Controlled experiment of underwater vision-based mapping: A preliminary evaluation

F Muhammad[1], Poerbandono[1], and H Sternberg[2]

[1]Hydrography Research Group, Faculty of Earth Sciences and Technology, Institut Teknologi Bandung, Jl. Ganesha 10, Bandung 40132, Indonesia
[2]Geodesy and Geoinformatics, Hafencity University of Hamburg, Henning Voscherau-Platz 1, 20457 Hamburg, Germany

E-mail: fickrie@gd.itb.ac.id, poerbandono@itb.ac.id, harald.sternberg@hcu-hamburg.de

Abstract. Underwater vision-based mapping (VbM) constructs three-dimensional (3D) map and robot position simultaneously out of a quasi-continuous structure from motion (SfM) method. It is the so-called simultaneous localization and mapping (SLAM), which might be beneficial for mapping of shallow seabed features as it is free from unnecessary parasitic returns which is found in sonar survey. This paper presents a discussion resulted from a small-scale testing of 3D underwater positioning task. We analyse the setting and performance of a standard web-camera, used for such a task, while fully submerged underwater. SLAM estimates the robot (i.e. camera) position from the constructed 3D map by reprojecting the detected features (points) to the camera scene. A marker-based camera calibration is used to eliminate refractions effect due to light propagation in water column. To analyse the positioning accuracy, a fiducial marker-based system –with millimetres accuracy of reprojection error– is used as a trajectory's true value (ground truth). Controlled experiment with a standard web-camera running with 30 fps (frame per-second) shows that such a system is capable to robustly performing underwater navigation task. Sub-metre accuracy is achieved utilizing at least 1 pose (1 Hz) every second.

1. Background

Underwater environment mapping known as bathymetric survey is important to provide information of seafloor topography at various details. Current technology employing high hydrographic echosounding equipment requires shipborne system to carry out such survey. However, in areas where shipborne system cannot enter non-navigable zones, such as very shallow water area, cave, or vulnerable archeological sites, an alternate system has to be considered. In accessing difficult underwater environments, underwater robotic systems have been developed to bring the sensor closer to the object for underwater mapping [1]. The underwater technology is rapidly being developed in recent decades for extending human capability in exploring ocean floor, inspection of human-made platform, marine conservation and many more. Underwater robot is classified to two types: Remotely Operated Vehicle (ROV) which is controlled by human with remote and Autonomous Underwater Vehicle (AUV) which has its own processing system on board for autonomous navigation [2].

The use of underwater robot is effective for monitoring underwater features, yet tends to deliver another issue, which is its difficulty in accurately positioning and mapping. As the robot travels in water medium, positioning with help of navigational satellite can no longer be used. This is due to the fact that

electromagnetic waves transmission from satellite attenuates very rapidly in water column. It is the particular intention that motivates the development of another techniques for underwater navigation [3]. At the present state, most of underwater mapping and navigation algorithms rely on acoustic sensor due to its ability to propagate efficiently in the water. Doppler velocity log (DVL), ultra-short baseline (USBL), Long baseline (LBL), and sonar are to name several notable methods [4]. This acoustic based techniques, however, pertain to parasitic backscattering effect due to presence of complex underwater features, such as vegetation and coral reefs [5].

In recent decades, visual sensor grows its recognition as an alternative approach for underwater mapping. This method is employing photogrammetry techniques to construct high accuracy three dimensional (3D) model up to centimeters level [6]. This approach yields to further approach e.g., object modelling, mapping, and identification known as Vision-based Mapping (VbM). Underwater VbM offers a beneficial solution for mapping of shallow seabed features as it is free from unnecessary parasitic returns which is found in sonar survey [7]. VbM is able to construct 3D model of detected object underwater by making quasi-continuous pictures from multiple known positions [8]. The principle of 3D model construction is based on structure from motion (SfM) photogrammetry, where the captured images are then used for estimating sensor position by projecting detected object to the camera scene [9]. In VbM, the SfM algorithm is applied throughout the system running or known as simultaneous localization and mapping (SLAM), where at the same time the system can estimate sensor location and mapping the environment detected by the sensor.

