Opti-Acoustic Stereo Imaging, System Calibration and 3-D Reconstruction

S. Negahdaripour, H. Sekkati, H. Pirsiavash
Electrical and Computer Engineering Dept.
University of Miami
Coral Gables, FL 33124-0620
[shahriar, sekkati]@miami.edu, h.pirsiavash@umiami.edu

Abstract

Utilization of an acoustic camera for range measurements is a key advantage for 3-D shape recovery of underwater targets by opti-acoustic stereo imaging, where the associated epipolar geometry of optical and acoustic image correspondences can be described in terms of conic sections. In this paper, we propose methods for system calibration and 3-D scene reconstruction by maximum likelihood estimation from noisy image measurements. The recursive 3-D reconstruction method utilized as initial condition a closed-form solution that integrates the advantages of so-called range and azimuth solutions. Synthetic data tests are given to provide insight into the merits of the new target imaging and 3-D reconstruction paradigm, while experiments with real data confirm the findings based on computer simulations, and demonstrate the merits of this novel 3-D reconstruction paradigm.

1. Introduction

Visual search, inspection and survey are critical in a number of underwater applications in marine sciences, maintenance and repair of undersea structures and homeland security. Video cameras, traditionally optical and more recently acoustic, provide suitable sensing technologies. However, dealing with environmental conditions that can change drastically with time and season, location, depth, etc., calls for novel methodologies and deployment strategies. As an example, extending visibility in naturally illuminated underwater images has been demonstrated by polarization-based image analysis that utilizes the image formation physics [15]. The method makes use of at least two images taken through a polarizer at different orientations (e.g., horizontal and vertical) to improve scene contrast and to accomplish color correction. Advantages can also come from the simultaneous use of different and complementary sensors to exploit their unique strengths and properties, while overcoming the shortcoming(s) and limitation(s) of each sensing modality.

Where visibility allows, potential integration of optical and acoustic information can enhance the performance in comparison to the processing of images from each sensor, alone. This multi-sensor fusion strategy has been explored for registering image data to known 3-D object models [5, 4], and to automatically navigate along natural contours on the sea floor, such as sea grass [1]. The key advantage here is the exploitation of valuable scene information from a *3-D sonar* [6].

In recent years, high-frequency 2-D acoustic cameras have emerged [13]; e.g., *D*ual-Frequency *ID*entification *SON*ar (DIDSON) [20] and BlueView based on blazed-array technology [19]. Video imagery from these systems provides high enough details that allow visual target recognition by human operators in search and inspection [3, 16]. The deployment in stereo configuration with an optical camera was recently proposed as a novel strategy for 3-D object reconstruction in underwater applications [11].One advantage over the use of 3-D acoustic cameras is the availability of visual data for target recognition and classification, in addition to 3-D geometric information. Investigation of some immediate fundamental problems has led to: 1) Establishing the epipolar geometry of the so-called *"opti-acoustic stereo imaging"*; 2) Derivation of certain *closed-form solutions* that utilize various combinations in three out of four constraints imposed by the corresponding acoustic and optical projections in two stereo views. Furthermore, computer simulations suggest improved 3-D reconstruction performance compared to triangulation in a traditional binocular system. Just as for optical systems, noisy *"opti-acoustic correspondences"* do not satisfy the epipolar geometry, and therefore 3-D reconstruction from any of these closed-form methods is *sub-optimal* with respect to the maximum-likelihood estimates (MLE) that take advantage of all four constraints [14]. Two approaches based on direct and indirect estimation of 3-D target points from noisy observations were shown to produce comparable re-

sults. Here, each MLE rests on the representation of range and bearing measurement noises by the Gaussian model.

This paper proposes improved methods for system calibration and 3-D reconstruction from opti-acoustic stereo imaging, relative to work previously reported in the literature [12, 14]. For example, the stereo calibration method in [14] relies on determining the pose of each camera relative to a planar calibration target. Being a critical step in establishing the epipolar geometry with high accuracy, we propose a more accurate method utilizing a minimum of 5 opti-acoustic correspondences to directly compute the relative pose of the two cameras. Next, a new recursive 3-D reconstruction method is proposed by reformulating the MLE problem with a revised noise model. Here, transformation from the sonar range-bearing measurements to a rectangular image form allows us to apply the Gaussian model to the uncertainty in the rectangular image positions. This corresponds to the modeling of noise in the range by the more suitable Raleigh distribution [7, 17]. The nonlinear estimation problem is solved iteratively by the application of Levenberg-Marquardt algorithm [10]. Since a good initial condition enhances the performance of recursive schemes, we also improve on the closed-form range and azimuth solutions in [12]. More precisely, we propose a new weighted average solution of these two earlier ones, by careful examination of their performances. The weighting function is chosen based on two important parameters associated with the optimum performance regions of the range and azimuth solutions, namely, the target distance and stereo baseline. Both are readily known based on stereo system geometry and sonar measurements. Finally, we report results from experiments with data collected in an indoor pool[1] and a 6'×12'×6' water tank facility. Utilizing these, we put into test the various theories of opti-acoustic imaging and epipolar geometry with real data.

Effective application of the 3-D reconstruction methods, say during the course of an online operation in sea, requires automatic robust and accurate matching of corresponding features in optical and acoustic images; this also sits at the heart of 3-D reconstruction from two or more 2-D optical views. The results in this paper utilize manually matched features, as we are mainly assessing the performances of the calibration and 3-D reconstruction methods. However, we are currently investigating a promising approach to the opti-acoustic correspondence problem. This study will explore various relevant complex problems, such as matching points at apparent edges, occlusion, etc.

