Peer-reviewed version available at *Remote Sens.* **2019**, *11*, 2093; doi:10.3390/rs11182093

Article

Scale Accuracy Evaluation of Optical Based 3D Reconstruction Strategies using Laser Photogrammetry

Klemen Istenič [1,2,*], Nuno Gracias [1], Aurélien Arnaubec[3], Javier Escartín[4] and Rafael Garcia [1]

[1] Underwater Robotics Research Center (CIRS), Computer Vision and Robotics Institute (VICOROB), University of Girona, Edifici P-IV, Campus de Montilivi, 17071 Girona, Spain; klemen.istenic@gmail.com (K.I.); ngracias@silver.udg.edu (N.G.); rafael.garcia@udg.edu (R.G)
[2] Coronis Computing, S.L., Science and Technological Park of UdG, Carrer Pic de Peguera, 15, 17003 Girona, Spain
[3] IFREMER, Ctr Mediterranee, Unité Syst. Marins, CS 20330, F-83507 La Seyne Sur Mer, France; aurelien.arnaubec@ifremer.fr
[4] Université de Paris, Institut de Physique du Globe de Paris, CNRS, F-75005, France; escartin@ipgp.fr
* Correspondence: klemen.istenic@gmail.com

Abstract: Rapid developments in the field of underwater photogrammetry have given scientists the ability to produce accurate 3-dimensional (3D) models which are now increasingly used in the representation and study of local areas of interest. This paper addresses the lack of systematic analysis of 3D reconstruction and navigation fusion strategies, as well as associated error evaluation of models produced at larger scales in GPS-denied environments using a monocular camera (often in deep-sea scenarios). Based on our prior work on automatic scale estimation of Structure from Motion (SfM)-based 3D models using laser scalers, an automatic scale accuracy framework is presented. The confidence level for each of the scale error estimates is independently assessed through the propagation of the uncertainties associated with image features and laser spot detections using a Monte Carlo simulation. The number of iterations used in the simulation was validated through the analysis of the final estimate behaviour. To facilitate the detection and uncertainty estimation of even greatly attenuated laser beams, an automatic laser spot detection method, mitigating the effects of scene texture, was developed, with the main novelty of estimating the uncertainties based on the recovered characteristic shapes of laser spots with radially decreasing intensities. The effects of four different reconstruction strategies resulting from the combinations of Incremental/Global SfM, and the *a priori*/*a posteriori* use of navigation data were analyzed using two distinct survey scenarios captured during the SUBSAINTES 2017 cruise (doi: 10.17600/17001000). The study demonstrates that surveys with multiple overlaps of non-sequential images result in a nearly identical solution regardless of the strategy (SfM or navigation fusion), while surveys with weakly connected sequentially acquired images are prone to produce broad-scale deformation (doming effect) when navigation is not included in the optimization. Thus the scenarios with complex survey patterns substantially benefit from using multi-objective BA navigation fusion. In all cases, the errors in the models are inferior to 5%, with errors often being around 1%. The effects of combining data from multiple surveys were also evaluated. The introduction of additional vectors in the optimization of multi-survey problems successfully accounted for offset changes present in the underwater USBL-based navigation data and thus minimize the effect of contradicting navigation priors. Our results also illustrate the importance of collecting a multitude of evaluation data at different locations and moments during the survey.

Keywords: Photogrammetry; Metrology; Underwater 3D Reconstruction; Structure-from-Motion; Navigation Fusion; Multi-Objective BA; Laser Scalers; Monte-Carlo Simulation; Uncertainty Estimation; Scale Drift Evaluation; Laser Spot Detection.

1. Introduction

Accurate and detailed 3D models of the environment are now an essential tool in different scientific and applied fields such as geology, biology, engineering, archaeology, among others. With advancements in photographic equipment and improvements in image processing and computational capabilities of computers, optical cameras are now widely used due to their low cost, ease of use, and sufficient accuracy of the resulting models for their scientific exploitation. The application of traditional aerial and terrestrial photogrammetry has greatly expanded in recent years, with commercial and custom build camera systems and software solutions enabling nearly black-box type of data processing (e.g., [1–4]).

These rapid developments have also significantly benefited the field of underwater photogrammetry. The ability to produce accurate 3D models from monocular cameras under unfavorable properties of the water medium (i.e., light attenuation and scattering, among other effects) [5], and advancements of unmanned underwater vehicles has given scientists unprecedented access to image the seafloor and its ecosystems from shallow waters to the deep ocean [6–9]. Optical seafloor imagery is now routinely acquired with deep-sea vehicles, and often associated with other geophysical data (acoustic backscatter, multibeam bathymetry) and water column measurements (temperature, salinity, chemical composition). High resolution 3D models with associated textures are thus increasingly used in the representation and study of local areas of interest. However, most remotely operated vehicles (ROVs) or autonomous underwater vehicles (AUVs) that are used nowadays in science missions have limited optical sensing capabilities, commonly consisting of a main camera used by the ROV-pilot, while larger workclass ROVs have additional cameras for maneuvering. Due to the nature of projective geometry, performing 3D reconstruction using only optical imagery acquired by monocular cameras results in a 3D model which is defined only up to scale, meaning that a unit is not necessary a standardized unit such as a meter [10]. In order to correctly disambiguate the scale, it is essential to use additional information in the process of model building. Predominantly, solutions in sub-aerial applications are based on the fusion of image measurement with robust and dependable satellite references, such as Global Navigation Satellite System (GNSS) [11–13], or ground control pointss (GCPs)[14–16], due to their accuracy and ease of integration. On the contrary, the water medium not only hinders the possibility of accurately establishing the control points, but also prevents the use of global positioning system (GPS) due to the absorption of electromagnetic waves. Hence the scale is normally disambiguated either using a combination of acoustic positioning (e.g., Ultra-Short BaseLine (USBL)) and inertial navigation system (INS) [17–19], or through the introduction of known distances between points in the scene [20].

In shallow water environments, researchers have often placed auxiliary objects (such as a scaling cube [21], locknuts [22], graduated bars [23], etc.) into the scene, and used the knowledge of their dimensions to scale the model *a posteriori*. Such approaches, while applicable in certain scenarios, are limited for small scale reconstructions, and for shallow water environments, due to the challenges in transporting and placing objects in deep-sea environments. Similarly, laser scalers have been used since late 1980s projecting parallel laser beams onto the scene to estimate the scale of the observed area, given known geometric setup of the lasers. Until recently, lasers have been mostly used in image-scaling methods, for measurements within individual images (e.g., Pilgrim et al. [24] and Davis and Tusting [25]). To provide proper scaling, we have recently proposed two novel approaches [26], namely fully- (FCM) and partially-calibrated method (PCM), to automatically estimate 3D model scale using a single optical image with identifiable laser projections. The proposed methods alleviate

Peer-reviewed version available at Remote Sens. 2019, 11, 2093; doi:10.3390/rs11182093

numerous restrictions imposed by earlier laser photogrammetry methods (e.g., laser alignment with the optical axis of the camera, perpendicularity of lasers with the scene), and removes the need for manual identification of identical points on the image and 3D model. The main drawback of these methods is the need for purposeful acquisition of images with laser projections, with the required additional acquisition time.

Alternatively, the model scaling can be disambiguated with known metric vehicle displacements (i.e., position and orientation from acoustic positioning, Doppler systems, and depth sensors [19,27,28]). As this information is recorded throughout the mission, such data is normally available for arbitrary segments even if they have not been identified as interesting beforehand. The classic range-and-bearing position estimates from acoustic-based navigation, such as USBL, have an uncertainty that increases with increasing range (i.e., depth) in addition to possible loss of communication (navigation gaps). Consequently, the scale information is inferred from data which is often noisy, poorly resolved, or both. Hence the quality of the final dataset is contingent on the strategy used in the fusion of image and navigation information. Depending on the approach, the relative ambiguity can cause scale drift, i.e. a variation of scale along the model, causing distortions [29]. Furthermore, building of large 3D models may require fusion of imagery acquired in multiple surveys. This merging often results in conflicting information among different dives, and affect preferentially areas of overlap between surveys, negatively impacting the measurements on the model (distances, areas, angles).

The need to validate the accuracy of optical-based 3D models has soared as the possibilities of using standard imaging systems increase and replace the need for more complex and dedicated reconstruction techniques (e.g., structured light). Numerous evaluations of this accuracy are available for aerial and terrestrial 3D models (e.g., [2,30–32]). Environmental conditions and limitations of underwater image acquisition preclude their transposition to underwater image acquisition and, to date, most underwater accuracy studies use known 3D models providing reference measurements.

Early studies [33–39] evaluated the accuracy of small-scale reconstructions (mainly on coral colonies), comparing model-based and laboratory-based volume and surface areas for specific corals. More recently, auxiliary objects (e.g., locknuts [22], graduated bars [23], special frames [40,41] and diver weights [42]) have been used to avoid removal of objects from the environment. Reported inaccuracies range from 0.85% to 17%, while more recent methods achieve errors as low as 2%-3% [22, 42]. Diver-based measurements and/or placement of multiple objects at the seafloor restricts the use of these methods to shallow-water or experimental environments, and hinder such approaches in deep sea environments (e.g., scientific cruises), where reference-less evaluation is needed instead, which has been performed in only a few experiments.

Ferrari et al. [39] evaluated their reconstruction method on a medium size reef area (400 m) and a 2 km long reef transect. Maximum heights of several quadrants within the model were compared to in situ measurements, coupled with an estimation of structural complexity (rugosity). The average accuracy in reef height was 82% $\pm$ 2%. This study split larger transects into approx 10 m long sections to reduce potential drift, and hence model distortion. Similarly, Gonzales et al. [43] obtained 85% accuracy in rugosity estimates from stereo imaging and compared with results from a standard chain-tape method, along a 2 km long transect. To the best of our knowledge, no other scale accuracy estimate of submarine large-area models has been published. Furthermore, although laser scalers are often used for qualitative visual scaling, they have never been used to evaluate the accuracy of underwater 3D models.

Objectives

While a growing body of literature supports that underwater optical-based 3D reconstruction is a highly efficient and accurate method at small spatial extents, there is a clear gap in the accuracy analyses of models produced at larger scales (often in deep-sea scenarios). Validation of 3D reconstruction methods, and associated error evaluation, are thus required for large underwater scenes and to allow

Peer-reviewed version available at Remote Sens. 2019, 11, 2093; doi:10.3390/rs11182093

quantitative measurements (distances and volumes, orientations, etc.) required for scientific and technical studies.

The main goal of this paper is to present an automatic scale accuracy estimation framework, applicable to models reconstructed from optical imagery and associated navigation data. We also evaluate various reconstruction strategies, often used in academic and private ROVs deep-sea surveys. The framework is based on the method recently presented by Istenič et al. [26] for automatic scale estimation of SfM-based 3D models.

First, we present several methods of 3D reconstruction using underwater vehicle navigation, to provide both scaling and an absolute geographic reference. Most commonly, SfM uses either an incremental or a global strategy, while the vehicle navigation may be considered *a priori* as part of the optimization process, or *a posteriori* after full 3D model construction. Here we compare four different strategies resulting from the combinations of Incremental/Global SfM, and the *a priori*/*a posteriori* use of navigation data. We discuss the impact of each of these strategies on the final 3D model accuracy.

Second, the four methods are evaluated to identify which is best suited to generate 3D models that combine data from multiple surveys, as it is often required under certain surveying scenarios. Navigation from different surveys may have significant offsets at the same location (x, y, z, rotation), show noise differences, or both. The changes between different acquisitions of a single scene are taken into account differently by each 3D reconstruction strategy.

Third, prior approaches to estimate model scale using laser scalers, namely FCM and PCM methods, are augmented with Monte Carlo simulations to evaluate the uncertainty of obtained scale estimates. Furthermore, the results are compared to estimates commonly used and suffering from parallax error.

Fourth, an automatic laser detection and uncertainty estimation method is presented. Accurate analyses requires a multitude of reliable measurements spread across the 3D model, whose manual annotation is extremely labor intensive, error-prone, and time consuming, when not nearly impossible. Unlike previous detection methods, our method detects centers of the lasers by considering the texture of the scene, and determines their uncertainty, which, to the best of our knowledge, has not been presented in the literature yet.

With the data from the SUBSAINTES 2017 cruise (doi: 10.17600/17001000; [44]) we evaluate the advantages and drawbacks of the different strategies to construct underwater 3D models, while providing quantitative error estimates. As indicated above, these methods are universal as they are not not linked to data acquired with specific sensors (e.g., laser systems, stereo cameras), and can be applied to standard imagery acquired with underwater ROVs. Hence, it is possible to process legacy data from prior cruises and with different vehicles and/or imaging systems. Finally, we discuss the best practices to conduct optical surveys, based on nature of targets and the characteristics of the underwater vehicle and sensors.

2. Optical-based Underwater 3D reconstruction

In this section we present a brief overview of the most important steps of the 3D reconstruction process for underwater applications, laying out our approach to evaluate the accuracy of the models.

Textured 3D models result from a set of sequential processing steps (Fig. 1). As scene geometry is computed entirely from the optical imagery, the end result directly depends on image quality and adequate survey strategy. Compared to sub-aerial imagery, the unfavorable properties of water medium (i.e., light attenuation and scattering effects) [5] cause blurriness of details, low image contrast and distance-dependent alteration of colors [45]. To prevent overly degraded data, acquisition is conducted at close range, significantly limiting the observation area of any single image, while significantly increasing the amount of data collected and processed. Keyframe selection and color correction are hence preprocessing steps used to minimize the degradation effects of water, and to remove unnecessary redundancies (i.e., images taken from similar poses). A concise set of pre-filtered images is then used to estimate the initial sparse 3D geometry of the scene and the camera

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 July 2019
Peer-reviewed version available at Remote Sens. 2019, 11, 2093; doi:10.3390/rs11182093

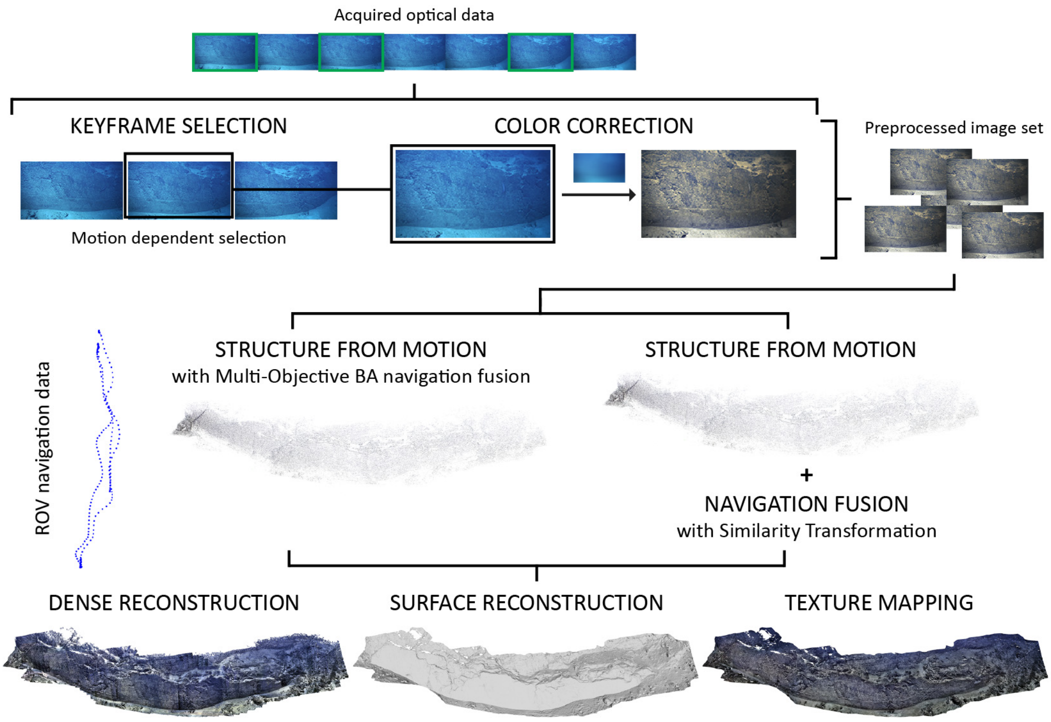


Figure 1. Flowchart of a 3D reconstruction process for underwater applications.

trajectory through a technique called Structure from Motion (SfM). The inherent scale ambiguity of the reconstructed 3D structure and camera motion from a set of images is addressed by either using the vehicle navigation *a priori* as part of the SfM optimization process (multi-objective BA), or *a posteriori* through an alignment with the reconstructed camera path using a similarity transformation. An accurate, high-detailed model description is subsequently obtained through an efficient patch-based stereo matching and fusion densification process, followed by a surface estimation from an unorganized noisy set of 3D points obtained earlier. A final photo-realistic 3D model uses a consistent high-quality texture from a seamless mapping of input images.

