Opti-Acoustic Stereo Imaging: On System Calibration and 3-D Target Reconstruction

Shahriar Negahdaripour, Hicham Sekkati, and Hamed Pirsiavash

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 two other closed-form solutions, referred to as the 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.

Index Terms—Sensor integration, stereovision, 3-D reconstruction, underwater sonar imaging.

I. 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 the visibility range in naturally illuminated underwater images has been demonstrated by polarization-based image analysis [19]. The method makes use of at least two images taken through a polarizer at different orientations (e.g., horizontal and vertical) for color correction and improved scene contrast. Advantages can also come from the simultaneous use of different and complementary sensors to exploit their unique strengths and properties, while overcoming the shortcomings and limitations of each sensing modality.

The authors are with the Electrical and Computer Engineering Department, University of Miami, Coral Gables, FL 33124-0620 USA (e-mail: shahriar@miami.edu; sekkati@miami.edu; sekkati@gmail.com; h.pirsiavash@umiami.edu; hpirsiav@ics.uci.edu).

Where visibility allows, potential integration of optical and acoustic information can enhance the performance in comparison with the processing of images from each sensor, alone. This multisensor fusion strategy has been explored for registering image data to known 3-D object models [5], 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 sonar cameras have emerged [17]; e.g., *D*ual-Frequency *ID*entification *SON*ar (DIDSON) [27] and BlueView based on blazed-array technology [26]. The video imagery from these systems provide high enough details to allow visual target recognition by human operators in search and inspection [3], [20]. In particular, many submersible platforms nowadays utilize a 2-D sonar camera for remote control in poor visibility, while an optical imaging system continues to serve as the navigation camera in clear waters. In other words, the deployment of both sonar and video cameras is becoming a standard practice to extend the utility and operation range of the subsea vehicles. This has motivated an investigation in deploying these multimodal cameras in stereo configuration—namely "*opti-acoustic stereo imaging*"—as a novel strategy for 3-D object reconstruction in underwater applications [13], [15]. This study addresses some fundamental problems: 1) establishing the epipolar geometry of opti-acoustic stereo images; 2) derivation of certain *closed-form solutions* that utilize the sonar and optical projection constraints of the two stereo views in various ways. Furthermore, the earlier computer simulations and our current experimental results point to improved 3-D reconstruction performance compared to triangulation in a traditional binocular system under a number of conditions: 1) small stereo baseline; 2) deteriorating visibility, where both optical cameras will suffer from inaccurate feature detection and localization. Furthermore, advantages over 3-D acoustic cameras include higher resolution and the availability of visual data for target recognition and classification.

Just as for optical systems, noisy "*opti-acoustic correspondences*" do not satisfy the epipolar geometry, and, therefore, 3-D reconstruction from any of the earlier closed-form methods is *sub-optimal* with respect to the maximum-likelihood estimates (MLE) that take advantage of redundant constraints. Two approaches based on direct and indirect estimation of 3-D target points from noisy observations were shown to produce comparable results [18]. Each method rests on the representation of range and bearing measurement noises by the Gaussian model.

This paper extends and improves on our earlier work on opti-acoustic stereo imaging system calibration and 3-D reconstruction [14], [18]. First, the previous stereo calibration method relies on determining the pose of each system relative
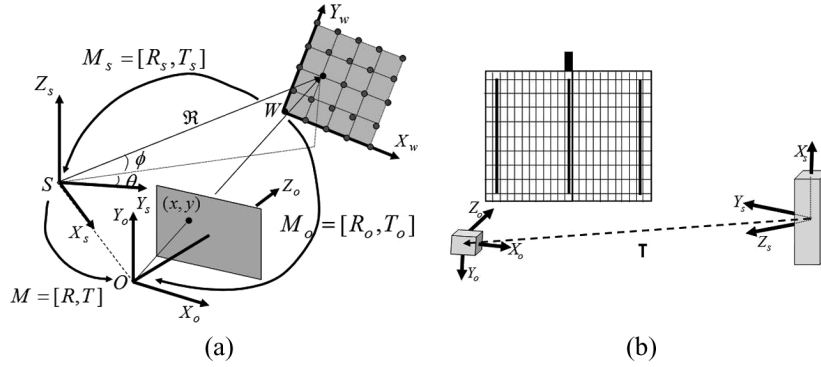
Fig. 1. (a) World, sonar, and optical coordinate systems and their transformations, and planar calibration grid with 2-D projection coordinates $(\Re, \theta)$ and (x, y) of a sample grid point in the acoustic and sonar cameras. (b) Typical geometry of opti-acoustic stereo cameras relative to world frame, aligned with axes of the planar calibration grid, in our real experiments.

to a planar calibration target. Being a critical step in establishing the epipolar geometry with high accuracy, we propose a new method by utilizing a minimum of five opti-acoustic correspondences to directly compute the relative pose of the two cameras. This gives a more accurate estimate compared to the previous indirect method. 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 range and bearing by the more suitable Rayleigh and uniform distributions, respectively [8], [23]. The nonlinear estimation problem is solved iteratively by application of the Levenberg–Marquardt algorithm [12]. Since a good initial condition enhances the performance of recursive schemes, we seek to improve on the closed-form range and azimuth solutions that were derived in previous work [14], [18]. By examining these solutions and their performances, we have devised as a new solution the weighted average of these two earlier ones. 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. By careful design of computer simulations, we establish performance baselines that are verified in experiments with real data collected under comparable conditions (in an indoor pool[1] and our own $6' \times 12' \times 6'$ water tank facility). Furthermore, the epipolar geometry, as well as methods for system calibration and for 3-D reconstruction, are tested with these real data. These experiments provide comparisons between the optic-acoustic and binocular stereo imaging systems.

Finally, it is noted that the effective application of our 3-D reconstruction methods during the course of an online underwater operation requires automatic robust and accurate matching of corresponding features in optical and sonar images. While we recently explored a geometric approach to the opti-acoustic correspondence problem [16], the results here are based on manually matched features because we are mainly concerned with

assessing the performance of calibration and 3-D reconstruction techniques.

The balance of the paper is organized as follows: Section II covers a review of sonar camera technology, and various background material on the projection model of an opti-acoustic system, the transformation between coordinate frames, the image measurements, and the stereo correspondence constraints. The epipolar constraint is described in Section III. We describe the new opti-acoustic stereo calibration method in Section IV. Section V is devoted to the 3-D reconstruction from opti-acoustic correspondences. Here, we study the closed-form range and azimuth solutions, devise the new weighted solution, and present the nonlinear optimization formulation that yields the MLE. Experiments with synthetic and real data are presented in Section VI, providing validation of proposed methods. Finally, we provide a summary of our contributions in Section VII.

II. PRELIMINARIES

A. Sonar Cameras

Sonar cameras produce an image by recording the reflected sound when the scene is insonified by acoustic pulses. In a 3-D sonar (e.g., Echoscope [25]), the back-scattered signals are collected by a 2-D array of transducers, and the image is formed from "beam signals," i.e., the echoes from *fixed steering directions*, specified by elevation ϕ and azimuth θ angles (see Fig. 1). The 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 a number of beams at varying bearings (azimuth angles), with range determined from the time of flight. In two existing technologies, namely, DIDSON and BlueView systems, a total of 512 range values are recorded within a fixed down-range (DR) window $[\Re_{\min} - \Re_{\max}]$ [m]. This is typically set according to object distances from the camera, thus establishing the down-range resolution. Formed with acoustic lenses and transducer curvature, DIDSON generates 96 beams with roughly $w_\theta = 0.3°$ azimuth and $w_\phi = 14°$ elevation widths. The transmitted beams cover a total field of view of $28.8°$ in the azimuth direction with $0.3°$ resolution; this translates to a cross-range (CR) of roughly

$0.5 * \text{DR}$ [m]. Built on blazed-array technology, the BlueView offers a larger 45-degrees cross-range field of view with $1°$ resolution. Treatment as a 2-D sonar is mainly because of the $\pm 7°$ uncertainty in the elevation of an imaged 3-D point. While our methods are applicable to any 2-D forward sector-scan sonar, we discuss the DIDSON video camera in the remainder of this paper, as our real data were acquired with this system.

The 2-D sonar cameras can produce high-quality images in turbid waters [2], [20]; 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)].