[1]Courtesy of colleagues from Teledyne Benthos who made available to us the use of their pool facility.

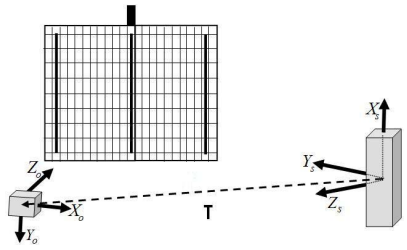


Figure 1. (Geometry of stereo cameras relative to world frame, aligned with axes of planar calibration grid.

2. Preliminaries

Acoustic Imaging: Acoustic cameras produce an image by recording the reflected sound, once the scene is insonified by acoustic pulse(s). In a 3-D sonar, e.g., Echoscope [18], the back-scattered signals are collected by a 2-D array of transducers, and the image is formed from "beam signals" – Echoes from *fixed steering directions*, specified by elevation ϕ and azimuth θ angles (see fig. 1). Range $\Re$ of 3-D points is determined from the round-trip travel time of the acoustic wave based on the peak of the beam signal.

The 2-D acoustic image formation is based on transmitting a number of beams at varying bearing (azimuth) angles, recording range based on time-of-flight. Two existing technologies, namely, DIDSON and BlueView, record a total of 512 range values within a fixed down-range (DR) window $[\Re_{\text{min}} - \Re_{\text{max}}]$ [m]; set to image objects within known distances from the camera, and thus establishing down-range resolution. Formed with acoustic lenses and transducer curvature, DIDSON generates 96 beams with roughly $w_\theta = 0.3$ [deg] azimuth and $w_\phi = 14$ [deg] elevation widths. The transmitted beams cover a total field of view of 28.8 [deg] in the azimuth direction, with 0.3 [deg] resolution; this translates to a cross-range (CR) of roughly 0.5∗DR [m]). Built on blazed-array technology, BlueView offers 45-degrees in cross-range field of view with 1 [deg] resolution. Treatment of each system as a 2-D sonar is because of the ± 7 [deg] uncertainty in the elevation of an imaged 3-D point. Our methods are applicable to any 2-D forward sector-scan sonar, though DIDSON is discussed in the remainder, as our real data were acquired with this system.

These 2-D acoustic cameras produce high-quality images in turbid waters [2, 16], however, the small elevation width of the transmitted beam limits the coverage area. Therefore, the target is typically viewed at relatively small grazing angles to increase the likelihood of imaging object features with distinct sonar returns in each frame; see fig. 1(b).

Rectangular & Spherical Coordinates: A 3-D point $\mathbf{P}$ may be expressed by rectangular or spherical coordinates, $[X, Y, Z]^T$ or $[\theta, \phi, \Re]^T$, respectively, where θ and ϕ are azimuth and elevation angles, and $\Re$ is the range. The relationship between rectangular and spherical coordinates and

the inverse transformation are

$$\mathbf{P}=\begin{bmatrix}X\\Y\\Z\end{bmatrix}=\Re\begin{bmatrix}\cos\phi\sin\theta\\\cos\phi\cos\theta\\\sin\phi\end{bmatrix}\begin{cases}\Re &=\sqrt{X^2+Y^2+Z^2}\\\theta &=\tan^{-1}\left(\frac{X}{Y}\right)\\\phi &=\tan^{-1}\left(\frac{Z}{\sqrt{X^2+Y^2}}\right)\end{cases} \quad (1)$$

Coordinate Frames and Transformation: Let $\mathbf{P}_o = (X_o, Y_o, Z_o)^T$ and $\mathbf{P}_s = (X_s, Y_s, Z_s)^T$ denote the coordinates of a 3-D world point $\mathbf{P}$ in the rectangular coordinate frames. Without loss of generality, the optical coordinate system is taken as the world reference frame. The relative pose of the two cameras is expressed by a rigid body motion transformation, comprising a 3×3 rotational matrix and a 3-D translational vector $\mathbf{M} = [\mathbf{R}, \mathbf{T}]$, where

$$\mathbf{P}_s = \mathbf{R}\mathbf{P}_o + \mathbf{T} \quad (2)$$

Image Measurements: We assume that the 2-D position (x, y) of the image of a 3-D scene feature $\mathbf{P}$ in the optical view satisfy the perspective projection model. Including the range and azimuth measurements of $\mathbf{P}$ in the acoustic image, we collectively have the opti-acoustic image measurements:

$$\begin{cases}x = f\frac{X_o}{Z_o}\\y = f\frac{Y_o}{Z_o}\\\Re=\sqrt{(\mathbf{r}_1^T\mathbf{P_o}+T_x)^2+(\mathbf{r}_2^T\mathbf{P_o}+T_y)^2+(\mathbf{r}_3^T\mathbf{P_o}+T_z)^2}\\\theta=\tan^{-1}\left((\mathbf{r}_1^T\mathbf{P_o}+T_x)/(\mathbf{r}_2^T\mathbf{P_o}+T_y)\right)\end{cases} \quad (3)$$

where $\mathbf{r_i}$ denotes i-th row of $\mathbf{R}$. A rectangular sonar image with symmetric coordinate units $\mathbf{p}_s = (x_s, y_s)$ can be constructed based on the following transformation:

$$\mathbf{p}_s = \begin{bmatrix}x_s\\y_s\end{bmatrix} = \Re\begin{bmatrix}\sin\theta\\\cos\theta\end{bmatrix} \quad (4)$$

It readily follows that

$$\Re = \sqrt{x_s^2 + y_s^2} \quad \text{and} \quad \theta = \tan^{-1}(x_s/y_s) \quad (5)$$

Transformation of optical image positions to computer coordinates is readily achieved by a linear mapping based on the camera intrinsic parameters [**?**]. A similar linear mapping is applied to construct the sonar image from $\{x_s, y_s\}$, by specifying arbitrarily either the row or column dimension.