The underwater environment presents a big challenge to apply VbM, where bias is often detected due to the existence of suspended material such as microbial life forms or microparticle sediments [7]. The most common bias from the light transmission is light refraction due to light absorption and scattering phenomenon in water medium [10]. The light refraction effect may lead to the wrong measurement of depth, where the light transmission is bending towards the upper direction or lower direction of the seabed [11].

In this research, we investigate our work from using VbM algorithm to map the underwater object and use the mapped object to locate the position of the robot. Our work is designed for small-scale testing with only the preliminary adjustment of bias effect from light penetration underwater for VbM robustness is discussed. Our further research is aimed to tackle mentioned issue of VbM application underwater. We designed the environment for ideal condition, which means the effect of water clarity, current, and light exposure are not discussed in this research**.** The accurate knowledge of the underwater robot positioning and object mapping will benefit many useful applications besides charting such as conservation, inspection, 3D reconstruction, and surveillance.

2. Material and Methods

2.1. Experimental Setup

A wide-angle camera is needed to get wide area coverage of visual input. The use of wide-angle camera benefits the system to track more object to be used as feature extraction. In this research, we use the standard web-camera with 640 × 480 digital resolution, recording at 30 frame per second (fps), and 75° field of view (FOV). To display the image taken by camera to a consumer laptop, USB-cable connection is employed. The adaptation of underwater VbM is based on modified SLAM algorithm that can do robust navigation in outdoor environment [12].

To adapt the system in underwater environment, the camera shall be able to submerge in underwater condition. Hence we design a simple box housing made of plastic for deploying the camera. Figure 1 shows the general complex environment setup with some objects such as toys, fiducial (photogrammetric) marker, and camera calibration pattern are placed in the bottom of basin. That the small scale testing is aimed in this experiment, the dimension is constructed with 28 x 68 cm area having depth of 18 cm. To be noted that this setup is excluding underwater bias effect caused by light exposure and water properties. Thus, only underwater refraction effect due to light penetration will be discussed in this research.

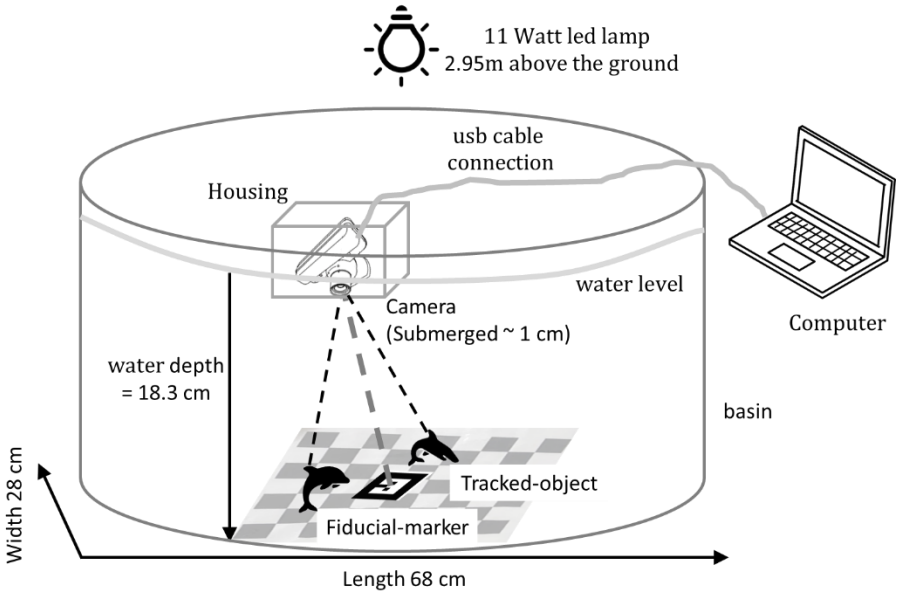

Figure 1. Environment setup for underwater VbM system testing.

2.2. Underwater Refraction

Challenge of using VbM in underwater condition is the refraction of light in water column that will produce distorted image. Prior to VbM system running, refraction adjustment shall be performed, which implies camera calibration techniques. One of the suitable techniques in performing refraction adjustment underwater is by applying Multiview camera calibration methods. This technique, however, requires a lot of images to be processed, so that all distortion from all angle and position can be computed. The multiview method extends the pinhole camera projection function which supports lenses exhibiting barrel radial distortion [13]. The radial distortion model transforms r to r', with the camera parameters of focal length (f_u, f_v), principal point or image coordinate of projection error ($\boldsymbol{c_x}$, $\boldsymbol{c_y}$) and distortion (ω) are assumed to be known parameters from the camera or so-called intrinsic parameters.

