Multi-View Underwater 3D Reconstruction using a Stripe Laser Light and an Eye-in-Hand Camera

Antonio Peñalver, J. Javier Fernández, Jorge Sales, Pedro J. Sanz
Computer Science and Engineering Department
University of Jaume-I
12071 Castellón de la Plana, Spain
e-mail: {penalvea, fernandj, salesj, sanzp}@uji.es

***Abstract*—Autonomous manipulation in unestructured underwater scenarios is a high challenging skill that has been poorly studied and is becoming more and more important in the last years. One of the main problems regarding the autonomous manipulation, is to find out the characteristics of the object which is going to be manipulated. This paper presents a new approach to obtain an accurate 3D reconstruction of this object. This approach consists in attaching a laser stripe emitter and a camera in the forearm of a robotic arm. Moving the arm, the laser scans the scene where the object is and, at the same time, the camera records the scan. Thanks to the arm and the position of the camera, the scene can be reconstructed from different views and from a position close to the object. The recorded images are processed to obtain the 3D position of the part of the scene projected by the laser. Before the intervention, a process of calibration is needed to calculate the relationship between each part of the system. Furthemore, in order to reduce the time of processing of the images recorded during the scan, an optimization algorithm is presented which consists in discarding, before the processing, the pixels of the image which do not contain relevant information. The approach herein presented and the optimization algorithm are tested using an underwater simulator.**

***Keywords*—3D Reconstruction; Underwater Laser Scanning; Autonomous Grasping; Laser-Camera Calibration, optimization algorithm**

I. Introduction

In recent years, exploration of the oceans and shallow waters is attracting the interest of many companies and institutions all around the world. In some cases because of the ocean valued resources, in others because of the knowledge that it houses for scientists, and also for rescue purposes. Remote Operated Vehicles (ROV's) are currently the most extended machines for doing these tasks. In this context, expert pilots remotely control the underwater vehicles from support vessels. However, due to the high costs and control problems involved in ROV-based missions, the trend is to advance towards more autonomous systems, i.e. Intervention Autonomous Underwater Vehicles (I-AUV's). One of the most challenging problems in the migration from tele-operated to autonomous vehicles is to automate grasping and manipulation tasks, mainly due to the limited availability of sensors, and to the continuous motion that naturally affects floating vehicles in the water.

In the autonomous underwater manipulation context, there exist only a few research projects. It is worth mentioning some of them. The project AMADEUS [1], in which two underwater robot arms were mounted on a cell to demonstrate bimanual skills. Nevertheless, this functionality demonstrated on a fixed frame, was never demonstrated on board an AUV, and hence with a mobile base. The TRIDENT project [2], demonstrated the capability to autonomously survey an area and recover an object from the seafloor, but still with some interaction with a human operator and its operations were restricted to shallow waters. In the context of that project, a framework to grasp objects with the user interaction just for the grasp planning phase was presented in [3]. This approach, focused on increasing autonomy using 3D data, is further developed with this framework.

The present research has been conducted within the Spanish Project TRITON [4], which finished at the end of 2014. The two scenarios considered in TRITON, which will be carried out by an autonomous underwater vehicle with a robotics arm, are the following:

- The vehicle docks to an underwater permanent observatory and it performs operations on an intervention panel, turning a valve and plugging/unplugging a connector. The structure is considered as a well known object, thus limiting the amount of information needed from the operator and increasing its autonomy.
- The robot performs a search and recovery intervention of a known object (i.e. black-box, amphora, etc.). This paper is focused in the recovery phase of this last scenario.

Concerning the second scenario, it is desirable to find out some quantitative measurements such as position, size, area, volume, shape, etc, in order to recover an object.

Different methods exist for obtaining 3D information from a given scene. They can be generally classified, according to the used sensor, into sonar-based, use of laser rangefinders, vision-based and use of structured light [3]. The main problem encountered when performing underwater 3D reconstructions is that underwater conditions are characterized by non-uniform lighting and poor visibility due to scattering and absorption in the medium [5].

