Turbid-water Subsea Infrastructure 3D Reconstruction with Assisted Stereo

R. Detry, J. Koch, T. Pailevanian, M. Garrett, D. Levine, C. Yahnker, M. Gildner
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA
Email: renaud.j.detry@jpl.nasa.gov

***Abstract*—This paper studies underwater perception and short-range scene reconstruction. Our work enables autonomous manipulation behaviors that support the autonomous maintenance of subsea infrastructure. We present a system that leverages assisted stereo to reconstruct the geometry of textured or untextured structures immersed in turbid water. Our package projects a random binary pattern in the cameras' field of view, which facilitates stereopsis in areas that are not naturally textured. We discuss the design and assembly of the package, and we quantify the accuracy and coverage of our method in turbid water ranging from 1 to 2.5 NTU.**

I. Introduction

To date, most underwater robotic manipulation tasks — for instance subsea infrastructure maintenance or ship inspection, are entirely guided by human operators who command the robots via joysticks and a camera-based monitoring systems. Unfortunately, teleoperation performances quickly degrade in applications affected by low bandwidth, long delays, and for tasks that require a fast operational pace. These difficulties can be addressed by allowing the robot to handle low-level manipulation operations autonomously, which allows the operator to focus on task-level robot supervision. For instance, the robot would manage actions such as turning a valve or plugging a cable. In turn, the operator would provide discrete directives, for instance by clicking a valve in an image and hitting a "turn" button. A key requirement for manipulation autonomy is the ability to reconstruct the 3D geometry of the robot's surroundings. Scene geometry is key to allow the robot to plan where and how to move its manipulators to perform a task, and to execute the task safely without contacting unmodeled obstacles. While scene reconstruction is a well-studied problem in clear media (air, other transparent gases, or vacuum), its application to scattering environments such as water, fog, or rain remains a poorly understood problem. Stereopsis and 3D reconstruction algorithms are less forgiving than human sight. Water and suspended particles scatter light, leading to blur and halo effects, and absorb light at a wavelength-dependent rate, leading to color distortion and signal degradation. Stereopsis and machine learning also struggle to cope with the low-resolution and spatially distorted video feeds that are traditionally used for teleoperation, and typically require higher-grade cameras with fixed optics, global shutter and digital transmission.

We present a solution that reconstructs the geometry of textured or untextured structures immersed in turbid water, with cm-scale fidelity within a 2m wide workspace. Our contributions include the design and realization of a submersible assisted-stereo camera system, a stereo system that reconstructs 3D structure from natural or artificial texture, and an experiment that quantifies reconstruction metrics relatively to water turbidity.

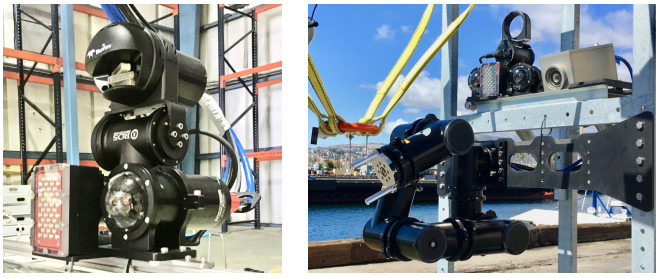

Fig. 1. Sensor head and articulated robot. Left: sensor head combining two cameras, an LED light (red-colored housing) and a sonar (above the light and cameras). Right: Cameras, pattern projector (grey metal housing), robot arm.

