Structured Light and Stereo Vision for Underwater 3D Reconstruction

Miquel Massot-Campos*, Gabriel Oliver-Codina*, Hashim Kemal†, Yvan Petillot†, Francisco Bonin-Font*
* Departament de Matemàtiques i Informàtica, Universitat de les Illes Balears
Cra. Valldemossa, km. 7,5 07122 Palma de Mallorca (Spain)
† Oceans Systems Laboratory, Heriot Watt University
Riccarton, Currie EH14 4AS, United Kingdom

***Abstract*—Stereo vision and structured light are compared in a common underwater environment with known dimensions and objects. Two different sensors are mounted on top of a Cartesian robot that moves with a known and programmed trajectory. The resulting point clouds from each sensor are compared in terms of distance from point to point, and measurements in the scanned objects, to determine which sensor is best suited depending on the environment and the survey purpose.**

The conclusions show that a stereo based reconstruction is best suited for long, high altitude surveys, always depending on having enough texture and light, whereas a structured light reconstruction can be better fitted in a short, close distance approach where accurate dimensions of an object or structure are needed.

I. Introduction

There is a direct classification in underwater three-dimensional sensors given its behaviour in the medium. Active sensors are those that either illuminate, project or cast a signal to the environment to help, enhance or measure the data to gather. An example of an active system is structured light, where a pattern is projected onto the object or scene to reconstruct.

Passive systems get measurable signals in the underwater environment without altering or changing the scene or media. An example of passive system include stereo vision, where two monocular cameras look for common features for a posterior 3D triangulation. Generally, imaging-based sensors are the only passive sensors for 3D reconstruction.

Three-dimensional data can be obtained measuring the distance from two or more devices (either sensors or receivers) with some known parameters. This method is called triangulation. For a typical projector and camera rig with a known baseline (e.g. the distance from the projector to the camera) the computation of a 3D point is done by identifying the same image feature in the camera image and its correspondence in the projector image. The lines formed from the illuminated point to the camera and to the projector, as well as the line joining both, form a triangle. This triangle, the baseline, and the angles formed are used to compute the three-dimensional position of that feature in space.

Triangulation-based sensors are limited by the field of view of both the emitter and the receiver. These devices tend to be better at close distances and worse at far. Also, the bigger is the baseline, the better is the z resolution [1].

Different sensors technology exist that make use of this way of computing 3D information, for example: Structured Light (SL), Laser Stripe (LS) and Photometric stereo (PhS) are active imaging based, Structure from Motion (SfM) and passive stereo vision (SV) are passive imaging based.

In this paper we present a comparison between two common triangulation sensors, one passive and one active. Stereo vision and structured light are compared in a common underwater environment. Both systems are used to recover the 3D structure of a surveyed area.

The article is structured as follows: first the related work is presented in section II. Then, in section III, the sensors used and how the comparison is made are explained. In section IV the experimental results are shown and finally, in section V the conclusions and the future work are discussed.

II. Related work

In the literature, some authors have focused on close-range underwater 3D reconstructions of objects [2], comparing structured light and passive stereo. Erič *et al* [3] explored that field from the point of view of the documentation of underwater heritage sites.

Regarding stereo-based underwater reconstruction, Hogue *et al.* [4] used a Bumblebee stereo combined to a inertial unit in a underwater housing, called *Aquasensor*, and used it to reconstruct a sunken barge. In [5], Beall *et al.* reconstructed coral reefs in Bahamas. Negre *et al.* [6] performed 3D reconstruction of underwater environments using a Graph SLAM approach in a micro AUV equipped with two stereo rigs.

Inglis and Roman constrained stereo correspondences using multibeam sonar [7]. From *Hercules* ROV, navigation data, multibeam and stereo were preprocessed to reduce error and then the sonar and optical data were mapped into a common coordinate system. They backprojected the range data comming from the sonar to the camera image and limited the available z range the stereo correspondence algorithm had.

On the structured light side, the use of an underwater stripe scanning system was initially proposed by Jaffe and Dunn in [8] to reduce backscattering. Tetlow and Spours [9] showed in their article a laser stripe system with an automatic threshold setup for the camera. In their results, they achieved resolutions up to five millimetres at a distance of three meters.

