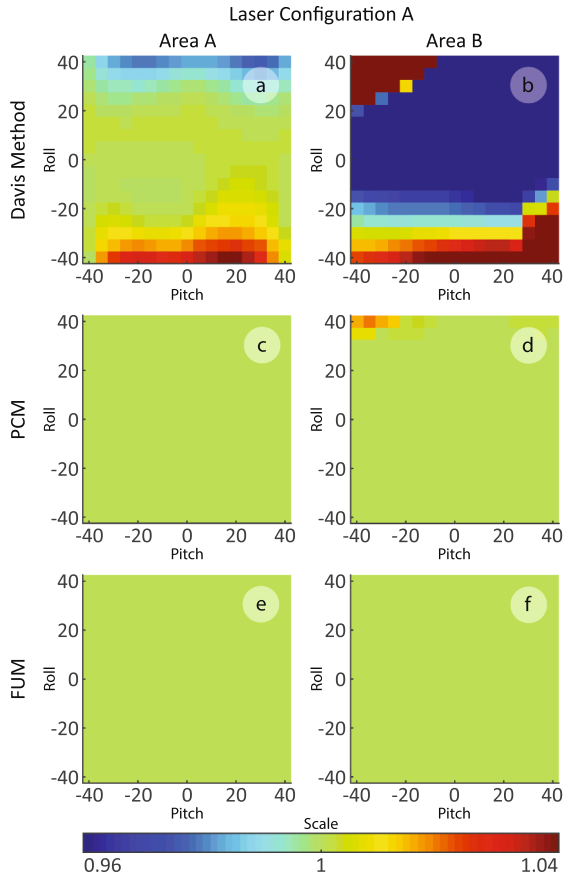

Fig. 11. Estimated scales of the model at a smooth (area A) and rough area (area B) with various perspective angles and constant camera-scene distance ($d = 3$ m) using Davis and Tusting (1991), partially constrained (PCM) and fully unconstrained method (FUM). Lasers were aligned with the optical axis (configuration A).

have been positioned at an equidistance of 10 cm from the camera center. For configuration *B* two pairs of lasers, with a 10 cm perpendicular distance between the beams, have been used, positioned vertically and horizontally. The pairs are perfectly parallel but not aligned with the optical axis of the camera. Each of the pairs has been used independently to test the two most common scenarios, with laser scalers positioned either below or at the side of the camera. As both produced similar results we only present the results for the horizontal pair.

Finally, the configuration *C* reflects a real laser configuration used during the 2017 SUBSAINTES cruise (doi:10. 17600/17001000) (Escartín et al., 2017). The laser set-up in the ROV VICTOR (IFREMER) used for image acquisition during this cruise was slightly misaligned, while the laser origins are placed at an approximately equal distance of 16.5 cm with slight rotation around the z-axis of the camera.

3.2. Terrain roughness

We first compare the results of estimated scales on two different types of terrain (smooth – *Area A* and rough – *Area B*) acquired from a variety of perspective angles and laser configurations. Fig. 11 presents the results obtained using laser configuration *A* and with our two proposed methods (FUM and PCM) as well as with the Davis approach.

Comparing the errors among the methods, we notice that the Davis and Tusting method is capable of estimating the correct scale only if the flatness assumption is only slightly violated, i.e. the area is nearly flat and the perspective angle is not too large (Fig. 11a). As that is not the case on rough terrain (Fig. 11b), the estimated scale varies significantly with different perspective angles, confirming the strong dependency of this method on scene geometry. On the other hand, our two methods correctly compensate for any changes in the viewing angle and terrain roughness. The laser direction approximation assumed in PCM does, however, cause a slight error – up to 1.5% in extreme cases (e.g., rough terrain and large perspective angle – Fig. 11d), a situation in which the camera-scene distance discrepancy between the two laser points is strongly boosted. The fact that the scale is correctly estimated in all cases, clearly shows the ability of the FUM to correctly compensate for the effects of terrain roughness and perspective angle (Figs. 11e and 11f). Additionally, it is important to emphasize again that image-scaling methods require an additional association between the image points and the model in order to be able to estimate the scale. In our tests, we assumed perfect association, which is nearly impossible to achieve as it is a manual error-prone process. The actual results in real cases are therefore expected to be even worse.

In scenarios in which the lasers are not perfectly aligned with the camera (i.e., laser configurations *B* and *C*), the image-scaling methods become unusable as the errors increase dramatically and unpredictably. For this reason, we only present the results of our proposed methods (FUM and PCM) for the remaining two configurations. Similarly, we limit the presented results to the rough terrain, as the methods will perform better (or equally) on flat areas.

As seen in Fig. 12a and c, both of our methods obtain good results with a laser configuration *B*, in which the lasers are mounted parallel to each other. As in the previous cases, the partial method exhibits slight errors due to the assumed laser direction approximation. Analysis of data collected using laser configuration *C*, shows that the partial method fails, with results strongly affected by the irregularities in the parallelism. Instead, the full method (Fig. 12d) correctly compensates these irregularities and yields correct results.

3.3. Laser direction approximation

To illustrate the influence that a difference of camera-scene distances (measured at the two points hit by the laser beams) has on the results of the partial method at various distances, we estimated the scale using 10, 000 randomly-selected points across the model (Fig. 13a). For each point, the camera has been positioned at a distance

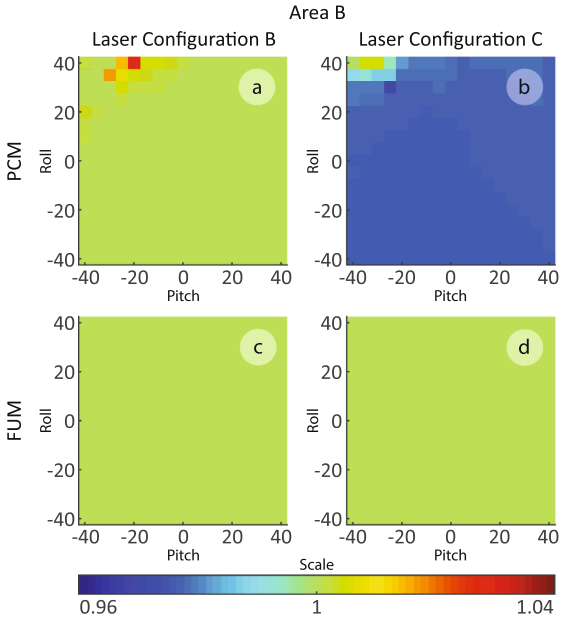

Fig. 12. Estimated scales of the model at a rough area (area B) with various perspective angles and constant camera-scene distance ($d = 3$ m) using partially constrained (PCM) and fully unconstrained method (FUM). Lasers were in configuration B and C.

