The best of two worlds: reprojecting 2D image annotations onto 3D models

Marin Marcillat[1]*, Loïc Van Audenhaege[1,2], Catherine Borremans[1], Aurelien Arnaubec[3], Lénaick Menot[1]

[1] BEEP/LEP/, Ifremer, Plouzane, Bretagne, France
[2] Ocean Bio Geosciences, National Oceanography Centre (NOC), Southampton, United Kingdom
[3] PRAO, Ifremer, Toulon, Provence-Alpes-Côte d'Azur, France

*Corresponding Author:
Marin.Marcillat@ifremer.fr

Abstract

Imagery has become one of the main data sources for investigating seascape spatial patterns. This is particularly true in deep-sea environments, which are only accessible with underwater vehicles. On the one hand, using collaborative web-based tools and machine learning algorithms, biological and geological features can now be massively annotated on 2D images with the support of experts. On the other hand, geomorphometrics such as slope or rugosity derived from 3D models built with structure from motion (sfm) methodology can then be used to answer spatial distribution questions. However, precise georeferencing of 2D annotations on 3D models has proven challenging for deep-sea images, due to a large mismatch between the raw navigation produced by underwater vehicles and the reprojected navigation resulting from bundle adjustment. In this article, we propose a streamlined, open-access processing pipeline to reproject 2D image annotations onto 3D models using ray tracing. Using four underwater image datasets, we evaluated the accuracy of annotation reprojection on 3D models and compared it to annotation geolocation available from the raw navigation. Features were georeferenced to centimetric accuracy, a 100-fold improvement over raw data geolocation. The combination of photogrammetric 3D models and accurate 2D annotations would allow the construction of a 3D representation of the landscape providing new insights into understanding species microdistribution and biotic interactions.

Introduction

The development of remote cameras, towed by research vessels or mounted on underwater platforms, was used early on for underwater exploration especially in the deep sea (e.g. Lonsdale et al. 1977). Compared to physical sampling of the fauna, imaging is non-intrusive and non-destructive, while allowing direct observation of the seabed over continuous areas (Tunnicliffe, 1990; Beisiegel et al., 2017). As a result, imaging has become a primary source of data to investigate across several scales of spatial patterns (i.e. 10s of m to kms) of seabed

Gelöscht: from
Gelöscht: inherited from
Gelöscht: inherited
Kommentiert [REV1]: Forgive me but your pipeline is not new, neither by approach, intent of use nor by the method. It has been implemented at least in the widely used software Agisoft Metashape for many years, where it is an integral part of the processing pipeline itself. What you CAN claim is that you provide an open access, if not open source alternative to this proprietary workflow.
Gelöscht: new
Gelöscht: 3D
Gelöscht: and could provide

geomorphology and benthic megafaunal communities, especially those poorly accessible and/or vulnerable (e.g. Girard et al., 2020; Marcon et al., 2014; Robert et al., 2017; van den Beld et al., 2017). Typical ecological investigations make use of geological and biological annotations made in images (Matabos et al., 2017; Schoening et al., 2017). This annotation task is complicated by the fact that, in most cases, the fauna cannot be identified down to the species level, making it susceptible to annotator bias (Durden et al., 2016). The recent development of image-based catalogues of fauna and seascape features (e.g. Althaus et al., 2013; Howell et al., 2020) and, the integration of these typologies into web-based annotation tools of 2D images has been widely used to mitigate identification bias by standardizing and remotely reviewing the categorization of large annotation sets (e.g., Langenkämper et al., 2017).

For spatial investigation, the compilation of those annotations into space requires coordinates typically measured by a georeferencing device (e.g. GPS for terrestrial studies; ref). However, the lack of precision of image positioning yielded by underwater vehicle navigation considerably lowers the spatial resolution at which deep-sea investigations are performed. Commonly, a submarine platform gets its position from the ship through an acoustic signal (usually the Ultra Short Baseline system, USBL), which precision decreases as a function of distance. Depending on the system used and its calibration, the accuracy may range from 1% to 0.1% of the slant distance. Consequently, at 1,000 m depth, the accuracy of the position is in the range of 1 m to 10 m. Recent advances in computer vision can provide access to an optimal repositioning (i.e. precision < 1 m) of underwater vehicles based on sequences of overlapping images. For instance, using photogrammetry, images can be re-assembled using feature-matching algorithm also refining *a posteriori* the position of the underwater camera. As a result, the relative positioning of features over the resulting 3D model of the seabed can be as precise as 1 cm or even less (Palmer et al., 2015; Istenič et al., 2019). Furthermore, a 3D model provides a numerical terrain model (NTM) of a centimetric to millimetric precision, enabling a high-resolution mapping of the seabed bathymetry from which geomorphometric descriptors, such as slope and rugosity, can be rapidly and quantitatively derived (Wilson et al., 2007; Gerdes et al., 2019). Those objective terrain metrics are especially of importance when considering them as driving ecological variables (Robert et al., 2017; Price et al., 2019). In addition, with the development of computer vision and photogrammetry, an overlapping image set allows reconstruction of underwater scenes in three dimensions (i.e., incuding overhangs and cavities), advancing seascape ecology from 2.5D to 3D (Kwasnitschka et al., 2013; Lepczyk et al., 2021).

