On Reconstruction of 3-D Volumetric Models of Reefs and Benthic Structures from Image Sequences of a Stereo Rig

Hongsheng Zhang and Shahriar Negahdaripour
Underwater Vision & Imaging Lab
Electrical and Computer Engineering Department
University of Miami
Coral Gables, FL 33124-0640

***Abstract*– Acquiring 3-D computer models of objects from images has been an active computer vision research topic. A common approach uses overlapping images from different views around the object, and attempts to fuse all the information to construct the model by volume intersection. Most current methods have in common that the camera poses for these views are known. To achieve this, a monocular camera circles the target object on a robot arm, or equivalently, views a target on a rotating platform (e.g., a turntable), while the images are acquired at known intervals. We are interested in the deployment of a submersible platform equipped with *stereo vision* to construct 3-D volumetric models of benthic habitat objects, e.g., coral reefs and other seafloor structures. In this application, we would automatically acquire continuous 360-degrees stereo views of object(s) of interest, by processing the images online to estimate and control the camera trajectory. The emphasis of this paper is to address the estimation of camera trajectory and views. In particular, we show that the redundancy from binocular cues improves the robustness and accuracy. Upon completion of the image data acquisition, a bundle adjustment formulation is proposed to recompute the trajectory and the stereo rig poses, as well as to generate object model(s).**

I. INTRODUCTION

Remotely Operated Vehicles (ROVs) are the most common tools for ocean exploration and discovery missions, and can become very effective and efficient tools, replacing tedious state-of-art diver-based operations, for regular ecological surveys, e.g., monitoring and inspection of deep coral reefs sites and (or) ground truthing benthic habitat surveys with high-resolution maps. Video survey has already become an important component of ROV operations in recent years, and the construction of photo-mosaics from video or snap shots, as two-dimensional (2-D) visual maps or composite images covering a large field of view, is now the most widely recognized application (e.g., [7, 8, 15, 17, 18, 28]. However, many ecologically significant measurements– e.g., surface area, rugosity, relief, and colony morphology– are directly dependent on quantitative 3-D structural measurements, and cannot be derived from 2-D mosaics. Moreover, such 3-D measurements need to be made at multiple spatial scales. For example, the appropriate length scale for measurements of spatial heterogeneity depends on the size of the organisms of interest. Small fishes will respond to spatial complexity on smaller scales than large fishes will. Methods to measure these parameters have one or more limitations: 1) Carried out manually through tedious procedures; 2) Limited to linear measurements, e.g., height or diameter, along transects; 3) Restricted by fundamental depth constraints on duration of diving operations. [4, 26, 27]. For example, accurate measurements of reef surface area from underwater photographs has been reported [4], using *manual* photogrammetry methods [1, 14].

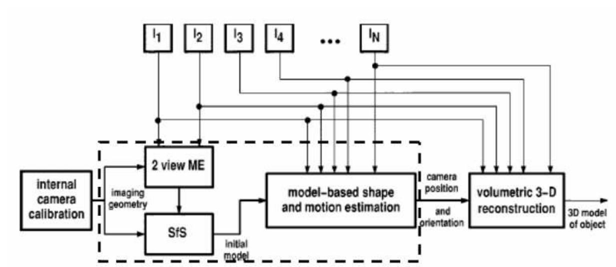

Figure 1: Block diagram taken from [5], to show the processes in the reconstruction of camera trajectory and 3-D volumetric model. Our SFS method described here for motion and structure estimation, based on the quadranocular constraints, will replace the modules within the dashed box.

This paper deals with developing the capability to automatically 1) acquire 360-degree views of corals and other objects of interest in situ with an ROV controlled directly with video, and 2) reconstruct ROV trajectory and the stereo rig poses with high precision. These computations directly affect the accuracy in the construction of the 3-D volumetric model(s) of the object(s) of interest. Most earlier techniques for construction of a 3-D model depend on the knowledge of camera geometry relative to the target, when the images are recorded [2, 19, 11, 13, 5, 3, 22, 16, 23, 6]. Under these conditions, accurate reconstruction has been demonstrated utilizing solely a single camera.

While techniques to estimate both the camera trajectory and object model based on monocular views have been proposed, reported results either use data with known camera trajectory or simple camera motions. In our application, we deal with a platform that constantly experiences motions with 6 degrees of freedom. We seek methods to recover the camera position and pose automatically from the video data. Binocular cues 1) provide redundancy to improve robustness and accuracy, 2) simplify some of the underlying 3-D shape/structure and motion estimation problems, and 3) are important from an operational perspective to automate and enhance ROV positioning for visual servo during image acquisition. To put this contribution into proper perspective, fig. 1 is a block diagram taken from [5], to show the processes

0-933957-31-9

in the reconstruction of camera trajectory and 3-D volumetric model. The method described here for motion and structure estimation, based on the quadranocular constraints, can replace the modules within the dashed box and plays a central role in accurate 3D reconstruction.

