Mini-UUV Cousteau II: Automated georeferencing of real-time 3D models from ORB-SLAM

Suitable for underwater scenarios using indirect GPS information

Michael BOMMHARDT-RICHTER, HTW Dresden, Germany
Marco BLOCK-BERLITZ, HTW Dresden, Germany
Vanessa BRÜLL, HTW Dresden, Germany

Abstract: The semi-autonomous control of the BlueRov2-based mini-submarine Cousteau II uses ORB-SLAM's real-time 3D reconstruction pipeline. Since the point clouds obtained by ORB-SLAM have only a relative reference, an automatic, indirect georeferencing method was developed for a complete mapping in UTM coordinates. Since the mini-submarine communicates with the base via a radio link, the method uses the GPS in-formation available in the buoy and derives the position and orientation of the mini-submarine from the current movement patterns and dive depths. The Cousteau II mini submarine is equipped with a downward looking sonar and an camera. The sonar is used to measure the distance to the bottom and the camera is used to obtain a 3D point cloud of the surroundings. The process starts by calibrating the position of the mini submarine with respect to the buoy. This is done by fitting a circle to the points observed by the mini submarine during a short calibration dive. The position of the mini-submarine is then estimated from the current location of the buoy and the orientation of the mini-submarine is estimated from the heading of the buoy. The estimated position and orientation of the mini-submarine are used to georeference the point clouds obtained by ORB-SLAM. This method is described in detail in the following paper.

Keywords: *Semi-Autonomous UUV—3D Reconstruction—Georeferencing Underwater—Self-Localization—UTM Coordinates*

CHNT Reference: Bommhardt-Richter, M., Block-Berlitz, M., and Brüll, V. (2022). 'Mini-UUV Cousteau II: Automatically georeferencing real-time 3D models from ORB-SLAM, in Börner, W., Rohland, H., Kral-Börner, C. and Karner, L. (eds.) *Proceedings of the 25th International Conference on Cultural Heritage and New Technologies, held online, November 2020.* Heidelberg: Propylaeum.
doi:10.11588/propylaeum.1045.c14493

Motivation and Introduction

The development of camera technology into high-resolution, compact, handy, and inexpensive devices has reached a high standard in recent years. Some of these cameras can be used in underwater applications. From videos and single image recordings, realistic 3D models can be reliably reconstructed by Structure-from-motion (SfM) methods (Pruno et al., 2015). Automated planning software is already available for multicopter applications (Block et al., 2018a). The costs for the combination of a multicopter with integrated camera and the corresponding free software for mission planning are far below 1000 Euros. There is currently no comparable solution in the underwater domain that allows automated, cost-effective underwater documentation. Indeed, some systems could perform such tasks at much higher costs (Yang et al., 2019).

However, due to the high price and partly also the actual size of the device, they are not suitable for every application. Especially in the underwater area, where the documentation method is still complicated and expensive (Moisan et al., 2015), the use of videogrammetry offers excellent advantages due to the low hardware usage and the flexibility and robustness of the method (Gehmlich and Block, 2015).

Within the framework of the SMWK-funded research project “Archaeonomous”, Manio, an unmanned underwater vehicle (UUV) has already been successfully developed, which can reliably record smaller underwater areas in an archaeological context semi-autonomously with various camera systems (see Fig. 1, left) (Block et al., 2018b; Bommhardt-Richter et al., 2018, Block et al., 2018c).

