STEREOSCOPIC SYSTEM FOR 3-D SEABED MOSAIC RECONSTRUCTION

A. Leone, G. Diraco, C. Distante

IMM-CNR, Via Prov.le per Arnesano, Lecce (Italy)
{alessandro.leone, giovanni.diraco, cosimo.distante}@le.imm.cnr.it

ABSTRACT

This paper presents an inexpensive framework for 3-D seabed mosaic reconstruction, based on an asynchronous stereo vision system when simplifying motion assumptions are used. In order to achieve a metric reconstruction some knowledge about the scene is recovered by a simple and reliable calibration step. The major issue in calibration come from the asynchronism that complicate the proper frames selection. To overcome this problem a stereo frames selection based on Epipolar Gap Evaluation (EGE) is proposed. Stereo disparity maps are evaluated by using both local and global approaches. To deal with brightness constancy model violation, zero-mean normalized cross-correlation is used as similarity measure in local approach, whereas a histogram equalization is necessary in global approach in order to improve min-cut based algorithms. Experimental results validate the proposed framework, allowing to define 3-D mosaics having visual quality similar to those obtained by using specialized hardware.

Index Terms— Stereo vision, 3-D mosaicking, 3-D reconstruction, asynchronous stereo, stereo matching.

1. INTRODUCTION

In underwater environment the optic sensing techniques include shape from stereopsis [4], shape from photometric stereo [1], shape from motion [2] and active stereo [3]. Shape from stereopsis require that two stereo frames are acquired at the same time, so that triggered frame grabber or other techniques that guarantee synchronism in stereo sequence are required. However, synchronized stereo sequences use expensive apparatus and present same technical limitations, such as low frame rate and poor images resolution. In this work, we explore the use of a more inexpensive asynchronous stereo sequence. In order to achieve a metric reconstruction a calibration phase is necessary, and the asynchronism in stereo sequence represents an issue for proper calibration frames selection. To overcome this problem we propose a new stereo frames selection scheme based on the Epipolar Gap Evaluation (EGE).

Depth and ego-motion estimation lead to the ill-posed stereo correspondence problem. Although for ego-motion estimation a sparse set of correspondence points is sufficient, depth estimation requires dense correspondences. In underwater environment normalized cross-correlation is generally employed for its robustness to brightness gain and offset than others measures [7]. In dense stereo matching, the best performing algorithms use either the Belief Propagation [8, 9] or Graph Cuts [10, 11]. However, these algorithms are tested on standard data sets under restricted conditions and/or controlled environments, such as a small disparity range, and image sequences that satisfy the brightness constancy assumption. To deal with the stereo correspondence problem, local and global matching methods are evaluated, with adoption of a suitable similarity measure (ZNCC) and image enhancement technique (CLAHE).

The 3-D mosaic reconstruction requires two-frames registration and then ego-motion estimation. Normally, motion and structure are jointly estimated with recursive approaches: the extended Kalman Filter represents the most popular approach, although the computation cost grows in size with the amount of features, deterring real-time implementations [2]. Iterative closest point (ICP) algorithm is used for 3-D registration [14]. Zhang [4x] proposes the epiflow framework based on the integration of motion and stereo epipolar geometries for ego-motion tracking. Our registration approach intends to conjugate both Epiflow and ICP, under a simplified motion model.

Document is organized as follow. In section 2 the main aspects of system are presented: shape acquisition, depth estimation and 3-D mosaicking. In subsequent sections each above step is outlined in detail.

2. OVERALL SYSTEM ARCHITECTURE

The main building blocks of 3-D seabed mosaic reconstruction system are (a) shape acquisition, (b) depth estimation and (c) 3-D mosaicking and surface reconstruction.

In Fig. 1 the overall system architecture is depicted. The block 1 deals with shape acquisition, including EGE asynchronous stereo sequence calibration, lens distortion correction and histogram equalization. Depth estimation is accomplished in blocks 2 and 3, where the former regards the disparity map estimation and the latter computes the subsequent depth map.