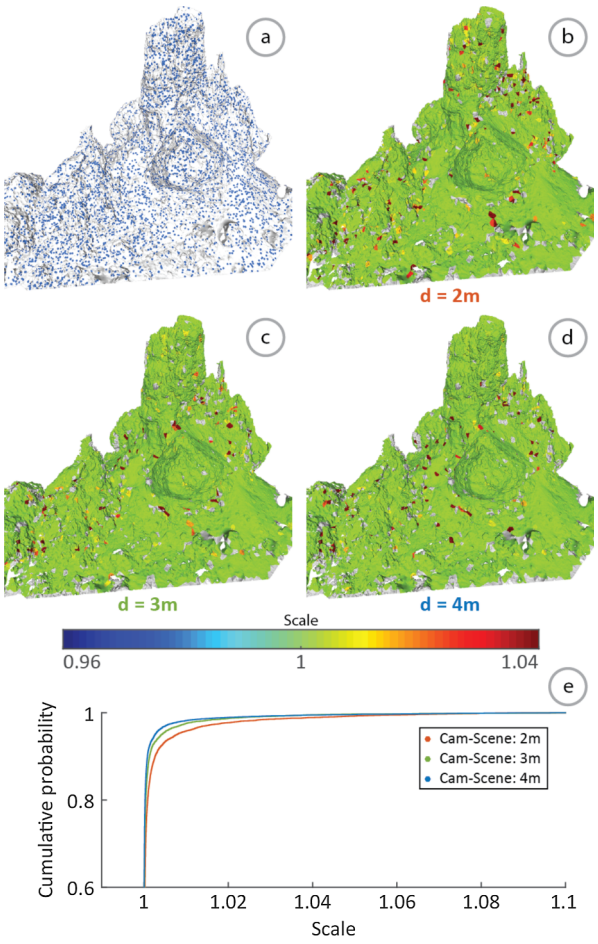

Fig. 13. (a) 10,000 random points used for estimating the scale across the model; (b–d) Estimated model scales at various camera-scene distances with laser configuration B using a partially constrained method; (e) Cummulative probability distribution of estimated scales.

d in the direction of the normal of the surface. Results obtained at three distances (2 m, 3 m and 4 m), illustrated in Fig. 13b–d, show that the error decreases with increasing distance of the camera (i.e., larger *d*). This is especially visible in rougher areas, such as the top of the hydrothermal vent and the areas near previously mentioned area *B*. In those areas, larger discrepancies between the camera-scene distances measured at the detected laser spots are expected, as the probability of laser beams hitting different parts of the model is much higher. Therefore the result indicates that the increased camera-scene distance decreases the effect of difference of distances on the accuracy of the results. We also document the cumulative distribution functions of these estimated scales obtained at various camera-scene distances (Fig. 13e), from which it is noticeable that a higher percentage of points with scales closer to the anticipated value of 1.0 is obtained the further the camera is from the scene.

The relation between the average camera-scene distance and its difference measured at the two points of laser beam-scene intersections can be clearly observed in Fig. 14, which shows the estimated scale vs. the difference of distances, with color coded average camera-scene distance. As expected, the error in the estimation grows with the increase in the distance discrepancies between the points. Furthermore, we can see that the increase follows a parabola-shaped functions determined by the camera-scene distance. Short distances define a narrow parabola, and cause an increase in the error that is larger than that for longer distances. This indicates that a distance discrepancy between the two points of laser – scene intersections causes a greater error in the result when the camera is near the scene. The shape and steepness of

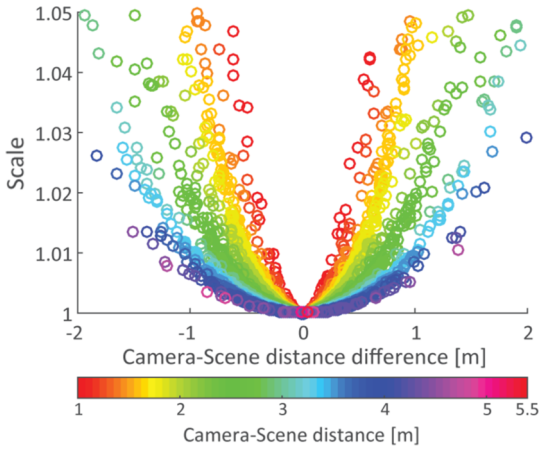

Fig. 14. Estimated model scale (using partially constrained method at 10,000 random points) with respect to the difference of the camera-scene distances measured at the two points of laser beam intersection with the model color coded by the average camera-scene distance per image. Narrower parabolas at shorter distances indicate that the accuracy of the partially constrained method is more affected by the difference of the camera-scene distances at the two laser points than at longer distances.

the parabolas is dependent on the displacement of the lasers from the camera origin, as well as their orientation with respect to the optical axis of the camera.

3.4. Noise

As collected data is never noise-free, we performed an additional analysis to evaluate the effects that the expected noise in feature and laser spot detection have on the scale estimation process. The various values for noise assumed in laser spot detection intend to represent the dispersion and absorption of the laser beam in the water medium. The greater the absorption and/or dispersion, the less certain the detection of the laser positions will be. The experiment was performed on *area B* of the model, with camera angles ranging from − 15° to 15° in pitch and roll; the range of view geometries which give consistent results in the ideal scenario (Fig. 11). The observation distributions were modelled by assuming multivariate Gaussian distributions with dimension-independent noise for both feature and laser spot detections. For 2D features, the values were set matching those normally obtained in underwater scenarios (Garcia and Gracias, 2011) $\sigma_f = \{0.5\text{px}, 1.0\text{px}\}$, while laser detection noise was defined by assuming 95% accuracy of peak detection within one or two pixels $\sigma_l = \{0.25\text{px}, 0.5\text{px}\}$. As feature matches themselves are normally corrupted with a certain level of outliers, we have also performed experiments with various inlier/outlier ratios (Campos et al., 2015) $r = \{0\%, 10\%, 20\%\}$. Each of the tests were repeated 500 times.

The resulting distributions of estimated scales with parallel and free laser configurations (i.e., configurations *B* and *C*) are presented in Table 1 with a subset of the results shown in Fig. 15. Given that the FUM requires only a single laser to obtain a scale estimate, results from separate lasers were fused by computing their average. The effect of such averaging can be identified in Table 1, where the results for a single laser (FUM – single) are shown side by side with the final averaged result (FUM – all).