This paper is organized as follows. In section II, we introduce our notation and review some background material. In section III, a novel "quadranocular constraint" is introduced, establishing the framework for 3-D reconstruction from stereo cues. In section IV, we cover motion-stereo methods for recovery of 3-D trajectory and scene reconstruction, including a new quadranocular constrained-based solution and its use for online navigation, as well as a stereo bundle adjustment algorithm based on binocular cues. In section V, experiment results illustrate the improved robustness and accuracy. Section VI summarizes the main contributions of this paper.

II. BACKGROUND

A. Notation

The stereo rig is illustrated in Figure 2. Each camera is described by the pinhole camera model. A 3-D point is denoted $M = [x, y, z]^T$ and its image $m = [u, v]^T$. The homogeneous coordinates of m and M are $\tilde{m} = [u, v, 1]^T$ and $\tilde{M} = [X, Y, Z, 1]^T$ respectively. At instant t_0, the two stereo cameras are modelled $P_{l,0}$ and $P_{r,0}$:

$$\begin{aligned} P_{l,0} &= K_l \left[\begin{array}{c|c} I_{3\times3} & 0_{3\times1} \end{array} \right] \\ P_{r,0} &= K_r \left[\begin{array}{c|c} R_s & t_s \end{array} \right] \end{aligned} \tag{1}$$

where K_l and K_r are the calibration matrix of the left and right cameras, respectively, and R_s and t_s are the rotation and translation of the right camera relative to the left camera. With the motion of the vehicle, a sequence of stereo images are taken at sample times $i, i = 1..M$. Accordingly, the camera projection matrices are $P_{l,i}$ and $P_{r,i}$:

$$\begin{aligned} P_{l,i} &= K_l \left[\begin{array}{c|c} R_i & t_i \end{array} \right] \\ P_{r,i} &= K_r \left[\begin{array}{c|c} R_s R_i & R_s t_i + t_s \end{array} \right] \end{aligned} \tag{2}$$

where R_i and t_i describe the motion of the stereo rig at instant i relative to the initial position. If we define the homogeneous transformation matrix H_i:

$$H_i = \begin{bmatrix} R_i & t_i \\ 0_{1\times3} & 1 \end{bmatrix} \tag{3}$$

we can then write

$$\begin{aligned} P_{l,i} &= P_{l,0} H \\ P_{r,i} &= P_{r,0} H \end{aligned} \tag{4}$$

Finally, $F_{l,i}$ and $F_{r,i}$ denote the fundamental matrices of the left and right camera, respectively (see section).

B. Camera Calibration

Several images of a chessboard coplanar pattern provide the internal camera parameters K_l and K_r, utilizing the camera calibration toolbox from Intel Open CV. From the set of paired images of the same calibration target at the same instant by both cameras, the external camera parameters R_s,t_s are obtained. After initial camera calibration, we assume the camera parameters to remain constant through the sequence. To simplify the computations, stereo rectification is applied to align vertical image lines and thus set

$$K_l = K_r \quad R_s = I_{3\times3} \quad \text{and} \quad t_s = [0, 1, 0]^T \tag{5}$$

Here, t_s is scaled to unit length. In motion estimation, all translation vectors are scaled accordingly across the whole sequence. However, external calibration gives the absolute translation between the two cameras in Euclidean space. Emphasis is placed on that Euclidean reconstruction, not just metric reconstruction, of the trajectory.

C. Motion Stereo Constraints

For a pair of camera models P_1 and P_2, the fundamental matrix can be extracted [9]:

$$F = [e_2]_\times P_2 P_1^+ \tag{6}$$

where $^+$ denotes pseudo-inverse, and e_2 is the epipole for the camera center of the first camera C_1; i.e. $e_2 = P_2 C_1$ and $P_1 C_1 = \mathbf{0}$.

For the left camera from (1) and (2), correspondences $\tilde{m}_{l,0_j}, \tilde{m}_{l,i_j}, j = 1..N_i$ satisfy the epipolar constraint

$$\begin{gathered} \tilde{m}_{l,i_j}^T F_{l,i} \tilde{m}_{l,0_j} = 0 \\ F_{l,i} = K_l^{-T} E_{l,i} K_l^{-1} \\ E_{l,i} = [t_i]_\times R_i \end{gathered} \tag{7}$$

Similarly for the right camera as in (1) and (2), correspondences $\tilde{m}_{r,0_j}, \tilde{m}_{r,i_j}, j = 1..N_i$ satisfy

$$\begin{gathered} \tilde{m}_{r,i_j}^T F_{r,i} \tilde{m}_{r,0_j} = 0 \\ F_{r,i} = K_l^{-T} E_{r,i} K_l^{-1} \\ E_{r,i} = \left[-R_s R_i R_s^T t_s + t_s + R_s t_i \right]_\times R_s R_i R_s^T \end{gathered} \tag{8}$$

$E_{l,i}$ directly encodes what we treat as the motion of the stereo rig; motion expressed in the coordinate system of the left camera.