As to be seen, both local and global method are provided in stereo disparity calculation. The block numbered as 4 is used in registration of 3-D reconstructed frames, estimating system ego-motion, pre-alignment and ICP registration refinement. Finally, triangular Delaunay interpolation and rendering are placed in block 5.

3. EPIPOLAR GAP EVALUATION

The epipolar rectification simplifies the stereo disparity estimation, so that it is widely used as preliminary step before stereo matching [8]. Dealing with an asynchronous stereo sequence, we propose a new approach for stereo frames selection based on Epipolar Gap Evaluation (EGE).

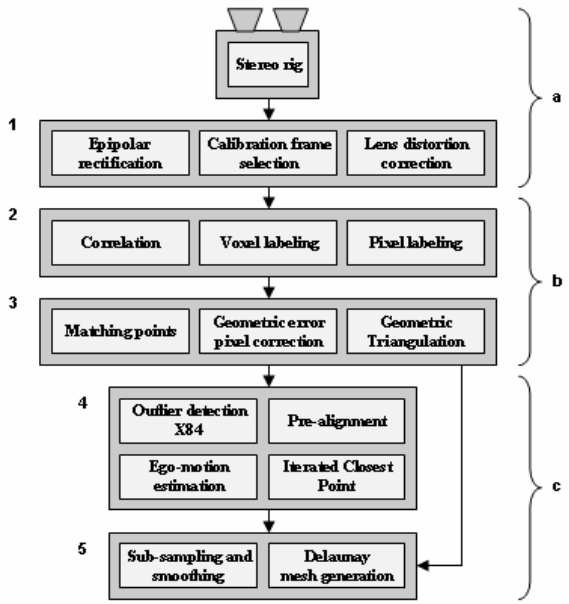


Fig. 1. Overall system architecture. Each functional block is organized in modules. *a* section groups shape acquisition functionalities, *b* section deals with depth estimation and *c* section with 3-D mosaicking and rendering.

Unlike others techniques based on view synthesis [5] or time estimation between views [6], our approach makes use of only calibration parameters encoded in camera matrices. The stereo sequence timing is represented in Fig. 2, where effective acquisition time is placed beside ideal acquisition time. Overall calibration error is given by:

$$\varepsilon = \varepsilon_{syn} + \varepsilon_{cal} + \varepsilon_{match} \tag{1}$$

where ε_{syn} is the synchronization error, ε_{cal} is the calibration algorithm error and ε_{match} is the error made in calibration pattern points selection. The calibration error minimization takes place by evaluating the epipolar gap for each stereo pairs in which calibration pattern is shot.

Preliminarily, it is necessary to estimate a set of marker points placed on calibration pattern. To do this a two step algorithm is used, in which at first step the calibration pattern is isolated using well-known morphological operations (4-connected pixel labeling for small object removing and interclass variance

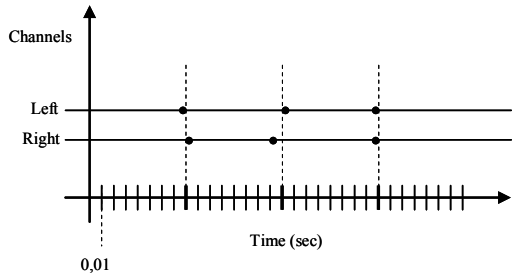


Fig. 2. Stereo sequence acquisition timing. Frame rate is 25 fps (0,04 *sec*). Dashed lines denote ideal acquisition time instants, while dots denote effective acquisition time instants.

minimization with Otsu's method). The second algorithmic step consists in the detection of parallel lines passing through centroids of the square pattern boxes. For lines detection Hough's method is used with selecting the only eight lines passing through four centroids and parallel to each other under a given degree of tolerance (Fig. 3.a). Subsequently, the parallel lines knowledge is used to divide the border pixels of each pattern square into four distinct sets depending on their position, i.e. North, East, South and West with respect to each centroid position. This four border points sets are used with a robust line fitting to estimate lines passing through rectangles edges and intersecting in rectangles vertexes (Fig. 3.b). Then, the position of rectangles vertexes in calibration pattern is used in calibration activity to estimate camera parameters.

