Extraction of Three Dimensional Structure From Optical Image Sequences

Mohcine Boudhane[1,2], Hamza Toulni[1] and Benayad NSIRI[1]
[1]University Hassan II, Faculty of Sciences, Ainchock B.P 5366 Maarif 20000, Casablanca, Morocco
[2]Faculty of Computer Science and Electrical Engineering, University of Applied Sciences, Grenzstr. 5, 24149 Kiel, Germany
Emails: { mrboudhane@gmail.com, hamza.toulni07@etude.univcasa.ma, benayad.nsiri@enst-bretagne.fr }

***Abstract*—Underwater imaging is a big challenge to discover and explore this environment. In this article, we proposed a 3D reconstruction method underwater using multi view 2D images in order to recovery of 3D shape from one or two 2D images. The conventional stereo approach based on perspective camera model cannot be directly applied and instead we used a spherical camera model to depict the relation between 3D point and its corresponding observation in the image. The stereo analysis problem consists of the following steps: image acquisition, camera modeling, feature extraction, image matching and depth determination.**

The sensor system generates a depth map in which each pixel represents a distance to the sensor. As a result, it can detect not only the presence of objects in the danger zone, but also their relative size and position. The system is implemented to reconstruct 3d scene using two or several three dimensional captures. It is also created to allow users to control the view perspective and obtain a better intuition of how the scene is rebuilt. Experiment results showed that the proposed approach offer a good preservation of the structure in sub-sea environment.

I. Introduction

With the growth experienced in recent years by the techniques of computer image processing, 3D reconstruction has been increasingly incorporating automated systems of different nature. three dimensional object reconstruction tasks are often domain sensitive requiring that the design of algorithms or points of interest extraction [1]. It s one such application domain important for:

- Image retrieval.
- Object recognition.
- Geo-localization for naval transportation.
- Robotics.
- Security underwater.
- · · ·

Various techniques have been described for acquiring a sequence of two dimensional images and reconstructing them into final three dimensional result. Some techniques is based on laser or light propagation such as **LIDAR** systems (**L**ight **D**etection **R**anging) [2]. Time-of-flight scanner is one of **LIDAR** methods based on camera and light propagation. In which the camera measures the amount of time it takes for light signal to reach and reflect off a subject. A single pulse of light is given, and for each point in the image the elapsed time is recorded, and distance is calculated relative to the speed of light [3]. However, the water property provoke a high attenuation of the light. This effect reduce the power of this method in subsea environment.

Different sensing modalities are used for obtaining a three dimensional representation underwater. Sonar imaging is often preferred over others. It s ability to capture in low visibility and the low coast in the use make it the most important sensor. Saucan in [4] propose a model-based adaptive filters are employed to achieve high-resolution 3D images of the sea bottom in real time. Authors in [5] build system able to estimate the 3-D positions of selected features (contained in a sequence of underwater 2-D acoustic images) by means of a tracking procedure and a triangulation operation. A sonar is a most useful sensor for 3d reconstruction. However, this technology miss many features in the pattern of the objects detected.

Camera is one of sensors used in this task. Many recent applications are aimed at deploying an underwater 3d image construction system. With the advent of new, low-cost 3D sensing hardware such as the Kinect [6], and continued efforts in advanced point cloud processing [7], 3D perception using camera gains more and more importance in robotics, as well as underwater object reconstruction or retrieval.

Compared with current state-of-the-art technology, the greater benefit of stereo camera technology as used in the scene construction and analysis, a stereo camera has considerable cost advantages over currently approved technologies. Two standard industrial cameras are fixed at a defined angle to one another. The adjustable lenses can be quickly adapted to fit the required scenario.

In this article, we consider the problem of 3D reconstruction and analysis using underwater images. The work is focused on estimating three-dimensional structures from two-dimensional optical image sequences in underwater environment, which may be coupled with local motion signals. In the following, we describe in Section 2 the system overview of our model, in which we describe the proposed approach. Section 3 gives the experimental result, where we demonstrate the performance of our model. Section 4 conclude.