III. QUADRANOCULAR CONSTRAINTS

Suppose that we have a collection of 3-D feature points $M = \{M_j, j = 1..N\}$. At each instant i, only a subset of N_i points, denoted $M_i = \{M_{i_j}, j = 1..N_i\} \subseteq M$ can be imaged due to occlusion, comprising the point correspondences between stereo images at times 0 and i. The image point correspondences are denoted $\{m_{l,0_j}, m_{l,i_j}, m_{r,0_j}, m_{r,i_j}\}, j = 1..N_i$. From the projection model, homogeneous coordinates of the image point and 3-D point are related by

$$\begin{aligned} \tilde{m}_{l,0_j} &= P_{l,0} \tilde{M}_{i_j} \\ \tilde{m}_{l,i_j} &= P_{l,i} \tilde{M}_{i_j} \\ \tilde{m}_{r,0_j} &= P_{r,0} \tilde{M}_{i_j} \\ \tilde{m}_{r,i_j} &= P_{r,i} \tilde{M}_{i_j} \end{aligned} \tag{9}$$

These constitute 4 vector constraints, which we call the *quadranocular motion-stereo constraints*, relating the observations $\{m_{l,0_j}, m_{l,i_j}, m_{r,0_j}, m_{r,i_j}\}, j = 1..N_i$ to the 3-D points M_{i_j} through the homogeneous transformation H_i.

This is a generalization of trifocal tensor geometry for point correspondences in three views, which has become the basis for many recent 3-D reconstruction techniques; e.g., [20, 21, 9].

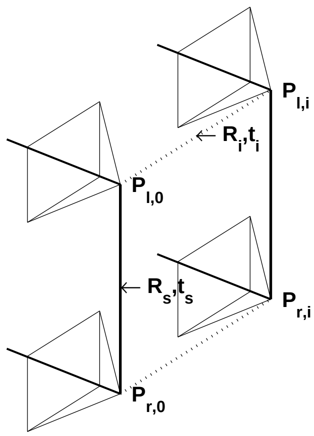

Figure 2: Horizontal and vertical disparities from the motion of the stereo rig.

IV. 3-D TRAJECTORY & SCENE RECONSTRUCTION

A. Traditional Motion-Stereo Approach

Estimating $F_{l,i}$ in (7) from the linear 8-point algorithm [12, 24], it is straightforward to compute $E_{l,i}$ from $F_{l,i}$. Then the stereo rig motion R_i and t_i is determined as described in [9]. Next, the camera matrices $P_{l,i}$ and $P_{r,i}$ are computed from (1). Through triangulation with known $P_{l,0}$ and $P_{l,i}$, 3-D points $M_{m,i_j}, j = 1..N_i$ are reconstructed. Similarly triangulation for $P_{l,0}$ and $P_{r,0}$ leads to another reconstruction of the same 3-D points $M_{s,i_j}, j = 1..N_i$. Here, subscripts m and s denote motion and stereo. However, the translation estimate t_i from the essential matrix E_i is scaled to unit length by some unknown factor, denoted $\hat{t}_i$, while we have normalized the baseline of the stereo cameras by a known factor. (These two scaling factors are not the same.) We need to determine the unknown factor s_i to scale the unit translation vector $\hat{t}_i$: $t_i = s_i \hat{t}_i$. It is noted that the factor is also the scaling that relates the reconstructions of the same set of 3-D points from motion and stereo cues:

$$s_i M_{m,i_j} = M_{s,i_j} \tag{10}$$

In a similar fashion as in [10], s_i can be estimated from a least square solution:

$$s_i = \frac{\sum_{j=1}^{N_i} M_{s,i_j}^T M_{m,i_j}}{\sum_{j=1}^{N_i} M_{s,i_j}^T M_{s,i_j}} \tag{11}$$

This solves the scaling ambiguity of the motion problem, giving us absolute camera positions along the trajectory.