B. Rectangular and 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. Referring to Fig. 1, the relationship between rectangular and spherical coordinates and the inverse transformation are

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

C. Coordinate Frames and Transformation

Let $\boldsymbol{P}_o = (X_o, Y_o, Z_o)^T$ and $\boldsymbol{P}_s = (X_s, Y_s, Z_s)^T$ denote the coordinates of a 3-D world point $\mathbf{P} = (X, Y, Z)^T$ in the rectangular coordinate frames, denoted $\mathcal{O}$, $\mathcal{S}$, assigned to the optical and sonar cameras [see Fig. 1(a)]. These two coordinate systems are related to the world reference frame by a rigid body motion transformation, comprising a 3×3 rotational matrix and a 3-D translational vector $\mathbf{M_o} = [\mathbf{R_o}, \mathbf{T_o}]$ and $\mathbf{M_s} = [\mathbf{R_s}, \mathbf{T_s}]$. The relative pose of the optical and acoustic cameras can be expressed in terms the two transformations with respect to the world reference frame

$$\mathbf{M} = [\mathbf{R}, \mathbf{T}] = \left[\mathbf{R_s}\mathbf{R_o^{-1}}, \mathbf{T_s} - \mathbf{R_s}\mathbf{R_o^{-1}}\mathbf{T_o}\right] \tag{2}$$

where

$$\boldsymbol{P}_s = \boldsymbol{R}\boldsymbol{P}_o + \boldsymbol{T}. \tag{3}$$

D. 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 satisfies the perspective projection model

$$\begin{cases} x = f\frac{X_o}{Z_o} = f\frac{\boldsymbol{r}_{o1}^T\mathbf{P} + T_{ox}}{\boldsymbol{r}_{o3}^T\mathbf{P} + T_{oz}} \\ y = f\frac{Y_o}{Z_o} = f\frac{\boldsymbol{r}_{o2}^T\mathbf{P} + T_{oy}}{\boldsymbol{r}_{o3}^T\mathbf{P} + T_{oz}} \end{cases} \tag{4}$$

where $\mathbf{T_o} = [T_{ox}, T_{oy}, T_{oz}]$, $\boldsymbol{r}_{oi}$ ($i = 1 : 3$) denotes rows of $\mathbf{R_o}$, and f is the focal length of the optical camera. The range

and azimuth measurements of $\mathbf{P}$ in the acoustic image are given by

$$\begin{cases} \Re = \sqrt{X_s^2 + Y_s^2 + Z_s^2} \\ \quad = \sqrt{\left(\boldsymbol{r}_{s1}^T\mathbf{P} + T_{sx}\right)^2 + \left(\boldsymbol{r}_{s2}^T\mathbf{P} + T_{sy}\right)^2 + \left(\boldsymbol{r}_{s3}^T\mathbf{P} + T_{sz}\right)^2} \\ \theta = \tan^{-1}(X_s/Y_s) = \tan^{-1}\left(\left(\boldsymbol{r}_{s1}^T\mathbf{P} + T_{sx}\right)/\left(\boldsymbol{r}_{s2}^T\mathbf{P} + T_{sy}\right)\right) \end{cases} \tag{5}$$

where $\mathbf{T_s} = [T_{sx}, T_{sy}, T_{sz}]$, and $\boldsymbol{r}_{si}$ ($i = 1 : 3$) denote rows of $\mathbf{R_s}$. A rectangular sonar image with symmetric coordinate units $\boldsymbol{p}_s = (x_s, y_s)$ can be constructed based on the following transformation:

$$\boldsymbol{p}_s = \begin{bmatrix} x_s \\ y_s \end{bmatrix} = \Re \begin{bmatrix} \sin\theta \\ \cos\theta \end{bmatrix}. \tag{6}$$

It readily follows that

$$\Re = \sqrt{x_s^2 + y_s^2} \quad \text{and} \quad \theta = \tan^{-1}\frac{x_s}{y_s}. \tag{7}$$

E. Stereo Correspondence Constraint

The relationship between opti-acoustic correspondences $\boldsymbol{p}_o = (x, y, f)$ and $\boldsymbol{p}_s = (x_s, y_s)$ 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 (3), and can be expressed in the form

$$\begin{bmatrix} x_s \\ y_s \end{bmatrix} = (1/\cos\phi)\left(\left(\frac{Z_o}{f}\right)\begin{bmatrix} \boldsymbol{r}_1 \cdot \boldsymbol{p}_o \\ \boldsymbol{r}_2 \cdot \boldsymbol{p}_o \end{bmatrix} + \begin{bmatrix} T_x \\ T_y \end{bmatrix}\right) \tag{8}$$

where $\mathbf{r_i}$ denotes ith row of $\boldsymbol{R}$. The dependent unknown ϕ can be eliminated by noting that

$$\cos\phi = \sqrt{1 - \left(\frac{Z_s}{\Re}\right)^2} = \sqrt{1 - (Z_o/\Re(\boldsymbol{r}_3 \cdot \boldsymbol{p}_o/f) + T_z/\Re)^2}. \tag{9}$$

Finally, we arrive at

$$\boldsymbol{p}_s = \sqrt{\frac{1}{1 - (Z_o/\Re(\boldsymbol{r}_3 \cdot \boldsymbol{p}_o/f) + T_z/\Re)^2}} \\ * \left(Z_o\begin{bmatrix} \boldsymbol{r}_1 \cdot \boldsymbol{p}_o/f \\ \boldsymbol{r}_2 \cdot \boldsymbol{p}_o/f \end{bmatrix} + \begin{bmatrix} T_x \\ T_y \end{bmatrix}\right). \tag{10}$$

III. EPIPOLAR GEOMETRY

The epipolar geometry is fundamental to 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 detail in [15], it is useful for completeness to summarize some relevant results.

Referring back to (3), we can establish the coordinate system of one camera as the reference frame. Here, we select the optical camera coordinate system, without loss of generality. Next, we

rewrite the projection equations with the optical camera frame as as the world reference

$$\begin{cases} x = f\frac{X_o}{Z_o} \\ y = f\frac{Y_o}{Z_o} \\ \Re = \sqrt{\left(\boldsymbol{r}_1^T\mathbf{P_o} + T_x\right)^2 + \left(\boldsymbol{r}_2^T\mathbf{P_o} + T_y\right)^2 + \left(\boldsymbol{r}_3^T\mathbf{P_o} + T_z\right)^2} \\ \theta = \tan^{-1}\left(\left(\boldsymbol{r}_1^T\mathbf{P_o} + T_x\right)/\left(\boldsymbol{r}_2^T\mathbf{P_o} + T_y\right)\right) \end{cases}$$
$$(11)$$

The epipolar constraint in an opti-acoustic stereo system—establishing the relationship between projections $\boldsymbol{p}_s$ and $\boldsymbol{p}_o$ of the same scene point $\boldsymbol{P}$ [15]—is derived by manipulating (3) and (11)

$$\mathcal{C}(\boldsymbol{p}_o, \boldsymbol{p}_s) = \boldsymbol{p}_o^T \mathbf{U}(\boldsymbol{p}_s)\boldsymbol{p}_o = 0. \tag{12}$$

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

$$\mathbf{U}(\boldsymbol{p}_s) = (x_s\tilde{T}_y - y_s\tilde{T}_x)^2\mathbf{I} + \left(\|\tilde{\mathbf{T}}\|^2 - 1\right)\mathbf{a}\mathbf{a}^T$$
$$+ (x_s\tilde{T}_y - y_s\tilde{T}_x)(\mathbf{a}\tilde{\mathbf{T}}^T\mathbf{R} + \mathbf{R}^T\tilde{\mathbf{T}}\mathbf{a}^T) \quad (13)$$

where $\mathbf{a} = (y_s\mathbf{r_1} - x_s\mathbf{r_2})^T$ and $\tilde{\mathbf{T}} = [\tilde{T}_x, \tilde{T}_y, \tilde{T}_z] = \boldsymbol{T}/\Re$. As noted, the match $\boldsymbol{p}_o$ in the optical image of an acoustic-image point $\boldsymbol{p}_s$ lies on a conic section. It often becomes necessary to establish the match $\boldsymbol{p}_s$ of an optical image point $\boldsymbol{p}_o$. It has also been shown that the epipolar geometry in the acoustic camera satisfies the following constraint [15]:

$$\Re = \sqrt{N_\theta/D_\theta} \tag{14}$$

where

$$N_\Re(\theta) = (\boldsymbol{\theta}^T\Upsilon\boldsymbol{T})^2 + (\tilde{z}^T\Upsilon\boldsymbol{T})^2 \quad D_\Re(\theta) = (\tilde{z}^T\Upsilon\boldsymbol{\theta})^2. \tag{15}$$

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 $\boldsymbol{v} = \boldsymbol{R}\boldsymbol{p}_o$ (such that $\Upsilon\boldsymbol{x} = \boldsymbol{v} \times \boldsymbol{x}$, for any vector $\boldsymbol{x}$).