Accuracy of the measurements performed in 3D models, required for quantitative studies (precise measurement of distances and volumes, etc.), depends on the strategy used for optical-based 3D reconstruction, in addition to data quality itself. Four different approaches are often used:

A) Incremental SfM with *a posteriori* navigation fusion;
B) Global SfM with *a posteriori* navigation fusion;
C) Incremental SfM with multi-objective BA navigation fusion;
D) Global SfM with multi-objective BA navigation fusion.

2.1. Keyframe selection

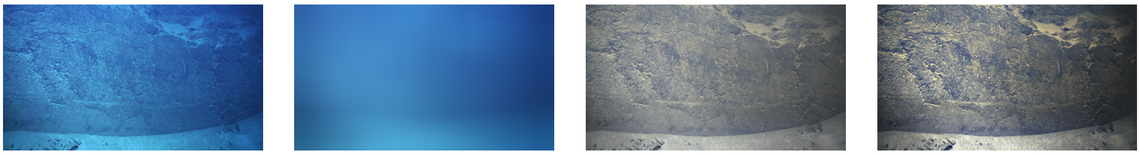
Surveying for underwater 3D models often produce redundant imagery so as to insure adequate imaging of areas that are of difficult access. Discarding unnecessary images is important to both reduce the computational time, and to minimize the possibility of unreliable depth estimations [10]. Commonly used time-dependent image selection (e.g., selecting a frame every n-th second) is often not suited; surveys with significant speed changes and/or distance to the scene lead to over or under filtering of images. Instead, we use an approach with implicit detection of frames with similar vantage points [46] through estimates of feature displacements between consecutive frames (e.g., Lucas–Kanade tracking algorithm [47]). For sufficiently dense image sets (e.g., video acquisitions), sharpness may be used for further selection (e.g., variance of Laplacian [48]).

Peer-reviewed version available at *Remote Sens.* **2019**, *11*, 2093; doi:10.3390/rs11182093

2.2. Color correction

Owing to the non-uniform absorption of the visible light spectrum over its frequency components [49,50], underwater images are typically bluish/greenish and present low contrast [51]. Minimizing the effects of the water medium not only benefits human perception and interpretation of the scene, but also improves the quality and quantity of successful feature matches between image pairs [52], thus increasing the quality of the final model.

Accurate color information recovery depends on the knowledge of the physical image formulation process model which is rarely available in its completeness. Alternatively, color enhancing methods (e.g., Bianco et al. [53] can remove the attenuation effects, as well as the color cast introduced by an unknown illuminant (Fig. 2).



(a) (b) (c) (d)

Figure 2. (**a**) Original UW image. (**b**) Chromatic components (α, β) of the estimated local illuminant. (**c**) White balanced image. (**d**) Final enhanced image.

2.3. Sparse Reconstruction

A sparse set of 3D points (the structure), and the camera parameters (motion) can be estimated from multiple projections of the same 3D point in overlapping images through the equations of projective geometry using SfM.

2.3.1. Feature detection and matching

As the structure and motion parameters are inferred entirely from feature points, robustness of detection and matching across the image set is important. In our approach, salient 2-dimensional (2D) points are detected as accelerated KAZE (AKAZE) local features [54], and described using a Modified-SURF descriptor [55], which was selected for its scale and local affine invariance properties. Feature association across the image set is performed over image pairs using descriptor matching with an additional geometric filtering procedure (e.g., fundamental/essential matrix [10]). To avoid an empirical selection of the inlier/outlier threshold in robust estimation techniques, the parameter-free A Contrario Ransac (AC-RANSAC) [56] is used to automatically determine the model meaningfulness by a statistical balance between the tight fitting of data and the number of the inliers.

With high number of images, the potential image pairs can be restricted either by pose (if navigation is available) or by image retrieval strategies [57,58].

2.3.2. Structure from Motion

Structure from Motion is a method in which the structure is jointly estimated with the motion of the camera from a noisy set of 2D features and their previously identified correspondences. The structure is expressed as a sparse set of 3D points $\mathcal{X} = \{X_k \in \mathbb{R}^3 \,|\, k = 1 \ldots L\}$, while camera motion is represented with the set of projection matrices $\mathcal{P} = \{P_i = [\boldsymbol{R}_i^T \,|\, -\boldsymbol{R}_i^T \boldsymbol{t}_i] \,|\, i = 1 \ldots N\}$, where $P_i \in \mathbf{SE}(\mathbf{3})$ defines the projection from world to camera frame. Additionally, intrinsic camera parameters $\mathcal{K} = \{K_i \mid i = 1 \ldots N\}$ can be considered in the optimization, leading to lower complexity of the problem and thus improving the results.

Due to the non-linearity in the projection process, a non-linear optimization, Bundle Adjustment (BA), is required. The solution is obtained by formulating a non-linear least squares (NLS) problem, which can be efficiently solved using iterative methods such as Levenberg-Marquardt (LM) [59]. The cost function to be minimized is normally an image-based error, consisting of the sum of squared

Peer-reviewed version available at Remote Sens. 2019, 11, 2093; doi:10.3390/rs11182093

re-projection errors (Eq. 1), defined as the distance between the 2D feature observations $\mathcal{F}_j = \{x_j \mid j = 1 \dots M\}$ of the 3D points X_k and their corresponding projections onto the images.

$$\mathcal{E}_{(v)} = \sum_{j=1}^{M} \left\| x_j - \mathrm{proj}(K_i, P_i, \mathbf{X}_k) \right\|^2 . \tag{1}$$

The LM algorithm only guarantees to find a local minimum of the optimizing function, making it extremely sensitive to the initial parameter estimate. The different strategies proposed to initialize these parameters can be broadly classified as either: *incremental* or *global*.

Incremental SfM expands model reconstruction one image at the time, allowing for a gradual estimate of parameters for the newly added points and cameras. Each image is registered by solving the Perspective-n-Point (PnP) problem, followed by a triangulation to augment the set of scene points. After each addition, intermediate BA can be performed to propagate and minimize the error of intermediate reconstructions. Incremental approaches are broadly used given that the intermediate partial reconstructions enable a more robust detection of outliers and thus decrease the chance of convergence to a wrong local minimum. However, when no prior information about the scene is available, the initialization step of decomposing the fundamental/essential matrix is critical, as a poor selection of the seed pair of images can quickly force the optimization to a non-recoverable state. Furthermore, as the method inherently gives disproportionate weight to images used at the beginning of the process, it can result in error accumulation. This may produce significant drift and fail to reconstruct the scene in the form of a single connected model. In our tests, the method of Moulon et al. [60,61] was used with a contrario model estimation.

Global SfM considers instead the entire problem at once, with full BA performed only at the end. To alleviate the lack of partial reconstructions, that identifies possible outliers, the parameter initialization is split into two sequential steps (i.e., rotation and translation estimation), the first one being more robust to a small number of outliers. This mitigates the need for intermediate non-linear optimizations, as camera and scene points are estimated simultaneously in a single iteration. It also ensures an equal treatment of all the images and consequently equal distribution of the errors. The methods rely on averaging relative rotations and translations, thus requiring images to have overlap with multiple other images, to ensure meaningful constraints and mutual information. As a consequence, the reconstruction from a sparsely connected set of images will result in distorted or even multiple disconnected components. In our test Moulon et al. [61,62] method was used.

2.4. Navigation Fusion

Joint reconstruction of 3D structure and camera motion from a set of images acquired by a single camera is an inherently ill-conditioned problem, with a solution determined only up to an unknown scale [10]. The estimated parameters can be multiplied by an arbitrary factor, resulting in an equal projection of the structure on the images. A metric solution thus requires known measurements [20] or metric vehicle displacements (navigation/inertial priors) [19,27,28]. Depending on the availability of synchronization between the camera and the navigation, priors $\mathcal{C} = \{C_i \mid i = 1 \dots N\}$ extracted from the ROV/AUV's navigation, can either be used in a multi-sensor fusion approach or to align the reconstructed camera path via a similarity transformation.

2.4.1. Multi-objective BA

When navigation priors are available for a significant proportion of images, then this information can be incorporated in the optimization through a multi-sensor fusion approach. The fusion is defined as a multi-objective optimization consisting of re-projection ($\mathcal{E}_{i(v)}$) and navigation fit errors ($\mathcal{E}_{i(n)} = T_i - C_i$). Most commonly, there is no unique solution that would simultaneously optimize both

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 July 2019

Peer-reviewed version available at Remote Sens. 2019, 11, 2093; doi:10.3390/rs11182093

objectives, but instead exists a hyper-surface of Pareto optimal solutions[1]. Such solution space can be defined as a weighted compound function of the two objectives [63]. Assuming that both re-projection and navigation fit errors are independent and Gaussian, it is statistically optimal to weight the errors by their variance [64,65]:

$$\mathcal{E} = \frac{1}{M\sigma_v^2}\sum_{j=1}^{M}\left\|\mathcal{E}_{j(v)}\right\|^2 + \frac{1}{N\sigma_n^2}\sum_{i=1}^{N}\left\|\mathcal{E}_{i(n)}\right\|^2 = \sum_{j=1}^{M}\left\|\mathcal{E}_{j(v)}\right\|^2 + \frac{M}{N}\lambda^2\sum_{i=1}^{N}\left\|\mathcal{E}_{i(n)}\right\|^2, \tag{2}$$

where $\lambda = \sigma_v/\sigma_n$ indicates the ratio between the two covariances, representing the noise variance of each sensor measurement and M and N are the number of re-projection and navigation prior terms.

The selection of the preferred solution on the Pareto Frontier [66] crucially depends on the knowledge of the ratio of variances, often unknown, in different units, or both (e.g., pixels vs. meters). In those cases, the weight can be selected empirically or through automatic weight determining methods.

For bi-objective optimizations, Michot et al. [63] have shown that the L-Curve criterion is the preferred selection method. This criterion is based on plotting the trade-off between the cost of the objectives using different weights, represented in log-log space. This plot has a typical L-curve shape, with two prominent segments. Each term dominating a segment (flat and vertical part) is used to detect the "corner" separating the two, essentially identifying a neutral objective dominance. The associated weight is considered to be the optimal, and representative of the ratio between the covariances of the sensors. Lying between two nearly flat segments, it can be easily identified as the point with maximum curvature.

2.4.2. Similarity Transformation

Alternatively, the navigation data can be used in an *a posteriori* step of re-scaling and geo-referencing. A similarity transformation, which minimizes the sum of differences between the reconstructed camera poses and their navigation priors, is applied to the reconstructed model. Depending on the survey pattern, this method can be used even in cases when the camera is not synchronized with the navigation data. If the reconstructed path can be unambiguously matched to the path given by the navigation data, then the associations between the cameras and navigation poses can be determined through finding the closest points between the paths.

2.5. Dense Reconstruction

To accurately describe the scene geometry in high detail, a dense representation is computed using the method of Shen [67]. For each image reconstructed in SfM, a depth-map is computed, and subsequently refined to enforce consistency over neighboring views. Initial depth-map estimates are generated by projecting points of the sparse reconstruction and interpolating intermediate depths with Delaunay triangulation. Using assigned reference images (i.e., images with a similar viewing direction and suitable baseline), each depth-map is improved by iterative spatial propagation and random assignment operations. The depth of each pixel is refined with information of neighboring pixels, subsequently reducing the discrepancy between the local window around the pixel and the projected patch on the reference image. This assumes that neighboring pixels likely have similar depths. Once estimated, depth maps are merged into a single (dense) set of 3D points. Points with high photometric inconsistencies are removed to suppress those violating the visibility constraints, efficiently reducing noise and outliers in the final dense representation of the 3D scene geometry.

---

[1] Pareto optimal solutions refer to solutions of objectives functions that can be improved solely by degrading a different objective.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 July 2019
Peer-reviewed version available at Remote Sens. 2019, 11, 2093; doi:10.3390/rs11182093

2.6. Surface and Texture Reconstruction

The obtained 3D sparse and dense point clouds are an unorganized and noisy scene description. A final photo-realistic 3D model requires an estimate of the surface and a consistent high-quality texture by seamlessly mapping input images. As underwater optical-based reconstructions are inevitably corrupted by both noise and outliers due to poor imaging conditions [5], an approximation-based surface reconstruction method is used [68]. It computes the most probable surface, given the available sampling of the scene, modeling the surface as an interface between the free and full space as opposed to directly using the input points. This method efficiently mitigates noise discrepancies and yields a robust reconstruction of weakly represented surfaces. The reconstruction is completed by estimating the texture with a two step method [69]. Initially, each mesh triangle is assigned the best representative image through an energy minimization process, which attempts to minimize color discontinuities between neighboring regions. This method prefers close, focused and orthogonal high-resolution views as well as similar adjacent patches. To mitigate texture inconsistencies due to inaccuracies in the estimation of camera poses and the scene, as well as unreconstructed occluding objects, an additional photo-consistency check is employed. Finally, any significant color discontinuities between neighboring regions are addressed by per-vertex-based globally optimal luminance correction as well as with Poisson image editing [70].

3. Model Evaluation Framework

Estimating the scale accuracy of 3D models reconstructed from underwater optical imagery and robot navigation data is of paramount importance since the input data is often noisy and erroneous. The noisy data commonly leads to inaccurate scale estimates and noticeable variations of scale within the model itself, which precludes the use of such models for their intended science applications. Real underwater scenarios usually lack elements of known sizes that could be readily used as size references to evaluate the accuracy of 3D models. However laser scalers are frequently used during underwater image collection to project laser beams onto the scene and can be used to provide such size reference.

The framework builds upon two methods for scale estimation of SfM-based 3D models using laser scalers, that were recently introduced [26]. We extend the scale estimation process by including it into a Monte Carlo (MC) simulation, where we propagate the uncertainties associated with the image features and laser spot detections through the estimation process.

As the evaluated models are built with metric information (e.g., the vehicle navigation data, dimension of auxiliary objects), their scale is expected to be consistent with the scale provided by the laser scaler (s_L). Therefore, any deviation from the expected scale value ($s = 1.0$) can be regarded as an inaccuracy of the scale of the model (ϵ_s). The error thus represents the percentage for which any spatial measurement using the model will be affected.

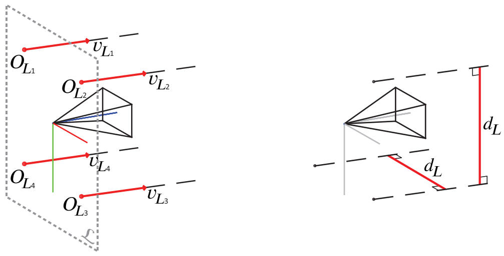
$$\epsilon_s = s_L - 1.0 = \frac{m}{\hat{m}} - 1.0\,, \tag{3}$$

where m and $\hat{m}$ represent a known metric quantity and its model based estimate.

