

3-D Object Modeling from Occluding Contours in Opti-Acoustic Stereo Images

M. Babaee
Department of Computer science
Technical University of Munich
Munich, Germany
babaee@mytum.de

S. Negahdaripour
ECE Department
University of Miami
Coral Gables, FL 33146
nshahriar@miami.edu

Abstract—Utilizing in situ measurements to build 3-D volumetric object models under variety of turbidity conditions is highly desirable for marine sciences. Feature-based structure from motion and stereo methods can become ineffective under poor visibility, where matches cannot be reliably identified in low contrast optical images. To address this, a solution is proposed by utilizing a multi-modal stereo imaging technique with coincident optical and sonar cameras. Moreover, the challenges of establishing the complex opti-acoustic correspondences are avoided, by employing 2-D occluding contours that are the images of 3-D object rims as structural features. As we circle our target object, we process matching 2-D apparent contours to construct the corresponding 3-D object rim, and compute the trajectory of the stereo rig by opti-acoustic bundle adjustment in order to transform the 3-D object rims into registered samples of the object surface in a reference coordinate system.

In addition to exploiting range measurements from sonar that offer unique advantages, the proposed paradigm enables computing both the 3-D positions and local surface normals of 3-D contours, leading to improved object reconstruction accuracy. We demonstrate the performance of our method based on the 3-D surface rendering of certain objects, imaged by an underwater opti-acoustic stereo system.

I. INTRODUCTION

Feature-based structure from motion and stereo are selected approaches among “Shape from X” techniques in computer vision that have been explored extensively for building 3-D object and (or) scene models in underwater, e.g., natural reef objects, shipwrecks, and other subsea structures; e.g., [2], [7], [18], [24], [25], [27], [34]. Unfortunately, the effectiveness of these methods diminishes quickly with rising turbidity level. The multi-modal opti-acoustic stereo imaging is a rather recent paradigm, targeted to address the drawbacks of poor visibility by exploiting complementary information from the two sensing modalities [17], [19]. However, the new difficulty becomes solving the correspondence problem in multi-modal opti-acoustic stereo pairs.

As an example, Fig. 1 (a,b) shows selected underwater images of reef objects with either low contrast or nonuniform illumination. It is rather difficult to either identify a large enough number of distinct features or to robustly compute dense correspondences from stereo pairs. However, we also note from the opti-acoustic stereo pair in

(d,d) that while object-level correspondences can be readily established, identifying point feature correspondences is a formidable task.

Motivated by the need for alternative approaches that circumvent the dense correspondence problem, this work employs a multi-modal stereo imaging system under a rather special configuration, namely, where the centers of the two cameras’ coordinate systems and their axes align. While coincident centers is difficult to achieve exactly, it is sufficient to maintain a negligible separation relative to the target distance with parallel coordinate frames. Under this scenario, the feature matching problem is circumvented by utilizing 2-D apparent contours associated with the same 3-D occluding contour (viewed by the two cameras from the same vantage point). Referring to Fig. 1(c), the occluding contours are readily identified, e.g., by the application of Laplacian of Gaussian (LoG) operator. Next, circling the object, imaging 2-D apparent contours in the two images and matching them, and computing the trajectory of the stereo rig by opti-acoustic bundle adjustment, we can reconstruct 3-D object rims in the form of registered samples of the object surface in a reference coordinate system. In theory, recording opti-acoustic views continuously enables perfect 3-D object reconstruction. However, for relatively smooth surfaces, a sufficient number of discrete 3-D contours, corresponding to sparse stereo pairs, can often give a reasonably accurate reconstruction by interpolation.

The proposed methodology is not only novel in the underwater arena, but also offers many advantages over the related approaches based on a monocular camera for the terrestrial applications [3], [4], [8], [11], [12], [28], [33], [35], motivated by the classic paper by Koenderink [13], which unveiled many novel findings and ideas. For example, traditional binocular vision offers little advantage for 3-D object reconstruction by utilizing occluding contours, while the proposed multi-modal approach enables us to devise a simple practical 3-D object reconstruction solution based on: 1) exploiting redundant and complementary visual cues in 2-D optical and sonar images at each position of the sensor platform; 2) making use of images at each time instant, without the need to estimate second derivatives to compute depth from three (or more) nearby views (as commonly