II. System overview

Our brain is such a clever organ; it can do something that we don't know how it does. It catches a small difference between two images captured by our eyes and recognizes the depth and distance of the objects. In this chapter we propose a technique for viewing pictures in three dimensions. When you are looking for example at a stereogram, you can imagine that you are viewing the real scene from a window. Size, depth, and distance are perceptible as when viewing the original.

Our eyes are separated by a distance of about 6 to 7cm. It makes a difference in the point of view of each eye, and therefore the aspect of every scene is slightly different in the eyes. When these two different pictures fuse in the brain, it makes a 3D scene (Figure 1). The same principled have been applied in our method. The stereo camera capture a scene into different views. from these views we form the three dimensional structure included on it.

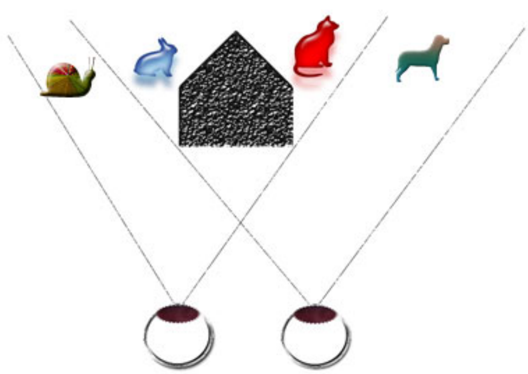

Figure 1. The Eyes 3D view.

Objectif: Given a set of flow fields or displacement vectors from a moving stereo camera over time, determine:

- The sequence of camera poses.
- Given optical flow or point correspondences, compute 3D motion (translation-rotation) and shape (depth).
- Recovery of 3D shape from 2D images.

In the following, we show different steps of the proposed approach (figure 2).

A. Image acquisition

The two dimensional images acquisition is critical step in the process for two main reasons. First, because the sequence of images will be assembled into a three dimensional image, the acquisition geometry must be known exactly to avoid distortion. Second, The localization of positions in space must not interfere with the performance of the system. Preprocessing of the input views is required. In this step we receive the raw 2D images. The pre-processing is done using bilateral filter.

B. Pre-processing

The use of bilateral filter is very important to the raw depth map to smooth it while preserving the borders. Then we project the pixel $\mathbf{u}$ to obtain a 3D point $\mathbf{p}$. With the projection of all the pixels, we generate a vertex map,

$$\mathbf{V}_k(\mathbf{u}) = D_k(\mathbf{u})\mathbf{K}^{-1}(\mathbf{u}), \quad (1)$$

where $\mathbf{K}^{-1}(\mathbf{u})$ is the ray corresponding to the pixel $\mathbf{u}$. Since the depth measurements is a regular map, we can compute the normal vectors $\mathbf{N}_k(\mathbf{u})$ using the neighbours.

C. Features selection

In this step we select the Keypoints (point of interest) in order to compute them. Keypoints identify a small number of locations where computing feature descriptors is likely to be most effective. Robust feature detection and stereo matching are crucial to building a good 3D model. Simple correlation-based features, such as Harris corners [9] or Shi & Tomasi features [10], are commonly used in vision-based SFM and SLAM, and others application in 3d scene reconstruction and analysis. In the proposed method, we use SIFT detector for feature extraction. because it s able to give better option for matching visual features from varying poses.

D. Camera calibration

Good Calibration of stereo platform is important to create a high-quality 3D reconstruction. Our stereo platform was calibrated by placing two cameras in the ocean floor and recording video of it from various angles. Both cameras were calibrated independently using the Camera Calibration toolbox [11].

$$K = \begin{bmatrix} f_x & s & c_x \\ & f_y & c_y \\ & & 1 \end{bmatrix} \quad (2)$$

where

$$\begin{cases} f_x \text{ and } f_y \text{ are focal lengths.} \\ s \text{ is the the skew.} \\ (c_x, c_y) \text{ is the position.} \end{cases}$$

