Sonar Visual Inertial SLAM of Underwater Structures

Sharmin Rahman, Alberto Quattrini Li, and Ioannis Rekleitis

***Abstract*— This paper presents an extension to a state of the art Visual-Inertial state estimation package (OKVIS) in order to accommodate data from an underwater acoustic sensor. Mapping underwater structures is important in several fields, such as marine archaeology, search and rescue, resource management, hydrogeology, and speleology. Collecting the data, however, is a challenging, dangerous, and exhausting task. The underwater domain presents unique challenges in the quality of the visual data available; as such, augmenting the exteroceptive sensing with acoustic range data results in improved reconstructions of the underwater structures. Experimental results from underwater wrecks, an underwater cave, and a submerged bus demonstrate the performance of our approach.**

I. INTRODUCTION

This paper presents a real-time simultaneous localization and mapping (SLAM) algorithm for underwater structures combining visual data from a stereo camera, angular velocity and linear acceleration data from an Inertial Measurement Unit (IMU), and range data from a mechanical scanning sonar sensor.

Navigating and mapping around underwater structures is very challenging. Target domains include wrecks (ships, planes, and buses), underwater structures, such as bridges and dams, and underwater caves. The primary motivation of this work is the mapping of underwater caves where exploration by human divers is an extremely dangerous operation due to the harsh environment [1]. In addition to underwater vision constraints—e.g., light and color attenuation—cave environments suffer from the absence of natural illumination. Employing robotic technology to map caves would reduce the cognitive load of divers, who currently take manual measurements. The majority of underwater sensing for localization is based on acoustic sensors, such as ultrashort baseline (USBL) and Doppler Velocity Logger (DVL). However, such sensors are usually expensive and could possibly disturb divers and/or the environment. Furthermore, such sensors do not provide information about the structure of the environment.

In recent years, many vision-based state estimation algorithms have been developed using monocular, stereo, or multi-camera system for indoor, outdoor, and underwater environments. Such algorithms result in cheaper solutions for state estimation. Vision-based systems can be characterized as incremental, when there is no loop closure, termed Visual Odometry (VO) systems, and full vision-based SLAM systems [2].

The authors are with the Computer Science and Engineering Department, University of South Carolina, Columbia, SC, USA srahman@email.sc.edu, [albertoq,yiannisr]@cse.sc.edu

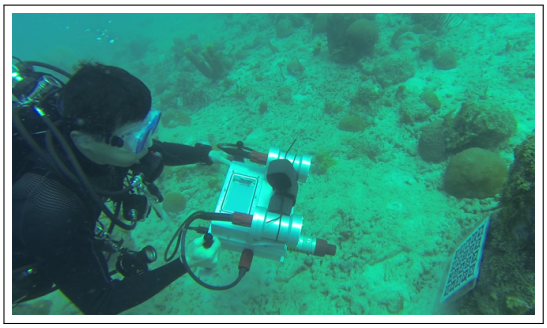


Fig. 1. The custom made sensor suite collecting data for the calibration of the visual, inertial, and acoustic range data.

Employing most of the available vision-based state estimation packages in the underwater domain is not straightforward due to many challenges. In particular, blurriness and light attenuation result in features that are not as clearly defined as above water. Consequently, different vision-based state estimation packages result in a significant number of outliers or complete tracking loss [3], [4]. In such a challenging environment, our preliminary work on using visual data and a video light for mapping an underwater cave [1] resulted in the successful reconstruction of a 250 meter long cave segment.

Vision can be combined with IMU and other sensors in the underwater domain for improved estimation of pose [5]. The open source package OKVIS [6] uses vision with IMU demonstrating superior performance. More recently, ORB-SLAM has been enriched with IMU [7] to recover scale for a monocular camera. In this paper, we propose a robust vision-based state estimation algorithm combining inertial measurements from IMU, stereo visual data, and range data from sonar, for underwater structure mapping domains.

Two general approaches have been employed for fusing inertial data into pure visual odometry. In the first approach, based on *filtering*, IMU measurements are used for state propagation while visual features are used for the update phase. The second approach, relying on *nonlinear optimization*, jointly optimizes all sensor states by minimizing both the IMU error term and the landmark reprojection error. Recent nonlinear optimization based Visual-Inertial Odometry (VIO) algorithms [6], [7] showed better accuracy over filtering approaches with comparable computational cost.

In this paper, a *tightly-coupled nonlinear optimization* method is employed to integrate IMU measurements with stereo vision and sonar data; see Fig. 1 for the underwater sensor suite used during calibration of both camera intrin-

 DOI: 10.1109/ICRA.2018.8460545

sics and extrinsics, required for good performance of VIO approaches.

The idea is that acoustic range data, though sparser, provides robust information about the presence of obstacles, where visual features reside; together with a more accurate estimate of scale. To fuse range data from sonar into the traditional VIO framework, we propose a new approach of taking a visual patch around each sonar point, and introduce extra constraints in the pose graph using the distance of the sonar point to the patch. The proposed method operates under the assumption that the visual-feature-based patch is small enough and approximately coplanar with the sonar point. The resulting pose-graph consists of a combination of visual features and sonar features. In addition, we adopt the principle of *keyframe*-based approaches to keep the graph sparse enough to enable real-time optimization. A particular challenge arises from the fact that the sonar features at an area are sensed after some time from the visual features due to the sensor suite configuration. Experimental data were collected from an artificial shipwreck in Barbados, the Ginnie ballroom cavern at High Springs, in Florida; and a submerged bus in North Carolina. In all cases, a custom sensor suite employing a stereo camera, a mechanical scanning profiling sonar, and an IMU was used.