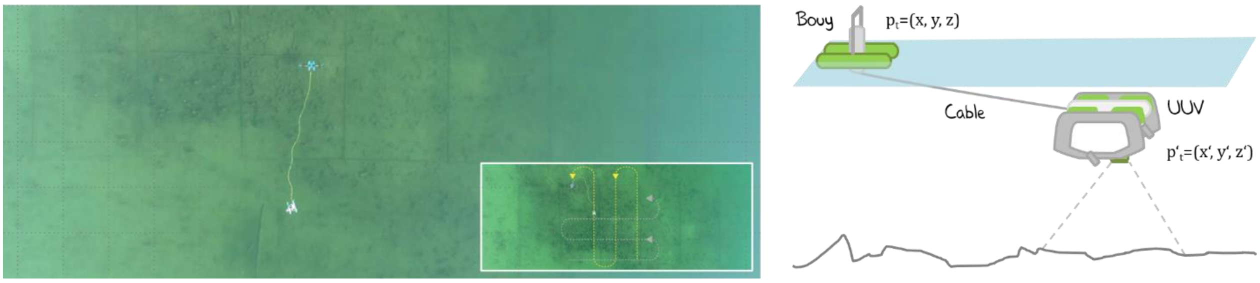


Fig. 1. Left: The Manio mini-submarine system was developed to record underwater areas in an archaeological context with cameras for later videogrammetric 3D reconstruction. For the base, a BlueROV2 was used, which was equipped with additional arms for the attachment of diving lamps and cameras and a self-developed radio buoy. Here we see Manio taking videos of a UNESCO World Heritage Site in Mondsee/Austria 2018 Right: The idea of automatic georeferencing is based on the simple idea that in a favourable situation during the recording, when the mini-submarine (UUV) is continuously moving in one direction for a particular time, the position from the buoy equipped with a GPS receiver can be used. These positions are given an exceptionally high weighting and stored accordingly. Based on the used and known cable length and the diving depth determined by the mini-submarine, the position and orientation of the UUV can be derived underwater. © Authors.

The automatic georeferencing is an important milestone in the localization of the surveyed underwater models, as GPS reception underwater is not directly possible. The idea presented in this article, an indirect solution is based on the positional relationship between the mini-submarine and the buoy at a constant speed (see Fig. 1, right).

The article is structured as follows: Section 2 introduces the theory and related work. Here the problems of georeferencing underwater and the current mini-submarine setup are explained. Section 3 introduces the method for indirect position determination, where selected keyframes are identified and then used for a transformation in UTM coordinates. In section 4, the first experiments and evaluations are presented.

Theory and related work

Reliable underwater georeferencing is still one of the biggest challenges in the 3D reconstruction of underwater areas. A solution, based on the indirect use of GPS information above water, and the structure of the current underwater system are briefly described below.

Georeferencing in underwater scenarios

Underwater georeferencing is still a challenge since underwater GPS data cannot be received directly. Long waves are the only electromagnetic waves that can propagate in water to a certain depth

(Tipler and Mosca, 2007). In the mini-submarine sector, therefore, only long waves with a shallow frequency of less than 30 kHz are used (Farr et al., 2010). Since GPS signals are emitted in the frequency spectrum above 1000 MHz, self-localization underwater is therefore only possible with indirect solutions.

For example, a GPS position measured above water can be derived from the position underwater by means of a stick (the length of which is known) (Scholz et al., 2016). Many different solutions were also developed and investigated in the Archaeonautic project (Block et al., 2018a).

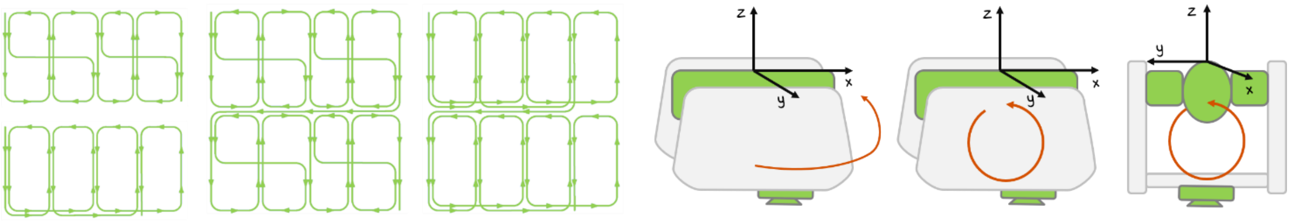


Fig. 2. Left: Here, different recording strategies can be seen, which have a high degree of potential loop closures and are based on classical recording strategies with multicopters (Wittchen 2017). Right: This figure shows the different measurable orientations of the mini-submarine along the respective coordinate axes. On the left, the yaw along the Z-axis is shown, which is particularly relevant for the determination of a straight track on the X/Y-plane. In the middle, is the rotation along the Y-axis, which may indicate a change in the diving depth. On the right, rolling represents the rotation along the X-axis. © Authors.

Real-time 3D reconstruction underwater

Real-time 3D reconstruction using Simultaneous Localization and Mapping (SLAM) provides a promising approach to solve the localization problem [8, 10]. The resulting point clouds can provide first information about the positioning underwater and the driven distances. Since the method depends on the detection of loop closures to minimize the error rate, the acquisition strategy must be designed accordingly (see Fig. 2, left) (Wittchen, 2017).