B. 3-D Reconstruction from quadranocular Constraints

Six sets of 3-D reconstructions can be performed from triangulation on any two of the four projections (C_4^2) of the same 3-D points. The solution in the last section uses only 2 sets to estimate motion and structure. In order to arrive at optimal estimates consistent across the 4 projections, we propose a general computing framework based on the quadranocular constraints. Assuming that the image measurement noises are Gaussian, we seek the Maximum Likelihood solution: Estimate R_i,t_i, hence stereo projection matrices $P_{l,i}$,$P_{r,i}$, and 3-D points $M_{i_j}, j = 1..N_i$ to minimize the squared distance between the reprojected points and actual measurements for the 4 views:

$$E(R_i, t_i, M_i) = e_{l,i} + e_{r,i} + e_{l,0} + e_{r,0} \tag{12}$$

where

$$\begin{aligned} e_{l,i} &= \sum_{M \in M_i} \|P_{l,i}\tilde{M} - \tilde{m}_{l,i_j}\|_2^2 \\ e_{r,i} &= \sum_{M \in M_i} \|P_{r,0}\tilde{M} - \tilde{m}_{l,i_j}\|_2^2 \end{aligned} \tag{13}$$

Note that $\tilde{M}$ is the representation of M by homogenous coordinates. Minimization is carried out using the *Levenberg-Marquardt* algorithm [25] with the initial solution given in Section III(A). The estimates R_i, t_i ($i = 1..M$) directly define the trajectory of the stereo rig (vehicle).

To provide a comparison with performance based on three views, we can determine a solution based on *left trinocular constraints*:

$$E(R_i, t_i, M_i) = e_{l,i} + e_{l,0} + e_{r,0} \tag{14}$$

or similarly, *right trinocular constraints*:

$$E(R_i, t_i, M_i) = e_{r,i} + e_{l,0} + e_{r,0} \tag{15}$$

C. Online Navigation and Trajectory Following

Before describing the trajectory following strategies, it is noted that several important problems are not addressed in this discussion, including trajectory planning, feature tracking, and online object reconstruction and occlusion detection. The primary purpose is to demonstrate the framework for the application of the quadranocular constraints for automatic navigation by online processing of the stereo data. The 3-D volumetric modelling of the coral reef or other benthic structures of interest can be carried out most optimally once the entire data set is collected; see fig. 1 and also the next section).

After the first few sample times, an initial set of 3-D global interest points $\{M_i\}$ is established[1]. These points are tracked during the execution of the trajectory, as the ROV circles the targets of interest to acquire the stereo data. Let $S = T(0)$, $T(k)$, and $T(k+1)$ denote the start, current, and the next target positions of the ROV (see fig. 3); we treat the start point as our very first target position. Consider the path from $T(k)$ to $T(k+1)$, which may be planned based on a number of criteria[2]. For example, this may be a section of the complete trajectory that circles above the objects of interest while maintaining "the best view" of these objects[3]. Let C denote

[1] Exact detailed procedure, which is strongly tied to the application or types of objects to be reconstructed, is beyond the scope of this paper.

[2] This, among many factors, depends on the features and objects to be reconstructed.

[3] Here again, "best view" has to be defined based on the application. Examples, among many, may be to minimize the number of occluded features, or to maintain a minimum number of features from the original set within the field of view.

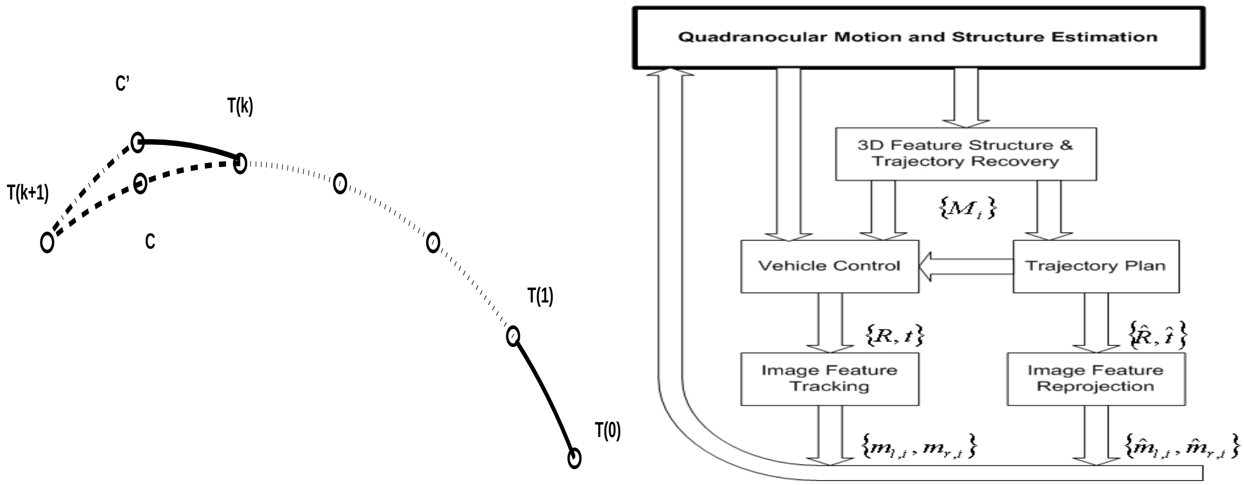

Figure 3: Online navigation– Left diagram shows various points along ROV trajectory defined in the text. The bottom shows shows the processing. A trajectory is planned based on current vehicle position relative to 3-D targets of interest (defined by 3-D points $\{M_i\}$). Projections $\{\hat{m}_{l,j}, \hat{m}_{r,j}\}$ are calculated from planned motion $\{\hat{R}, \hat{t}\}$, while $\{m_{l,j}, m_{r,j}\}$ are actual projections detected from the stereo images (after the vehicle motion $\{R, t\}$). Applying quadranocular constraints to these 4 projections, ROV position and pose, and 3-D features are updated. The loop repeats again.

