Preprints.org

Article

Refraction-aware Structure-from-Motion for Airborne Bathymetry

Alexandros Makris * , Vassilis C. Nicodemou , Evangelos Alevizos , Iason Oikonomidis , Dimitrios D. Alexakis ,
Anastasios Roussos

Article

Refraction-Aware Structure-from-Motion for Airborne Bathymetry

Alexandros Makris [1],*, Vassilis C. Nicodemou [1], Evangelos Alevizos [2], Iason Oikonomidis [1], Dimitrios D. Alexakis [2] and Anastasios Roussos [1]

[1]   ICS FORTH; e-mail@ics.forth.gr
[2]   IMS FORTH; e-mail@ims.forth.gr
*   Correspondence: amakris@ics.forth.gr

Abstract: In this work, we introduce the first pipeline that combines a refraction-aware Structure-from-Motion (SfM) method with a deep learning model specifically designed for airborne bathymetry. We accurately estimate the 3D positions of the submerged points by integrating refraction geometry within the SfM optimization problem. This way, no refraction correction as post-processing is required. Experiments with simulated data that approach the real-world capturing conditions demonstrate that SfM with refraction correction is extremely accurate, with submillimeter errors. We integrate our refraction-aware SfM within a deep learning framework that also takes into account radiometrical information, developing a combined spectral and geometry-based approach, with further improvements on accuracy and robustness to different seafloor types, both textured and textureless. We conducted experiments with real-world data at 2 locations in the southern Mediterranean Sea, with varying seafloor types which demonstrate the benefits of refraction correction for the deep learning framework. We open-source our refraction-aware SfM, providing researchers in airborne bathymetry with a practical tool to apply SfM in shallow water areas.

Keywords: refraction-aware structure-from-motion; deep learning; shallow water bathymetry; drone imagery

1. Introduction

Shallow water bathymetry is an active field of research with modern drone imagery based methods performing in unprecedented detail. Spectral and geometry-based approaches are the most common methodologies that are adopted in the literature and sometimes are used in combination. Spectral approaches rely on the fact that light attenuation in the water column depends on the light's wavelength, therefore an estimation of the depth can be derived from ratios of different spectral bands [1–4].

Geometry-based approaches take advantage of the multiview image geometry to produce a 3D surface from corresponding points between successive images. Recent works have shown that they can provide promising results in airborne bathymetry [5,6]. Furthermore, geometry-based approaches are also particularly useful as an important complement to spectral-based approaches, offering additional information, especially in seafloor areas with rich texture [7–9], and providing training data for several spectral-based approaches [10]. They are based on the basic principle that when a point is viewed from multiple views that correspond to different drone positions, its 3D position can be recovered. Computationally, this is achieved through Structure-from-Motion (SfM) and bundle adjustment [11,12] in an optimization framework. However, since in airborne bathymetry the camera positions are in the air and the observed points are within the sea, the refraction phenomenon affects this process and introduces inaccuracies in the estimation of the standard SfM technique, which has been designed with the assumption that all observed points are in the air, or more generally, in a single, homogeneous medium. Despite this fact, all previous approaches are indeed applying standard SfM and, in order to compensate for the introduced errors, are adopting some form of "refraction correction" [5,10,13–15]. These corrections decrease the introduced errors,

but cannot eliminate them completely, since important information related to the real optical phenomenon has been lost in the process of applying the standard SfM which previous works use as a black box.

In this work, we fill this important gap by introducing a novel, refraction-aware SfM method specially designed for airborne bathymetry. Our method models the air-water refractive surface by integrating the refraction geometry within the SfM optimization problem. In this way, it directly yields accurate estimates of the 3D positions of the observed points at the seafloor, without requiring any refraction correction as post-processing. This methodology is combined with a convolutional neural network (CNN) for a complete pipeline that effectively manages to address the shallow bathymetry estimation problem regardless of the seafloor type. The deep learning framework takes also into account radiometrical information, developing a combined spectral and geometry-based approach. Our experiments include refraction-aware SfM evaluation with simulated data, which reveal that it is extremely accurate, leading to depth estimation errors that are smaller than 0.001m, as compared to a range of 0.22m to 1.5m (depending on water depth) achieved by the standard SfM. To test our complete pipeline we also conduct real world experiments in two different areas with different seafloor types both textured and textureless.

1.1. State of the Art Review

1.1.1. Structure-from-Motion

Structure-from-Motion (SfM) is a photogrammetric technique which is widely used for deriving shallow bathymetry using optical sensors typically from satellite or aerial imagery [16]. This technique produces a cloud of 3D points corresponding to points on the seafloor, based on point correspondences between consecutive overlapping images. Successful SfM results require optimal environmental conditions during data acquisition such as: increased water clarity, very low wave height, minimal breaking waves, and minimal sunglint [5,8]. Recent works try to relax some of these requirements by creating video composite images [9]. Additionally, SfM requires significant seafloor texture.

To perform multimedia photogrammetry, which means that the camera and the object of interest are not in the same optical media, requires the extension of standard photogrammetric imaging models. This is the case for the problems of underwater SfM and airborne seafloor reconstruction. For the case of airborne reconstruction of a shallow seafloor, light refraction due to the air-water interface should be taken into account. The reconstruction error caused by refraction directly depends on the water depth and on the incidence angles of the light rays and is significant for typical drone-based datasets. The main approaches in the literature for refraction correction are based on either machine learning or geometric algorithms. Machine learning based approaches are used to correct refraction effects in the 3D reconstruction space [5] or in the image space [14,17]. On the other hand, geometric approaches directly model the refraction effect to perform the reconstruction. The main difficulty with these geometric approaches is that triangulation of submerged points given stereo observations is theoretically not possible in the general case; exact solutions can only be obtained for very few particular points [18]. Due to these limitations, several approximation methods have been proposed. The method of [19] uses Snell's law to retriangulate the initially-obtained 3D points given a flat and known water surface. In [20], the camera intrinsic and extrinsic calibration parameters are computed using frames from the onshore part of the dataset while refraction is corrected according to the method of [19]. In [21], an improved triangulation algorithm is proposed for the case of non-intersecting conjugate image rays. The method of [6] is also based on standard SfM to obtain an initial point cloud, and uses a simplifying approximation to calculate the refraction correction for each point from each camera independently. More precisely, it assumes that the horizontal position of a point is not affected by refraction, thus only the depth is corrected. The average of these per-camera estimations is used, leading to inexact solutions. In the survey of [22], the authors apply the refraction correction from [6] for reconstructing very shallow areas (<2 m) without significantly improving the final bathymetry. In the works of [23,24], geometric models to handle refraction in SfM by including it in the bundle adjustment step are presented. By incorporating these models into

standard SfM pipelines, very accurate estimations can be achieved under certain basic assumptions of a planar refractive interface and a constant refractive index. In the aforementioned works this is only demonstrated in laboratory conditions. In [25], the authors present a small case study for applying these methods to airborne bathymetry without providing the details of the implementation and the source code. They process images above land areas for deriving camera intrinsic and extrinsic parameters, and then they use these parameters as fixed during processing images from further offshore. In this work we   provide an accurate refraction aware solution to the SfM problem, integration in a deep learning framework for airborne bathymetry, and an open source implementation.

1.1.2. Radiometry

The logarithmic band ratios approach introduced in [26] sets the standard for most of the empirical satellite-derived bathymetry (SDB) methods. Recent methods typically incorporate these features in a machine learning framework showing promising results. In [27–29], artificial neural networks on Landsat, IKONOS, and IKONOS-2 imagery are applied with promising results for water depths reaching 20m. In [30] the authors develop a CNN tested on Sentinel-2 imagery, trained with sonar and LIDAR bathymetry.

A few recent studies apply the methods originally developed for satellite-derived bathymetry on drone-based multispectral imagery. For instance approaches in [7,31,32] show good results with less than half a meter vertical errors. Accordingly, the authors of [10] applied a deep learning methodology for extracting bathymetry by incorporating spectral and photogrammetric features of airborne imagery while lately, in [15], a refraction correction approach for spectrally derived bathymetry is proposed.