The paper is structured as follows. The next section outlines related work. Section III present an overview of the experimental system used. Section IV describes the proposed method with special focus on the mathematical formulation for introducing sonar data in the OKVIS framework. Section V presents experimental results on datasets we collected in different underwater structures, validating the proposed method. The paper concludes with a discussion on lessons learned and directions of future work.

II. RELATED WORK

Compared to above water visual odometry techniques where GPS might be available (e.g., [8], [9]), visual odometry in underwater cave environment is a challenging problem due to the lack of natural light illumination and dynamic obstacles in addition to the underwater vision constraints i.e. light and color attenuation. There are not many works for mapping and localization in an underwater cave. Gary et al. [10] presented a 3D model of underwater cave using LIDAR and sonar data collected by DEPTHX (DEep Phreatic THermal eXplorer) vehicle having DVL, IMU, and a depth sensor for underwater navigation. Most of the underwater navigation algorithms [11]–[16] are based on acoustic sensors such as DVL, USBL, and sonar. Nevertheless, collecting data using DVL, sonar, and USBL while diving is expensive and sometimes not suitable in cave environment. In this context, vision-based state estimation could be used as it is cheaper and easily deployable; however, because of its incremental motion-based nature, it accumulates drift over time. Corke et al. [17] compared acoustic and visual methods for underwater localization showing the viability of using visual methods underwater in some scenarios.

In recent years many vision-based state estimation algorithms have been developed using monocular, stereo, or multi-camera system for indoor, outdoor and underwater environments. Monocular VO systems such as PTAM [18], Mono-SLAM [19], ORB-SLAM [20] are mainly feature-based—i.e., tracking features over a certain number of images. SVO [21] combines feature-based and direct methods to generate a fast and robust monocular VO. One consequence of monocular system is the loss of scale. Despite the variety of open source packages available, Quattrini Li et al. [4] provided a comparison of open-source state estimation algorithms on many datasets, showing insights on the challenges to adapt such methods to different domains.

Exploiting SLAM techniques in underwater environment is a difficult task due to the highly unstructured nature. To avoid scale ambiguity in monocular systems, stereo camera pairs are used. Salvi et al. [22] implemented a real-time EKF-SLAM incorporating a sparsely distributed robust feature selection and 6-DOF pose estimation using only calibrated stereo cameras. Johnson et al. [23] proposed an idea to generate 3D model of the seafloor from stereo images. Beall et al. [24] presented an accurate 3D reconstruction on a large-scale underwater dataset by performing bundle adjustment over all cameras and a subset of features rather than using a traditional filtering technique. A stereo SLAM framework named *selective* SLAM (SSLAM) for autonomous underwater vehicle localization was proposed in [25].

Vision is often combined with IMU and other sensors for improved estimation of pose. Oskiper et al. [26] proposed a real-time VO using two pairs of backward and forward looking stereo cameras and an IMU in GPS denied environments. Howard [27] presented a real-time stereo VO for autonomous ground vehicles. This approach is based on *inlier detection*— i.e., using a rigidity constraint on the 3D location of features before computing the motion estimate between frames. Konolige et al. [28] presented a real-time large scale VO on rough outdoor terrain integrating stereo images with IMU measurements. Kitt et al. [29] presented a visual odometry based only on stereo images using the trifocal geometry between image triples and a RANSAC-based outlier rejection scheme. Their method requires only a known camera geometry where no rectification is needed for the images. Badino et al. [30] proposed a new technique for improved motion estimation by using the whole history of tracked features for real-time stereo VO.

Hogue et al. [31] used stereo and IMU for underwater reconstruction. Stereo and IMU were used for VO in [32] and [33]. Sáez et al. [34] proposed a 6-DOF Entropy Minimization SLAM to create dense 3D visual maps of underwater environments using a dense 3D stereo-vision system and IMU; it is an offline method. Shkurti et al. [5] proposed a state estimation algorithm for underwater robot by combining information from monocular camera, IMU, and pressure sensor based on the multi-state constrained Kalman filter [35] .

There are also a body of work using low-cost sonar in underwater. Folkesson et al. [36] used a blazed array sonar

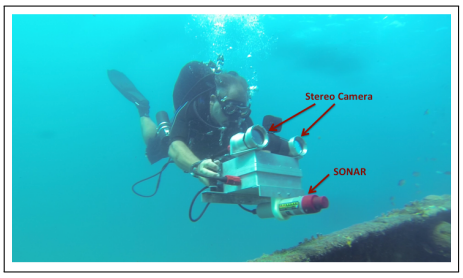


Fig. 2. Custom-made underwater sensor rig with depth sensor, IMU, stereo camera, and a mechanical scanning sonar.

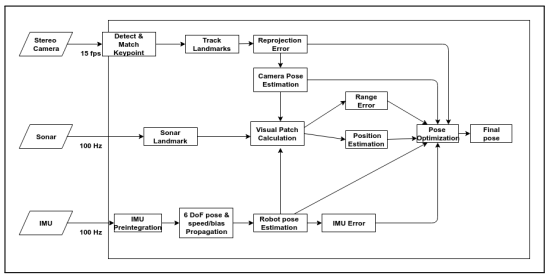


Fig. 3. The block diagram of the proposed algorithm.

for real-time feature tracking. A feature reacquisition system with a low-cost sonar and navigation sensors was described in [37].