978-1-4799-8736-8/15/$31.00 ©2015 IEEE

Kondo *et al.* [10] tested a LS system in *Tri-Dog I* AUV. Apart from using it for 3D reconstructions, they also tracked the image to govern the robot to keep a safe distance to the seabed, centering the laser line in the camera image by changing the depth of the vehicle. They reported a resolution of $40\ mm$ at three meters.

In [11] a system consisting on a camera, a laser line, a LED light and a clever shade was mounted on the AUV *Tuna Sand* to gather 3D information as well as imagery. The laser was pointed at the upper part of the image whilst the lightning is illuminating the lower part, so that there is enough contrast to detect the laser line. With this system, the authors performed 3D reconstructions in real intervention scenarios such as in hydrothermal sites and shipwrecks.

Zhang *et al.* projected a gray scale four-step sinusoidal fringe [12]. Therefore, it is a time multiplexing method using four different patterns. In their article, SL was compared to SV showing better behaviour in SL at textureless objects. Same reports were obtained projecting 20 different gray coded patterns in a pool [13]. An accuracy in z direction of 2% was achieved with this system.

Bruno *et al.*[14] also projected gray coded patterns with a final code shift of 4 pixel wide bands. The projector was only used to establish the correspondences and it was not involved in the triangulation. This system would be an hybrid between SL and SV.

Another way to triangulate information using structured light is to sweep a light plane. This light plane can be swept either using the available pixels in the projector of by moving the projector. Narasimhan and Nayar [15] sweeped a light plane into a tank with dilute milk and recovered 3D information even in high turbidity scenarios. The authors showed that in high turbidity medium and with conventional floodlight, it is impossible to see anything but backscattering. By narrowing the illuminated area to a light plane, the shapes of the objects in the distance could be picked out and therefore triangulated.

Different approaches to the common laser stripe scanning have been also reported. In [16] two almost-parallel laser stripes were projected to compute the distance between these lines captured from a camera, to know the distance to the target. These values were used as an underwater rangefinder. However 3D reconstruction was not the aim of the paper.

Some authors of this paper designed a laser-based structured light system (LbSLS) that enhances simpler laser stripe approaches by using more than two lines in the laser projector [17]. Their reconstructions can be also done in one camera shot, turning them out to be feasible for moving robots or for manipulation stages.

In this paper, the LbSLS will be compared to a stereo camera in a common underwater environment. Two reconstructions will be made and compared for the retrieval of object shapes and dimensions to assess which sensor is better suited depending on the purpose of the survey operation.

In our experiments, two datasets using this sensor and a stereo camera are presented, not only reconstructing small objects but also bigger areas that cannot be recovered by a single camera shot or pattern projection.

(a) PGR Bumblebee 2 stereo camera (b) Laser-based structured light system

Figure 1. Sensors compared. Note that the structured light system on the right is formed by a AVT Manta G-283C camera housed in a black acetal cylinder and a green laser housed in a white acetal cylinder.

III. Methodology

Two sensors are compared in this paper, a stereo camera and a structured light sensor based on a laser projector and a camera. The sensors are shown in figure 1. These two devices were mounted, one at a time, on a Cartesian robot in a pool.

The methodology used in the data processing from both sensors has been implemented using C++, ROS and OpenCV libraries.

A. Stereoscopy

The first sensor used is an off-the-shelf *FireWire* stereo camera *Bumblebee2* from *Point Grey Research*. This stereo rig includes two CCD 1024×768 px CCD color cameras. The stereo rig is mounted in an underwater housing with a flat optical port. Its focal length is $3.8\ mm$ out of water, $5\ mm$ in water.

The synchronized images are received in *Bayer* format and the computer processes their disparity, the rectified images and their corresponding point clouds. These are computed using the depth given by eq. (1) and the colour from the left rectified image. The stereo correspondence is solved using a block matching algorithm to obtain the disparity from image pairs.

$$z = \frac{b \cdot f}{d}, \tag{1}$$

where, for each pixel, b is the baseline, f the focal length, d its computed disparity and z its depth.