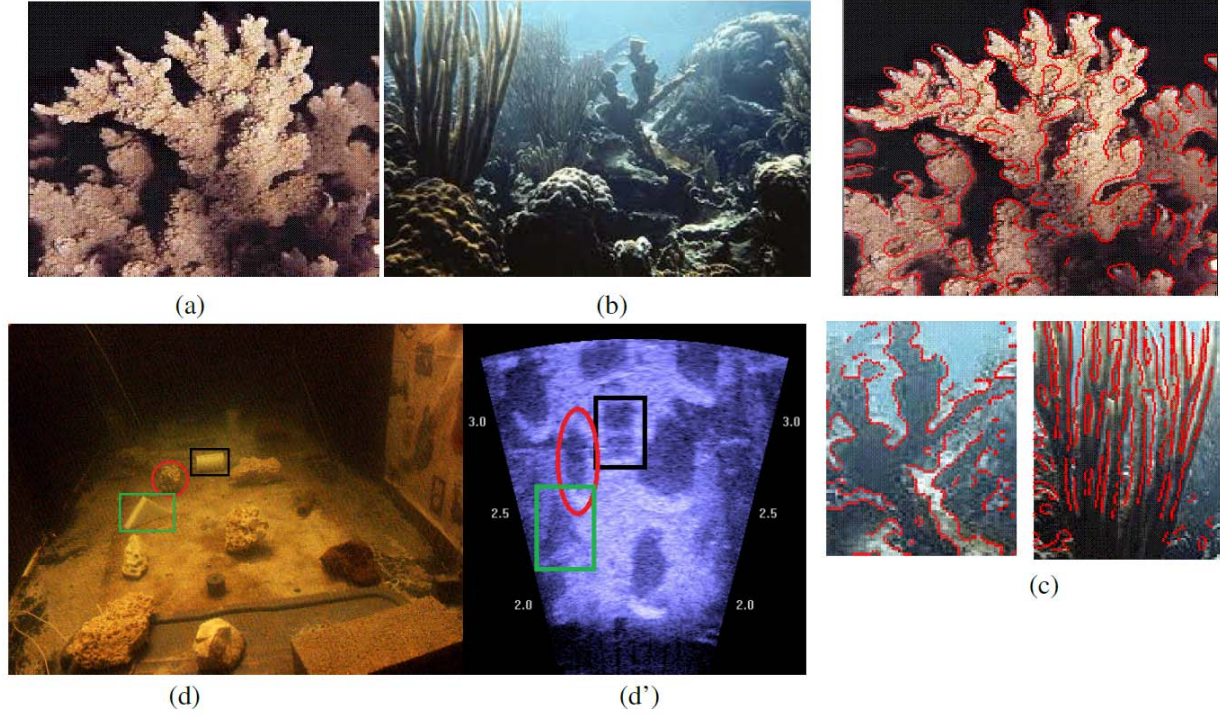


Fig. 1. (a,b) Two images of a reef objects and (c) extracted occluding contours by LoG operator. (d,d') Opti-acoustic stereo pair of a scene on a water tank bottom (optical image has been adjusted for lighting enhanced to improve visualization), with marked corresponding objects.

done when employing the occluding contours in monocular optical images); 3) overcoming the correspondence problem by making use of opti-acoustic epipolar geometry, with cameras at (roughly) the same position that image the same occluding contour; 4) enabling the recovery of both 3-D position and the surface normal at each point on the occluding rim; 5) relaxing the smooth surface assumption that is commonly made in earlier terrestrial methods based on monocular views, enabling application to complex shapes, and removing any restriction on the camera motion. Additionally, two other practical issues should not be overlooked. First, the visibility constraint from turbidity can be addressed by imaging objects at shorter distances. Next, the zero baseline requirement is a desirable feature for deployment on small submersible platforms, where the stereo rig dimension is no larger than the sonar size. We demonstrate the performance of our method based on the 3-D surface rendering of certain objects, imaged by an underwater opti-acoustic stereo system.

II. BACKGROUND

Fig. 2(a) depicts a scene surface point P with rectangular $\mathbf{P}_s = [X, Y, Z]^T$ and polar $[\mathfrak{R}, \theta, \varphi]$ coordinates that are related by

$$\mathbf{P}_s = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \mathfrak{R} \begin{bmatrix} \sin \theta \cos \phi \\ \cos \theta \cos \phi \\ \sin \phi \end{bmatrix} \quad (1)$$

The range $\mathfrak{R}$, azimuth angle θ , and elevation angle φ are given by

$$\mathfrak{R} = \sqrt{X^2 + Y^2 + Z^2} \quad \theta = \tan^{-1}(X/Y) \quad \varphi = \sin^{-1}(Z/\mathfrak{R}) \quad (2)$$

The acoustic signal transmitted by the sonar is reflected by scene surfaces, and the sonar beam-forming process transforms the backscattered acoustic energy from different ranges $\mathfrak{R}$ and azimuth angles θ into a 2-D so-called *beam-bin image* $I(\mathfrak{R}, \theta)$ [23]. A 2-D *polar image* $I(x_s, y_s)$ can be generated by transforming the two measurable polar coordinates of a 3-D point (namely, range and azimuth values $\{\mathfrak{R}, \theta\}$) to $\mathbf{s} = [x_s, y_s] = \mathfrak{R}[\sin \theta, \cos \theta]$, which defines points on the zero-elevation plane:

$$\mathbf{s} = \begin{bmatrix} x_s \\ y_s \end{bmatrix} = \begin{bmatrix} X \\ Y \end{bmatrix}_{\varphi=0} \quad (3)$$

Referring to Fig. 2(b), the sonar projection is modeled as the mapping of 3-D points $\mathbf{P}_s$ onto this plane, along meridians defined by the range and azimuth angles of the 3-D point. Here, the distinction from perspective projection in optical images is also highlighted.

We consider two sonar and optical cameras with the same projection center, and aligned coordinate systems, as depicted in Fig. 4(a). The occluding contours C projects onto corresponding contour generators C_o and C_s in an opti-acoustic stereo pair. Utilizing the conventional coordinate system for each camera as shown in (b), these are related by the extrinsic rotation ($\mathbf{R}$) and translational ($\mathbf{d}$)

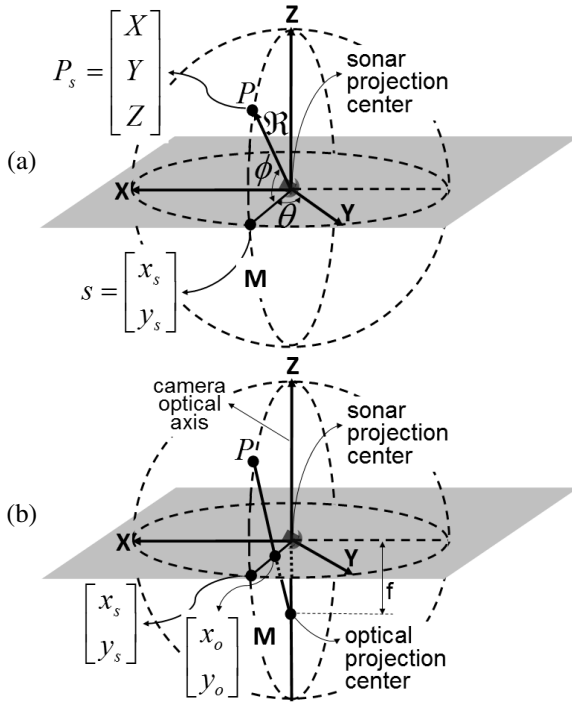


Fig. 2. (a) Sonar coordinate system and projecting 3-D point $\mathbf{P}$ onto image (zero-elevation plane). (b) Comparing sonar and optical projections of a 3-D point.

transformation parameters:

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}; \mathbf{d} = \mathbf{0} \quad (4)$$