Typical constraints in most empirical SDB studies include the requirement for ground-truth depth measurements and the difficulty of handling heterogeneous seafloors. The latter occurs particularly in shallow and relatively flat seafloor areas with mixed seafloor types (e.g.: sand, reefs, algae, seagrasses).

1.1.3. Deep Learning

Recent advancements in bathymetry have leveraged deep learning to improve accuracy and efficiency in mapping underwater topographies. Al Najar et al. [33], Benshila et al. [34], and Chen et al. [35] have utilized deep learning for bathymetry estimation from satellite and aerial imagery, with a focus on satellite-derived bathymetry and the integration of physics-informed models in nearshore wave field reconstruction and bathymetry mapping. Lumban-Gaol et al. [30] and Mandlburger et al. [10] have applied Convolutional Neural Networks (CNNs) to multispectral images for bathymetric data extraction. Lumban-Gaol et al. [30] utilized Sentinel-2 images in their CNN-based approach for coastal management, while Mandlburger et al. [10] developed BathyNet, which combines photogrammetric and radiometric methods for depth estimation from RGBC aerial images.

Addressing large-scale riverine bathymetry, Forghanioozroody et al. [36] and Ghorbanidehno et al. [37] demonstrate the efficiency of deep learning over traditional methods. They showcase the adaptability of deep learning techniques in varying environmental conditions. Jordt et al.'s work on refractive 3D reconstruction [38] and Sonogashira et al.'s study on deep-learning-based image superresolution [39] further exemplify the applicability of deep learning in accurately representing complex underwater landscapes and reducing the need for extensive depth measurements.

Alevizos et al.'s work [7] aligns closely with the current work's focus. The method of [7] integrates standard SfM (ignoring refraction) outputs with drone image band-ratios into a CNN achieving effective bathymetry prediction with significant accuracy. The study demonstrates its capability for handling different seabed types. However, since refraction is ignored the performance of the method is severely impacted, as we demonstrate in our experiments.

The aforementioned works underscore the role of deep learning in modern bathymetric applications, revealing a trend towards more precise, data-driven, and efficient methods for underwater mapping, essential for a range of marine and environmental purposes.

1.2. Contributions

In summary, our main contributions are the following:

- We implement a Refraction-aware SfM (R-SfM) pipeline within the OpenSfM framework. Refraction is taken into account in the bundle adjustment problem leading to very accurate solutions. We experimentally validate the pipeline using real and simulated data.

- We demonstrate that a CNN pipeline that combines the R-SfM provided geometric information with radiometric information achieves accurate depth estimations. We experimentally validate it on real world data. Our pipeline is especially designed to balance between R-SfM and radiometric-based estimations.

- We open source the R-SfM and CNN source code. In addition, we make the data available upon request for the benefit of the research community. (*Note for the reviewers: the source code and data will be available upon publication*).

2. Materials and Methods

We solve the bathymetry estimation problem using a CNN-based pipeline, see Figure 2. The employed CNN relies on SfM and radiometric features. The features are extracted from a set of drone-acquired images, which provide the main input for our system. For training, unmanned surface vehicle (USV) based bathymetric measurements are used. The trained CNN performs dense bathymetry estimation on the examined region. In the following, we describe our system's main components i.e. the refraction-aware SfM (R-SfM), the radiometric data processing, and the CNN model.
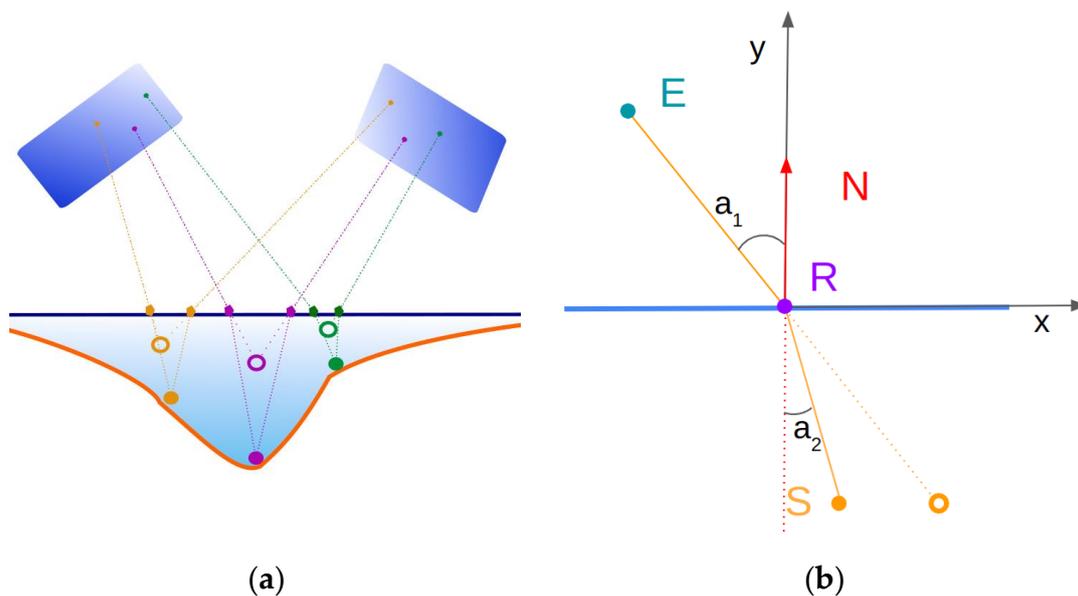

(**a**)   (**b**)

Figure 1. (**a**) Concept sketch of the proposed refraction aware structure from motion (R-SfM) method. Our method estimates the true point locations (solid points) during the bundle adjustment stage of SfM. This is in contrast to most other approaches which first perform standard SfM getting wrong position estimations (rings) and attempt to correct them in post processing. (**b**) Refraction Geometry. Point S is observed by E through a planar refractive interface with normal N. Point R is the apparent position of S on the interface. Snell's law determines the relationship between the angles a_1 and a_2.
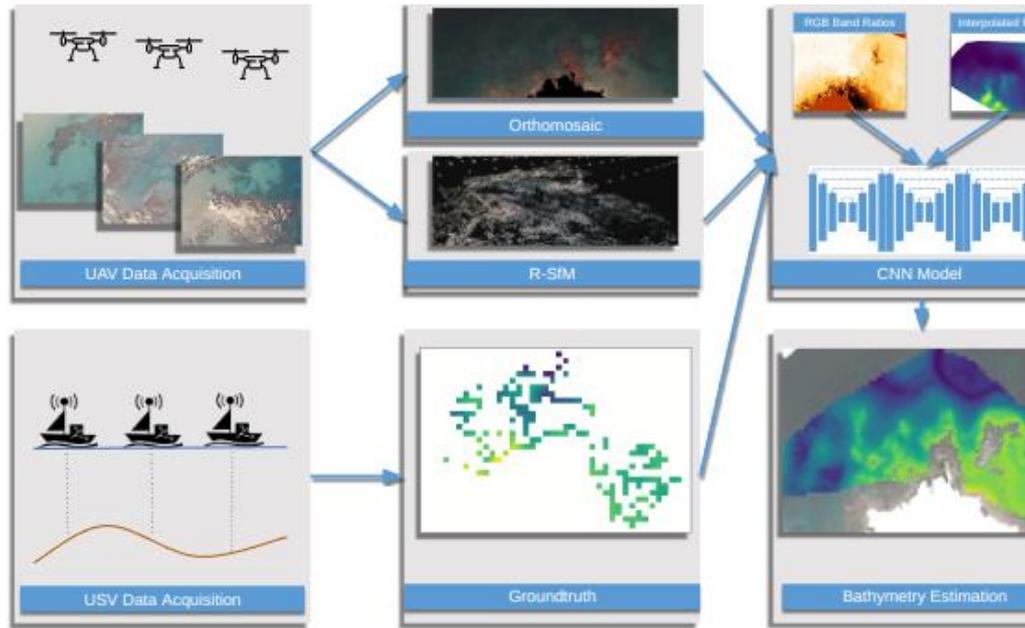
Figure 2. Method Pipeline. Drone acquired images are used as input to our R-SfM method. From the same images, the RGB band ratios are extracted and fed along with the R-SfM output to the CNN model. The grountruth for CNN training is obtained using sonar measurements performed by the USV. The CNN output is dense bathymetry.