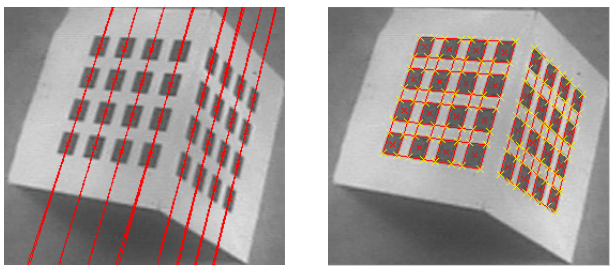
When calibration information are known as described above, for each h^{th} corresponding stereo images $\left(I_h^L, I_h^R\right)$ that shot the calibration pattern, calibration matrices $\left(\mathrm{P}_h^{\mathrm{L}}, \mathrm{P}_h^{\mathrm{R}}\right)$ and the related fundamental matrix $\mathbf{F}_h$ must be estimated in order to define the stereo geometry. At the other side, for each pairs of stereo images we want to reconstruct, the sets $\mathbf{M}^{\mathrm{L}}{}_{\mathrm{k}} \times \mathbf{M}^{\mathrm{R}}{}_{\mathrm{k}}$ of n_k feature-based matching [7] points ($m_{k,j}^L \in \mathrm{M}_{\mathrm{k}}^{\mathrm{L}}$ and $m_{k,j}^R \in \mathrm{M}_{\mathrm{k}}^{\mathrm{R}}$) are defined according to the following relation:

$$\begin{aligned} &\mathrm{M}_k^{\mathrm{L}} \times \mathrm{M}_k^{\mathrm{R}} = \left\{m_{k,1}^{\mathrm{L}}, \ldots, m_{k,n_k}^{\mathrm{L}}\right\} \times \left\{m_{k,1}^{\mathrm{R}}, \ldots, m_{k,n_k}^{\mathrm{R}}\right\} \\ &n_k = \left|\mathrm{M}_k^{\mathrm{L}}\right| = \left|\mathrm{M}_k^{\mathrm{R}}\right|, \; k = \{1, \ldots, r\}, \; r = 317 \text{ (frames)} \end{aligned} \tag{2}$$

Moreover, for each left image point $m_{k,j}^L \in \mathrm{M}_{\mathrm{k}}^{\mathrm{L}}$, in right image the corresponding epipolar line is $l_{k,j,h}^R = \mathrm{F}_h \cdot m_{k,j}^L$. In order to satisfy epipolar constraint the conjugated point $m_{k,j}^R \in \mathrm{M}_{\mathrm{k}}^{\mathrm{R}}$ must belong to this epipolar line. In this case, the h^{th} pair of calibration matrices $\left(\mathrm{P}_h^{\mathrm{L}}, \mathrm{P}_h^{\mathrm{R}}\right)$ is chosen according to the epipolar constraint, by evaluating the distance of point $m_{k,j}^R$ from epipolar line $l_{k,j,h}^R$:

$$\begin{aligned} &\mathrm{d}_{k,j}^h = \mathrm{dist}\left(m_k^R, l_{k,j,h}^R\right) \\ &\forall k = 1, \ldots, r, \; \forall j = 1, \ldots, n_k, \; \forall h = 1, \ldots, s \end{aligned} \tag{3}$$

Hence, fixed the calibration matrices $\left(\mathrm{P}_h^{\mathrm{L}}, \mathrm{P}_h^{\mathrm{R}}\right)$, for each set of correspondences $\mathbf{M}^{\mathrm{L}}{}_{\mathrm{k}} \times \mathbf{M}^{\mathrm{R}}{}_{\mathrm{k}}$, the *Maximum Epipolar Gap* (MEG) is defined as follow:



(a) (b)

Fig. 3. (a) Eight parallel lines passing through four collinear centroids. (b) Line fitting and vertex intersection.