The opti-acoustic epipolar geometry plays the main role in the fusion of 2-D optical and sonar contours in order to reconstruct the 3-D contour C at each position of the stereo rig. Utilizing the epipolar geometry given in [17], we can map each point on C_o to the corresponding epipolar curve in the sonar view. To elaborate, the projection ray of a point $\mathbf{o} = (x_o, y_o)$ on C_o , can be expressed by $\mathbf{P}_o = Z_o(x_o/f, y_o/f, 1)$. This is transformed to the sonar coordinate system using stereo calibration parameters $\{\mathbf{R}, \mathbf{d} \approx \mathbf{0}\}$:

$$\mathbf{P}_s(Z_o) = \mathbf{R}\mathbf{P}_o(Z_o) + \mathbf{d} \quad (5)$$

Calculating range $\mathfrak{R}$ and bearing θ in the sonar view:

$$\mathfrak{R}(Z_o) = \|\mathbf{P}_o\| \quad \theta(Z_o) = \tan^{-1}(X_s, Y_s) = \tan^{-1}(X_o, Z_o) \quad (6)$$

we finally determine the epipolar contour:

$$(x_s(Z_o), y_s(Z_o)) = \mathfrak{R}(\sin \theta, \cos \theta) \quad (7)$$

The contour can be confined to the range $Z_o = [Z_{\min}, Z_{\max}]$, if this is known a priori.

III. 3-D RECONSTRUCTION PROCESS

Fig. 3 depicts the flowchart of the proposed 3-D reconstruction technique comprises the steps of apparent image contour detection, 3-D occluding contour reconstruction, local normal estimation, view-to-view motion estimation and representation (alignment) of 3-D occluding contours in the reference frame, as well as dense reconstruction.

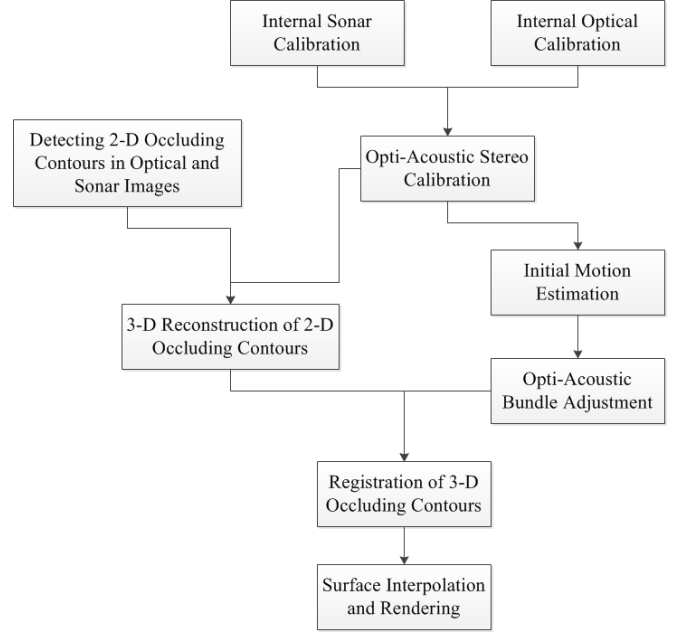


Fig. 3. Flowchart of proposed method.

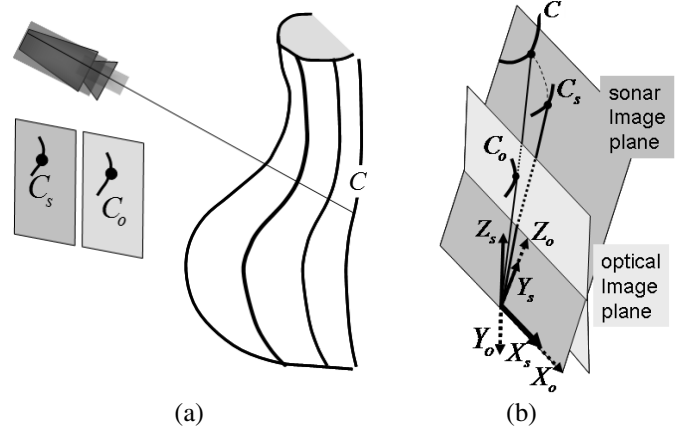


Fig. 4. (a) Occluding contours and (b) projection onto opti-acoustic stereo views yielding correspondences of 2-D apparent contours.

We assume a calibrated opti-acoustic stereo system, whereby we verify that the two cameras satisfy the assumed geometry: 1) baseline is sufficiently small (negligible) relative to target distances; 2) the cameras are aligned properly. In the appendix, we describe a simple method to make small adjustments to camera orientations to achieve the desired relative pose.

Imaging a target object in the optical camera, the points at which a visual ray is tangent to the surface (i.e., points of contour generator) are projected onto the optical image plane as occluding contour points. In the sonar image, this holds for points where the azimuth plane is tangent to the surface; see Fig. 4(b).

A. 2-D Contour Detection and 3-D Reconstruction

Occluding contours C_o and C_s at each stereo rig position are identified by edge detection. In typical configuration where the sonar is deployed at grazing incidence to max-

imize the diffuse backscatter component (relative to the specular component), objects cast a strong shadow on the background plane. This provides a strong cue in detecting the object and the occluding contour C_s in the sonar view; e.g., see Fig. 1(d'). Next, consider a point on either contour, say $\mathbf{o}$ on C_o . The correspondence $\mathbf{s}$ on C_s lies on the epipolar curve E_s , comprising points $(x_s(Z_o), y_s(Z_o))$ parameterized by the depth Z_o (as described in section II). The intersection with the occluding contour C_s gives the sonar match $(x_s(\widehat{Z}_o), y_s(\widehat{Z}_o))$, the simultaneously the depth $\widehat{Z}_o$, and the 3-D contour point $\widehat{\mathbf{P}}_o = \widehat{Z}_o(x_o/f, y_o/f, 1)$. Repeating for each point on C_o , we can determine the entire 3-D occluding contour C in the optical (or sonar coordinate system).

On important question is if the intersection can always be found. Alternatively, do degenerate scenarios exist where the intersection is either not defined or the solution is ill-conditioned (the two curves E_s and C_s either coincide or are nearly parallel at the sought after point). We have shown, in work to be published elsewhere, that degenerate configuration arises where the local slope of the occluding contour lies in a plane passing through the Y -axis of the optical camera; which coincides with the elevation Z -axis of the sonar.