A. Measurement Noise Model

Image positions are typically noisy observations of the true 2-D projections $\{\boldsymbol{p}_o, \boldsymbol{p}_s\}$, 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, but the impacts can be rather different. To explain, calibration involves use of known targets that are most suitable for feature detection and matching, and can be carried out in ideal visibility conditions. Furthermore, corresponding features can be matched manually to avoid gross mismatches. Measurement noise of the calibration process may be modeled as additive Gaussian:

$$\begin{cases} \hat{x} = x + n(0, \sigma_{xo}) \\ \hat{y} = y + n(0, \sigma_{yo}) \end{cases} \quad \begin{cases} \hat{x}_s = x_s + n(0, \sigma_{xs}) \\ \hat{y}_s = y_s + n(0, \sigma_{ys}) \end{cases} \tag{16}$$

where $n(0, \sigma_i)$ is a normal distribution with zero mean and variance σ_i $(i = xo, yo, xs, ys)$. The independent Gaussian model of the uncertainties in x_s and y_s translates to Rayleigh and uniform distributions for range and azimuth uncertainties in agreement with the model of the speckle noise [8], [23], which is assessed to be the dominant effect in high-frequency short-range

sonar imaging systems [21]. Thus, an advantage in working with $\boldsymbol{p}_s = (x_s, y_s)$ to represent the sonar image coordinate of a 3-D point $\boldsymbol{P}_s$ is that the image noise is suitably modeled as Gaussian.

In contrast to the calibration process, 3-D reconstruction makes use of features that are detected and matched automatically under a variety of environmental conditions, and thus outliers are inevitable. In our work, manual opti-acoustic matching allows us to avoid outliers, putting the emphasis on assessing the 3-D reconstruction accuracy of inliers. One should note that the reconstruction of each 3-D point depends only on the 2-D optical and acoustic projections of the same point (it does not depend on the information from any other point). Therefore, an outlier match affects only the accuracy of the corresponding 3-D point. As it becomes apparent, the assumed variances, representing the measurement uncertainties, weight the appropriate terms in the optimization algorithm. Therefore, one can study analytically the estimation degradation by increasing the measurement variances.

IV. CALIBRATION OF OPTI-ACOUSTIC SYSTEM

As for optical images, imperfections in an acoustic lens lead to image distortions and geometric deviations from the 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 [14].

The relative pose of the optical and acoustic 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.

The points on a plane satisfy the relationship

$$\frac{f}{Z_o} = -(\boldsymbol{n}_o \cdot \boldsymbol{p}_o) \tag{17}$$

where $\boldsymbol{n}_o = (n_{ox}, n_{oy}, n_{oz})^T$ 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 $\boldsymbol{n}_s = (n_{sx}, n_{sy}, n_{sz})^T$ in the sonar coordinate system. This is given by

$$\boldsymbol{n}_s = \frac{\boldsymbol{R}\boldsymbol{n}_o}{1 - \boldsymbol{T}^T\boldsymbol{R}\boldsymbol{n}_o}. \tag{18}$$

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 ways. We use the decomposition into three rotations about the axes of the coordinate system: $\boldsymbol{R}(\alpha_x, \alpha_y, \alpha_z) = \boldsymbol{R}_z(\alpha_z)\boldsymbol{R}_y(\alpha_y)\boldsymbol{R}_x(\alpha_x)$ where $\boldsymbol{R}_u(\alpha_u)$ denotes a rotation about axis u of the respective coordinate system by angle α_u. Substituting for Z_o from (17) into (10), each match provides two constraints as given in (10), in terms of nine unknowns: the six pose parameters $\boldsymbol{M} = [\boldsymbol{R}(\alpha_x, \alpha_y, \alpha_z), \boldsymbol{T}]$ and three parameters of the normal $\boldsymbol{n}_o$ of the calibration target

plane in the optical camera coordinate frame. We have redundancy with $N \geq 5$ correspondences. We can solve a nonlinear 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 acoustic projections. Assume an estimate $\hat{\boldsymbol{M}} = [\hat{\boldsymbol{R}}, \hat{\boldsymbol{T}}]$ and $\hat{\boldsymbol{n}}_o$ of the 9 sought after parameters, to initialize the nonlinear optimization algorithm. These are updated during each step of an iterative estimation process. For a feature $\hat{\boldsymbol{p}}_o$ in the optical image, we estimate the depth $\hat{Z}_o$ from the plane equation in (17). Computing the other two coordinates $\hat{X}_o$ and $\hat{Y}_o$ from (4), we have an estimate of the 3-D point $\boldsymbol{P}_o$. Utilizing (3), transformation to the sonar coordinate system with $\hat{\boldsymbol{M}} = [\hat{\boldsymbol{R}}, \hat{\boldsymbol{T}}]$ gives us the estimated position $\hat{\boldsymbol{P}}_s$. Next, we calculate for the acoustic match $\hat{\boldsymbol{p}}_s$ the elevation angle

$$\phi = -\gamma + \sin^{-1}\left(\frac{-1}{\sqrt{(\hat{n}_{sx}x_s + \hat{n}_{sy}y_s)^2 + \Re^2\hat{n}_{sz}^2}}\right) \quad (19)$$

where

$$\gamma = \tan^{-1}\left(\frac{\hat{n}_{sx}x_s + \hat{n}_{sy}y_s}{\Re\hat{n}_{sz}}\right). \quad (20)$$

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

$$e(\boldsymbol{M}, \boldsymbol{n}_o) = \sum_{i=1}^{n}(\boldsymbol{P}_o - \hat{\boldsymbol{P}}_o)^T\Sigma_{Po}^{-1}(\boldsymbol{P}_o - \hat{\boldsymbol{P}}_o)$$
$$+ \sum_{i=1}^{n}(\boldsymbol{P}_s - \hat{\boldsymbol{P}}_s)^T\Sigma_{Ps}^{-1}(\boldsymbol{P}_s - \hat{\boldsymbol{P}}_s). \quad (21)$$

The covariances Σ_{Po} and Σ_{Ps}—not given here because of the complexities of their expressions—are estimated by first-order approximation from the Jacobian of the solution with respect to the image projections and their uncertainties, and covariance of the estimation parameters. The nonlinear optimization problem in (21) has been solved by the Levenberg–Marquardt algorithm [12].

V. 3-D RECONSTRUCTION

Given an opti-acoustic correspondence, the corresponding 3-D point can be calculated in closed form [15]. However, an optimal solution in the maximum likelihood (ML) sense is derived from a nonlinear method that requires the application of iterative algorithms. Two such methods have been proposed based on indirect and direct formulations, assuming a Gaussian model for representing both the range and azimuth measurement uncertainties [18]. Both approaches provide comparable accuracy based on computer simulations with data corrupted by additive Gaussian noise. However, neither gives optimal estimates with real data, as the noise of an acoustic imaging system is not in agreement with the Gaussian model. Here, the direct method is revisited, reformulating the optimization problem by representing the uncertainty in the sonar image position $\boldsymbol{p}_s$ by the Gaussian model. As stated earlier, this corresponds to

the Rayleigh model for range measurements, as reported for acoustic image noise representation [8], [23].

Good initial estimates improve the convergence rate of recursive estimation methods. Examining the performance of the range and azimuth closed-form solutions proposed in [15], 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.

