2017 3rd International Conference on Control, Automation and Robotics

Evaluation of Underwater 3D Reconstruction Methods for Archaeological Objects: Case Study of Anchor at Mediterranean Sea

Yadpiroon Onmek, Jean Triboulet, Sebastien Druon, Arnuad Meline, Bruno Jouvencel
LIRMM Laboratory, University of Montpellier - CNRS UMR 5506,
161 rue Ada, 34090 Montpellier Cedex 5, France
e-mail: Omnec@lirmm.fr, Triboulet@lirmm.fr, Druon@lirmm.fr, Meline@lirmm.fr, Jouvencel@lirmm.fr

***Abstract*—The objective of this paper is to develop 3D underwater reconstruction of archeology object, which is based on a mono-camera. The underwater images are obtained from a calibrated camera system. We first solve the problem of image processing by applying the well-known filter, therefore to improve quality of underwater images. The features of interest between image pairs are selected by well-known methods: a FAST detector and FLANN descriptor. Subsequently, the RANSAC method is applied to reject outlier points. The putative inliers are matched by triangulation to produce sparse point clouds in 3D space, using a pinhole camera model and Euclidean distance estimation.**

Keywords-3D reconstruction; underwater archeology; monocamera; computer vision

I. INTRODUCTION

During recent years, 3D reconstruction from single camera vision systems has been a topic widely studied in the computer vision community, computer graphics, and photogrammetry. The essential 3D modeling applications are robot navigation, visual inspection, virtual reality, and so on. In the context of archeology, several researchers have been developing novel approaches to produce 3D models of objects as well as scenes.

Numerous methods exist, but they are not applicable to all objects and environments. Indeed, they depend on knowledge of the system as well as the environments. Especially in underwater environments, there are several limitations when working with underwater images. It is not easy to access and recover 3D information because of the possibility of encountering poor experimental conditions. More importantly, we must deal with various light conditions, loss of color and contrast in significant depth, the effects of several noises, as well as unclear water. The main purpose of this paper is the evaluation of 3D reconstruction methods for archeological objects (an Anchor object) in underwater environments. This project includes two parts: the first part correspond to the 3D reconstruction of constrained underwater environment with a video camera. The Second part is the multimodal aspect of the problem with Sonar system. The work presented in this paper focuses on the first part of the project and deals with the 3D reconstruction of the underwater environment with a video camera.

In this work we were working with The research group for naval archeology (GRAN). The underwater images sequences were acquired by divers at the La Grande Motte, Mediterranean Sea located in the Herault department in southern France. The 3D reconstruction pipeline is presented in Fig. 1.

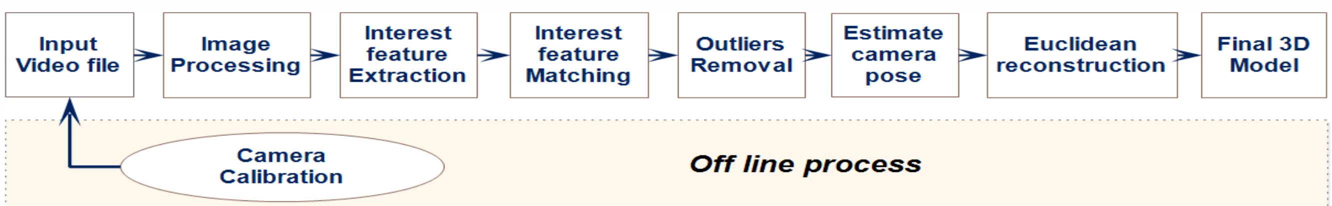

Figure 1. The 3D underwater reconstruction system overview

This paper is organized as follows: In section II a brief overview of 3D reconstruction problems is presented. Section III describes the method applied for 3D reconstruction in the case of an underwater environment. The evaluations of the system and results are described with noise-simulating underwater conditions in section IV. In the last section, discussion, conclusions as well as future works are presented