the next ROV position on the path where images will be acquired. Due to external disturbances, imperfect thrusters' responses, etc., the vehicle actually tracks an approximate but different path, ending at point C'. Among various strategies for positioning and trajectory following, two preferred methods give the absolute position of the ROV relative to targets of interest (rather that indirectly by integrating incremental motions). We can estimate the position and pose at C by processing the stereo pairs at C' and $T(0)$, utilizing our quadranocular constraints to these 4 views. In this way, the ROV gets its position relative to the start position. Alternatively, we may process the data at C' and $T(k+1)$, recalculating the path to reach the target destination. Here again, we apply the quadranocular constraints. The data at $T(k+1)$ comprises the predicted image positions of the features, which are calculated from the position and pose at $T(k)$ and the most current estimated 3-D positions of the features. Accordingly, the control system generates the proper thruster signals to move the vehicle along the revised path; e.g., as shown by the dashed line. Once the ROV reaches within a position threshold of $T(k+1)$, say $\hat{T}(k+1)$, the position and pose are updates from data at $T(0)$ and $\hat{T}(k+1)$. The 3-D positions of features are also revised by performing a bundle adjustment over previous stereo pairs at $T(0), \ldots, T(k+1)$.

The overall dynamic configuration for the 2nd strategy is shown fig. 3. Here $\{R, t\}$ denote the position of the ROV at C' and $\{\hat{R}, \hat{t}\}$ represent the position at $T(k+1)$. The concept of applying the quadranocular constraints to the data at C' and $T(k+1)$ is depicted in fig. 4. The circles denote the features detected in the stereo pairs at position C' and the predicted image positions at position $T(k+1)$ are shown by crosses (arrows simply demonstrate the correspondences).

D. Reconstruction by Stereo Bundle Adjustment

From quadranocular constraints, the position and pose of the stereo rig at each instant i are obtained independently. From this, the 3-D points can be readily reconstructed by triangulation. It is often the case that the 3-D estimates from multiple views are inconsistent. Another question arises: what is the accuracy from different views and how are the information integrated into a more robust estimate? Furthermore, the 3-D object model construction needs an accurate camera trajectory estimate for optimal results. To address these issues, we propose to minimize the reprojection error over the entire sequence in order to impose the global consistency of the trajectory and the 3-D scene:

$$\min_{R_i, t_i, M_j} \sum_{i=0}^{M} \{e_{l,i} + e_{r,i}\} \tag{16}$$

where $e_{l,i}$ and $e_{r,i}$ are defined in Eq. (13).

Minimization is performed with the *Levenberg-Marquardt* algorithm [25]. The initial solution for the motions is given by the individual estimates given in section III. Since for target reconstruction, the estimates of 3-D points at various sample instants do not usually coincide, a bundle adjustment strategy is adopted [14, 9]. The median of each feature's 3-D positions is chosen as the initial guess. In the sense of M-view geometry [9], our stereo bundle adjustment should be classified as a 2M-view bundle. However, the number of optimization parameters is the same as in the M-view since the relative motion between the stereo cameras is known. Therefore, the computation cost is much less than for the ordinary 2M-view bundle adjustment, and redundancy is achieved by using stereo cues, but at less than double the computation load.

V. EXPERIMENT RESULTS

An experiment with a number of artificial coral reef objects is described. A stereo video was acquired while circling the scene, using only 14 sparse pairs at 240x320 resolution for the calculations. Selection of a total of 32 feature points and establishing their correspondences over the entire sequence were done manually, as the primary purpose of this experiment is to test the positioning accuracy of various methods for trajectory and 3-D feature estimation. For each stereo pair, roughly 11-15 of the total point correspondences were processed. Figure 4 shows a sample stereo pair, where