A. Closed-Form Solutions

In the context of opti-acoustic stereo imaging, stereo triangulation deals with determining the 3-D point $\boldsymbol{P} = (X, Y, Z)$—or equivalently, the position in either camera reference frames, say $\boldsymbol{P}_o = (X_o, Y_o, Z_o)$—for any opti-acoustic correspondence $\boldsymbol{p}_o = (x, y, f)$ and $\boldsymbol{p}_s = (x_s, y_s)$. The asymmetrical form of the optical and acoustic projection models in (4) and (5) 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

$$\left(\left\|\frac{\mathbf{p}}{f}\right\|^2\right)Z_{\Re}^2 + \frac{2}{f}(\mathbf{T}^T\mathbf{R}\mathbf{p})Z_{\Re} - \left(\Re^2 - \|\mathbf{T}\|^2\right) = 0. \quad (22)$$

The correct solution must agree with the *azimuth solution*—the intersection of the optical ray with the azimuth plane

$$Z_\theta = f\frac{\tan\theta T_y - T_x}{(\mathbf{r_1} - \tan\theta\mathbf{r_2})\cdot\mathbf{p}}. \quad (23)$$

It should be noted that while these closed-form expressions yield the exact solution with perfect data, each gives a different estimate with noisy observations. In particular, neither solution is optimal in the maximum likelihood (ML) sense. In the next section, we present a formulation that directly addresses this shortcoming, however, it requires an iterative scheme to solve a nonlinear optimization problem. In this case, our closed-form solutions can provide the initial guess for the iterative MLE method. Clearly, identifying the closed-form solution with the best estimate with noisy data enhances the convergence of the recursive method.

One approach is the fusion of the two earlier solutions by weighted averaging

$$Z_m = \xi_t Z_\theta + (1 - \xi_t)Z_{\Re} \quad (24)$$

where the weight ξ_t is chosen to be a sigmoid function, and the transition 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 and how these depend on the imaging and environmental factors.

We have analytically derived and compared the first-order approximation to the variance of each solution, seeking to identify the relevant parameters. This analysis and extensive computer simulations with the range and azimuth solutions have revealed that the *target distance* and the *stereo baseline* are two critical factors. Luckily we know both quantities from the stereo system geometry and range measurements of the sonar camera. To demonstrate, we examine the inverse normalized error,
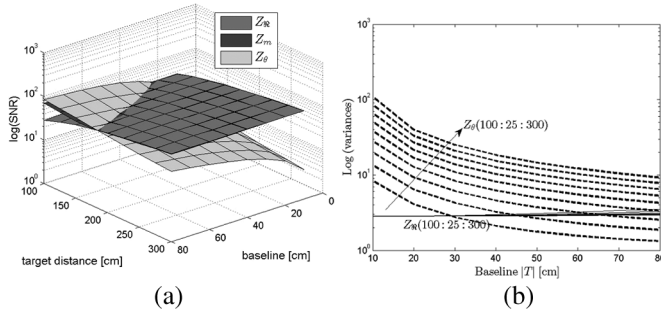
Fig. 2. (a) Average Reconstruction SNR over 36 symmetrically distributed points on a planar surface. (b) Error variances of azimuth and range solutions for varying baseline and range as 2-D iso-distance contours, allowing the selection of proper scaling to compute the weighted average solution Z_m.
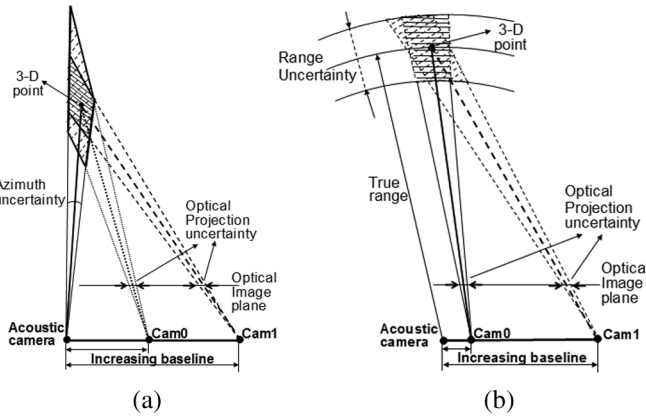

Fig. 3. Geometrical interpretation of sensitivities of closed-form (a) azimuth and (b) range solutions to size of baseline. Reconstruction uncertainty regions due to noisy optical image projection, and range and azimuth angle measurements from sonar camera are highlighted by dotted lines.
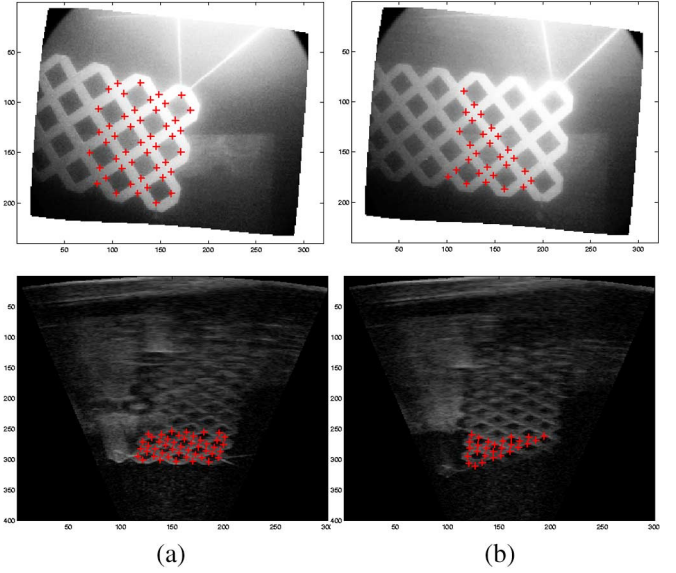

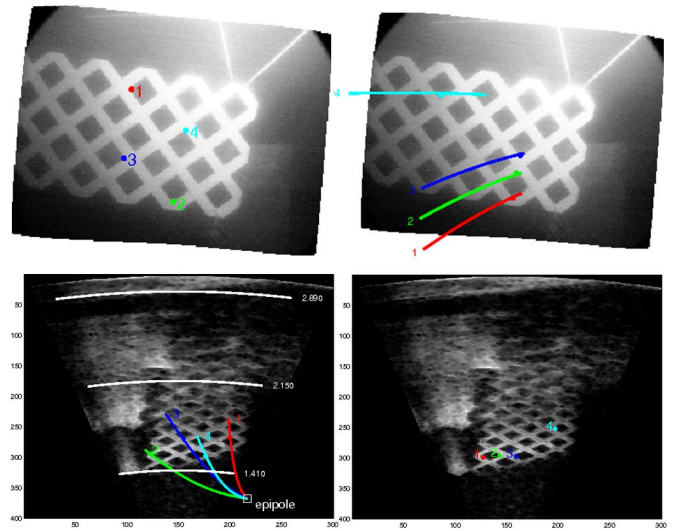Fig. 4. Matching features in stereo pairs for two data sets.

Fig. 5. Stereo calibration establishes epipolar geometry. For each feature in one image, corresponding epipolar contour in other view is computed from (12) and (14).

defined as the ratio of the true target distance Z_0 along the viewing direction of the optical camera to the mean reconstruction error (SNR of these solutions). For data, we utilize points on a plane viewed relative to the opti-acoustic stereo in the configuration shown in Fig. 1. We test with different baselines $T_x = (10, 20, \ldots, 80)$ [cm] and plane distances $Z_0 = (100, 125, \ldots, 300)$ [cm]. The noise variances are fixed to $\sigma_x = \sigma_y = \sigma_{xs} = \sigma_{ys} = 1$ [pix] for both optical and acoustic cameras. Finally, we compute the average reconstruction SNR over 36 uniformly distributed points on the plane, thus minimizing the effect of undesirable viewing angles for certain scene points.

Fig. 2(a) shows the reconstruction SNRs for the range and azimuth solutions, computed experimentally by ensemble averaging over hundreds of simulations with different noisy samples from the assumed distribution. As stated, these plots are in full agreement with analytical SNRs that are computed from the first-order approximation of variances [18]. Important conclusions are:

- the range solution is relatively insensitive to baseline size, while azimuth solution, as for traditional optical systems, improves with increased baseline;
- the range solution SNR improves with target distance.