ORB-SLAM3 is used for real-time 3D reconstruction in the mini-submarine Cousteau-II (Campos et al., 2020). As an extension to the previous version ORB-SLAM2 (Mur-Artal, 2017), the ATLAS algorithm was developed for the management and manipulation of the generated point clouds and maps. As soon as the tracking is lost, the keyframes recorded up to that point are regarded as a separate sub-map, which is stored in the ATLAS (Elvira et al., 2020). During the next initialization, a new submap is created. If similarities between the features of different sub-maps are recognized, they are combined to a common map. Furthermore, ORB-SLAM3 now offers several camera models, which makes pre-calibration and distortion correction necessary (Block, 2020).

The processing chain in the UUV starts with images with the reduced resolution of 432 × 420 pixels by a GoPro Hero 4 BE as input for ORB-SLAM3 and takes a resolution of 1080p for the later video-based 3D reconstruction in Archaeo3D (Block et al., 2018c).

2.3 Camera calibration and automatic image enhancement

For the single images captured by the underwater cameras, which are extracted from the videos, the distortions must first be corrected. This is done once by a camera calibration step with the selected camera and resolution setup (Block, 2020).

For further processing of the underwater images, it is beneficial to send the images through an automated color adjustment and image enhancement process. With the help of the JEnhancer (Block et al., 2017A) and JFeatureManager (Block et al., 2017B) modules, the number of features identified is significantly increased, which is necessary for the reduced 1080p capture strategy and the partly low-contrast underwater shots. The further 3D reconstruction steps in COLMAP (Schönberger and Frahm, 2016) or VisualSFM (Wu, 2013) thus provide significantly better results.

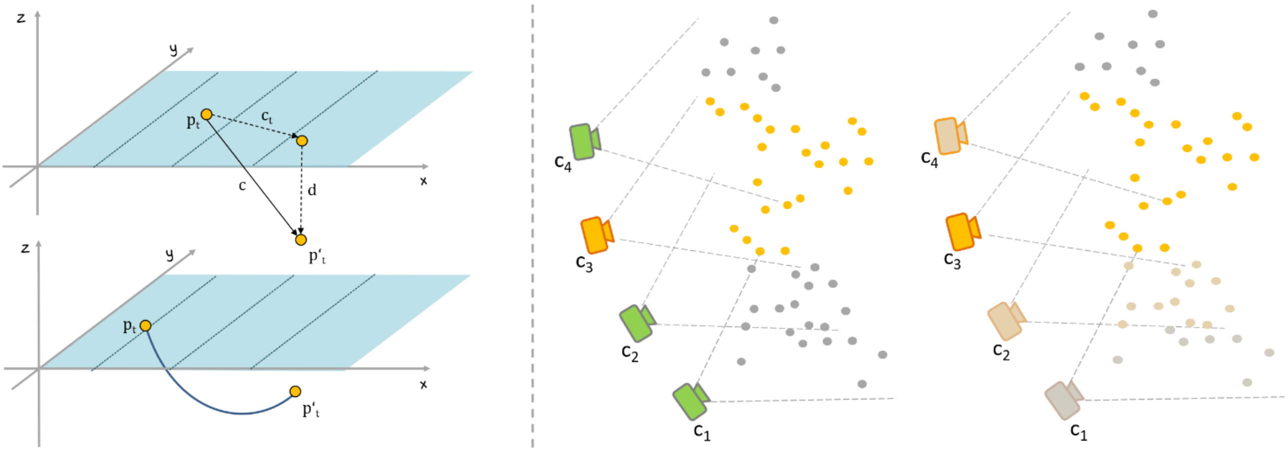


Fig. 3. Left above: The figure shows the model used for georeferencing. This describes an ideal application case in which the mini-submarine travels long enough at a constant speed in one direction to stretch the cable between the buoy and the UUV tight. Thus, the cable can be assumed to be a straight line between the two positions. Left below: Here, we see an alternative view where the cable is sagging in a resting state. This case can be better described with the catenary curve/model. Center/Right: The figure shows two examples of keyframe weighting and its visualization. The middle shows the currently pursued visualization, where the keyframes are classified binarily as good and bad. The good keyframe is highlighted in orange here. On the right, a variant with a stepwise weighting within an interval is shown, so that better keyframes are displayed with a stronger orange tone than worse ones that tend towards grey. © Authors.

Automatic georeferencing while recording

The mini-submarine Cousteau-II is equipped with a GoPro Hero 4 BE on the underside, which records videos in $1080p$ during the journey for later videogrammetric analysis. In parallel, a reduced video with a resolution of 432 × 420 pixels is delivered to the base station and serves as input for ORB-SLAM3. For the automatic georeferencing of the point cloud obtained in ORB-SLAM3, it is first necessary to select good keyframes the GPS position of which can be easily determined by the relative position to the buoy (Block et al., 2019).