To our knowledge, this system is pioneering cm-scale underwater 3D reconstruction for untextured objects. The underwater vision community has demonstrated multiple means of conducting underwater stereo reconstruction of the seafloor [1], [2], [3], [4]. While passive stereo is well-suited to reconstructing natural environments where texture is abundant, it is not applicable to man-made structures that exhibit uniformly-colored surfaces. Researchers have studied the applicability of structured light to seafloor reconstruction, by shining a laser sheet across the robot's field of view [5], [6]. While this approach can cope with uniform surfaces, its application to robot manipulation is difficult because of the lengthy capture process during which the laser sheet scans the scene. This method also requires the scene to remain static for the duration of the scan. Bruno et al. [7] have studied the applicability of assisted stereo to underwater 3D reconstruction, but their approach requires the scene to be illuminated with 16 different patterns. As for structured light, this approach takes multiple seconds to capture the images required for scene reconstruction. Our approach is closest in sprit to that of Bruno at al. [7]. By contrast to their work, our approach allows the robot to capture and reconstruct a scene in less than a second, by only requiring one pattern to be projected onto the scene.

II. Underwater Scene Reconstruction

We have designed, assembled, and tested an underwater stereo head (Fig. 1) composed of two cameras, a light,

an imaging sonar, and a projector, and we have developed software for reconstructing 3D scene geometry from stereo images. We reconstruct depth with assisted stereo: The package projects a random binary "checkerboard" pattern in the cameras' field of view, which facilitates stereopsis in areas that are not naturally textured. We show that with a carefully-selected set of optical, imaging, lighting, acoustic and image-processing components, we are able to provided cm-accuracy 3D scene reconstruction in turbid water.

A. Hardware

To adequately support autonomous manipulation tasks, we require a system that is capable of providing a 3D reconstruction across a view angle of 50 degrees for surfaces that are 0.5m to 3m far from the camera, with 2cm accuracy at 2m range. To match those requirements, we selected two cameras with 1024x768 pixels on a 1/3 inch sensor and 3.5mm focal-length optics. The cameras are PointGrey gigabit-ethernet Flea cameras (FL3-GE-08S2C-C), equipped with Kowa LM3NCM lenses with a 3.5mm focal length and F2.4 aperture. We encased the cameras behind two polycarbonate domeports of 39mm inner radius, 3.8mm thickness, rigidly fixed on a plate with a 20cm baseline. The domeports mitigate distortions resulting from water-polycarbonate-air refractions, by feeding the cameras with light beams that traverse the port perpendicularly to the locally-tangent plane. An underwater domeport acts as a diverging lens [8], [9], producing a virtual upright image in front of the dome, at a distance that depends on the refraction indices of water, air, and polycarbonate, and on the radius and thickness of the port [8]. Given the dome dimensions listed above, the virtual image of an object situated at a 1m distance from one of our cameras is created at a distance of 82mm from the apex of the dome. The virtual image of an object at infinity is 10mm far from the apex. We performed an out-of-water lens configuration whereby we adjusted the lens' focus until it produced a sharp image of a calibration target situated at a distance of 85mm from the apex of the dome.

Each camera has a 68-degree horizontal field of view. The effective stereo field of view of the setup is 51 degrees at 0.5m range, and 65 degrees at 3m. Assuming a 0.5-pixel stereo correlation accuracy, the theoretical depth error has a standard deviation of 0.8mm at 0.5m, and 13mm at 2m. We trigger the two cameras with a single step signal generated by an Arduino, to produce closely-synchronized stereo pairs.

We calibrated the intrinsic and extrinsic parameters of our vision head with measurements collected underwater. We achieved a 0.3-pixel model reprojection error, which is within the expected range for underwater vision. The vision head achieves 3D point triangulation with an average error of 0.7mm at 0.5m. This assessment results from the triangulation of the corners of a planar checkerboard of know geometry across 70 images taken in a fishtank under optimal lighting and negligible turbidity. This number represents an upper bound for the accuracy that is potentially achievable with dense stereo reconstruction.