Stereo Correspondence Constraint: The relationship between opti-acoustic correspondences $\mathbf{p}_o = (x, y, f)$ and $\mathbf{p}_s = (x_s, y_s)$ is the fundamental constraint not only for 3-D reconstruction, but also for other relevant problems, such as stereo calibration. This is derived from the transformation in (2), and can be expressed in the form

$$\begin{bmatrix}x_s\\y_s\end{bmatrix} = \frac{1}{\cos\phi}\left(\left(\frac{Z_o}{f}\right)\begin{bmatrix}\mathbf{r}_1\cdot\mathbf{p}_o\\\mathbf{r}_2\cdot\mathbf{p}_o\end{bmatrix}+\begin{bmatrix}T_x\\T_y\end{bmatrix}\right) \quad (6)$$

The dependent unknown ϕ can be eliminated by noting that

$$\cos\phi = \sqrt{1-\left(\frac{Z_s}{\Re}\right)^2} = \sqrt{1-\left(\frac{(Z_o/f)(\mathbf{r}_3\cdot\mathbf{p}_o)+T_z}{\Re}\right)^2} \quad (7)$$

Finally, we arrive at

$$\mathbf{p}_s = \sqrt{\frac{\Re^2}{\Re^2-(Z_o/f(\mathbf{r}_3\cdot\mathbf{p}_o)+T_z)^2}} * \left(\left(\frac{Z_o}{f}\right)\begin{bmatrix}\mathbf{r}_1\cdot\mathbf{p}_o\\\mathbf{r}_2\cdot\mathbf{p}_o\end{bmatrix}+\begin{bmatrix}T_x\\T_y\end{bmatrix}\right) \quad (8)$$

Measurement Noise Model: Image positions are typically noisy observations of the true 2-D projections $\{\mathbf{p}_o, \mathbf{p}_s\}$, denoted $\{\hat{\mathbf{p}_o}, \hat{\mathbf{p}_s}\}$ here. The measurement noise directly affects the accuracy in the solutions of the calibration and 3-D reconstruction methods. Image measurement noise may be modeled as additive Gaussian:

$$\begin{cases}\hat{x} = x + n(0, \sigma_{xo})\\\hat{y} = y + n(0, \sigma_{yo})\end{cases} \quad \begin{cases}\hat{x}_s = x_s + n(0, \sigma_{xs})\\\hat{y}_s = y_s + n(0, \sigma_{ys})\end{cases} \quad (9)$$

where $n(0, \sigma_i)$ is a normal distribution with zero mean and variance σ_i ($i = xo, yo, xs, ys$). Independent Gaussian model of the uncertainties in x_s and y_s translates to Raleigh distribution for range uncertainties commonly assumed for sonar imaging systems [7, 17].Thus, an advantage in working with $\mathbf{p}_s = (x_s, y_s)$ for sonar image coordinate representation of a 3-D point $\mathbf{P}_s$ is that the image noise is suitably modeled as Gaussian.

3. Epipolar Geometry

The epipolar geometry is fundamental to the 3-D reconstruction from calibrated stereo views. For example, it allows us to solve the correspondence problem as a 1-D search along the epipolar contours. While the epipolar geometry of an opti-acoustic system has been explored in details in [11], it is useful for completeness to summarize some relevant results.

The epipolar constraint in an opti-acoustic stereo system – establishing the relationship between projections $\mathbf{p}_s$ and $\mathbf{p}_o$ of the same scene point $\mathbf{P}$ [11] – is derived by manipulating (2) and (3):

$$\mathcal{C}(\mathbf{p}_o, \mathbf{p}_s) = \mathbf{p}_o^T\mathbf{U}(\mathbf{p}_s)\,\mathbf{p}_o = 0 \quad (10)$$

The 3×3 symmetric matrix $\mathbf{U}$ is given by

$$\mathbf{U}(\mathbf{p}_s) = (x_s\tilde{T}_y - y_s\tilde{T}_x)^2\mathbf{I} + \left(\|\tilde{\mathbf{T}}\|^2 - 1\right)\mathbf{a}\mathbf{a}^T + (x_s\tilde{T}_y - y_s\tilde{T}_x)(\mathbf{a}\tilde{\mathbf{T}}^T\mathbf{R} + \mathbf{R}^T\tilde{\mathbf{T}}\mathbf{a}^T) \quad (11)$$

where $\mathbf{a} = (y_s\mathbf{r_1} - x_s\,\mathbf{r_2})^T$ and $\tilde{\mathbf{T}} = [\tilde{T}_x, \tilde{T}_y, \tilde{T}_z] = \mathbf{T}/\Re$. As noted, the match $\mathbf{p}_o$ in the optical image of an sonar-image point $\mathbf{p}_s$ lies on a conic section. It often becomes

necessary to establish the match $\mathbf{p}_s$ of an optical image point $\mathbf{p}_o$. It has also been shown that the epipolar geometry in the sonar image satisfies the following constraint [11]:

$$\Re = \sqrt{N_\theta / D_\theta} \tag{12}$$

$$N_\Re(\theta) = \left(\boldsymbol{\theta}^T \Upsilon \mathbf{T}\right)^2 + \left(\tilde{z}^T \Upsilon \mathbf{T}\right)^2 \quad D_\Re(\theta) = \left(\tilde{z}^T \Upsilon \boldsymbol{\theta}\right)^2 \tag{13}$$

Here, $\tilde{z} = (0, 0, 1)^T$, $\boldsymbol{\theta} = (\sin\theta, \cos\theta, 0)^T$, and Υ is a 3×3 skew-symmetric matrix defined in terms of components of $\upsilon = \mathbf{R}\mathbf{p}_o$ (such that $\Upsilon\mathbf{x} = \upsilon \times \mathbf{x}$, for any vector $\mathbf{x}$).