Determining position and orientation relative to the buoy

The UUV based on the BlueRov2 has a relative position system and thus allows the collection of data on acceleration forces for the pitch, roll and yaw axes (see Fig. 2, right). The information obtained about the orientation of the UUV together with its speed is used in a model (see Fig. 3, left above) to determine the position of the mini-submarine relative to the known position of the buoy on the water surface. If the UUV is travelling at a dive depth d and a cable length of c with the attached buoy (see Fig. 1, right), we define a favourable time when the system has moved constantly in one direction for the time t_c. The time can be estimated with:

$$t_c = 2 \cdot \frac{\|\overrightarrow{v_t}\|}{\sqrt{c^2 - d_t^2}}$$

Currently, there is a horizontal distance c_t on the water surface between mini-submarine and buoy of:

$$c_t = \sqrt{c^2 - d_t^2}$$

With this distance c_t, the known position of the buoy $\overrightarrow{p_t}$ and the current speed of the mini-submarine (with buoy) $\overrightarrow{v_t}$ the position of the UUV can then be determined as follows:

$$\overrightarrow{p_t'} = \overrightarrow{p_t} + norm(\overrightarrow{v_t}) \cdot c_t$$

If the prerequisites that the system has been continuously moving in one direction for a time t_c are not given for this position determination, an alternative approach based on the catenary curve/model can be considered (Forster, 2017). In this case, the cable between mini-submarine and buoy describes a sagging chain (catenoid), the arc length of which can be described as follows:

$$l = \int_{x_1}^{x_2} \sqrt{1 + y'^2 dx}$$

Here x_1 and x_2 denote the x-coordinates of points P and P', respectively.

Weighting and prioritizing keyframes

The relative position determination of the mini-submarine and the resulting transformation of the map points into the UTM coordinate system depends on the conditions mentioned in the previous section. These conditions are used to weight the keyframes generated by ORB-SLAM3 to clarify which measured positions are to be preferred to obtain the most accurate reference. The weighting of the keyframes directly influences the weighting of the map points considered by the keyframes. The approach followed in this project is that every point that is considered by a keyframe to be good is also rated good (see Fig. 3, center/right).

This simplifying assumption disregards the number of camera positions from which a point is viewed, and how many of these positions are considered good. If these aspects are also included in the evaluation, it is advisable to use a finer graded evaluation within an interval of [0,1] to be able to represent this fine graduation of the weighting accordingly.

Selecting keyframes

The weighting of the keyframes takes place once during the integration of an image in ORB-SLAM3. To determine a good triangle for georeferencing, it is essential that the three selected keyframes span as large an area as possible (see Fig. 4, left).

We use the simple way to determine the area with the determinant or the sine formula in relation to a triangle $\Delta P_1 P_2 P_3$:

$$F_{\Delta P_1 P_2 P_3} = \frac{1}{2} \|(P_2 - P_1) \times (P_3 - P_1)\|$$

The keyframes that can now be considered can be found at the edge of the generated ORB-SLAM point cloud (see Fig. 4, right).

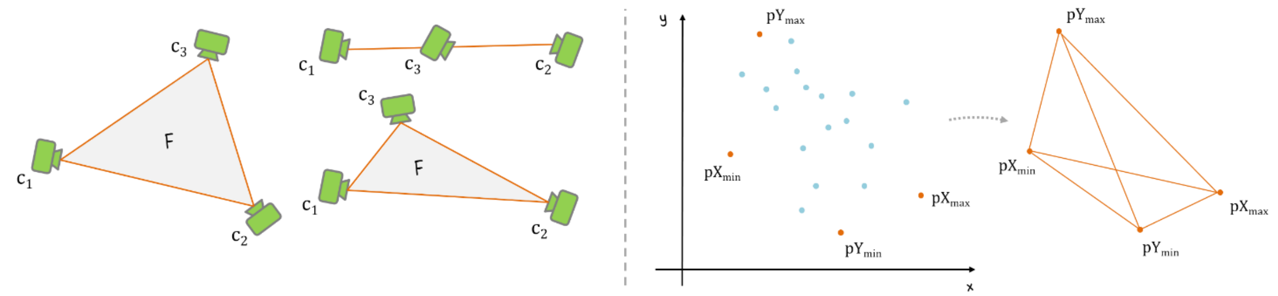


Fig. 4. Left: This figure shows the cases to be considered for the transformation of the map points. For the determination of the transformation matrix, three keyframes serve as vertices of a triangle, whereby the area of the said triangle should be as large as possible. In the worst case, the camera positions form a straight line, so that it can be assumed that the resulting transformation matrix leads to the largest distortions of map points in the UTM coordinate system. Right: The figure shows an example of the assumption used to determine the vertices for the triangle. The approach followed here is that the maxima/minima of the X/Y plane have the greatest distance to each other and thus three of the four points can span a triangle with the largest possible area. On the left side, are the relations of the points to each other, which have to be determined to define the corner points for the triangle. © Authors.