The object projected to the camera is formed by intersection of light rays from the objects through of lens (projection centre) [14]. The perspective camera projection should then consider the distortion parameter estimation which visualizes with radial r (distance) and tangential distortion r' (angle) factors. The equation expects to model the distortion parameters from selected sensors. The distortion parameters will later be used as camera calibration parameter input that has to be included in VbM system to get the correct feature point positions on image scene (u_D and v_D):

$$\begin{pmatrix} u_D \\ v_D \end{pmatrix} = \begin{pmatrix} u \\ v \end{pmatrix} + \begin{bmatrix} f_u & 0 \\ 0 & f_v \end{bmatrix} \frac{r'}{r} \begin{pmatrix} \frac{x}{z} \\ \frac{y}{z} \end{pmatrix} \tag{1}$$

where x, y, and z is coordinate of detected features (object).

The camera calibration underwater is challenging task where 1.9 pixels or root mean square error (RMSE) is obtained in this experiment. Figure 2 shows the calibration procedure, where camera submerged in water is necessary to measure the camera calibration parameters. The camera calibration parameters are obtained by scanning calibration checkboard underwater containing 11 × 7 square box with size of each square is 1.98 cm. After camera calibration procedure is succeeded, the virtual box with blue color shall show a perfectly fit alignment to the calibration checkboard. The misalignment of virtual box and calibration checkboard is indicating camera calibration error that defines as RMSE pixel.

The optimum RMSE to perform VbM task normally lies in range of 0.2 to 0.5 pixels. The large RMSE from camera calibration may result in underestimate or overestimate water depth due to the light propagation in water medium.

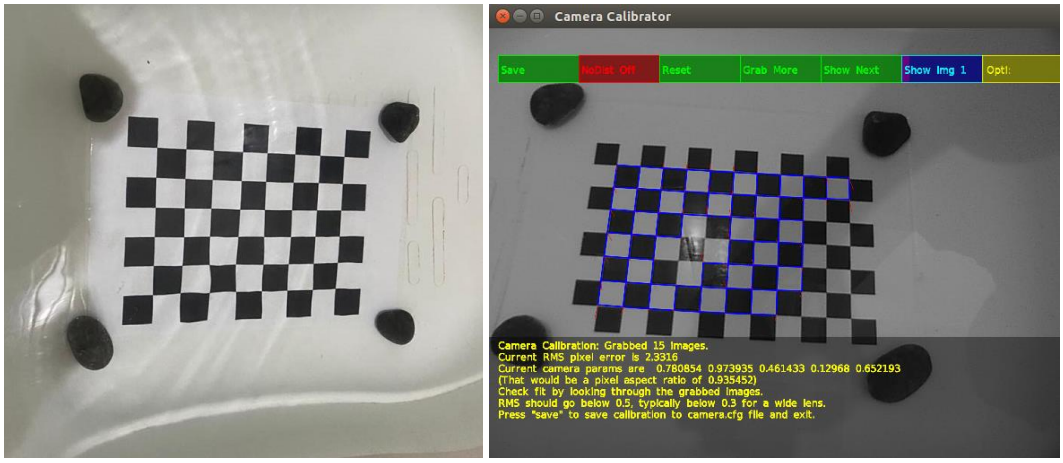

Figure 2. Illustration of calibration procedure with checkboard pattern (size of 11x7 square and 1.98 cm): before calibration procedure is performed (left) and after calibration procedure is performed (right).