II. RELATED WORKS

This section briefly discusses some earlier works related to the detection of features of interest, matching point-sets as well as the reconstruction of 3D models. Comparative performances of methods have been published, assessing the detection performance features and image matching algorithms. The system of [13] compared several well-known feature detectors and descriptors. To compare each combination performance objectively, the effects of JPEG compression, zoom and rotation, blur, viewpoint as well as illumination variations have been investigated in terms of precision and recall values. Similarly, works from [16], proposed to investigate the performance of the SIFT and HARRIS methods. They simulated a noise filter in surface images to compare the percentage of inliers. More recent systems built on this same approach [1] considered methods to improve color quality and contrast of underwater images that do not need a prior knowledge of the scene. Finally, SIFT and SURF descriptors were used to compare the computation time.

The procedure of some researchers was to create a digital model and physical replicas [2] using undetermined images to estimate 3D urban models. They give no information about the devices that took the pictures, but their methods engage a prior knowledge of the scene. In the same way,[19] proposed a method for 3D object reconstruction to elaborate the 3D model from pictures in an automated way. Concerning robust photometric implementation, a bundle adjustment is used to create a standard package. Another essential point in [8] presented a multi-view stereo algorithm capable of computing high quality reconstruction of a range of scenes from large, shared, multiuser picture collections available on the Internet. This capability opens the possibility of computing accurate geometric models of several sites such as cities and landscapes.

Focusing on underwater reconstruction, [15] stated that to carry out the 3D reconstruction of natural underwater scenes from images obtained from calibrated single cameras, underwater constraints in the camera distortion model must be taken into account.

Consequently, [12] combined a 3 DOF inertial sensor and a calibrated stereo ring to estimate the trajectory and produced a 3D dense map. [4] used a stereo camera to capture images and reconstruct indoor environments for robot navigation.

III. METHODS

A. Camera Calibration

In 3D computer vision fields, camera calibration is a mandatory step, and it is an important task for Euclidean reconstruction [22]. The process of camera calibration is to impose the characteristics of the transformation between an object in 3D space and the 2D image observed by a camera. The transformation includes the characteristics of the camera, such as intrinsic parameters including focal length, the principal points of the camera as well as distortion, and extrinsic parameters to present the orientation and camera locations, such as a rotation matrix and translation vector. 1) Intrinsic parameter: In this work we used Pinhole camera model. The intrinsic matrix transforms 3D camera coordinates to 2D homogeneous image coordinates. This perspective projection is modeled by the ideal pinhole camera. The intrinsic matrix is parameterized by [11].

$$K = \begin{bmatrix} fx & s & Cx \\ 0 & fy & Cy \\ 0 & 0 & 1 \end{bmatrix} \tag{1}$$

where fx and fy are focal lengths, s is the skew, which is 0 in our work, and the principal point (cx, cy). The extrinsic calibration parameters consist of the $3x3$ rotation matrix R and the $3x1$ translation vector t which describe the pose of the camera movement.

B. Image Processing

In case of underwater images sometimes need to improve or filter images before 3D reconstruction processing. Because, it has noise in the images that effect from light, deep or sand, dirty water.

1) Gaussian filter: In image processing, a Gaussian blur [17] (also known as Gaussian smoothing) is the result of blurring an image by a Gaussian function. It is a widely used effect in graphics software, typically to reduce image noise and reduce detail. The visual effect of this blurring technique is a smooth blur resembling that of viewing the image through a translucent screen, distinctly different from the bokeh effect produced by an out-of-focus lens or the shadow of an object under usual illumination. Gaussian smoothing is also used as a pre-processing stage in computer vision algorithms in order to enhance image structures at different scale space representation and scale space implementation.

2) Median filter: The median filter is a nonlinear digital filtering technique [17], it is used to remove “salt and pepper” noise. The template size slider defines how much filtering takes place. Median filtering is a nonlinear method used to remove noise from images. It is widely used as it is very effective at removing noise while preserving edges. It is particularly effective at removing salt and pepper type noise. The median filter works by moving through the image pixel by pixel, replacing each value with the median value of neighboring pixels. The pattern of neighbors is called the “window”, which slides, pixel by pixel over the entire image 2 pixels, over the entire image. The median is calculated by first sorting all the pixel values from the window into numerical order, and then replacing the pixel being considered with the middle (median) pixel value.