Prepare and perform georeferencing

As a starting point for the desired georeferencing, coordinates are available in two different reference systems. The camera positions generated by the ORB-SLAM algorithm at the time of acquisition and the associated tracked points are present in a three-dimensional, Cartesian coordinate system. The scaling in the monocular vision method can be defined arbitrarily at each initialization. In contrast, the positions of the buoy are given as latitude and longitude in decimal degrees in the geographic system.

To bring both into a comparable space for georeferencing, the Universal-Transverse-Mercator-Projection (UTM) (Volkmann, 2015) is used (see Fig. 5).

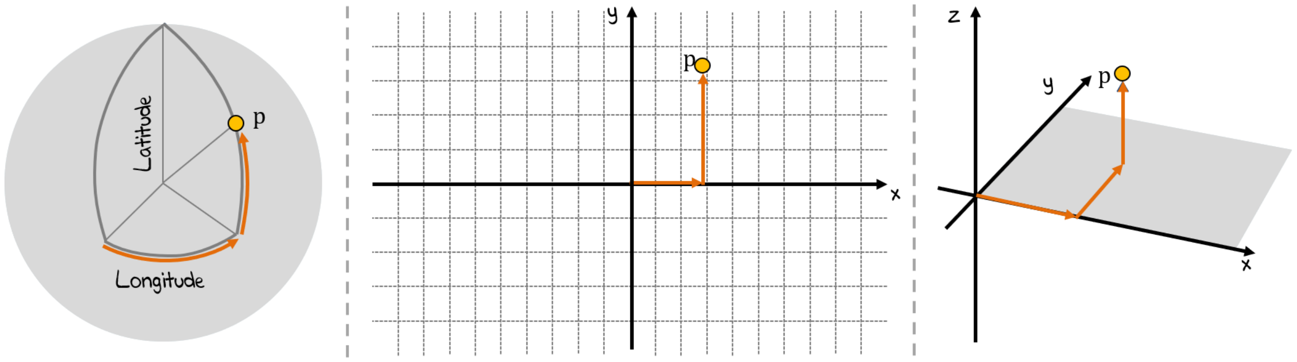
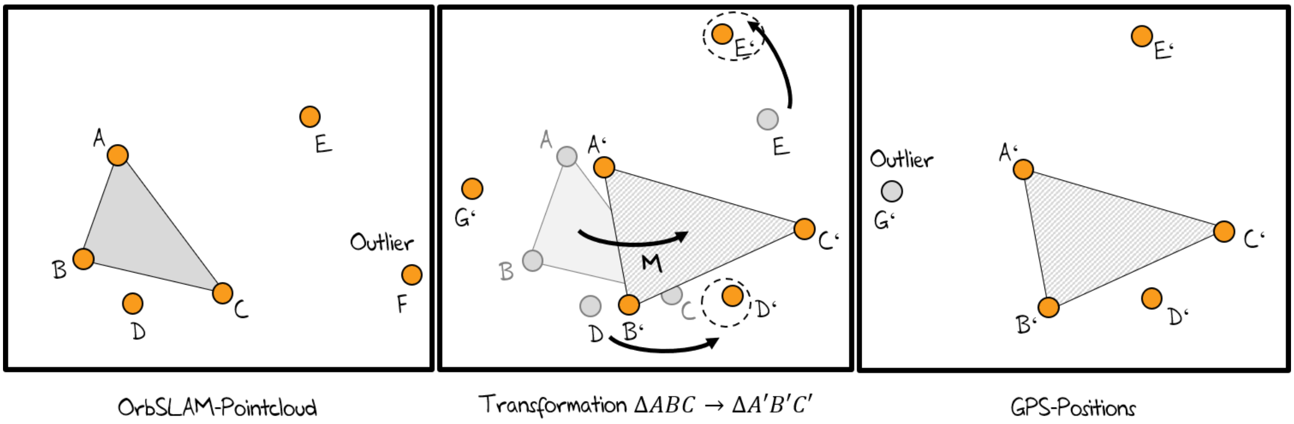