Differently from other work, our proposed system, as described in the next section, includes a sonar in a new configuration to improve the reconstruction of underwater structures with a focus on caves. As such, a new method for integrating such data is presented.

III. SYSTEM OVERVIEW

The sensor suite employed for underwater structures reconstruction is a custom-made stereo rig, shown in Figure 2. In particular, the current system consists of the following components:

- two IDS UI-3251LE cameras,
- IMAGENEX 831L Sonar,
- Microstrain 3DM-GX4-15 IMU,
- Bluerobotics Bar30 pressure sensor,
- Intel NUC.

The two cameras are synchronized via an Arduino-like board, and they are able to capture at 15 frames per second with a resolution of 1600×1200 pixels. The sonar can provide data at a maximum of $6\,\mathrm{m}$ range, scanning in a plane over $360°$, with angular resolution of $0.9°$. A complete scan at $6\,\mathrm{m}$ takes about $4\,\mathrm{s}$. Note that the sonar provides for each measurement (beam) 255 intensity values, that is, at $6\,\mathrm{m}$ maximum range, $6/255\,\mathrm{m}$ is the distance between each returned intensity value. Clearly, higher response means a more likely presence of an obstacle. Sediment on the floor, porous material, and multiple reflections result in a multimodal distribution of intensities. The IMU produces linear accelerations and angular velocities in three axes at a frequency of $100\,\mathrm{Hz}$. The depth sensor produces depth measurements at $15\,\mathrm{Hz}$. The latter has not been used as data were collected from the same depth.

The hardware was designed with cave mapping as the target application. As such, the sonar scanning plane is parallel to the image plane. At first, the sensor suite is carried by divers. As a future design, we plan to mount it on an Autonomous Underwater Vehicle (AUV). In particular the hardware used is compatible with the Aqua AUV [38] and by mounting the scanning sonar on the robot identical sensing capabilities are provided. To enable easy processing of data, ROS [39], [40] has been used to record data in bag files.

IV. PROPOSED METHOD

Figure 3 presents an overview of the proposed approach. Data from the different sensors is combined to produce an accurate estimate of the state of the sensor suite. More specifically, the proposed method estimates the state $\mathbf{x}_R$ of the robot R by minimizing a joint estimate of the reprojection error, the IMU error term, and the sonar range error. The coordinate frames for camera, IMU, sonar, and world are denoted as C, I, S, and W respectively. The state vector contains the robot position ${}_W\mathbf{p}_{WI}^T = [{}_Wp_x, {}_Wp_y, {}_Wp_z]^T$, the robot attitude expressed by the quaternion $\mathbf{q}_{WI}^T$, the linear velocity ${}_W\mathbf{v}_{WI}^T$, all expressed in world coordinates; in addition the state vector contains the gyroscopes bias $\mathbf{b}_g$, and the accelerometers bias $\mathbf{b}_a$. Thus, Eq. (1) represents the state $\mathbf{x}_R$:

$$\mathbf{x}_R = [{}_W\mathbf{p}_{WI}^T, \mathbf{q}_{WI}^T, {}_W\mathbf{v}_{WI}^T, {\mathbf{b}_g}^T, {\mathbf{b}_a}^T]^T \tag{1}$$

The error-state vector is defined in minimal coordinates while the perturbation takes place in the tangent space; see Eq. (2):

$$\delta\boldsymbol{\chi}_R = [\delta\mathbf{p}^T, \delta\mathbf{q}^T, \delta\mathbf{v}^T, \delta{\mathbf{b}_g}^T, \delta{\mathbf{b}_a}^T]^T \tag{2}$$

which represents the error for each component of the state vector with a transformation between tangent space and minimal coordinates [41].

A. Cost Function

The joint nonlinear optimization cost function $J(\mathbf{x})$ for the reprojection error $\mathbf{e}_r$ and the IMU error $\mathbf{e}_s$ is adapted from the formulation of Leutenegger et al. [6] with an addition for the sonar error $\mathbf{e}_t$:

$$J(\mathbf{x}) = \sum_{i=1}^{I=2}\sum_{k=1}^{K}\sum_{j\in\mathcal{J}(i,k)} {\mathbf{e}_r^{i,j,k}}^T \mathbf{P}_r^k \mathbf{e}_r^{i,j,k} + \sum_{k=1}^{K-1} {\mathbf{e}_s^k}^T \mathbf{P}_s^k \mathbf{e}_s^k + \sum_{k=1}^{K-1} {e_t^k}^T \mathbf{P}_t^k e_t^k \tag{3}$$

where i denotes the camera index—i.e., left or right camera in a stereo camera system with landmark index j observed in the k^{th} camera frame. $\mathbf{P}_r^k$, $\mathbf{P}_s^k$, and $\mathbf{P}_t^k$ represent the information matrix of visual landmark, IMU, and sonar range measurement for the k^{th} frame respectively.

B. Error Terms Formulation

The reprojection error function for the stereo camera system and IMU error term follow the formulation of Leutenegger et al. [6]. Reprojection error describes the difference between a keypoint measurement in camera coordinate frame and the corresponding landmark projection according to the stereo projection model. Each IMU error term combines all accelerometer and gyroscope measurements by the *IMU preintegration* between successive camera measurements and represents both the robot *pose*, *speed*, and *bias* errors between the prediction based on the previous state and the actual state.