These can be verified by the geometric interpretation of the estimation uncertainty, depicted in Fig. 3. Let us assume a fixed noise level in the range and azimuth measurements. The reconstruction uncertainty region shrinks with increasing baseline for the azimuth solution, but remains relatively constant for the range solution. Furthermore, it can be readily verified graphically that the reconstruction error of the range solution remains steady with increasing target range; thus, the SNR of the range solution increase with target distance. The same information is conveyed in Fig. 2(b) in the form of 2-D iso-distance contour plots.

We are now ready to formalize these findings by assigning the weighting factor in (24). Generally speaking, the azimuth solution is weighted more heavily for larger baselines, while the range solution would contribute more heavily for larger target
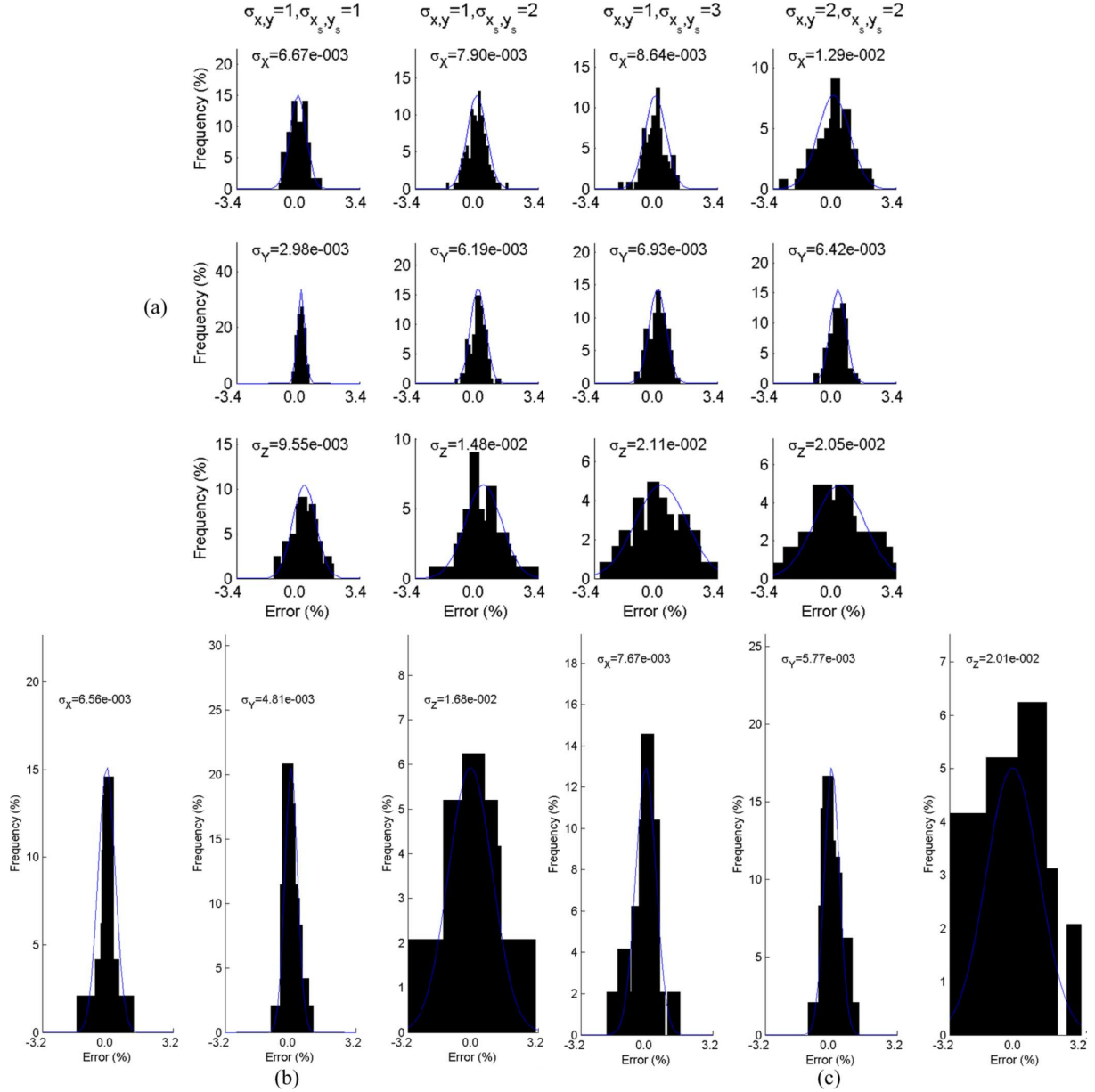
Fig. 6. (a) Computer simulations of 3-D reconstruction error, assuming the same imaging and target configurations as for real data. Experiments are for various optical and acoustic noise levels $\sigma_{x,y}$ and $\sigma_{xs,ys}$, to check consistency with results for real data (b), (c), and to examine degradation rate with increasing noise level. For real data, reconstruction is inferior when representing range and bearing noise as Gaussian (c), according to model in [18].

distances. Defining the weight in terms of the ratio of the baseline to the target distance can serves this objective. We have selected the sigmoidal form

$$\xi_t = \left(1 + e^{-\left(\|\boldsymbol{T}\|/\bar{Z} - k_o\right)}\right)^{-1} \qquad (25)$$

where $\bar{Z} = (Z_{\Re} + Z_\theta)/2$. The threshold $k_o = \|\boldsymbol{T}\|/Z_c$ is set by determining, for a given stereo baseline, the *critical depth* Z_c where the depth and azimuth solutions have equal variances [see Fig. 2(b)]. This threshold can be precalculated and stored in a lookup table. As one can verify from Fig. 2(a), the weighted closed-form solution Z_m closely coincides with the range and azimuth solutions, where either has the better performance.

B. Maximum Likelihood Formulation

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

$$\text{Minimize } \mathcal{E}(\boldsymbol{P}_o) = (X - \hat{X})^T \Sigma^{-1}(X - \hat{X}) \qquad (26)$$

where $\Sigma = E[(X - \hat{X})(X - \hat{X})^T]$. Here, we utilize the projection model in (11) with 3-D points expressed in the optical

camera coordinate frame. It is reasonable to assume independency among components of the measured vector $\hat{X}$. This allows us to write Σ as a diagonal matrix with elements σ_i, and thus we can write

$$\mathcal{E}(\boldsymbol{P}_o) = \frac{(x - \hat{x})^2}{\sigma_x^2} + \frac{(y - \hat{y})^2}{\sigma_y^2} + \frac{(x_s - \hat{x}_s)^2}{\sigma_{x_s}^2} + \frac{(y_s - \hat{y}_s)^2}{\sigma_{y_s}^2}. \tag{27}$$

This nonlinear least-squares optimization problem is efficiently solved using the Levenberg–Marquardt algorithm [12].

C. First-Order Error Analysis

The reconstruction accuracy may be assessed by the covariance of the estimate $\boldsymbol{P} = (X, Y, Z)^T$. This can be expressed in terms of the covariance of the measurements $\mathbf{V} = (x, y, x_s, y_s)$

$$\Sigma_{\boldsymbol{P}} = \mathrm{E}\left\{(\boldsymbol{P} - \mathrm{E}\{\boldsymbol{P}\})(\boldsymbol{P} - \mathrm{E}\{\boldsymbol{P}\})^T\right\} = J\Sigma_{\mathrm{V}}J' \tag{28}$$

where Σ_{V} is the covariance of the observation vector, and the Jacobian $J = \partial\boldsymbol{P}/\partial\mathbf{V}$ can be determined analytically or estimated numerically. While this approximation may not always provide a tight bound on the estimation error, it is relatively accurate for nondegenerate scene geometries and positions with respect to the stereo system (as verified in our computer simulations).

VI. EXPERIMENTS

A. Calibration

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

The calibration method described in Section IV has been applied to each of the two opti-acoustic stereo pairs, depicted in the columns of Fig. 4; superimposed are the correspondences. For any feature in one image, we can determine whether the corresponding epipolar contour passes through the matching feature in the other view. For the data in column (b), this has been verified for four sample points in each image and the epipolar curve in the other stereo view, computed from (12) and (14) (see Fig. 5).

B. 3-D Reconstruction

To assess the 3-D reconstruction performance, we start with computer simulations with noisy data (perfect data that is corrupted with noise from a known distribution). This is to gain insight into the performance of an opti-acoustic stereo imaging system under various conditions. We then present results from other experiments with real data sets.