2.1. Refraction Aware Structure from Motion

SfM algorithms typically solve a bundle adjustment problem which is based on a geometric scene model. In this section, we describe the employed SfM algorithm [40] and the geometric model [41] that allows the handling of refractive interfaces (see Figure 1).

Our SfM implementation follows the incremental SfM pipeline [42,43]. The input is a set of drone acquired images and their approximate poses extracted using GPS/IMU sensors, and 3D coordinates and image locations of the ground control points (GCPs).

The first part of the method generates a set of point tracks in the images using feature detection and matching. Keypoint detection is performed using the Hessian affine region detector [44]. For feature description, a histogram of oriented gradients type descriptor similar to SIFT is used [45].

The second part uses the tracks to estimate the 3D positions of the corresponding points along with the camera's intrinsic and extrinsic parameters. The reconstruction starts with two images with a large intersection. Their respective poses and 3D points are recovered through triangulation and are refined by bundle adjustment. Subsequently, new images are added to the reconstructed scene sequentially by finding the camera pose that minimizes the reprojection error of the already reconstructed 3D points to the new image (resectioning). A local bundle adjustment is performed after each iteration to decrease accumulated errors. A global bundle adjustment is conducted after a predefined number of images have been added and also at the end of the reconstruction to refine all camera poses and scene points.

The bundle adjustment consists of refining a set of initial camera parameters and 3D point locations by minimizing the so-called reprojection error, i.e. the distance between the image projections of the estimated 3D points and the detected image points. Let n be the number of 3D points and m the number of images. Each camera j is parameterized by a vector a_j and each 3D point i by a vector b_i. The minimization of the reprojection error can thus be written as follows:

$$min_{a,b} \sum_{i=1}^{n} \sum_{j=1}^{m} d(Q(a_j, b_i), x_{ij})^2, (1)$$

where $Q(a_j, b_i)$ is the projection of point i on image j and d denotes the Euclidean distance between the image points.

To account for the effects of refraction, we consider it in the bundle adjustment algorithm by locating the apparent position of the submerged points on the surface. Let a 3D point S observed by E through a planar refractive interface P_i with normal N. Let R be the apparent position of S on the interface P_i. The refraction occurs on the plane P_r which is the plane perpendicular to the interface P_i that includes points $E(e_x, 0)$, $S(s_x, s_y)$, and $R(0, r_y)$. According to Snell's law the incident and refracted ray angles are on P_r are given by: $sina_1 = rsina_2$ where r is the refraction ratio between the two media (see Figure 1).The location of R is calculated by solving the following equation:

$$f(r_y) = Nr_y^4 - 2Ns_yr_y^3 + (Ns_y^2 + \frac{s_x^2}{r^2} - e_x^2)r_y^2 + 2e_x^2s_yr_y -$$

$$e_x^2s_y^2, (2),$$

where: $N = \frac{1}{r^2} - 1$.

Equation (2) has 4 solutions but only one lies on $[0, s_y]$ which is the location of R. The bundle adjustment problem can be solved in the presence of a refractive interface by using (2) to calculate the apparent point on the interface R which corresponds to point S. R is then projected in the image as in classical bundle adjustment [40].

2.2. Radiometric Data Processing

Structure-from-motion provides bathymetry only in areas with rich seafloor texture, leaving areas with homogeneous seafloor un-reconstructed. To complement it, we rely on the empirical SDB method of [26] which employs the logarithmic band ratios from multispectral data. The main concept of this approach is that light attenuation in the water column is exponential for wavelengths in the visible spectrum. We follow an approach similar to [7] in order to prepare the imagery for use with the CNN model.

2.3. Deep Learning Model

The deep learning model used in this work, follows the design and architecture choices described in Alevizos et al.'s work [7]. Specifically, the model is a Convolutional Neural Network (CNN) based on a stacked-hourglass [46] backbone. As shown in this work [7] this model is robust in depth estimation from multichannel images for   shallow bathymetry, since this type of model was specially designed to find dominant features in multichannel inputs. This architecture's success has also been proven in other domains of depth acquisition, such as depth estimation for hands [47] and faces [48]. Moreover, the specific model can be characterized as lightweight, with a relatively low number of parameters. Since the use case of this work is similar to the one studied in [7], we chose the number of stacked hourglass networks to be 6.

3. Results

3.1. Data Collection Platforms

3.1.1. Onshore Survey

Prior to the drone surveys a set of ground control points (GCPs) were uniformly distributed and measured along the coastline of each study area. The GCPs were measured with a Global Navigation Satellite System Real-Time Kinematics (GNSS RTK) GPS [49] for achieving high accuracy (± 2cm). This level of accuracy is crucial in drone surveys that produce imagery with centimeter-scale spatial resolution while the onboard GPS sensor has a horizontal accuracy of approximately two meters. Thus, the GCPs are used for accurate orthorectification of the point-clouds and 2D reflectance mosaics.

3.1.2. Drone Platform

The drone platform comprises a commercial DJI Phantom 4 Pro drone equipped with a 20 Mpixel RGB camera. The Phantom 4 Pro has a maximum flight time of 28 minutes and carries an on-board camera that shoots 4k 60 frames/second video.   Drone imagery was processed with SfM techniques for producing an initial bathymetric surface (see the following section).   The raw imagery values were converted to reflectance by using a reference reflectance panel and Pix4D software. The blue (B) and green (G) bands correspond to shorter wavelengths (460±40 nm and 525±50 nm respectively) and thus show deeper penetration through the water column. The red (R) band corresponds to 590±25 nm, which is highly absorbed in the first 1-2 meters of water depth, but assists in emphasizing extremely shallow areas.

3.1.3. USV Platform

The USV used is a remote-controlled platform mounted with an Ohmex BTX single-beam sonar with an operating frequency of 235 kHz. The sonar is integrated with a mounted   RTK-GPS sensor and it collects altitude-corrected bathymetry points at 2 Hz sampling rate. All USV data were collected on the same date as drone imagery to avoid temporal changes in bathymetry. The RTK-GPS measurements provide high spatial accuracy, which is essential in processing drone-based imagery with a pixel resolution of a few centimeters.

3.2. Dataset

3.2.1. Simulated Data

We generated a simulated dataset to thoroughly test our R-SfM method with complete ground truth. The data consists of a point cloud representing the points of the seabed and a set of camera poses. The 3D points are projected into each camera taking into account the refraction effect for those which are underwater. The point correspondences between camera poses which are used during the reconstruction are considered known.

We generated 3 different simulated areas (see Table 1 and Figure 3). For the first area (named GS) we used a simple ridge-shape geometry to represent the seabed and   generated the drone trajectory over the area. Five different versions of this area were produced with different maximum depths (0-20m). For the other two areas (named RSA and RSB) we used real data from actual surveys to extract the seabed and the drone trajectory. Depth values provided by the USV were interpolated to generate these more realistic setups. The drone trajectories were extracted from the GPS data.

Table 1. Simulated dataset specification.