As expected, the uncertainty of estimated scales increases with the increasing noisiness of the laser detections, as each estimation is directly influenced by displacements in laser spot positions. Comparison of these results show that with noisy data the PCM method performs better than the FUM with a single laser point, but worse when multiple laser points are used instead. This occurs due to the averaging of independent scale estimates. As each laser produces a result that is

Table 1
The results obtained with various methods (PCM, FUM) with different levels of noise induced into the location of detected features and laser spots.

Cam-Scene	Configuration B		Configuration C	
distance [m]	PCM	FUM – all	FUM – single	FUM – all
$\sigma_f = 0.5$, $\sigma_l = 0.25$				
2	1.0 ± 0.0014	1.0 ± 0.0010	1.0 ± 0.0019	1.0 ± 0.0010
3	1.0 ± 0.0022	1.0 ± 0.0015	1.0 ± 0.0028	1.0 ± 0.0014
4	1.0 ± 0.0030	1.0 ± 0.0021	1.0 ± 0.0034	1.0 ± 0.0017
$\sigma_f = 1.0$, $\sigma_l = 0.25$				
2	1.0 ± 0.0014	1.0 ± 0.0010	1.0 ± 0.0019	1.0 ± 0.0010
3	1.0 ± 0.0022	1.0 ± 0.0015	1.0 ± 0.0028	1.0 ± 0.0014
4	1.0 ± 0.0030	1.0 ± 0.0021	1.0 ± 0.0034	1.0 ± 0.0017
$\sigma_f = 0.5$, $\sigma_l = 0.5$				
2	1.0 ± 0.0028	1.0 ± 0.0020	1.0 ± 0.0038	1.0 ± 0.0020
3	1.0 ± 0.0044	1.0 ± 0.0031	1.0 ± 0.0056	1.0 ± 0.0028
4	1.0 ± 0.0059	1.0 ± 0.0042	1.0 ± 0.0069	1.0 ± 0.0034

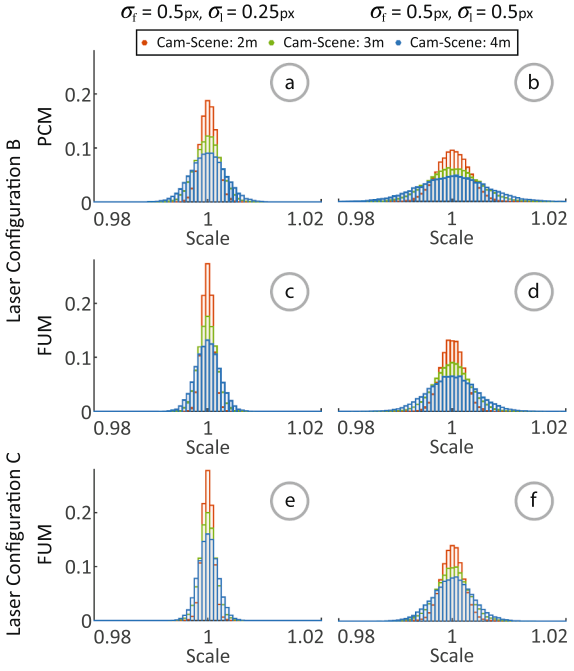

Fig. 15. Distributions of estimated model scales with partially constrained and fully unconstrained methods at various noise levels induced into the location of detected features and laser spots. The results obtained at different camera-scene distances are depicted with (2 m – red; 3 m – green; 4 m – blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

independently affected by noise, the subsequent averaging reduces its effect.

To some extent this can also be observed in the partial method with the simultaneous use of two laser points, which explains the improved results over the full method with the single laser. It is also clear that the uncertainty of the scaling estimate also increases with the camera-scene distance, which is expected as errors on the image are magnified when projected further from the camera.

In contrast, the noise affecting the feature points used in the pose estimation, does not significantly influence the final scaling results. This is due to the use of BA in the pose optimization, which is a maximum likelihood estimator when the image error is zero-mean and normally distributed, as is the case in our tests. Similarly, the effects of outliers are mitigated by the use of a robust estimation method AC-RANSAC

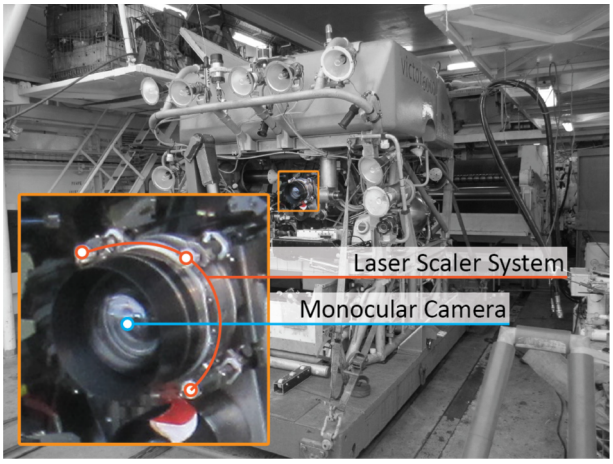

Fig. 16. ROV VICTOR 6000 (IFREMER) with enlarged camera and laser scaler system.

(Moisan et al., 2012). As the outliers do not follow a specific pattern, the iterative procedure successfully identifies and removes spurious matches, and hence the final estimate is unaffected. It is important to note that while the results obtained might indicate a method which is extremely robust to any discrepancy in the feature points, the approach is still vulnerable to (a) outliers that obey the estimated geometric model; to (b) the possibility of having a set of feature points which can be explained with multiple camera poses, or to both (a) and (b). However, this vulnerability can be reduced to a level that is not of practical concern, by ensuring that the set of features is well spread throughout the image.

3.5. Real scenario

The fully unconstrained method was used on a real dataset collected during the SUBSAINTES cruise (doi: 10.17600/ 17001000). Throughout the cruise, extensive seafloor imagery was collected using the ROV VICTOR 6000 (IFREMER) (Michel et al., 2003) with a mounted monocular camera (Sony FCB-H11 with corrective optics and dome port), and a laser scaler with four laser beams positioned around the camera (Fig. 16). The intrinsic parameters of the camera were determined using a standard calibration procedure (Bouguet, 2008) assuming a pinhole model together with the 3rd degree radial distortion model. Once calibrated, the camera parameters were kept constant through the entire acquisition process.

One of the main goals of this cruise is to identify, map, and measure indicators of displacement at the seafloor associated with a recent submarine earthquake (Escartín et al., 2016) that occurred in the French Antilles, offshore from Les Saintes Islands in 2004 (Feuillet et al., 2004). These traces are visible in outcrops of an active submarine fault scarp at depths of up to ~1000 m below sea level, that has been systematically mapped and surveyed. Imagery was used to obtain ~30 3D models, that will ultimately be used to conduct measurements of displacement associated with the 2004 earthquake. Therefore proper scaling is required to enable accurate geological measurements.