B. Surface Normal Estimation

The unit surface normal $\widehat{\mathbf{n}}$ at a point $\mathbf{P}$ on contour generator C is perpendicular to the visual ray, and to the apparent contour. Expressed in the optical coordinate system, the normal $\widehat{\mathbf{n}}_o$ can be recovered from the optical image using the ray direction $\mathbf{o}$ and the local tangent to the occluding contour, $\mathbf{t}_o$:

$$\widehat{\mathbf{n}}_o = \frac{\mathbf{o} \times \mathbf{t}_o}{\|\mathbf{o} \times \mathbf{t}_o\|} \quad (8)$$

Projecting onto the sonar image, this is given by $\widehat{\mathbf{n}}_s^o = \mathbf{R}\widehat{\mathbf{n}}_o$. Similarly, we can determine an estimate $\widehat{\mathbf{n}}_s^s$ directly from the sonar view by computing the direction perpendicular to both the contour C_s and to the azimuthal plane:

$$\widehat{\mathbf{n}}_s^s = \frac{\mathbf{s} \times \mathbf{t}_s}{\|\mathbf{s} \times \mathbf{t}_s\|} \quad (9)$$

where $\mathbf{t}_s$ is the local tangent of contour C_s .

The accuracy of either estimate is directly related to two key factors: 1) localization accuracy of detected contour points; 2) calibration accuracy. In particular, the medium visibility in underwater imagery plays a critical role in optical image contrast and sharpness, as well as the impact of back-scatter and small-angle forward scattering. Given optical and sonar uncertainty measures σ_o and σ_s , respectively, we may express the solution as a weighted sum of the two estimates:

$$\mathbf{n}_s = \frac{\sigma_o}{\sigma_o + \sigma_s} \widehat{\mathbf{n}}_s^o + \frac{\sigma_s}{\sigma_o + \sigma_s} \widehat{\mathbf{n}}_s^s \quad (10)$$

This yields both the 3-D contour C_i and the local normal $\widehat{\mathbf{n}}_i$ at each position $i = 1, 2, \dots, N$ of the sonar rig (in either coordinate system).

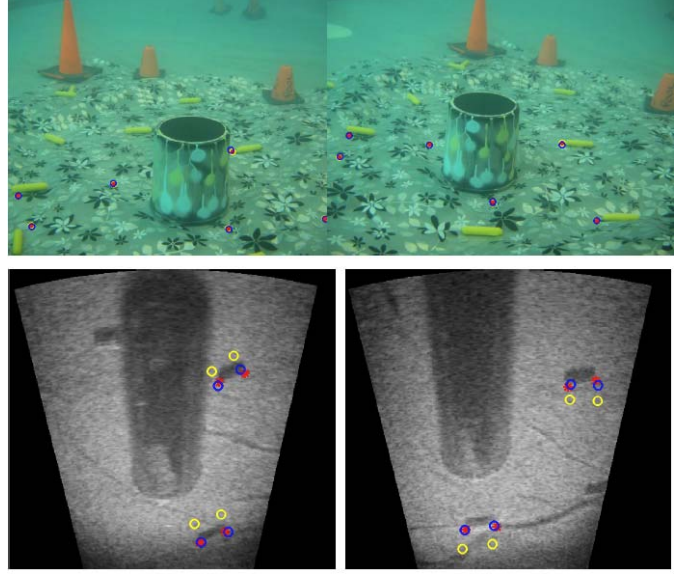


Fig. 5. Sample stereo pair highlights use of sonar matches to resolve scale ambiguity in optical sequences. A few optical correspondences (red dots) are re-projected by an arbitrary scale of estimated 3-D motion and structure (yellow dots). While optical features match, sonar features do not. This is resolved by utilizing correct scaling estimated from sonar features in 3-D estimation (blue dots).

C. Registration of 3-D Occluding Contours

The construction of object model makes use of 3-D contours $C(\mathbf{m}_i)$ from $i = 1, 2, \dots, N$ images over 360 [deg] views, where $\mathbf{m}_i$ is a 6-D parametrization of each stereo rig pose in some reference frame. Using $\mathbf{m}_i$ for these N positions, transformation of $C(\mathbf{m}_i)$ to the common reference frame yields *oriented 3-D contour points* (i.e., 3-D surface point samples with known local orientation). From this, we can readily construct a dense representation by applying a surface interpolation technique (e.g., [29]).

The estimation of camera pose $\mathbf{m}_i$ along its trajectory by tracking image features is the main focus of visual odometry in robotics (e.g., [5], [22], [26]). This has received renewed interest with the emergence of real-time techniques within the VisualSLAM framework (e.g., [6]). For a monocular vision system, only up-to-scale estimation is feasible due to the well-known inherent scaling ambiguity. In our application, the availability of 2-D opti-acoustic feature tracks enables both resolving this ambiguity, and improving the estimation by utilizing redundant visual motion cues from the two modalities [20]. In particular, availability of range measurements from sonar play a critical role in overcoming the ill-conditioned nature of 3-D reconstruction based on triangulation with two (nearly) parallel optical rays for a small/negligible baseline. Instead, “opti-acoustic stereo triangulation” involves an optical ray and a range sphere, which is a highly well-posed problem.

As detailed next, we compute the optimal estimate of the sensor platform trajectory by opti-acoustic bundle ad-

justment (utilizing the projections of scene features in the optical and sonar views). The iterative bundle adjustment scheme is initialized with the coarse frame-to-frame 3-D motion estimates.

1) *Trajectory Initialization*: We compute the initial pose and trajectory of the sensor package by integrating the frame-to-frame motions from opti-acoustic stereo data, that are computed as follows [20]:

- Compute the Euclidean homography $\mathbf{H}_i$ for consecutive optical views, which can be performed using a minimum of 4 features on the planar sea floor [9]. Alternatively, we can estimate the essential matrix $\mathbf{Q}_i$ from minimum of 5 non-planar points [21], and select among up to 10 solutions by utilizing additional features.
- Determine the up-to-scale motion $\{\mathbf{R}_i, k\mathbf{T}_i; \mathbf{n}_i/k\}$ from the decomposition of homography $\mathbf{H}_i$ [14], [31] (or essential matrix $\mathbf{Q}_i$ [9], [15], [32]).
- Applying the planar solution, select both motion scale and the true solution (from two possible ones [10], [14]) by employing a minimum of one sonar correspondence.
- Improve the (scaled) frame-to-frame motion by iteratively minimizing the reprojection errors in both the optical and sonar views (using solution from earlier steps as initial guess).