2.3. Underwater VbM

In order to navigate vehicle movement underwater, precise measurement of vehicle attitudes such as translation and rotation is integrated into navigation algorithm. The algorithm implementation is based on calculating the first position X, predict the next position, update the position from visual features, detecting new features (landmark), then environment mapping [15], in such that:

$$X(t_n) = [x, \dot{x}, \ddot{x}, \alpha, \omega\ ] \tag{2}$$

Where body position x, velocity $\dot{x}$, acceleration $\ddot{x}$, linear acceleration α, and angular velocity ω are the parameters to estimate the final position [16]. These parameters form transformation parameters of rotation R and translation t matrix known as wrap function that connects initial camera scene to the next camera scene. Prior to 3D map construction, VbM shall perform initialization stage to estimate feature depth by capturing identical features known as feature-based method and detecting intensity values known as appearance-based method from two or more camera positions [17]. The appearance-based method on SLAM has been introduced by Semi-fast Visual Odometry (SVO) algorithm, where directly working with intensity (grey) values to identify a feature detected on-camera scene [18].

The 3D point p detection on every camera scene is connected with projection function T. The navigation estimation is based on SLAM algorithm that uses rigid body transformation to connect every camera scene from one position to the next position [19]. Figure 3 exhibits the underwater object being scanned and projected to the camera scene, where reprojection error needs to be minimized in each scene to estimate the accurate camera pose. The 3D point reprojection contains measurement noise that influences the image intersection to create 3D visualization. This error propagates from one camera scene to the next camera scene. The uncertaint of camera pose is then sum of uncertainty of camera pose and translation multiply with camera focal length, so that it will be always increasing in parallel of additional number of camera scenes taken:

$$C_n = f(C_{n-1}, T_n) \tag{3}$$

With C is camera pose uncertainty, T is Translation, f is focal length, and n is total data.To have better camera pose estimation, as described in [20] error metric ΔZ of feature detection from visual

recognition on every image scene needs to be minimized as can be seen in the following equation:

$$X = \sum_i \left\| x_i - T_i p_i \right\|^2 \Big|_{min\ x_i} \rightarrow X = \sum_i \|\Delta Z_i\|^2 \Big|_{min\ \Delta Z_i} \quad (4)$$

With every camera scene *i* correspondences with *X* as maximum likelihood of 3D points estimation in 2D image, *x* as the point coordinate in 2D image connecting with projection function *K* and *p* as the detected point from world environment. The connection of every camera scene is connected with wrap function consists of rotation *R* and translation *t*.

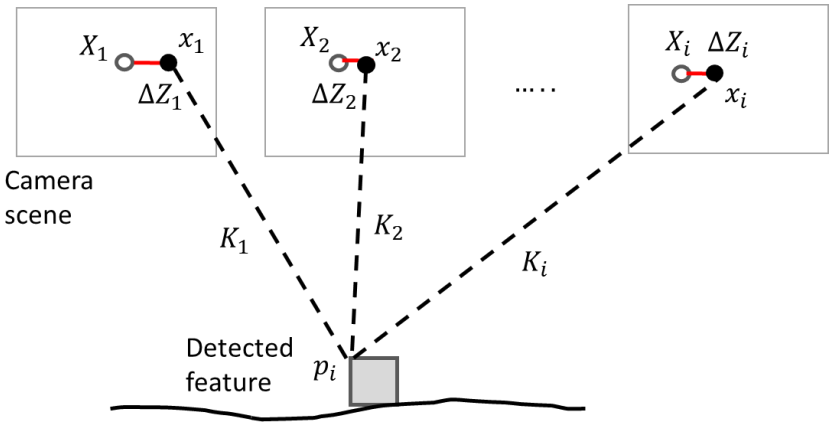

Figure 3. Illustration of feature tracking with VbM from multiple images: where *X* is 3D point projection based on feature point tracking *p*. Note: The projection of point *p* to camera scence has error metric ΔZ that has to be minimized in each camera position to have the best positioning accuracy.

The calculated camera position and mapped environment will be stored in local camera coordinate system which may provide data misinterpretation. In order to assess the VbM accuracy, true value of measurement shall be set. The true value or ground truth is based on fiducial (photogrammetric) marker measurement with known dimension. The dimension of fiducial marker is beneficial because it will provide metric-based coordinate frame with milimeter projectior error [21]. Figure 4 shows the scenario of system testing and relation between ground truth (marker frame) and VbM frame. In order to bring VbM data to marker coordinate system, the similarity transformation implying seven parameters is used [22]. Parameters used in similarity transformation consist of rotation *R* and translation *t* in x, y and z axis and scale *S* between one coordinate to other coordinate systems as can be seen in Eq. (5).