For each frame, 3D points are computed using pixel disparities and colour. For consecutive frames, the 3D points have been filtered using a voxel grid filter, and finally accumulated using the pose of the Cartesian robot

If there is not enough texture in the images, the correspondence will not be solved and the disparity image will show a blank area, where the correspondence was missed. Thus, the generated point cloud will also miss some points.

B. Structured light

The second sensor used is a Laser-based Structured Light sensor [17] formed by an off-the-shelf *Ethernet* camera *Manta G-283c* from *Allied Vision Technologies*, 1920×1440 px CCD

color camera equipped with a 12 mm optics (16 mm in water), and a 532 nm 5 mW green laser with a Diffractive Optical Element (DOE) in front of the laser beam. This DOE shapes the laser dot into a set of 25 parallel lines with a brighter dot in the central line and a horizontal field of view of 25° (19° in water). Both the laser and the camera have also a flat optical port in their underwater housing.

The system processing is based on triangulation. The laser projects an uncoded pattern of 25 parallel lines in the environment and the camera recovers the deformed pattern. The deformed lines are then labelled (decoded) in the image to correspond to the original pattern. Each of these lines can be represented as a plane π_n that crosses the laser origin and the line projected in the environment. Once the plane is known by previous calibration and the deformed line pattern has been identified in the image, its three dimensional coordinates can be computed by solving (3). Each point belonging to the projected pattern and recovered as a pixel in the camera can be represented as a line $r(t)$ that crosses the camera focal point and the plane π_n.

$$\pi_n \quad : \quad Ax + By + Cz + D = 0, \tag{2}$$

$$p(t) \quad = \quad \left(\frac{u - c_x}{f_x}\, t, \frac{v - c_y}{f_y}\, t, t \right), \tag{3}$$

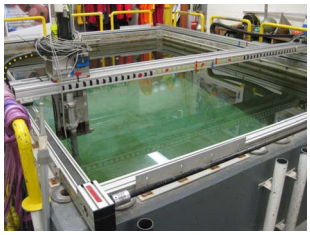
$$t \quad = \quad \frac{-D}{A\frac{u-c_x}{f_x} + B\frac{v-c_y}{f_y} + C}, \tag{4}$$

where (f_x, fy) is the camera focal length in x and y axes. (c_x, c_y) is the central pixel in the image. (u, v) is the detected laser peak pixel in the image.

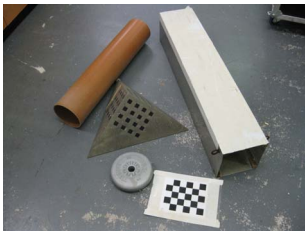
The labelling relies on the detection of the central laser point in the image or in the detection of 25 line crossings in image rows on each frame. This detections are needed to correctly relate the detected points to its corresponding laser plane. If the central point is not detected or absent in the image and there are not enough line crossings, then there will not be 3D points for that image frame, as the sensor is working as a one-shot reconstruction sensor [17].

When the lines are detected in the image, the computer does not know which line corresponds to which plane or index. This is called the correspondence problem. If 25 line crossings are detected in any row, a Floodfill algorithm is used on each crossing to label the lines. Similarly, if the central dot is found in the image, then its correspondence is automatically determined to be the central line, e.g. index 13 of 25 lines. Once these operations are complete, then the unlabelled line segments are guessed using the available correspondences in the same row. For example, if there is a unlabelled segment in between of two segments labelled 1 and 3, the unlabelled can be correctly guessed as index 2.

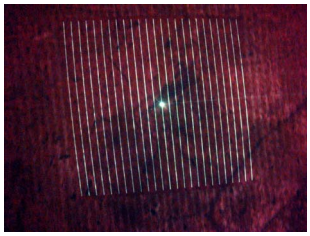
As in stereoscopy, the detected laser points can be placed in the 3D space by backprojection. The reconstructed point cloud is the accumulation of the different laser detections at their capture position. For consecutive frames, as with stereo point clouds, the points have been filtered using a voxel grid filter, and accumulated using the updated pose of the robot.