Concerning the different kinds of methods to reconstruct a scene, sonar-based approaches are the most extended because acoustic devices do not suffer from turbidity [6]. However, apart from being the less cheaper devices, optical techniques provide higher accuracy [7]. Laser rangefinders are not commonly used probably because of the light absorption problem. Stereo vision is the cheapest alternative, and can also

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

provide the 3D reconstruction of the color information. But this approach is not useful in turbid waters, on untextured floors or in the darkness. Even increasing the light with artificial light could only worsen the situation since the light is strongly backscattered by the suspension particles in the water [6]. Structured light is also a cheap alternative and it allows imaging at greater range and higher contrast, thus reducing the backscattering. It produces interesting results in close range 3D acquisitions. Objects that have dimensions of less than one meter may be effectively acquired because the equipment can stay close to the objects, thus reducing the backscattering effect generated by the particles suspended in water [6].

This paper is organized as follows. In the next section, a new approach to reconstruct an underwater scene is detailed. Section III describes how to calibrate the system used in this approach to obtain a good reconstruction. The process of reconstruction is presented in Section IV. Section V presents a method for optimizing the process of reconstruction. Once the system is completely explained, some experiments are shown in Section VI before the conclusions.

II. Multi-View Laser Reconstruction

This paper presents a new Multi-View Laser approach to reconstruct an underwater scene. The scene typically contains an object (i.e. the target) that is desired to be autonomously recovered. In order to choose the best device to perform the reconstruction, the only assumption made is that the target is close enough to the system so it can be manipulated. Therefore, the level of scattering and absorption could be high. Taking into account the worst cases, the structured light device would be the best option because it performs well with poor visibility if the target is close to the system.

Concerning the structured light devices, it is also important the pattern that they project on the scene. There are a lot of patterns used, from a simple straight line [3] to a complex patterns like 15 slits pattern [8]. Complex patterns have the advantage of allowing a non dense reconstruction in just one shot, but the scattering effect increases because of the wider water illuminated. Another drawback could be the complexity of the process of matching the illuminated pixels with the projected pattern, in order to obtain its 3D positions. On the other hand, with a simple straight line, it is needed to scan the whole scene. This process is more time consuming but the obtained point cloud is much more dense.

The last important topic to take into account is the color of the projected light. The color spectra between blue and green is the optimum to avoid absorption and backscattering.

In our approach, we simulate a green laser that projects a straight line. This laser is placed on the forearm of the robotic arm. So, moving the arm, the laser light scans the scene like in [3]. The novelty of our approach is in the position of the camera, that now is placed on the forearm of the robotic arm (eye-in-hand configuration) (see Fig.1). In this new approach both the camera and the laser projector move along the arm during the scan of the scene. Up to now, the cameras were placed in a fixed position while the reconstruction.

With this approach, the object could be reconstructed from different point of views, obtaining a more complete 3D model

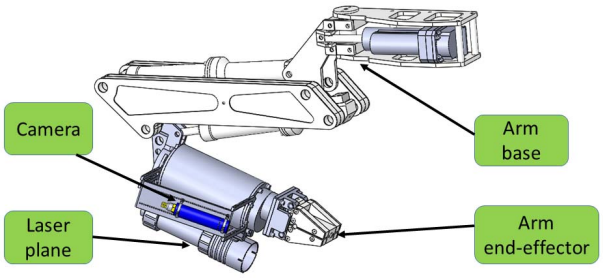

Fig. 1. Eye-in-Hand configuration to autonomously reconstruct underwater scenes.

of the object. Another advantage could be that using the arm, the camera can be closer to the object than in the usual approaches, obtaining more precision in the reconstruction.

III. Calibration

A good calibration of the whole system is a transcendental step for an accurate reconstruction. A little error in the calibration of any of the parts of the system, could lead to a big error in the position of each point of the 3D point cloud that represents the scene. In the proposed calibration methodology, which consists on three stages, a stereo camera is needed.