3.1. Scale Estimation

The two methods, namely fully-calibrated method (FCM) and partially-calibrated method (PCM) are both suitable for different laser scaler configurations. FCM permits an arbitrary position and orientation for each of the lasers in the laser scaler, at the expense of requiring a full *a priori* knowledge of their geometry relative to the camera (Fig. 3a). On the other hand, the laser-camera constraints are significantly reduced for using the PCM method. The laser origins have to be equidistant to the camera center and laser pairs have to be parallel (Fig. 3b). As opposed to prior image scaling methods [24,25], the lasers do not have to be aligned with the optical axis of the camera.

Both methods exploit images with visible intersections of laser beams with the scene beyond the simple location of the laser spots. The model scale is estimated through a three step process: laser



(a) (b)

Figure 3. (**a**) Fully- and (**b**) partially-calibrated setup consisting of an optical camera and lasers, with the required information marked in red.

detection, pose estimation and scale estimation (Fig. 4). The two initial steps are identical in both methods; First, a laser detection method determines the locations of laser spots on an image; Second, the pose of the camera (wrt. the 3D model) at the time of image acquisition is estimated through a feature-based localization process.

The initial camera extrinsic values (and optionally camera intrinsics) are obtained by solving an PnP problem [71] using 3D-2D feature pairs. Each pair connects an individual image feature and a feature associated with the sparse set of points representing the model. As these observations and matches are expected to be noisy and can contain outliers, the process is performed in conjunction with a robust estimation method A-Contrario Ransac (AC-RANSAC) [56]. The estimate is further refined through a non-linear optimization (BA) minimizing the re-projection error of known (and fixed) 3D points and their 2D observation on the image.

The camera pose and location of the laser spots are lastly used either to estimate the position of the laser origin so as to produce the recorded result (FCM), or to estimate the perpendicular distance between the two parallel laser beams (PCM). As these predictions are based on the 3D model, they are directly affected by its scale, and can therefore be used to determine it through a comparison with *a priori* known values. As shown through an extensive evaluation in our previous work, both FCM and PCM can be used to estimate model scale regardless of the camera view angle, camera-scene distance, or terrain roughness [26]. The use of a maximum likelihood estimator (BA) and a robust estimation method (AC-RANSAC), the final scale estimation is minimally affected by noise in the detection of feature positions and the presence of outlier matches .

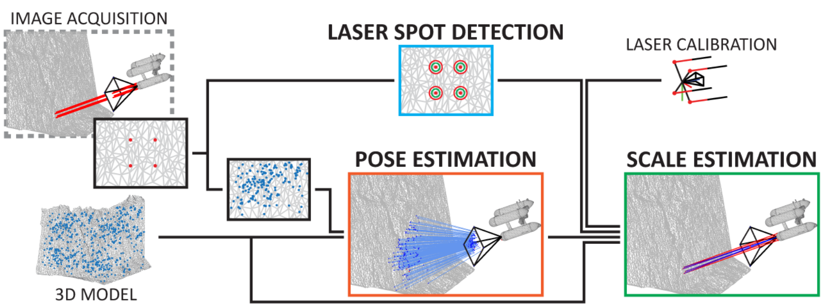
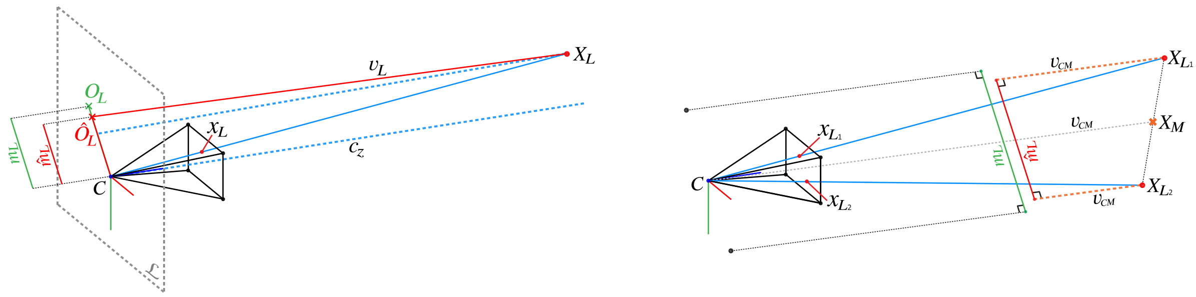


Figure 4. Flowchart of the scale estimation process depicting three crucial steps in scale estimation: laser spot detection, pose estimation, and scale estimation.

In the fully-calibrated method, the knowledge of the complete laser geometry is used (origins O_L and directions v_L) to determine the position of laser emission $\hat{O}_L$, and so as to produce the results observed on the image (Eq. 4). The laser origins $\hat{O}_L$ are predicted by projecting 3D points X_L, representing the location of laser beam intersections with the model, using a known direction of the beam v_L. As the points X_L had to be seen by the camera, i.e. be in the line-of-sight of the camera, their positions can be deducted by a ray casting procedure using a ray starting in the camera center and passing through the laser spot x_L detected in the image. The final scale estimate can then be determined by comparing the displacement of the $\hat{m}_L = \|\hat{O}_L\|$ with its *a priori* known value $\|O_L\|$.

$$\hat{O}_L = \boldsymbol{P}X_L - \frac{\boldsymbol{P}X_L \cdot c_z}{v_L \cdot c_z} v_L \,, \tag{4}$$

where $\boldsymbol{P}$ is defined as the projection from world to camera frame and c_z represents the optical axis of the camera.



(a) (b)

Figure 5. Scale estimation using (**a**) fully-calibrated and (**b**) partially-calibrated approach, based on the 3D model and optical image depicting the laser beam projection on the scene intersection with the scene.

Alternatively, the partially-calibrated method can be used when laser pairs are parallel but with unknown relation with the camera. As opposed to other image scaling methods, laser alignment with the optical axis of the camera is not required, allowing its application to numerous scenarios in which strict rigidity between camera and lasers is undetermined or not maintained (e.g., legacy data). To overcome the lack of information about the direction of the laser beams wrt. the camera, equidistance between the laser origins and the camera center is exploited. Laser beam direction is thus approximated with the direction of the vector connecting the camera center and the middle point between the two points of lasers intersections with the model v_{CM}. As we have showed in our previous work [26], this approximation can lead to small scaling errors in the most extreme cases where the depth discrepancy between two points on the model is disproportionally large compared to the camera-scene distance. As underwater surveys are always conducted at sufficiently large safety distances, this scenario is de facto absent in underwater reconstructions.

3.2. Uncertainty Estimation

Uncertainty characterization of each scale estimate is key for quantitative studies (precise measurement of distances and volumes, orientations, etc.), as required in marine science studies where accurate metrology is essential (such as in geology, biology, engineering, archaeology and others). The effect of uncertainties of input values on the final estimate is evaluated using a MC simulation method. The propagation through the process is modelled by repetitions of computation of the same quantities, while statistically sampling the input values based on their probability distributions. Final uncertainty estimate in scale is derived from the independently computed values.

Figure 6 depicts the complete MC simulation designed to compute the probability distribution of an estimated scale error, computed from multiple laser observations in an image. We assume that the sparse 3D model points, associated with the 2D features in the localization process, are constant and thus noise free. On the other hand, uncertainty of the imaging process and feature detection is characterized using the re-projection error obtained by the localization process. We also account for the plausible uncertainty in the laser calibration and laser spot detection, with each laser being considered independently.

Peer-reviewed version available at Remote Sens. 2019, 11, 2093; doi:10.3390/rs11182093

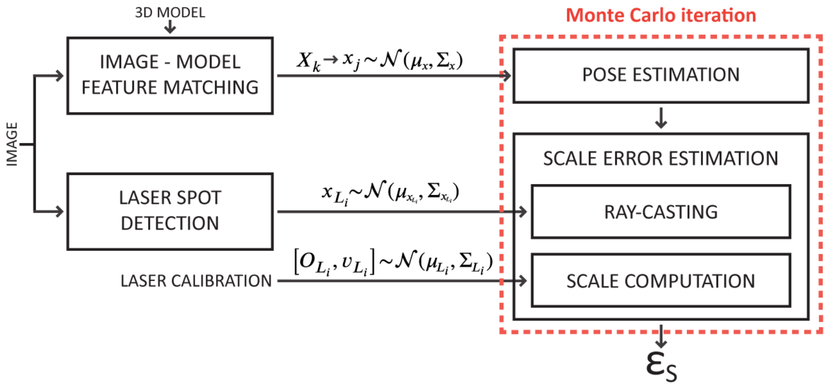


Figure 6. Monte Carlo simulation scheme used for propagating input uncertainties through the process of scale error estimation.

4. Laser Spot Detection

The accurate quantification of scale errors affecting 3D models derived from imagery requires numerous reliable measurements that have to distributed throughout the model. As scale estimates are obtained by exploiting the knowledge of laser spot positions on the images, the quantity and quality of such detections directly determines the number of useful scale estimates. Furthermore, to properly estimate the confidence levels of such estimated scale, the uncertainty of the laser spot detections needs to be known.

The laser beam center is commonly considered to be the point with the highest intensity in the laser spot, as the luminosity of laser spots normally overpowers the texture of the scene. However, due to the properties of the water medium, the laser light can significantly attenuate on its path to the surface, before being reflected back to the camera. In such cases, the final intensity of the beam reaching the camera might be overly influenced by the texture at the point of the impact (Fig. 7). As such, performing manual accurate annotations of laser spots tends to be extremely challenging and labor intensive, and even impossible in certain cases.

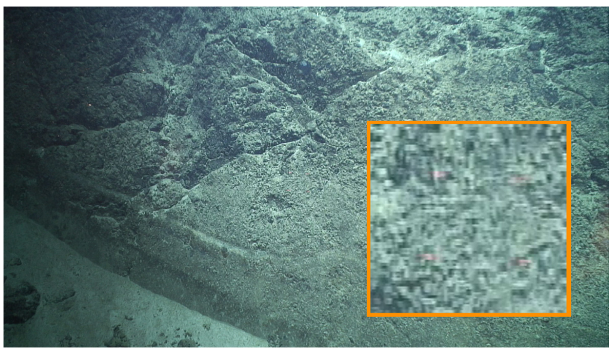


Figure 7. Example of image used for scale error evaluation with enlarged laser area.

Considerable attention has been given to the development of the image processing components of laser scanners, namely on laser line detection [72,73], while the automatic detection of laser dots from underwater laser scalers has only been addressed in few studies. Rzhanov et al. [74] developed a toolbox (The Underwater Video Spot Detector - UVSD), with a semi-automatic algorithm based on a Support Vector Machine (SVM) classifier. Training of this classifier requires user-provided detections. Although the algorithm can provide a segmented area of the laser dot, this information is not used for uncertainty evaluation. More recently, [75] presented a web-based, adaptive learning laser point detection for benthic images. The process comprises a training step using k-means clustering on color features, followed by a detection step based on a k-nearest-neighbor (kNN) classifier. From this training on laser point patterns the algorithm deals with a wide range of input data, such as the cases

Peer-reviewed version available at Remote Sens. 2019, 11, 2093; doi:10.3390/rs11182093

of having lasers of different wavelengths, or acquisitions under different visibility conditions. Neither the uncertainty in laser point detection nor the laser line calibration are addressed by this method.

To overcome the lack of tools capable of detecting and estimating the uncertainty in laser spot detection, while producing robust and accurate detections, we propose a new automatic laser detection method. To mitigate the effect of laser attenuation on the detection accuracy, scene texture is considered while estimating the laser beam center. We use a Monte Carlo simulation to estimate the uncertainty of detections, consider the uncertainty of image intensities.

4.1. Detection

To determine laser spot positions on any image, the first step is a restriction of the search area to a patch where visible lasers are expected (Fig. 8a). While not compulsory, this restriction minimizes false detections and reduces computational complexity and cost. The predicted area may be determined from the general pose of lasers with respect to the camera, and from the range of distances to the scene.

An auxiliary image is used to obtain a pixel-wise aligned description of the texture in the patch. This additional image is assumed to have been acquired at a similar distance, and with laser spots either absent or in different positions. This ensures visually similar texture information at the positions of the laser spots. The requirement is easily achievable for video acquisitions, as minor changes in camera pose sufficiently change the positions of the lasers. The appropriate auxiliary patch is determined using normalized cross correlation in Fourier domain [76] using the original patch and the auxiliary image. The patch is further refined using a homography transformation estimated by enhanced correlation coefficient maximization [77] (Fig. 8b). Potential discrepancies caused by the changes of the environment between acquisitions of the two images, are further reduced using histogram matching. Once estimated, the texture is removed from the original patch to reduce the impact of the texture on the laser beam spots. A low-pass filter further reduces noise and effect of other artifacts (e.g., image compression), before detection using color thresholding (e.g., red color) in the HSV (Hue, Saturation, Value) color space (Fig. 8d). Pixels with low saturation values are discarded as hue can not be reliably computed. The remaining pixels are further filtered using mathematical morphology (opening operation). The final laser spots are selected by connected-component analysis (Fig. 8e).

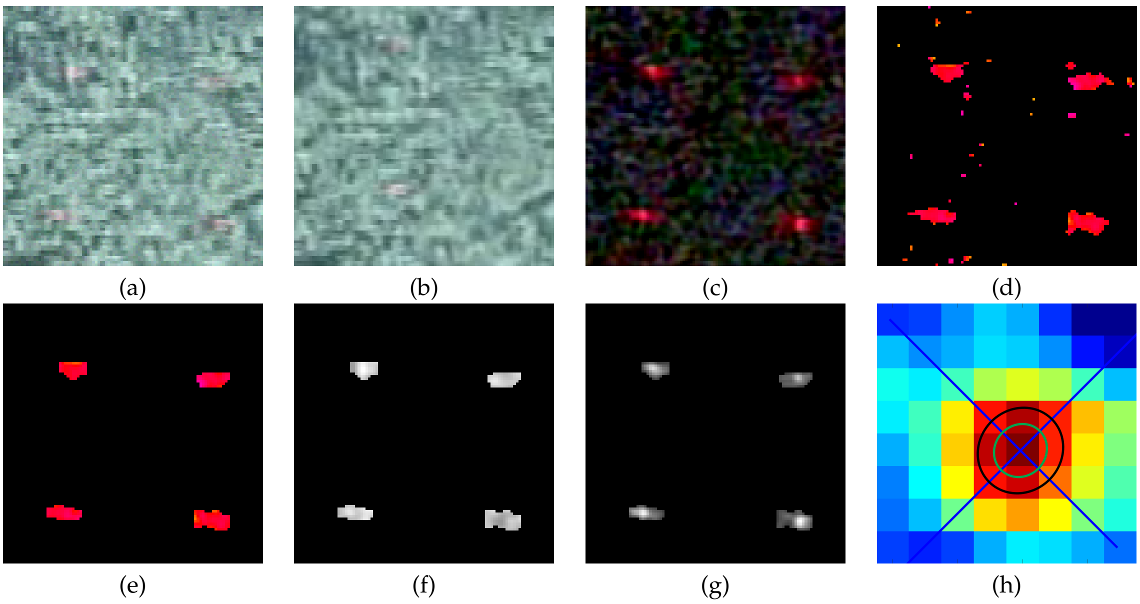


Figure 8. Laser spot detection: (**a**) predicted ROI of original image; (**b**) aligned auxiliary patch; (**c**) ROI after the removal of texture information (intensity x5); (**d**) potential laser pixels after color thresholding; (**e**) filtered laser pixels; (**f**,**g**) estimated laser beam luminosity without/with texture removal; (**h**) detected laser spot with detection uncertainty.