C. Computer Simulations

Computer simulations allow us to assess the performance of the 3-D reconstruction method under a variety of conditions.
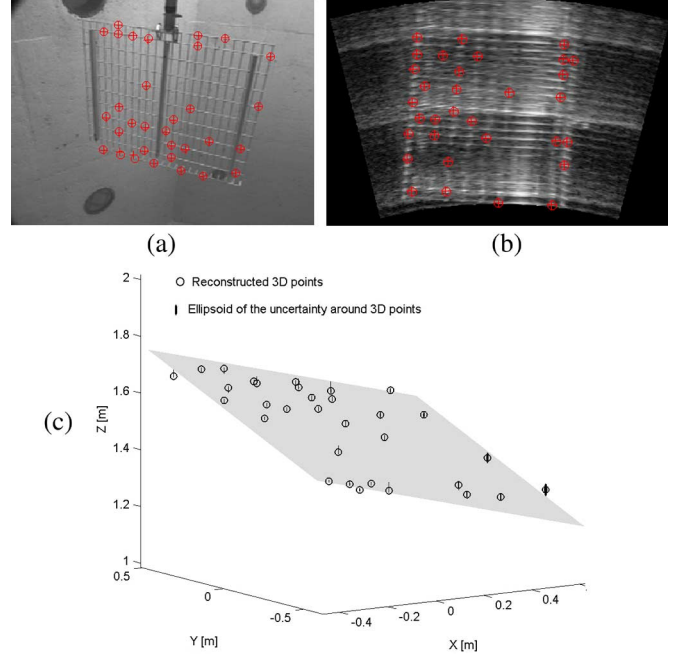

Fig. 7. Sample opti-acoustic stereo pair, where circles depict matching points in (a), (b) used for 3-D reconstruction, while crosses are projections of 3-D reconstructed points. (c) 3-D reconstructed points and the estimated planar surface during *a priori* stereo calibration.

Specifically, the goals are: 1) to examine the agreement between the numerical simulations with ground truth data (corrupted with additive noise) and results from experiments with real data, and 2) to quantify the degradation in performance for higher noise levels. To meet the first objective, computer simulations have been carried out with the camera and target geometries that are determined from the calibration for the first real data (see next section). For optical measurement noise, we have assumed zero-mean Gaussian noise with variances of $\sigma_{x,y} = 1$, 2 [pix]. For the acoustic images, we have tested with variances $\sigma_{xs,ys} = 1, 2, 3$ [pix]. The MLE method is applied to determine the estimates for both the synthetic and real data.

Fig. 6(a) depicts the 3-D reconstruction error distributions for a number of points on the planar grid. Among these, the cases $\sigma_{x,y} = 1$ [pix] and $\sigma_{xs,ys} = 2$ [pix] are most comparable with the results for real data; see (b). These uncertainty levels are more or less consistent with the quality of the real data and the accuracy of feature positions. The results for the larger noise levels of $\sigma_{x,y} = 1, 2$ and $\sigma_{xs,ys} = 2, 3$ [pix] simulate less favorable conditions of optical and acoustic visibility and clutter. The noted graceful degradation of performance is the desired behavior. Finally, the results in (c) are obtained by applying the MLE algorithm while representing the range and azimuth measurement uncertainties by the Gaussian model [18]. Not only is the performance inferior compared to (b), the reconstruction errors deviate from the Gaussian distribution.

Overall, the reconstruction errors are much smaller for the X and Y components 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, the 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; see Fig. 1(a).
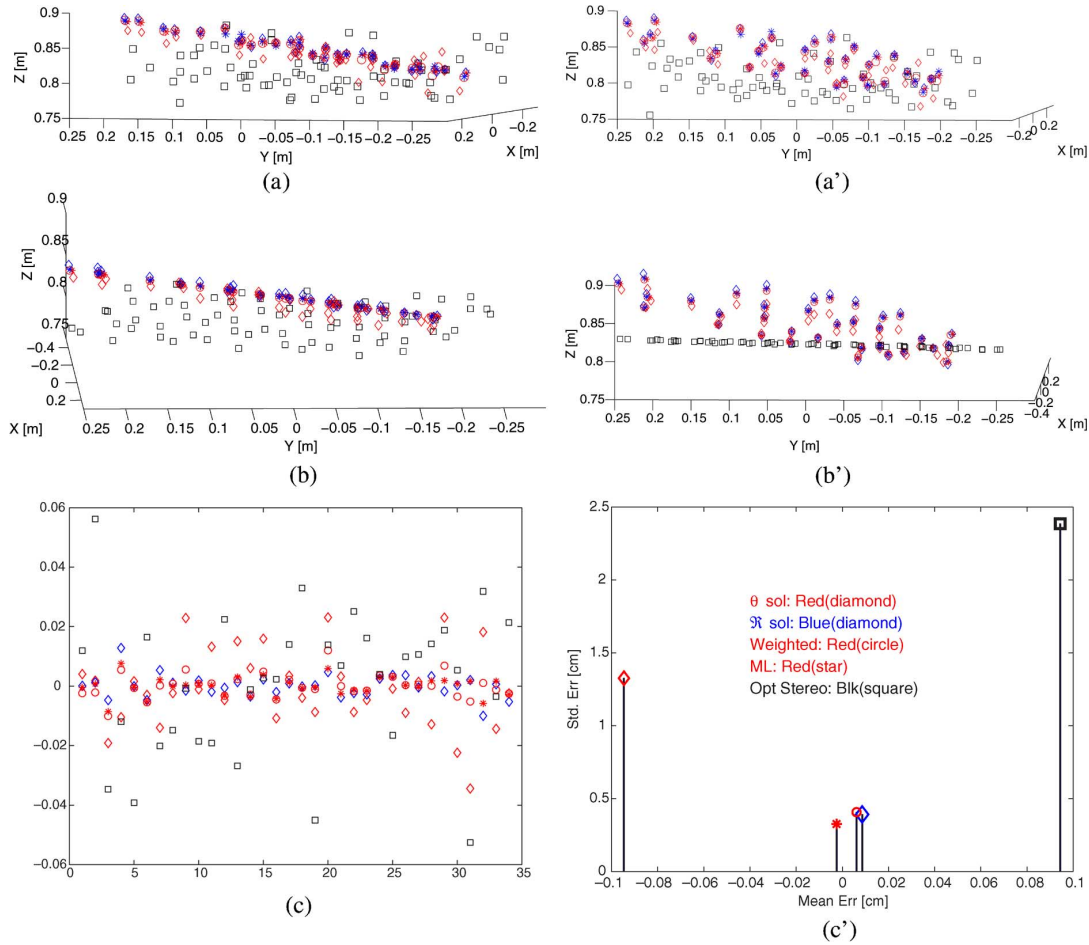
Fig. 8. Comparison of 3-D reconstruction for traditional binocular optical and opti-acoustic stereo imaging. Opti-acoustic stereo pair is given in column (b) of Fig. 4. (a,a') Two views of reconstructed 3-D point to examine how well the reconstructed points lie on a plane. (b,b') Two views of best fitting planes for various reconstruction methods. Perpendicular distances of 3-D points from corresponding best fitting plane (c), with their means and variances (c').

D. Real Data

The first data set is comprised of the stereo views of a planar metallic grid, collected in an indoor pool; see Fig. 7(a) and (b). 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 acoustic camera. Acoustic images have been corrupted by multiple returns from the water surface and various pool walls; see Fig. 7(b-b"). Each stereo pair depicts the matched features (circles) and the projections of the reconstructed 3-D points (crosses).

While we do know the ground truth, we use as the estimation error the distance of each reconstructed point from the plane. For the plane, we have used the estimate from the calibration process, which is independent of the reconstruction 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. The reconstructed points and uncertainty ellipsoids, as well as the plane of the grid computed by our calibration algorithm have been given in (c). Referring back to Fig. 6(b), the estimated 3-D points are within 3.5% of their distances to the

optical cameras (utilizing the plane reconstruction from calibration as the ground truth).