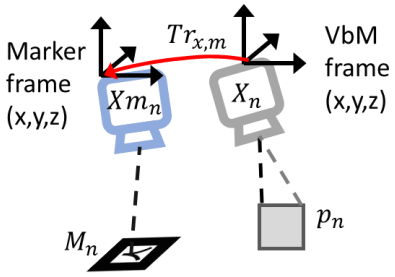

Figure 4. Illustration of marker measurement with single-camera: with *Xm* is camera position based on marker tracking *M*, *X* is camera position based on object point tracking *p*, *Tr* is transformation function for coordinate alignment between two systems.

$$X_c = S_{xyz} \, . \, R_{xyz} X_g + t_{xyz} \tag{5}$$

The transformation parameters is categorized as unknown parameters with known parameters are camera coordinate from marker tracking and camera coordinate from 3D object tracking. The least square method is suitable to calculate the unknwon parameters with accuracy of milimeter level [23]. The necessity of the same coordinate system is to have correct evaluation of VbM system. The evaluation is based on the calculation of root mean square error (RMSE) between camera coordinate *X* and marker coordinate *Xm*.

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(X - X_m)^2} \tag{6}$$

3. Results and Discussion

3.1. System Testing

The algorithm implementation and test are performed in underwater with rich features for visual detection. The test environment enables the system to run continuously for around 550 seconds long without losing its position. To scan prepared feature, the camera sensor is submerged in underwater condition up to circa 1 cm from the water level. The system's setting is running with a 75° wide-angle PS3-eye camera (webcam) and able to capture circa 2 frames per second (FPS).

Figure 5 shows the experiment result with the VbM system is able to extract around 2000 feature points. These feature points, however, contain a lot of noises/spikes mainly caused by the miscalculation of depth. Camera calibration result is one of the factors that causing the depth miscalculation as it is exploited by Anwer et all [24]. For visual comparison of the tracked feature points, the environment is scanned by time of flight (ToF) method that use RGB-Depth camera while the water basin has not been filled by water (dry condition). The visual comparison shows similarity of object detected that big object such as slope can be clearly visible, while small object such as toys needs to be further proceeded by filtering the point clouds. The use of ToF is challenging in underwater conditions, where infrared laser is not able to penetrate the water column so that the system is only suitable for dry environment at the moment.

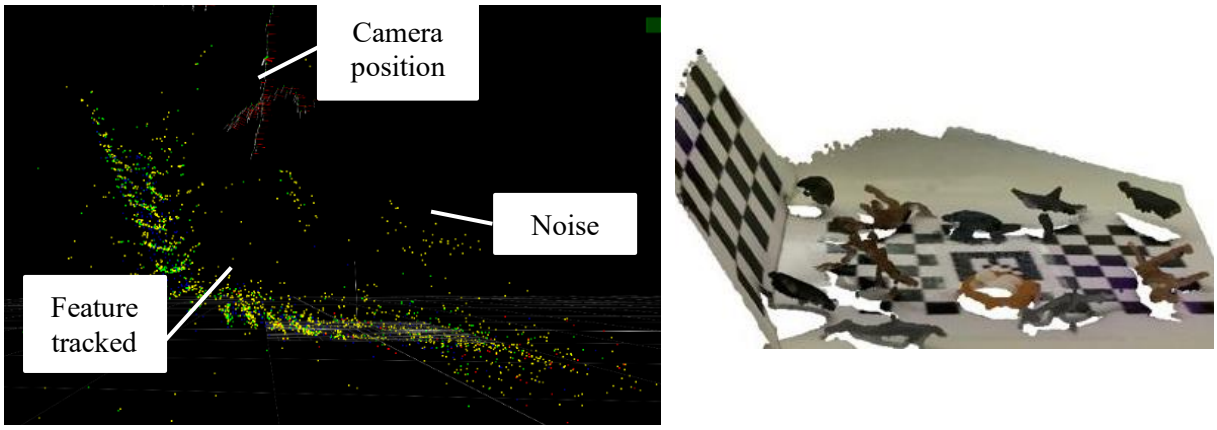

Figure 5. VbM feature tracking using SLAM (left) and VbM using ToF camera (right).