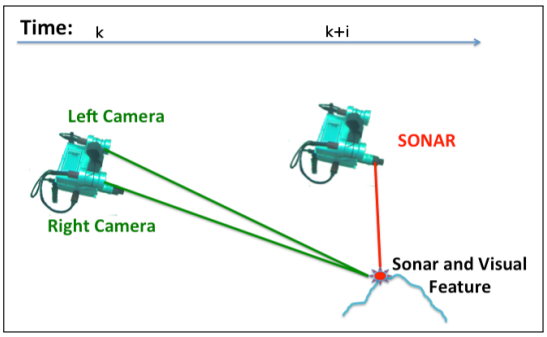


Fig. 4. The relationship between sonar measurement and stereo camera features. A visual feature detected at time k is only detected by the sonar with a delay, at time $k+i$, where i depends on the speed the sensor suite is moving.

In the presented system, the sonar measurements are used to correct the robot *pose* estimate as well as to optimize the landmarks coming from both vision and sonar. Due to the low visibility of underwater environments, when it is hard to find visual features, sonar provides features with accurate scale. A particular challenge is the temporal displacement between the two sensors, vision and sonar. Figure 4 illustrates the structure of the problem: at time k some features are detected by the stereo camera; it takes some time (until $k+i$) for the sonar to pass by these visual features and thus obtain a related measurement. To address the above challenge, visual features detected in close proximity to the sonar return are grouped together and used to construct a patch. The distance between the sonar and the visual patch is used as an additional constraint.

For computational efficiency, the sonar range correction only takes place when a new camera frame is added to the pose graph. As sonar has a faster measurement rate than the camera, only the nearest *range* to the robot *pose* in terms of timestamp is used to calculate a small patch from *visual landmarks* around the sonar landmark for that given *range* and *head_position*. Algorithm 1 shows how to calculate the *range error* $\mathbf{e}_t^k$ given the robot position ${}_W\mathbf{p}^k$ and the sonar measurement $\mathbf{z}_t^k$ at time k.

The sonar returns *range* r and *head_position* θ measurements, which are used to obtain each sonar landmark ${}_W\boldsymbol{l}_S = [l_x, l_y, l_z]$ by a simple geometric transformation in world coordinates:

$$ {}_W\boldsymbol{l} = (\boldsymbol{T}_{WI}\boldsymbol{T}_{IS}[r\cos(\theta), r\sin(\theta), 0]_S), \tag{4} $$

Algorithm 1 SONAR Range Error Algorithm

Input: Estimate of robot position ${}_W\mathbf{p}^k$ at time k
Sonar measurement $\mathbf{z}_t^k = [r, \theta]$ at time k
List of current visual landmarks, $\mathcal{L}_v$
Distance threshold, T_d
Output: Range error e_t^k at time k

/*Compute sonar landmark in world coordinates*/
1: ${}_W\boldsymbol{l} = \boldsymbol{T}_{WI}\boldsymbol{T}_{IS}[r\cos(\theta), r\sin(\theta), 0]_S)$
/*Create list of visual landmarks around sonar landmark*/
2: $\mathcal{L}_S = \emptyset$
3: **for** (every $\boldsymbol{l}_i$ in $\mathcal{L}_v$) **do**
/*Compute Euclidean distance from visual landmark to sonar landmark*/
4: $d_S = \|{}_W l - l_i\|$
5: **if** ($d_S < T_d$) **then**
6: $\mathcal{L}_S = \mathcal{L}_S \cup \boldsymbol{l}_i$
7: **end if**
8: **end for**
9: $\hat{r} = \|{}_W\hat{\mathbf{p}}_{WI} - \text{mean}(\mathcal{L}_S)\|$
10: **return** $r - \hat{r}$

where $\boldsymbol{T}_{WI}$ and $\boldsymbol{T}_{IS}$ are the respective transformation matrices used to transform the sonar measurement from the sonar coordinates to the world coordinates. More specifically, $\boldsymbol{T}$ represents a standard affine transformation matrix (rotation and translation). $\boldsymbol{T}_{IS}$ represents the transformation from the sonar frame of reference to the IMU reference frame, and $\boldsymbol{T}_{WI}$ represents the transformation from the inertial (IMU) frame to the world coordinates. Consequently, the sonar range prediction is calculated using Lines 2-9 of Algorithm 1:

$$ \hat{r} = \|{}_W\hat{\mathbf{p}}_{WI} - \text{mean}(\mathcal{L}_S)\| \tag{5} $$

where $\mathcal{L}_S$ is the subset of visual landmarks around the sonar landmark. As mentioned above, the concept behind calculating the *range error* is that, if the sonar detects any obstacles at some distance, it is more likely that the visual features would be located on the surface of that obstacle, and thus will be at approximately the same distance. Thus, the error term is the difference between the two distances. Note that we approximate the visual patch with the centroid (mean($\mathcal{L}_S$)), to filter out noise on the visual landmarks.