However, the georeferencing system of 3D models conflicts with the acoustic-based relocation of annotations made on 2D images, hence producing two spatially incompatible datasets. A possible solution to cope with the mismatch between an acoustic-based positioning inherited from vehicle navigation and an optical-based positioning inherited from photogrammetry would still be to annotate the 3D model or the derived orthomosaics, instead of raw images. However, due to the additional reprojection step involved in their calculation, 3D models and associated orthomosaics typically have a slightly lower resolution than the raw 2D images, thus reducing the detectability of small organisms, and possibly biasing the observed community composition (Thornton et al.,

Kommentiert [REV2]: Reference missing

Gelöscht: inherited

Gelöscht: from

Kommentiert [REV3]: Its less about resolution (i.e., the precision that comes through the 2D imagery, but more about the accuracy of the observations. The resolution argument is more valid when comparing acoustic navigation to optical microbathymetry.

Gelöscht: whose

Gelöscht: with

Gelöscht: depth

Gelöscht: A

Kommentiert [REV4]: This part is unclear as it is missing a very important point that the paper further on neglects: Unlike e.g. swath echo sounding, structure from motion yields an internally consistent spatial network of camera poses and corresponding terrain features without the need of any external positional data. Nevertheless, it heavily depends both on the precise knowledge of intrinsic optical parameters and, to some degree especially if there is no info on IN SITU camera calibration under pressure, on the spatial distribution of features in 2D (and, thus 3D) space from which to infer or validate the intrinsics. The result is that, in reality, the uncertainty of the photogrammetric reconstruction may be on the order of meters. So: who is right, the acoustic nav or the photogrammetry? This is why you can only aim for internal consistence, and this is the important point of the exercise, not for absolute positioning.
As for positioning accuracy, given the acoustic nav, this is definitely governed by economies of scale: the larger the DTM, the generally larger the amount of nav samples and the smaller the positional error relative to the extent of the DTM. This is why there are fascinating effects connected with really large DTMs.

Kommentiert [REV5]: I believe the more common terminus technicus is digital terrain model, or DTM…

Gelöscht: from

Gelöscht: mathematically

Gelöscht: s

Gelöscht: explanatory

Gelöscht: for ecological investigation

Gelöscht: reconstructing

Gelöscht: lower the

Gelöscht: of

Gelöscht: , eventually flattened over the mesh

Gelöscht: hence

Gelöscht: ,

2016). Because we identified the lack of existing method to mutualize the benefit of 2D image annotation on web-based annotation tools and photogrammetric outcomes (i.e. optimal navigation and objective terrain descriptors), we propose an innovative workflow to transform 2D annotations in a 3D georeferencing system (Marcillat et al., 2023). This involves the development of a function allowing re-projection of annotations made in the open-access web-based image annotation tool Biigle (Langenkämper et al., 2017) onto 3D models produced with the freely available photogrammetry software Matisse3D (Arnaubec et al., 2023). The precision of the method has been assessed in different deep seascapes, with different camera settings and vehicles, and has been compared to an available georeferencing method from Biigle. Furthermore, we explored the use of this workflow for overlap management in image set building.

Materials & Methods

Study sites

Four datasets were used to assess the accuracy of annotation reprojection onto 3D models (Table 1). The four datasets represented different geological settings acquired using two different vehicles and cameras. Downward-looking cameras were mounted on the remotely operated vehicle (ROV) *Victor 6000* to map three hydrothermal vent sites and on the hybrid remotely operated vehicle (HROV) *Ariane* to map a cold-water coral (CWC) reef. The underwater vehicles were flown at a constant altitude to acquire parallel photo transects to map the respective seafloor structures. Tui Malila is a hydrothermal vent field surrounded by a complex basalt and is substratum located on a fast dorsal in the center of the Lau back-arc basin (South-West Pacific; ref). Eiffel Tower and White Castle are two vent edifices surrounded by a mild-slope terrain consisting mainly of slab and are located in the Lucky Strike vent field on the Mid-Atlantic Ridge (Ondréas et al., 2009). The most recent dataset was acquired in a cold-water coral reef located on a large (10s of meters) and mostly flat terrace in a submarine canyon of the Bay of Biscay. The reef consists of isolated colonies of *Madrepora oculata* growing on a matrix of dead coral infilled with soft sediments.

Table 1: Different datasets used during the reprojection error evaluation.

3D reconstructions and image annotations

For each of the four datasets, 3D models were reconstructed using Matisse3D (Arnaubec et al., 2023). Prior to reconstruction and annotation, the images were corrected for underwater attenuation and non-uniform illumination using the Matisse3D preprocessing module in order to improve feature matching outcome. For the reconstruction, images were also resized to 4 Mpx to cut down reconstruction speed.

Matisse3D performs feature detection and matching using the SIFT algorithm and removes outliers using the RANSAC model based on the fundamental matrix (see Arnaubec et al., 2015). Bundle adjustment then uses the images to reconstruct the 3D points detected by the SIFT algorithm. During bundle adjustment, the position of the camera relative to the scene is modelled

Gelöscht: new and

Kommentiert [REV6]: Please state for every site its xyz extent in meters so we get an idea of how large the sites are. I would like to read this here, not in a table which makes total sense for all the background info.

Gelöscht: active vent chimneys and periphery as well as the CWC reef…

Kommentiert [REV7]: This is a rock type, not a landform. Do you mean a lava field? To be safe, write: volcanic seafloor

Kommentiert [REV8]: Unclear, delete

Kommentiert [REV9]: This is totally unclear. Are you talking about fast spreading ridges? Or a horst structure? Better not get into geoscientific terminology here. ☺

Kommentiert [REV10]: Ref missing.