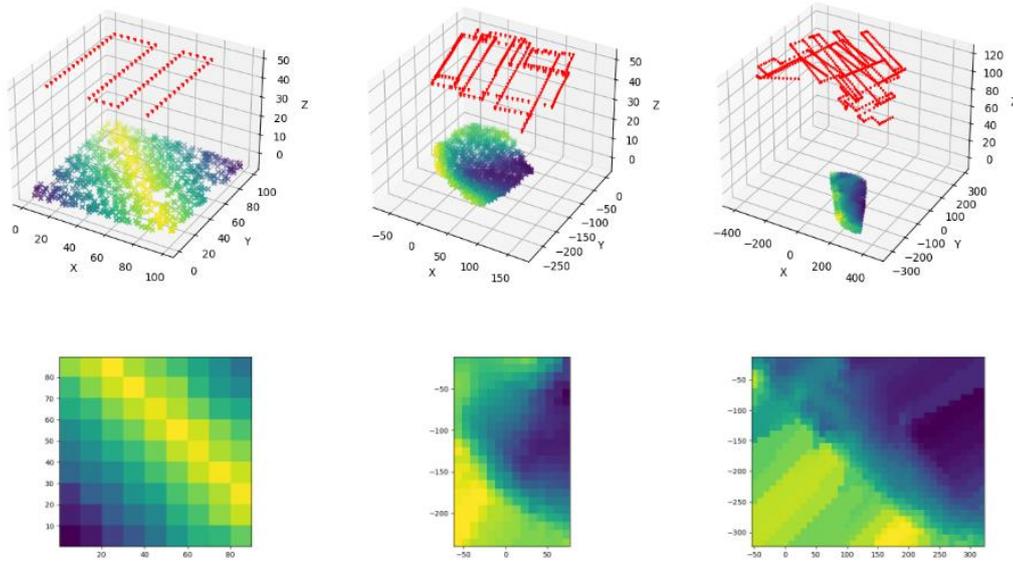|	**GS**	**RSA**	**RSB**
Flight Altitude	50m	50m	120m
Area	100mx100m	200mx150m	400mx200m
Max Depth	0-20m	3.5m	5m

Figure 3. Simulated dataset areas: **GS**, **RSA**, **RSB**. Color coding represents water depth, darker is deeper.

3.2.2. Real Data

Real-world data were captured from two sites, Kalamaki and Plakias, both located on the island of Crete, Greece. The data consist of a set of UAV acquired images and USV acquired depth measurements. The USV measurements are shown in Figure 4. In the Kalamaki area, 810 drone images and 2145 depth measurements were acquired, while in the Plakias area 440 UAV images and 1838   USV points were acquired.

Considering that the tidal range on the island of Crete is a maximum ±0.2 m [50] and drone data acquisition has a one-hour difference with USV data, we infer that the tidal effect is completely negligible on the USV and drone data.

3.2.3. Error Metrics

In the following, we define the metrics that we use to evaluate the methods on the datasets presented above. Given N ground truth-estimated point pairs (x_i, y_i) the root-mean-square error (**RMSE)** and **R²** metrics are defined as:

$$RMSE = \sqrt{\frac{SSR}{N}}, \ R^2 = 1 - \frac{SSR}{SST},$$

where,

$$SSR = \sum_i \ \ (x_i - y_i)^2, \ \ SST = \sum_i \ \ (x_i - \underline{x})^2$$

with $\underline{x}$ the mean of the ground truth points.

3.3. Deep Learning Model Training

The training procedure of the CNN consists of a data preprocessing step and the training routine. The data preprocessing required constitutes the formulation of multichannel input tiles and their respective labeled output, which the network utilizes together. The tiling procedure in the acquired data was followed for several reasons. Firstly, the depth acquired measurements are not dense over the whole imaged area, and therefore the training can only be performed in specific parts of the imaged area. The tiling allows one to select only the annotated regions and exclude the rest. For the data preprocessing, the training input was created from tiles of the acquired RGB data of size 128x128 pixels that include four channels: three channels for the logarithmic band ratios (Blue/Green,

Blue/Red and Green/Red) and one for the R-SfM surface. The labeled output tiles of the training set consist of single channeled 128x128 pixel tiles of interpolated USV depth that depict the depth values of the respective input. In the case of R-SfM and USV data, a thin-plate spline interpolation is applied within the region of each patch in order to fill each pixel with usable data, since the original data is composed of sparse point measurements. The training/testing sets split randomly in the ratio of 60/40 while the USV testing points neighboring less than 3 meters from the training patches were discarded, to minimize the effect of spatial auto-correlation during testing.

The whole CNN framework was implemented in PyTorch and the training routine is formulated in 100 epochs. For optimizing the network in each epoch we use the Adam optimizer [51] with learning rate of 10^{-5} and batch size of 8 patches. In each epoch the multi-channeled input traverses through the network. Each stacked hourglass of the model provides a refined estimated version of the output depth. The validity of the output of each stack is quantified with an error metric of the loss function between the computed output and the defined ground truth USV labeled depth. After computing the loss, the network's weights are updated using backpropagation.

3.4. Experimental Results

In this section, we present the experimental results for the methods proposed in the paper. The results include validation of the R-SfM method using simulated and real data. The CNN method is trained only on real data to prove the beneficial use of R-SfM data acquired from our method over other acquisitions in real world cases.

3.4.1. Simulated Data

We thoroughly test our proposed R-SfM approach on the simulated datasets described in the previous section. We compute the RMSE metric between reconstructed and GT point locations for all points (since their locations are known). Additionally, the RMSE metric is computed for the estimated and GT camera locations.

The results demonstrate that in the case of simulated data with full observability the proposed R-SfM approach successfully reconstructs the scene in all simulated scenes with very small errors (<1e-05m), (see Table 2). For the classical SfM where refraction effects are not considered, the errors range from 0.18m to 0.35m. The refraction influence on the camera pose estimation for SfM negatively affects the overall quality of the reconstruction since this type of pose errors are accumulated from frame to frame.

To highlight the influence of water depth in the estimation, we generated 5 different versions of the GS scene by placing the same slope in different depths. The maximum water depths range from 0 m to 20m and the results are shown in Table 3. As is expected the encountered depths negatively influence the classical SfM method, its reconstruction error grows with increasing depth. The proposed R-SfM method is not influenced by the water depths, the reconstruction errors are negligible for all tested depths. We should note however that in reality water depth influences the visibility of the seafloor therefore the method can only be used in shallow water (depths 0-20m).

Table 2. RMSE of the reconstructed points for the 3 simulated scenes. GS05 contains depths of up to 5m.

	GS05	RSA	RSB
SfM	0.22m	0.35m	0.18m
R-SfM	7e-09m	5e-05m	7e-05m

Table 3. RMSE of the reconstructed points for the different maximum depths: GS00 contains depths in (-5m, 0m) so no points are submerged in water, GS05 contains points with depths in (0m, 5m), etc..

	GS00	GS05	GS10	GS15	GS20
Min/Max Depth	-5m/0m	0m/5m	5m/10m	10m/15m	15m/20m
SfM	7e-09m	0.22m	0.36m	1.37m	1.54m
R-SfM	7e-09m	5e-05m	1e-05m	3e-05m	5e-05m