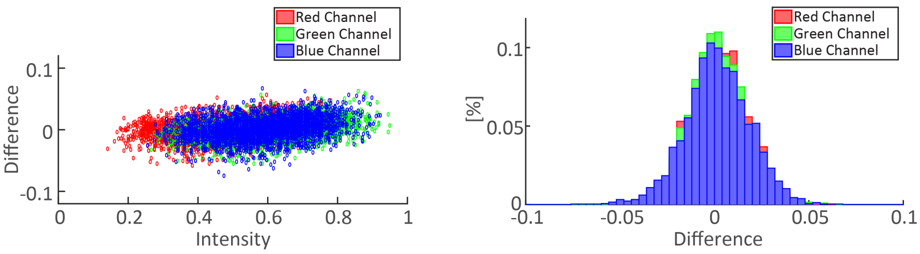
Peer-reviewed version available at *Remote Sens.* **2019**, *11*, 2093; doi:10.3390/rs11182093

Once the effects of the scene texture have been eliminated, the highest intensity point may be assigned to the laser beam center. In our procedure, the beam luminosity is characterized by the V channel of the HSV image representation. Figures 8f and 8g depict the estimate of the laser beam luminosity without and with the texture removal. Our proposed texture removal step clearly recovers the characteristic shape of the beam, with radially decreasing intensity from the center. Fitting a 2D Gaussian distribution to each laser spot allows us to estimate the center of the beam, assuming a 95% probability that the center falls within the top 20% of the luminance values (Fig. 8h).

4.2. Uncertainty

Given that the estimation of the laser center is based on color information, it is important to consider the effect the image noise. Depending on the particularities of the image set, image noise is the result of the combined effects of sensor noise, image compression and motion blur, among others. In our approach, the image noise is characterized by comparing the same area in two images taken a fraction of a second apart, where the sensed difference can be attributed to noise rather than an actual change in the environment. As shown in figure 9, the lack of correlation between image noise and pixel intensity levels or color channel supports our assumptions. Furthermore, the histogram of differences shows that noise can be well described by a Gaussian distribution.

For a final estimate of confidence levels of detection, we propagate the uncertainty of image intensities through the laser detection process using MC simulation. At each iteration we add noise independently to each pixel before the described laser spots detection. The iterations yield a set of independent detections, which are joined into a final laser spot detection represented by the sum of 2D Gaussians [78]. If the laser is not detected in $> 80\%$ of iterations, the detection is considered unstable and discarded. A set of laser spot detections obtained by a MC simulation is shown in figure 10 together with the final joined estimation. Red and green ellipses represent 66% and 95% confidence levels for independent detections, while blue and cyan indicate the final (combined) uncertainty.



(a) (b)

Figure 9. Characterization of image noise: (**a**) image noise vs. pixel intensity; (**b**) distribution of noise per color channel.

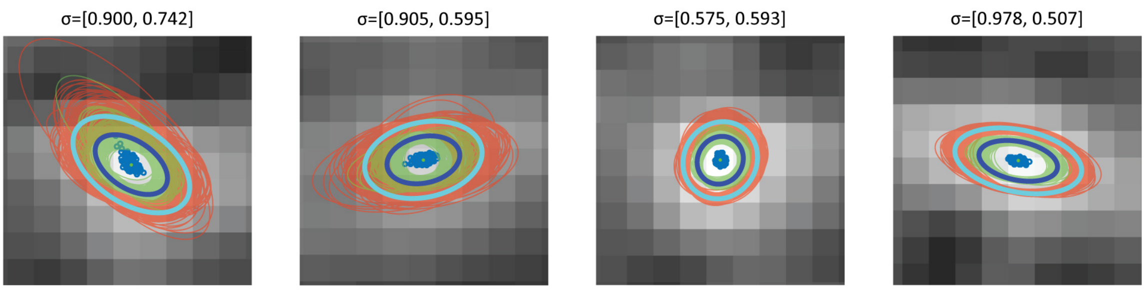
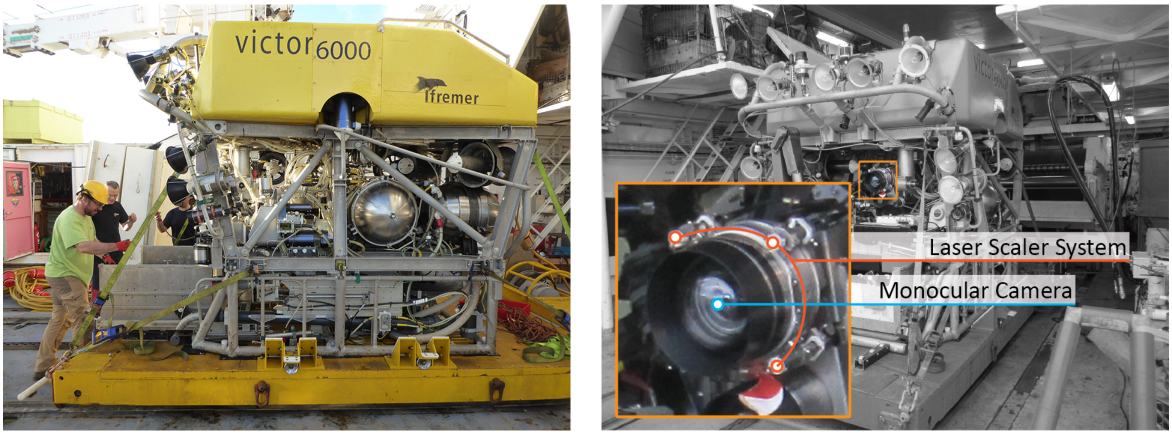


Figure 10. Examples of detected laser spot with uncertainty estimated through MC simulation for image shown in Fig. 8. Individual detections and uncertainties are depict with blue dots and red/green ellipses, while final uncertainty estimate is blue and cyan.

5. Dataset

During the SUBSAINTES 2017 cruise (doi: 10.17600/17001000) [44] an extensive seafloor imagery was acquired with the ROV VICTOR 6000 (IFREMER) [79]. The cruise targeted tectonic and volcanic features off Les Saintes Islands (French Antilles), at the same location as that of the model published in an earlier study [80], and derived from imagery of the ODEMAR cruise (doi: 10.17600/13030070). One of the main goals of this cruise was to study geological features associated with a recent earthquake, to measure the associated displacement along a fault rupture, while expanding a preliminary study that presented a first 3D model where this kind of measurements was performed [80]. To achieve this, the imagery was acquired at more than 30 different sites along $\sim$20 km, at the base of a submarine fault scarp. This is therefore one of the largest sets of image-derived underwater 3D models acquired with deep-sea vehicles to date.

The ROV recorded HD video with a monocular camera (Sony FCB-H11 camera with corrective optics and dome port) at 30Hz, and with a resolution of 1920×1080 (Fig. 11). Intrinsic camera parameters were determined using a standard calibration procedure [81] assuming a pinhole model with the 3rd degree radial distortion model. These camera parameters are kept constant through the entire acquisition process. Onboard navigation systems included a Doppler velocity log (Workhorse Navigator®), fibre-optic gyrocompass (OCTANS), depth sensor (Paroscientific Digiquartz®) and a long-range USBL acoustic positioning system (POSIDONIA®) with a nominal accuracy of about 1% of the depth. As the camera was positioned on a pan-and-tilt module lacking synchronization with the navigation data, only the ROV position can be reliably exploited.



(a) (b)

Figure 11. (**a**) ROV VICTOR 6000 (IFREMER) [79]. (**b**) Enlarged camera and laser system.

To date, 3D models at more than 30 geological outcrops throughout the SUBSAINTES study area have been built. Models vary in length between ~10 m and ~300 m horizontally, and extend vertically up to 30 m. Here we select two out of the 30 models (FPA and AUTT28), representative both of different survey patterns and spatial extents and complexity. Concurrently, evaluation data were collected with the same optical camera centered around a laser scaler consisting of 4 laser beams. For both selected datasets, numerous laser observations were collected, ensuring data spanning throughout the whole area. This enabled us to properly quantify the potential scale drifts within the models.

5.1. FPA

The first model (named FPA), extends laterally 33 m and 10 m vertically, and corresponds to a subvertical fault outcrop at a water depth of 1075 m. The associated imagery was acquired in a 10 min 51 s video recording during a single ROV dive (VICTOR dive 654). To fully survey the outcrop, the ROV conducted multiple passes over the same area. In total 218 images were selected and successfully processed to obtain the final model shown in Fig. 12.

Peer-reviewed version available at Remote Sens. 2019, 11, 2093; doi:10.3390/rs11182093

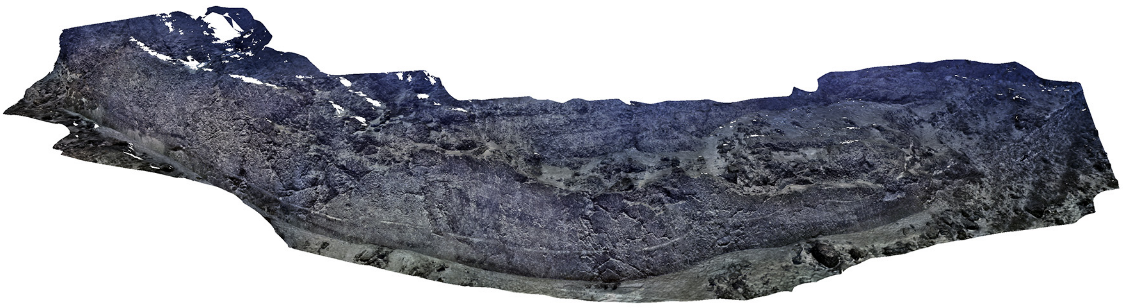


Figure 12. Textured 3D model of FPA area.

5.2. AUTT28

The second model (named AUTT28) is larger and required a more complex surveying scenario, as often encountered in real oceanographic cruises. Initially, the planned area of interest was recorded during VICTOR dive 654. Following a preliminary onboard analysis of the data, a vertical extension of the model was required, which was subsequently surveyed during VICTOR dive 658. This second survey also partially overlapped with the prior dive, with overlapping images acquired at a closer range and thus providing higher textural detail. The survey also included a long ROV pass with the camera nearly parallel to the vertical fault outcrop, an extremely undesirable imaging setup. This second 3D model is the largest constructed in this area, covering a sub-vertical fault scarp spanning over 300 m laterally and 10 m vertically, with an additional section of about 30 m in height from a vertical ROV travel. This model is thus well suited to evaluate scaling errors associated with drift as it includes several complexities (survey strategy and geometry, multiple dives, extensive length and size of the outcrop). After keyframe selection, 821 images were used out of a combined 1 h 28 min and 19 s of video imagery.

Figure 13. Textured 3D model of AUTT28 area.

5.2.1. Multi-Objective BA Weight Selection

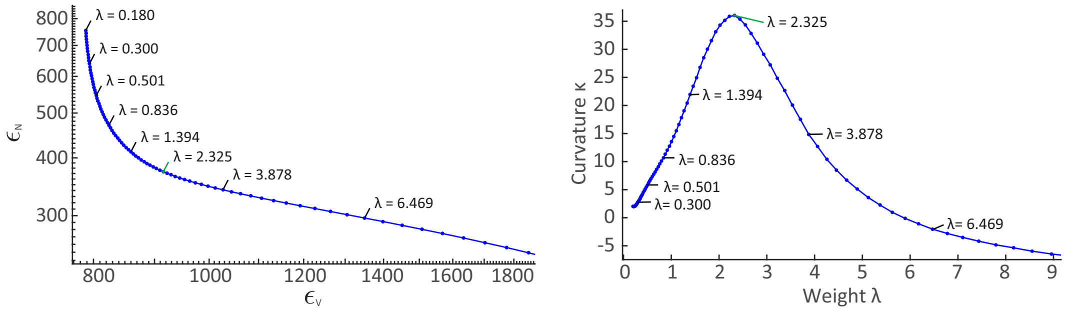
Models built with *a priori* navigation fusion through the multi-objective BA strategy require a weight selection which represents the ratio between re-projection and navigation fit errors. As uncertainties of the two quantities are in different units and, more importantly, not precisely known, this selection must be done either empirically or automatically. Due to the tedious and potentially ambiguous trial-and-error approach of empirical selection, the weight was determined using L-Curve analysis.

The curve, shown in figure 14a, uses the FPA dataset and 100 BA repetitions with weights λ spanning from 0.18 to 18. As predicted, the shape of the curve resembles an "L", with two dominant parts. The point of maximum curvature is determined to identify the weight with which neither objective has dominance (Fig. 14b). As noise levels of the camera and navigation sensors do not significantly change between the acquisition of different datasets, the same optimal weight $\boldsymbol{\lambda = 2.325}$ was used in all our multi-objective optimizations.

5.2.2. Multi-Survey Data

As is often the case in real cruise scenarios, the data for AUTT28 model was acquired in multiple dives (Fig. 15). When combining the data, it is important to consider the consequences of the merger.

Peer-reviewed version available at *Remote Sens.* **2019**, *11*, 2093; doi:10.3390/rs11182093



(a) (b)

Figure 14. (**a**) L-Curve for FPA dataset. (**b**) Curvature of L-Curve (shown on a smaller segment of weights for bigger clarity).

Optical imagery can be simply combined, given the short time period of time in between the two dives in which no significant differences are expected to happen in the scene. In contrast, the merging of navigation data may be challenging; ROV navigation is computed using smoothed USBL and pressure sensor data, with expected errors in acoustic positioning being approx. 1% of depth. As data was collected at roughly 1000 m depth, the expected nominal errors are of $\sim$10 m, or more in areas of poor acoustic conditions (e.g., close to vertical scarps casting acoustic shadows or reverberating acoustic pings). These errors, however, do not represent the relative uncertainty between nearby poses, but a general bias of the collected data for a given dive. While constant within each dive, the errors can differ between the dives over the same area, and are problematic when data from multiple dives are fused. Models built with data from a single dive will only be affected by a small error in geo-referencing, while multi-survey optimization may have to deal with contradicting navigation priors; images taken from identical positions would have different acoustic positions, with offsets in the order of several meters or higher.

This is overcome by introducing an additional parameter to be estimated, in the form of a 3D vector for each additional dive, representing the difference between USBL-induced offsets. Each vector is estimated simultaneously with the rest of the parameters in the SfM. For the case of AUTT28, the offset between the dives 654 and 658 was estimated to be (-2.53 m, 1.64 m, -0.02 m) in the x (E-W), y (N-S) and z (depth) directions, respectively. The disproportionately smaller z offset is due to the fact that the pressure sensor yields inter-dive discrepancies that are orders of magnitude smaller than the USBL positions.

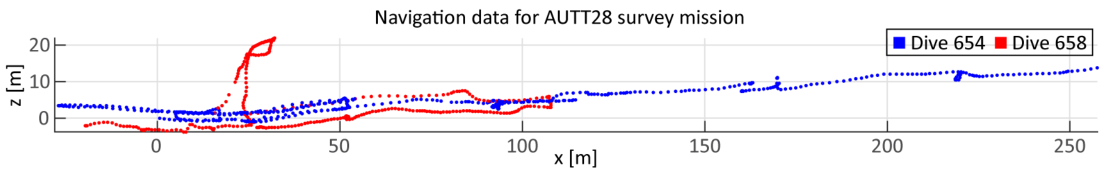


Figure 15. Navigation data for AUTT28 model merged from multiple dives (654 - blue; 658 - red).

5.3. Laser Calibration

The evaluation data was collected during multiple dives separated by days, and with camera and lasers being mounted and dismounted several times. While laser scaler mounting brackets ensured that the laser origins remained constant, the directions with respect to the camera changed slightly with each installation. Due to operational reasons, calibration data was not collected before each dive. However, the origins of the lasers remained fixed throughout the cruise, leaving as the only unknown in our setup the inter-dive variations in laser directions (relative to the camera and with respect to each other).