C. Detection and Extraction

The next step is to find the correspondence between the first and second images. In other words, when a scene or an object must be reconstructed in 3D, detection and matching points in the images are the most crucial factors for model accuracy. The 3D model will be of low quality or even completely wrong if the feature extraction and matching steps introduce errors. We investigated several feature detectors and many of them are very good, such as the basic HARRIS corner detector proposed by Harris and Stephens [10], which is an operator to detect corners in images. And the SIFT method (Scale Invariant Feature Transform) proposed by [14] which is a local detector and descriptor. Also, the SURF method (Speed up Robust Features) proposed by [9], is a detector and descriptor that is invariant to change of scale or rotation. It uses integral images, which results in a significant performance boost. However, as we are planning to work with a video file and multiple images in our next work, in this experiment we chose the FAST method to detect the point of interest and used the FLANN method that is available in the OpenCV open source for matching the points of interest that we identified from the detection step. For more detail are shown in below.

1) FAST detector method: A feature from the accelerated segment test (FAST) was originally proposed by [5] [6]. This is a well-known corner detection method. The main idea is to combine edge and point-based tracking systems to emphasize the problem of a real-time 3D model based on tracking systems. The edge and point-based systems complement each other and can establish a rather robust system. To carry out a corner detector, a circle of 16 pixels around the corner candidate is considered. The test is performed on a Bresenham circle. The classification of the positive and negative corner is based on the pixels which produce the extreme information and is measured using the ID3 algorithm [18] to determine whether it is a corner. Non-maxima suppression is subsequently applicable on the sum of the absolute difference between the pixels in the circle as well as the center pixel [20]

D. *Suppression of False Matching*

The RANSAC algorithm proposed by [7] is intended to verify the matches between key points. The concept of this algorithm is to interpret a method and find its parameters with N subsets of n random data. From these N estimates of the model, classifying matching points as excellent or weak matches is possible. The pairing validation of points is based on the measurement of error between the projected point from the first image onto the second image as well as the points matched in the second image. The algorithm is used as follows: selecting a set of eight random points and estimating the fundamental matrix, then calculating for each point the distance between the projected point in the second image and the epipolar line. If the distance exceeds a certain threshold, the points are rejected. The remaining items are grouped in K set. The process is iterated N times and the set with the greatest number of elements is selected. Finally, the fundamental matrix with these points is estimated.

E. *Camera Pose Estimation*

In the next step, when we have a matching point between first and second images. Then we can use them to find the camera movement or we call camera pose. To begin with the fundamental matrix F is a 3x3 matrix which relates corresponding points in stereo images. In epipolar geometry, with homogeneous image coordinates, x and x', of corresponding points in a stereo image pair, Fx describes a line (an epipolar line) on which the corresponding point x on the other image must lie.

$$x^T F x = 0 \quad (2)$$

Then the essential matrix method proposed by [11] is used. The aim of this method is to obtain the pose of the second camera with respect to the first camera. The essential matrix E is calculated from fundamental F and calibration K matrices obtained previously.

$$E = K^t \times F \times K \quad (3)$$

Calculating the single value decomposition *SVD* [21] of the essential matrix grants the camera pose (*R* and *t*). In the first place is defined the projection matrix $P = K[I|0]$ for the first camera and $P2 = K[R|t]$ for the second one. Subsequently, the rotation matrix R and the translation vector t are estimated.

F. *3D Reconstruction*

We used all the parameters we have applied previously, including the essential matrix that obtains the pose of the camera from the first image to the second image, the projection matrix P for the first images and P2 for the second ones. Subsequently, the rotation matrix R and the translation vector t are estimated. Finally, a triangulation from the inliers found earlier and the projection matrices is used to create the 3D reconstruction. The 3D model is then retained to remove some irregular points that are locally isolated. The 3D Delaunay triangulation was used to create a mesh texture as well as rendering on the 3D model.