Kommentiert [REV11]: Again, from a geoscientific point, this is very unhappy wording. “On a gentle slope mainly made up of volcanic talus…”

Kommentiert [REV12]: Give precise extent

Gelöscht: improve

Gelöscht: modeled

and georeferenced by fitting these camera positions to those provided by the navigation system. The camera position computed during bundle adjustment is referred to here as the optical navigation. The number of points in the model is then increased using the dense matching method to create a dense cloud. Finally, the surface is reconstructed using the Poisson surface reconstruction algorithm. The resulting 3D models have an average resolution of 5 mm. Optical navigation can differ significantly from the raw navigation. Disjoint images were selected from the ROV dive on Tui Malila vent site using the disjoint mosaic function of Matisse3D (non-overlapping image georeferencing using raw navigation and altitude data). All disjoint images were annotated for fauna (e.g. on each image all recognizable individuals were tagged with points, and all patches were delimited with polygons) using the image annotation web service Biigle (Langenkämper et al., 2017).
We then investigated the effect of inaccuracies in the raw navigation on the quality of the annotations by estimating image overlap and annotation duplicates. For the purpose of reprojection error evaluation, a subset of 20 images was randomly selected in each of the four dataset and manually checked for non-overlap. On each image, four recognizable features were annotated with points using Biigle. Annotations were then exported from Biigle using the csv report scheme. These points are hereafter referred to as "ground control points", knowing that, unlike in terrestrial photogrammetry involving ground control points of precise georeferencing, our ground control points would actually be less precise than the corresponding DTM coordinates.

Annotation reprojection

The annotations were re-projected onto the 3D model using the camera position and rotation information (the extrinsic parameters) and the optical characteristics of the camera (the intrinsic parameters) resulting from the photogrammetric recunstruction. This information is stored in a *sfm_data* binary file generated by OpenMVG (Moulon et al., 2016), the 3D reconstruction library implemented in Matisse3D. For the reprojection of each camera line of sight we used the Blender python library (Blender Community, 2018) and the Blender Photogrammetry Importer (Bullinger, Bodensteiner & Arens, 2023). In this process, the annotation features were projected onto the 3D model using ray tracing (Figure 1). For each of these features, a ray is launched from the camera viewpoint towards the 3D model. The 3D reprojected position of a 2D feature corresponds to the first intersection between the ray and the model. If the reprojected annotation is a polygon, each vertex of the polygon is reprojected. If a ray does not hit the model, e.g. if the corresponding area in the image is not properly modeled, the annotation is discarded (Figure 1). A Python implementation of this process has been developed: (Marcillat et al., 2023).

Figure 1: Principle of annotations reprojection

Kommentiert [REV13]: A ref would be good, you reference all the other algorithms. Unlike us other noobs who use commercial software, you have the advantage of knowing what's under the hood!

Kommentiert [REV14]: Would you want to work with the terminus of ground surface density (GSD)?

Kommentiert [REV15]: This is your commonplace main argument but it does not work to lead to your next point. You want to say: In parallel, we prepared a data set entirely based on 2D imagery and raw acoustic navigation data.

Gelöscht: s,

Kommentiert [REV16]: Folks, this is exactly what you do in Metashape! The outcome is a shapefile, so don't even think of claiming a workflow into GIS software! ☺

Kommentiert [REV17]: New paragraph here!

Gelöscht: ¶

Gelöscht: ".

Gelöscht: from the 3D model

Gelöscht: intrinsic

Gelöscht: , OpenMVG (Moulon et al., 2016)

Gelöscht: ,

Gelöscht: the

Gelöscht: s

Gelöscht: were

Gelöscht: simulated using

Kommentiert [REV18]: Talking about pixels is misleading

Gelöscht: T

Gelöscht: pixels

Gelöscht: then

Gelöscht: pixels

Gelöscht: in the 3D scene containing

Gelöscht: the

Gelöscht: l

The local 2D uv coordinates of the annotation features in the images were thus transformed into global 3D coordinates (georeferenced using the WGS84 datum). The footprint of each image on the 3D model was also determined as 3D polygons by reprojecting the image corner coordinates in the same way as the annotation points.

Offset evaluation

To compare this method with the annotation geolocation available in Biigle (georeferencing based on the ROV's raw navigation, heading and elevation), annotation location reports of the ground controls were generated in Biigle. These geolocated positions were then compared to the reference positions on the 3D textured models, and the distance calculated.

Reprojection accuracy was assessed by identification of the ground control points within the textured 3D models. The distance between features on the textured 3D model and the same features after reprojection of the 2D images was determined calculating the Euclidean distance in 3D (Figure 2). The distance is used as a proxy for the error measurement of the annotation reprojection.

Figure 2: Error evaluation process.
Left: Annotation of a feature on the 2D raw image in Biigle. Right: Measurement of the reprojection error as the distance between the reprojected annotation point and the reference position of the feature on the 3D model.

Results

Raw versus optical navigation

The comparison between optical and raw navigation on the Tui Malila vent site highlights the discrepancy between raw image positioning based on raw navigation and 3D model positioning based on optical navigation (Figure 3). Transects that should be parallel and separated by a fixed distance are actually overlapping in some cases. Consequently, images that were supposedly non-overlapping according to the raw navigation actually overlapped on the 3D model (Figure 3), and 29% of the annotations were duplicated.

Figure 3: Impact of navigational inaccuracies.
(A) Discrepancy between raw navigation and optical navigation (from photogrammetry). (B) 2D disjoint mosaic obtained from raw navigation data. (C) Imprints of the same images reprojected onto the photogrammetric model. Disjoint images of the 2D mosaic actually overlap.

