NOAA Office of Exploration and Research (OER)
Award #NA160AR0110195

"Real-Time 3D Reconstruction from ROV Camera Arrays of Opportunity"

PI: Aaron Marburg
University of Washington
Applied Physics Laboratory
Seattle, WA
amarburg@apl.washington.edu
206-685-8461

Funding: $105,720
Period of Performance: 9/1/2016 – 8/30/2018

Progress Report for Period: 3/1/2017-8/30/2018

Prepared By: ______________________________ 27 September 2018
Signature of Principal Investigator Date

Work Progress:

Previous work periods focused on indexing the ROV videos provided by OER, identifying short periods / scenes which might be suitable for 3D reconstruction, and the processing of those sections with the post-processed (offline) 3D reconstruction software *Agisoft Photoscan*. Taking *Photoscan* as representative of the current state-of-the-art in non-domain-specific reconstruction software performs several important functions:

1. Provides a sanity check on the suitability of the identified segments of ROV video for reconstruction. The ROV video has unique properties in terms of having relatively low contrast and limited color range which might impact the performance of reconstruction algorithms. There were also concerns on the impact of camera zoom and unmeasured camera motion (pan/tilt) on reconstruction.
2. Assuming reconstruction is feasible, the resulting point clouds provide a ground truth of both scene structure and camera motion for comparison with online algorithms.
3. Provides an estimate of the camera intrinsic properties (focal length, image center, and distortions). These parameters are necessary for reconstruction, and can be estimated either through direct measurement or can be included as additional dependent variables in the reconstruction process. The former requires in situ measurement of camera properties. This can be challenging for underwater cameras, as the focal length and distortions are influenced by refraction at the air-water interface at the viewport. This option is essentially impossible given the singular nature of the D2.

   The latter case is less preferable in general because it introduces a further degree of freedom in the bundle adjustment and can induce optimization instability. In this case, however, intrinsic estimation during bundle adjustment is the best available option. Performing this estimation within the more-mature Photoscan software means the intrinsics can be taken as a constant in the less mature online reconstruction software.

The reconstruction process was performed independently for two segments within dive EX1605L3_DIVE07 (Chamarro Seamount) and two segments within dive EX1605L3_DIVE22 (Saipan Channel / Romeo and Juliet) which had been identified previously as suitable for 3D reconstruction. The former two feature the ROV moving upslope over a relatively low detail rocky bottom. These sections were previously shown to be suitable for reconstruction, however the lack of re-viewings of scenes reduced confidence in the geometric correctness of the resulting reconstruction. The latter two show portions of B-29 debris on a sandy bottom and were thought to be the most suitable for reconstruction as the segments most closely resembled conventional terrestrial 3D reconstruction. To simplify analysis, segments were selected during portions of each dive where the camera was at a single zoom level – without metadata it cannot be confirmed whether the zoom levels were in fact equal (i.e., at equal focal length), however it is assumed these video segments were captured during periods where the main science camera was zoomed all the way out.

Intrinsic camera properties from the two segments in dive 22 were found to be relatively similar, and this was taken as confirmation both that the intrinsics had been well-estimated and that the camera focal length was at a preset stop – either a limit of travel or a pre-programmed position. The intrinsic camera properties from the two dive 7 segments were similar but with some variance. This variance might be attributed to the relatively low detail in the video, leading to persistent errors in feature tracking.

With these ground truth segments in hand, work continued on feeding the video and intrinsic properties into the online reconstruction software. As a post-processing reconstruction software package, Photoscan is designed to work most efficiently with a fixed set of images which are handled in an atemporal manner. During processing, a camera trajectory or path may become apparent, but the ordering of the images is not required for Photoscan. In this way, Photoscan reveals the origins of modern 3D photogrammetry in aerial photography, where surveys are collected as a relatively sparse set of high-resolution images with a fixed overlap. Photoscan obtains the best results by considering the entirety of the image set at once, looking for all possible overlaps and solving the full scene optimization, potentially at great computational cost.

In contrast, the goal of this project is to demonstrate *online* (and eventually realtime) reconstruction. In contrast to conventional photogrammetry, online reconstruction software takes images as a temporally-ordered stream – a sequence of images or a video – and attempts incremental reconstruction, adding each new image to its overall model. In this way, online reconstruction provides more immediate feedback about the relationship between the current video input and model quality. In exchange, the absolute quality of the final reconstruction may be compromised by optimizations which allow the software to run in realtime, for example is may reduce image size dramatically before processing, or may discard data it believes to be redundant.