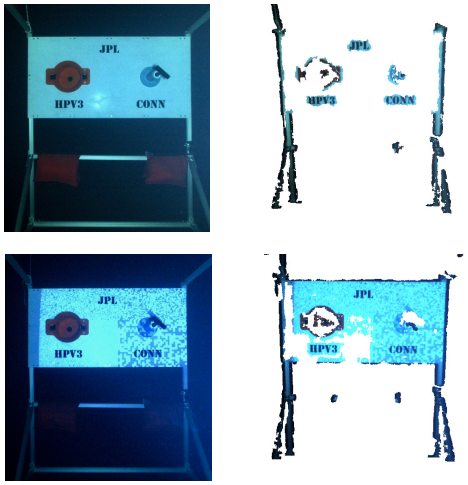

Fig. 2. The top-right image shows a camera image of a task panel subjected to homogeneous lighting. The lack of natural texture on the panel makes it hard for stereo to compute depth, leading to gaps in the panel's 3D reconstruction shown in the top-right image. The bottom-left image shows the same panel subjected to an artificially-projected pattern. The frequency (or block size) of the pattern increases through the bottom-left, top-left, top-right, and bottom-right quadrants of the panel. The bottom-right image illustrates that for all but the bottom-left quadrant, the texture projected on the panel is sufficient to allow stereo to compute depth.

As mentioned above and further discussed in the next section, we reconstruct 3D geometry using a technique referred to as assisted stereo, whereby we project a random pattern in the field of view of the cameras. The projector adds texture to homogeneous surfaces and facilitates the work of stereopsis. We use a Texas Instruments DLP4710 projector of 600 lumens, encased in a waterproof container with a flat frontal port. The projector is controlled by a Raspberry Pi installed in the same container, and connected to a deckbox via ethernet (and a separate power line). To enable operations in high-turbidity conditions, we have mounted a large LED spotlight between the two cameras, within the blind spot of their joint field of view. This spotlight is not used in the experiments below.

B. Scene reconstruction

We solve dense stereo reconstruction with a standard block matching algorithm (BM), or with its semi-global variant SGBM [10], with a correlation window of 9 pixels and limiting the disparity search range to depths of 0.80 to 2 meters. To compute a dense depth map, we undistort and rectify raw images based on the cameras' intrinsic and extrinsic parameters computed during calibration, and we compute a dense disparity map from the rectified images. This process provides us with a distance (in meters) between the left camera's optical center and all resolvable textured surfaces within the stereo system's field of view.

Block matching works by identifying corresponding patches in left and right images, and triangulating the 3D position of the source from camera geometry. The main difficulty with stereo is to identify unique left-right correspondences. Unique correspondences are abundant when a scene is highly

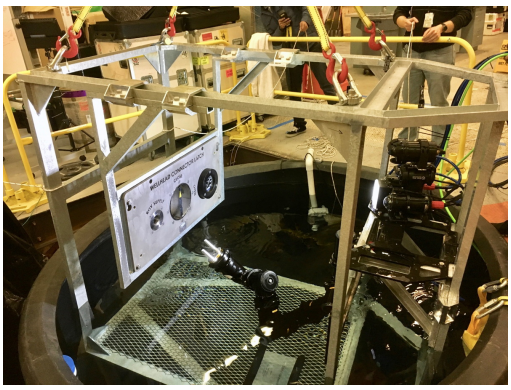

Fig. 3. Perception head, test panel, and robot manipulator mounted on a rigid frame in the relative configuration used for the experiment of Fig. 6.

textured, but the steel and painted components of subsea structures are typically uniform in color and surface properties, making stereo matching unreliable. We alleviate this problem by projecting our own texture in the cameras' field of view, which drastically enhances stereo on homogeneous surfaces (Fig. 2).

III. Experiments

We evaluated the applicability of our system by reconstructing the panel of Fig. 3. We attached the vision head and panel to a rigid steel structure, such that the panel stands at 1.1m in front of the camera. We captured panel images in varying turbidity conditions, with different shutter speeds and projected patterns, and measured the accuracy and coverage of stereo reconstructions across all combinations of conditions.

We varied turbidity by dissolving solid clay in the tank shown in Fig. 3 (bottom), and circulating the water with a pump. After homogeneous turbidity was achieved, we stopped the pump and captured batches of images approximately every 8 minutes for a total of 33 hours as the clay settled at a natural pace. We monitored turbidity with a spectrophotometer produced by Scan Messtechnik GmbH installed next to the cameras. The spectrophotometer provides NTU and FTU readings. NTU readings throughout the test ranged from 1 to 2.5. Fig. 4 shows a plot of the turbidity readings averaged through windows of 8 minutes.