Figure 4 presents the summarized offsets of raw navigated versus photogrammetrically determined feature positions for all four data sets considered. The median distances were comparable among datasets, ranging from 2.35 m at White Castle to 4.67 m at the Coral Garden. The offsets were however variable within datasets. The interquartile ranges (IQR) varied from 2.57 m at Eiffel Tower to 3.28 m at the coral garden. The differences were unrelated to the

Gelöscht: s (in pixels, 2D)
Gelöscht: :
Gelöscht: , 3D
Gelöscht: imprint
Gelöscht: edge
Gelöscht: pixels
Kommentiert [REV19]: You compare two methods and data sets. This gives you an offset! They also both have an error, one of which (the photogrammetric error) you don't talk about because it can almost never be assessed precisely. State this, you are already in stage two of the review process so no one will stop you for admitting the truth. Coming back to the error of the usbl, you could qualitatively assess this by identifying the positional offset between identical features on overlapping tracks which you should have as you were mowing the lawn.
Gelöscht: Error
[1] verschoben (Einfügung)
Gelöscht: finding back
Gelöscht: annotations
Kommentiert [REV20]: Important but the whole paragraph is very unclearly written, I hope I understood you right.
Gelöscht: 3D
Gelöscht: measured
Gelöscht: using
[1] nach oben verschoben: To compare this method with the annotation geolocation available in Biigle (georeferencing based on the ROV's raw navigation, heading and elevation), annotation location reports of the ground controls were generated in Biigle. These geolocated positions were then
Gelöscht: a recognizable
Gelöscht: on the 3D textured model
Kommentiert [REV21]: It took me a night to understand ... [1]
Gelöscht: from the Chubacarc dive
Gelöscht: allows
Gelöscht: ing
Gelöscht: ,
Gelöscht: ,
Kommentiert [REV22]: Again, please provide an idea of ... [2]
Kommentiert [REV23]: Illustrate this in the figure by ... [3]
Gelöscht: For the four datasets, the distance between the ... [4]
Gelöscht: coral
Gelöscht: garden
Kommentiert [REV24]: What is the significance of this ... [5]

vehicle. The median distance and IQR were comparable at the coral garden, mapped with HROV, and the three vent sites mapped with the ROV.

Figure 4: Distribution of distances between annotations of ground control points on raw images and their reference positions on 3D models for different sites, according to two georeferencing methods: (A) geolocalisation based on raw navigation, and (B) reprojection based on optical navigation.

Annotation reprojections

Of the 320 ground control points that were annotated on the raw images to assess the precision of reprojections, 293 could be successfully reprojected onto the 3D model. Points that failed to reproject mostly corresponded to imperfections in the 3D models (17 missing points at Tui Malila, 5 at Eiffel Tower, 3 at the coral garden, and 2 at White Castle). For instance, where there are 'holes' (non-reconstructed areas in the model), light rays can pass through without hitting the model.

The distance between the reprojected ground controls points and their reference position on the 3D models emphasize the precision achieved by reprojection (Figure 4). The median distances at the coral garden (1.1*10e-2 m), White Castle (9.6*10e-3 m), and Eiffel Tower (1.3*10e-2 m) were similar, as well as their IQR, ranging from 1.2*10e-2 m at White Castle to 2.3*10e-2 m at Eiffel Tower. At Tui Malila, which is the topographically most complex site, the median distance and the IQR were higher, at respectively 8.6*10e-2 m and 2.1*10e-1 m.

Discussion

In the context of growing interest in 3D seafloor ecology (Lepczyk et al., 2021; Swanborn et al., 2022; Pulido Mantas et al., 2023), structure from motion (sfm) allows to reconstruct textured 3D models with centimeter resolution using underwater ROV/AUV imagery. Hereby, we present the first tool for reprojection of 2D annotation on 3D models. This tool implemented in a dedicated interface 'Chubacapp' provides multiple benefits for improving ecological investigations.

The resolution of 3D models allows generating small-scale geomorphometrics (e.g. slope, roughness, bathymetric position index…), which can provide new insights on species distribution patterns, especially in complex three-dimensional ecosystems and over continuous spatial scales (Price et al., 2019; Robert et al., 2020). For deep-sea studies, however, the use of high-resolution environmental variables necessitates the georeferencing of faunal observations with an equal precision which remains a challenge. Annotation georeferencing using raw navigation (using ROV position attitude and altitude) relies heavily on navigation precision. The distance we observed between raw-navigation and optical-navigation based positioning, with a median positioning error across datasets of 3.01 m, is coherent with the acoustic navigation precision. The median error does not vary much between datasets but variations can be as high as

Kommentiert [REV25]: What is the significance of this statement? Meaning the two vehicles performed similarly? Did you mow the lawn on all those data sets? You actually don't show the data sets anywhere in the paper, please correct this with an overvief figure showing the models with their corresponding camera poses/tracks. In the Matisse paper, you thankfully expand on the difference of lawnmower to 6 DOF surveys. How is this reflected in this study?

Kommentiert [REV26]: The way you describe it, I neither understand the figure nor the paragraph above that discusses it, even though your data might make total sense. Please make it much clearer what you mean by geolocalization and reprojection, and why are there two plots and not just one showing the summarized offsets between the two methods on each dive?

Kommentiert [REV27]: I believe you said this already.

Kommentiert [REV28]: No, the accuracy.