It is noted that the final iterative scheme does not require opti-acoustic correspondences, solely the correspondences over consecutive optical and sonar views, separately.

2) *Optimized Trajectory by Bundle Adjustment*: Bundle adjustment is the gold standard for estimation the camera trajectory [30]. In our application, the method is generalized to incorporate visual constraints from both the optical and sonar image sequences. The Maximum Likelihood Estimation is revised by defining the cost function:

$$\sum_{i,j=1}^{N,M} V_{i,j}^o d(P^o(\mathbf{m}_j^o, \mathbf{P}_i), x_{i,j}^o) + V_{i,j}^s d(P^s(\mathbf{m}_j^s, \mathbf{P}_i), x_{i,j}^s) \quad (11)$$

where 1) s and o subscripts are associated with the sonar and optical measurements; 2) P is the projection model; 3) $\mathbf{m}_j$ models the j -th camera pose in N views; 4) $\mathbf{P}_i$ represents a 3-D point (from total of N), and $x_{i,j}$ is the 2-D projection of i -th point in j -th view; 5) d is some distant measure, L-2 norm in our results; 6) $V_{i,j}$ is the $\{0,1\}$ visibility mask (set to one for a point $\mathbf{P}_i$ visible in j -th view, and zero, otherwise). This can be generalize to a non-binary value that also accounts for the data reliability, e.g., a fixed weight for relative optical-to-sonar data uncertainty in its simplest form. (In our results, no weighting was incorporated.) Finally, it is noted that the sonar and optical camera poses are constrained by the fixed configuration of the stereo rig, namely, $\mathbf{m}_j^s = \mathbf{R}\mathbf{m}_j^o$.

IV. EXPERIMENTS

The finite size of waterproof housing for underwater camera deployment prohibits the realization of coinciding optical and sonar cameras, exactly. However, we can approximate the desired zero-baseline assumption by minimizing the baseline while imaging targets at a relatively large distance (giving a large target distance-to-baseline ratio).

The optical camera is mounted on top of a DIDSON camera in our setup, which rotates around the target while collecting stereo images. An experiment is described, comprising stereo pairs at roughly every 25 [deg]. (Both translational and rotational motions vary from one pose to the next.) We estimate the sensor platform trajectory by opti-acoustic bundle adjustment described in Section III-C2, utilizing 1-2 dozen optical and a handful of sonar features. The optical features have been tracked by utilizing the SURF descriptor [1]. For sonar features, the same or other common descriptors for feature detection and matching (e.g., SIFT [16]) perform poorly. While this step is a component of the entire 3-D reconstruction system, it does not directly relate to primary theoretical contribution, namely, the 3-D reconstruction of oriented occluding contour points from opti-acoustic stereo data. Therefore, we have manually matched these few sonar features over consecutive frames. In practice, robust acoustic features (deflectors) can be positioned strategically over the object to be imaged and modeled, to facilitate the automated sonar feature detection and tracking.

Feature \ View	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	*	*	*	*	*	*	*	*						*	*
2	*	*	*	*	*	*	*	*						*	*
3	*	*	*	*	*	*	*	*				*	*	*	*
4	*	*	*	*	*	*	*	*			*	*	*	*	*
5	*	*	*	*	*	*	*	*					*	*	*
6	*	*	*	*	*	*	*	*						*	*
7	**	**	**	**	**	*	**	**			*	*	**	**	**
8	**	**	**	**	**	*	**	**	*	*	*	*	*	*	**
9	*	**	**	**	**	*	*	*					*	*	**
10	**	**	**	*	**	**	*	*	*	*	*	*	**	**	**
11	**	**	**	**	**	**	*	*	*	*		**	**	**	**
12	*	*	*					*	*	*	*		*	*	*
13	*	*									*		*	*	*
14	*	*											*	*	*
15	*	*								*	*	*	*	*	*
16	*	*		*	*	*	*	*					*	*	*
17	*	*	*	*	*	*	*	*	*				*	*	*
18	*	*	*	*	*								*	*	*
19	*	*	*	*	*								*	*	*
20	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
21	*	*			*	*	*	*	*	*	*	*	*	*	*
22	*	*											*	*	*
23	*	*						*			*	*	*	*	*
24	*	*							*	*	*	*	*	*	*
25	*	*			*	*	*	*	*				*	*	*
26	*	*	*	*	*								*	*	*
27	*														*

TABLE I
VISIBILITY MATRIX DEPICTING VIEWS IN WHICH EACH FEATURE IS VISIBLE IN OPTICAL (BLACK) AND SONAR (RED) VIEWS.

Fig. 5 depicts 2 samples from 15 opti-acoustic pairs. Three different objects, namely, the plastic cover of a