The 3D models have been reconstructed using an adapted 3D reconstruction procedure consisting of multiple open-source solutions (OpenMVG (Moulon et al., 2018; Moulon et al., 2013), OpenMVS (Shen, 2013; Jancosek and Pajdla, 2014), MVS-Texturing (Waechter et al., 2014)) as described in (Hernández et al., 2016). Fig. 17 depicts one such model, named FPA, which has been reconstructed from a total of 218 images with a resolution of 1920 × 1080. This particular outcrop was already imaged during a prior cruise (ODEMAR, doi:10.17600/ 13030070) (Escartín et al., 2016).

As the FPA model was reconstructed using only optical images

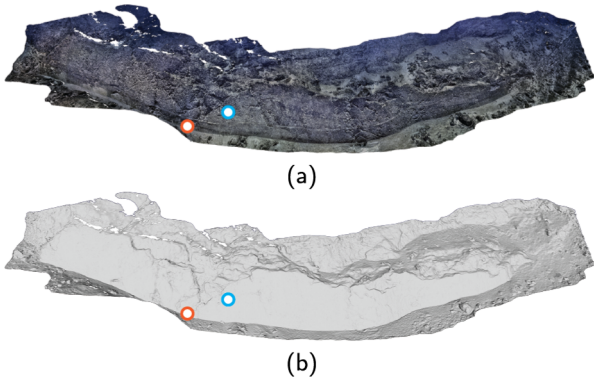

Fig. 17. (a) Textured and (b) triangle mesh representations of FPA 3D model, with marked areas of evaluation.

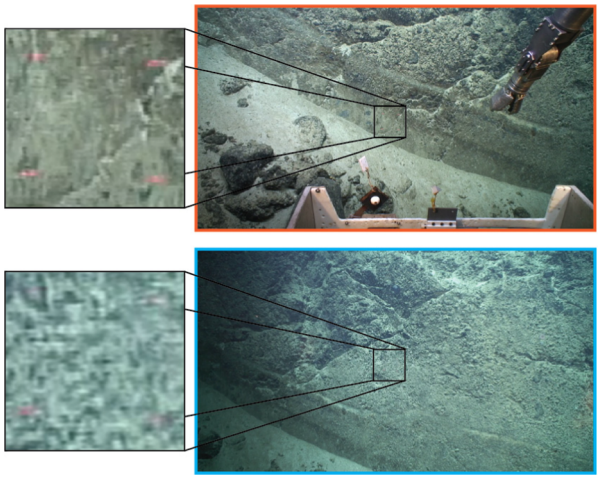

Fig. 18. Example of images from the two areas of evaluation with visible laser projections on the scene.

acquired by a monocular camera, the proper scale of the model can be obtained using images containing laser beams projected on the surface of the scene by applying one of our proposed methods. During the SUBSAINTES cruise, such images were collected in addition to the ones already used in the reconstruction process. Six images with clearly noticeable laser spots (Fig. 18) have been selected from the center of the 3D model, at two different locations as indicated in Fig. 18. The images were collected at camera-scene distances of approximately 3 m and 4 m respectively while keeping the camera intrinsic parameters constant and equal to the ones used in the acquisition process. Subsequently, the laser spot's locations have been marked manually (with the guidance of simple color thresholding) with an expected error that was on average between 1 px and 2 px. Due to multiple changes in the vehicle payload throughout the cruise, the lasers became misaligned and therefore a fully unconstrained method was used to obtain the scale of the model.

Given that the setup consisted of four lasers, the FUM method computed four independent estimates of the model's scale per image. As we have shown in the previous experiments, averaging these independent results further reduces the effects of errors in the detection processes, leading to a better constrained final solution. The scaling results for each of the 6 selected images are presented in Table 2 and Fig. 19. In this figure, the scale estimates obtained for each laser beam are depicted as circles, while the final estimate per image is marked with a black cross (x). The average of all the values obtained is additionally shown by a red dashed line.

The average value of the scale of the FPA model estimated per image was 0.237 ± 0.0008 which represents 0.3% of the scale value. The obtained result means that each unit in the current model is equal to

Table 2
Estimated FPA model's scale using fully unconstrained method and a simplistic direct 3D approach. Reported numbers represent the ratio between the model's unit and a meter – each measurement has to be multiplied with the inverse of the ratio to obtain a metric result.

	Cam-Scene	FUM (per laser)				FUM	Direct 3D
	distance [m]	L_1	L_2	L_3	L_4	(all)	(all)
1	3.05	0.234	0.239	0.237	0.236	0.237 ± 0.002	0.235 ± 0.009
2	3.06	0.236	0.239	0.236	0.238	0.237 ± 0.002	0.236 ± 0.008
3	3.05	0.237	0.237	0.235	0.236	0.236 ± 0.001	0.235 ± 0.008
4	3.90	0.239	0.241	0.236	0.236	0.238 ± 0.003	0.236 ± 0.013
5	3.91	0.238	0.239	0.237	0.234	0.237 ± 0.002	0.236 ± 0.013
6	3.60	0.238	0.236	0.236	0.233	0.236 ± 0.002	0.234 ± 0.010

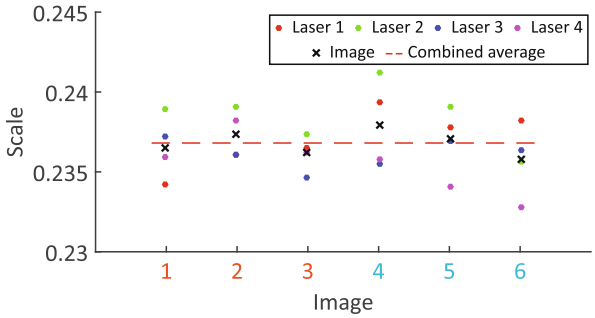

Fig. 19. Estimated scale factors for FPA model, per laser and per image, using our fully unconstrained method. Colour of image numbers (x axis) corresponds to locations shown in Fig. 18. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

0.237 m or alternatively, the model has to be scaled by a factor 4.22 to obtain a metric result. This implies that on a measurement of 25 cm on a model with lateral and vertical dimensions of 33 m by 10 m is expected to have an uncertainty, due to scaling error introduced by the proposed method, of approximately 1 mm. It is important to note that this estimate is based on the assumption of adequate calibration of the camera and of the camera-laser system. Given that these methods estimate the scale from image information, it is impossible to decouple the error reported by the method with the actual error of the model without the validation with the external measurements. A sign of such problems would be disproportionately large variations of scale estimates from individual lasers within the same image, as each result would be affected differently by the erroneous calibration. Comparing the deviations of scale estimates for image sets 1-3 and 4-6, the correlation between increasing camera-scene distance and increased uncertainty is apparent and consistent with previous result from generated data.