4. Calibration of Opti-Acoustic Stereo System

As for optical images, imperfections of an acoustic lens lead to image distortions and geometric deviations from ideal image model. A method for the *intrinsic calibration* of a DIDSON camera has been devised, determining the lens distortion parameters by utilizing one or more images of a known planar grid [12].

The relative pose of the optical and sonar cameras can be established by *extrinsic* or *stereo calibration*, allowing us to exploit the epipolar geometry in reducing the correspondence problem to a 1-D search along the epipolar curves. To do this, we also utilize a target with prominent opti-acoustic features that can be readily matched, ideally automatically but manually if necessary. Again, we can utilize a planar grid. Manual or manually guided feature matching is acceptable as calibration is often carried out as an off-line process, computing results that are later applied for 3-D reconstruction in online applications.

It can be readily shown that points on a plane satisfy the relationship

$$f/Z_o = -(\mathbf{n}_o \cdot \mathbf{p}_o) \tag{14}$$

where $\mathbf{n}_o = (n_{ox}, n_{oy}, n_{oz})^T$ is the inward surface normal in the optical camera coordinate system, and Z_0 is the distance to the plane along the Z axis of the optical camera. For calibration, we need the surface normal $\mathbf{n}_s = (n_{sx}, n_{sy}, n_{sz})^T$ in the sonar coordinate system. It can be shown that this is given by

$$\mathbf{n}_s = \frac{\mathbf{R}\mathbf{n}_o}{1 - \mathbf{T}^T \mathbf{R}\mathbf{n}_o} \tag{15}$$

In establishing the relative pose of stereo cameras, orthogonality of the rotation matrix $\mathbf{R}$ has to be enforced. This can be achieved in several forms. We use the decomposition into 3 rotations about the axes of the coordinate system: $\mathbf{R}(\alpha_x, \alpha_y, \alpha_z) = \mathbf{R}_z(\alpha_z)\mathbf{R}_y(\alpha_y)\mathbf{R}_x(\alpha_x)$ where $\mathbf{R}_u(\alpha_u)$ denotes a rotation about axis u of the respective coordinate system by angle α_u. Each match provides two constraints as given in (8), in terms of 9 unknowns: The 6 pose parameters $\mathbf{M} = [\mathbf{R}(\alpha_x, \alpha_y, \alpha_z), \mathbf{T}]$ and 3 parameters of the normal $\mathbf{n}_o$ of the calibration target plane in the optical camera coordinate frame. We have redundancy with $N \geq 5$ correspondences. We can solve a non-linear optimization problem based on a suitable error measure.

We have adopted a modified implementation that minimizes the 3-D distances between the reconstructions of planar grid points from the optical and sonar projections: Assume an estimate $\hat{\mathbf{M}} = [\hat{\mathbf{R}}, \hat{\mathbf{T}}]$ and $\hat{\mathbf{n}}_o$ of the 9 sought after parameters. Comprising the initial condition of our nonlinear optimization algorithm, these are updated during each step of an iterative estimation process. For a feature $\hat{\mathbf{p}_o}$ in the optical image, we estimate the depth $\hat{Z}_o$ from the plane equation in (14). Computing the other two coordinates $\hat{X}_o$ and $\hat{Y}_o$ from (3), we have an estimate of the 3-D point $\mathbf{P}_o$. Utilizing (2), transformation to the sonar coordinate system with $\hat{\mathbf{M}} = [\hat{\mathbf{R}}, \hat{\mathbf{T}}]$ gives us the estimated position $\hat{\mathbf{P}}_s$. Next, we calculate for the sonar match $\hat{\mathbf{p}_s}$ the elevation angle

$$\phi = -\gamma + \sin^{-1}\left(\frac{-1}{\sqrt{(\hat{n}_{sx}x_s + \hat{n}_{sy}y_s)^2 + \Re^2\hat{n}_{sz}^2}}\right) \tag{16}$$

where

$$\gamma = \tan^{-1}\left(\frac{\hat{n}_{sx}x_s + \hat{n}_{sy}y_s}{\Re\hat{n}_{sz}}\right) \tag{17}$$

The coordinate $\mathbf{P}_s$ of the 3-D point in the sonar would be obtained from (1). Transforming to the optical coordinate system with $\hat{\mathbf{M}} = [\hat{\mathbf{R}}, \hat{\mathbf{T}}]$ yields $\hat{\mathbf{P}}_o$. The estimation problem is solved by minimizing

$$\begin{aligned} e(\mathbf{M}, \mathbf{n}_o) = \; & \textstyle\sum_{i=1}^{n} (\mathbf{P}_o - \hat{\mathbf{P}}_o)^T \Sigma_{Po}^{-1} (\mathbf{P}_o - \hat{\mathbf{P}}_o) + \\ & \textstyle\sum_{i=1}^{n} (\mathbf{P}_s - \hat{\mathbf{P}}_s)^T \Sigma_{Ps}^{-1} (\mathbf{P}_s - \hat{\mathbf{P}}_s) \end{aligned} \tag{18}$$