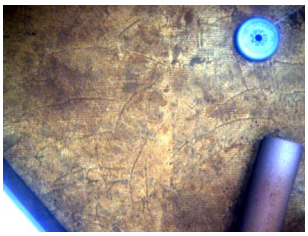
(a) Pool at HWU with the cartesian robot on top.



(b) Objects used to verify the 3D reconstruction accuracy.



(c) Image frame from the laser dataset.



(d) Left image frame from the stereo dataset.

Figure 2. Experimental setup. The objects shown were deployed in the pool and the sensors shown were moved in a lanw-moving pattern survey with a cartesian robot.

IV. Experimental results

The sensors were mounted, one at a time, on the cartesian robot and performed a 3 by 2 m lawn-moving survey while recording the imagery. In the case of the LbSLS, the sensor was mounted at 0.7 m of the pool bottom and the spacing between lines was 0.2 m. For the stereo camera, the spacing was 0.5 m and was mounted at 1.3 m.

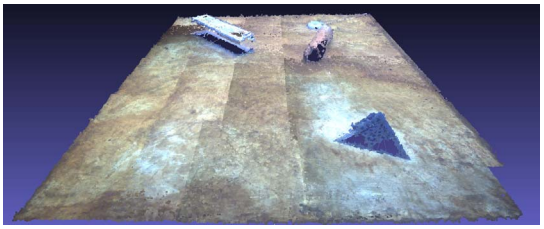
The pool is 4 meters long by 3 meters wide and 2 meters deep, and the Cartesian robot is mounted around the pool. It is able to carry an object on its mounting bracket and move it in the space without changing its orientation. This robot also is able to provide the position of the carried element in real time, avoiding the neeed to compute any visual odometry or dead reckoning on the acquired data. The position of the Cartesian robot is therefore treated as ground truth.

The LbSLS was attached using a pan and tilt unit to the robot. The pan and tilt unit helped to perform a better calibration of the system, and allowed us to set the camera totally perpendicular to the pool bottom. However, this unit has a considerable weight that increased the inertia of the robot and slightly overshot the robot's position. Whereas the stereo camera was simply attached using a lightweight pole.

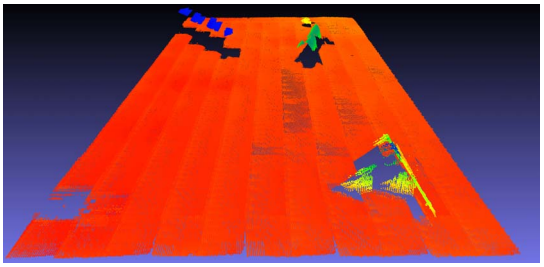
Four different objects were deployed in the pool to test the 3D reconstruction, as explained in section IV-A.

The pool, the cartesian robot, and the objects can be seen in figures 2(a) and 2(b). In figures 2(c) and 2(d) one frame from each dataset are shown.

The resulting three-dimensional reconstructions can be seen in figures 3(a) and 3(b). Note that for structured light the pool lights had to be dimmed to sense the laser, thus no RGB data has been captured. Depth is represented in a colour scale, being blue closer and red farther from the camera.



(a) Stereo reconstruction. 685872 points.



(b) Structured light reconstruction. Blue is closer and red is farther from the camera. 370261 points.

Figure 3. Sensor 3D reconstructions, available at http://srv.uib.es/pointclouds

Figure 4. Pool mosaic built from a subset of 40 images (left camera of stereo dataset). From [18].

A. Object dimensions and accuracy

In the pool we dropped four objects with known dimensions to check the accuracy of both systems. These objects were a $1\ m$ long $0.2 \times 0.2\ m$ square pipe, a $\varnothing 0.16$ by $0.75\ m$ round pipe, a $\varnothing 0.2$ by $0.04\ m$ lifting wheel with a $\varnothing 0.03\ m$ central hole in it, and a $0.207\ m$ tall triangular pyramid with a isosceles basis, whose side length is $0.51\ m$. In figure 4, a color mosaic of the pool bottom can be seen, with the object placed in their reconstructed positions.

In table I the true measurements from these objects and the measurements extracted from both pointclouds are shown.

The obtained measurements from the point cloud have been manually measured using CloudCompare software [19].