The analyses of scaling deviations computed for each laser with respect to the final estimated scale per image (Fig. 20) shows that independent evaluations deviate by about 0.6% with a maximum

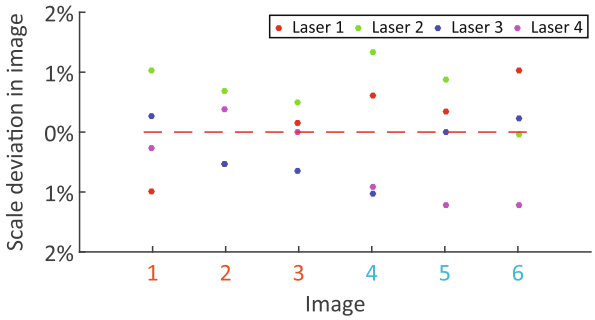

Fig. 20. Deviation of estimated FPA model's scales, using our fully unconstrained method FUM, and for each laser in each image.

deviation of 1.3% for laser 2 in image 4. These results are again in agreement with the results previously computed with the validation data on the hydrothermal vent in Fig. 8.

To further show the robustness and usefulness of our approach, we compare our results to the ones that would have been obtained if our method was not available. As the non-alignment of lasers with the optical axis would have prevented the use of both image-scaling methods (Pilgrim et al. (2000) and Davis and Tusting (1991)), the only option available would have been a manual and somewhat simplistic approach still widely used in laser photogrammetry (Kocak et al., 2004; E. Rowe and Dawson, 2008; Robert et al., 2017; Pilgrim et al., 2000). This involves manual identification of laser beam intersection points with the scene on the 3D model, and assuming pair-wise Euclidean distances to be the actual distances between the laser pairs. In order to compare our results with the best possible outcome of this simplistic approach, we determined the points on the model using a ray-casting technique, effectively completely eliminating the extremely error-prone human step. The results averaged over 4 laser pairs are presented in the last column of Table 2 (Direct 3D). We can see that the results of different laser pairs are much more incoherent (4.3% deviation compared to 0.6% in the case of fully unconstrained method). We also note that the results of such a simplistic method are extremely dependent on the perspective angle of the camera, and the degree of misalignment of the lasers, as well as errors induced by manual point selection. As shown with the validation tests, our fully unconstrained method remains unaffected.

4. Conclusions

This paper introduced two novel methods for automatic scaling of SfM-based 3D reconstructions using laser scalers, methods that are applicable for routine underwater surveys with ROVs or AUVs. Both methods were validated using a series of generated datasets based on an underwater 3D model derived from submarine field imagery, and showed its applicability in real scenarios using a dataset collected during a recent cruise (SUBSAINTES 2017).

The two approaches presented here, namely the fully and partially constrained methods, overcome a multitude of restrictions imposed by prior laser photogrammetry methods (e.g., laser alignment with the optical axis of the camera, perpendicularity of laser beams with the scene). These methods, within the step of pose estimation, also remove the need for manual identification of identical points on the image and 3D model, an extremely time-consuming and error-prone processing step.

Each of the two methods is designed to address the different types of laser setup, encompassing the variety of most commonly used setups in real underwater scenarios. The fully unconstrained method is applicable to arbitrary laser setups, with known geometric relations between the camera and the lasers. The ability to compensate for any misalignments enables accurate scaling in a wider variety of circumstances, such as the manipulation of equipment between surveys during a cruise and precluding strict parallelism. We thus propose a partially constrained method, which significantly reduces the camera-laser rigidity constraints, that may be otherwise too restrictive in real scenarios. This approach requires parallel lasers but alleviates the need for a time-consuming calibration process. The partially constrained method can thus be used to accurately and automatically scale 3D models built with data acquired using ROVs, including smaller shallow-water ones. Pre-calibrated underwater laser scalers are readily available nowadays, and need only to be placed near the optical camera.

To robustly validate the performance of the methods, a real 3D model of an underwater hydrodynamic vent was used to generate laser and image information as it would have been obtained from various laser configurations, camera viewing angles and camera-scene distances. We tested our methods with three laser configurations (i.e., aligned with the optical axis of the camera; parallel but misaligned with the optical axis; and freely oriented) which can account for nearly all possible laser setups in real seafloor surveying situations using ROVs and AUVs. The initial evaluation was performed on two different types of terrain (smooth and rough), and demonstrated the advantages provided by the two proposed approaches relative to previously used image-scaling methods. Our methods can be used in the field, with misaligned or freely oriented lasers, and with extreme camera angles during image acquisitions, reaching up to 40° in both pitch and roll.

While the fully unconstrained method yielded robust results under all tested circumstances, the partially constrained method was affected by a slight error (2.9% in the most extreme case) due to the approximation used for determining the laser direction. We further analyzed the effect of the approximation by evaluating 10, 000 randomly selected points. We demonstrate that scaling errors depend on the difference of camera-scene distances between the two points of laser beam – scene intersection, and that this effect decreases with an increasing camera-scene distance. The consequences of inevitable noise in feature and laser-spot detection uncertainty were also examined, together with the effects of potential errors in feature matching (outliers). Due to the specificity of the algorithms used, the noise and potential outliers in the feature detection and matching process did not have a significant effect on the results, while the noise induced in the position of laser spots did directly influence the estimations. As expected, increases in camera-scene distance results in higher errors in the estimation, as the displacements are magnified with distance. Additionally we compared the results obtained from a single laser measurement with the average obtained from all and demonstrated that such fusion further reduces the effects of noise.

It is important to acknowledge that the achieved accuracy in the simulated scenario should be regarded as theoretical accuracy, given that the camera and camera-laser calibrations used in generating the data were ideal. Errors in the calibration will affect the final result, in such a way that, without external validation (using known measurements) it becomes impossible to decouple the error reported by the method and the actual error of the model. We note that a sign of such problems would be large deviations of scale estimated obtained by independent lasers/laser pairs within the same image.

Finally we report on the application of the fully unconstrained method to determine the scale of a model built using images from a geologic outcrop, recorded during the SUBSAINTES cruise. Six images with clearly visible laser spots have been selected from two different model locations, and used to independently determine the scale of the model. The average scale estimated using our fully unconstrained method was 0.237 with the standard deviation of 0.3% between the results from various images. The average deviation of estimated scales by independent lasers was 0.6% with the maximum deviation of 1.3%. We also documented data showing that images acquired at a longer camera-scene distance exhibited bigger deviations of estimated scales, as predicted from the validation test results.