Peer-reviewed version available at *Remote Sens.* **2019**, *11*, 2093; doi:10.3390/rs11182093

In a normal calibration process, laser information is computed by fitting individual lines through sets of 3D points that are known to lay on the laser beams. Points are typically acquired by identifying laser intersections with a surface at a range of known distances. Given that in our case laser origins are known and fixed, we are only interested in individual laser directions. As these do not encapsulate scale information (as opposed to laser origins), the points used for individual line fittings do not necessarily have to be in metric scale, although they do have to be affected by the same scale factor.

In our case, the set of points lying along the laser beams is obtained from images with detected laser spots. Their 3D position can be determined from the camera pose, and a ray-casting procedure, as laser spots represent the projection of laser intersection with the scene onto the image. As our interest is solely in the laser directions, any model, regardless of its potential scale error can be used. However, it is important to avoid models with scale drift, or to use data from multiple models with different scales. Moreover, to maximize the conditioning of line fitting, the selection of a model with the widest depth range of such intersection points is important. This is the case for the AUTT28 model built using Global SfM and multi-objective BA, selected here. The global nature of the SfM and internal fusion of navigation data is predicted to most efficiently reduce a potential scale drift. As noisy laser detections are used to obtain the 3D points utilized in the calibration, laser spot uncertainties were propagated to obtain the associated uncertainty of the estimated laser direction. A MC simulation with a 1000 repetitions was used.

The evaluation data were collected on dives 653, 654 and 658. As no camera and laser dismounting/mounting occurred between dives 653 and 654, there are two distinct laser setups: one for dives 653 and 654 and one for dive 658. Figure 16 depicts all laser intersections with the scene (for both AUTT28 and FPA models), as well as the calibration results, projected onto an image plane. Intersections detected in 3D model AUTT28 are depicted in black, while those from 3D model FPA are shown in orange. Similarly, the squares and circles represent dives 653/654 and dive 658, respectively. The projections of the final laser beam estimations are presented as solid and dotted lines. The figure shows a good fit of estimated laser beams with the projections of the intersections, both in the AUTT28 and FPA models. The adequate fit to the vast majority of AUTT28 points shows that the model used in the calibration had no significant scale drift. Furthermore, the fitting of the FPA related points, which were not used in the calibration and are affected by a different scale factor, confirms that calibration of laser directions is independent of the 3D model used, and of different scalings. The broad spread of the black points relative to the orange ones also confirms that the choice of the AUTT28 over the FPA model was adequate for this analysis. Lastly, it is worth reiterating that it was not possible to combine the data from all the models for calibration, as they are affected by a different scale factors.

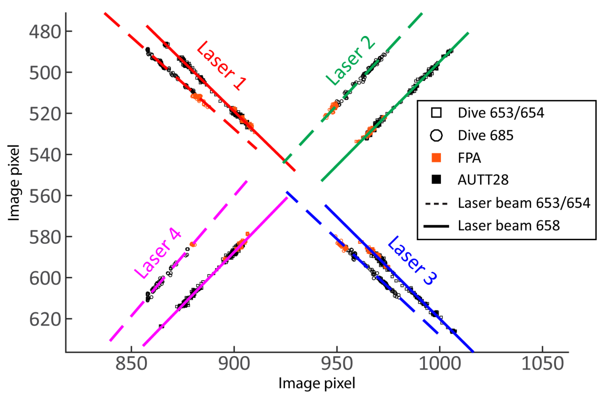


Figure 16. Calibration results for dives 653/654 and 658. Solid and dotted lines represent the projections of estimate laser beams on the image plane, while projected laser intersections with the scene are depicts as squares/circles.

Peer-reviewed version available at *Remote Sens.* **2019**, *11*, 2093; doi:10.3390/rs11182093

6. Results

As introduced above, the evaluation of the scale accuracy was performed for four different optical-based 3D reconstruction strategies: *A*) Incremental SfM with *a posteriori* navigation fusion; *B*) Global SfM with *a posteriori* navigation fusion; *C*) Incremental SfM with multi-objective BA navigation fusion; *D*) Global SfM with multi-objective BA navigation fusion. The models for each of the two datasets (FPA and AUTT28) were built using each of the four strategies, and subsequently evaluated on multiple segments spread across the observed area.

Using the model evaluation framework and laser spot detection method presented above, the scale accuracy and its associated uncertainties were automatically estimated using more than 550 images. To minimize the effects of possible false laser spot detections, only images with at least two confidently detected laser points were used. Furthermore, any images exhibiting excessive variation of the estimated scale between the individual lasers were discarded, as scale can be assumed to be locally constant.

6.1. Scale accuracy estimation

During accuracy evaluation, the scale error ϵ_s is estimated for each image independently. The final per-image scale error and its uncertainty are estimated through a MC simulation, with input variables (features, laser spot locations and laser calibration) sampled according to their probability distributions. The repeated computation with noisy data thus results in an equal number of final scale error estimates per laser. Figure 17 shows one example of such estimation, together with the selected intermediate results of the evaluation process. As each MC iteration encapsulates the complete evaluation process (image localization, ray-casting, origin estimation and scale error evaluation), intermediate distributions presented in Fig. 17 are only shown for illustration, and are not used as distribution assumptions in the process itself.

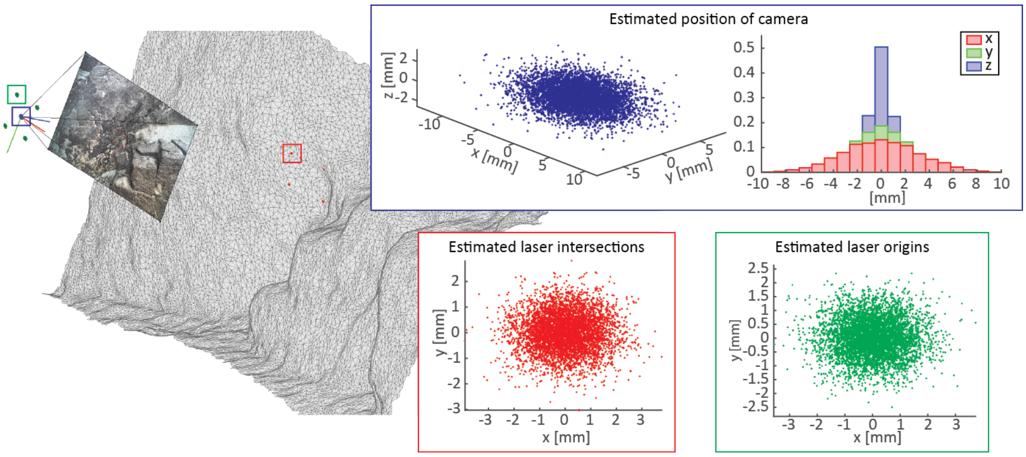


Figure 17. Intermediate results of a scale estimation procedure.

To satisfactorily represent the complexity of the process, 5000 iterations were used for each estimation. Figure 18 shows the evolution of the estimated scale error with associated uncertainty under increasing number of samples. After 500 iterations, the errors exhibit only minor fluctuations, and after 1500 iterations there is no noticeable difference. Hence, our selection of 5000 iterations is more than adequate to encapsulate the distribution of noise.

To show the advantages of our fully-calibrated approach compared to previously available methods or our partially-calibrated method, scale estimates obtained for each laser/laser pair are compared. Given the non-alignment of lasers with the optical axis of the camera, the majority of previous image-scaling methods (e.g. 24,25) are not applicable. The only available option is thus a simplified approach where the Euclidean distance between a pair of 3D points (laser intersection points with the scene) is assumed to be the actual distance between the laser pair.

Peer-reviewed version available at *Remote Sens.* **2019**, *11*, 2093; doi:10.3390/rs11182093

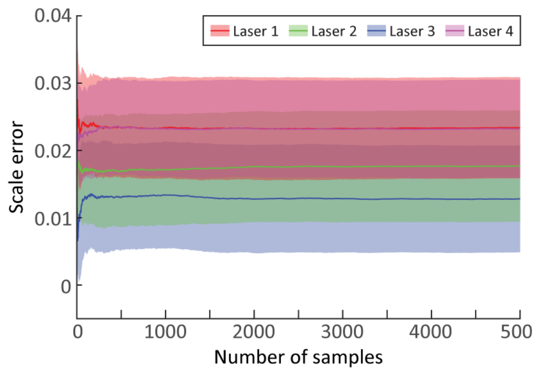


Figure 18. Evolution of scale error estimate with increasing number of MC iterations.

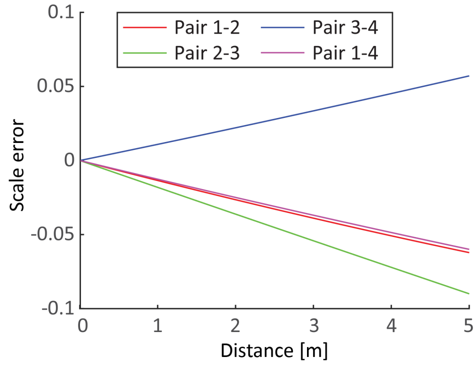


Figure 19. Error induced in scale error estimate due to disregarding non-parallelism of laser beams.

Results using different lasers (Fig. 20) show that the FCM method produces the most consistent results. This is expected as the estimation process considers both individual laser directions and the geometry of the scene. The effect of scene geometry is clear when Figs. 20a and 20b are compared. The slightly slanted angle together with the uneven geometry of the scene causes a large variation in the scale error estimates by the individual laser pairs. Similarly, the comparison of Figs. 20b and 20c shows the effect of inaccurate assumption of laser parallelism. This error depends on the camera-scene distance as shown in Fig. 19. It is evident that the overestimation of laser pair 3-4 and underestimation of other laser pairs can be explained by the use of oversimplified laser geometry. To validate this assumption, the results of the partially-calibrated method were corrected by the expected errors (at $d = 2m$) induced by disregarding non-parallelism of laser beams (Fig. 20d). While the result is nearly identical to that from a FCM method (Fig. 20c), we note that the scale error in Fig. 20c is computed for each laser individually, while the partially-calibrated method considers laser pairs instead, and hence minor discrepancies.

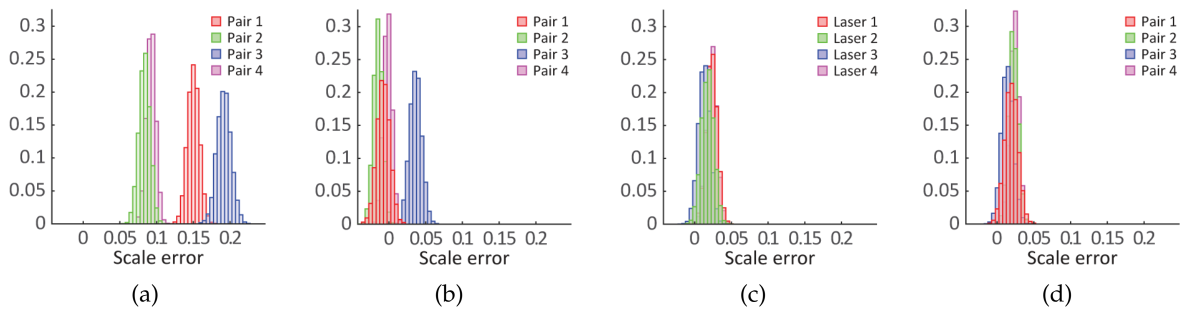


Figure 20. Estimated scale error per laser using different methods of computation: (**a**) Simplistic; (**b**) partially-calibrated method; (**c**) fully-calibrated method; (**d**) partially-calibrated method corrected for errors induced by non-parallelism of laser beams.

6.2. FPA

The accuracy of the FPA model was analyzed using 148 images (432 lasers). To represent the results concisely, measurements are grouped into 7 segments based on their model position (Fig. 21 and Table 2). To ensure that the scale of the model did not vary within each segment, the maximum distance of any laser detection to the assigned segment center was set to 1 m.

FPA covers a relatively small area, imaged with multiple passes providing redundancy that promotes model accuracy. It is thus expected to have only minor variations in scale error between areas. Figure 22 depicts the distribution of estimated scale errors for all four methods of 3D model

Peer-reviewed version available at Remote Sens. 2019, 11, 2093; doi:10.3390/rs11182093

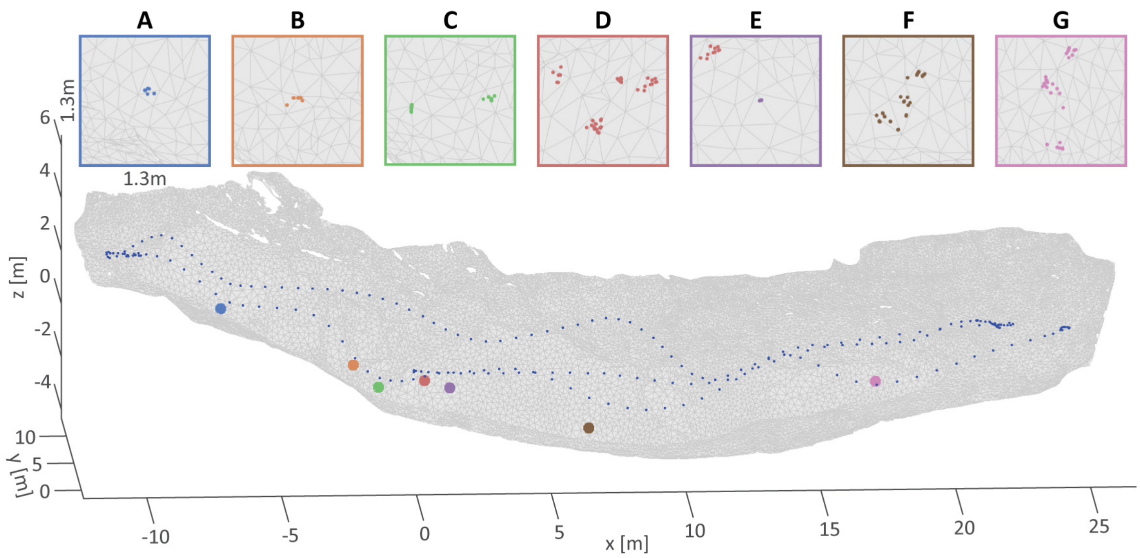


Figure 21. 3D reconstruction with the distribution of laser observations per segment for FPA area.

Table 1. Distribution of laser observations per segment for FPA area.

	A	B	C	D	E	F	G
# Images	12	6	39	40	12	24	15
# Lasers	29	12	137	103	33	53	36
Laser distance (min/max) [m]	3.23/3.29	4.44/4.46	3.03/3.50	3.58/4.01	3.59/3.61	3.19/3.37	3.19/3.79

construction. The comparison of results shows that accuracy does not significantly differ. The scale error varies between -1% and -5% with estimated uncertainties of around $\pm 3\%$. The highest errors occur at the borders of the model. As expected, uncertainty is closely related to the camera-scene distance, as small uncertainties in the laser direction translate to larger discrepancies at larger distances.

Table 2. Estimated scale errors (%) per segment for different reconstructions of FPA area (values represent mean value with standard deviation).