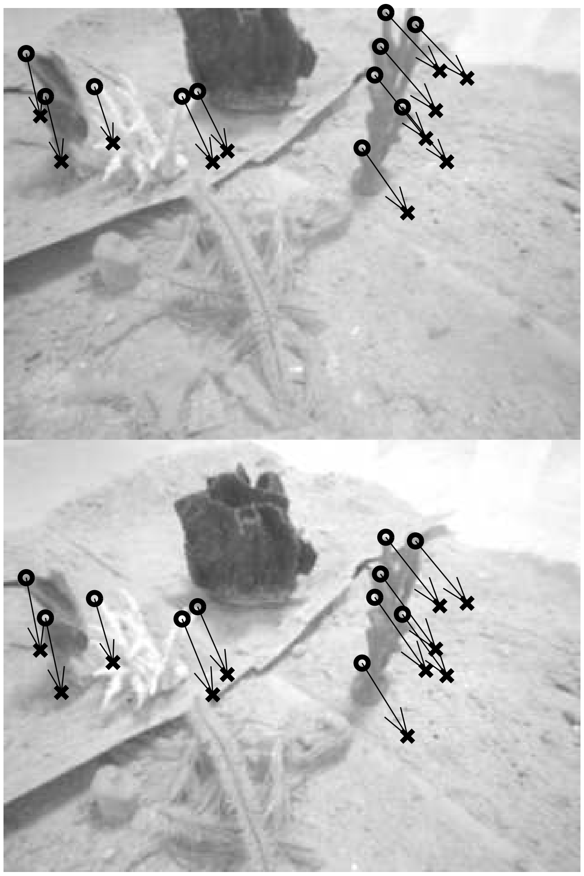

Figure 4: Circle “O” marks are the observed images $\{m_{l,j}, m_{r,j}\}$ of 3-D points $\{M_i\}$ on the actual trajectory. Points marked by X are the images $\{\hat{m}_{l,j}, \hat{m}_{r,j}\}$ of the same 3-D points on the planned trajectory. These constitute the data processed based on the quadranocular constraints to compute the control signal for navigation.

the 11 selected features visible in these views are marked by circles. The vertical stereo baseline was determined to be $|t_s| = 99.7mm$ by calibration. Since establishing the ground truth with sufficient accuracy was difficult, the camera positions and 3-D points reconstructed from the stereo bundle adjustment over the entire data set was used as the baseline for quantitative assessment. More precisely, we determine how close(good) the estimates from various methods are when processing data from only two positions. Three Different approaches for trajectory reconstruction are compared against the stereo bundle adjustment method, and the results are depicted in fig (5). The errors for various methods have been given in Table 1. Two trajectories were determined by processing the left and right sequences separately– corresponding to the videos from two independent monocular cameras, while the scale ambiguity was resolved by employing the estimate from the bundle adjustment method. These, shown in (a) and (b), have the largest errors. Trinocular constraints, applied to a stereo pair at the first view and a single view at position i is the second method [9]; Two results, given in (c) and (d) for each of $\{P_{l,0}, P_{r,0}, P_{l,i}\}$ and $\{P_{l,0}, P_{r,0}, P_{r,i}\}$, provide improved estimates over the ones from the monocular sequences. Finally, we have applied our quadranocular constraints to $\{P_{l,0}, P_{r,0}, P_{l,i}, P_{r,i}\}$ for the final estimate, shown in (e). As we expected, we obtain the best results here.

Next comparison is in terms of the computed positions of the selected feature points. As stated earlier, the estimates from various views are expected to be inconsistent. This means that, for each feature, we obtain a “cluster”, whose size is directly ties to the inaccuracy of estimates from various views. In (a’) to (e)’, we have shown these clusters (dots) with the corresponding estimate from the bundle adjustment method (circle).

While the methods based on trinocular and quadranocular constraints provide the best results, the improvement from the latter is demonstrated quantitatively in the results of table. . Interestingly, for the same method, sequences involving the right camera consistency give the worse results. This, yet to be resolved,is expected to be due to a sync problem in the stereo cameras.

	$\frac{1}{M}\sum_i^M g_i$ (mm)	$\frac{1}{M}\sum_i^M \theta_i$ (degree)
left camera	21.3	9.0
right camera	43.5	10.0
left trinocular	18.2	1.4
right trinocular	23.6	1.9
quadranocular	13.1	1.1

Table 1: Each entry is the difference in solutions of camera positions and poses by bundle adjustment and each listed method. First column is the average distance error over all the positions along the trajectory. Second column gives the average angle error between the pose directions at position.

VI. CONCLUSION

We have presented quadranocular constraints as a framework for more accurate 3-D reconstruction– structure estimation and positioning– for online ROV navigation from motion-stereo cues. A stereo bundle adjustment is the proposed method for off-line 3-D reconstruction. Quadranocular constraints involve the information in two pairs of stereo images from nearby views. The stereo bundle adjustment incorporates cues from $2M$ views in M positions, without significant increase in the computation load (the number of optimization parameters is the same as the ordinary M views bundle adjustment algorithm). Experimental results show the improvement in exploiting quadranocular constraints over alternative approaches which utilize partial information from available views. Calibrating the stereo rig enables Euclidean reconstruction of trajectory and object structure [9]. These developments are essential prerequisite to the reconstruction of 3-D volumetric models of natural objects from 360 views, which is our objective in ongoing work.

Acknowledgement: The first author is supported by the UoM Graduate Fellowship program.