3.4.2. Real Data

We evaluated the sparse SfM reconstructions using real data from the Kalamaki dataset. The RMSE and R^2 metrics are measured between the reconstructed points and the GT data provided by the USV. The comparison is performed by locating the nearest reconstruction point for each USV datapoint. If the distance between these two points on the horizontal plane is below a threshold, set to 1m, then the pair is considered for the metrics calculations otherwise it is discarded. The results clearly demonstrate the benefits of R-SfM over plain SfM. As shown in the graphs of Figure 5, SfM fails to reconstruct the scene. Accumulated camera pose estimation errors lead to very high reconstruction errors. Most of the reconstructed seabed points have depths outside the ground-truth points range. On the other hand, R-SfM successfully reconstructs the scene. The RMSE in this real world dataset is significant (RMSE=0.75m) compared to the error of the simulated datasets. However, this is expected due to the inaccuracies in camera pose initialization from the GPS, and the noisy underwater measurements (point tracks).
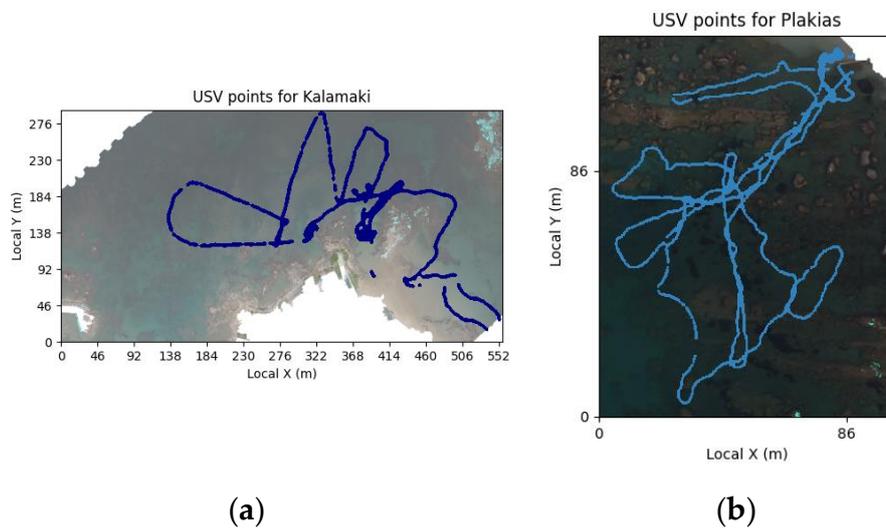

(a)   **(b)**

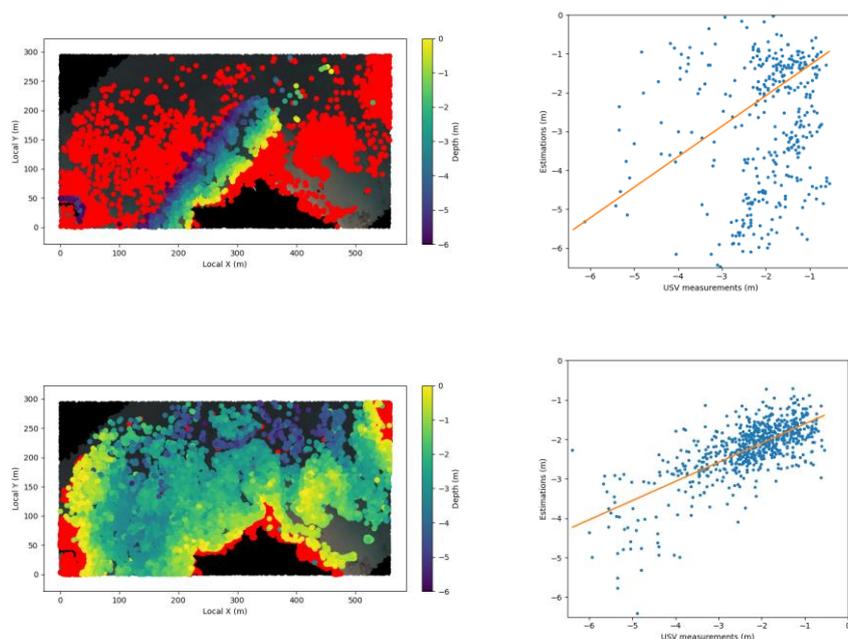**Figure 4.** USV measurements (blue dots). (**a**) Kalamaki bay, (**b**) Plakias.

Figure 5. Sparse SfM (top) / R-SfM (bottom) estimations and corresponding scatter plots of test points on Kalamaki dataset.The red points on the estimation maps represent points with depth outside of the colormap range [-6m, 0m].

Many works show that deep learning/CNN methods achieve the best results in-depth estimation bathymetry through learning [7,33–35]. Since CNN-based approaches are very sensitive to the data used for training [52,53], any significant change in the data will also significantly change the desired accuracy of the results. Therefore we can evaluate the quality of our R-SfM method by employing the computed data in a CNN model.

The same can also be said when using traditional Machine Learning methods. Models   acquired from training Support Vector Machines (SVMs) or Random Forests (RFs) trained on different types of data can also show which data are more dominant for training. All of the above scenarios are examined in the following subsections.

To fully assess the fidelity of the computed R-SfM, we performed a number of experiments to show its beneficial use   in CNN based depth estimation scenarios. We trained the previously described stacked hourglass CNN architecture on our multi-channeled input using a combination of different cases. Specifically, we trained different models with different RGB, SfM, and R-SfM input combinations.

In Table 4 we report the RMSE and R^2 for each model when evaluated on the respective test set. In Figures 6 and 7 we show the CED curves and scatter plots while Figure 9a shows the depth residuals. The first row containing the proposed R-SfM channel together with RGB as input yields the best results as expected. We can observe that RGB data plays a paramount role in estimation, yielding good results when utilized independently (row 3). But when combined with SfM we notice a deterioration in the results (row 2). This can be justified as the SfM channel has many erroneous samples as can be seen in Figure 5. Therefore when using SfM alone for training we get the worst results (row 4), while when using R-SfM alone gives moderately good results (row 5), close to the ones obtained by utilizing the RGB only.

Table 4. Comparison with alternative CNN pipelines and sparse reconstructions: RMSE and R^2 of reconstructed points on the Kalamaki dataset for different channels. See also Figure 6 (CED curves) and Figure 7 (scatter plots).

Method	RMSE	R^2
CNN R-SfM and RGB	0.36m	0.84
CNN SfM and RGB [similar to our previous paper]	0.70m	0.23
CNN RGB only	0.59m	0.56
SfM (intermediate sparse reconstruction)	2.71m	-6.44
R-SfM (intermediate sparse reconstruction)	0.75m	0.48

To further prove the proposed R-SfM's beneficial quality, we also trained an SVM and an RF model with different channel combinations. The results are reported in Table 5. Again, we can observe the performance gain achieved when using the proposed R-SfM channel (rows 2,3) instead of the SfM channel (rows 5,6).

Table 5. Comparison with traditional ML methods **RMSE and R^2** of reconstructed points on the Kalamaki dataset for various ML models and training channels.

Method	RMSE	R^2
CNN RGB+R-SfM	0.36m	0.84
SVM RGB+R-SfM	0.65m	0.47
RF RGB+R-SfM	1.69m	-6.18
CNN RGB+SfM	0.70m	0.23
SVM RGB+SfM	2.60m	-5.51
RF RGB+SfM	2.86m	-6.80

In addition to the single coast results, we performed an initial cross coast evaluation of our proposed method. We trained the CNN on the Kalamaki dataset and tested on the Plakias dataset. These results are shown in Figures 8 and 9b. We achieved an RMSE of 0.82m and an R^2 of 0.02. This level of accuracy, despite being far from perfect, indicates that the method can operate in a novel site without the need for retraining on that site. Note however that the full exploitation of this capability would require data from multiple sites for training.
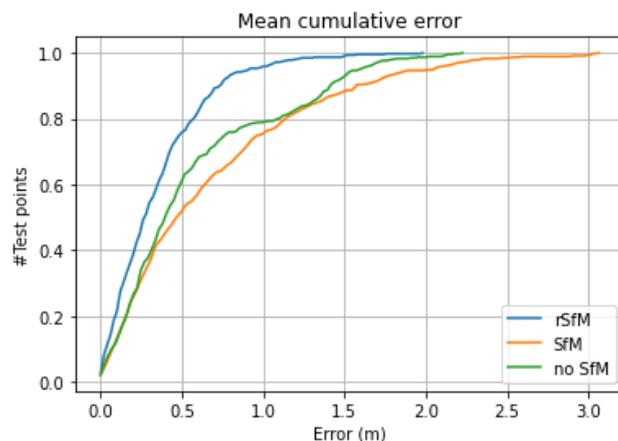
13