We have estimated of the covariances Σ_{Po} and Σ_{Ps} analytically based on the first-order approximation:

$$\Sigma_{Px} = (\partial\mathbf{P}_x/\mathbf{p}_x)\, \Sigma_{px}\, (\partial\mathbf{P}_x/\mathbf{p}_x)^T \quad x = \{s, o\} \tag{19}$$

where $\Sigma_{px} = \sigma_{px}^2 \mathbf{I}_{2\times 2}$ is set based on the uncertainty σ_{px} ($x = \{s, o\}$) in image positions (assumed to be equal in row and column directions). The nonlinear optimization problem in (18) has been solved by the Levenberg-Marquardt algorithm [10].

5. 3-D Reconstruction

Given an opti-acoustic correspondence, the corresponding 3-D point can be calculated in closed form. However, an optimal solution in the maximum likelihood (ML) sense is derived from a nonlinear method. As this requires the application of iterative algorithms, we seek a good initial estimate to improve the convergence rate. Examining the performance of the so-called range and azimuth closed-form solutions proposed in [11], we can devise an improved weighted average that takes advantage of conditions when these two solutions perform best. This serves to initialize

our iterative direct method. Clearly, identifying the closed-form solution with the best estimate with noisy data enhances the convergence of the recursive method.

Closed-form Solutions: In the context of opti-acoustic stereo imaging, stereo triangulation deals with determining the 3-D point $\mathbf{P} = (X, Y, Z)$ – or equivalently, the position in either camera reference frames, say $\mathbf{P}_o = (X_o, Y_o, Z_o)$ – for any opti-acoustic correspondence $\mathbf{p}_o = (x, y, f)$ and $\mathbf{p}_s = (x_s, y_s)$. The asymmetrical form of the optical and sonar projection models in (3) leads to derivation of various closed-form solutions, each with a particular geometric interpretation.

The *Range solution*, the intersection of the optical ray with the range sphere, is computed from the positive solution of

$$(\|\frac{\mathbf{p}_o}{f}\|^2)Z_{\Re}^2 + \frac{2}{f}(\mathbf{T}^T\mathbf{R}\,\mathbf{p_o})Z_{\Re} - (\Re^2 - \|\mathbf{T}\|^2) = 0 \quad (20)$$

The correct solution is the one in agreement with the so-called *azimuth solution* – the intersection of the optical ray with the azimuth plane:

$$Z_\theta = f\frac{\tan\theta\, T_y - T_x}{(\mathbf{r_1} - \tan\theta\mathbf{r_2})\cdot\mathbf{p}_o} \quad (21)$$

The fusion of the two earlier solutions by weighted averaging gives

$$Z_m = \xi_t Z_\theta + (1-\xi_t)Z_{\Re} \quad (22)$$

where the transition in the weight ξ_t, chosen in the form of a sigmoid function, takes into account the characteristics of the range and azimuth solutions. It then becomes necessary to establish conditions under which one solution outperforms the other, and to determine if/how these depend on imaging and environmental factors. This has been established analytically, and verified by computer simulations, based on the first-order approximation to the variance of each solution

Fig. 2 shows the variances of the range and azimuth solutions for various range and baselines in the form of 2-D iso-distance contour plots. Formalizing these findings in assigning the weighting factor in (22), azimuth solution is weighted more heavily for larger baselines, while range solution would contribute more heavily for larger target distances. Defining the weight in terms of the ratio of the baseline to the target distance serves the objective:

$$\xi_t \quad = (1 + e^{-(\|\mathbf{T}\|/\bar{Z} - k_o)})^{-1} \quad (23)$$

where $\bar{Z} = (Z_{\Re} + Z_\theta)/2$. The threshold $k_o = \|\mathbf{T}\|/Z_c$ is set by determining, for a given stereo baseline, the so-called critical depth Z_c where the depth and azimuth solutions have equal variances; see fig. 2(b). This threshold can be pre-calculated and stored in a lookup table.

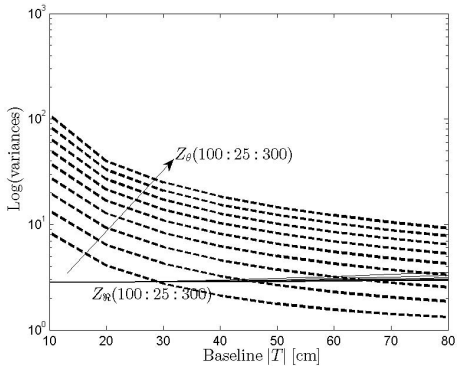


Figure 2. Error Variances of azimuth and range solutions for varying baseline as 2-D iso-distance contours, allowing the selection of proper scaling to compute the weighted average solution Z_m.