1) In the first stage, the intrinsic and extrinsic (${}^{S}\mathbf{M}_{eye}$) parameters of the stereo and eye-in-hand camera are obtained using a checkerboard pattern [9].
2) In the second stage, with the aid of a marker placed in the gripper of the arm, the transformation between stereo cam and the end-effector (${}^{S}\mathbf{M}_{E}$), is calculated [10]. Then, knowing this relationship and also the relationship between the stereo camera and the eye-in-hand camera (${}^{S}\mathbf{M}_{eye}$), it is easy to compute the transformation between the end-effector and the eye-in-hand camera ${}^{E}\mathbf{M}_{eye} = ({}^{S}\mathbf{M}_{E})^{-1} * {}^{S}\mathbf{M}_{eye}$.
3) For the third stage, the laser needs to be switched on. Then, using the stereo camera, the 3D position of the pixels projected by the laser are obtained by triangulation. With those 3D points, the RANSAC algorithm is used to determine the planar parameters of the laser plane [11]. These parameters are referenced to the stereo camera. But using the transformation between the end-effector and the stereo camera, it is possible to reference this plane with respect to the end-effector.

After the three stages, the eye-in-hand camera has been calibrated and the transformation between the eye-in-hand camera and the laser with respect to the end-effector (${}^{E}\mathbf{M}_{L}$, ${}^{E}\mathbf{M}_{eye}$) has been obtained.

IV. 3D Reconstruction

Once the system is over the scene to reconstruct, the robotic arm is moved at a constant velocity between two predefined joint positions. Using this movement, the laser is projected over the scene while the eye-in-hand camera captures images of the scene.

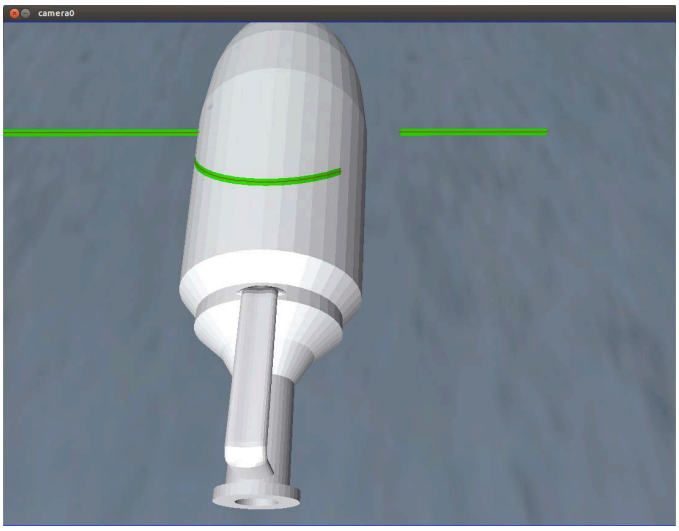

Fig. 2. Image recorded by the eye-in-hand camera during the reconstruction. The red points inside, the green laser, are the pixels illuminated by the centroid of the laser.

In order to extract the 3D information of the scene using those images, it is necessary to find out the pixels which are illuminated by the laser. Then, the 3D position of each pixel is triangulated taking into account the relation between the different parts of the system in the moment the image was captured.

The process of detection of the pixels illuminated by the laser from an image, and the process of triangulating their 3D position are better explained hereinafter.

A. Laser Peak Detection

The process of detecting which pixels of an image are illuminated by a laser is commonly known as *Laser Peak Detection*.

In our methodology, the image is converted from RGB to HSV color model. Then, the pixels which are out of a predefined threshold of huge, saturation and value, are discarded (this threshold has been obtained experimentally). Next, four operations are applied over the thresholded image. These operations are: erosion, dilation followed by dilation and erosion. With these operations, the pixels detected as a laser which are far from other pixels detected are discarded.

Due to the fact that the laser pattern is a straight line and that the eye-in-hand camera is placed parallel to that line, there can be just a point illuminated by the centroid of the laser at each column of the image. So, for each column of the image, the pixel, which is into the threshold, with the highest intensity is selected. Then, the *center of masses* algorithm [12] is applied to this pixel and the five pixels above and below it in the same column, in order to obtain, with a subpixel accuracy, the pixel illuminated by the centroid of the laser (see Fig. 2).

B. 3D Triangulation

Once the pixels illuminated by the centroid of the laser are detected, their 3D position must be calculated. For that, the methodology proposed in [3] is used.