Fig. 5. Left: The geographic coordinate system, in which latitude and longitude of the position of the buoy are given. Both values are available as decimal degrees. The center of the earth or geoid, which is used for the mathematical representation of the earth, is also the coordinate origin. Center: The universal transverse Mercator projection, in which the earth is divided into individual zones. The X-axis lies on the equator and the Y-axis on the central meridian of the respective UTM zone with a false easting of 500.000 meters. The coordinates here are usually given in meters (Bill, 2016). Right: Example of a coordinate system generated by ORB-SLAM, which is the only one that automatically gives values on the Z-axis for each point compared to the other systems, in which this information has to be derived from an additional height value. Since the monocular vision method is used, the scaling of the map and thus the unit of the axes cannot be derived from the video data and must be arbitrarily defined at each initialization (Mur-Artal and Tardós, 2017). © Authors.



OrbSLAM-Pointcloud

Transformation $\Delta ABC \rightarrow \Delta A'B'C'$

GPS-Positions

Fig. 6. The figure shows an exemplary affine transformation of a triangle from one 2D space to another. On the left, we see the ORB-SLAM point cloud. Here the three good keyframe positions A, B and C are selected. On the right side, we see the corresponding GPS positions A', B' and C' of these keyframes. © Authors.

For the projection of the geographical data, it is essential to determine in advance in which UTM zone and on which hemisphere (northern or southern) the photographs will be taken, so that the correct $6°$ wide UTM zone can be selected for the projection and thus the projection distortions are as small as possible. After conversion, the coordinates of the buoy are in a Cartesian and metric representation, which simplifies the relative positioning of the mini-submarine.

A more significant challenge is the transformation of the ORB-SLAM coordinates into the UTM reference system since the exact parameters of the map generated by the algorithm are not selected until initialization. Since every keyframe stored in the map has a homogeneous pose, an affine transformation is used to determine a transformation matrix (see Fig. 6).

If the point set P, which comprises the positions of the keyframes of the previously determined triangle ΔABC, and the point set Q of the UTM positions $\Delta A'B'C'$ assigned to these keyframes are given relative to the buoy, the transformation can be determined according to the following solution using Singular Value Decomposition (Brunton and Kutz, 2019):

1. Find the centroid $\overrightarrow{c_P}$ of P and the centroid $\overrightarrow{c_Q}$ of Q.
2. Determine the optimal rotation $R = VU^T$ between the centered point sets using Singular Value Decomposition (SVD) given $[U, S, V] = SVD(H)$ and $H = (P - \overrightarrow{c_P})(Q - \overrightarrow{c_Q})^T$. The matrix H is called the covariance matrix.
3. Finally, the translation $\vec{t}$ must still be determined, with $\vec{t} = \overrightarrow{c_Q} - R \times \overrightarrow{c_P}$.

Now we can transform all ORB-Slam points $\vec{p} \in MP$ (including P) in the UTM coordinate system with $\overrightarrow{p'} = \left(R \cdot (\vec{p} - \overrightarrow{c_P})\right) + \overrightarrow{c_Q}$. The assumption is that the further away the map points are from the selected triangular area, the more their position is distorted by the transformation.

Experiments and evaluations

Several test runs were carried out in the reservoir in Oberwartha/Germany. The functionality of the newly mounted sensors with simultaneous recording should also be examined. A cable with a fixed length of $c = 1\,m$ was used between the buoy and the UUV. During the recordings, the positions of the buoy were recorded with the sensor U-blox NEO-6M GPS. This allows the position of the boat to be easily determined and stored at the keyframes in ORB-SLAM.

Out of a concrete sequence of 3,600 frames in the video, 445 keyframes with corresponding GPS positions were identified. The point cloud MP in ORB-Slam contained 29,207 points. As expected, the transformed (georeferenced) set of points is located correctly in the reservoir in Oberwartha (see Fig. 7).