Consequently, the sonar error $e_t^k(\mathbf{x}_R^k, \mathbf{z}_t^k)$ is a function of the robot state $\mathbf{x}_R^k$ and can be approximated by a normal conditional probability density function f and the conditional covariance $\mathbf{Q}(\delta\hat{\boldsymbol{\chi}}_R^k|\mathbf{z}_t^k)$, updated iteratively as new sensor measurements are integrated:

$$ f(e_t^k|\mathbf{x}_R^k) \approx \mathcal{N}(\mathbf{0}, \mathbf{R}_t^k) \tag{6} $$

The information matrix is:

$$ \mathbf{P}_t^k = \mathbf{R}_t^{k\,-1} = \left(\frac{\partial e_t^k}{\partial \delta\hat{\boldsymbol{\chi}}_R^k} \mathbf{Q}(\delta\hat{\boldsymbol{\chi}}_R^k|\mathbf{z}_t^k) \frac{\partial e_t^k}{\partial \delta\hat{\boldsymbol{\chi}}_R^k}^T \right)^{-1} \tag{7} $$

The Jacobian can be derived by differentiating the expected *range* measurement $\hat{r}$ (Eq. (5)) with respect to the robot pose:

$$\frac{\partial e_t^k}{\partial \delta \hat{\boldsymbol{\chi}}_R^k} = \left[\frac{-l_x + {}_W p_x}{r}, \frac{-l_y + {}_W p_y}{r}, \frac{-l_z + {}_W p_z}{r}, 0, 0, 0, 0 \right] \tag{8}$$

The estimated error term is added in the nonlinear optimization framework (Ceres [42]) in a similar manner of the IMU and stereo reprojection errors.

V. EXPERIMENTS

The proposed approach has been tested in numerous challenging environments. In the following, experimental results from three representative scenarios are presented. For each dataset, a description is provided together with the results of the proposed state estimation approach. In addition, a short discussion of the challenges encountered during the field trials is included.

One of the first datasets was collected at an artificial shipwreck in Barbados; see Fig. 5(a). The initial deployment of the sonar sensor suffered from a configuration where data was collected at a very slow rate and at a maximum range of one meter. However, even with this configuration, the floor of the shipwreck is visible, which suggests that the sensor suite can be used even in less structured environments, such as coral reef regions. Figure 5(b) shows a top view of the trajectory together with sonar and visual features. Figure 5(c) presents a side view, where the vertical pole visible in the back of Fig. 5(a) is visible on the left side. Note that Fig. 5(c) shows the trajectory of the camera going slightly upwards, although the image frame of Fig. 5(a) shows the floor being horizontal. The shipwreck sits on the sea floor with an inclination, a fact that the IMU was able to capture from the calculation of the gravity vector.

We also collected a short segment from inside a cavern in Ginnie Springs, in Florida (USA). Such footage provided preliminary data from an underwater cave environment; see Fig. 6(a). The video light utilized was providing illumination on only part of the scene. Figures 6(b), 6(c) present two views of the trajectory together with the visual and sonar features. The reconstruction shows both visual landmarks and sonar points giving a sense of the cavern as the diver was swimming around. In this experiment, the sonar was configured at higher rate with maximum range of $6\,\mathrm{m}$. However, because of the light and environment characteristics—i.e., the scene was not uniformly illuminated—the visual features were sparse.

Finally, the inside of a sunken bus was mapped at Fantasy Lake Scuba Park, NC, USA; see Fig. 7(a). The image quality was quite poor due to the many particulates in the water. A top view is presented in Fig. 7(b) where the trajectory of the sensor as it entered the bus and traverse through its length is clear. Figure 7(c) presents a side view of the same results. Gaps on the sonar data are visible corresponding to areas where the windows of the bus were. In addition, at the right side of the figure the three steps of the bus are outlined.

In all environments, the images contain a significant amount of blur (softness) which clearly increases with distance. Moreover, dynamic obstacles, such as fishes, but more importantly floating particles that reflect back with high intensity, were present in all datasets; see Fig. 8.

In such challenging environments, it is very hard to get ground truth. However, the trajectory and the distance covered qualitatively resembled the one followed by the diver. Furthermore, the sonar landmarks were indeed used to correct the pose estimate, allowing the optimization to converge and keeping the error very low. Compared to OKVIS that uses just stereo images and IMU measurements, all the results in the datasets show more features mapped, e.g., several rings in the cavern, indicating the improved mapping of underwater structures.