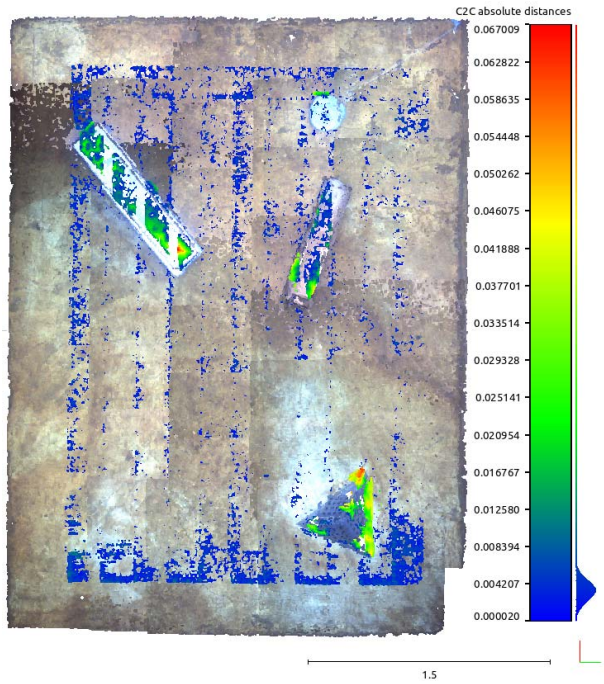


Figure 5. Distance from laser point cloud to stereo point cloud. Measurements in meters. Deep blue is almost zero error whilst green to red colour transition means higher error.

B. Distance between point clouds

Finally, the two point clouds have been registered using ICP to compare the distance from one point cloud to the other. As the stereo point cloud has a higher number of points, it has been taken as the reference model, and the laser point cloud as the test model. Therefore, the distances were computed from the points in the laser dataset to the closest point in the stereo.

In figure 5 the distance for each laser point to the closest stereo point is shown. In figures 6(a) to 6(d) the objects close-ups can be seen. Note that if there are no stereo points in one area but laser, then the error for the laser points will be big even if the points are well located. Otherwise, if there are stereo points but no laser points, there is no error. Note this behaviour in figure 6(a).

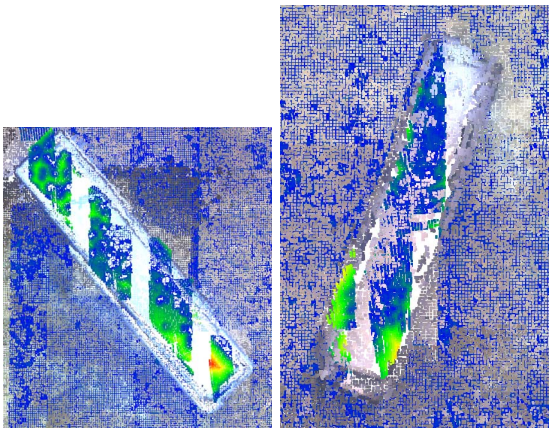
In figure 7 the frequency plot of the distance error is shown. A Gaussian distribution has been fitted with a mean of $3.56\ mm$, and a standard deviation of $1.256\ mm$. The long right side tail has been split at $0.01\ m$ in figure 7(b). The number of bins is the same as in figure 7(a), but the vertical scale has been resized to better show the distance distribution.

V. Conclusions and Future Work

The results shown comparing the two point clouds are promising for the LbSLS. Notice how in table I the z dimension accuracy is better for the LbSLS than for the stereo. However, in the other two dimensions this relationship is not

Table I
OBJECT MEASUREMENTS FOR THE DIFFERENT SENSORS

Object	**True dimensions (m)**	**Laser dimensions (m)**	**Stereo dimensions (m)**
Square pipe ($L \times H \times W$)	$1 \times 0.2 \times 0.2$	$1.038 \times 0.2102 \times 0.2396$	$1.034 \times 0.1967 \times 0.2113$
Round pipe ($D_{ext} \times L$)	$\varnothing 0.16 \times 0.75$	$\varnothing 0.16 \times 0.8390$	$\varnothing 0.1651 \times 0.7542$
Wheel ($D_{ext} \times D_{int} \times H$)	$\varnothing 0.20 \times \varnothing 0.03 \times 0.04$	$\varnothing 0.1886 \times \varnothing 0.045 \times 0.0511$	$\varnothing 0.1973 \times \varnothing 0.024 \times 0.0630$
Pyramid ($L \times H$)	0.51×0.207	0.5020×0.2064	0.517×0.2186