	A	B	C	D	E	F	G
Global SfM w/ Similarity T.	-3.6 ± 2.9	0.9 ± 3.2	-1.1 ± 1.9	-1.2 ± 3.4	-1.4 ± 2.8	-4.0 ± 2.9	-4.0 ± 3.4
Incremental SfM w/ Similarity T.	-3.6 ± 2.9	0.9 ± 3.2	-1.0 ± 1.9	-1.2 ± 3.4	-1.4 ± 2.8	-4.0 ± 2.9	-4.0 ± 3.5
Global SfM w/ multi-objective BA	-4.7 ± 2.8	0.7 ± 3.2	-1.3 ± 1.9	-1.2 ± 3.4	-1.4 ± 2.8	-3.1 ± 2.9	-2.2 ± 3.5
Incremental SfM w/ multi-objective BA	-4.7 ± 2.9	0.7 ± 3.2	-1.3 ± 1.9	-1.2 ± 3.4	-1.4 ± 2.8	-3.2 ± 2.9	-2.1 ± 3.5

6.3. AUTT28

For model AUTT28, the evaluation data (images containing projected laser spots) were gathered during VICTOR dives 654 and 658, after the video acquisition of data used for 3D model creation. A total of 432 images with 1378 laser measurements were selected and grouped in 6 distinct sections throughout the 3D model, as shown in Table 3 and Fig. 23. Dive 654 covered a longer vertical path (blue dots), while dive 658 (red dots) surveyed an additional horizontal segment together with parts of the area already viewed using dive 654. The higher density of red points indicates that the ROV observed the scene at a closer range during dive 658, requiring a higher number of images to obtain the necessary overlap compared to dive 654.

The comparison of results shows that the models built using *a posteriori* navigation fusion (Figs. 24a and 24b) are significantly impacted by scale drift ($\sim 15\%$), and that this impact is nearly identical regardless of the use of global or incremental SfM approaches. The gradual scale sliding observed is caused by inherit scale ambiguity of the two-view image pair geometry when BA is solely dependent

Peer-reviewed version available at Remote Sens. 2019, 11, 2093; doi:10.3390/rs11182093

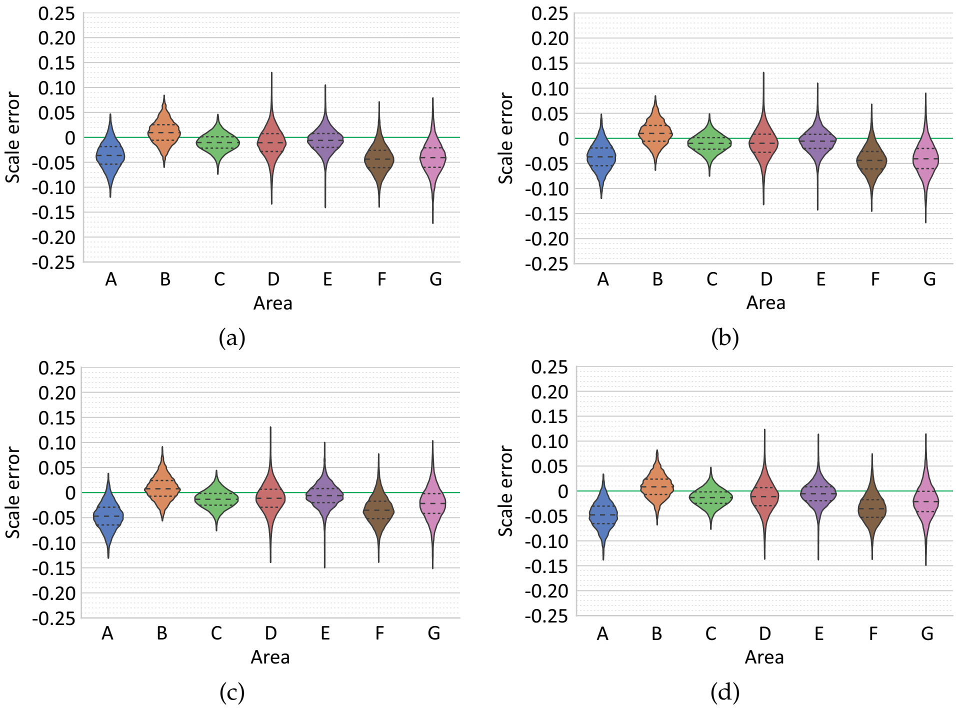


Figure 22. Estimated scale errors per segment for model FPA: (**a**/**b**) Global / Incremental SfM with similarity transformation navigation fusion; (**c**/**d**) Global / Incremental SfM with multi-objective BA navigation fusion.

Table 3. Distribution of laser observations per segment for AUTT28 area.

	A	B	C	D	E	F
# Images	30	47	20	46	261	28
# Lasers	97	169	51	165	812	84
Laser distance (min/max) [m]	1.95/2.25	2.13/2.67	2.90/3.36	3.21/3.89	1.70/4.14	3.63/3.79

on visual information. While this might not have been as obvious in the previous case, the long single pass of the camera, as performed in dive 654, introduces in this particular model numerous consecutive two-view image pairs, magnifying the scale drift. As shown in Figs. 24c and 24d, additional constraints in the BA (e.g., navigation data) reduce ambiguity and, ultimately, nearly eliminate scale drift. Overall, scale error of the model built with global SfM using multi-objective BA is less than 1% with nearly zero scale drift, while a model built with incremental SfM approach showed a 2% scale drift along its 300 m length. It is worth noting that the observed difference in scale estimates are within the uncertainty levels of the estimations, and therefore inconclusive.

Table 4. Estimated scale errors (%) per segment for different reconstructions of AUTT28 area (values represent mean value with standard deviation).

	A	B	C	D	E	F
Global SfM w/ Similarity T.	-6.4 ± 2.3	-6.3 ± 1.9	-4.5 ± 2.4	-1.1 ± 2.0	2.1 ± 2.2	9.2 ± 2.9
Incremental SfM w/ Similarity T.	-6.3 ± 2.3	-6.1 ± 1.9	-4.1 ± 2.5	-0.8 ± 2.0	2.3 ± 2.3	9.3 ± 2.8
Global SfM w/ multi-objective BA	0.7 ± 2.4	0.7 ± 2.0	0.8 ± 2.6	-0.2 ± 2.0	1.7 ± 2.3	0.9 ± 2.6
Incremental SfM w/ multi-objective BA	-0.6 ± 2.4	-0.1 ± 2.1	1.2 ± 2.6	2.0 ± 2.2	1.7 ± 2.3	0.6 ± 2.7

Peer-reviewed version available at Remote Sens. 2019, 11, 2093; doi:10.3390/rs11182093

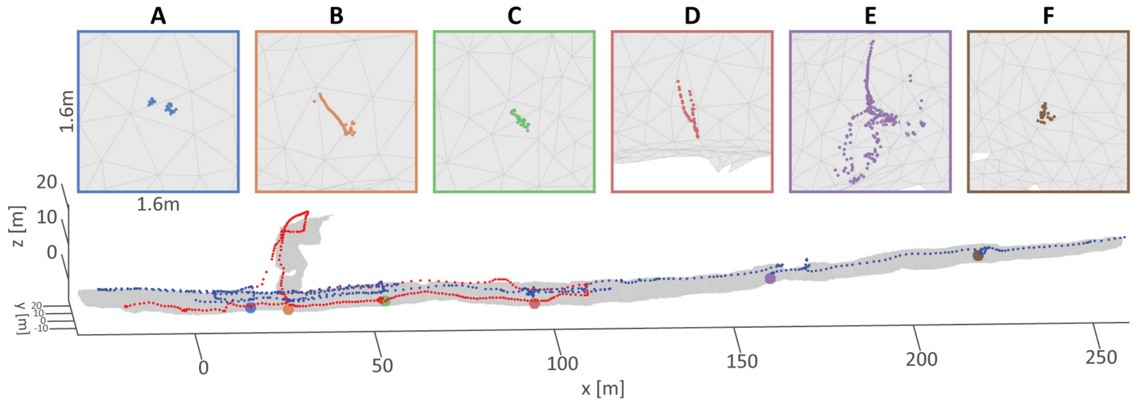


Figure 23. 3D reconstruction with the distribution of laser observations per segment for AUTT28 area. Red and blue dots correspond to VICTOR dives 654 and 658.

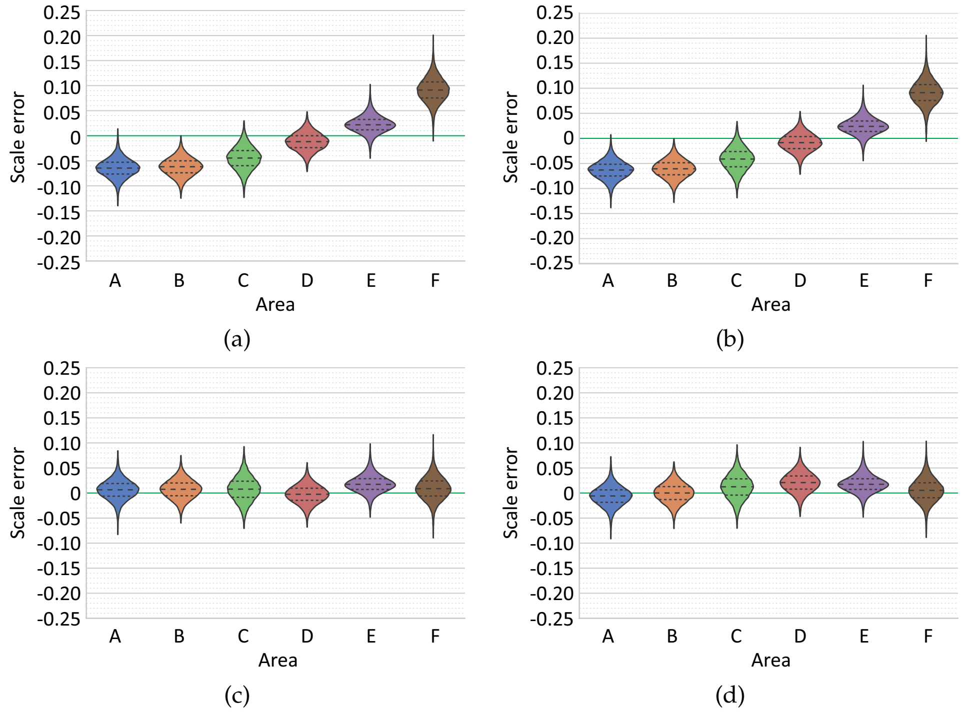


Figure 24. Estimated scale errors per segment for model AUTT28: (**a**/**b**) Global / Incremental SfM with similarity transformation navigation fusion. (**c**/**d**) Global / Incremental SfM with multi-objective BA navigation fusion.

6.3.1. Multi-Objective BA vs Similarity Transformation navigation fusion

The effects of different navigation fusion strategies are demonstrated through the comparison of two reconstructions obtained using Global SfM with multi-objective BA and with similarity transformation (Fig. 25). The reconstructions diverge on the outer parts of the model, consistent with a "doming" effect. A broad-scale systematic deformation produces a reconstruction that appears as a rounded-vault-distortion of a flat surface. This effect is a result of a rigorous re-projection error minimization of a loosely interconnected longer sequence of images taken from a nearly parallel direction combined with slight inaccuracies in modelling of the radial distortion of the camera [14]. As for scale drift, additional non-vision related constraints can reduce this distortion and the associated error.

Peer-reviewed version available at Remote Sens. 2019, 11, 2093; doi:10.3390/rs11182093

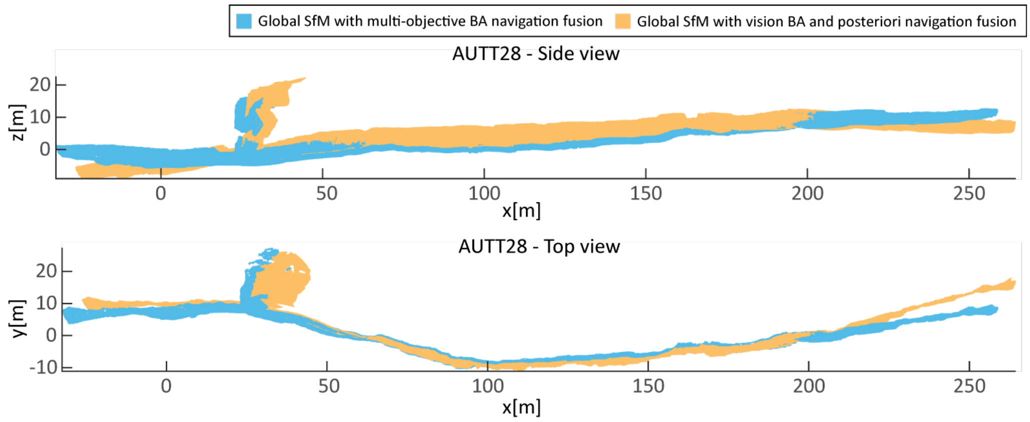


Figure 25. Comparison of navigation fusion strategies on the reconstruction of 3D models.

6.3.2. Multi-Survery data fusion

As explained in section 5.2.2, the multi-mission data fusion can cause contradictory navigation priors during optimization. We address this by expanding the optimization problem with an additional 3D vector, representing the possible USBL offset between the recorded navigation data of the two dives. To examine the effects of this offset compensation on model construction, an additional model was constructed using raw navigation data (i.e., without offset compensation). Figure 26 depicts errors in the camera pose estimates with respect to their navigation priors, and show a concentration of errors in areas imaged during both dives (Fig. 15), where navigation priors of the two dives are incoherent. The errors dramatically decrease with the introduction of an offset, yielding an improved fitting solution. Alternatively, incoherences can cause model distortions to compensate for contradicting priors, as shown by abrupt changes of scale (area D in Fig. 27).

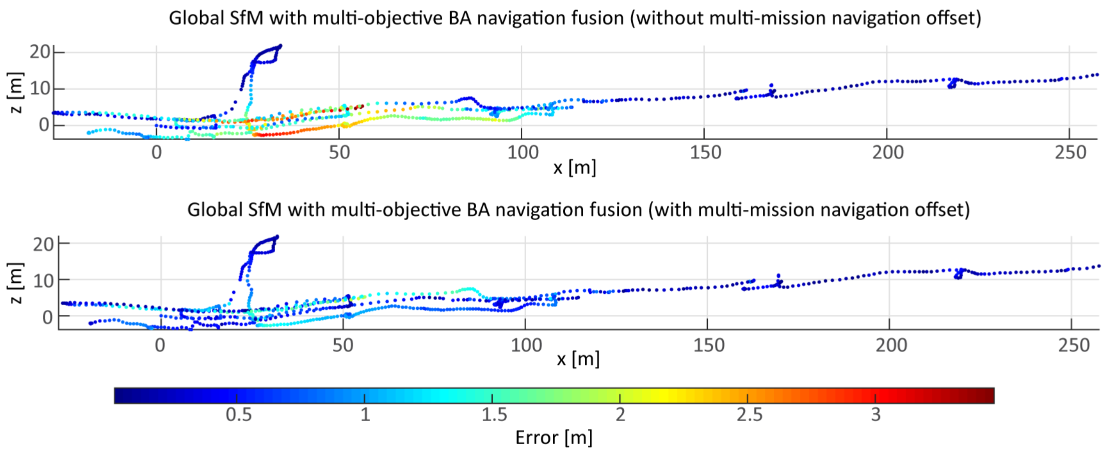


Figure 26. Comparison of multi-survey data fusion strategies on the estimated camera path.