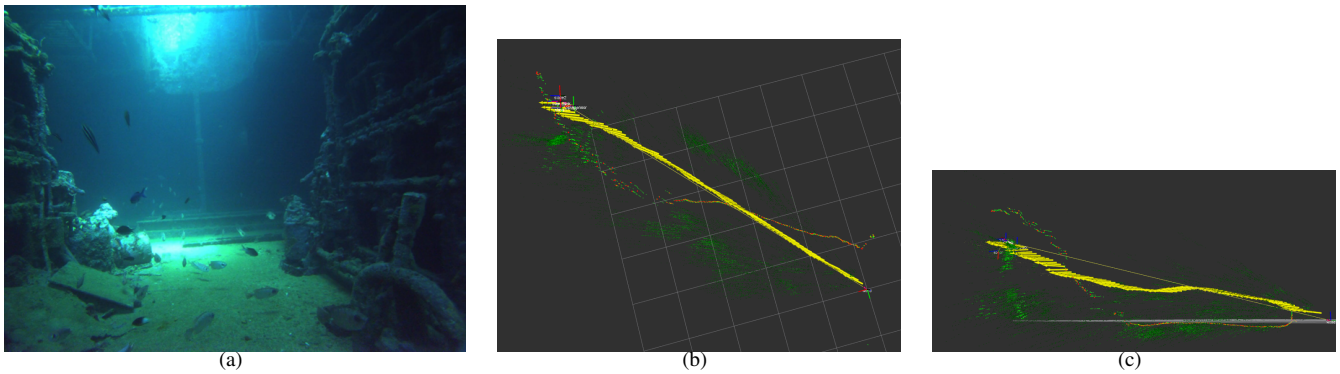
VI. CONCLUSIONS

As vision-based state estimation achieves a certain degree of maturity, more sensors are being integrated. Extending the well studied problem of Visual Inertial integration, we introduce a new sensor, a mechanical scanning sonar, which returns range measurements based on acoustic information. While the primary motivation of our work has been the mapping of underwater caves [1], the technique was tested in different environments, including a shipwreck at the clear waters of Barbados, artificial wrecks in the lakes of the Carolinas, and a cavern. A novel approach of merging sonar points with visual features is used to extend the pose graph generated for applying a global nonlinear optimization. The integration of the range data in the popular optimizer of Ceres [42] resulted in scale estimation improvements.

During the different experiments, it became clear that a minimum visibility and clarity in the visual data is required for basic performance; however, the data used degraded to a degree not often seen in typical datasets used for testing VO or VIO approaches. Moreover, the use of a strong video light, while necessary in the cave environment, it requires careful calibration of its position in order not to saturate the camera. Furthermore, different surfaces resulted in different reflectance properties of the acoustic signal; we are currently analyzing the sonar data over different materials to improve the quality.

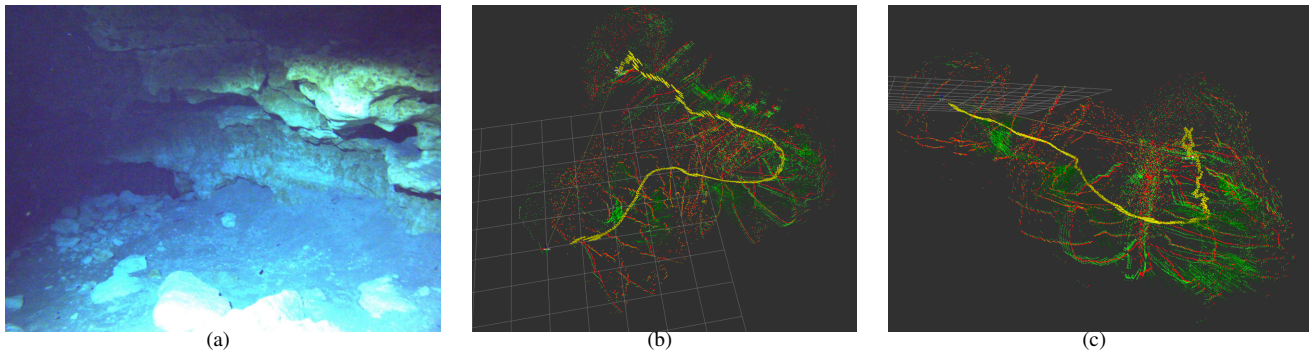
Future work, besides more data collection, will incorporate the stereo features obtained by the use of a strong video light during the data collection process. The robustness of the features introduced by the artificial light in a cave environment was presented in the work by Weidner et al. [1]. Preliminary work have demonstrated that even low-level ambient light cancels the effect of the artificial light, making the approach viable only inside caves. Furthermore, data from a depth sensor will be added in the proposed framework to account for vertical motions. Currently the majority of the data collected were from the same depth thus reducing the impact of such sensor. In addition, techniques to improve the quality of the images will be investigated.

Integration of multiple sensors will improve the quality of the estimation in addition to the density of the reconstruction.



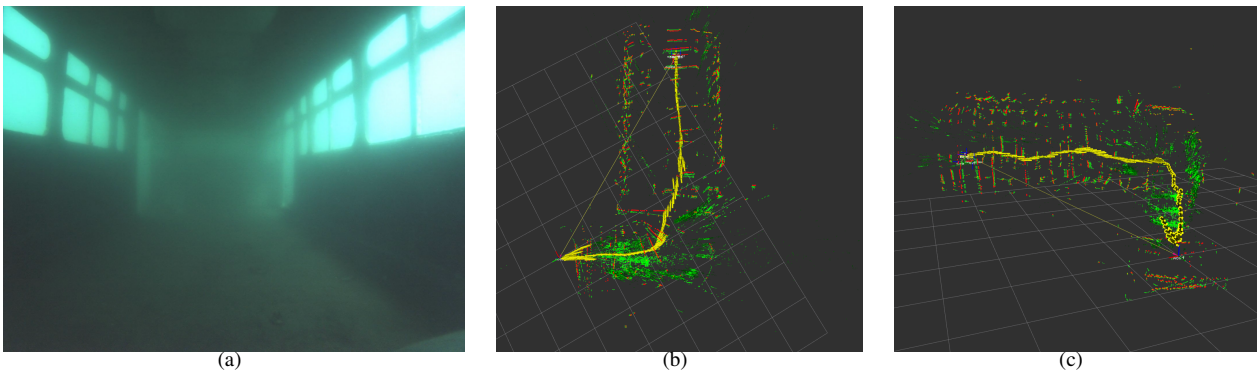
(a) (b) (c)

Fig. 5. Bajan Queen artificial reef (shipwreck) in Carlisle Bay, Barbados. (a) Sample image of the data collected inside the wreck (beginning of trajectory). (b) Top view of the reconstruction. (c) Side view of the reconstruction.