Figure 6. Cumulative error between the CNN model trained on RGB plus: R-SfM, simple SfM and without any SfM (RGB only) on Kalamaki dataset.
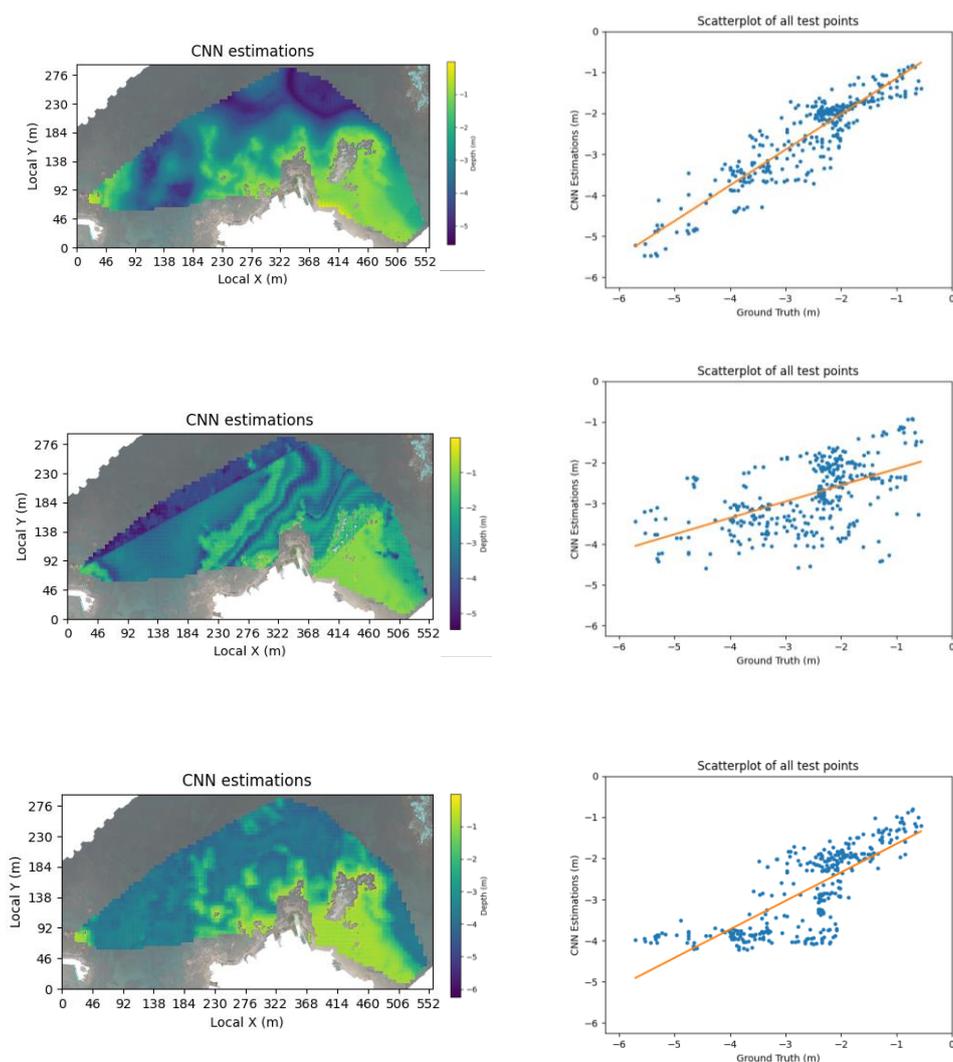

Figure 7. CNN R-SfM and RGB, CNN simple SfM and RGB [our previous paper], CNN RGB only. estimations and corresponding scatter plots of test points on Kalamaki dataset.
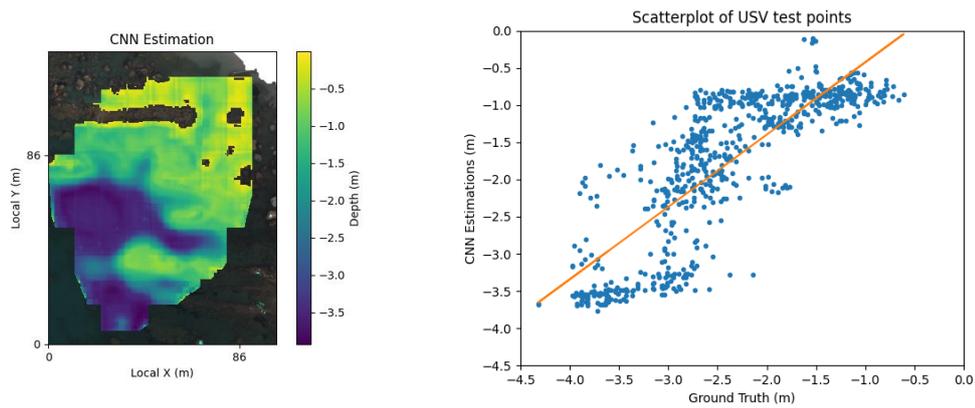
Figure 8. Bathymetry and scatter plot for our full pipeline, Trained: Kalamaki, Test: Plakias.
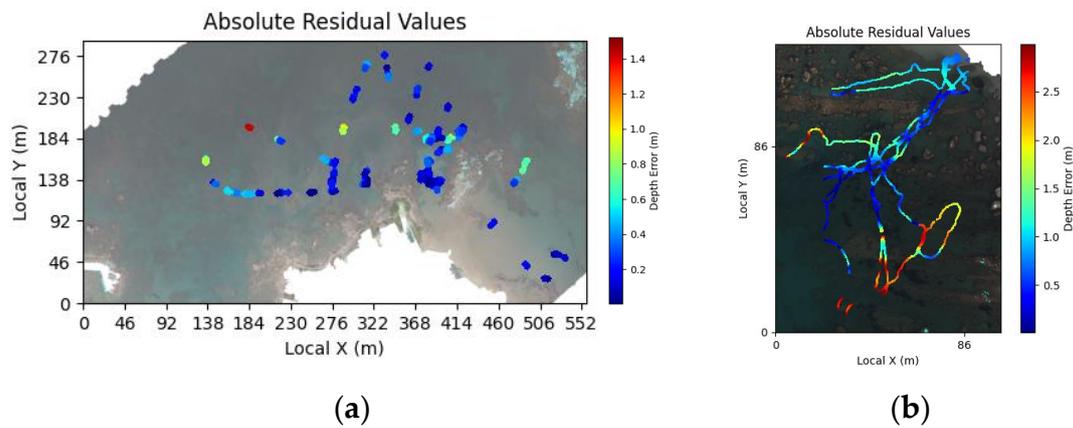

(a)   **(b)**

Figure 9. Absolute depth residuals for the points with USV measurements, achieved by our full pipeline. (**a**) Trained: Kalamaki train set, Test: Kalamaki test set. (**b**) Trained: Kalamaki whole dataset, Test: Plakias whole dataset.

4. Discussion

We developed and experimentally validated R-SfM, a shallow water sparse reconstruction approach that models the refraction effects. Experiments with simulated data demonstrate that R-SfM can be extremely accurate. Real-world validation of the method demonstrated its clear superiority over plain SfM. However, the nature of the real world data presented a challenge for R-SfM as well which can be seen from the much larger reconstruction errors compared to the simulated data. Camera pose estimation errors which accumulate in large reconstruction areas is the main cause for this discrepancy and we plan to address it in future work.

To further evaluate the quality of the proposed R-SfM, we employed the resulting R-SfM as a guiding role in a multi-channeled CNN framework, to demonstrate its beneficial usage over other types of data, such as regular SfM and RGB information. Our experiments revealed the superiority of the proposed R-SfM, and showed that the best results are achieved when R-SfM is combined with RGB data. The same pattern was observed when these data were also used in other ML models, further verifying their benefit.

In future work, we will consider the acquisition of a much larger dataset including multiple sites to account for the variability of the coastal areas. This will allow us to train a CNN that will be able to perform in new sites without the need for retraining. Furthermore, based on the statistics of this larger dataset, we plan to develop a methodology for realistic synthetic dataset generation. The latter

will provide ample data for CNN training, or even for larger architectures such as Vision Transformers (ViT) [54], resulting in more robust and accurate predictions.