A 2D image point (u_l, v_l) in homogeneous coordinates taken from the left camera with an associated disparity d can then be reprojected to 3D coordinates in the left camera coordinate frame by:

$$p_i = Q. + \begin{bmatrix} u_l \\ v_l \\ d \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} \quad (3)$$

where Q is the reprojection matrix resulting from the rectification, and T_x is the baseline of the stereo platform

$$\begin{bmatrix} 1 & 0 & 0 & -c_x \\ 0 & 1 & 0 & -c_y \\ 0 & 0 & 0 & f \\ 0 & 0 & \frac{-1}{T_x} & \frac{c_x - c'_x}{T_x} \end{bmatrix} \tag{4}$$

The 3D coordinates with respect to the left cameras coordinate frame are then given by

$$(\frac{x}{w}, \frac{y}{w}, \frac{z}{w})$$

E. Disparity matrix

Depth map is an image or image channel that contains information relating to the distance of the surfaces of scene objects from a viewpoint (shown equation 4).

F. Motion analysis (using optical flow)

The approach to motion analysis presented in [8] tries to determine the instantaneous displacement of every pixel between successive images. If one assumes that the intensity change is due only to camera and objects movements, the instantaneous speed of the pixel satisfies the following optical-flow:

$$\frac{\partial I(x,y,t).u}{\partial x} + \frac{\partial I(x,y,t).v}{\partial y} + \frac{\partial I(x,y,t)}{\partial t} = 0, \tag{5}$$

where $I(x, y, t)$ is the image intensity, and the partial derivatives are generally estimated from the image by using spatial and temporal differences. If we want (1) to be a good approximation, the 2-D speed should be low, usually one or two pixels per frame.

G. Point cloud

A point cloud is just what it sounds like: a cloud of points in 3d space that contains one or more channels of data (lighting, occlusion, area, etc) at each point. Point clouds are important because they are the immediate precursor to a brick map, which is a useful structure for caching data as a 3d texture. A point cloud is a data structure used to represent a collection of multi-dimensional points. These points usually represent the X, Y, and Z geometric coordinates of an underlying sampled surface.

H. 3D reconstruction

Here we obtain the final structure of the 3D object.

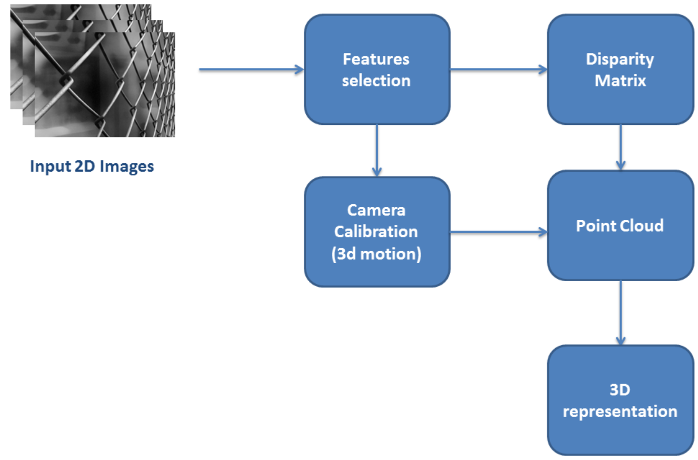

Figure 2. 3D reconstruction process.

III. Experimental results

We have developed a system for real-time monitoring based on smart algorithms presented. The system is implemented on a standard PC. we use a different images on different size and complexity.

Framework used:

- `OpenCV`: (Open Source Computer Vision) is a library of programming functions mainly aimed at real-time computer vision.
- `PCL library (Point cloud library)`: PCL is a large scale, open project for 2D/3D image and point cloud processing (in C++, w/ new python bindings).

The method consist to build a three dimensional representation of objects. this method refers to the process of estimating three-dimensional structures from two-dimensional image sequences which may be coupled with local motion signals. To find correspondence between images, features such as corner points (edges with gradients in multiple directions) are tracked from one image to the next. From these features, we made the camera calibration to find the 3D motion in rotation and translation. The surfaces of the object is represented by Disparity matrix (depth map). This last refers to the apparent pixel difference or motion between a pair of stereo image.
From the depth map and calibration matrix we obtain so called point cloud, that show a object structure as several points. Then we obtain the surface of the constructed object (Figure 3). Figure 4 show an example that resume all steps

IV. Conclusion

In this paper, we presented a method for efficient stereo matching and 3D reconstruction to be used in a more complex

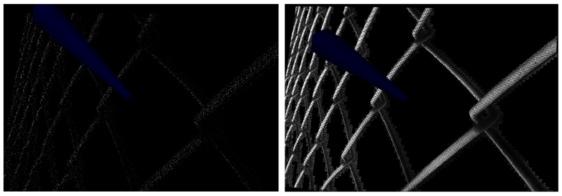

Figure 3. An example of representation 3D. The left image is Point cloud representation and the image in the right side is the 3D representation.