After a properly calibration of the system the images given by the camera are undistorted. So, a pixel with row and column coordinates r and c, defines a line in projective coordinates given by the column vector:

$$\mathbf{l} = (l_x, l_y, 1) = \left(\frac{c - c_0}{p_x}, \frac{r - r_0}{p_y}, 1 \right) \tag{1}$$

Being c_0, r_0, p_x and p_y, the camera intrinsic parameters, i.e. the principal point in pixels and the focal length to pixel size ratio. If a pixel in the image belongs, in addition, to the projected laser plane, the intersection of the camera ray with the laser plane gives the 3D coordinates of the point. The line defined by the camera ray can be expressed with its parametric equation as:

$$\mathbf{P} = (X, Y, Z) = \lambda \mathbf{l} \tag{2}$$

If, in addition, a 3D point, $\mathbf{P}$, belongs to the laser plane, it holds that:

$$(\mathbf{P} - \mathbf{P_0})^T \mathbf{n} = 0; \tag{3}$$

where $\mathbf{n}$ is the plane normal given in camera coordinates, and $\mathbf{P_0}$ is a 3D point that belongs to the plane. Merging equations (2) and (3) leads to:

$$\lambda = \frac{\mathbf{P_0}^T \mathbf{n}}{\mathbf{l}^T \mathbf{n}}$$

and the final 3D coordinates of the point are given by $\mathbf{P} = \lambda \mathbf{l}$.

In order to compute these equations, it is necessary to know the laser plane equation with respect to the camera (given by the point $\mathbf{P_0}$ and a normal $\mathbf{n}$), which was obtained in the calibration process.

With that methodology, the 3D positions of the points are related to the camera (${}^{eye}\mathbf{M}_P$), but with the eye-in-hand approach, the camera is in continuous movement. In order to be able to join all the obtained 3D points in the same point cloud, those positions must be related to a fixed frame. The base of the arm could be a good frame to refer the points because it is fixed, and knowing the joint values of the arm in the moment the image was captured and using the Direct Kinematics of the arm, it is easy to refer the point to the base (${}^{B}\mathbf{M}_P$).

$${}^{B}\mathbf{M}_{eye} = DK(q_0 q_1 ... q_n)$$

$${}^{B}\mathbf{M}_P = {}^{B}\mathbf{M}_E {}^{E}\mathbf{M}_{eye} {}^{eye}\mathbf{M}_P$$

V. Optimization

Due to the limited memory capacity on the usual computers used in the AUVs, it is not possible to record all the captured images and to process them afterwards. So, when an image is captured, it is processed immediately, and the next images are discarded until the process is complete. It means that the more time you spend extracting 3D information from one image, the less dense your final 3D point cloud will be.

Moreover, the aim of the reconstructions proposed in this paper is to discover the characteristics of an object that is desired to be manipulated and it is supposed that the object is inside the arm workspace. It means that the part of the scene out of the arm workspace is not important. It is even desirable that the system does not reconstruct that part to avoid outliers and to keep the final point cloud as lightweight as possible without lossing any detail of the target.

In order to reduce the time of processing the images and taking into account the above-mentioned, a new method has been developed. This method consists on calculating the pixels of the image which can be discarded because in case they are illuminated by the laser, the part of the scene illuminated is out of the workspace of the arm (see Fig. 3). As a result, the images given to the laser peak detector are smaller and therefore their processing is faster.

Before the begin of the intervention, the user must introduce two parameters: (1) The max_radius which is the radius of the smallest sphere that wraps all the arm workspace, having as a center the base of the arm, and (2) the min_radius which is the radius of the bigger sphere with the base of the arm as a center, and that does not intersect with any part of the arm workspace.

The method, which is explained below, has 4 steps and needs to be applied once for each column of each image that is going to be processed.