Kommentiert [REV29]: Unless the journal makes you use this notation, please write 0.011m

Gelöscht: most

Gelöscht: for

Gelöscht: seascape

Kommentiert [REV30]: This is more like a part of the abstract, the end of the intro or the conclusions. Not needed for discussion. Dive right in.

Gelöscht: However

Gelöscht: makes necessary

Kommentiert [REV31]: Again: isn't the term naïve, or a priori, better than raw, given that you always filter?

Gelöscht: that

Gelöscht: to reach for deep-sea studies

Gelöscht: geolocalisation

Gelöscht: based

3.28 m within datasets. With such an approximation in the positioning of features on the deep seafloor, abiotic and biotic interactions occurring at scales smaller than 10 m are out of our understanding.

Typically, as a result of the low navigation resolution, most seabed studies made use of annotations directly performed over continuous image sets either processed into textured 2D orthomosaics or 3D models in the case of complex environments (Pizarro & Singh, 2003; Marsh et al., 2012; Bodenmann et al., 2017; Simon-Lledó et al., 2019; Mitchell & Harris, 2020; Girard et al., 2020). On top of that, annotations of 3D models can be challenging due to the difficulty of displaying large high-resolution models and because of the longer duration required for drawing 3D geometries compared to 2D annotations. As a result, photogrammetry investigation typically focused on a subset of easily discernible organisms (e.g., Thornton et al., 2016) or on areas of a few 10s of m^2 (e.g., Lim et al., 2020; Mitchell & Harris, 2020). Moreover, during the texturing of photogrammetry outputs, images are distorted and the definition is lowered. Hence, smaller organisms may be missed, misidentified, and might require a time-consuming confirmation in the original image. To avoid biasing community composition insights, identification must be performed on the highest resolution of images and on an annotation platform allowing an easy and remote review by taxonomic experts (e.g., the largo tool in Biigle: Langenkämper et al., 2017).

As the tool developed in this study provided a satisfactory repositioning of annotations at the centimeter scale, it demonstrates its whole potential by combining the better of two worlds. 2D annotations of higher-resolution images allow an optimization of annotation georeferencing on 3D continuous models of the benthic habitat. 3D textured models allow an acceleration of the annotation process. Ultimately, this will facilitate exhaustive megafaunal community characterizations over large continuous spatial extents of 100s of m^2, as those large image sets are increasingly collected with autonomous underwater vehicles (Thornton et al., 2016).

It should be noted however that 3D models are not exempt from errors. In our datasets, 3% of image annotations could not be reprojected on 3D models due to model imperfections. Many improvements could be made to the quality of 3D models to avoid non-reprojected points. Manual transects can cause these gaps were the ROV is too fast or too high (or too low). "Survey" mode, the ROV autonomous guidance mode, has proven to operate more efficiently when mapping large areas in multiple transects, and should be generalized. However, this autopilot mode can be prone to Doppler Velocity Log (DVL) dropouts, particularly in complex terrains with large bathymetric variations. Topography can also drive erroneous reprojection of annotations over 3D models. In fact, the higher distance error (~m) observed for the Tui Malila site dataset could be related to a higher topographic complexity because of the rocky basaltic terrain exhibiting faults. Those faults could clearly limit the 'hit' of the ray traced from the downward-looking camera in near vertical setting, thus reprojecting the annotations a few decimeters away in the fault. In the near future, specific 3D mapping instruments such as LIDAR, multibeam scanning sonars or stereo camera will significantly improve the speed and

Gelöscht: lower

Kommentiert [REV32]: What is the point you are trying to make? The wording is unclear.

Kommentiert [REV33]: Yes, the lack of a powerful photogrammetric editor is indeed a big problem, but you could reference existing solutions such as potree, VRGS or commercial systems such as Metashape.

Gelöscht: From

Gelöscht: taxa

Gelöscht: , it

Gelöscht: s

Gelöscht: From textured

Gelöscht: , it

Gelöscht: s

Gelöscht: ed

Kommentiert [REV34]: In this paper, you are ignoring DVL as a data source altogether I am afraid. I trust you looked at Kwasnitschka 2013 for their pretty basic way of cross correcting DVL with USBL to arrive at a locally much more precise acoustic navigation track. It doesn't need Kalman Filtering and hardware wise it has been standard for three decades. I am not asking you to redo your entire analytics incorporating DVL, but I point out that you might have different results incorporating it in the a priori nav solution. Please find a way to address this that highlights the state of the art and doesn't invalidate your results.

Gelöscht: light

Gelöscht: hence

the accuracy of the reconstruction and the overall quality of 3D mapping, as well as providing an accurate optical navigation.

Still, one of the main bottlenecks in faunal imaging studies remains the annotation step of 2D images (Matabos et al., 2017). For image and video annotation, many online and collaborative tools have emerged (e.g. Biigle: Langenkämper et al., 2017, Squidle+: Bewley et al., 2015, VARS: Schlining & Stout, 2006), and the latest developments in assisted feature annotation have been integrated (Zurowietz et al., 2018). Citizen science platforms (Deep sea spy (Matabos et al., 2018), Zooniverse (Simpson, Page & De Roure, 2014)) also allow a significant increase in the amount of images processed. Our ability to adapt the workflow to Biigle demonstrates that in principle it is flexible against any of the annotation platform mentioned above. Furthermore, automated detection by machine learning looks very promising to speed up the process.