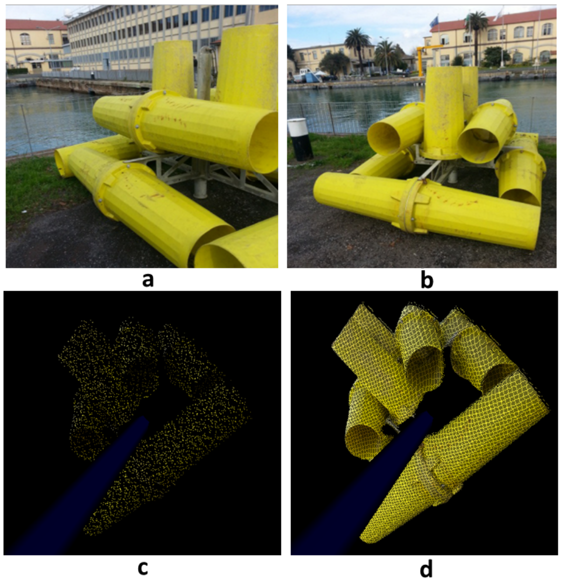

Figure 4. An example of representation 3D.
(a) left image, (b) right image, (c) object constructed on point cloud, (d) 3d object.

system in underwater environment. This algorithm was done and tested successfully on scenes with different shapes. The output of this algorithm provides good details in the constructed objects. In our future work, we will extend this work in the for scene analysis and object identification.

REFERENCES

[1] Fahiem, M.A. Haq, S.A. ; Saleemi, F. ”A Review of 3D Reconstruction Techniques from 2D Orthographic Line Drawings ” Geometric Modeling and Imaging, 2007. GMAI ’07, pp. 60-66.

[2] M. Kada, L. McKinley ”3D BUILDING RECONSTRUCTION FROM LIDAR BASED ON A CELL DECOMPOSITION APPROACH ”

[3] I. Gavriel, and G. Yahav. ”3D Imaging in the Studio and Elsewhere.” Proceedings of SPIE. 4298 (2003): 48.

[4] A.Saucan, C.Sintes, T.Chonavel, J.M. Le Caillec ”Model-Based Adaptive 3D Sonar Reconstruction in Reverberating Environments”IEEE TRANSACTIONS ON IMAGE PROCESSING 2015. Vol 24

[5] A.Trucco ”Extraction of 3-D Information From Sonar Image Sequences” 2003 IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICSPART B: CYBERNETICS, VOL. 33, NO.4

[6] Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, and Andrew Fitzgibbon ”KinectFusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera” ACM Symposium on User Interface Software and Technology 2011.

[7] R.B. Rusu, S. Cousins ”3D is here: Point Cloud Library (PCL)” PCL tutorial www.poitcloudlibrary.com.

[8] Jepson, A.D. Fleet, D.J. ; El-Maraghi, T.F. ”Robust online appearance models for visual tracking” Pattern Analysis and Machine Intelligence, IEEE Transactions on (Volume:25 , Issue: 10) p.1296-1311 .

[9] C. Harris, M. Stephens, ”A combined corner and edge detector” Proceedings of Alvey Vision Conference, pp.147151.

[10] Shi J., Tomasi C. , ”Good features to track” IEEE Conference on Computer Vision and Pattern Recognition (1994) pp.593-600.

[11] C. Beall, B. J. Lawrence, V. Ila and F. Dellaert ”3D Reconstruction of Underwater Structures” 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems. pp4418-4423, toolbox is available in: http://www.vision.caltech.edu/bouguetj/calib_doc/