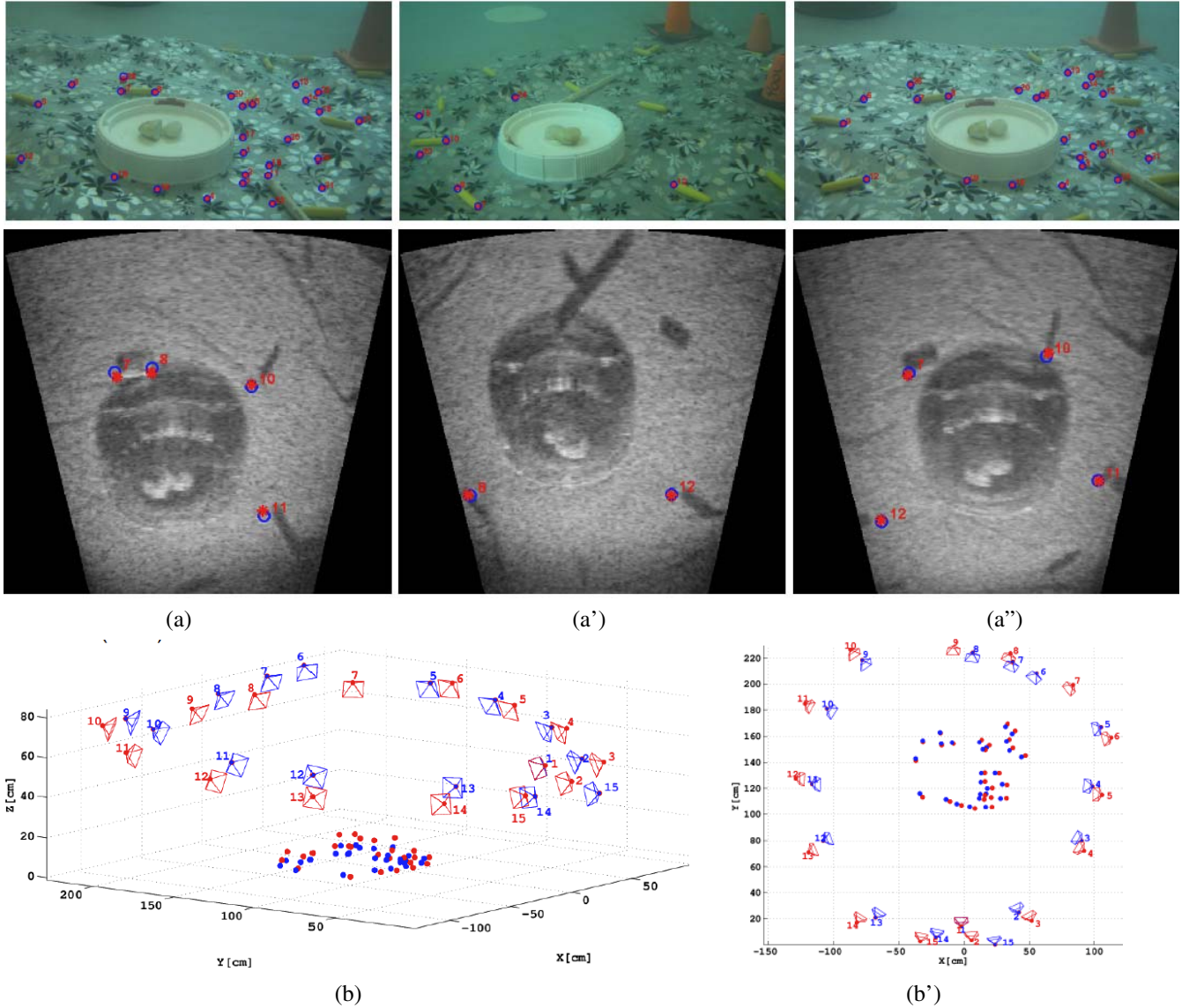


Fig. 6. (a,a',a'') Sample stereo pair highlights use of sonar matches to resolve scale ambiguity in optical sequences. A few optical correspondences (red dots) are re-projected by an arbitrary scale of estimated 3-D motion and structure (yellow dots). While optical features match, sonar features do not. This is resolved by utilizing correct scaling estimated from sonar features in 3-D estimation (blue dots). (b,b') Two views of estimated 3-D feature positions and camera poses, before (blue) and after (red) bundle adjustment.

cylindrical container and two stones, are considered, among which the plastic cover is in the degenerate configuration (due to vertical object rims that are nearly parallel to the elevation axis of the sonar).

The small number of correspondences depicted by red dots are utilized to demonstrate how we resolve the motion scaling ambiguity in monocular optical sequences. First, the up-to-scale motion and 3-D scene feature positions are calculated from the optical image correspondences. For demonstration purposes only, we arbitrarily fix the scale to determine the absolute translational motion and 3-D feature positions, and to transform the 3-D motion and feature positions to the sonar coordinate system. Next, the 3-D points are reprojected onto all 4 views (yellow

circles). It is noted that the errors are small for the 2 optical images (which are solely due to the estimation inaccuracy), but rather large in the sonar views (primarily due to inaccurate motion scaling). Incorporating the sonar matches in the estimation process [20], the absolute 3-D motion and feature positions can be recalculated. This yields the reprojected points depicted by the blue circles, fitting well with the optical and sonar features in the 4 consecutive opti-acoustic stereo views.

Fig. 6 depicts 3 more views and all of the correspondences (red dots) on the flat bottom, used in determining the poses of the stereo rig. Table I gives the visibility mask. The markers show if point i is visible in the j -th optical (black) and sonar (red) views. It is immediately noted that

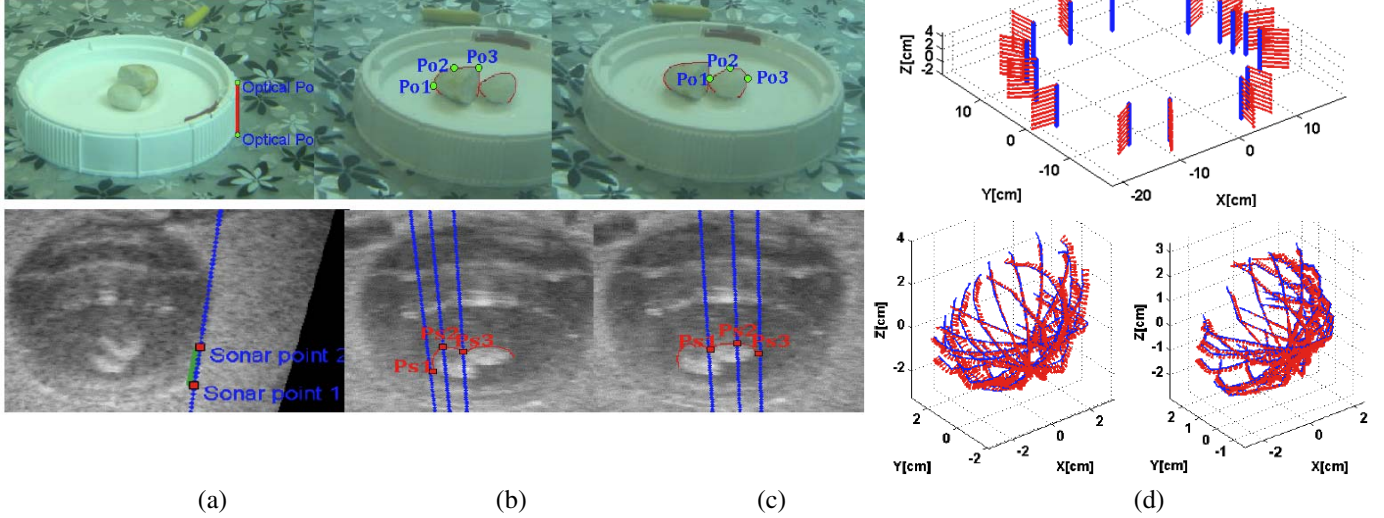


Fig. 7. Sample occluding contours of (a) cylindrical cover in both views, and two corresponding points; (b,c) same contours for two stones, with three selected correspondences on each object. (d) Registered 3-D contours and normals of all three objects.