(a) Square pipe (b) Round pipe

(c) Wheel (d) Pyramid

Figure 6. Object distance close-ups. Deep blue is almost zero error whilst green to red colour transition means higher error.

so clear. This may have been caused due to the weight of the pan and tilt unit that carried the sensor. Its heavy weight increased the inertia of the system, overshot it, and caused an offset from the sensor position to the cartesian robot. This can be noted in the square pipe (figure 6(a)) when we compare the forward movement reconstruction to the backward movement. The four sensor travels can be paired 2 by 2, as there are 2 made in forward movement and 2 in backward. This happens throughout the course of the robot in the pool.

Regarding the comparison of the two point clouds, the distance error mean is low if compared to the distance from the sensor to the objects ($3.56\ mm$ at a distance of $0.7\ m$, a 0.51% error), and although there are points with an error greater than $1\ cm$, their number is low (6750 points compared to 370261 points, a 1.8%).

With these results, we can conclude that LbSLS and stereo systems can be used for 3D reconstruction taking into account several particularities. With a stereo camera one can get a good general overview of the underwater scenario, missing small details and accurate distances, whilst with LbSLS one can get sharper and clearer details, losing colour information.

Note the top of the square pipe in the stereo dataset. There are no points because of the lack of texture. The pipe is plain white and the block matching algorithm is unable to find any correspondence between left and right frames.

As a result, for long, high altitude surveys where there is enough texture and visibility, stereo data may be enough to recover the overall shape of the underwater environment. But prior to a manipulation where precise measurements and distances are desired, a small LbSLS can provide the information required.

Future work includes comparing these systems in different turbidity scenarios, together with a multibeam sonar, and finally mount and test the LbSLS on a AUV in a real underwater scenario.

ACKNOWLEDGMENTS

This work has been partially supported by grant BES-2012-054352(FPI) and contract DPI2011-27977-C03-02 by the Spanish Ministry of Economy and Competitiveness, FEDER Funding and by Govern Balear (Ref 71/2011).

REFERENCES

[1] J. S. Jaffe, "Computer modeling and the design of optimal underwater imaging systems," *IEEE Journal of Oceanic Engineering*, vol. 15, pp. 101–111, Apr. 1990.

[2] G. Bianco, A. Gallo, F. Bruno, and M. Muzzupappa, "A comparative analysis between active and passive techniques for underwater 3D reconstruction of close-range objects.," *Sensors (Basel, Switzerland)*, vol. 13, pp. 11007–31, Jan. 2013.

[3] M. Eric, R. Kovacic, G. Berginc, M. Pugelj, Z. Stopinsek, and F. Solina, "The impact of the latest 3D technologies on the documentation of underwater heritage sites," in *2013 Digital Heritage International Congress (DigitalHeritage)*, pp. 281–288, IEEE, Oct. 2013.

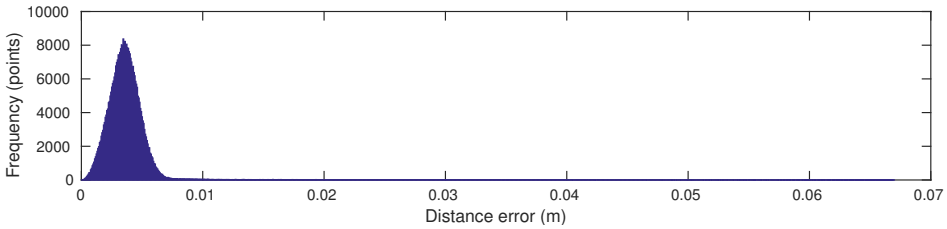
[4] A. Hogue, A. German, and M. Jenkin, "Underwater environment reconstruction using stereo and inertial data," *2007 IEEE International Conference on Systems, Man and Cybernetics*, 2007.