We projected four random binary patterns of increasing block size: 4×4, 8×8, 12×12 and 24×24 pixels respectively. For reference, the projected image consisted of 1824×984 pixels. The top row of Fig. 9 shows the pattern with 4×4 pixel blocks, the bottom row shows the pattern with the largest blocks, 24×24 pixels. Smaller blocks lead to more complex texture, which should facilitate stereo matching. However, turbid water scatters light and induces blur, which may obstruct small-block patterns. The experiments below showed that within our turbidity range, the 24px block pattern consistently outperformed other patterns.

We captured images at five different shutter speeds — 0.01s, 0.02s, 0.04s, 0.08s, and 0.16s. Shorter exposures typically lead to noisier images. However, long exposures are not always acceptable, as they increase sensitivity to motion blur. We implemented a software gain controller which selects the maximum camera gain that provides fewer than 0.1% of saturated pixels across both camera images. The same gain is then set for both cameras.

Every 8 minutes of the 33 hours of the test, we captured 20 stereo pairs, corresponding to the 20 combination of 4 patterns and 5 shutter speeds. We evaluated our results by measuring the accuracy and coverage of the reconstruction of the task panel. We built a ground-truth reconstruction of the panel from manual measurements, yielding the ground-truth point cloud shown in Fig. 5 (left). The ground-truth point cloud had an average 1.2cm spacing between points.

We define coverage as the fraction of ground-truth points that are within 2cm of a reconstructed point. Let $G = \{g_i\}_{i\in[1,M]}$ be the set of points forming the ground-truth cloud, and $P = \{p_i\}_{i\in[1,N]}$ a point cloud acquired via stereo. We first define G' as the subset of G that have a neighbor in P at a distance of 2cm or less

$$G' = \left\{ g \in G : \min_{p\in P} |s - g| < 2\text{cm} \right\}. \tag{1}$$

Coverage is expressed as

$$c = \frac{|G'|}{|G|}. \tag{2}$$

Coverage varies between 0 and 1. Low coverage values indicate gaps in the reconstruction. We define the accuracy of P as the root mean square error of the pairs established in Eq. 1:

$$a = \sqrt{\frac{1}{|G'|} \sum_{g\in G'} \min_{p\in P} (g-p)^2} \tag{3}$$

By restricting our measure of accuracy to pairs of points from G and P that are at most 2cm apart, we allow this metric to remain informative in low-coverage cases where only a small number of scene points are reconstructed.

Fig. 6 show plots of accuracy and coverage as functions of time, for BM and SGBM stereo, shutter speeds of 0.01s, 0.04s, and 0.16s, and 4px and 24px block patterns. In all cases, the larger block pattern (in orange) leads to better results than the smaller one (in green), leading to an accuracy of approximately 6mm and a coverage of 100% across the board in low turbidity.

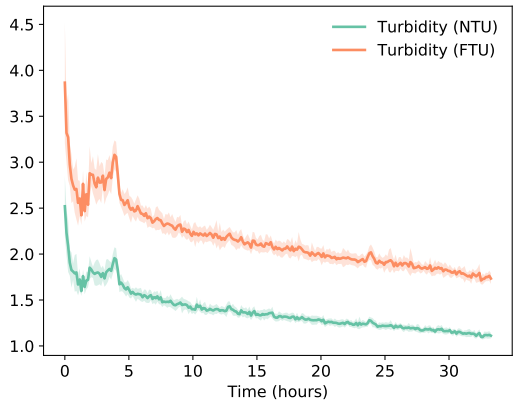

Fig. 4. Turbidity readings through a 33-hour test. Our turbidity sensor provides approximately three measurements per minute. We divided the data in time slices of 8 minutes. This figure plots the per-slice average turbidity. Standard deviation within a slice is plotted in filled lighter color. Both plots exhibit a large variance between 1 and 4 hours. We have not been able to identify the cause. We note however that the monotonicity of the plots of Fig. 6 indicates that this variance may not be related to a global change of tank turbidity, but rather to a phenomenon local to the area where the spectrophotometer was mounted. The sensor was mounted behind the cameras, approximately 50cm away from their viewports.

