Trinocular System for 3D Motion and Dense Structure Estimation

R. Campos, J. Ferrer, M. Villanueva, T. Nicosevici, L. Magí and R. Garcia
Computer Vision and Robotics Group
University of Girona, Spain
Email: {rcampos, jferrerp, miki, tudor, llmagi, rafa}@eia.udg.edu

Abstract - *The relief of the seafloor is an important source of data for many scientists. In this paper we present an optical system to deal with underwater 3D reconstruction of complex scenes. This system is formed by three cameras that take images synchronously at a constant frame rate. We use the images taken by these cameras to build dense 3D reconstructions that will provide detailed information about the structure of the observed scene. We use a stereo tracking system to estimate the motion of the trinocular rig through the recorded sequence and we later apply a Bundle Adjustment refinement to the computed trajectory to minimize the accumulated drift. Using the obtained trajectory, we generate a dense map of the observed scene by registering the different dense local reconstructions in a unique and composite 3D map representing the entire surveyed area.*

Keywords - *3D motion, 3D reconstruction, trinocular system.*

I. INTRODUCTION

The seafloor is one of the most unexplored areas of the world. Underwater images are a very rich source of data for scientists who study biological and geological processes in these areas. Among other uses, underwater imagery can be used to construct composite images, referred to as photo-mosaics.

Photo-mosaics [4] are widely used in many different study areas such as geological surveys, mapping, detection of temporal changes in benthic communities, etc. Up to date, mosaicing has been extensively carried out in 2D, providing a high resolution map of the surveyed area. However, most of the regions of interest for the scientific community are located in areas with 3-dimensional relief.

Traditionally, acoustic systems such as multibeam sonars are used to estimate the 3D relief of the seafloor. Although they provide satisfactory results, they are intended for a general modeling of the seafloor relief, but their resolution (in the order of meters) is far from that obtained using video cameras (up to the order of millimetres).

In this paper we present an optical system aimed to model complex 3D structures of the ocean floor. Compared to the multibeam sonar, this system is intended for obtaining a more detailed 3D reconstruction of the surveyed area. It will be the basis for automatic detection of changes in the morphology of many underwater objects in the future, allowing the understating of geologic, tectonic and sedimentologic processes in the seafloor.

II. DESCRIPTION OF THE IMAGE ACQUISITION SYSTEM

The light is attenuated by the medium where it is going trough. Our system works in an underwater environment, so we have to bear in mind that when light propagates through a fluid, its intensity decreases following an exponential decay. This fact is caused by the phenomena of absorption and scattering. Given these poor lightning conditions, the cameras in our image capturing system must have high light sensitivity. For this reason, we have selected three Rolera XR cameras for our system. Their pixel size ($13.7 \times 13.7 \mu m$) provides a very good light sensitivity. Additionally, we decided to use a Schneider's F1.2 optics to minimize light absorption in the lenses.

The image acquisition system, hereafter called *trinocular* system, is basically composed by three cameras, a processing unit (a standard PC) to control the image acquisition process and inputs of the user, two hard-drives to store the images and an optional LCD screen for user visual feedback. All these elements are encapsulated inside four cylindrical steel containers: one for the processing unit and one for each camera. This trinocular system was designed to be operated in two ways: (i) manipulated by a scuba diver, who will interact with the system using a set of buttons (reed switches), or (ii) coupled to an underwater robot. Fig. 1 illustrates the trinocular system on its scuba diver configuration.

The images taken by the 3 cameras serve as input for a stereo reconstruction algorithm. For this reason, it is of the utmost importance that the cameras capture the images in a synchronized fashion. In order to achieve this goal, we use the standard Parallel Port of the processing unit to simultaneously drive three trigger signals. By means of additional circuitry, these signals are buffered and distributed to the trigger port of each camera.

III. IMAGE PROCESSING

A. Epipolar Geometry

The most important part in a 3D reconstruction algorithm based on stereo images is the detection of image correspondences. A correspondence is a pair of points, one in each im-

(a) General view (b) Front view (c) Back view

Fig. 1. Images of the trinocular system's final version.

age, which represents the projections of the same 3D point. Solving the correspondence problem allows the reconstruction of this point in 3D, by using a stereo triangulation.

In order to get the 3D position of a point, the stereo system must first be calibrated, hence determining the extrinsic and intrinsic camera parameters [3].