1) The finality of the first step is to find the line which is the intersection between the laser plane and the plane which crosses the center of projection of the camera and the pixels of the column that is being processed. For this, two pixels of this column are selected and using the *3D Triangulation* algorithm explained in the Section II, the 3D positions of the two points that would be illuminated by the laser and seen in these pixels are calculated. These two points ($\mathbf{P_1}$, $\mathbf{P_2}$) belong and define ($\mathbf{L_1} = \overrightarrow{\mathbf{P_1P_2}}$) the line we was looking for.
2) The second step consist in calculating the intersection point between $\mathbf{L_1}$ and the two spheres with center the base of the arm and with radius max_radius and min_radius. Being $\mathbf{A} = (A_x, A_y, A_z)$ and $\mathbf{B} = (B_x, B_y, B_z)$ two points that belong to a line and $\mathbf{C} = (C_x, C_y, C_z)$ the center of a sphere and r its radius, the intersection points between the line and the sphere can be calculated by:

$$a = (B_x - A_x)^2 + (B_y - A_y)^2 + (B_z - A_z)^2$$

$$\begin{aligned} b = 2((&B_x - A_x)(A_x - C_x) + \\ &(B_y - A_y)(A_y - C_y) + \\ &(B_z - A_z)(A_z - C_z)) \end{aligned}$$

$$c = (A_x - C_x)^2 + (A_y - C_y)^2 + (A_z - C_z)^2 - r^2$$

$$\delta = b^2 - 4ac$$

$$d = \begin{cases} \nexists & \text{if } \delta < 0 \\ \frac{-b}{2a} & \text{if } \delta = 0 \\ \frac{-b \pm \sqrt{\delta}}{2a} & \text{if } \delta > 0 \end{cases}$$

$$\mathbf{intersection\_point} = (\mathbf{B} - \mathbf{A})d + \mathbf{A}$$

So, using the points obtained in the last step (P_1 and P_2) and the radius introduced by the user, the intersection by L_1 and the two spheres can be calculated.
3) In the third step, it is calculated in which rows of the image the points obtained in the previous step would be seen. For each sphere, there may be two intersection points, so the point that is placed behind the camera is discarded.
4) Finally, for each column, there are two points, one which intersects with the bigger sphere and the other with the smaller one. Merging equations (1) and (2) leads to:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} \lambda \frac{c - c_0}{p_x} \\ \lambda \frac{r - r_0}{p_y} \\ \lambda \end{pmatrix}$$

And the row is given by:

$$r = \frac{Y p_y}{Z} + r_0$$

Once the two rows for each column are calculated, the pixels which are not between these two rows are discarded and the resultant image is sent to the laser peak detector algorithm.

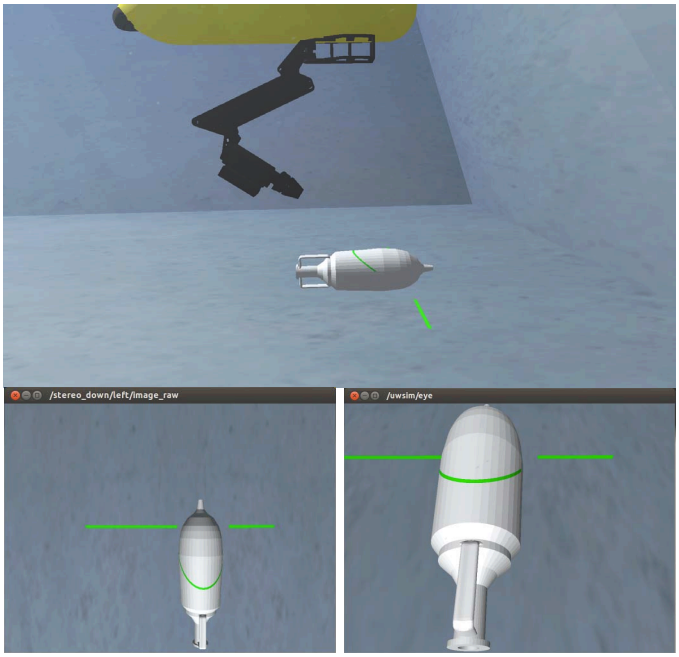

Fig. 4. ARM 5E with a stripe laser projector during a scan (top); Image recorded from the fixed camera (left). Image recorded from the eye-in-hand camera (right)

VI. Experiments

In order to validate the functionalities of the system, the underwater simulator UWSim [13] has been used. The model of the robotic manipulator light-weight ARM 5E [14] has been loaded to the simulator with a laser projector attached to its forearm. The laser projector simulates a stripe pattern which consists of a green rectangle of 1024 pixels wide and 10 pixels high. Two cameras are also loaded with the arm. One of them simulates the usual reconstruction approach. It has been placed in the base of the arm facing downwards and it is fixed during the scans. The other one, which simulates the multi-view approach, has been attached to the forearm of the robotic arm (see Fig.4).