Maximum Likelihood Formulation: We want to formulate an optimization problem to compute the Maximum Likelihood Estimate (MLE) of a 3D point from the noisy opti-acoustic correspondences $\hat{\mathbf{p}}_o = (\hat{x}, \hat{y}, f)^T$ and $\hat{\mathbf{p}}_s = (\hat{x_s}, \hat{y_s})$. Representing the measurement uncertainties as zero-mean Gaussian, the MLE is determined by minimizing the Mahalanobis distance between the vectors $X = (x, y, x_s, y_s)^T$ and $\hat{X} = (\hat{x}, \hat{y}, \hat{x_s}, \hat{y_s})^T$:

$$\text{Minimize } \mathcal{E}\,(\mathbf{P}_o) = (X - \hat{X})^T\Sigma^{-1}(X - \hat{X}) \quad (24)$$

where $\Sigma = E[(X - \hat{X})(X - \hat{X})^T]$. Here, we utilize the projection model in (3) with 3-D points expressed in the optical camera coordinate frame. It is reasonable to assume independency among components of the measured vector $\hat{X}$, allowing us to write Σ as a diagonal matrix with elements σ_i. This leads to

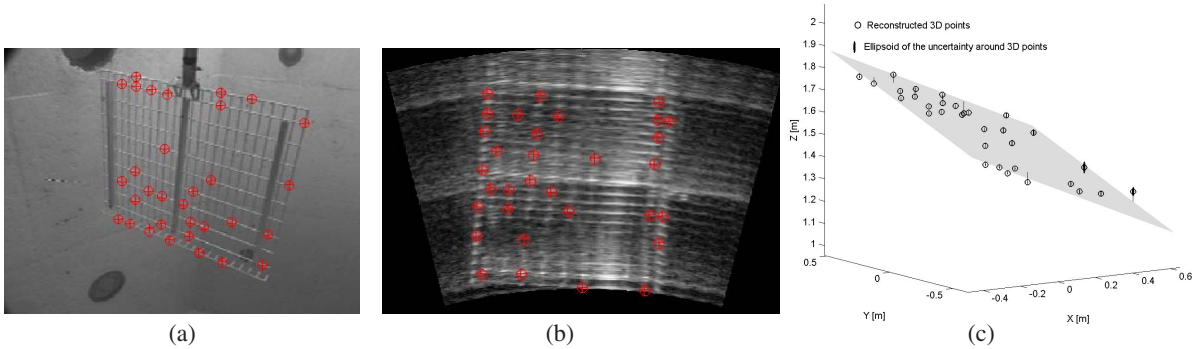
$$\text{Min } \mathcal{E} = \frac{(x-\hat{x})^2}{\sigma_x^2} + \frac{(y-\hat{y})^2}{\sigma_y^2} + \frac{(x_s-\hat{x_s})^2}{\sigma_{x_s}^2} + \frac{(y_s-\hat{y_s})^2}{\sigma_{y_s}^2} \quad (25)$$

This nonlinear optimization problem is efficiently solved using the Levenberg-Marquardt algorithm [10].

6. Experiments

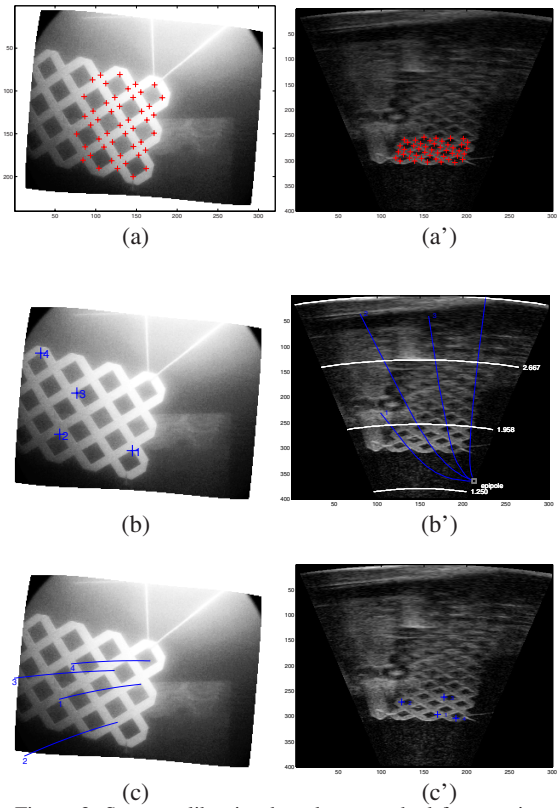
Calibration: The relative pose of the stereo cameras, determined by exterior calibration, fixes the epipolar geometry. The immediate advantage is that the match of a feature in one image can be located by a 1-D search along the corresponding epipolar curve in the other stereo view. We start with results from sample experiments in the calibration of opti-acoustic stereo cameras in different configurations. In addition to the verification of epipolar geometry, we have utilized these results in assisting us to manually establish image correspondences in the 3-D reconstruction experiments.

Fig. 3 depicts a sample data set used in applying the calibration method described in section 4. We have shown the



(a) (b) (c)

Figure 4. Circles depict matching points in optical (a) and sonar (b) views used for 3D reconstruction, while crosses are the projections of 3D reconstructed points. (c) 3D reconstructed points are depicted with the estimated planar surface as determined by calibration.



(a) (a') (b) (b') (c) (c')

Figure 3. Stereo calibration based on matched features in stereo pairs (a,a') establishes the epipolar geometry. For a feature in each image, we can compute the corresponding epipolar contour in the other image according to (10) and (12).

opti-acoustic stereo pair, superimposed by the selected correspondences used in calibration. We can examine if, for any given feature in one image, the corresponding epipolar contours pass through the matching feature in the other view. We have selected 4 sample feature points in each image, and computed the corresponding epipolar curve in the other stereo view according to (10) and (12); see fig. 3.