When a pair of cameras are looking at the same scene from different points of view, there is a relationship between the 3D points and their 2D projections. This results in a series of restrictions given by the *epipolar geometry* [5]. Basically, given a point location in one image, its matching position in the other image will be lying along the *epipolar line*, reducing the search space from two dimensions (all image) to one dimension (one line that crosses the image). In order to speed up the search process even more, we rectify the images corresponding to each camera pair [1]. Once the images have been rectified, the epipolar lines become horizontal, making the search process more efficient in terms of computation.

B. Match Propagation

Since we want to obtain a scene reconstruction with high level of detail, we need to obtain as much correspondences as we can. In order to achieve this, we use a variant of the Match Propagation (MP) algorithm [7].

This algorithm uses a stereo pair of images and a set of initial correspondences. This propagation is based on the assumption that the neighborhood around valid correspondences still shows sufficient texture to contain other reliable matches. A first approach of the algorithm appeared in [8], where Lhuillier suggested the idea of applying a region growing scheme (similar to that of image segmentation) to image matching. The algorithm was then refined and described in more detail by Lhuillier and Quan in [7].

To obtain a reliable set of seed correspondences, we use state-of-the-art algorithms for point detection, description and matching (e.g. SIFT [10], SURF [2] etc.). In order to ensure the quality of the initial seed matches, we use the known geometry of the stereo rig for outlier rejection [11]. Next, dense correspondences are detected around the seed matches by applying cross-correlation [7] in the rectified image pairs. The detection of dense correspondences is described hereafter.

We first apply a variant of the MP method that performs a deeper search for matches and uses the information from the third camera to avoid outliers. Our proposal starts with a previously defined set of reliable correspondences from the rectified left and right image pair, along with their corresponding correlation score. At each iteration, we extract the correspondence with the maximum correlation score and we open a window with a predefined size around each point in the left and right images. Then, for each pixel within the left image window, we open a second window around its correspondent position in the right image window. This second window is defined by a tolerance threshold for the epipolar line (height) and dispariti tolerance threshold (width). Given the pixels in the left image, we compute their correlation score with the ones from the second window in the right image. The pixel pairs with a correlation score higher than a given threshold are considered as valid correspondences. The remaining pixel pairs are given a *second chance*, computed their correlation score with their predicted position in the third image. We do this second check because the correspondence search process using correlation is sensitive to illumination and camera view point changes.

Given that we only know the position in the left and right image of the three cameras, we need to compute the position of the match in the third view. In order to do this, we use the properties of the epipolar geometry. Given a calibrated stereo rig, the epipolar geometry establishes that, given a point in an image, its correspondence in the second image will lie somewhere along the epipolar line. Our three-camera system can be seen as three separated stereo imaging systems (i.e. formed by camera pairs 0-1, 0-2, 1-2). As we know the location of a match in the left and right images, we obtain two epipolar lines in the upper image. By definition, both epipolar lines must contain the correspondence, so its intersection will give us the estimated position of the point in the third camera. These relations between cameras must be computed in the original images, prior to the rectification process. For this, in each image we must know the homographies containing the transformations between the rectified images corresponding to the two possible stereo pairs and the original images.

Knowing the position of the point in the third camera, we compute the correlation score between the left and the upper rectified images, and we check if it passes another (a slightly more permissive) threshold. An illustrative outline for this process is illustrated in Fig. 2.

If the correspondence passes one of the two checks, we save it in a local array, along with the correlation score of the first matching attempt if it passed the check, or the mean of both scores if it passed the second test.

Once we have processed all the pixels within the windows corresponding to the current correspondence, we repeat the process using the correspondence with the second-best correlation score. The process is finished one all the correspondences have been analyzed.