Short exposures (top row) unsurprisingly lead to worse accuracy than longer ones. The top row of Fig. 7 shows an image taken with a 0.01s shutter speed, with turbidity at its highest. The only part of the scene that can be reconstructed from that image is the contour of the center valve. With the same turbidity and pattern, a longer exposure leads to approximately 50% of coverage (Fig. 7, bottom row).

Standard block matching and SGBM were on par with one another. We note that we did not filter the SGBM output in Fig. 7, Fig. 8 and Fig. 9, which explains the speckles that appear over the background. Points that are reconstructed outside of the surface occupied by the panel in the image do not influence accuracy or coverage. The 24px block pattern outperformed the 4px pattern across the board.

Fig. 8 shows that 2.5 hours into the test, when turbidity read approximately 1.6 NTU, a shutter speed of 0.04s or longer provided coverage above 95%, and a RMSE below 7mm. Fig. 9 illustrates low-turbidity performance.

The standard deviation of the mean absolute error across all inliers (1) of a given reconstruction was consistently of the order of 5mm, regardless of turbidity, pattern block size or shutter speed.

IV. Discussion

This work is motivated by autonomous underwater robotic operations, where accurate 3D reconstruction is a key asset. Our objective is to develop software that allows the robot of Fig. 3 to turn the panel's valve without supervision. The experiment of the previous section informs us on the minimum acceptable exposure time for a given accuracy, at a certain turbidity, which in turn informs us on the pace of scene motion

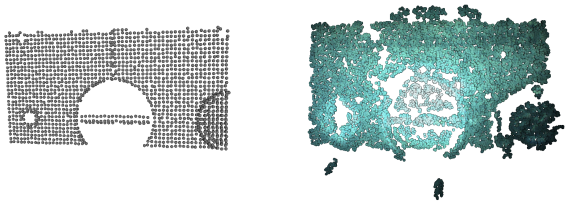

Fig. 5. Ground truth structure (left) and stereo point cloud (right). Ground truth was established via manual measurements. The ground truth panel only contains points that are within the projection cone of the pattern projector, and within the frontal plane of the panel.

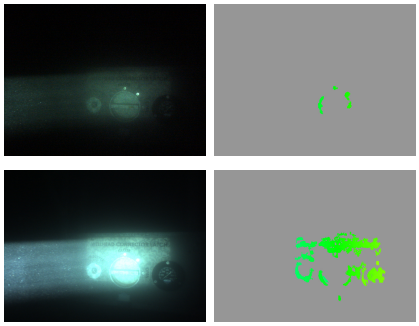

Fig. 7. Camera image (left) and disparity (right) captured at $t = 8$ minutes (high turbidity), with shutter speeds of 0.01s (top) and 0.16s (bottom). Both top and bottom images were captured while projecting the pattern with the largest block size (24px).

Fig. 8. Camera image (left) and disparity (middle and right) captured at $t = 2.5$ hours (medium turbidity), with a 0.16s shutter speed and 24px block pattern. The middle image illustrates a disparity map computed with standard block matching, while the rightmost image was computed with SGBM. Green pixels are approximately 1.1m far from the cameras. Red pixels are less than a meter from the cameras; blue pixels are further than 1.2m.