$$\delta_{k,h} = \max_{j=1,\dots,n_k} \left\{ d^{h}_{k,j} \right\} \tag{4}$$

MEG is used to filter stereo couples, dropping those for which the corresponding MEG exceed a given threshold Θ. The amount of rejected couples will be:

$$N_h = \left| \left\{ k = 1,\dots, r \mid \delta_{k,h} > \theta \right\} \right| \tag{5}$$

Finally, the chosen calibration couples will be:

$$\left(\mathrm{P}^{\mathrm{L}}_{u}, \mathrm{P}^{\mathrm{R}}_{u} \right) \text{ with } u = \arg \min_{h=1,\dots,s} \left\{ N_h \right\} \tag{6}$$

4. STEREO DISPARITY

For disparity maps calculation both local and global methods are explored. The difficulties came from the stereo disparity algorithms and theirs underlying assumption of the brightness constancy, systematically violated in underwater environment [12, 13]. As local method the normalized zero-mean cross-correlation (ZNCC) approach are employed, defined as follow:

$$\mathrm{ZNCC}(x,y,d) = \frac{\sum_{w_x,w_y \in W} \left(\mathrm{I_L}(x+w_x, y+w_y) - \bar{\mathrm{I}}_{\mathrm{L}x,y} \right) \cdot \left(\mathrm{I_R}(x+w_x+d, y+w_y) - \bar{\mathrm{I}}_{\mathrm{R}x,y} \right)}{\sqrt{\left(\sum_{w_x,w_y \in W} \left(\mathrm{I_L}(x+w_x, y+w_y) - \bar{\mathrm{I}}_{\mathrm{L}x,y} \right)^2 \right) \cdot \left(\sum_{w_x,w_y \in W} \left(\mathrm{I_R}(x+w_x+d, y+w_y) - \bar{\mathrm{I}}_{\mathrm{R}x,y} \right)^2 \right)}} \tag{7}$$

and then

$$\mathrm{Disparity(x, y)} = \max_{d_{\mathrm{MIN}} \le d \le d_{\mathrm{MAX}}} \mathrm{ZNCC}(x,y,d) \tag{8}$$

The adopted global methods are both voxel and pixel coloring [11]. In this case a preliminary image enhancement based on the Contrast Limited Adaptive Histogram Equalization (CLAHE) is used in order to mitigate the effects of brightness constancy model violation on global matching algorithm (block size of 20 by 20 is used). In this work, the adopted voxel coloring uniformity parameter is defined as follow:

$$\lambda_{v,v'} = \begin{cases} 3\lambda & \text{if } \Delta\mathrm{I}(\mathrm{p},\mathrm{p}') < 5 \\ \lambda & \text{otherwise} \end{cases} \tag{9}$$

where

$$\Delta\mathrm{I}(\mathrm{p},\mathrm{p}') = |\mathrm{I(p)-I(p')}| \tag{10}$$

and I(·) is a function that gives intensity of its argument pixel. Results are reported in Fig. 4. Stereo disparity quality is evaluated through ground truth. CLAHE enhancement with ZNCC correlation measure don't give more improvements since cross-correlation is quite insensitive to brightness variations (Figures 4.c and 4.d). Rather, voxel and pixel coloring are heavily affected by brightness variations, so that CLAHE yields considerable differences (Figures 4.e and 4.f).

5. PRE-ALIGNMENT AND REGISTRATION

3-D mosaicking phase starts from two-frames reconstruction, passing through registration and disparity fusion, and finishes with surface reconstruction and rendering. The disparity fusion can be addressed by the iterative closest point (ICP) algorithm for 3-D registration [14], requiring the computation of 3-D transformation between two disparity maps.