(a) (b) (c)

Fig. 6. Underwater cave, Ballroom Ginnie cavern at High Springs, FL, USA. (a) Sample image of the data collected inside the cavern. (b) Top view of the reconstruction. (c) Side view of the reconstruction.



(a) (b) (c)

Fig. 7. Sunken bus, Fantasy Lake Scuba Park, NC, USA. (a) Sample image of the data collected from inside the bus. (b) Top view of the reconstruction. (c) Side view of the reconstruction, note the stairs detected by visual features at the right side of the image.

A variety of domains will be affected with underwater archaeology and speleology being the primary areas. The resulting technology will be integrated to existing AUVs and ROVs for improving their autonomous capabilities.

ACKNOWLEDGMENT

This work was made possible through the generous support of National Science Foundation grants (NSF 1513203, 1637876). The authors are grateful to the University of South Carolina and in particular of the College of Engineering and Computing for the generous support.

REFERENCES

[1] N. Weidner, S. Rahman, A. Quattrini Li, and I. Rekleitis, "Underwater Cave Mapping using Stereo Vision," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 5709–5715.

[2] D. Scaramuzza and F. Fraundorfer, "Visual odometry [tutorial]," *IEEE Robotics Automation Magazine*, vol. 18, no. 4, pp. 80–92, 2011.

[3] A. Quattrini Li, A. Coskun, S. M. Doherty, S. Ghasemlou, A. S. Jagtap, M. Modasshir, S. Rahman, A. Singh, M. Xanthidis, J. M. O'Kane, and I. Rekleitis, "Vision-based shipwreck mapping: on evaluating features

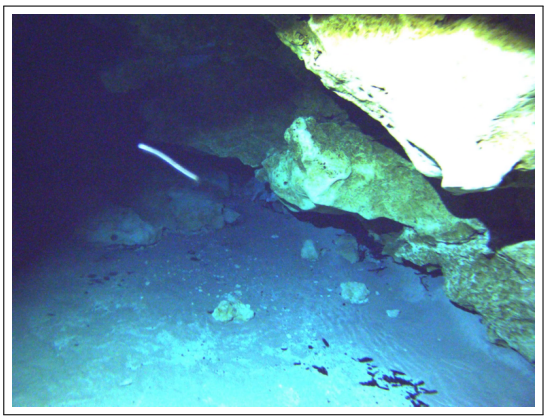


Fig. 8. A small particle reflecting back at high speed generating a blurry streak. In addition light reflecting back from a nearby surface completely saturates the camera.

quality and open source state estimation packages," in *MTS/IEEE OCEANS Monterey*, Sept. 2016, pp. 1–10.

[4] A. Quattrini Li, A. Coskun, S. M. Doherty, S. Ghasemlou, A. S. Jagtap, M. Modasshir, S. Rahman, A. Singh, M. Xanthidis, J. M. O'Kane, and I. Rekleitis, "Experimental Comparison of open source Vision based State Estimation Algorithms," in *International Symposium of Experimental Robotics (ISER)*, 2016, pp. 775–786.

[5] F. Shkurti, I. Rekleitis, M. Scaccia, and G. Dudek, "State estimation of an underwater robot using visual and inertial information," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2011, pp. 5054–5060.

[6] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based visual–inertial odometry using nonlinear optimization," *The International Journal of Robotics Research*, vol. 34, no. 3, pp. 314–334, 2015.

[7] R. Mur-Artal and J. D. Tardós, "Visual-inertial monocular SLAM with map reuse," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 796–803, 2017.

[8] M. Agrawal and K. Konolige, "Real-time localization in outdoor environments using stereo vision and inexpensive gps," in *International Conference on Pattern Recognition (ICPR)*, vol. 3, 2006, pp. 1063–1068.

[9] J. Rehder, K. Gupta, S. Nuske, and S. Singh, "Global pose estimation with limited gps and long range visual odometry," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2012, pp. 627–633.

[10] M. Gary, N. Fairfield, W. C. Stone, D. Wettergreen, G. Kantor, and J. M. Sharp Jr, "3d mapping and characterization of sistema Zacatón from DEPTHX (DEep Phreatic THermal eXplorer)," in *Proceedings of KARST08: 11th Sinkhole Conference ASCE*. ASCE, 2008.

[11] J. J. Leonard and H. F. Durrant-Whyte, *Directed sonar sensing for mobile robot navigation*. Springer Science & Business Media, 2012, vol. 175.

[12] C.-M. Lee, P.-M. Lee, S.-W. Hong, S.-M. Kim, W. Seong, *et al.*, "Underwater navigation system based on inertial sensor and doppler velocity log using indirect feedback Kalman filter," *International Journal of Offshore and Polar Engineering*, vol. 15, no. 02, 2005.

[13] J. Snyder, "Doppler Velocity Log (DVL) navigation for observation-class ROVs," in *OCEANS 2010 MTS/IEEE SEATTLE*, 2010, pp. 1–9.

[14] H. Johannsson, M. Kaess, B. Englot, F. Hover, and J. Leonard, "Imaging sonar-aided navigation for autonomous underwater harbor surveillance," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2010, pp. 4396–4403.

[15] P. Rigby, O. Pizarro, and S. B. Williams, "Towards geo-referenced auv navigation through fusion of usbl and dvl measurements," in *OCEANS 2006*, 2006, pp. 1–6.