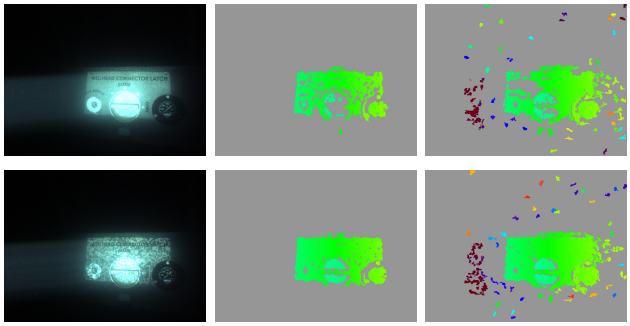

Fig. 9. Camera image (left) and disparity (middle and right) captured at $t = 33$ hours (low turbidity), with a 0.04s shutter speed, and 4px and 24px block pattern for the top and bottom row respectively. In each row, the middle image illustrates a disparity map computed with standard block matching, while the rightmost image was computed with SGBM. Green pixels are approximately 1.1m far from the cameras. Red pixels are less than a meter from the cameras; blue pixels are further than 1.2m.

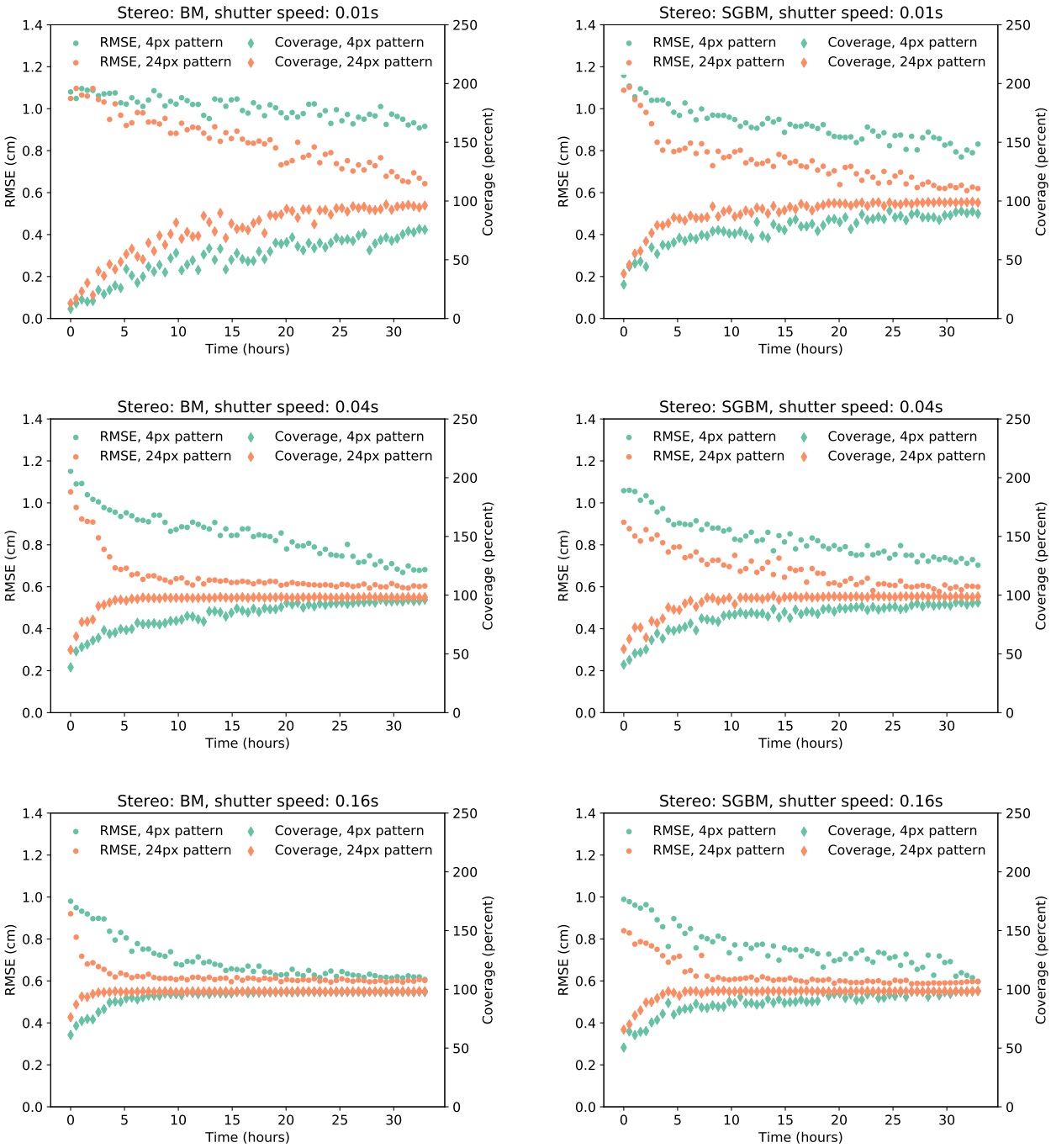

Fig. 6. Accuracy (RMSE) and coverage of stereo reconstructions with different shutter speeds and projected patterns. The plots show a systematically-sampled subset of the data for clarity. Refer to Fig. 4 to relate time to turbidity.

our system can cope with (shorter exposures allow for faster motion).

In parallel to the work discussed above, we are designing an experiment where depth is computed via the piecewise-planar stereo algorithm of Roser et al. [3]. This algorithm first computes a sparse set of support point via feature matching. Then, the algorithm computes the depth of the remaining pixels with a mixture of dense stereo and a pixel-wise depth prior parametrized by the depth of pixels where a support point has been computed. We are strengthening our reconstruction algorithm by pruning support points that do not correlate with acoustic measurements, leading to a 20% accuracy improvement according to preliminary tests.

V. CONCLUSIONS

We presented a system that leverages assisted stereo to reconstruct the geometry of textured or untextured structures immersed in turbid water. Our package projects a random binary pattern in the cameras' field of view, which facilitates stereopsis in areas that are not naturally textured. We discussed the design and assembly of the package, and we quantified the accuracy and coverage of our method in turbid water ranging from 1 to 2.5 NTU. Our experiments showed that our system can achieve cm-scale accuracy in a large range of turbidity conditions. This work will ultimately enable autonomous manipulation behaviors that support the autonomous maintenance of subsea infrastructure.