The trajectory comparison between VbM and ground truth can be seen in the Figure 6, where both systems show similar pattern of camera movement. The navigation using free feature tracking in VbM system shows more robust positioning in comparison with navigation with pre-programmed marker detection. In some scenarios, where marker body is not fully covered by camera scene, the system is unable to locate the camera that leads to position loss in navigating camera body. The system, however, is able to do relocalization by detecting marker position in the next camera scene. This issue is rarely happened in VbM due to its continuity of free feature tracking that is not relying only on pre-programmed marker for feature detection and sensor positioning.

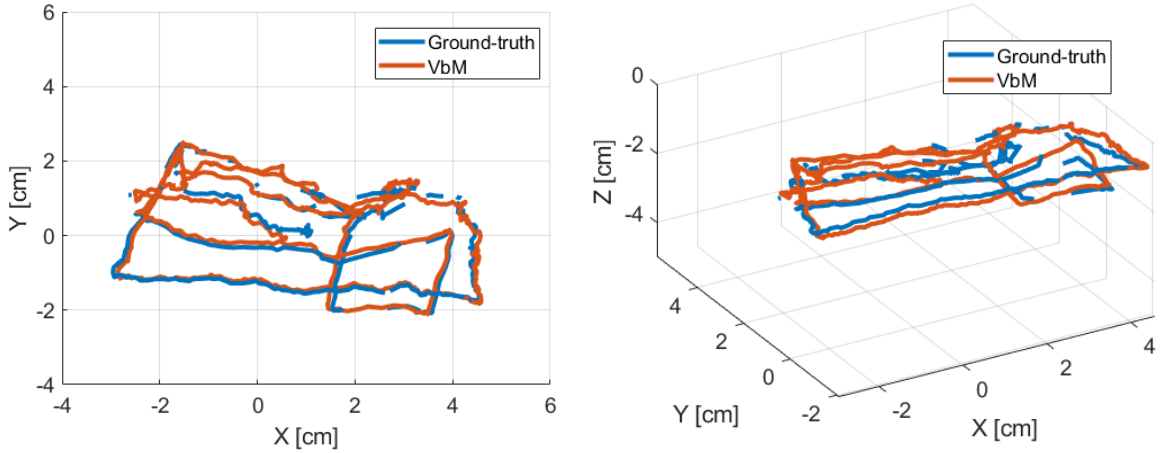

Figure 6. VbM trajectory from underwater testing against ground truth: views in 2D (left) and views in 3D (right).

3.2. Position Drift Analysis

The VbM creates map on local coordinate frame, which may lead to miscalling and misorientation of the camera scene trajectory. This misleading information related to global coordinate frame and scale is bound to presenting position drift over time to time in continuous tracking. the alignment of trajectory helps to adjust this issue, where around 150 frames out from around 1000 frames are used to delineate the transformation parameters. The illustration of position error of VbM comparing with ground truth position over time can be seen in Figure 7. Table 1 shows that the VbM system is able to perform small-scale navigation task with up to sub-centimeters error and RMSE of 0.11 cm, 0.15 cm, and 0.23 cm for

X, Y, and Z-axis respectively. The position drift showed as distance/radius indicates that the system is very stable during run-test with maximum drift is 0.71 cm.

Table 1. Summary of position drift (error)*.

	Max. (cm)	Min. (cm)	Mean (cm)	RMSE (cm)	σ (cm)
X	0.34	-0.28	0.02	0.11	0.11
Y	0.34	-0.41	-0.01	0.15	0.15
Z	0.33	-0.69	-0.11	0.23	0.21
Radius	0.71	0.04	0.27	0.09	0.12