The results of our two methods were also compared to those that would have been obtained without the availability of our method. Due to laser non-alignment with the optical axis of the camera, the only approach possible would be a somewhat simplistic method which involves manual identification of laser intersection points with the 3D model, and assumes that the pair-wise Euclidean distances are the actual distances between the laser pairs. To predict the best possible outcome, we automatically determined these correspondences, alleviating any additionally induced errors. The results from the simplistic scale method show a much more significant deviation than that of our method (4.3% vs. 0.6%, respectively). Based on our results we also stress that the results of such simplistic methods are extremely dependent on the perspective angle of the camera and the degree of misalignment of the lasers, which is not the case for our fully unconstrained method. Finally, these methods can be used universally as they are based on standard sensors available for ROVs and AUVs (cameras and laser scalers), do not require any dedicated hardware, and can be applied to

legacy data.

Although the presented methods are designed to be independent of the laser spot detection approach used, we showed that its performance directly influences the scale estimation accuracy. In the reported results, we identified the location of the spots manually albeit with the help of simple color thresholding. While relatively accurate, this manual process is time consuming. An effort is currently ongoing aimed at automating the detection of the laser spots, which will facilitate the ability to perform scale estimation on a larger number of images.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This study is based on results from the MOMARSAT 2015 and SUBSAINTES 2017 cruises, that deployed the ROV VICTOR 6000 (IFREMER, France) for image acquisition used here. These cruises (ship and ROV time) were funded by the French Ministry of Research. We commend the work of the crew, officers, and engineers that participated on these cruises and made possible this data acquisition. Partial funding was provided by the European Union's Horizon 2020 project ROBUST (grant agreement 690416-H2020-CS5-2015-onestage) (to K. Istenič), project Eurofleets Plus (grant agreement 824077), the Spanish Ministry of Education, Culture and Sport under project UDRONE CTM2017-83075-R (to R. Garcia and N. Gracias), the ANR SERSURF Project (ANR-17-CE31-0020, France) (to J. Escartín and A. Arnaubec), and the Institut de Physique du Globe de Paris (to J. Escartín).

References

Agarwal, S., Snavely, N., Simon, I., Seitz, S.M., Szeliski, R., 2009. Building rome in a day. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 72–79. https://doi.org/10.1109/ICCV.2009.5459148.

Anderson, K., Gaston, K.J., 2013. Lightweight unmanned aerial vehicles will revolutionize spatial ecology. Front. Ecol. Environ. 11, 138–146. https://doi.org/10.1890/120150.

Bergmann, M., Langwald, N., Ontrup, J., Soltwedel, T., Schewe, I., Klages, M., Nattkemper, T.W., 2011. Megafaunal assemblages from two shelf stations west of svalbard. Mar. Biol. Res. 7, 525–539. https://doi.org/10.1080/17451000.2010.535834.

Bingham, B., Foley, B., Singh, H., Camilli, R., Delaporta, K., Eustice, R., Mallios, A., Mindell, D., Roman, C., Sakellariou, D., 2010. Robotic tools for deep water archaeology: Surveying an ancient shipwreck with an autonomous underwater vehicle. J. Field Robot. 27, 702–717. https://doi.org/10.1002/rob.20350.

Bodenmann, A., Thornton, B., Ura, T., 2017. Generation of high-resolution three-dimensional reconstructions of the seafloor in color using a single camera and structured light. J. Field Robot. 34, 833–851. https://doi.org/10.1002/rob.21682.

Bouguet, J.Y., 2008. Camera calibration toolbox for matlab (2008). http://www.vision.caltech.edu/bouguetj/calib_doc1080. Accessed: January 22 2018.

Caimi, F.M., Tusting, R.F., 1987. Application of lasers to ocean research and image recording systems. In: Proceedings of the International Conference on LASERS, STS Press McLean, Virginia. pp. 518–524.

Campos, R., Garcia, R., Alliez, P., Yvinec, M., 2015. A surface reconstruction method for in-detail underwater 3d optical mapping. Int. J. Robot. Res. 34, 64–89.

Campos, R., Gracias, N., Ridao, P., 2016. Underwater multi-vehicle trajectory alignment and mapping using acoustic and optical constraints. Sensors 16, 387. https://doi.org/10.3390/s16030387.

Cocito, S., Sgorbini, S., Peirano, A., Valle, M., 2003. 3-d reconstruction of biological objects using underwater video technique and image processing. J. Exp. Mar. Biol. Ecol. 297, 57–70. https://doi.org/10.1016/S0022-0981(03)00369-1.

Comex.SA, Orus 3d subsea photogrammetry. https://comex.fr/en/orus3d. (accessed: September 6 2019).

Davis, D., Tusting, R., 1991. Quantitative Benthic Photography using Laser Calibrations. Undersea World, San Diego, California.

E. Rowe, L., Dawson, S., 2008. Laser photogrammetry to determine dorsal fin size in a population of bottlenose dolphins from doubtful sound, New Zealand. Aust. J. Zool. 56, 239–248. https://doi.org/10.1071/ZO08051.

Eltner, A., Schneider, D., 2015. Analysis of different methods for 3d reconstruction of natural surfaces from parallel-axes uav images. Photogram. Rec. 30, 279–299. https://doi.org/10.1111/phor.12115.

Escartín, J., Le Friant, A., Feuillet, N., 2017. Subsaintes cruise report, n/o l'atalante - rov victor - auv asterx. https://campagnes.flotteoceanographique.fr/campagnes/17001000/, doi:https://doi.org/10.17600/17001000.

Escartín, J., Leclerc, F., Olive, J.A., Mevel, C., Cannat, M., Petersen, S., Augustin, N., Feuillet, N., Deplus, C., Bezos, A., et al., 2016. First direct observation of coseismic slip and seafloor rupture along a submarine normal fault and implications for fault slip history. Earth Planet. Sci. Lett. 450, 96–107. https://doi.org/10.1016/j.epsl.2016.06.024.

Feuillet, N., Beauducel, F., Jacques, E., Tapponnier, P., Delouis, B., Bazin, S., Vallée, M., King, G., 2004. The mw = 6.3, november 21, 2004, les saintes earthquake (guadeloupe): Tectonic setting, slip model and static stress changes. J. Geophys. Res.: Solid Earth 116. https://doi.org/10.1029/2011JB008310.

Garcia, R., Campos, R., Escartín, J., 2011. High-resolution 3d reconstruction of the seafloor for environmental monitoring and modelling. In: 2011 IEEE/RSJ International Conference on Proc. Intelligent Robots and Systems (IROS).