Alternatively, Zhang [4] proposes the epiflow framework based on the integration of motion and stereo epipolar geometries to track stereo correspondences in passive navigation. Epiflow algorithm exploits both geometric and photometric constraints, while only the former is employed by the ICP algorithm. Our registration approach allows to conjugate both Epiflow and ICP, under a simplified motion model. A pre-alignment [15] phase is adopted for registration simplification purpose, based on motion assumptions such as pitch and roll absence. Two-frames reconstructions are pre-aligned making use of a-priori knowledge about system motion. We suggest an ego-motion estimation through four-frames (i.e. two consequently stereo couples) feature tracking, evaluating corresponding images points displacements and scaling according to the effective retina plane pixel dimension. Given two frame couples (I^L_1, I^R_1) and (I^L_2, I^R_2), correspondences between consecutive frame I^L_1, I^L_2 are estimated with robust feature-based matching [7]. Instead, correspondences between stereo coupled frame I^L_N, I^R_N are estimated with previously calculated stereo disparity maps. Once determined this four correspondent feature-points sets, ego-motion tracking is

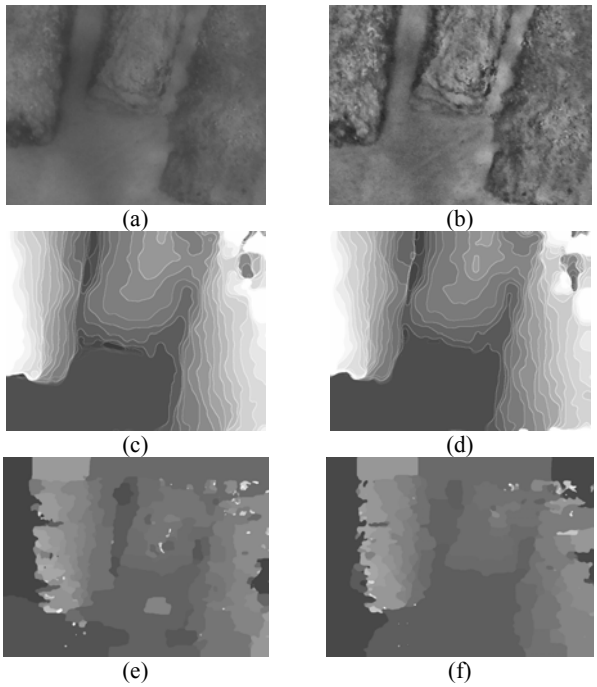


Fig. 4. (a) original rectified image, (b) CLAHE enhanced rectified image, (c) ZNCC based disparity map obtained from original image (a), (d) ZNCC based disparity map obtained from enhanced image (b), (e) voxel based disparity map obtained from original image (a), (f) voxel based disparity map obtained from enhanced image (b).

performed by mean its centroids displacement through consecutive frames. Subsequently, ICP is used for registration refinement. Fig. 5 shows the reconstructed trajectory following by the navigation equipment during acquisition and the finally 3-D mosaic of seabed surface (two subsided column are shown).

6. CONCLUSIONS

In this paper, we presented a whole framework for seabed 3-D mosaic reconstruction. The three mainly troublesome aspects considered are asynchronous stereo acquisition, depth estimation and 3-D mosaicking. Firstly, we explored the use of an inexpensive asynchronous stereo sequence, proposing a new scheme for accurate calibration frames selection. Secondly, disparity map calculation in both point of view, local and global, are considered. Brightness constancy violation is treated adopting cross-correlation and histogram equalization, respectively. Results are evaluated using a ground truth. Thirdly, we proposed a four-frames features tracking scheme for ego-motion estimation, that combines epiflow advantages with ICP registration refinements. Current ongoing and future work involves improvements on ICP module, more accurate ground truth results evaluation and the implementation an of Extendend Kalman Filter based ego-motion and structure recovery for off-line 3-D mosaicking.