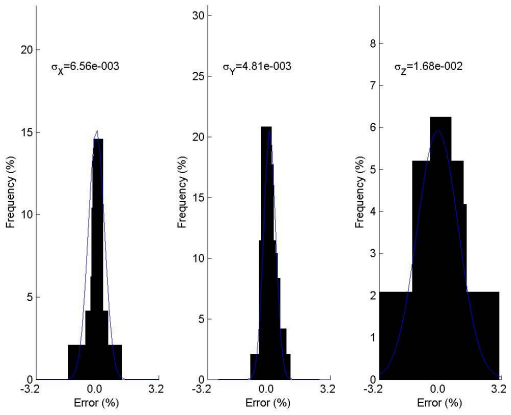


Figure 5. Reconstruction errors of planar grid points.

Once can readily verify the accuracy of these correspondences.

3-D Reconstruction: We present results from experiments with two real data sets. The calibration method has been employed in each case, before applying the 3-D reconstruction method of section 5.

The first data set comprise a stereo pair of a planar metallic grid at different orientations, collected in an indoor pool; see fig. 4. The grid is at a distance of roughly 1.5 [m] from the optical camera, which is located at about 2.7 [m] to the left of the sonar camera. The sonar images have been corrupted by multiple returns from the water surface and various pool walls; see 4(b). Each stereo pair depicts the matched features (circles) and the projections of the reconstructed 3-D points (crosses). The reconstructed points and uncertainty ellipsoids, as well as the plane of the grid computed by our calibration algorithm have been given in (c).

We have used for the estimation error the distance of each reconstructed point from the plane. While we do not know perfect ground truth, we have used the estimate from the calibration process, which is independent of the recon-

struction technique. (Recall that the calibration method gives both the stereo configuration and the normal of the target plane in each camera coordinate frame.) Here, we expect that the estimate from calibration is reasonably accurate since it is determined from a MLE formulation with a reasonably large number of opti-acoustic correspondences. Referring to fig. 5, the estimated 3-D points are within 3.5% of their distances to the optical cameras, utilizing the reconstruction of the plane by the calibration algorithm as ground truth. Overall, the reconstruction errors are much smaller for the X and Y coordinates than for the Z component of the 3-D points. This behavior is reminiscent of binocular optical stereo imaging with (nearly) parallel cameras; In this example, X_s and Y_s axes of the sonar system are nearly aligned with the $-Y_o$ and $-X_o$ axes of the optical camera.

The next data set comprises a stereo pair, collected in a water tank with better acoustic insulation; see fig. 6. The scene comprise an artificial reef and a toy lobster, each hung on a rope in front of a plastic planar grid. The grid is placed at an average distance of about 1-1.1 [m] along the viewing direction (Z_o axis) of the optical camera, which is positioned at a distance of about 1.2 [m] to the left of the sonar camera. Here again, X_s and Y_s axes of the sonar system are nearly aligned with the $-Y_o$ and $-X_o$ axes of the optical camera. First, certain matching grid points were chosen to calibrate the stereo system. Next, some of these and other grid points, selected points on the two objects – artificial reef and toy lobster – and a few on the supporting ropes were manually matched for 3-D reconstruction. These points have been numbered so they can be readily referenced. While grid corners could be readily matched, the epipolar geometry was employed in assisting us to establish the correspondences. Fig. 6 depicts two views of the reconstructed points. First, we can verify that the grid points depicted by black circles lie on a single plane. Next, various distances of points on each object as well as the distances of various objects agree well with manual measurements.

7. Summary and Conclusions

We have studied the 3-D reconstruction of objects in underwater by opti-acoustic stereo imaging – a paradigm to integrate information from optical and sonar camera systems with overlapping views. We have proposed and analyzed methods for system calibration and target scene reconstruction. Our calibration technique employs a minimum of 5 correspondences from features on a planar grid to compute the relative pose of the stereo cameras.

The asymmetrical nature of the optical and sonar projection equations and the redundant constraints from an opti-acoustic correspondence for the reconstruction of corresponding 3-D points lend themselves to the derivation of different closed-form solutions. Two such solutions based on independent employment of the range and azimuth measurements have simple geometric interpretations in the context of "triangulation" within the opti-acoustic stereo imaging framework. Neither solution provides an optimum estimate in the maximum likelihood sense with noisy data, and thus we have formulated a standard nonlinear optimization problem for computing the MLE of 3-D target points from opti-acoustic correspondences. Since the solution is determined iteratively, convergence can be enhanced by initialization with a good initial condition. This is obtained from a weighted average of our two closed-form solutions: With the proposed formula for the weighting function, this gives an estimate that fully utilizes the advantages of each of the two solutions for a larger range of imaging conditions.

Results from two experiments have been given to assess the performance of the 3-D reconstruction method, revealing the potentials of this novel paradigm for underwater 3-D object reconstruction in a wider range of environmental conditions. We are currently exploring a promising approach for addressing the correspondence problem, aimed at devising a robust opti-acoustic stereo matching method. This is a critical component of our efforts, aimed at bringing to bear a complete computer system for the 3-D reconstruction of underwater objects.

Acknowledgement: This work is based on research supported by ONR under grant N000140510717. Views, opinions and conclusions of the authors are not necessarily shared and endorsed by ONR.

References

[1] C. Barat, M.J. Rendas, "Exploiting natural contours for automatic sonar-to-video calibration," *Proc. Oceans* Volume 1, Brest, France, June, 2005, pp. 271–275.

[2] Belcher, E.O., Fox, W.L.J. and Hanot, W.H., "Dual-frequency acoustic camera: a candidate for an obstacle avoidance, gap-filter, and identification sensor for untethered underwater vehicles," *Proc. MTS/IEEE Oceans02*, vol 4, 2002, pp. 2124–2128.