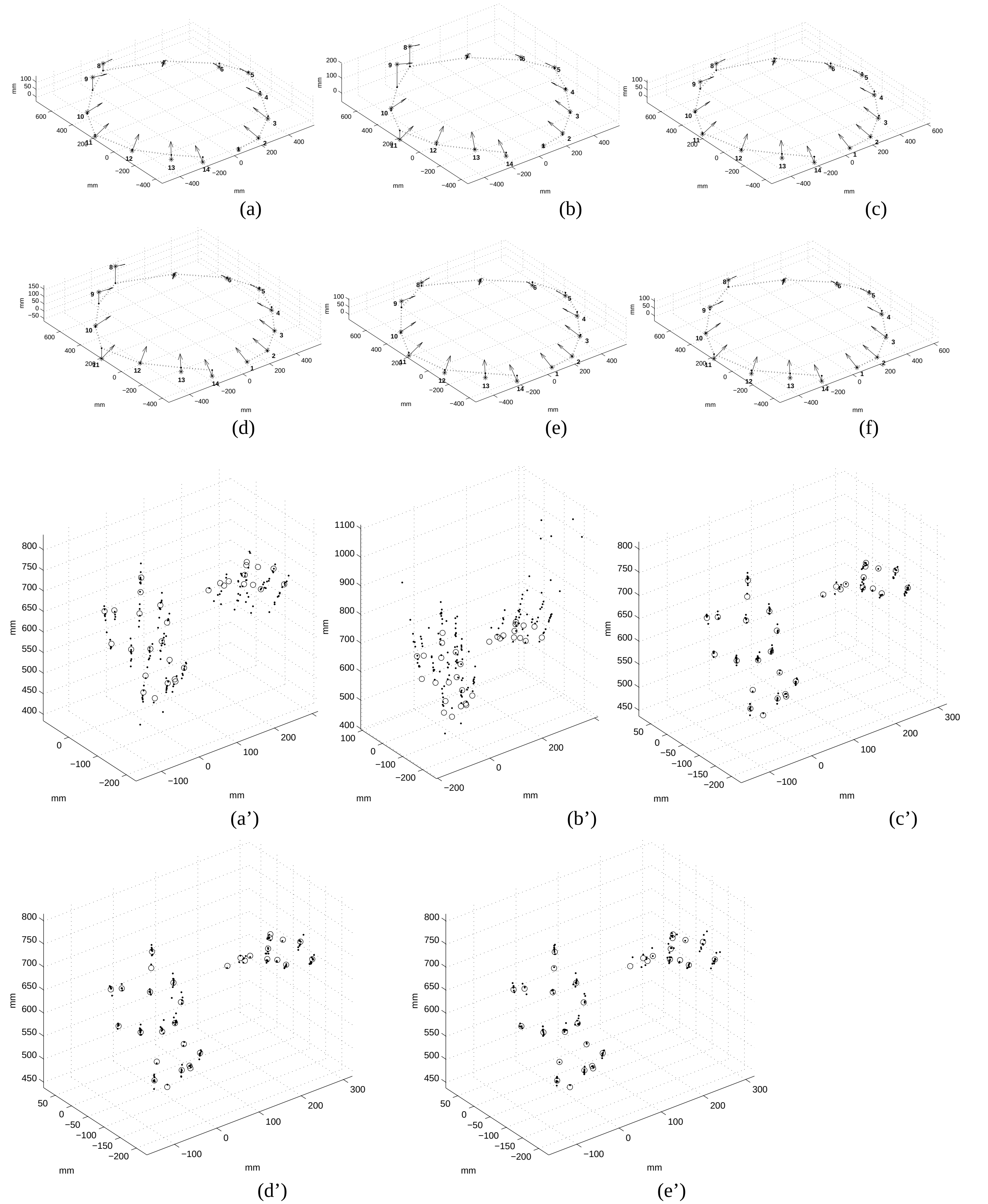

Figure 5: Estimation of trajectory and 3-D feature positions from various methods: (a,a’) left-camera sequence; (b,b’) right-camera sequence; (c,c’) Trinocular constraints with stereo data at time 0 with left frame at i; (d,d’) Trinocular constraints with stereo data at time 0 and right frame at i; (e,e’) quadranocular constraints with stereo data at times 0 and i; (f) trajectory from bundle adjustment over the entire data set. For trajectories, each camera position is shown with a star (*), and corresponding pose with an arrow. Projection of the positions onto XY plane (circles) have been connected with dashed lines. For 3-D points, estimates of each feature’s 3-D position from various views form a cluster, and estimated 3-D points based on bundle adjustment are shown with circles. See text for details.

References

[1] K.B. Atkinson, *Close range photogrammetry and machine vision*, Whittles Publishing, 1996.

[2] B.G. Baumgart, *Geometric modeling for computer vision*, Ph.D. thesis, Stanford University, 1974.