6.3.3. Scale error estimation methods

To recover high-resolution and precise information from 3D models (lengths, areas, volumes) it is important to use the most accurate method. As the non-alignment of lasers with the optical axis of the camera prevents the use of previous image-scaling methods (e.g., Pilgrim et al. [24], Davis and Tusting [25]), two other methods could be used instead. Minor misalignments of laser scalers may be discarded for simplicity or lack of sufficiently distributed calibration data. In such case both, our partially-calibrated approach and simplified direct 3D method, that assumes an equivalence of the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 July 2019

Peer-reviewed version available at Remote Sens. 2019, 11, 2093; doi:10.3390/rs11182093

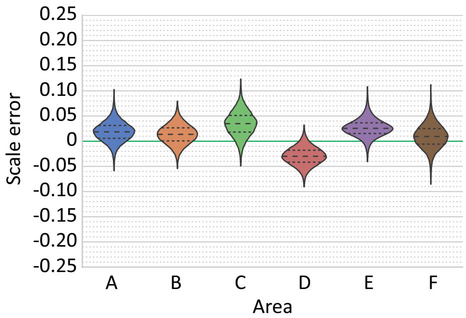
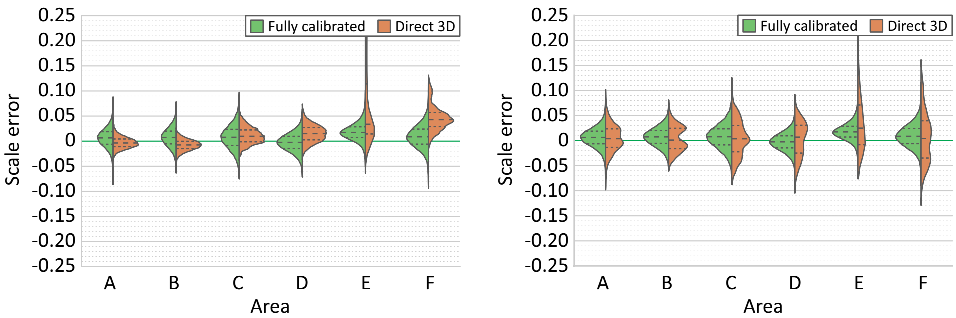


Figure 27. Estimated scale errors per segment for AUTT28 model built with Global SfM with multi-objective BA navigation fusion without an additional offset vector.

Euclidean distance between the points of laser intersections and the beams themselves, could be used for the evaluation.

For this comparison the model with least scale drift was selected (Global SfM and multi-objective BA navigation fusion) to emphasize the effects of different methods on the results. Furthermore, as the simplistic direct 3D and partially-calibrated method assume laser-pair parallelism, the analysis of these two methods was performed on data consisting of only laser pairs that were the closest to being parallel (Figs. 28a and 29a), as well as on the complete dataset (Figs. 28b and 29b), to show the effect that non-parallelism of laser beams may have on the different methods.

As expected, in comparison to the simplistic approach (orange) (Fig. 28a), our method (green) is less impacted by the range of camera-scene angles and distances, whereas the spread of the estimated values within each segment for the direct approach correlates directly with the span of camera-scene distances. Although varying distances themselves do not play a role, they do however increase the probability of both having different camera-surface angles, and of violating the surface flatness requirement.

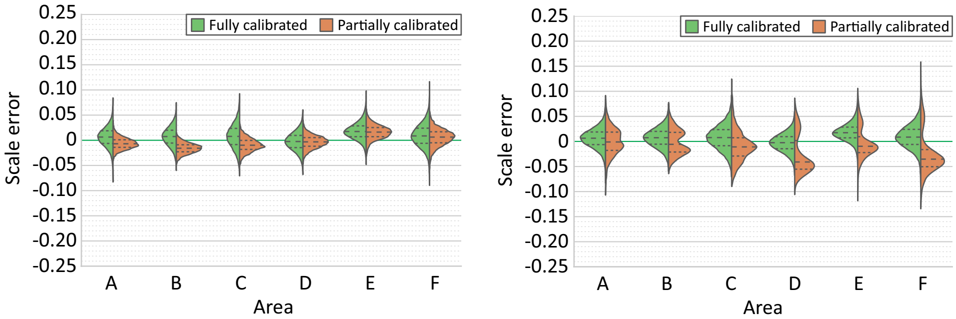


(a) (b)

Figure 28. Comparison of estimated scale errors computed with fully-calibrated and simplistic direct 3D method using: (**a**) only nearly-parallel laser pairs; (**b**) all laser pairs.

In contrast, the analysis of the results of the partially-calibrated approach (Fig. 29a) confirms that this method is unaffected by changes of camera angle and scene roughness. As expected, the results in sections D, E and F are nearly identical, with discrepancies in sections A, B and C. Sections A, B, and C were evaluated using data collected during dive 658, while D, E, F in dive 654, and we attribute this discrepancy to the marginally larger error in non-parallelism of the laser configuration used during dive 658 than that of dive 654. This is clearly shown when the results are computed on the data from all laser pairs (Fig. 29b), as non-parallelism of different laser pairs causes significant variation in the results. Segments acquired at closer ranges (A, B and C), and therefore less affected by the errors in parallelism, have smaller errors than those of segments D and F, which are evaluated at a larger distances. While similar multi-modal distributions appear in the results of the simple direct

3D method, the clear multi-modal peaks are suppressed by the effects of camera-surface angles and roughness of the surface model.



(a) (b)

Figure 29. Comparison of estimated scale errors computed with fully- and partially-calibrated method using: (**a**) only nearly-parallel laser pairs; (**b**) all laser pairs.

7. Conclusions

In this study the scale error evaluation of four most commonly used optical-based 3D reconstruction strategies of underwater scenes is presented. This evaluation seeks to determine the advantages and limitations of the different methods, and to provide a quantitative estimate of model scaling and the precision of measurements performed on them for quantitative studies (distances, areas, volumes, etc.). The analysis was performed on two data sets acquired during a scientific cruise (SUBSAINTES 2017) with a scientific ROV (VICTOR6000), and therefore under realistic deep-sea fieldwork conditions. For models built using multi-objective BA navigation fusion strategy, an L-Curve analysis was performed to determine the optimal weight between competing objectives of the optimization. Furthermore, the potential offset in navigation when using USBL-based positioning from different dives was addressed in a representative experiment.

Building upon our previous work, the lack of known measurements readily available in large scale models was overcome with the fully-calibrated method, which exploits laser projections onto the scene from laser scalers, which are common in deep-sea ROVs. The confidence level for each of the scale error estimates was independently assessed with a propagation of the uncertainties associated with image features and laser spot detections using a Monte Carlo simulation. The number of iterations used in the simulation to satisfactorily represent the complexity of the process was validated through the analysis of the final estimate behaviour. The comparison of the results show that the fully-calibrated method is more consistent and accurate than the two other plausible approaches, i.e. partially-calibrated and simplistic direct 3D method. We also note that by limiting the data to parallel laser pairs (dive 654), the partially-calibrated method produced similar results. Therefore, the PCM approach can be used when the relation between parallel lasers and the camera is not known. This opens its use in numerous scenarios where strict rigidity between the camera and lasers is not maintained or determined (e.g., legacy data).

As each scale error estimate characterizes an error at a specific area of the model, independent evaluations across the models enable efficient determination of potential scale drifts. To obtain a sufficient number of accurate laser measurements, an automatic laser spot detector was also developed. By mitigating the effects of scene texture, a much larger amount of accurate detections was possible, even with greatly attenuated laser beams. Furthermore, the recovery of characteristic shapes of laser spots with radially decreasing intensities enabled additional determination of the uncertainty of laser spot detections. In total, the scale errors have been evaluated on a large set of measurements in both models (432/1378) spread across them.

The effects of different reconstruction strategies were analyzed using two distinct survey scenarios. The first model (FPA dataset) was acquired with multiple passes over the same areas. Overlap of non-sequential images restricted the potential solution of the optimization problem to a nearly identical solution regardless of the strategy (SfM or navigation fusion). In a second model (AUTT28 dataset), data were acquired during two separate surveys, and includes a long single pass with the camera oriented nearly parallel to the vertical wall. The results demonstrate that surveys with weakly connected sequentially acquired images are prone to produce broad-scale deformation (doming effect) in the final model. Rigorous minimization of the re-projection error, combined with the projective scale ambiguity, bends the model, and can further lead to drift in the scale estimate. While navigation fusion strategy did not play a role in the first model (FPA), the results of this second model (AUTT28) demonstrate the advantage of using multi-objective BA navigation fusion to process data with more complex survey patterns. Furthermore, the introduction of additional vectors in the optimization of multi-survey problems successfully accounted for offset changes present in the underwater USBL-based navigation data and thus minimize the effect of contradicting navigation priors.

Finally, in surveys over a single dive and with multiple overlapping regions, the reconstruction strategy is to a first order irrelevant, while more complex scenarios significantly benefit from optimization including the navigation data. In all cases, the errors in the models are inferior to 5%, with errors often being around 1%.

Acquisition of calibration data (points collected at large range of distances) is also critical. Depending on laser setup, a modification of laser geometry is possible (e.g., during the process of diving due to pressure changes). As minor discrepancies in parallelism can cause significant offsets at the evaluating distance, to perform a calibration in the field is desirable (e.g., approach of the scene illuminated with laser beams). Furthermore, our results also indicate that it is important to collect multitude of evaluation data at different locations and moments during the survey.

Author Contributions: Conceptualization, K.I., N.G., J.E. and R.G.; Data curation, K.I. and A.A.; Funding acquisition, J.E. and R.G.; Investigation, K.I.; Methodology, K.I., N.G. and R.G.; Software, K.I. and A.A.; Supervision, N.G., J.E. and R.G.; Writing–original draft, K.I.; Writing–review & editing, N.G., J.E. and R.G.

Funding: Partial funding was provided by the European Union's Horizon 2020 project ROBUST (grant agreement 690416-H2020-CS5-2015-onestage) (K.I.), project Eurofleets Plus (grant agreement 824077), the Spanish Ministry of Education, Culture and Sport under project CTM2017-83075-R (N.G. and R.G.), the ANR SERSURF Project (ANR-17-CE31-0020, France) (J.E. and A.A.), and the Institut de Physique du Globe de Paris (J.E.). The cruise, ship and ROV time were funded by the French Ministry of Research

Acknowledgments: This study is based on data acquired during the SUBSAINTES 2017 cruise, that deployed the ROV VICTOR 6000 (IFREMER, France) for image acquisition. We commend the work of the crew, officers, and engineers that participated on SUBSAINTES 2017 cruise, making possible the acquisition of data.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lucieer, A.; de Jong, S.M.; Turner, D. Mapping landslide displacements using Structure from Motion (SfM) and image correlation of multi-temporal UAV photography. *Progress in Physical Geography: Earth and Environment* **2014**, *38*, 97–116. doi:10.1177/0309133313515293.
2. Javernick, L.; Brasington, J.; Caruso, B. Modeling the topography of shallow braided rivers using Structure-from-Motion photogrammetry. *Geomorphology* **2014**, *213*, 166–182. doi:10.1016/j.geomorph.2014.01.006.
3. Marteau, B.; Vericat, D.; Gibbins, C.; Batalla, R.J.; Green, D.R. Application of Structure-from-Motion photogrammetry to river restoration. *Earth Surface Processes and Landforms* **2017**, *42*, 503–515. doi:10.1002/esp.4086.
4. Mathews, A.J.; Jensen, J.L.R. Visualizing and Quantifying Vineyard Canopy LAI Using an Unmanned Aerial Vehicle (UAV) Collected High Density Structure from Motion Point Cloud. *Remote Sensing* **2013**, *5*, 2164–2183. doi:10.3390/rs5052164.
5. Campos, R.; Garcia, R.; Alliez, P.; Yvinec, M. A surface reconstruction method for in-detail underwater 3D optical mapping. *The International Journal of Robotics Research* **2014**, pp. 64–89. doi:10.1177/0278364914544531.

Peer-reviewed version available at *Remote Sens.* **2019**, *11*, 2093; doi:10.3390/rs11182093

6. Pizarro, O.; Friedman, A.; Bryson, M.; Williams, S.B.; Madin, J. A simple, fast, and repeatable survey method for underwater visual 3D benthic mapping and monitoring. *Ecology and Evolution* **2017**, *7*, 1770–1782. doi:10.1002/ece3.2701.
7. Bingham, B.; Foley, B.; Singh, H.; Camilli, R.; Delaporta, K.; Eustice, R.; Mallios, A.; Mindell, D.; Roman, C.; Sakellariou, D. Robotic tools for deep water archaeology: Surveying an ancient shipwreck with an autonomous underwater vehicle. *Journal of Field Robotics* **2010**, *27*, 702–717. doi:10.1002/rob.20350.
8. Rossi, P.; Castagnetti, C.; Capra, A.; Brooks, A.; Mancini, F. Detecting change in coral reef 3D structure using underwater photogrammetry: critical issues and performance metrics. *Applied Geomatics* **2019**, pp. 1–15. doi:10.1007/s12518-019-00263-w.
9. Escartín, J.; Leclerc, F.; Olive, J.A.; Mevel, C.; Cannat, M.; Petersen, S.; Augustin, N.; Feuillet, N.; Deplus, C.; Bezos, A.; others. First direct observation of coseismic slip and seafloor rupture along a submarine normal fault and implications for fault slip history. *Earth and Planetary Science Letters* **2016**, *450*, 96–107. doi:10.1016/j.epsl.2016.06.024.
10. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*, 2 ed.; Cambridge University Press: New York, NY, USA, 2003. doi:10.1017/CBO9780511811685.001.
11. Soloviev, A.; Venable, D. Integration of GPS and vision measurements for navigation in GPS challenged environments. IEEE/ION Position, Location and Navigation Symposium, 2010, pp. 826–833. doi:10.1109/PLANS.2010.5507322.
12. Mian, O.; Lutes, J.; Lipa, G.; Hutton, J.; Gavelle, E.; Borghini, S. Accuracy assessment of direct georeferencing for photogrammetric applications on small unmanned aerial platforms. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* **2016**, *40*, 77. doi:10.5194/isprs-archives-XL-3-W4-77-2016.
13. Forlani, G.; Dall'Asta, E.; Diotri, F.; Cella, U.M.d.; Roncella, R.; Santise, M. Quality Assessment of DSMs Produced from UAV Flights Georeferenced with On-Board RTK Positioning. *Remote Sensing* **2018**, *10*. doi:10.3390/rs10020311.
14. James, M.R.; Robson, S. Mitigating systematic error in topographic models derived from UAV and ground-based image networks. *Earth Surface Processes and Landforms* **2014**, *39*, 1413–1420. doi:10.1002/esp.3609.
15. Eltner, A.; Schneider, D. Analysis of Different Methods for 3D Reconstruction of Natural Surfaces from Parallel-Axes UAV Images. *The Photogrammetric Record* **2015**, *30*, 279–299. doi:10.1111/phor.12115.
16. Mertes, J.; Zant, C.; Gulley, J.; Thomsen, T. Rapid, quantitative assessment of submerged cultural resource degradation using repeat video surveys and Structure from Motion. *Journal of Maritime Archaeology* **2017**, *12*, 91–107. doi:10.1007/s11457-017-9172-0.
17. Sedlazeck, A.; Koser, K.; Koch, R. 3D reconstruction based on underwater video from ROV Kiel 6000 considering underwater imaging conditions. OCEANS 2009-EUROPE, 2009, pp. 1–10. doi:10.1109/OCEANSE.2009.5278305.
18. Pizarro, O.; Eustice, R.M.; Singh, H. Large Area 3-D Reconstructions From Underwater Optical Surveys. *IEEE Journal of Oceanic Engineering* **2009**, *34*, 150–169. doi:10.1109/JOE.2009.2016071.
19. Campos, R.; Gracias, N.; Ridao, P. Underwater Multi-Vehicle Trajectory Alignment and Mapping Using Acoustic and Optical Constraints. *Sensors* **2016**, *16*, 387. doi:10.3390/s16030387.
20. Garcia, R.; Campos, R.; Escartín, J. High-resolution 3D reconstruction of the seafloor for environmental monitoring and modelling. Proc. Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on, 2011.
21. Cocito, S.; Sgorbini, S.; Peirano, A.; Valle, M. 3-D reconstruction of biological objects using underwater video technique and image processing. *Journal of Experimental Marine Biology and Ecology* **2003**, *297*, 57–70. doi:10.1016/S0022-0981(03)00369-1.
22. Kalacska, M.; Lucanus, O.; Sousa, L.; Vieira, T.; Arroyo-Mora, J. Freshwater fish habitat complexity mapping using above and underwater structure-from-motion photogrammetry. *Remote Sensing* **2018**, *10*, 1912. doi:10.3390/rs10121912.
23. Neyer, F.; Nocerino, E.; Gruen, A. MONITORING CORAL GROWTH-THE DICHOTOMY BETWEEN UNDERWATER PHOTOGRAMMETRY AND GEODETIC CONTROL NETWORK. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **2018**, *42*, 2. doi:10.5194/isprs-archives-XLII-2-759-2018.