*Notes: RMSE=Root Mean Squared Error, σ=standard deviation

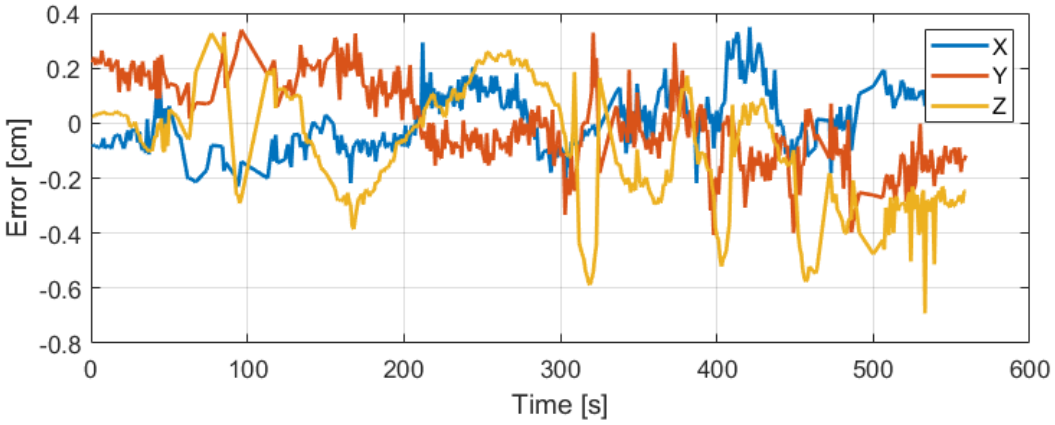

Figure 7. Position drift (error) from comparison of VbM and groundtruth.

4. Conclusion

In this presented paper, we are able to perform underwater test of an experimental VbM system. Such a system is constructed by 18 cm depth artificial basin having an area of 28 cm × 68 cm. The test works by submerging a standard low-cost web camera into the water to detect the prepared features. As it is shown by the experiment reported here, a digital camera having 640×480 pixel running at 30 fps is capable in performing navigational task by estimating the camera positions at secondly basis. The accuracy is within centimeters level. It should be noted that sufficient quantity of underwater landmarks are prepared. This enables the estimation of the feature points from its field of view. It seems that light propagation underwater plays major rule in transforming depth calculated by the system, as we still spot underestimated water depth of circa 2 cm. In addition to that, the large RMSE of the calibration result , i.e. up to 1.9 pixels, may play major role in underestimating water depth in our experiment.

5. Further Works

The use of VbM in underwater environment is an interesting topic to be discussed since it is offering the ability to create and store 3D object model with higher accuracy comparing to traditional sonar measurement. The underwater environment, however, presents a great challenge to apply VbM for large areas and long journey of measurement. The presence of noise due to light intensity change from one position to the next position may result in feature tracking loss. Another major issue is the water medium that often leads to miss-measurement of depth due to light propagation underwater. We are keen to work with this issue in the future by applying noise and bias correction, such as contrast adjustment and further underwater camera calibration, to make the system more robust. The integration of active sensor

such as laser distance measurement that found ToF camera will also provide direct depth measurement and 3D modeling for future works.