[3] E. Boyer, *Object models from contour sequences*, Proc. European Conf. Computer Vision, 1996.

[4] J.C. Bythell and J. Lee P. Pan, *Three-dimensional morphometric measurements of reef corals using underwater photogrammetry techniques*, Coral Reefs **20** (2001), no. 3.

[5] P. Eisert, E. Steinbach, and B. Girod, *Automatic reconstruction of stationary 3-d objects from multiple uncalibrated camera views*, IEEE TR-CSVT, Special Issue on 3D Video Technology **10** (2000), no. 3.

[6] A.W. Fitzgibbon, G. Cross, and A. Zisserman, *Automatic 3d model construction for turn-table sequences*, Proc. ECCV Workshop 3D Structure from Multiple Images of Large-Scale Environments (Freiburg), 1998.

[7] S. Fleischer, H.W. Wang, S.M. Rock, and M.J. Lee, *Video mosaicking along arbitrary vehicle paths*, Proc. Symp. AUV Technology (Monterey, CA), June 1996.

[8] N. Gracias and J. Santos-Victor, *Underwater mosaicing and trajectory reconstruction using global alignment*, Proc. Oceans'01 (Honolulu USA), 2001.

[9] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*, Cambridge U. Press, 2000.

[10] B.K.P. Horn, H.M. Hilden, and S. Negahdaripour, *Closed-form solution of absolute orientation using orithnormal matrices*, Journal of the Optical Society of America **5** (1987), no. 4.

[11] K. N. Kutulakos and Steven Seitz, *A theory of shape by space carving*, Technical Report TR692, Computer Science Dept., U. Rochester, May 1998.

[12] H.C. Longuet-Higgins, *A computer algorithm for reconstructing a scene from two projections*, Nature **293** (1981), no. 10.

[13] W. Matusik, C. Buehler, R. Raskar, S.J. Gortler, and L. McMillan, *Image-based visual hulls*, Siggraph 2000, Computer Graphics Proceedings (Kurt Akeley, ed.), ACM Press / ACM SIGGRAPH / Addison Wesley Longman, 2000.

[14] E. M. Mikhail, J. S. Bethel, J. C. McGlone, and C. McGlone, *Introduction to modern photogrammetry*, John Wiley & Sons, 2001.

[15] S. Negahdaripour and X. Xu, *Mosaic-based positioning and improved motion estimation methods for automatic navigation of submersible vehicles*, IEEE Trans. Oceanic Eng. **27** (2002), no. 1.

[16] W. Niem and J. Wingbermühle, *Automatic reconstruction of 3d objects using a mobile monoscopic camera*, Proc. Int. Conf. Recent Advances 3D Imaging and Modeling (Ottawa, Canada), 1997.

[17] C. Roman and H. Singh, *Estimation of error in large area underwater photomosaics using vehicle navigation data*, Proc. Oceans (Honolulu, HI), 2001.

[18] Y. Rzhanov, L.M. Linnett, and R. Forbes, *Underwater video mosaicking from seabed mapping*, Proc. Int. Conf. Image Proc. (Vancouver, CA), 2000.

[19] Steven Seitz and Charles Dyer, *Photorealistic scene reconstruction by voxel coloring*, International Journal of Computer Vision **25** (1999), no. 3.

[20] A. Shashua, *Trilinearity in visual recognition by alignment*, Proc. 3rd European Conference on Computer Vision (Stockholm), vol. 1, 1994.

[21] A. Shashua and M. Werman, *On the trilinear tensor of three perspective views and its underlying geometry*, Proc. 3th International Conference on Computer Vision (Boston), 1995.

[22] S. Sullivan and J. Ponce, *Automatic model construction, pose estimation, and object recognition from photographs using triangular splines*, Proc. International Conference on Computer Vision, 1998.

[23] R. Szeliski, *Rapid octree construction from image sequences*, Proc. Computer Vision, Graphics and Image Processing, 1993.

[24] R.Y. Tsai and T.S. Huang, *Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces*, IEEE Trans. Pattern Analysis Machine Intelligence **6** (1984), no. 1.

[25] S. Teukolsky W. Vetterling W. Press, B. Flannery, *Numerical recipes in* C, Cambridge U. Press, 1988.

[26] I.M. Williams and J.H.J. Leach, *Measurements of benthic species using drop-video platforms- comparative uses of stereo and single video systems*, Proc. Australian Marine Science Assoc. (Sydney, Australia), 1999.

[27] A.J. Woods, *The application of stereoscopic video to underwater remotely operated vehicles*, Proc. APPEA 97 conference (Melbourne), 1997.

[28] X. Xu and S. Negahdaripour, *Vision-based motion sensing for underwater navigation and mosaicing of ocean floor images*, Proc. Oceans (Halifax), October 1997.