The online reconstruction software used in this project is based on LSD-SLAM (Large-Scale Direct Simultaneous Localization and Mapping), a software package developed by Dr. Jacob Engels as part of his Doctoral thesis at TU Bremen in 2014-2016. SLAM broadly describes the algorithmic process of building a 3D world model online while simultaneously estimating the cameras motion. As of this writing, most SLAM algorithms sit at a level of NASA Technology Readiness Level 4-5, with competent hand-tuned algorithms being demonstrated widely in academia, but few commercially viable solutions. Those commercialized products which do exist are thoroughly hand-tuned for the application, although the authors believe generalized or easily adapted SLAM implementations on the near horizon.

The focus within this project period was on making the LSD-SLAM software sufficiently robust to allow processing of the ROV video. Large portions of the software were refactored to improve readability and functioning, to make use of modern C++ language constructs, and to improve robustness. The software was also adapted to the specifics of this data set, including development of a new frontend which can work directly with the uncompressed HD videos provided by OER, and which can index across multiple videos seamlessly by indexing to an absolute position within a dive. LSD-SLAM was also adapted to import Photoscan intrinsic camera properties directly and to correctly implement the Photoscan distortion and pinhole projection models.

LSD-SLAM was able to correctly process and model the two video segments from dive 22. For initial testing, the software was not run in a true realtime mode (where the video frames arrive at a consistent 30 frames per second and the software must keep up), but in a pseudo-realtime mode where LSD-SLAM processed frames as fast as possible. In this mode it currently runs at approx. 5 frames per second with full HD resolution input images with the understanding that this can be accelerated through further algorithmicoptimization, use of a faster computer and/or reduction in input image resolution. The latter step was not taken initially in part due to concerns about incorrect re-scaling of the Photoscan-derived intrinsics parameters. This scaling can be determined mathematically, or the intrinsics could be determined by processing scaled images in Photoscan, however this step was not taken in the time available.

While successful, this process revealed a major deficit in LSD-SLAM: it's relatively poor feature set for checkpointing or saving results. While it produces a live 3D model when starting from the beginning of the video, the Dr. Engel's published version does not provide an option to export that point cloud, understandably since this would be an expensive I/O operation and would require updating every time the pointcloud was improved. The final stage in LSD-SLAM improvement under this project was development of rudimentary 3D model and camera trajectory export tools for LSD-SLAM.

Final results and comparison to Photoscan results are being compiled for inclusion in the final project report.

Presentations and Publications:

Preliminary results have been presented at the 2018 Ocean Hack Week, hosted at UW, and at the Workshop on Remotely Deployed, Interactive, Deep Ocean Robotic Vehicles: "RESIDENT-AUVs, also hosted at UW. A technical paper is in preparation for the IEEE/MTS Oceans conference in 2019, with a targeted expansion to the IEEE Journal of Ocean Engineering.

Work Progress and Plan:

This report covers the final project period, final reporting is in progress.

Expenditures:

PLANNED EXPENDITURES FOR REPORTING PERIOD VERSUS ACTUAL EXPENDIT			
NOAA Grant No.:	**NA160AR0110195**		
Institution Name:	**University of Washington Applied Physics Laboratory**		
Lead PI:	**A. Marburg**		
Report Period	**3/1/2018-8/31/2018**		
Award Period:	**9/1/2016-8/31/2018**		
Award Amount:			
	Expenditures Planned	**Actual Expenditures**	
	For This Period	**For This Period**	**Difference**
Salaries & Wages	$ 21,787.42	$ 21,787.42	$ -
Staff Benefits	$ 12,018.08	$ 12,018.08	$ -
Travel	$ 251.83	$ 251.83	$ -
Services		$ -	$ -
Supplies		$ -	$ -
Equipment		$ -	$ -
Other	$ 12,124.45	$ 12,124.45	$ -
Indirect Cost	$ 8,774.53	$ 8,774.53	$ -
Total	$ 54,956.31	$ 54,956.31	$ -
Please Explain Any Significant Differences			