References

[1] G. Alfonso and E. Cuan-urquizo, "applied sciences Autonomous Underwater Vehicles : Localization , Navigation , and Communication for Collaborative Missions," 2020.
[2] M. Ferrera, "Monocular Visual-Inertial-Pressure Fusion for Underwater Localization and 3D Mapping.," *PhD Thesis Information, Struct. Syst. Univ. Montpellier*, 2019.
[3] P. Woock and C. Frey, "Deep-sea AUV navigation using side-scan sonar images and SLAM," *Ocean. IEEE Sydney, Ocean. 2010*, pp. 1–8, 2010, doi: 10.1109/OCEANSSYD.2010.5603528.
[4] S. Rahman, A. Q. Li, and I. Rekleitis, "SVIn2 : An Underwater SLAM System using Sonar , Visual , Inertial , and Depth SVIn2 : An Underwater SLAM System using Sonar , Visual , Inertial , and Depth Sensor," no. February 2020, 2019, doi: 10.1109/IROS40897.2019.8967703.
[5] P. Corke, C. Detweiler, M. Dunbabin, M. Hamilton, D. Rus, and I. Vasilescu, "Experiments with Underwater Robot Localization and Tracking," *Proc. ICRA. IEEE*, pp. 4556–4561, 2007.
[6] T. Van Damme, "Computer Vision Photogrammetry for underwater archaeological site recording in a low-visibility environment," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.*, vol. 40, no. 5W5, pp. 231–238, 2015, doi: 10.5194/isprsarchives-XL-5-W5-231-2015.
[7] F. Guth, L. Silveira, S. Botelho, P. D. Jr, and P. Ballester, "Underwater SLAM : Challenges , state of the art , algorithms and a new biologically-inspired approach," *5th IEEE RAS EMBS Int. Conf. Biomed. Robot. Biomechatronics*, 2014, doi: 10.1109/BIOROB.2014.6913908.
[8] D. S. Casagrande, "Real-Time Featureless Visual Odometry for Sea Floor Imaging with a Lagrangian Float," *Master Thesis Ocean Eng. Univ. Rhode Island, USA*, p. Paper 159, 2013.
[9] N. Micheletti, J. H. Chandler, and S. N. Lane, "Structure from Motion (SfM) Photogrammetry Photogrammetric heritage," vol. 2, pp. 1–12, 2015.
[10] H. Maas, "On the Accuracy Potential in Underwater/Multimedia Photogrammetry," *Sensors*, pp. 18140–18152, 2015, doi: 10.3390/s150818140.
[11] R. Rofallski and T. Luhmann, "Fusion von Sensoren mit optischer 3D-Messtechnik zur Positionierung von Unterwasserfahrzeugen," in *Hydrographie 2018 – Trend zu unbemannten Messsystemen*, Lindau, 2018, pp. 223–233.
[12] S. M. Weiss, "Vision Based Navigation for Micro Helicopters," *PhDThesis Electr. Eng. Information, ETH Zurich, Switz.*, no. Diss. ETH NO. 20305, 2012.
[13] F. E. E. Devernay and O. Faugeras, "Straight lines have to be straight: automatic calibration and removal of distortion from scenes of structured enviroments," *Mach. Vis. Appl.*, vol. 13, no. 1, pp. 14–24, 2001, doi: 10.1007/PL00013269>.
[14] D. Scaramuzza and F. Fraundorfer, "Visual Odometry part I," *IEEE Robot. Autom. Mag.*, no. December, 2011.
[15] F. Rothganger and M. Muguira, "SLAM Using Camera and IMU Sensors," *Sandia Rep.*, no. January, 2007.
[16] R. Jäger, "Geodetic Challenges on Precise Positioning & Navigation using GNSS Raw Data on Smartphones along the NAVKA-Disy RaD Project PREGON-X," *Lect. Notes Ref. Syst. Hochschule Karlsruhe - Tech. und Wirtschaft*, 2016.
[17] B. D. Scaramuzza and F. Fraundorfer, "Visual Odometry," *IEEE Robot. Autom. Mag.*, no. December, 2011, doi: 10.1109/MRA.2011.943233.
[18] C. Forster, M. Pizzoli, and D. Scaramuzza, "SVO : Fast Semi-Direct Monocular Visual Odometry," *Int. Conf. Robot. Autom.*, 2014, doi: 10.1109/ICRA.2014.6906584.
[19] D. Scaramuzza and F. Fraundorfer, "Tutorial: Visual odometry," *IEEE Robot. Autom. Mag.*, vol. 18, no. 4, pp. 80–92, 2011, doi: 10.1109/MRA.2011.943233.
[20] M. Pollefeys and T. Sattler, "3D Photography : Bundle Adjustment and SLAM Structure-From-Motion," no. March, 2014.

[21] H. Kato and M. Billinghurst, "Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System," *2nd Int. Work. Augment. Real. (IWAR 99).*, pp. 85–94, 1999, doi: 10.1109/IWAR.1999.803809.
[22] Reiner Jaeger, "Three-dimensional similarity transformation between two cartesian systems and state vectors of coordinates x1 and x2," *Lect. Notes Ref. Syst. Hochschule Karlsruhe - Tech. und Wirtschaft*, 2013.
[23] F. Muhammad, J. Zwiener, and R. Jaeger, "Evaluation and Software Implementations in Out-/Indoor Navigation based on Visual Odometry for Mono Cameras," Hochschule Karlsruhe, 2016.
[24] A. Anwer, S. Saad, A. Ali, S. Member, A. Khan, and F. Mériaudeau, "Underwater 3-D Scene Reconstruction Using Kinect v2 Based on Physical Models for Refraction and Time of Flight Correction," vol. 5, 2017.