Gelöscht: recently

Gelöscht: with

Gelöscht: to

Although some experiments on automatic recognition of 3D features have been carried out (De Oliveira et al., 2021), most detection models remain actually developed for 2D images and videos (Katija et al., 2022). Reprojection may allow the use of well-proven generic 2D image detection Convoluted Neural Networks (CNN) for 3D annotation generation and vice-versa. On the one hand, 3D reprojection positions could generate 3D annotation sets for machine learning training. On the other hand, once a feature has been manually annotated on a single image and reprojected onto the 3D model, that 3D position can be used to locate that feature back on images taken from different angles. Multiple crops of the same object could then be generated and served as a training dataset.

Conclusions

This study demonstrated that 3D photogrammetry and reprojected annotations have several advantages when combined.

- Annotated images, using the popular and collaborative Biigle software, are raw and at full scale allowing optimal morphospecies categorization.
- Image overlap and annotation duplicates are avoided thanks to optical navigation. Annotations are precisely georeferenced to a centimetric scale.
- The 3D model provides access to high-resolution topographic metrics to quantify annotation distribution over several spatial scales. This technique could be applied on a very large scale to particularly complex terrains such as hydrothermal vents, canyons, cliffs, or coral reefs.
- We expect the developed workflow to considerably speed up the generation of annotation sets for 3D and for deep-learning purposes.

Kommentiert [REV35]: I don't fully follow this point. You annotate, then you validate against the DTM and afterwards you may filter for duplicates. But the duplicate annotation is made, unless you have a realtime workflow, which I understand you do not… Or?

Gelöscht: explain

Gelöscht: s

Gelöscht: continuous

Gelöscht: fasten

Supplementary Materials

The Python implementation of annotation reprojection, as well as other useful tools such as 3D topographic metrics calculation, disjoint image selection or model inference import into Biigle, are available on GitHub (Marcillat et al., 2023). Images, navigation data and 3D reconstructions are available on request.

Acknowledgements

We are grateful to all crews during the different sampling cruises for their help and support. We would also like to thank D. Langenkämper and M. Zurowietz for their support and the ongoing development of Biigle.

References

Althaus F, Hill N, Edwards L, Ferrari R. 2013. CATAMI Classification Scheme for scoring marine biota and substrata in underwater imagery - A pictorial guide to the Collaborative and Automation Tools for Analysis of Marine Imagery and Video (CATAMI) classification scheme, Version 1. *http://catami.org/classification*.

Arnaubec A, Ferrera M, Escartín J, Matabos M, Gracias N, Opderbecke J. 2023. Underwater 3D Reconstruction from Video or Still Imagery: Matisse and 3DMetrics Processing and Exploitation Software. *Journal of Marine Science and Engineering* 11:985. DOI: 10.3390/jmse11050985.

Arnaubec A, Opderbecke J, Allais A-G, Brignone L. 2015. Optical mapping with the ARIANE HROV at IFREMER: The MATISSE processing tool. In: *OCEANS 2015 - Genova*. 1–6. DOI: 10.1109/OCEANS-Genova.2015.7271713.

Beisiegel K, Darr A, Gogina M, Zettler ML. 2017. Benefits and shortcomings of non-destructive benthic imagery for monitoring hard-bottom habitats. *Marine Pollution Bulletin* 121:5–15. DOI: 10.1016/j.marpolbul.2017.04.009.

van den Beld IMJ, Bourillet J-F, Arnaud-Haond S, de Chambure L, Davies JS, Guillaumont B, Olu K, Menot L. 2017. Cold-Water Coral Habitats in Submarine Canyons of the Bay of Biscay. *Frontiers in Marine Science* 4. DOI: 10.3389/fmars.2017.00118.

Bewley M, Friedman A, Ferrari Legorreta R, Hill N, Hovey R, Barrett N, Pizarro O, Figueira W, Meyer L, Babcock R, Bellchambers L, Byrne M, Williams S. 2015. Australian sea-floor survey data, with images and expert annotations. *Scientific Data* 2. DOI: 10.1038/sdata.2015.57.

Blender Community online. 2018. *Blender - a 3D modelling and rendering package*. Stichting Blender Foundation, Amsterdam: Blender Foundation.

Bodenmann A, Thornton B, Nakajima R, Ura T. 2017. Methods for quantitative studies of seafloor hydrothermal systems using 3D visual reconstructions. *ROBOMECH Journal* 4:22. DOI: 10.1186/s40648-017-0091-5.

Bullinger S, Bodensteiner C, Arens M. 2023. A Photogrammetry-based Framework to Facilitate Image-based Modeling and Automatic Camera Tracking. In: 106–112.

De Oliveira LM, Lim A, Conti L, Wheeler A. 2021. 3D Classification of Cold-Water Coral Reefs: A Comparison of Classification Techniques for 3D Reconstructions of Cold-Water Coral Reefs and Seabed. *Frontiers in Marine Science* 8:640713. DOI: 10.3389/fmars.2021.640713.

Durden JM, Bett BJ, Schoening T, Morris KJ, Nattkemper TW, Ruhl HA. 2016. Comparison of image annotation data generated by multiple investigators for benthic ecology. *Marine Ecology Progress Series* 552:61–70.

Gerdes K, Martínez Arbizu P, Schwarz-Schampera U, Schwentner M, Kihara TC. 2019. Detailed Mapping of Hydrothermal Vent Fauna: A 3D Reconstruction Approach Based on Video Imagery. *Frontiers in Marine Science* 6.