Garcia, R., Gracias, N., 2011. Detection of interest points in turbid underwater images. In: OCEANS 2011 IEEE-Spain, IEEE. pp. 1–9.

Hartley, R., Zisserman, A., 2003. Multiple View Geometry in Computer Vision, second ed. Cambridge University Press, New York, NY, USA. https://doi.org/10.1017/CBO9780511811685.001.

Hernández, J.D., Istenič, K., Gracias, N., Palomeras, N., Campos, R., Vidal, E., Garcia, R., Carreras, M., 2016. Autonomous underwater navigation and optical mapping in unknown natural environments. Sensors 16, 1174. https://doi.org/10.3390/s16081174.

Jancosek, M., Pajdla, T., 2014. Exploiting visibility information in surface reconstruction to preserve weakly supported surfaces. Int. Scholarly Res. Notices 2014. https://doi.org/10.1155/2014/798595.

Javernick, L., Brasington, J., Caruso, B., 2014. Modeling the topography of shallow braided rivers using structure-from-motion photogrammetry. Geomorphology 213, 166–182. https://doi.org/10.1016/j.geomorph.2014.01.006.

Kalacska, M., Lucanus, O., Sousa, L., Vieira, T., Arroyo-Mora, J., 2018. Freshwater fish habitat complexity mapping using above and underwater structure-from-motion photogrammetry. Remote Sensing 10, 1912. https://doi.org/10.3390/rs10121912.

Ke, T., Roumeliotis, S.I., 2017. An efficient algebraic solution to the perspective-three-point problem. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7225–7233.

Kocak, D.M., Caimi, F.M., Jagielo, T.H., Kloske, J., 2002. Laser projection photogrammetry and video system for quantification and mensuration. In: OCEANS '02 MTS/IEEE, pp. 1569–1574 vol 3. doi:https://doi.org/10.1109/OCEANS.2002.1191869.

Kocak, D.M., Jagielo, T.H., Wallace, F., Kloske, J., 2004. Remote sensing using laser projection photogrammetry for underwater surveys. In: IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium, vol 2., pp. 1451–1454 doi:https://doi.org/10.1109/IGARSS.2004.1368693.

Lourakis, M., Zabulis, X., 2013. Accurate scale factor estimation in 3d reconstruction. In: International Conference on Computer Analysis of Images and Patterns, Springer. pp. 498–506.

Mertes, J., Zant, C., Gulley, J., Thomsen, T., 2017. Rapid, quantitative assessment of submerged cultural resource degradation using repeat video surveys and structure from motion. J. Maritime Archaeol. 12, 91–107. https://doi.org/10.1007/s11457-017-9172-0.

Mian, O., Lutes, J., Lipa, G., Hutton, J., Gavelle, E., Borghini, S., 2016. Accuracy assessment of direct georeferencing for photogrammetric applications on small unmanned aerial platforms. Int. Arch. Photogramm. Remote Sensing Spatial Inf. Sci. 40, 77. https://doi.org/10.5194/isprs-archives-XL-3-W4-77-2016.

Michel, J.L., Klages, M., Barriga, F.J., Fouquet, Y., Sibuet, M., Sarradin, P.M., Siméoni, P., Drogou, J.F., et al., 2003. Victor 6000: design, utilization and first improvements. In: The Thirteenth International Offshore and Polar Engineering Conference, International Society of Offshore and Polar Engineers.

Moisan, L., Moulon, P., Monasse, P., 2012. Automatic homographic registration of a pair of images, with a contrario elimination of outliers. Image Process. On Line 2, 56–73. https://doi.org/10.5201/ipol.2012.mmm-oh.

Moulon, P., Monasse, P., Marlet, R., 2013. Global fusion of relative motions for robust, accurate and scalable structure from motion. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3248–3255. https://doi.org/10.1109/ICCV.2013.403.

Moulon, P., Monasse, P., Marlet, R., Others, Openmvg. an open multiple view geometry library. https://github.com/openMVG/openMVG. Accessed: January 11 2018.

Neyer, F., Nocerino, E., Gruen, A., 2018. Monitoring coral growth-the dichotomy between underwater photogrammetry and geodetic control network. Int. Arch. Photogramm. Remote Sensing Spatial Inf. Sci. 42, 2. https://doi.org/10.5194/isprs-archives-XLII-2-759-2018.

Pilgrim, D.A., Parry, D.M., Jones, M.B., Kendall, M.A., 2000. Rov image scaling with laser spot patterns. Underwater Technol. 24, 93–103. https://doi.org/10.3723/175605400783259684.

Pizarro, O., Eustice, R.M., Singh, H., 2009. Large area 3-d reconstructions from underwater optical surveys. IEEE J. Oceanic Eng. 34, 150–169. https://doi.org/10.1109/JOE.2009.2016071.

Pizarro, O., Friedman, A., Bryson, M., Williams, S.B., Madin, J., 2017. A simple, fast, and repeatable survey method for underwater visual 3d benthic mapping and monitoring. Ecol. Evol. 7, 1770–1782. https://doi.org/10.1002/ece3.2701.

Remondino, F., El-Hakim, S.F., Gruen, A., Zhang, L., 2008. Turning images into 3-d models. IEEE Signal Process. Mag. 25, 55–65. https://doi.org/10.1109/MSP.2008.923093.

Ribas, D., Palomeras, N., Ridao, P., Carreras, M., Mallios, A., 2011. Girona 500 auv: From survey to intervention. IEEE/ASME Trans. Mechatronics 17, 46–53.

Robert, K., Huvenne, V.A., Georgiopoulou, A., Jones, D.O., Marsh, L., Carter, G.D., Chaumillon, L., 2017. New approaches to high-resolution mapping of marine vertical