VI. ACKNOWLEDGEMENTS

The research described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Copyright 2018 California Institute of Technology. U.S. Government sponsorship acknowledged.

REFERENCES

[1] 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.

[2] M. Massot-Campos and G. Oliver-Codina, "Optical sensors and methods for underwater 3d reconstruction," *Sensors*, vol. 15, no. 12, pp. 31 525–31 557, 2015.

[3] M. Roser, M. Dunbabin, and A. Geiger, "Simultaneous underwater visibility assessment, enhancement and improved stereo," in *IEEE International Conference on Robotics and Automation*, 2014.

[4] A. Sedlazeck, K. Koser, and R. Koch, "3d reconstruction based on underwater video from rov kiel 6000 considering underwater imaging conditions," in *OCEANS 2009-EUROPE*. IEEE, 2009, pp. 1–10.

[5] A. Bodenmann, B. Thornton, R. Nakajima, H. Yamamoto, and T. Ura, "Wide area 3d seafloor reconstruction and its application to sea fauna density mapping," in *Oceans-San Diego, 2013*. IEEE, 2013, pp. 1–5.

[6] C. Roman, G. Inglis, and J. Rutter, "Application of structured light imaging for high resolution mapping of underwater archaeological sites," in *OCEANS 2010 IEEE-Sydney*. IEEE, 2010, pp. 1–9.

[7] F. Bruno, G. Bianco, M. Muzzupappa, S. Barone, and A. Razionale, "Experimentation of structured light and stereo vision for underwater 3d reconstruction," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 4, pp. 508–518, 2011.

[8] F. A. Jenkins and H. E. White, *Fundamentals of optics*. Tata McGraw-Hill Education, 1937.

[9] A. Jordt, *Underwater 3D reconstruction based on physical models for refraction and underwater light propagation*. Universitätsbibliothek Kiel, 2014.

[10] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 30, no. 2, pp. 328–341, 2008.