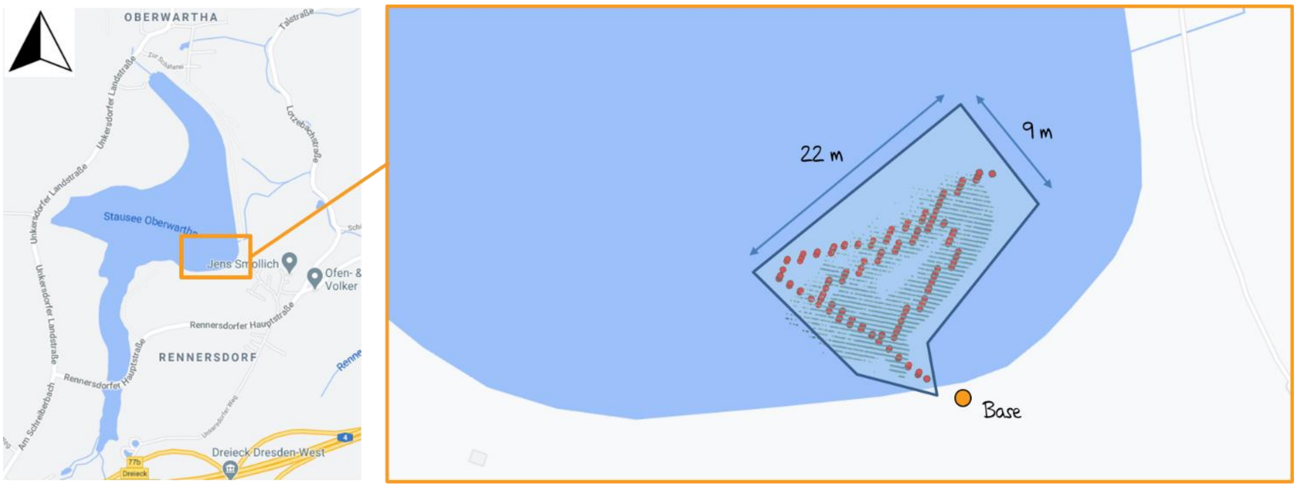


Fig. 7. Left: The reservoir in Oberwartha in Google Maps Layer with its approx. 30 ha offers a good opportunity to carry out different recording strategies. Right: A video sequence of 445 keyframes with the corresponding GPS positions (red dots) was extracted from a specific test dive. The resulting point cloud with 29,207 points (blue points) was successfully georeferenced afterwards by the presented method. © information Authors, Maps Layer from Google Maps.

5. Conclusion and future work

In this article, a robust method for automatic georeferencing of underwater point clouds is presented, which indirectly transfers the GPS information of the buoy on the surface into underwater coordinates. Currently, georeferencing is implemented as a final step of ORB-SLAM3 but shall be applied step by step during the real-time 3D reconstruction to visualize the resulting point cloud and the current or already sailed positions of the submarine on a map.

Funding

The initial project was funded by the Sächsisches Staatsministerium für Wissenschaft, Kultur und Tourismus and is now continued within a PhD-Scholarship of the European Social Fond in Germany.

Conflict of Interests Disclosure

The authors declared no conflict of interest.

Author Contributions

Conceptualization, Formal Analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing: Bommhardt-Richter M., Block-Berlitz M., Brüll V.
Data curation: Bommhardt-Richter M., Brüll V.
Funding acquisition: Bommhardt-Richter M., Block-Berlitz M.
Project Administration: Block-Berlitz M.
Supervision: Block-Berlitz M.
Validation: Bommhardt-Richter M., Brüll V.

References

Bill, R. (2016). *Grundlagen der Geo-Informationssysteme*, 6th edition. Offenbach-Berlin: Wichmann Verlag.

Block, M. (2020). *Warum sich der Dino furchtbar erschreckte: Beleuchtung und Rendering mit Java, LWJGL, OpenGL, OpenCV und GLSL*, Book, 2nd edition, Berlin: vividus Wissenschaftsverlag.

Block, M., Bommhardt-Richter, M., and Bochmann, H. (2019). ‘*Robust radio link solution for a semi-autonomous underwater vehicle*’, *24th Conference on Cultural Heritage and New Technologies (CHNT 24) in Vienna/Austria*.

Block M., Gehmlich B., Bommhardt-Richter M., Görsch, N., Bochmann, H., Haferland, A., Gunnarsson, K., Löw, C., and Dworsky, C. (2018a). ‘*Using Aerial Documentation for Underwater Georeferencing at UNESCO Pile Dwellings “See am Mondsee“ in Austria*’, Congress Visual Heritage, 23rd Conference on Cultural Heritage and New Technologies (CHNT 23) in Vienna/Austria.

Block, M., Bommhardt, M., Gehmlich, B., Wittchen, D., Burggraf, D., Zeisberg, S., Görsch, N., Gunnarsson, K., Bochmann, H., and Haferland, A. (2018b). *Mini-U-Boot Manio: Echtzeitbasierte, semi-autonome Unterwasserdokumentationen*, Technical Report, TR DD-2018-01, FIITS, HTW Dresden, Germany.