[5] C. Beall, B. J. Lawrence, V. Ila, and F. Dellaert, "3D reconstruction of underwater structures," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 0448111, pp. 4418–4423, IEEE, Oct. 2010.

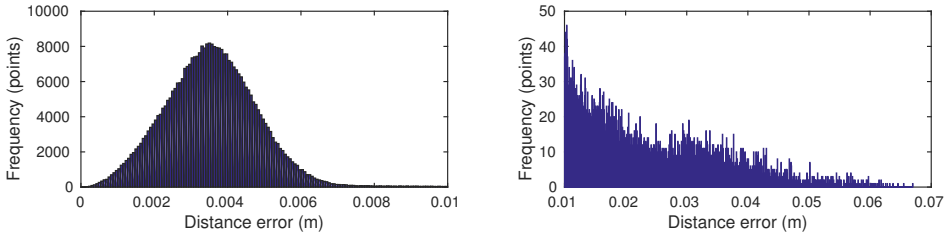
[6] P. L. Negre Carrasco, F. Bonin-Font, and G. Oliver-Codina, "Stereo Graph-SLAM for Autonomous Underwater Vehicles," in *Proceedings of the 13th International Conference on Intelligent Autonomous Systems (IAS 2014)*, 2014.

[7] G. Inglis and C. Roman, "Sonar constrained stereo correspondence for three-dimensional seafloor reconstruction," in *OCEANS'10 IEEE SYDNEY*, pp. 1–10, IEEE, May 2010.

[8] J. S. Jaffe and C. Dunn, "A Model-Based Comparison Of Underwater Imaging Systems," *Proceedings of the Ocean Optics*, vol. IX, no. 925, pp. 344–350, 1988.



(a) Distance error containing all the laser points.



(b) Distance error less (left) or greater (right) than $0.01\ m$.

Figure 7. Distance error frequency plots.

[9] S. Tetlow and J. Spours, "Three-dimensional measurement of underwater work sites using structured laser light," *Measurement Science and Technology*, vol. 1162, 1999.

[10] H. Kondo, T. Maki, T. Ura, Y. Nose, T. Sakamaki, and M. Inaishi, "Structure tracing with a ranging system using a sheet laser beam," in *Proceedings of the 2004 International Symposium on Underwater Technology (IEEE Cat. No.04EX869)*, pp. 83–88, IEEE, 2004.

[11] A. Bodenmann, B. Thornton, T. Nakatani, and T. Ura, "3D colour reconstruction of a hydrothermally active area using an underwater robot," in *OCEANS 2011 IEEE - Spain*, 2011.

[12] Q. Zhang, Q. Wang, Z. Hou, Y. Liu, and X. Su, "Three-dimensional shape measurement for an underwater object based on two-dimensional grating pattern projection," *Optics & Laser Technology*, vol. 43, pp. 801–805, June 2011.

[13] N. Törnblom, *Underwater 3D Surface Scanning using Structured Light*. PhD thesis, Uppsala Universitet, 2010.

[14] 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, pp. 508–518, July 2011.

[15] S. Narasimhan and S. Nayar, "Structured Light Methods for Underwater Imaging: Light Stripe Scanning and Photometric Stereo," in *Proceedings of OCEANS 2005 MTS/IEEE*, pp. 1–8, Ieee, 2005.

[16] C. Cain and a. Leonessa, "Laser based rangefinder for underwater applications," *2012 American Control Conference (ACC)*, pp. 6190–6195, June 2012.

[17] M. Massot-Campos and G. Oliver-Codina, "Underwater laser-based structured light system for one-shot 3D reconstruction," in *Proc. IEEE Sensors*, vol. 2014, 2014.

[18] E. Garcia-Fidalgo, A. Ortiz, F. Bonnin-Pascual, and J. P. Company, "Vessel Visual Inspection: A Mosaicing Approach," Tech. Rep. A-01-2015, Department of Mathematics and Computer Science, University of the Balearic Islands, Palma de Mallorca, March 2015.

[19] "CloudCompare: 3d point cloud and mesh processing software." http://www.cloudcompare.org/. Accessed: 2015-03-20.