The next experiment utilizes the pair in Fig. 4(a), collected in a water tank with better acoustic insulation. In addition, we have acquired a third image with an optical camera, allowing us to determine the scene geometry independently from binocular disparity cues (image not shown here due to space limitations). A plastic planar grid is placed at an average distance of about 0.85 [m] along the viewing direction (Z_o axis) of the optical camera, which is positioned at a distance of about 1.22 m to the left of the acoustic camera. The binocular stereo baseline is 13.6 [cm]. Here again, the 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.

Fig. 8 depicts various results. In (a,a'), we have given two views of various 3-D estimates: closed-form range, azimuth and weighted solutions, ML method, the estimate from calibration, as well as binocular disparity; the legend for various symbols is given in (c'). The two chosen views are best for examining how well the reconstructed points lie on a plane, one for all of the estimates from various opti-acoustic methods and one for the optical stereo system. From these plots, we conclude that the estimates from various opti-acoustic methods are relatively consistent, and rather different from those based on the optical
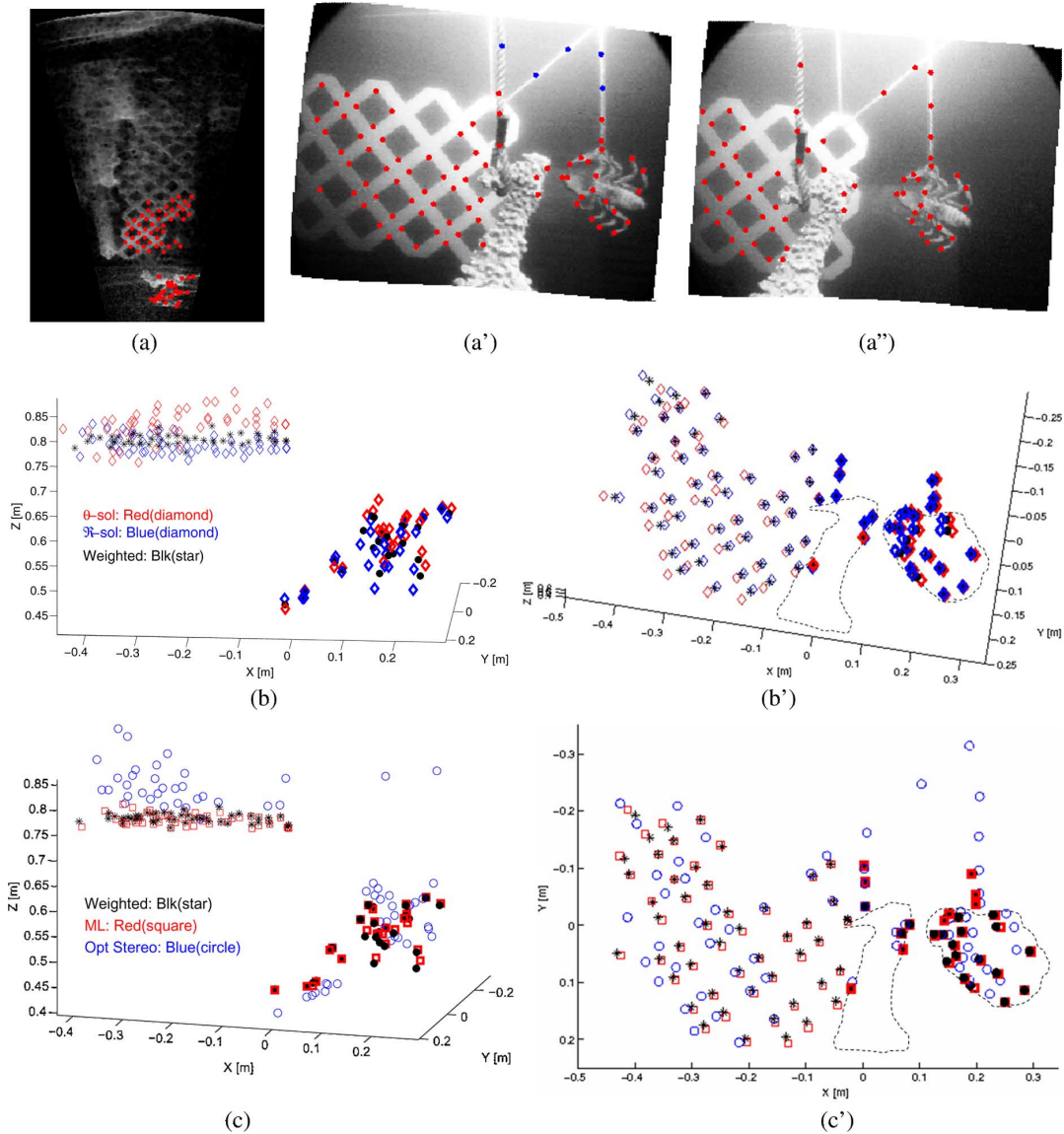
Fig. 9. Comparison of 3-D reconstruction for traditional binocular optical and opti-acoustic stereo imaging. (a,a',a") data set with correspondences; (b,b') 2 views of reconstructed 3-D points based on azimuth, range, and weighted solutions. (c,c',c") 3 views of reconstructions for weighted and ML solutions and traditional binocular optical stereo.

stereo system (which also exhibit the largest variations). The tilt of the plane, by the visual inspection of the optical image, appears to agree with the opti-acoustic estimate. To verify further, we have computed for each method the best plane passing through the estimated 3-D points, and again displayed from two different views in (b,b'). For each method, we have computed, and depicted in (c), the distances of all of the 3-D points from the best plane. The mean and variances of these distances in (c') confirm the earlier conclusions. A notable observation is that the discrepancies between the opti-acoustic and binocular stereo estimates are smaller for the points closer to the optical cameras, where the latter is expected to provide more accurate results.

The last experiment also utilizes both opti-acoustic and binocular optical stereo data [see Fig. 9(a,a',a")]. The scene is comprised of an artificial coral reef and a plastic toy lobster, each hung on a rope in front of the planar grid. The optical stereo images are horizontally rectified, and the correspondences were established manually. These are comprised of

points also imaged in the sonar view (red), with five additional matching points on the supporting ropes (blue). The opti-acoustic correspondences, for points on the lobster and the reef, were established by employing the epipolar curves.

Fig. 9(b,b') depicts two views of the reconstructed points based on the azimuth, range and weighted average solutions. The structure of the planar grid is emphasized in the near-top view, while the near-frontal view allows the comparison among solutions for each of the three objects. The solutions are consistent for nearby objects (reef and lobster). For the more distant grid points, the range and azimuth solutions deviate and the relative weight from (25) favors the former estimate that is expected to be more accurate for larger ranges. In (c,c'), two views of the reconstructions based on the weighted solution, ML estimate, and binocular disparity method are compared. The binocular stereo system seems to give better results for the nearby objects, namely the lobster and the lobster, but becomes less accurate for the distant planar grid.
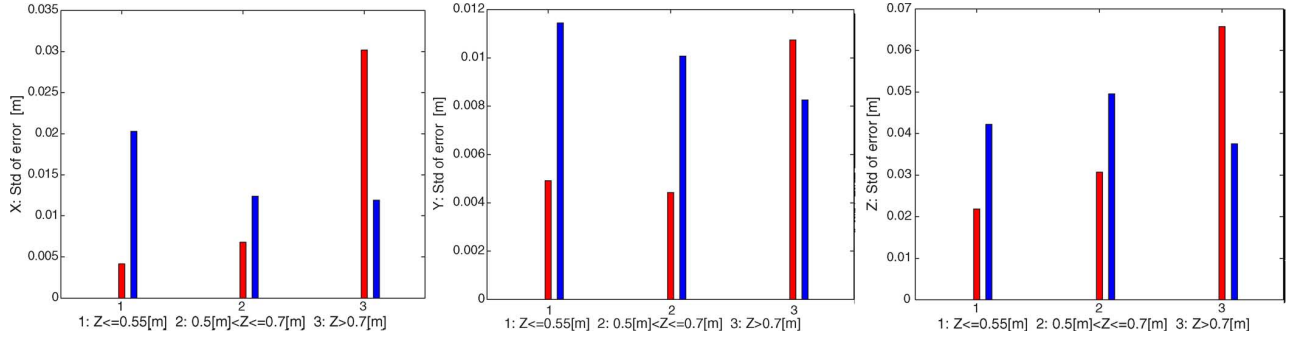
Fig. 10. Standard deviation of X, Y and Z components of the solution for three groups of points at distances $Z \leq 0.55$ [m], 0.55 [m] $< Z \leq 0.7$ [m] and $Z > 0.7$ [m] from optical camera based on computer simulations with noisy samples from Gaussian distributions $\sigma_x = \sigma_y = 1$ and $\sigma_{xs} = \sigma_{ys} = 2$ [pix]. (Red: Binocular Optical. Blue: Opti-acoustic).