Two kinds of experiments have been performed to show in the first case the advantages of the approach presented in this

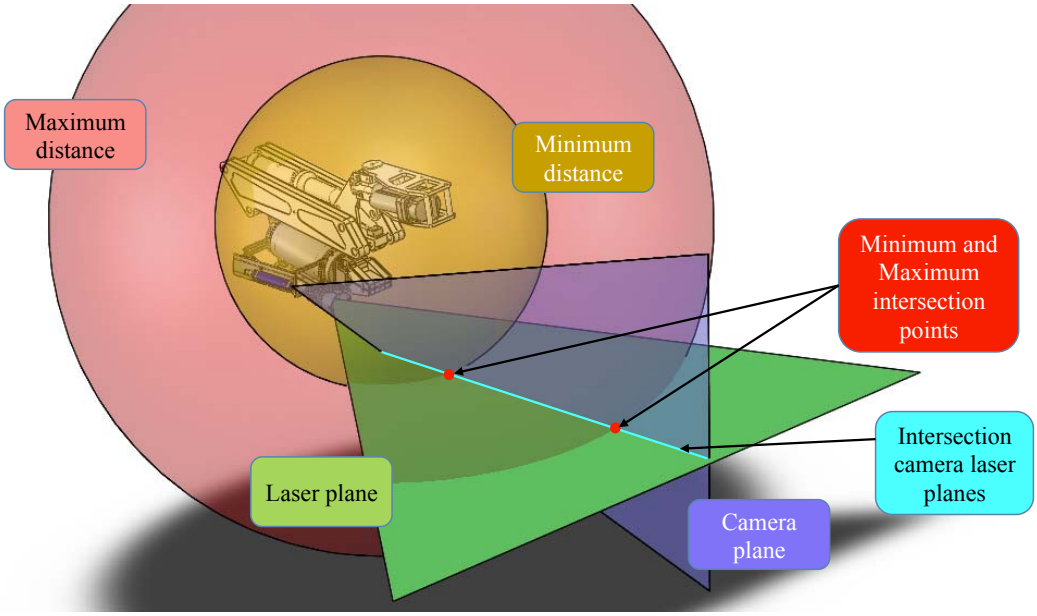

Fig. 3. Method to optimizate the time of extraction 3D iformation from an image.

paper with respect to the usual approaches and, in the second case, the time saved by the optimization algorithm during the laser peak detection.

A. *One view vs Multi-View Reconstruction*

Three different objects have been reconstructed using the fixed camera placed on the base of the arm and with the eye-in-hand camera, which is placed on the forearm of the arm (see Fig.5).

The first object is a skull. In the reconstruction of this object, it is easy to see how thanks to the position of the eye-in-hand camera, it can be close to the object during the reconstruction and it allows to obtain a much more accurate reconstruction. Moreover, it allows to clearly recognise difficult to reconstruct characteristics like the holes of the eyes and nose.

The second second object is a black-box mockup. In this case, the eye-in-hand approach allows to reconstruct the object from different views making possible the reconstruction of four sides of the black-box, compared to the two sides allowed by the fixed camera.

The last object is an amphora. The reconstruction of the handle of the amphora, which could be really important in a recovery intervention, needs a combination of both advantages previously explained. The camera needs to be close to the object to be enough accurate to reconstruct it, and different views are needed to obtain all the important characteristics of the object.

The main drawback of the multi-view approach is the time needed to perform the different scans.

B. *Optimization Test*

In order to test the optimization algorithm, the processing time of an image has been measured using different camera resolutions and different reconstruction limits (see Fig.6).