Girard F, Sarrazin J, Arnaubec A, Cannat M, Sarradin P-M, Wheeler B, Matabos M. 2020. Currents and topography drive assemblage distribution on an active hydrothermal edifice. *Progress in Oceanography* 187:102397. DOI: 10.1016/j.pocean.2020.102397.

Howell KL, Davies JS, Allcock AL, Braga-Henriques A, Buhl-Mortensen P, Carreiro-Silva M, Dominguez-Carrió C, Durden JM, Foster NL, Game CA, Hitchin B, Horton T, Hosking B, Jones DOB, Mah C, Laguionie Marchais C, Menot L, Morato T, Pearman TRR, Piechaud N, Ross RE, Ruhl HA, Saeedi H, Stefanoudis PV, Taranto GH, Thompson MB, Taylor JR, Tyler P, Vad J, Victorero L, Vieira RP, Woodall LC, Xavier JR, Wagner D. 2020. A framework for the development of a global standardised marine taxon reference image database (SMarTaR-ID) to support image-based analyses. *PLOS ONE* 14:e0218904. DOI: 10.1371/journal.pone.0218904.

Istenič K, Gracias N, Arnaubec A, Escartín J, Garcia R. 2019. Scale Accuracy Evaluation of Image-Based 3D Reconstruction Strategies Using Laser Photogrammetry. *Remote Sensing* 11:2093. DOI: 10.3390/rs11182093.

Katija K, Orenstein E, Schlining B, Lundsten L, Barnard K, Sainz G, Boulais O, Cromwell M, Butler E, Woodward B, Bell KLC. 2022. FathomNet: A global image database for enabling artificial intelligence in the ocean. *Scientific Reports* 12:15914. DOI: 10.1038/s41598-022-19939-2.

Kwasnitschka T, Hansteen TH, Devey CW, Kutterolf S. 2013. Doing fieldwork on the seafloor: Photogrammetric techniques to yield 3D visual models from ROV video. *Computers & Geosciences* 52:218–226. DOI: 10.1016/j.cageo.2012.10.008.

Langenkämper D, Zurowietz M, Schoening T, Nattkemper TW. 2017. BIIGLE 2.0 - Browsing and Annotating Large Marine Image Collections. *Frontiers in Marine Science* 4. DOI: 10.3389/fmars.2017.00083.

Lepczyk CA, Wedding LM, Asner GP, Pittman SJ, Goulden T, Linderman MA, Gang J, Wright R. 2021. Advancing Landscape and Seascape Ecology from a 2D to a 3D Science. *BioScience* 71:596–608. DOI: 10.1093/biosci/biab001.

Lim A, Wheeler AJ, Price DM, O'Reilly L, Harris K, Conti LA. 2020. Influence of benthic currents on cold-water coral habitats: a combined benthic monitoring and 3D photogrammetric investigation. *Scientific Reports* 10:19433. DOI: 10.1038/s41598-020-76446-y.

Marcillat M, Menot L, Van Audenhaege L, Borremans C. 2023. Chubacapp, an open tool-box to process images, 3D models and annotation data from ROV and AUV. DOI: 10.5281/zenodo.7716971.

Marcon Y, Sahling H, Allais A-G, Bohrmann G, Olu K. 2014. Distribution and temporal variation of mega-fauna at the Regab pockmark (Northern Congo Fan), based on a comparison of videomosaics and geographic information systems analyses. *Marine Ecology* 35:77–95. DOI: 10.1111/maec.12056.

Marsh L, Copley JT, Huvenne VAI, Linse K, Reid WDK, Rogers AD, Sweeting CJ, Tyler PA. 2012. Microdistribution of faunal assemblages at deep-sea hydrothermal vents in the Southern Ocean. *PloS One* 7:e48348. DOI: 10.1371/journal.pone.0048348.

Matabos M, Borremans C, Bossard P, Tourolle J, Sarrazin J. 2018. Deep sea spy: a collaborative annotation tool. *Instrumentation Viewpoint*:41–42.

Matabos M, Hoeberechts M, Doya C, Aguzzi J, Nephin J, Reimchen TE, Leaver S, Marx RM, Branzan Albu A, Fier R, Fernandez-Arcaya U, Juniper SK. 2017. Expert, Crowd, Students or Algorithm: who holds the key to deep-sea imagery 'big data' processing? *Methods in Ecology and Evolution* 8:996–1004. DOI: 10.1111/2041-210X.12746.

Mitchell EG, Harris S. 2020. Mortality, Population and Community Dynamics of the Glass Sponge Dominated Community "The Forest of the Weird" From the Ridge Seamount, Johnston Atoll, Pacific Ocean. *Frontiers in Marine Science* 7. DOI: 10.3389/fmars.2020.565171.

Moulon P, Monasse P, Perrot R, Marlet R. 2016. OpenMVG: Open multiple view geometry. In: *International workshop on reproducible research in pattern recognition*. Springer, 60–74.

Ondréas H, Cannat M, Fouquet Y, Normand A, Sarradin PM, Sarrazin J. 2009. Recent volcanic events and the distribution of hydrothermal venting at the Lucky Strike hydrothermal field, Mid-Atlantic Ridge. *Geochemistry, Geophysics, Geosystems* 10. DOI: 10.1029/2008GC002171.

Palmer L, Franke K, Martin A, Sines B, Rollins K, Hedengren J. 2015. Application and Accuracy of Structure from Motion Computer Vision Models with Full-Scale Geotechnical Field Tests. *Geotechnical Special Publication*:2432–2441. DOI: 10.1061/9780784479087.225.