[3] E.O. Belcher, D.G. Gallagher, J.R. Barone, and R.E. Honaker, "Acoustic lens camera and underwater display combine to provide efficient and effective hull and berth inspections," *Proc. Oceans'03*, San Diego, CA, September, 2003, pp. 1361-1367.

[4] U. Castellani, A. Fusiello, V. Murino, L. Papaleo, E. Puppo, M. Pittore, "A complete system for on-line 3D modelling from acoustic images," *Signal Proc. Image Comm.*, Vol 20(9-10), 2005, pp. 832–852

[5] A. Fusiello, and V. Murino, "Augmented scene modeling and visualization by optical and acoustic sensor integration," *IEEE T. Visual. Comp. Graphics*, Vol 10(5), Nov-Dec, 2004, pp. 625-636.

[6] R.K. Hanson, and P.A. Anderson, “A 3-D underwater acoustic camera – properties and application,” in: P. Tortoli, L. Masotti (Eds.), Acoustic Imaging, Plenum Press, New York, 1996, pp. 607–611.

[7] E. Jakeman, R.J. Tough, “Generalized K distribution: a statistical model for weak scattering,” *Jour. Opt. Soc. Amer.*, vol 4, September 1987, pp. 1764–1772.

[8] K. Kim, N. Neretti, and N. Intrator, “Non-iterative Construction of Super-resolution Image from an Acoustic Camera Video Sequence,” *Proc. CIHSPS*, 2005, pp. 105-111.

[9] D.M. Kocak, F.M. Caimi, ”The current art of underwater imaging with a glimpse of the past and vision of the future,” *Marine Technology Society Journal*, Vol 39(3), 2005, pp. 5–26.

[10] D. Marquardt, ”An algorithm for least-squares estimation of nonlinear parameters,” *J. Society Indust. and Applied Math.*, Vol 11(2), 1963, pp. 431–441.

[11] S. Negahdaripour, ”Theoretical foundations for opti-acoustic stereo imaging,” *Proc. IEEE Oceans*, Brest, France, 2005 (to appear in *IEEE Trans. PAMI*).

[12] S. Negahdaripour, ”Calibration of DIDSON forward-scan acoustic video camera,” *Proc. IEEE/MTS Oceans*, Washington, DC, August, 2005, pp. 1287-1294.

[13] L.J. Rosenblum, B. Kamgar-Parsi, E.O. Belcher, and O. Engelsen, ”Acoustic imaging: The reconstruction of underwater objects,” *IEEE Visual.*, 1991, pp. 94–101.

[14] H. Sekkati, and S. Negahdaripour, ”Direct and indirect 3-D reconstruction from opti-acoustic stereo imaging,”*Proc. 3DPVT*, Chapel Hill, NC, June, 2006.

[15] Y.Y. Schechner, and N. Karpel, “Clear underwater vision,” *Proc. IEEE CVPR*, Vol 1, 2004, pp. 536–543.

[16] R. L. Thompson, “Blazed array sonar systems - a new technology for creating low-cost, high-resolution imaging sonar systems for fisheries management, *Proc. Puget Sound Georgia Basin Res. Conf.*, 2005.

[17] R.F. Wagner, S.W. Smith, J.M. Sandrik, H. Lopez, “Statistics of speckle in ultrasound B-scans,” *IEEE T. Sonics and Ultras.*, vol 30, May 1983, pp. 156–163.

[18] http://www.codaoctopus.com/3d_ac_im/index.asp

[19] http://www.blueviewtech.com/

[20] http://www.soundmetrics.com/

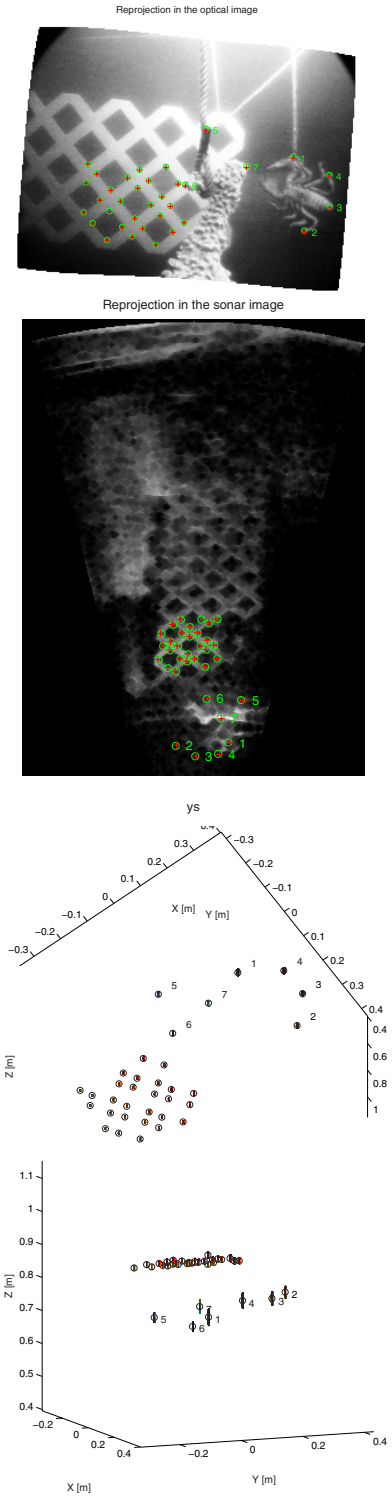


Figure 6. Stereo pairs with matched features for water-tank data set, and two views of the 3-D reconstructed object points.