Author Contributions: Conceptualization, A.R. and D.A.; methodology, A.R. and D.A., A.M., V.N., I.O.; software, A.M., V.N.; validation, A.M.., V.N.; formal analysis, A.M..; investigation, X.X.; resources, E.A., D.A.; data curation, E.A., A.M., V.N., A.R.; writing—original draft preparation, all authors; writing—review and editing, all authors; visualization, A.M.; supervision, A.R. and D.A.; project administration, A.R. and D.A.; funding acquisition, A.R., D.A. All authors have read and agreed to the published version of the manuscript.

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

References

1. Misra, A.; Vojinovic, Z.; Ramakrishnan, B.; Luijendijk, A.; Ranasinghe, R. Shallow Water Bathymetry Mapping Using Support Vector Machine (SVM) Technique and Multispectral Imagery. *Int. J. Remote Sens.* **2018**, *39*, 4431–4450, doi:10.1080/01431161.2017.1421796.

2. Sagawa, T.; Yamashita, Y.; Okumura, T.; Yamanokuchi, T. Satellite Derived Bathymetry Using Machine Learning and Multi-Temporal Satellite Images. *Remote Sens.* **2019**, *11*, 1155, doi:10.3390/rs11101155.

3. Alevizos, E.; Alexakis, D.D. Evaluation of Radiometric Calibration of Drone-Based Imagery for Improving Shallow Bathymetry Retrieval. *Remote Sens. Lett.* **2022**, *13*, 311–321, doi:10.1080/2150704X.2022.2030068.

4. Zhou, W.; Tang, Y.; Jing, W.; Li, Y.; Yang, J.; Deng, Y.; Zhang, Y. A Comparison of Machine Learning and Empirical Approaches for Deriving Bathymetry from Multispectral Imagery. *Remote Sens.* **2023**, *15*, 393, doi:10.3390/rs15020393.

5. Agrafiotis, P.; Skarlatos, D.; Georgopoulos, A.; Karantzalos, K. SHALLOW WATER BATHYMETRY MAPPING FROM UAV IMAGERY BASED ON MACHINE LEARNING. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *XLII-2/W10*, 9–16, doi:10.5194/isprs-archives-XLII-2-W10-9-2019.

6. Dietrich, J.T. Bathymetric Structure-from-Motion: Extracting Shallow Stream Bathymetry from Multi-View Stereo Photogrammetry. *Earth Surf. Process. Landf.* **2017**, *42*, 355–364, doi:10.1002/esp.4060.

7. Alevizos, E.; Nicodemou, V.C.; Makris, A.; Oikonomidis, I.; Roussos, A.; Alexakis, D.D. Integration of Photogrammetric and Spectral Techniques for Advanced Drone-Based Bathymetry Retrieval Using a Deep Learning Approach. *Remote Sens.* **2022**, *14*, 4160, doi:10.3390/rs14174160.

8. Slocum, R.K.; Parrish, C.E.; Simpson, C.H. Combined Geometric-Radiometric and Neural Network Approach to Shallow Bathymetric Mapping with UAS Imagery. *ISPRS J. Photogramm. Remote Sens.* **2020**, *169*, 351–363, doi:10.1016/J.ISPRSJPRS.2020.09.002.

9. Wang, E.; Li, D.; Wang, Z.; Cao, W.; Zhang, J.; Wang, J.; Zhang, H. Pixel-Level Bathymetry Mapping of Optically Shallow Water Areas by Combining Aerial RGB Video and Photogrammetry. *Geomorphology* **2024**, *449*, 109049, doi:10.1016/j.geomorph.2023.109049.

10. Mandlburger, G.; Kölle, · Michael; Nübel, · Hannes; Soergel, · Uwe BathyNet: A Deep Neural Network for Water Depth Mapping from Multispectral Aerial Images. **2021**, *1*, 71–89, doi:10.1007/s41064-021-00142-3.

11. Triggs, B.; McLauchlan, P.F.; Hartley, R.I.; Fitzgibbon, A.W. Bundle Adjustment — A Modern Synthesis. In *Vision Algorithms: Theory and Practice*; Triggs, B., Zisserman, A., Szeliski, R., Eds.; Lecture Notes in Computer Science; Springer Berlin Heidelberg: Berlin, Heidelberg, 2000; Vol. 1883, pp. 298–372 ISBN 978-3-540-67973-8.

12. Lourakis, M.I.A.; Argyros, A.A. SBA: A Software Package for Generic Sparse Bundle Adjustment. *ACM Trans. Math. Softw.* **2009**, *36*, 1–30, doi:10.1145/1486525.1486527.

13. Agrafiotis, P.; Skarlatos, D.; Georgopoulos, A.; Karantzalos, K. DepthLearn: Learning to Correct the Refraction on Point Clouds Derived from Aerial Imagery for Accurate Dense Shallow Water Bathymetry Based on SVMs-Fusion with LiDAR Point Clouds. *Remote Sens.* **2019**, *11*, 2225, doi:10.3390/rs11192225.

14. Agrafiotis, P.; Karantzalos, K.; Georgopoulos, A.; Skarlatos, D. Correcting Image Refraction: Towards Accurate Aerial Image-Based Bathymetry Mapping in Shallow Waters. *Remote Sens.* **2020**, *12*, doi:10.3390/rs12020322.

15. Lambert, S.E.; Parrish, C.E. Refraction Correction for Spectrally Derived Bathymetry Using UAS Imagery. *Remote Sens.* **2023**, *15*, 3635, doi:10.3390/rs15143635.

16. Cao, B.; Fang, Y.; Jiang, Z.; Gao, L.; Hu, H. Shallow Water Bathymetry from WorldView-2 Stereo Imagery Using Two-Media Photogrammetry. *Eur. J. Remote Sens.* **2019**, *52*, 506–521, doi:10.1080/22797254.2019.1658542.

17. Agrafiotis, P.; Karantzalos, K.; Georgopoulos, A.; Skarlatos, D. Learning from Synthetic Data: Enhancing Refraction Correction Accuracy for Airborne Image-Based Bathymetric Mapping of Shallow Coastal Waters. *PFG – J. Photogramm. Remote Sens. Geoinformation Sci.* **2021**, *89*, 91–109, doi:10.1007/s41064-021-00144-1.

18. Murase, T.; Tanaka, M.; Tani, T.; Miyashita, Y.; Ohkawa, N.; Ishiguro, S.; Suzuki, Y.; Kayanne, H.; Yamano, H. A Photogrammetric Correction Procedure for Light Refraction Effects at a Two-Medium Boundary. *Photogramm. Eng. Remote Sens.* **2007**, *73*, 1129–1136, doi:10.14358/PERS.73.9.1129.

19. Wimmer, M. Comparison of Active and Passive Optical Methods for Mapping River Bathymetry. Thesis, Wien, 2016.

20. Mandlburger, G. A CASE STUDY ON THROUGH-WATER DENSE IMAGE MATCHING. In Proceedings of the The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences; Copernicus GmbH, May 30 2018; Vol. XLII–2, pp. 659–666.

21. Cao, B.; Deng, R.; Zhu, S. Universal Algorithm for Water Depth Refraction Correction in Through-Water Stereo Remote Sensing. *Int. J. Appl. Earth Obs. Geoinformation* **2020**, *91*, 102108, doi:10.1016/j.jag.2020.102108.

22. David, C.G.; Kohl, N.; Casella, E.; Rovere, A.; Ballesteros, P.; Schlurmann, T. Structure-from-Motion on Shallow Reefs and Beaches: Potential and Limitations of Consumer-Grade Drones to Reconstruct Topography and Bathymetry. *Coral Reefs* **2021**, *40*, 835–851, doi:10.1007/s00338-021-02088-9.

23. Maas, H.-G. On the Accuracy Potential in Underwater/Multimedia Photogrammetry. *Sensors* **2015**, *15*, 18140–18152, doi:10.3390/s150818140.