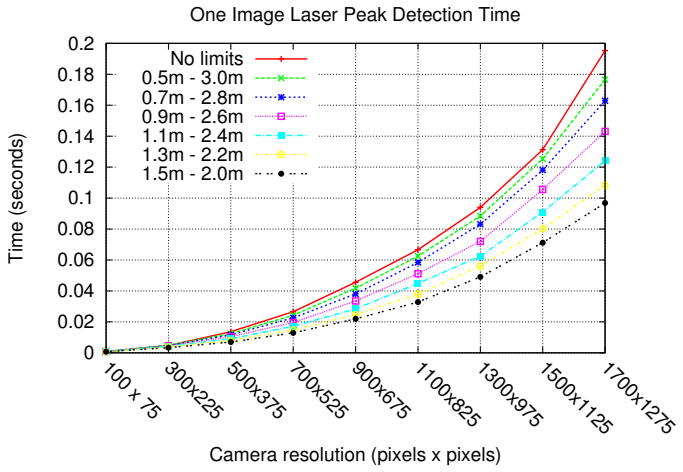

Fig. 6. Test of the optimization algorithm

The red line shows the average time of processing an image without using the optimization algorithm. The rest of the lines show the average of the time of calculating the pixels that can be discarded plus the time of processing the resulting image.

The graphic demonstrates that even when the limits are distant, the time saved when processing a smaller image is higher than the time spent in calculating the pixels that can be discarded.

VII. Conclusion

This paper has presented a new approach to autonomously obtain an accurate 3D reconstruction of an object in an underwater scenario. The approach consist in attaching a laser stripe emitter and a camera in the forearm of a robotic arm. While the arm is moving, the laser scans and the camera records the scene. The position of the camera makes possible to record the scene from different points of views and from

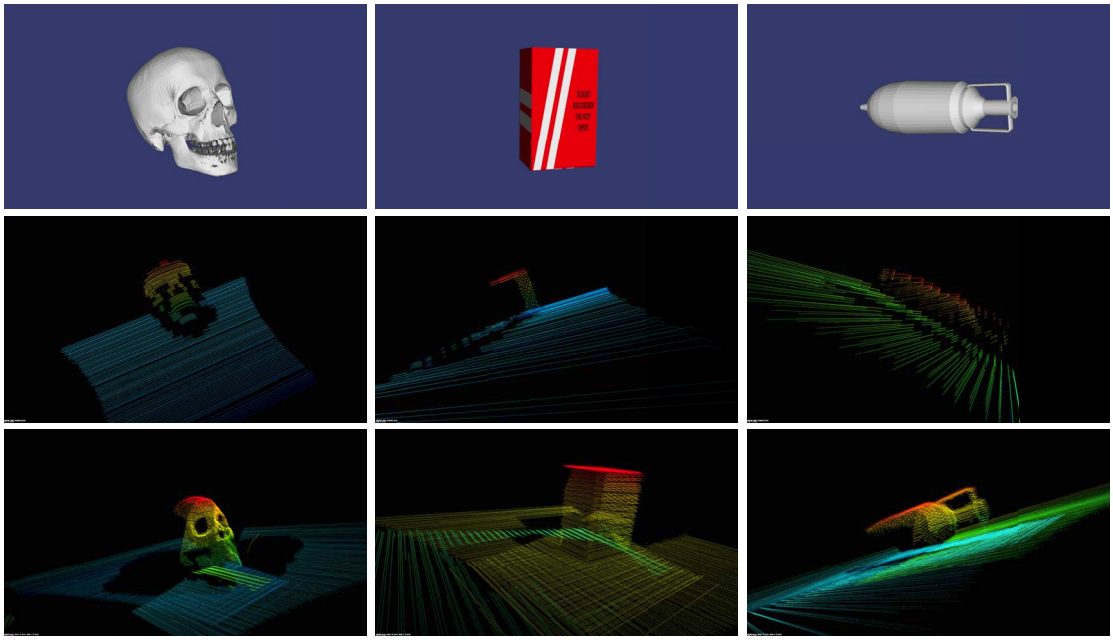

Fig. 5. Reconstruction of three objetcs: skull, black-box and amphora (left to right). Original model, fixed camera and multi-view approach camera (top to bottom)

positions close to the object. A laser peak detector and 3D reconstruction algorithms have been used to obtain a 3D point cloud from the images. An optimization algorithm has also been presented to reduce the time of processing those images. All the previous methods have been validated in an underwater simulator. From the obtained results, we can conclude that this approach improves the state of the art in this particular context. Different shaped objects was reconstructed in a very accurate and precise manner, and also reducing the computation requirements that usually requires those kind of reconstruction techniques

ACKNOWLEDGMENT

This work was partly supported by Spanish Ministry of Research and Innovation DPI2011-27977-C03 (TRITON Project), by Generalitat Valenciana PhD grant ACIF/2014/298 and by Universitat Jaume I grant PID2010-12.