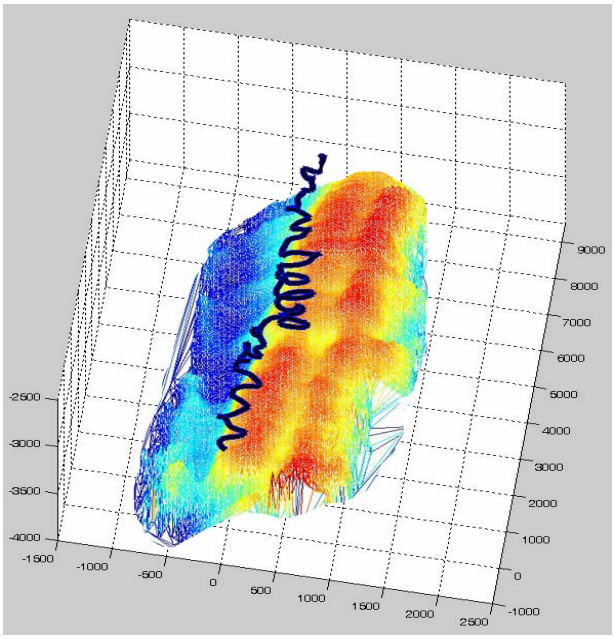


Fig. 5. Final 3-D mosaic and system trajectory.

REFERENCES

[1] S. Negahdaripour, H. Zhang, and X. Han, "Investigation of Photometric Stereo Method for 3-D Shape Recovery from Underwater Imagery," *Oceans MTS/IEEE*, vol. 2, pp. 1010-1017, Oct. 2002.

[2] A. Khamene, H. Madjidi, and S. Negahdaripour, "3-D Mapping of Sea Floor Scenes by Stereo Imaging," *Oceans MTS/IEEE*, vol. 4, pp. 2576-2583, Nov. 2001.

[3] S.G. Narasimhan, S.K. Nayar, B.Sun, and S.J. Koppal, "Structured Light in Scattering Media," *Proc. of the Tenth IEEE International Conference on Computer Vision*, vol. 1, pp. 420-427, Oct. 2005.

[4] H. Zhang, "Automatic sensor platform positioning and 3-d target modeling from underwater stereo sequences," PhD Thesis, Coral Gables, Florida, Dec 2005.

[5] Marcus Svedman, Luis Goncalves, Niklas Karlsson, Mario Munich, and Paolo Pirjanian, "Structure from Stereo Vision using Unsynchronized Cameras for Simultaneous Localization and Mapping," *IEEE International Conference on Intelligent Robots and Syste*ms, *Intelligent Robots and Systems,* pp. 3069-3074, Aug. 2005.

[6] I. Reid, and A. Zisserman, "Goal-directed video metrology," *Proc. of 4th European Conf. on Computer Vision*, In R. Cipolla and B. Buxton, editors, Cambridge, vol. 2, pp. 647-658, 1996.

[7] A. Leone, C. Distante, A. Mastrolia, and G. Indiveri, "A fully automated approach for underwater mosaicking," *Oceans MTS/IEEE*, Boston, Sep 2006.

[8] D. Scharstein, and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Stereo and Multi-Baseline Vision IEEE Workshop*, pp. 131-140, Dec 2002.

[9] J. Sun, H. Shum, and N. Zheng, "Stereo matching using belief propagation," *Proc. of European Conference on Computer Vision*, pp. 510-524, Copenhagen, May 2002.

[10] S. Birchfield, and C. Tomasi, "Multiway cut for stereo and motion with slanted surfaces," *Proc. of International Conference on Computer Vision*, vol. 22, pp. 489-495, Corfu, Greece, Sep 1999.

[11] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, pp. 1222-1239, 2001.

[12] R. Bolles, H. Baker, and M. Hannah, "The JISCT stereo evaluation," *Proc. DARPA Image Understanding Workshop*, pp. 263-274, 1993.

[13] I. Cox, S. Roy, and S. Hingorani, "Dynamic histogram warping of image pairs for constant image brightness," *Proc. International Conference on Image Processing*, vol. 2, pp. 2366-2369, October 1995.

[14] P. Besl, and N. McKay, "A method for registration of 3D shapes," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, pp. 239-256, 1992.

[15] V. Murrino, L. Ronchetti, U. Castellani, and A. Fusiello, "Reconstruction of Complex Environment by Robust Pre-aligned ICP," *3-D Digital Imaging and Modeling*, *Third International Conference on*, pp. 187-194, May 2001.