IV. EXPERIMENT AND RESULTS

A camera/camcorder and a sonar system were used for these experiments. For the image pair of archeological objects, in this experiment we used an Anchor object. Firstly, we used the mono camera to take a video of the object in an underwater environment. Secondly, we extracted the file video into an image sequence. Thirdly, as underwater images are normally not very clear, and therefore there is a problem with light, color, and dirty water, we need to improve the quality of the images by using an image processing technique. Fourthly, we choose the image pair for the experiment based on there being enough movement between the image pairs and also enough overlap between them. Finally, the features of interest are detected in each of the images and the points of interest between the image pair are matched. These features are triangulated from the 2D image into 3D point clouds. The OpenCV open source was used to develop an application to carry out 3D reconstruction, while OpenGL was used for visualization of the 3D model. Moreover, we have the information from the sonar system that Anchor object is located about 12 meter at the seafloor.

A. *Camera Calibration*

In this experiment, the camera used is a Nikon model D7000. The resolution used for picture acquisition is 1920x1080 pixels. The camera model in this experiment is a pinhole. Our single camcorder was calibrated by placing a calibration chessboard on the lake floor and recording a video of it from various angles. Then the camera calibration platform based on [22] and [3] was used. The intrinsic parameters are estimated for each camera, such as the focal length, the principal point, as well as the distortion.

B. *Image Processing*

As in our original underwater images in this experiment, from the video taken in the lake, there is some effect from white points. So we need to improve our original images to achieve sufficiently good quality to be able to extract and find the points of interest in the images. Initially, we modified the images by gathering the Median filter kernel size 9, in order to reduce the white and black points (salt and pepper points). In addition, we applied a *5x5* Gaussian kernel to modify blurred images.

C. 3D Model Reconstruction

1) Detection and Extraction: Robust feature detection and feature matching are crucial to building a robust 3D model. The matching step is the essential point of the 3D reconstruction. To begin the detection, the FAST method is used to identify the detected features of interest in the stereo images. The FLANN method is used to match those features in image pairs. Finally, the RANSAC method is used to reject inconsistent matches. Inlier features are points that have a correct match with the initial image pairs.

TABLE I. THE RESULT FOR DETECTION, EXTRACTION AND MATCHING POINT.

Image	Detection	Matching	Final Inliers	% of Inliers
First Image	4,797	1,767	876	49.58
Second Image	3,499			

The results in Table I show that the features of interest detected on the first and second underwater images are more numerous. However, not all of the features of interest can be matched between image pairs. And several false matches are represented. Finally, the RANSAC method was used to remove false matched points

2) Reconstruction: to produce the 3D stereo model, a triangulation is performed using the inliers and the projection matrix found previously. Then, the mesh structure is achieved through a 3D Delaunay triangulation. OpenGL is used for rendering and textures of the 3D model. The results of rendering and texture are presented. The final underwater 3D reconstruction model of the filtered images is shown in Fig. 2.

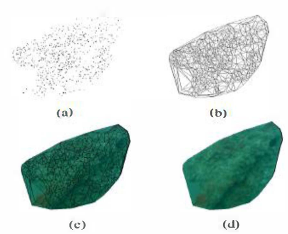

Figure 2. The 3D model of Anchor underwater image. (a) 3D points (b) Delaunay from 3D points (c) Delaunay with texture (d) 3D model of underwater images.

Figure 3. The final 3D reconstruction model of Anchor object.

We present the result of the underwater 3D model of Anchor underwater images, turning it in other views to show the model, as seen in Fig. 3.

V. CONCLOUSION

The goal of this work was to achieve a robust 3D reconstruction of archeological objects with a single camera camcorder system in an underwater environment. To begin with, image processing was used to improve the quality of the underwater image. Then, to achieve this purpose, we have to make sure that the feature points and the matching method are robust enough to the noise condition. Finally, the Euclidian reconstruction method is used to create a 3D model of the archeological object.

In future work, we plan to work with a video of the archeological object obtained with a stereo system instead of a single camera system and to take numerous views of the object to carry out 3D point clouds. Then tracking and 3D mapping techniques will be used to produce a more precise underwater 3D reconstruction model. Moreover, we will apply the fusion of this 3D information with a sonar map of an underwater archeological site.