REFERENCES

[1] D. M. Lane, J. B. C. Davies, G. Casalino, G. Bartolini, G. Cannata, G. Veruggio, M. Canals, C. Smith, D. J. O'Brien, M. Pickett *et al.*, "Amadeus: advanced manipulation for deep underwater sampling," *Robotics & Automation Magazine, IEEE*, vol. 4, no. 4, pp. 34–45, 1997.

[2] P. J. Sanz, P. Ridao, G. Oliver, G. Casalino, Y. Petillot, C. Silvestre, C. Melchiorri, and A. Turetta, "TRIDENT: An european project targeted to increase the autonomy levels for underwater intervention missions," in *OCEANS'13 MTS/IEEE conference*, San Diego, CA, 2013.

[3] M. Prats, J. Fernández, and P. Sanz, "Combining template tracking and laser peak detection for 3D reconstruction and grasping in underwater environments," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, 2012, pp. 106–112.

[4] N. Palomeras, A. Peñalver, M. Massot-Campos, G. Vallicrosa, P. L. Negre, J. J. Fernández, P. Ridao, P. J. Sanz, G. Oliver-Codina, and A. Palomer, "I-AUV docking and intervention in a subsea panel," in *Intelligent Robots and Systems (IROS 2014), IEEE/RSJ International Conference on*, Sept 2014, pp. 2279–2285.

[5] Y. Schechner and N. Karpel, "Clear underwater vision," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 1, June 2004, pp. I–536–I–543 Vol.1.

[6] C.-C. Wang, S.-W. Shyue, and S.-H. Cheng, "Underwater structure inspection with laser light stripes," in *Underwater Technology. Proceedings of the 2000 International Symposium on*, 2000, pp. 201–205.

[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] T. Kuroki, K. Terabayashi, and K. Umeda, "Construction of a compact range image sensor using multi-slit laser projector and obstacle detection of a humanoid with the sensor," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, Oct 2010, pp. 5972–5977.

[9] T. Svoboda, D. Martinec, and T. Pajdla, "A convenient multi-camera self-calibration for virtual environments," *PRESENCE: Teleoperators and Virtual Environments*, vol. 14, no. 4, pp. 407–422, August 2005.

[10] A. Peñalver, J. Pérez, J. J. Fernández, J. Sales, P. J. Sanz, J. C. García, D. Fornas, and R. Marín, "Autonomous intervention on an underwater panel mockup by using visually-guided manipulation techniques," in *Proceedings of the 19th World Congress of the International Federation of Automatic Control (IFAC)*, Cape Town, Africa, Aug. 2014, pp. 5151–5156.

[11] G. Inglis, C. Smart, I. Vaughn, and C. Roman, "A pipeline for structured light bathymetric mapping," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, Oct 2012, pp. 4425–4432.

[12] J. Forest, J. Salvi, E. Cabruja, and C. Pous, "Laser stripe peak detector for 3D scanners. a FIR filter approach," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, vol. 3, Aug 2004, pp. 646–649 Vol.3.

[13] M. Prats, J. Pérez, J. J. Fernández, and P. J. Sanz, "An open source tool for simulation and supervision of underwater intervention missions," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*. IEEE, 2012, pp. 2577–2582.

[14] J. Fernandez, M. Prats, P. Sanz, J. Garcia, R. Marin, M. Robinson, D. Ribas, and P. Ridao, "Grasping for the seabed: Developing a new underwater robot arm for shallow-water intervention," *Robotics Automation Magazine, IEEE*, vol. 20, no. 4, pp. 121–130, Dec 2013.