structures. Sci. Rep. 7, 9005.
Rossi, P., Castagnetti, C., Capra, A., Brooks, A., Mancini, F., 2019. Detecting change in coral reef 3d structure using underwater photogrammetry: critical issues and performance metrics. Appl. Geomatics 1–15. https://doi.org/10.1007/s12518-019-00263-w.
Rovco, 2019. High definition 3d visualisation. https://www.rovco.com/technology/3d-photogrammetry. Accessed: September 6 2019.
Rzhanov, Y., Mamaenko, A., Yoklavich, M., 2005. Uvsd: software for detection of color underwater features. In: Proceedings of OCEANS 2005 MTS/IEEE, vol. 3, pp. 2189–2192. doi:https://doi.org/10.1109/OCEANS.2005.1640089.
Schoening, T., Kuhn, T., Bergmann, M., Nattkemper, T.W., 2015. Delphi—fast and adaptive computational laser point detection and visual footprint quantification for arbitrary underwater image collections. Front. Mar. Sci. 2, 20. https://doi.org/10.3389/fmars.2015.00020.
Sedlazeck, A., Koser, K., Koch, R., 2009. 3d reconstruction based on underwater video from rov kiel 6000 considering underwater imaging conditions. In: OCEANS 2009-EUROPE, pp. 1–10. doi:https://doi.org/10.1109/OCEANSE.2009.5278305.
Shen, S., 2013. Accurate multiple view 3d reconstruction using patch-based stereo for large-scale scenes. IEEE Trans. Image Process. 22, 1901–1914. https://doi.org/10.1109/TIP.2013.2237921.
Snavely, N., Seitz, S.M., Szeliski, R., 2008. Modeling the world from internet photo collections. Int. J. Comput. Vision 80, 189–210. https://doi.org/10.1007/s11263-007-0107-3.
Soloviev, A., Venable, D., 2010. Integration of gps and vision measurements for navigation in gps challenged environments. In: IEEE/ION Position, Location and Navigation Symposium, pp. 826–833. doi:https://doi.org/10.1109/PLANS.2010.5507322.
Spaenlehauer, A., Fremont, V., Sekercioglu, Y.A., Fantoni, I., 2017. A loosely-coupled approach for metric scale estimation in monocular vision-inertial systems. In: 2017 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), pp. 137–143. https://doi.org/10.1109/MFI.2017.8170419.
Storlazzi, C.D., Dartnell, P., Hatcher, G.A., Gibbs, A.E., 2016. End of the chain? rugosity and fine-scale bathymetry from existing underwater digital imagery using structure-from-motion (sfm) technology. Coral Reefs 35, 889–894. https://doi.org/10.1007/s00338-016-1462-8.
Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W., 1999. Bundle adjustment – a modern synthesis. In: Vision Algorithms: Theory and Practice. Springer, Berlin, Germany, pp. 298–372. https://doi.org/10.1007/3-540-44480-7_21.
Tusting, R., Davis, D., 1986. Non-conventional techniques for sampling and collecting marine organisms. In: Proceedings of the Pacific Congress on Marine Technology, PACON'86, pp. 12–18.
Tusting, R.F., Davis, D., 1993. Improved methods for visual and photographic benthic surveys.
Tusting, R.F., Davis, D.L., 1992. Laser systems and structured illumination for quantitative undersea imaging. Mar. Technol. Soc. J. 26, 5–12.
Waechter, M., Moehrle, N., Goesele, M., 2014. Let there be color! large-scale texturing of 3d reconstructions, in: Computer Vision–ECCV. Springer, Berlin, Germany, pp. 836–850. doi:https://doi.org/10.1007/978-3-319-10602-1_54.
Wakefield, W.W., Genin, A., 1987. The use of a canadian (perspective) grid in deep-sea photography. Deep Sea Research Part A. Oceanogr. Res. Pap. 34, 469–478. https://doi.org/10.1016/0198-0149(87)90148-8.
Wallace, L., Lucieer, A., Malenovsky, Z., Turner, D., Vopenka, P., 2016. Assessment of forest structure using two uav techniques: a comparison of airborne laser scanning and structure from motion (sfm) point clouds. Forests 7. https://doi.org/10.3390/f7030062.
Zhang, J., Singh, S., 2015. Visual-inertial combined odometry system for aerial vehicles. J. Field Robot. 32, 1043–1055. https://doi.org/10.1002/rob.21599.

Klemen Istenič: received his Diploma degree in Computer Science from the University of Ljubljana, Slovenia in 2013 and a joint M.Sc. degree with distinction in Computer Vision and Robotics (European Master ViBOT) from Heriott-Watt University, UK, University of Girona, Spain, and the University of Burgundy, France in 2015. He is currently pursuing a Ph.D. in Technology at the University of Girona as a member of the Underwater Robotics Research Center (CIRS), part of the Computer Vision and Robotics Institute, as well as a member of an European Academy for Marine and Underwater Robotics (EU FP7 Marie Curie ITN network no 608096 – Robocademy). His research focuses on 3D mapping, color restoration and change detection using optical data in underwater scenarios.

Dr. Nuno Gracias: was awarded the Ph.D. degree in 2003 from the Technical University of Lisbon, Portugal. From 2004 to 2006 he was a post-doctorate fellow at the University of Miami. Since 2006 he has been a member of the Computer Vision and Robotics Group (ViCOROB) of the University of Girona. His research interests span the areas of underwater optical mapping, and navigation and guidance of autonomous underwater robots, image processing and classification. Dr. Gracias has authored more than 80 articles in peer-review journals and scientific conferences, and co-supervised 3 PhD and 8 MSc theses. He is adjunct faculty at the department of Marine Geosciences of the University of Miami, and member of the editorial board of the Journal of Intelligent and Robotic Systems.

Dr. Aurélien Arnaubec: received his Ph.D degree from the Universé Aix-Marseille III,Marseille, France, in 2012. He did his Ph.D. degree at the French Aerospace Laboratory, Office National d'Etudes et Recherches Aérospatiales, (ONERA), Salon Cedex Air, France, and in the Physics and Image Processing Group, Fresnel Institute, Marseille, France where his main research interests was radar imaging and statistical signal processing for remote sensing. Then from 2012 to 2014 he did a postdoc at Ifremer, la Seyne sur Mer, France, where he developed image processing techniques for optical mapping. He has now joined Ifremer PRAO team (Positionning, Robotic, Acoustic and Optics) where he is in charge of all underwater optical systems and image processing software.

Dr. Javier Escartín: received his PhD from the MIT/WHOI Joint Program in 1996 (USA). He is now a CNRS Senior Research Scientist at the Institute de Physique du Globe de Paris – Université de Paris, where since 2017 he leads the Marine Geosciences Group. He is also Associated CNRS Professor at the ENS of Paris. His research focuses on deep-sea exploration of the seafloor to understand geological processes, such as tectonism, volcanism, and hydrothermal activity. For this research he employs acoustic and optical mapping of the seafloor acquired with deep-sea vehicles.

Dr. Rafael Garcia: received a M.S. degree in computer engineering from the Universitat Autonoma de Barcelona in 1994, and a Ph.D. in Computer Engineering from the Universitat de Girona (Spain) in 2001. He is the founder and director of the Underwater Vision Lab (UVL), within the Computer Vision and Robotics Group in the Department of Computer Architecture at the University of Girona. His main research interests are underwater robotics and computer vision. He is particularly interested in understanding how to make underwater vehicles sense their environment in order to carry out autonomous surveys. He has published more than 190 technical contributions, including journal papers and conference proceedings. Dr. Garcia is a Member of the IEEE.