ACKNOWLEDGMENT

The first author greatly thanks to the Royal Thai Government for providing PhD scholarship.

REFERENCES

[1] Mahiddine Amine, Seinturier Julien, Bo Jean-marc, Drap Pierre, and Merad Djamal. Performances analysis of underwater image preprocessing techniques on the repeatability of sift and surf descriptors. In WSCG 2012: 20th International Conference on Computer Graphics, Visualization and Computer Vision, 2012.

[2] L. Barazzetti. A trip to rome: Physical replicas of historical objects create in a fully automated way from photos. the First int.conf. on Humancomputer interaction, tourism and cultural heritage, HCITOCH'10, (Berlin, Heidelberg), pages 63–72, 2011.

[3] J. Y. Bouguet and P. Perona. Camera calibration from points and lines in dual-space geometry. technical report, California Institute of Technology, 1998.

[4] J. Lim A. Ranganathan M. Pollefeys C. Haene, C. Zach. Stereo depth map fusion for robot navigation. in IEEE/RSJ Int. Conf. on Intelligent

[5] Robots and Systems (IROS), 2011.

[6] Reid Porter Edward Rosten and Tom Drummond. Faster and better: A machine learning approach to corner detection. IEEE Trans. Pattern Analysis and Machine Intelligence, 32:105–119, 2010.

[7] Tom Drummond Edward Rosten. Machine learning for high-speed corner detection. Computer VisionECCV 2006, pages 430–443, 2006.

[8] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM, 24(6):381395, 1981.

[9] M. Goesele, N. Snavely, B. Curless, H. Hoppe, and S. M. Seitz. Multiview stereo for community photo collections. Proceedings of ICCV 2007, Rio de Janeiro, Brasil, 14-20 October 2007.

[10] Tuytelaars H. Bay and L. Van Gool. Surf: Speeded up robust features. In European Conference on Computer Vision, 1(2), May 7-13 2006.

[11] C. Harris and M. Stephens. A combined corner and edge detector. In Proceedings of the Alvey Vision Conference 1998, pages 147–151, 1998.

[12] R. Hartley and A. Zisserman. Multiple view geometry in computer vision. Cambridge University Press, 2003.

[13] A. Hogue, A. German, J. Zacher, and M. Jenkin. Underwater 3d mapping: Experiences and lessons learned. computer and robot vision, Canadian conference, page 24, 07-09 June 2006.

[14] Sahin Isik and Kemal Ozkan. A comparative evaluation of wellknown feature detectors and descriptors. International Journal of Applie Mathematics, Electonics and Computers of computer, 3(1), 2015.

[15] David G. Lowe. Distinctive image feature from scale-invariant keypoints. International Journal of Computer vision, 60(2):91–110, 2004.

[16] A. Meline, J. Triboulet, and B. Jouvencel. A camcorder for 3d underwater reconstruction of archeological objects. OCEANS 2010, 20 - 23 September 2010.

[17] A. Meline, J. Triboulet, and B. Jouvencel. Comparative study of two 3d reconstruction methods for underwater archaeology. Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on, pages 740 – 745, 2012. ISBN: 2153-0858.

[18] Mark S. Nixon and Alberto S. Aguado. Feature extraction and image processing. Academic Press, 2008.

[19] J.R. Quinlan. Induction of decision trees. Maching Learning, pages 81–106, 1986.

[20] N. Snavely, Seitz S. M., and R. Szeliski. Photo tourism: exploring photo collections in 3d. ACM Trans. Graph, 25:835–846, July 2006.

[21] Tuytelaars T. and K. Mikolajczyk. Local invariant feature detectors: A survey. foundat. Trends Comput. Graphics Vision, pages 177–280, 2008.

[22] Andreas Rechtsteiner Luis M. Rocha Wall, Michael E. Singular value decomposition and principal component analysis. Springer, pages 91–109, 2003.

[23] Zhengyou Zhang. A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11), November 2000.