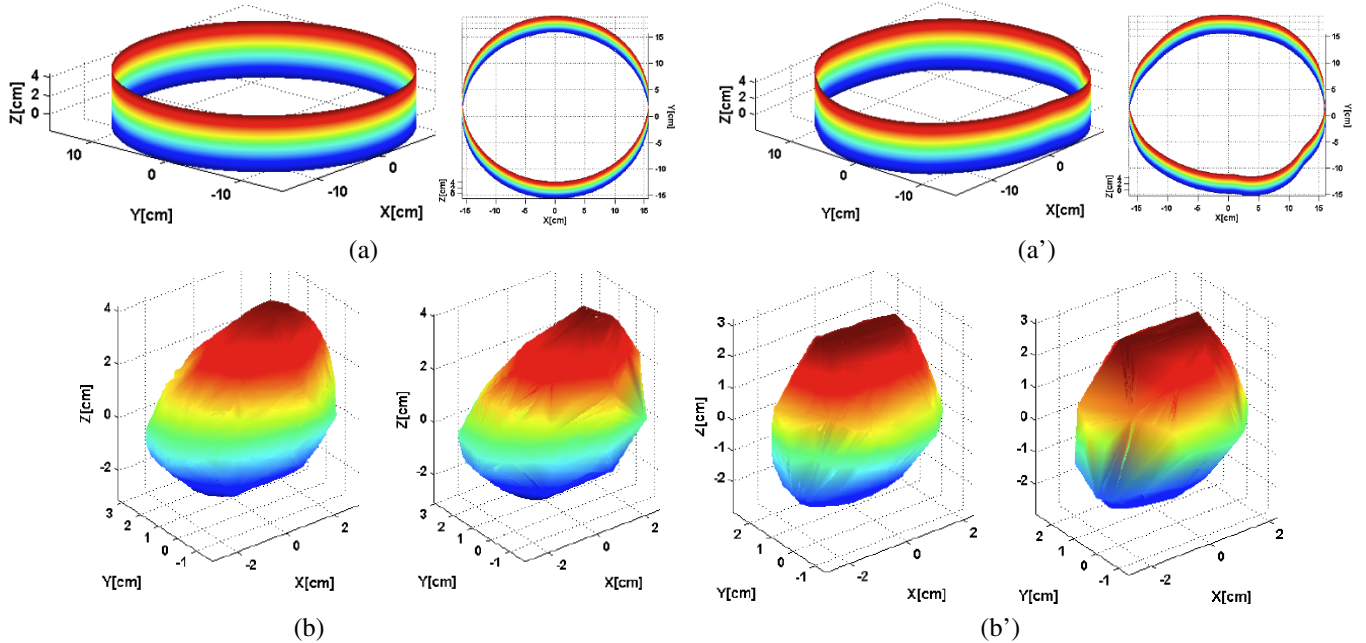


Fig. 8. Reconstructions of plastic cover based on (a) proposed method and (a') only 3-D contour positions without surface normal. (b) and (c) same results for the 2 stones.

each sonar view contains no more than 5 features. The blue dots show the reprojected points after the calculation of each frame-to-frame motion. The estimate poses for the 15 viewpoints along the trajectory and the positions of the 3-D features are given in Figs. 6(b,b'), both before (blue) and after (red) bundle adjustment.

Fig. 7(a) shows the right occluding contour of the plastic cylindrical cover in both views, where the degenerate configuration of stereo cameras relative to the target can be readily verified; the occluding contour becomes aligned with the epipolar line, giving no intersection point. In this case, we have manually matched the two end points of

the optical contour to their correspondences in the sonar image by exploiting the shadow and highlight boundary cues. Computing the positions of these end points in 3-D, we can then determine the 3-D object rim (straight line) and consequently match all other contour points in the two views (by reprojection). Additionally, we can readily determine the local surface normal based on the position of each apparent contour point, and the local contour tangents; see (8). In Fig. 8(d), we depict the registered 3-D oriented occluding counter set for all 15 views of the cylinder.

In Fig. 7(b,c), we have given the contours of the two stones, with three selected correspondences on each object.

Figs. 7(d) depict the estimated 3-D contours and surface normals of these objects in a common reference frame. In Fig. 8, we have given various reconstructions. In particular, we have compared for various objects the solutions based on the proposed method with those where only the 3-D contour positions are used, without knowledge of surface normals; e.g., compare (a) and (a'). The same two methods for the reconstructions of stones 1 and 2 are given in (b) and (b'), respectively (left view is our method and right view is the solution without surface normal information). The 3-D reconstruction accuracy is readily confirmed for the cylindrical plastic cover, and by visual size and shape inspection for the two stones. These results also support the significance of local surface normal information in improving the 3-D object reconstruction.

V. SUMMARY

This paper presents a solution for the 3-D volumetric modeling of underwater objects from their 2-D apparent contours. Utilizing the opti-acoustic stereo imaging paradigm with coincident camera configuration enables us to exploit some unique advantages in the integration of multi-modal visual cues and data: 1) generating models of natural and man-made objects in situ under much higher turbidity levels that existing feature-based monocular and (or) stereo vision techniques may allow; 2) theoretically, building continuous models by reconstructing the 3-D contours at each position of the sensor platform; 3) improving reconstruction accuracy based on both 3-D position and surface normals at each point. Results of experimental with real data demonstrate the potential application of our method, and together with the theoretical results on degenerate configurations based on target pose relative to the stereo cameras (reported elsewhere) enable devising effective strategies in stereo rig path planning to enhance data acquisition and accurate reconstruction.