Pizarro O, Singh H. 2003. Toward large-area mosaicing for underwater scientific applications. *IEEE Journal of Oceanic Engineering* 28:651–672. DOI: 10.1109/joe.2003.819154.

Price DM, Robert K, Callaway A, Lo lacono C, Hall RA, Huvenne VAI. 2019. Using 3D photogrammetry from ROV video to quantify cold-water coral reef structural complexity and investigate its influence on biodiversity and community assemblage. *Coral Reefs* 38:1007–1021. DOI: 10.1007/s00338-019-01827-3.

Pulido Mantas T, Roveta C, Calcinai B, Di Camillo C, Gambardella C, Gregorin C, Coppari M, Marrocco T, Puce S, Riccardi A, Cerrano C. 2023. Photogrammetry, from the Land to the Sea and Beyond: A Unifying Approach to Study Terrestrial and Marine Environments. *Journal of Marine Science and Engineering* 11:759. DOI: 10.3390/jmse11040759.

Robert K, Huvenne VAI, Georgiopoulou A, Jones DOB, Marsh L, D. O. Carter G, Chaumillon L. 2017. New approaches to high-resolution mapping of marine vertical structures. *Scientific Reports* 7:9005. DOI: 10.1038/s41598-017-09382-z.

Robert K, Jones DO, Georgiopoulou A, Huvenne VA. 2020. Cold-water coral assemblages on vertical walls from the Northeast Atlantic. *Diversity and Distributions* 26:284–298.

Schlining BM, Stout NJ. 2006. MBARI's Video Annotation and Reference System. In: *OCEANS 2006*. 1–5. DOI: 10.1109/OCEANS.2006.306879.

Schoening T, Durden J, Preuss I, Albu AB, Purser A, Smet BD, Dominguez-Carrió C, Yesson C, Jonge D de, Lindsay D, Schulz J, Möller KO, Beisiegel K, Kuhnz L, Hoeberechts M, Piechaud N, Sharuga S, Treibitz T. 2017. Report on the Marine Imaging Workshop 2017. *Research Ideas and Outcomes* 3:e13820. DOI: 10.3897/rio.3.e13820.

Simon-Lledó E, Bett BJ, Huvenne VAI, Schoening T, Benoist NMA, Jones DOB. 2019. Ecology of a polymetallic nodule occurrence gradient: Implications for deep-sea mining. *Limnology and Oceanography* 64:1883–1894. DOI: 10.1002/lno.11157.

Simpson R, Page KR, De Roure D. 2014. Zooniverse: observing the world's largest citizen science platform. In: *Proceedings of the 23rd International Conference on World Wide Web*. WWW '14 Companion. New York, NY, USA: Association for Computing Machinery, 1049–1054. DOI: 10.1145/2567948.2579215.

Swanborn DJB, Huvenne VAI, Pittman SJ, Woodall LC. 2022. Bringing seascape ecology to the deep seabed: A review and framework for its application. *Limnology and Oceanography* 67:66–88. DOI: 10.1002/lno.11976.

Thornton B, Bodenmann A, Pizarro O, Williams SB, Friedman A, Nakajima R, Takai K, Motoki K, Watsuji T, Hirayama H, Matsui Y, Watanabe H, Ura T. 2016. Biometric assessment of deep-sea vent megabenthic communities using multi-resolution 3D image reconstructions. *Deep Sea Research Part I: Oceanographic Research Papers* 116:200–219. DOI: 10.1016/j.dsr.2016.08.009.

Tunnicliffe V. 1990. Observations on the effects of sampling on hydrothermal vent habitat and fauna of Axial Seamount, Juan de Fuca Ridge. *Journal of Geophysical Research: Solid Earth* 95:12961–12966. DOI: 10.1029/JB095iB08p12961.

Wilson MFJ, O'Connell B, Brown C, Guinan J, Grehan A. 2007. Multiscale Terrain Analysis of Multibeam Bathymetry Data for Habitat Mapping on the Continental Slope. *Marine Geodesy* 30:3–35. DOI: 10.1080/01490410701295962.

Zurowietz M, Langenkämper D, Hosking B, Ruhl HA, Nattkemper TW. 2018. MAIA—A machine learning assisted image annotation method for environmental monitoring and exploration. *PLOS ONE* 13:e0207498. DOI: 10.1371/journal.pone.0207498.

Kommentiert [REV36]: Some refs incomplete

Seite 5: [1] Kommentiert [REV21] Reviewer 14.11.23 10:29:00

It took me a night to understand this figure. Please choose a different representation for the 3D model, e.g. an oblique model rendering. Plot the photogrammetric and the raw navigated camera pose with their images as a sprite (the plugin lets you do that) over the model, show the projecting rays, and point out the Euclidian distance. This would be a highly citable figure, but the current one is misleading as it suggests we are comparing two sets of U/V coordinates.

Seite 5: [2] Kommentiert [REV22] Reviewer 14.11.23 10:36:00

Again, please provide an idea of the spread of the USBL data and of its systematic offset. It is obvious that there has been some processing done on the USBL data, but you don't state that. Please also see my comment below on DVL data

Seite 5: [3] Kommentiert [REV23] Reviewer 14.11.23 10:39:00

Illustrate this in the figure by linking a few features and labeling their offsets in meters

Seite 5: [4] Gelöscht Reviewer 14.11.23 10:41:00

Seite 5: [5] Kommentiert [REV24] Reviewer 14.11.23 10:44:00

What is the significance of this parameter?