24. Pilgrim, D.A.; Parry, D.M.; Jones, M.B.; Kendall, M.A. ROV Image Scaling with Laser Spot Patterns. *Underwater Technology* **2000**, *24*, 93–103. doi:10.3723/175605400783259684.
25. Davis, D.; Tusting, R. Quantitative benthic photography using laser calibrations. *Undersea World, San Diego, California* **1991**.
26. Isteničč, K.; Gracias, N.; Arnaubec, A.; Escartin, J.; Garcia, R. Automatic Scale Estimation of Structure from Motion based 3D Models using Laser Scalers. *arXiv e-prints*, [arXiv:1906.08019].
27. Nornes, S.M.; Ludvigsen, M.; Ødegard, Ø.; SØrensen, A.J. Underwater Photogrammetric Mapping of an Intact Standing Steel Wreck with ROV. *IFAC-PapersOnLine* **2015**, *48*, 206–211. doi:10.1016/j.ifacol.2015.06.034.
28. Warren, M.; Corke, P.; Pizarro, O.; Williams, S.; Upcroft, B. Visual sea-floor mapping from low overlap imagery using bi-objective bundle adjustment and constrained motion. Australasian Conference on Robotics and Automation; , 2012.
29. Strasdat, H.; Montiel, J.; Davison, A.J. Scale drift-aware large scale monocular SLAM. *Robotics: Science and Systems VI* **2010**, *2*, 7. doi:10.15607/RSS.2010.VI.010.
30. Fonstad, M.A.; Dietrich, J.T.; Courville, B.C.; Jensen, J.L.; Carbonneau, P.E. Topographic structure from motion: a new development in photogrammetric measurement. *Earth Surface Processes and Landforms* **2013**, *38*, 421–430. doi:10.1002/esp.3366.
31. Thoeni, K.; Giacomini, A.; Murtagh, R.; Kniest, E. A comparison of multi-view 3D reconstruction of a rock wall using several cameras and a laser scanner. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* **2014**, *40*, 573. doi:10.5194/isprsarchives-XL-5-573-2014.
32. Kalacska, M.; Chmura, G.; Lucanus, O.; Berube, D.; Arroyo-Mora, J. Structure from motion will revolutionize analyses of tidal wetland landscapes. *Remote Sensing of Environment* **2017**, *199*, 14 – 24. doi:10.1016/j.rse.2017.06.023.
33. Bythell, J.; Pan, P.; Lee, J. Three-dimensional morphometric measurements of reef corals using underwater photogrammetry techniques. *Coral reefs* **2001**, *20*, 193–199. doi:10.1007/s003380100157.
34. Courtney, L.A.; Fisher, W.S.; Raimondo, S.; Oliver, L.M.; Davis, W.P. Estimating 3-dimensional colony surface area of field corals. *Journal of Experimental Marine Biology and Ecology* **2007**, *351*, 234–242. doi:10.1016/j.jembe.2007.06.021.
35. Cocito, S.; Sgorbini, S.; Peirano, A.; Valle, M. 3-D reconstruction of biological objects using underwater video technique and image processing. *Journal of Experimental Marine Biology and Ecology* **2003**, *297*, 57–70. doi:10.1016/S0022-0981(03)00369-1.
36. McKinnon, D.; He, H.; Upcroft, B.; Smith, R.N. Towards automated and in-situ, near-real time 3-D reconstruction of coral reef environments. OCEANS'11 MTS/IEEE KONA. IEEE, 2011, pp. 1–10. doi:10.23919/OCEANS.2011.6106982.
37. Lavy, A.; Eyal, G.; Neal, B.; Keren, R.; Loya, Y.; Ilan, M. A quick, easy and non-intrusive method for underwater volume and surface area evaluation of benthic organisms by 3D computer modelling. *Methods in Ecology and Evolution* **2015**, *6*, 521–531. doi:10.1111/2041-210X.12331.
38. Gutiérrez-Heredia, L.; D'Helft, C.; Reynaud, E. Simple methods for interactive 3D modeling, measurements, and digital databases of coral skeletons. *Limnology and Oceanography: Methods* **2015**, *13*, 178–193. doi:10.1002/lom3.10017.
39. Ferrari, R.; McKinnon, D.; He, H.; Smith, R.; Corke, P.; Gonzalez-Rivero, M.; Mumby, P.; Upcroft, B. Quantifying multiscale habitat structural complexity: a cost-effective framework for underwater 3D modelling. *Remote Sensing* **2016**, *8*, 113. doi:10.3390/rs8020113.
40. Capra, A.; Dubbini, M.; Bertacchini, E.; Castagnetti, C.; Mancini, F. 3D reconstruction of an underwater archaelogical site: Comparison between low cost cameras. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **2015**, *40*, 67–72. doi:10.5194/isprsarchives-XL-5-W5-67-2015.
41. Guo, T.; Capra, A.; Troyer, M.; Grün, A.; Brooks, A.J.; Hench, J.L.; Schmitt, R.J.; Holbrook, S.J.; Dubbini, M. ACCURACY ASSESSMENT OF UNDERWATER PHOTOGRAMMETRIC THREE DIMENSIONAL MODELLING FOR CORAL REEFS. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences* **2016**, *41*. doi:10.5194/isprsarchives-XLI-B5-821-2016.
42. Raoult, V.; Reid-Anderson, S.; Ferri, A.; Williamson, J. How reliable is Structure from Motion (SfM) over time and between observers? A case study using coral reef bommies. *Remote Sensing* **2017**, *9*, 740. doi:10.3390/rs9070740.

43. Gonzalez-Rivero, M.; Bongaerts, P.; Beijbom, O.; Pizarro, O.; Friedman, A.; Rodriguez-Ramirez, A.; Upcroft, B.; Laffoley, D.; Kline, D.; Bailhache, C.; Vevers, R.; Hoegh-Guldberg, O. The Catlin Seaview Survey – kilometre-scale seascape assessment, and monitoring of coral reef ecosystems. *Aquatic Conservation: Marine and Freshwater Ecosystems* **2014**, *24*, 184–198. doi:10.1002/aqc.2505.
44. Escartín, J.; Le Friant, A.; Feuillet, N. SUBSAINTES Cruise Report, N/O L'Atalante - ROV VICTOR - AUV AsterX, 2017. doi:10.17600/17001000.
45. Bryson, M.; Johnson-Roberson, M.; Pizarro, O.; Williams, S.B. True Color Correction of Autonomous Underwater Vehicle Imagery. *Journal of Field Robotics* **2015**. doi:10.1002/rob.21638.
46. Hernández, J.D.; Istenič, K.; Gracias, N.; Palomeras, N.; Campos, R.; Vidal, E.; Garcia, R.; Carreras, M. Autonomous underwater navigation and optical mapping in unknown natural environments. *Sensors* **2016**, *16*, 1174. doi:10.3390/s16081174.
47. Bouguet, J.Y. Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm. *Intel Corporation* **2001**, *5*, 4.
48. Prados, R.; Garcia, R.; Gracias, N.; Escartin, J.; Neumann, L. A novel blending technique for underwater gigamosaicing. *Oceanic Engineering, IEEE Journal of* **2012**, *37*, 626–644. doi:10.1109/JOE.2012.2204152.
49. Jaffe, J.S. Computer modeling and the design of optimal underwater imaging systems. *Oceanic Engineering, IEEE Journal of* **1990**, *15*, 101–111. doi:10.1109/48.50695.
50. Mobley, C.D. *Light and water: radiative transfer in natural waters*; Academic press, 1994.
51. Bianco, G.; Gallo, A.; Bruno, F.; Muzzupappa, M. A comparison between active and passive techniques for underwater 3D applications. *International Archives of the Photogrammetry. In: Remote sensing and spatial information sciences* **2011**, *38*. doi:10.5194/isprsarchives-XXXVIII-5-W16-357-2011.
52. Andono, P.N.; Purnama, I.; Hariadi, M. UNDERWATER IMAGE ENHANCEMENT USING ADAPTIVE FILTERING FOR ENHANCED SIFT-BASED IMAGE MATCHING. *Journal of Theoretical & Applied Information Technology* **2013**, *52*. doi:10.1117/12.2503527.
53. Bianco, G.; Neumann, L. A fast enhancing method for non-uniformly illuminated underwater images. OCEANS-Anchorage, 2017. IEEE, 2017, pp. 1–6.
54. Alcantarilla, P.F.; Solutions, T. Fast explicit diffusion for accelerated features in nonlinear scale spaces. *IEEE Trans. Patt. Anal. Mach. Intell* **2011**, *34*, 1281–1298. doi:10.5244/C.27.13.
55. Agrawal, M.; Konolige, K.; Blas, M.R. Censure: Center surround extremas for realtime feature detection and matching. European Conference on Computer Vision. Springer, 2008, pp. 102–115. doi:10.1007/978-3-540-88693-8_8.
56. Moisan, L.; Moulon, P.; Monasse, P. Automatic homographic registration of a pair of images, with a contrario elimination of outliers. *Image Processing On Line* **2012**, *2*, 56–73. doi:10.5201/ipol.2012.mmm-oh.
57. Nister, D.; Stewenius, H. Scalable recognition with a vocabulary tree. Computer vision and pattern recognition, 2006 IEEE computer society conference on. Ieee, 2006, Vol. 2, pp. 2161–2168. doi:10.1109/CVPR.2006.264.
58. Schönberger, J.L.; Price, T.; Sattler, T.; Frahm, J.M.; Pollefeys, M. A vote-and-verify strategy for fast spatial verification in image retrieval. Asian Conference on Computer Vision. Springer, 2016, pp. 321–337. doi:10.1007/978-3-319-54181-5_21.
59. Triggs, B.; McLauchlan, P.F.; Hartley, R.I.; Fitzgibbon, A.W. Bundle adjustment – a modern synthesis. In *Vision algorithms: theory and practice*; Springer, 1999; pp. 298–372. doi:10.1007/3-540-44480-7_21.
60. Moulon, P.; Monasse, P.; Marlet, R. Adaptive structure from motion with a contrario model estimation. Asian Conference on Computer Vision. Springer, 2012, pp. 257–270. doi:10.1007/978-3-642-37447-0_20.
61. Moulon, P.; Monasse, P.; Marlet, R.; Others. OpenMVG. An Open Multiple View Geometry library. https://github.com/openMVG/openMVG.
62. Moulon, P.; Monasse, P.; Marlet, R. Global fusion of relative motions for robust, accurate and scalable structure from motion. Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 3248–3255. doi:10.1109/ICCV.2013.403.
63. Michot, J.; Bartoli, A.; Gaspard, F. Bi-objective bundle adjustment with application to multi-sensor slam. *3DPVT'10* **2010**, *3025*.
64. Aitken, A.C. On least squares and linear combination of observations. *Proceedings of the Royal Society of Edinburgh* **1936**, *55*, 42–48. doi:10.1017/S0370164600014346.

65. Svärm, L.; Oskarsson, M. Structure from motion estimation with positional cues. Scandinavian Conference on Image Analysis. Springer, 2013, pp. 522–532. doi:10.1007/978-3-642-38886-6_49.
66. Pareto, V. *Cours D'Economie Politique*; F. Rouge: Lausanne, Switzerland, 1896.
67. Shen, S. Accurate multiple view 3d reconstruction using patch-based stereo for large-scale scenes. *IEEE transactions on image processing* **2013**, *22*, 1901–1914. doi:10.1109/TIP.2013.2237921.
68. Jancosek, M.; Pajdla, T. Exploiting visibility information in surface reconstruction to preserve weakly supported surfaces. *International scholarly research notices* **2014**, *2014*. doi:10.1155/2014/798595.
69. Waechter, M.; Moehrle, N.; Goesele, M. Let there be color! Large-scale texturing of 3D reconstructions. In *Computer Vision–ECCV*; Springer, 2014; pp. 836–850. doi:10.1007/978-3-319-10602-1_54.
70. Pérez, P.; Gangnet, M.; Blake, A. Poisson image editing. ACM Transactions on Graphics (TOG). ACM, 2003, Vol. 22, pp. 313–318. doi:10.1145/882262.882269.
71. Ke, T.; Roumeliotis, S.I. An efficient algebraic solution to the perspective-three-point problem. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 7225–7233.
72. Caimi, F.M.; Kocak, D.M.; Dalgleish, F.; Watson, J. Underwater imaging and optics: Recent advances. OCEANS 2008, 2008, Vol. 2008-Supplement, pp. 1–9. doi:10.1109/OCEANS.2008.5289438.
73. Massot-Campos, M.; Oliver-Codina, G. Underwater Laser-based Structured Light System for one-shot 3D reconstruction. SENSORS, 2014 IEEE, 2014, pp. 1138–1141. doi:10.1109/ICSENS.2014.6985208.
74. Rzhanov, Y.; Mamaenko, A.; Yoklavich, M. UVSD: software for detection of color underwater features. Proceedings of OCEANS 2005 MTS/IEEE, 2005, pp. 2189–2192 Vol. 3. doi:10.1109/OCEANS.2005.1640089.
75. Schoening, T.; Kuhn, T.; Bergmann, M.; Nattkemper, T.W. DELPHI—fast and adaptive computational laser point detection and visual footprint quantification for arbitrary underwater image collections. *Frontiers in Marine Science* **2015**, *2*, 20. doi:10.3389/fmars.2015.00020.
76. Padfield, D. Masked FFT registration. Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, 2010, pp. 2918–2925. doi:10.1109/CVPR.2010.5540032.
77. Evangelidis, G.D.; Psarakis, E.Z. Parametric image alignment using enhanced correlation coefficient maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2008**, *30*, 1858–1865. doi:10.1109/TPAMI.2008.113.
78. Wan, E.A.; Van Der Merwe, R. The unscented Kalman filter for nonlinear estimation. Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373), 2000, pp. 153–158. doi:10.1109/ASSPCC.2000.882463.
79. Michel, J.L.; Klages, M.; Barriga, F.J.; Fouquet, Y.; Sibuet, M.; Sarradin, P.M.; Siméoni, P.; Drogou, J.F.; others. Victor 6000: design, utilization and first improvements. The Thirteenth International Offshore and Polar Engineering Conference. International Society of Offshore and Polar Engineers, 2003.
80. Escartín, J.; Leclerc, F.; Olive, J.A.; Mevel, C.; Cannat, M.; Petersen, S.; Augustin, N.; Feuillet, N.; Deplus, C.; Bezos, A.; Bonnemains, D.; Chavagnac, V.; Choi, Y.; Godard, M.; Haaga, K.; Hamelin, C.; Ildefonse, B.; Jamieson, J.W.; John, B.E.; Leleu, T.; Macleod, C.J.; Massot-campos, M.; Nomikou, P.; Paquet, M.; Rommevaux-Jestin, C.; Rothenbeck, M.; Steinfuhrer, A.; Tominaga, M.; Triebe, L.; Campos, R.; Gracias, N.; Garcia, R.; Andreani, M.; Vilaseca, G. First direct observation of coseismic slip and seafloor rupture along a submarine normal fault and implications for fault slip history. *Earth and Planetary Science Letters* **2016**, *450*, 96–107. doi:10.1016/j.epsl.2016.06.024.
81. Bouguet, J.Y. Camera calibration toolbox for Matlab (2008). *URL http://www. vision. caltech. edu/bouguetj/calib_doc* **2008**, *1080*.