24. Mulsow, C. A Flexible Multi-Media Bundle Approach. *Int Arch Photogramm Remote Sens Spat Inf Sci* **2010**, *38*, 472–477.

25. Mulsow, C.; Kenner, R.; Bühler, Y.; Stoffel, A.; Maas, H.-G. SUBAQUATIC DIGITAL ELEVATION MODELS FROM UAV-IMAGERY. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *42*.

26. Stumpf, R.P.; Holderied, K.; Sinclair, M. Determination of Water Depth with High-Resolution Satellite Imagery over Variable Bottom Types. *Limnol. Oceanogr.* **2003**, *48*, 547–556, doi:10.4319/lo.2003.48.1_part_2.0547.

27. Gholamalifard, M.; Kutser, T.; Esmaili-Sari, A.; Abkar, A.; Naimi, B. Remotely Sensed Empirical Modeling of Bathymetry in the Southeastern Caspian Sea. *Remote Sens.* **2013**, *5*, 2746–2762, doi:10.3390/rs5062746.

28. Liu, S.; Gao, Y.; Zheng, W.; Li, X. Performance of Two Neural Network Models in Bathymetry. *Remote Sens. Lett.* **2015**, *6*, 321–330, doi:10.1080/2150704X.2015.1034885.

29. Wang, L.; Liu, H.; Su, H.; Wang, J. Bathymetry Retrieval from Optical Images with Spatially Distributed Support Vector Machines. *GIScience Remote Sens.* **2019**, *56*, 323–337, doi:10.1080/15481603.2018.1538620.

30. Lumban-Gaol, Y.A.; Ohori, K.A.; Peters, R.Y. SATELLITE-DERIVED BATHYMETRY USING CONVOLUTIONAL NEURAL NETWORKS AND MULTISPECTRAL SENTINEL-2 IMAGES. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2021**, *XLIII-B3-2021*, 201–207, doi:10.5194/isprs-archives-XLIII-B3-2021-201-2021.

31. Rossi, L.; Mammi, I.; Pelliccia, F. UAV-Derived Multispectral Bathymetry. *Remote Sens.* **2020**, *12*, 3897, doi:10.3390/rs12233897.

32. Slocum, R.K.; Parrish, C.E.; Simpson, C.H. Combined Geometric-Radiometric and Neural Network Approach to Shallow Bathymetric Mapping with UAS Imagery. *ISPRS J. Photogramm. Remote Sens.* **2020**, *169*, 351–363, doi:10.1016/J.ISPRSJPRS.2020.09.002.

33. Al Najar, M.; Thoumyre, G.; Bergsma, E.W.J.; Almar, R.; Benshila, R.; Wilson, D.G. Satellite Derived Bathymetry Using Deep Learning. *Mach. Learn.* **2023**, *112*, 1107–1130, doi:10.1007/s10994-021-05977-w.

34. Benshila, R.; Thoumyre, G.; Najar, M.A.; Abessolo, G.; Almar, R.; Bergsma, E.; Hugonnard, G.; Labracherie, L.; Lavie, B.; Ragonneau, T.; et al. A Deep Learning Approach for Estimation of the Nearshore Bathymetry. *J. Coast. Res.* **2020**, *95*, 1011, doi:10.2112/SI95-197.1.

35. Chen, Q.; Wang, N.; Chen, Z. Simultaneous Mapping of Nearshore Bathymetry and Waves Based on Physics-Informed Deep Learning. *Coast. Eng.* **2023**, *183*, 104337, doi:10.1016/j.coastaleng.2023.104337.

36. Forghanioozroody, M.; Qian, Y.; Lee, J.; W. Farthing, M.; Hesser, T.; K. Kitanidis, P.; Darve, E. *Deep Learning Application to Large-Scale Bathymetry Estimation*; 2021;

37. Ghorbanidehno, H.; Lee, J.; Farthing, M.; Hesser, T.; Darve, E.F.; Kitanidis, P.K. Deep Learning Technique for Fast Inference of Large-Scale Riverine Bathymetry. *Adv. Water Resour.* **2021**, *147*, 103715, doi:10.1016/j.advwatres.2020.103715.

38. Jordt, A.; Köser, K.; Koch, R. Refractive 3D Reconstruction on Underwater Images. *Methods Oceanogr.* **2016**, *15–16*, 90–113, doi:10.1016/j.mio.2016.03.001.

39. Sonogashira, M.; Shonai, M.; Iiyama, M. High-Resolution Bathymetry by Deep-Learning-Based Image Superresolution. *PLOS ONE* **2020**, *15*, e0235487, doi:10.1371/journal.pone.0235487.

40. Jiang, S.; Jiang, C.; Jiang, W. Efficient Structure from Motion for Large-Scale UAV Images: A Review and a Comparison of SfM Tools. *ISPRS J. Photogramm. Remote Sens.* **2020**, *167*, 230–251, doi:10.1016/j.isprsjprs.2020.04.016.

41. Glaeser, G.; Schröcker, H.-P. *Reflections on Refractions*; 2000; Vol. 4, pp. 1–18;.

42. Mapillary OpenSfM 2022.

43. Adorjan, M. OpenSfM : A Collaborative Structure-from-Motion System. **2016**, 105 pages, doi:10.34726/HSS.2016.31424.

44. Mikolajczyk, K.; Tuytelaars, T.; Schmid, C.; Zisserman, A.; Matas, J.; Schaffalitzky, F.; Kadir, T.; Gool, L.V. A Comparison of Affine Region Detectors. *Int. J. Comput. Vis.* **2005**, *65*, 43–72, doi:10.1007/s11263-005-3848-x.

45. Tareen, S.A.K.; Saleem, Z. A Comparative Analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK. In Proceedings of the 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET); IEEE: Sukkur, March 2018; pp. 1–10.

46. Newell, A.; Yang, K.; Deng, J. Stacked Hourglass Networks for Human Pose Estimation. In *Computer Vision – ECCV 2016*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, 2016; Vol. 9912, pp. 483–499 ISBN 978-3-319-46483-1.

47. Nicodemou, V.C.; Oikonomidis, I.; Tzimiropoulos, G.; Argyros, A. Learning to Infer the Depth Map of a Hand from Its Color Image. In Proceedings of the 2020 International Joint Conference on Neural Networks (IJCNN); IEEE: Glasgow, United Kingdom, July 2020; pp. 1–8.

48. Jackson, A.S.; Bulat, A.; Argyriou, V.; Tzimiropoulos, G. Large Pose 3D Face Reconstruction from a Single Image via Direct Volumetric CNN Regression. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV); IEEE: Venice, October 2017; pp. 1031–1039.

49. Cho, H.-M.; Park, J.-W.; Lee, J.-S.; Han, S.-K. Assessment of the GNSS-RTK for Application in Precision Forest Operations. *Remote Sens.* **2023**, *16*, 148, doi:10.3390/rs16010148.

50. Mourtzas, N.; Kolaiti, E.; Anzidei, M. Vertical Land Movements and Sea Level Changes along the Coast of Crete (Greece) since Late Holocene. *Quat. Int.* **2016**, *401*, 43–70, doi:10.1016/j.quaint.2015.08.008.

51. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization 2017.

52. Recht, B.; Roelofs, R.; Schmidt, L.; Shankar, V. Do Imagenet Classifiers Generalize to Imagenet? In Proceedings of the International conference on machine learning; PMLR, 2019; pp. 5389–5400.

53. O'Mahony, N.; Campbell, S.; Carvalho, A.; Harapanahalli, S.; Hernandez, G.V.; Krpalkova, L.; Riordan, D.; Walsh, J. Deep Learning vs. Traditional Computer Vision. In *Advances in Computer Vision*; Arai, K., Kapoor, S., Eds.; Advances in Intelligent Systems and Computing; Springer International Publishing: Cham, 2020; Vol. 943, pp. 128–144 ISBN 978-3-030-17794-2.

54. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale 2021.