Once the MP algorithm is applied, the resulting correspondences are refined at subpixel accuracy. This is achieved by opening a window around one of the two points of each correspondence and computing the correlation score. We fit a

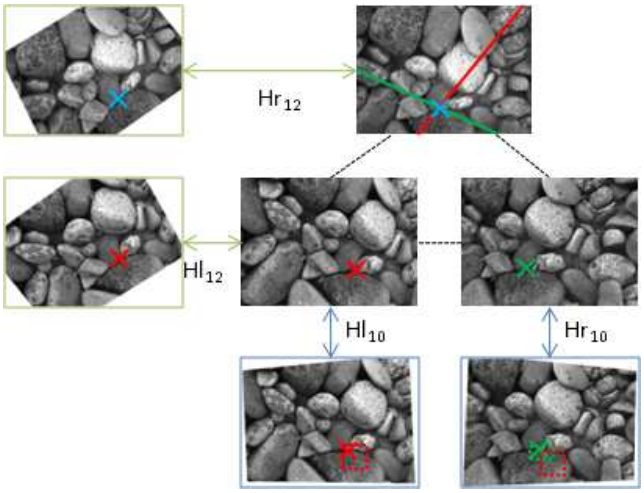


Fig. 2. Schematic of the worst-case iteration of the MP algorithm. We start with a match in the rectified stereo pair found at the bottom of the image, where we have opened the search windows. If the match does not pass the first correlation threshold, we use the rectification homographies to get the position in the original images, from where we can compute the epipolar lines and get the expected position of the match in the third image. Finally, we pass this match between the left and upper image of the rig and we perform the second check with their correlation score.

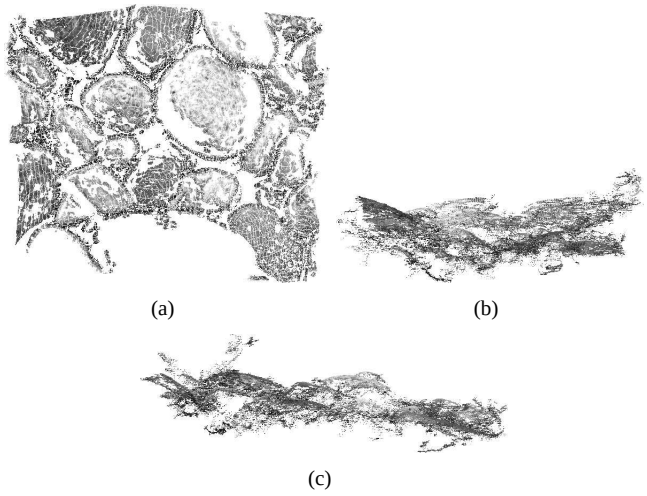


Fig. 3. Three views of a dense 3D map obtained from a triplet of images acquired by the trinocular system, where (a) is the plan, (b) the side view and (c) the main view.

quadratic-function surface to the resulting values. The position of the point is then given by the position of the peak in the quadratic function.

Finally, we obtain the 3D position corresponding to each match using triangulation. The result is a dense set of 3D points.

IV. CREATION OF A DENSE MAP

Given the dense reconstructions obtained in the previous section, we aim to obtain a composite 3D map by registering these local 3D point clouds.

In order to get this global registration, we use a stereo tracking algorithm (we only use a specific pair from the triplets) to recover the pose of the camera system corresponding to each time instant in the sequence. The algorithm detects robust correspondences between image pairs in the stereo system, and tracks them through time. Then, we use Horns' Absolute Orientation algorithm [6] to get the motion between pairs of images in the sequence.

However, the resulting trajectory suffers from the accumulated error. In order to minimize this error, we use a Sparse Bundle Adjustment method (based on [9]). The used cost function minimizes the reprojection error of the 3D points (i.e. the difference between the observed position of the point in an image and the reprojection of the corresponding 3D point in that image). It is shown in [12] that this mesurement is the Maximum Likelihood Estimator (MLE) for the BA problem.

Provided the trajectory of the trinocular system we register the dense sets of local points, obtaining the composite 3D map.

V. RESULTS

Fig. 3 illustrates the results obtained in a non-underwater environment from a triplet of images acquired with the trinocular system. The number of matches obtained for this specific case using the MP algorithm is 117768 from a total of 361920 pixels (image size is 696×520 pixels), corresponding to $32,54\%$ of the image plane. As it can be seen, the amount of matched (and reconstructed) points is sufficiently high to get fine structure details of the observed scene. Also, the number of outliers is sufficiently small in comparison with the number of inliers.