[16] A. Mallios, P. Ridao, D. Ribas, M. Carreras, and R. Camilli, "Toward autonomous exploration in confined underwater environments," *Journal of Field Robotics*, vol. 33, no. 7, pp. 994–1012, 2016. [Online]. Available: http://dx.doi.org/10.1002/rob.21640

[17] P. Corke, C. Detweiler, M. Dunbabin, M. Hamilton, D. Rus, and I. Vasilescu, "Experiments with underwater robot localization and tracking," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2007, pp. 4556–4561.

[18] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," in *IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, 2007, pp. 225–234.

[19] J.-P. Tardif, Y. Pavlidis, and K. Daniilidis, "Monocular visual odometry in urban environments using an omnidirectional camera," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2008, pp. 2531–2538.

[20] R. Mur-Artal, J. Montiel, and J. D. Tardós, "ORB-SLAM: a versatile and accurate monocular SLAM system," *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015.

[21] C. Forster, M. Pizzoli, and D. Scaramuzza, "SVO: Fast semi-direct monocular visual odometry," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2014, pp. 15–22.

[22] J. Salvi, Y. Petillo, S. Thomas, and J. Aulinas, "Visual SLAM for underwater vehicles using video velocity log and natural landmarks," in *OCEANS 2008*, 2008, pp. 1–6.

[23] M. Johnson-Roberson, O. Pizarro, S. B. Williams, and I. Mahon, "Generation and visualization of large-scale three-dimensional reconstructions from underwater robotic surveys," *Journal of Field Robotics*, vol. 27, no. 1, pp. 21–51, 2010.

[24] C. Beall, F. Dellaert, I. Mahon, and S. B. Williams, "Bundle adjustment in large-scale 3D reconstructions based on underwater robotic surveys," in *OCEANS 2011 IEEE-Spain*, 2011, pp. 1–6.

[25] F. Bellavia, M. Fanfani, and C. Colombo, "Selective visual odometry for accurate auv localization," *Autonomous Robots*, pp. 1–11, 2015.

[26] T. Oskiper, Z. Zhu, S. Samarasekera, and R. Kumar, "Visual odometry system using multiple stereo cameras and inertial measurement unit," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.

[27] A. Howard, "Real-time stereo visual odometry for autonomous ground vehicles," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2008, pp. 3946–3952.

[28] K. Konolige, M. Agrawal, and J. Sola, "Large-scale visual odometry for rough terrain," in *Robotics research*. Springer, 2010, pp. 201–212.

[29] B. Kitt, A. Geiger, and H. Lategahn, "Visual odometry based on stereo image sequences with ransac-based outlier rejection scheme." in *Intelligent Vehicles Symposium*, 2010, pp. 486–492.

[30] H. Badino, A. Yamamoto, and T. Kanade, "Visual odometry by multi-frame feature integration," in *IEEE International Conference on Computer Vision Workshops*, 2013, pp. 222–229.

[31] A. Hogue, A. German, and M. Jenkin, "Underwater environment reconstruction using stereo and inertial data," in *IEEE International Conference on Systems, Man and Cybernetics*, 2007, pp. 2372–2377.

[32] M. Hildebrandt and F. Kirchner, "Imu-aided stereo visual odometry for ground-tracking auv applications," in *OCEANS 2010 IEEE-Sydney*, 2010, pp. 1–8.

[33] S. Wirth, P. L. N. Carrasco, and G. O. Codina, "Visual odometry for autonomous underwater vehicles," in *OCEANS-Bergen, 2013 MTS/IEEE*, 2013, pp. 1–6.

[34] J. M. Sáez, A. Hogue, F. Escolano, and M. Jenkin, "Underwater 3D SLAM through entropy minimization," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2006, pp. 3562–3567.

[35] A. I. Mourikis and S. I. Roumeliotis, "A Multi-State Constraint Kalman Filter for Vision-aided Inertial Navigation," in *IEEE International Conference on Robotics and Automation*, 2007, pp. 3565–3572.

[36] J. Folkesson, J. Leonard, J. Leederkerken, and R. Williams, "Feature tracking for underwater navigation using sonar," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2007, pp. 3678–3684.

[37] M. F. Fallon, J. Folkesson, H. McClelland, and J. J. Leonard, "Relocating underwater features autonomously using sonar-based SLAM," *IEEE Journal of Oceanic Engineering*, vol. 38, no. 3, pp. 500–513, 2013.

[38] G. Dudek, M. Jenkin, C. Prahacs, A. Hogue, J. Sattar, P. Giguere, A. German, H. Liu, S. Saunderson, A. Ripsman, S. Simhon, L. A. Torres-Mendez, E. Milios, P. Zhang, and I. Rekleitis, "A visually guided swimming robot," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2005, pp. 1749–1754.

[39] M. Quigley, K. Conley, B. P. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "ROS: an open-source Robot Operating System," in *ICRA Workshop on Open Source Software*, 2009.

[40] J. M. O'Kane, *A Gentle Introduction to ROS*. Independently published, October 2013, available at http://www.cse.sc.edu/~jokane/agitr/.

[41] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, "On-manifold preintegration for real-time visual–inertial odometry," *IEEE Transactions on Robotics*, vol. 33, no. 1, pp. 1–21, 2017.

[42] S. Agarwal, K. Mierle, and Others, "Ceres Solver," http://ceres-solver.org, 2015.