REFERENCES

- [1] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. *Computer Vision—ECCV 2006*, pages 404–417, 2006.
- [2] C. Beall, B. Lawrence, V. Ila, and F. Dellaert. 3d reconstruction of underwater structures. In *Intelligent Robots and Systems, Proc.*, pages 4418–4423, 2010.
- [3] A. Blake and R. Cipolla. Robust estimation of surface curvature from deformation of apparent contours. *Computer Vision, Proc. European Conf.*, pages 465–474, 1990.
- [4] E. Boyer and M. Berger. 3d surface reconstruction using occluding contours. *Int. J. Computer Vision*, 22(3):219–233, 1997.
- [5] J. Campbell, R. Sukthankar, I. Nourbakhsh, and I. Pittsburgh. Techniques for evaluating optical flow for visual odometry in extreme terrain. In *Intell. Robots and Systems, 2004. Proc. Conf.*, pages 652–659. IEEE, 2004.
- [6] A. Davison, I. Reid, N. Molton, and O. Stasse. Monoslam: Real-time single camera slam. *IEEE T. Pattern Analysis Machine Intelligence*, 29(6):1052–1067, June 2007.
- [7] R. Eustice, H. Singh, and J. Leonard. Exactly sparse delayed-state filters for view-based slam. *IEEE T. Robotics*, 22(6):1100–1114, December 2006.
- [8] P. Giblin and R. Weiss. Reconstruction of surfaces from profiles. In *Computer Vision, Proc. First Int. Conf.*, pages 136–144, London, UK, 1987. IEEE.
- [9] R. Hartley and A. Zisserman. Multiple view geometry in computer vision. 2004.
- [10] J. Hay. Optical motions and space perception: An extension of gibsons analysis. *Psychology Review*, 73(6):550–565, 1966.
- [11] T. Joshi, N. Ahuja, and J. Ponce. Structure and motion estimation from dynamic silhouettes under perspective projection. In *Computer Vision, Proc. Int. Conf.*, pages 290–295, Boston, MA, June 1995.
- [12] K. Kang, J. Tarel, R. Fishman, and D. Cooper. A linear dual-space approach to 3d surface reconstruction from occluding contours using algebraic surfaces. In *Computer Vision, 2001. ICCV 2001. Proc. Eighth IEEE Int. Conference*, volume 1, pages 198–204. IEEE, 2001.
- [13] J. Koenderink. What does the occluding contour tell us about solid shape. *Perception*, 13(3):321–330, 1984.
- [14] H. Longuet-Higgins. The visual ambiguity of a moving plane. *Proc. Royal Society of London. Series B. Biological Sciences*, 223(1231):165–175, 1984.
- [15] H. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Readings in Computer Vision: Issues, Problems, Principles, and Paradigms*, MA Fischler and O. Firschein, eds, pages 61–62, 1987.
- [16] D. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Computer Vision*, 60(2):91–110, 2004.
- [17] S. Negahdaripour. Epipolar geometry of opti-acoustic stereo imaging. *Pattern Analysis and Machine Intelligence, IEEE T.*, 29(10):1776–1788, 2007.
- [18] S. Negahdaripour and H. Madjidi. Stereovision system for rov deployment in mapping of benthic habitats and seafloor structures. *IEEE J. Oceanic Engineering: Special issue on vision & image processing*, 28(4):625–650, October 2003.
- [19] S. Negahdaripour and A. Taatian. 3-d motion and structure estimation for arbitrary scenes from 2-d optical and sonar video. In *OCEANS 2008*, pages 1–8. IEEE, 2008.
- [20] S. Negahdaripour and A. Taatian. Integration of motion cues in 2-d optical and sonar video for 3-d positioning. *Computer Vision Image Understanding*, 114(8):928–941, August 2010.
- [21] D. Nister. An efficient solution to the five-point relative pose problem. *IEEE T. Pattern Anal. Machine Intelligence*, 26(6):756–770, 2004.
- [22] D. Nister, O. Naroditsky, and J. Bergen. Visual odometry. In *Computer Vision and Pattern Recognition, 2004. Proc. Conf. on*, volume 1, page 652:659. IEEE, 2004.
- [23] M. Ramirez. *On techniques for the formation and post processing of sonar images*. M.S. Thesis, Electronics Engineering – Metrology. Univ. Applied Sciences, Bremen, Germany, April 2008.
- [24] A. Sarafraz, S. Negahdaripour, and Y. Schechner. Enhancing images in scattering media utilizing stereovision and polarization. In *WACV'09, Proc.*, 2009.
- [25] H. Singh, C. Roman, O. Pizarro, and R. Eustice. Advances in high-resolution imaging from underwater vehicles. In *Intl. Symp. Robotics Research*, october 2005.
- [26] B. Steder, G. Grisetti, C. Stachniss, and W. Burgard. Visual slam for flying vehicles. *Robotics, IEEE T.*, 24(5):1088–1093, 2008.
- [27] Y. Swirski, Y. Schechner, B. Herzberg, and S. Negahdaripour. Caustereo: Range from light in nature. *Applied Optics*, 50(28):89–101, 2011.
- [28] R. Szeliski and R. Weiss. Robust shape recovery from occluding contours using a linear smoother. *Computer Vision and Pattern Recognition, Proc. IEEE Conf.*, 28(1):666–, 1993.
- [29] D. Terzopoulos. Image analysis using multigrid relaxation method. *IEEE T. Pattern Anal. Machine Intelligence*, 8(2):129–139, 1986.
- [30] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. Bundle adjustment: a modern synthesis. *Vision algorithms: theory and practice*, pages 153–177, 2000.
- [31] R. Tsai and T. Huang. Estimating three-dimensional motion parameters of a rigid planar patch. *Acoustics, Speech and Signal Processing, IEEE T.*, 29(6):1147–1152, 1981.
- [32] R. Tsai and T. Huang. Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *Pattern Anal. Machine Intell., IEEE T.*, 6(1):13–27, 1984.
- [33] R. Vaillant and O. Faugeras. Using extremal boundaries for 3-d object modeling. *IEEE T. Pattern Analysis and Machine Intelligence*, 14(2):157–173, 1992.
- [34] H. Zhang and S. Negahdaripour. EKF-based recursive dual estimation of structure & motion from stereo data. *IEEE J. Oceanic Engineering*, 35(2):424–437, 2010.
- [35] C. Zhao, R. Mohr, et al. Global three-dimensional surface reconstruction from occluding contours. *Computer Vision and Image Understanding*, 64(1):62–96, 1996.