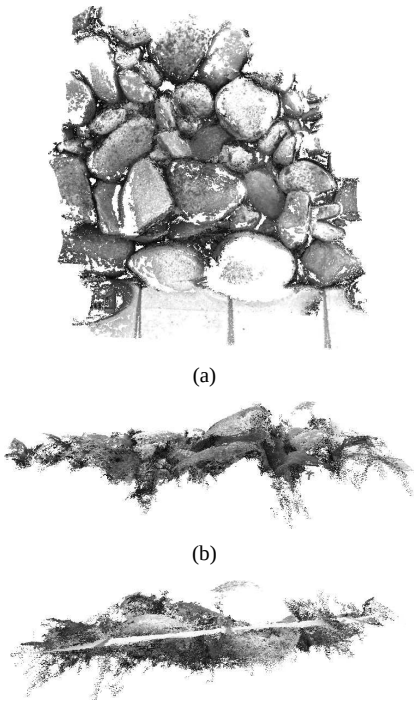
Besides, Fig. 4 shows the dense map obtained using 100 local reconstructions from 100 triplets. As we can see, the reconstruction detail increases, as multiple local dense maps of the same area are merged. In this case, areas that cannot be reconstructed from one view can be recovered from others. In addition, the covered zone is much bigger than the one that covered by a single triplet. The total number of reconstructed points for this map is 6042941.

VI. CONCLUSIONS

A new system for 3D modeling of the underwater sea-floor composed by both hardware and software proposals has been presented.

In the Hardware part, we have developed a stereo image acquisition system formed by three cameras that can deal with poor underwater light conditions. The system captures images in a synchronized fashion, at a constant frame rate.

On the other hand, the software consists in a processing scheme for obtaining accurate dense 3D maps from the cap-



(a)

(b)

(c)

Fig. 4. Three views of a dense 3D map obtained from a set of local dense 3D reconstructions, where (a) is the plan, (b) the side view and (c) the main view.

tured set of images: (i) it generates dense reconstructions of triplets of images by propagating the matches between images, and (ii) it registers the local reconstructions using camera trajectory estimations.

Unfortunately, we have not been able to acquire real underwater data to test the proposed methods. However, the results obtained in an outdoor environment are promising.

As future work, the resulting point cloud could be post processed in order to obtain a continuous surface through meshing. This would allow us to generate more realistic 3D models. Additionally, we study a series of techniques to remove the outliers in the 3D model that are not eliminated by traditional methods.

ACKNOWLEDGEMENTS

This work was supported in part by the Spanish Ministry of Education and Science grant CTM2007-64751 and in part by the European Union under Marie Curie RTN FREESUBNet project.

REFERENCES

[1] A. Fusiello, E. Trucco, and A. Verri, "A compact algorithm for rectification of stereo pairs," *Machine Vision and Applications*, vol. 12, no. 1, pp. 16–22, 2000.

[2] H. Bay, T. Tuytelaars, and L. J. V. Gool, "Surf: Speeded up robust features," in *Proceedings of the ECCV'06: European Conference on Computer Vision*, Graz, Austria, May 2006, pp. 404–417.

[3] J. Y. Bouguet, "Camera calibration toolbox for matlab," 2008. [Online]. Available: http://www.vision.caltech.edu/bouguetj/calib_doc/.

[4] J. Ferrer, N. Gracias, O. Delaunoy, and R. Garcia, "Creating Large and Accurate Mosaics of the Mid–Atlantic Ridge," in *Martech–Second International Workshop on Marine Technology*, vol. 6, Vilanova i la Geltrú, Spain, Nov. 15-16, 2007, pp. 99–100.

[5] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, ISBN: 0521540518, 2004.

[6] B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," *Journal of the Optical Society of America A*, vol. 4, no. 4, pp. 629–642, 1987.

[7] M. Lhuillier and L. Quan, "Match propagation for image-based modeling and rendering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, pp. 1140–1146, 2002.

[8] M. Lhuillier, "Efficient dense matching for textured scenes using region growing," in *the Ninth British Machine Vision Conference*, 1998, pp. 700–709.

[9] M. Lourakis and A. Argyros, "The design and implementation of a generic sparse bundle adjustment software package based on the Levenberg–Marquardt algorithm," 2004. [Online]. Available: citeseer.ist.psu.edu/lourakis04design.html

[10] D. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[11] P. D. Sampson, "Fitting conic sections to very scattered data: An iterative refinement of the Bookstein algorithm," *Computer Vision, Graphics and Image Processing*, vol. 18, pp. 97–108, 1982.

[12] B. Triggs, P. F. Mclauchlan, R. I. Hartley, and A. W. Fitzgibbon, *Bundle Adjustment – A Modern Synthesis*, January 2000, vol. 1883.