Block, M., Gehmlich, B., Bommhardt-Richter, M., and Bochmann, H. (2018c). ‘*Archaeological underwater documentations with Manio*’, Congress Visual Heritage, 23rd Conference on Cultural Heritage and New Technologies (CHNT 23) in Vienna/Austria.

Block, M., Dworsky, C., Löw, C., Gehmlich, B., Wittchen, D., Görsch, N., Suchowska, P., and Ducke, B. (2017a). *Underwater Videogrammetry with Adaptive Feature Detection at ‘See am Mondsee’, Austria*, in Journal of Studies in Digital Heritage, Indiana University, Vol. 1, No. 2, pp. 547–565.

Block, M., Gehmlich, B., and Hettmanczyk, D. (2017b). ‘*Automatic Underwater Image Enhancement using Improved Dark Channel Prior*’, in *Journal of Studies in Digital Heritage*, Indiana University, Vol. 1, No. 2, pp. 566–589.

Bommhardt-Richter, M., Block, M., and Gehmlich, B. (2018*). ‘ArchaeoPlanner: Planning semi-autonomous underwater campaigns*’, *Congress Visual Heritage, 23rd Conference on Cultural Heritage and New Technologies (CHNT 23) in Vienna/Austria*.

Brunton, S. L. and Kutz, J. N. (2019). *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*, Cambridge: University Press.

Campos, C., Elvira, R., Rodríguez, J. J. G., Montiel, J. M. M., and Tardós, J. D. (2020). *ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial and Multi-Map SLAM*, [preprint] arXiv:2007.11898.

Elvira, R., Montiel, J. M. M., and Tardós, J. D. (2020). *ORBSLAM-Atlas: a robust and accurate multi-map system*, IROS.

Farr, N., Bowen, A., War, J., Pontbriand, C., and Tivey, M. (2010). ‘*An integrated, underwater optical/acoustic communications system*’, *OCEANS 2010 IEEE-Sydney*, pp. 1–6.

Forster, O. (2017). *Analysis 2: Differentialrechnung im IRn, gewöhnliche Differentialgleichungen (Grundkurs Mathematik)*, 11. Auflage, Wiesbaden: Springer-Spektrum-Verlag.

Gehmlich, B. and Block, M. (2015). ‘*Diversity of Flight Strategies in UAV Recording*’, 20th International Conference on Cultural Heritage and New Technologies (CHNT 20), Vienna, Austria.

Moisan, E., Charbonnier, P., Foucher, P., Grussenmeyer, P., Guillemin, S., and Koehl, M. (2015*). ‘Building a 3D Reference Model for Canal Tunnel Surveying Using Sonar and Laser Scanning*’, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XL-5/W5, 2015 Underwater 3D Recording and Modeling.

Mur-Artal, R. and Tardós, J. D. (2017). *Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras*, IEEE Transactions on Robotics 33.5 (2017), pp. 1255–1262.

Pruno, E., Marcotulli, C., Vannini, G., and Drap, P. (2015). ‘Underwater Photogrammetry Methods for a Peculiar Case-Study: San Domenico (Prato-Italy)’, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* Volume XL-5/W5, 2015 Underwater 3D Recording and Modeling.

Scholz, R., Bekic, L., and Block, M. (2016). ‘*Geo-referencing from Photo-based Documentation in Freshwater and Marine Archaeology’, 21st Conference on Cultural Heritage and New Technologies (CHNT 21) in Vienna/Austria*.

Schönberger, J. L. and Frahm, J.-M. (2016). ‘*Structure-from-Motion Revisited’*, *Conference on Computer Vision and Pattern Recognition (CVPR).*

Tipler, P.A. and Mosca, G. (2007). *Physik. Für Wissenschaftler und Ingenieure*, Hrsg.: Dietrich Pelte. 2. Auflage. Spektrum akademischer Verlag.

Volkmann, A. (2015). *Archäologische Fundkartierung und Zeichnungs- Georeferenzierung mit QGIS*, Digital Humanities and Digital Cultural Heritage, Universität Heidelberg.

Wittchen, D. (2017). *Robuste Echtzeitlokalisierung von Unterwasseraufnahmen*, Diplomarbeit an der Hochschule für Technik und Wirtschaft Dresden im Bereich Künstliche Intelligenz.

Wu, C. (2013). *Towards Linear-time Incremental Structure from Motion*, 3DV 2013.

Yang, Yin, Chen Shuai, and Gui Fan (2019). ‘*The Technological Development and Prospect of Unmanned Surface Vessel’, IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC).*