The binocular stereo system seems to give better results for the nearby objects, namely the lobster and the reef, but becomes less accurate for the distant planar grid. To verify, we have examined the deviation in the solution with additive noise, taking the reconstructed geometry as ground truth. Furthermore, points were pooled into 3 groups based on their distances from the optical camera: $Z \leq 0.55$ [m], 0.55 [m] $< Z \leq 0.7$ [m] and $Z > 0.7$ [m]. Fig. 10 depicts the standard variation of the solution for the X, Y and Z components of estimated 3-D points, determined from 200 noisy samples from Gaussian distributions $\sigma_x = \sigma_y = 1$ and $\sigma_{xs} = \sigma_{ys} = 2$ [pix]. These results agree with our conclusions from the reconstructions in Fig. 9, as well as with the theoretical analysis of the relative reconstruction accuracy of opti-acoustic and binocular stereo systems, reported in earlier work [15]. Assuming a fixed feature localization uncertainty over the imge, of special importance as depicted in these plots is the fact that the opti-acoustic stereo estimation accuracy improves with increasing target distance.

VII. SUMMARY AND CONCLUSION

We have studied the 3-D reconstruction of underwater objects by opti-acoustic stereo imaging—a paradigm to integrate information from optical and acoustic camera systems with overlapping views [15]. We have proposed and analyzed methods for system calibration and target scene reconstruction. The calibration technique employs a minimum of five correspondences from features on a planar grid to compute the relative pose of the stereo cameras, but many more points over a large field of view in one or more stereo pairs are necessary to obtain accurate results.

The asymmetrical nature of the optical and acoustic 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.

The results from experiments with ground-truth synthetic data serve to assess the performance of our 3-D reconstruction method, and are verified with real data acquired under assumed imaging geometry. That is, our experiments with real data support the theoretical findings in reference to the performance of various opti-acoustic reconstruction methods, and in comparison to a binocular optical system for close-range and distant targets. Overall, we foresee good potential in the application of this novel paradigm for underwater 3-D object reconstruction in a wider range of environmental conditions. Complementing this work is a geometric solution to the opti-acoustic correspondence problem, aimed at devising a robust automatic stereo matching method [16]. This is a critical component of bringing to bear a complete computer system for the 3-D reconstruction of underwater objects.

REFERENCES

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

[2] E. O. Belcher, W. L. J. Fox, and W. H. Hanot, "Dual-frequency acoustic camera: A candidate for an obstacle avoidance, gap-filter, and identification sensor for untethered underwater vehicles," in *Proc. MTS/IEEE OCEANS*, 2002, vol. 4, 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," presented at the Oceans, San Diego, CA, Sep. 2003.

[4] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, pp. 381–395, 1981.

[5] A. Fusiello and V. Murino, "Augmented scene modeling and visualization by optical and acoustic sensor integration," *IEEE Trans. Vis. Comput. Graph.*, vol. 10, no. 5, pp. 625–636, Nov.–Dec. 2004.

[6] R. K. Hanson and P. A. Anderson, "A 3-D underwater acoustic camera—Properties and application," in *Acoustic Imaging*, P. Tortoli and L. Masotti, Eds.   New York: Plenum, 1996, pp. 607–611.

[7] J. Heikkilä and O. Silvén, "A four-step camera calibration procedure with implicit image correction," presented at the CVPR, 1997.

[8] E. Jakeman and R. J. Tough, "Generalized K distribution: A statistical model for weak scattering," *J. Opt. Soc. Amer.*, vol. 4, pp. 1764–1772, Sep. 1987.

[9] K. Kim, N. Neretti, and N. Intrator, "Non-iterative construction of super-resolution image from an acoustic camera video sequence," in *Proc. CIHSPS*, 2005, pp. 105–111.

[10] D. M. Kocak and F. M. Caimi, "The current art of underwater imaging with a glimpse of the past and vision of the future," *Marine Tech. Soc. J.*, vol. 39, no. 3, pp. 5–26, 2005.

[11] S. Kullback, *Information Theory and Statistics*.   New York: Dover, 1968, edition.

[12] D. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *J. SIAM*, vol. 11, no. 2, 1963.

[13] S. Negahdaripour, "Theoretical foundations for opti-acoustic stereo imaging," presented at the IEEE Oceans, Brest, France, 2005.

[14] S. Negahdaripour, "Calibration of DIDSON forward-scan acoustic video camera," presented at the IEEE/MTS Oceans, Washington, DC, Aug. 2005.

[15] S. Negahdaripour, "Epipolar geometry of optic-acoustic stereo imaging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1776–1788, Oct. 2007.

[16] S. Negahdaripour, H. Pirsiavash, and A. Taatian, "A geometric approach to solving the correspondence problem in opti-acoustic stereo imaging," presented at the Symp. UUST, Durham, NC, Aug. 2007.

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

[18] H. Sekkati and S. Negahdaripour, "Direct and indirect 3-D reconstruction from opti-acoustic stereo imaging," presented at the 3DPVT, Chappel Hill, NC, Jun. 2006.

[19] Y. Y. Schechner and N. Karpel, "Clear underwater vision," in *Proc. IEEE CVPR*, 2004, vol. 1, pp. 536–543.

[20] R. L. Thompson, "Blazed array sonar systems—A new technology for creating low-cost, high-resolution imaging sonar systems for fisheries management," presented at the Puget Sound Georgia Basin Res. Conf., 2005.

[21] A. Trucco, *Speckle Statistics.* Jan. 2007, Working paper (personal communication), Univeristy of Genoa, Italy.

[22] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," *IEEE J. R&A*, vol. 3, no. 4, Aug. 1987.

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

[24] X. Zhuang and R. M. Haralick, "A highly robust estimator for computer vision," in *Proc. Int. Conf. Pattern Recognition*, Atlantic City, NJ, Jun. 1990, pp. 545–550.

[25] [Online]. Available: http://www.codaoctopus.com/3d_ac_im/index.asp

[26] [Online]. Available: http://www.blueviewtech.com/

[27] [Online]. Available: http://www.soundmetrics.com/

Shahriar Negahdaripour received B.S., M.S., and Ph.D. degrees from the Massachusetts Institute of Technology, Cambridge, in 1979, 1980, and 1987, respectively.

He joined the Electrical Engineering Department, University of Hawaii, Honolulu, in January 1987 as an Assistant Professor. In August 1991, he joined the University of Miami, Coral Gables, FL, where he is currently a Professor of electrical and computer engineering. He has devoted much of his research career to the development of various vision technologies for underwater applications, starting with a Sea Grant College Program project on the automatic vision-based ROV station keeping, in 1988. He has been responsible for the development of other real-time vision-based ROV technologies including concurrent mosaic construction and mosaic-based positioning and navigation, and automatic stereovision-based navigation and mapping system for ship hull inspection.

Dr. Negahdaripour was a member of the 14-man team of U.S. researchers at the U.S.-France Workshop on Experimental Subsea Robotics in 1992 (Toulon, France), and has served as the General Co-Chair (with G. Medioni, USC) of the IEEE CVPR'91, held in Maui, HI, and the IEEE Int. Symposium on Computer Vision (with T. S. Huang, University of Illinois at Urbana-Champaign, in 1995, held in Miami, FL), and as a Technical Committee Member of ICPR'94 and CVPR'04. He co-organized with A. K. Jain of MSU the NSF workshop on "Challenges in Computer Vision: Future Directions in Research" in 1991 (Maui, HI) and with J. Yuh the NSF Workshop on "Future Research Direction in Underwater Robotics" in 1994 (Maui, HI). He served as a member of the National Academies Committee on Future Needs in Deep Submergence Sciences (2003) and is currently the Local Arrangement Co-Chair of CVPR'09, Miami, June 2009.

Hicham Sekkati recieved the B.S. degree in physics from Rabat, Morroco, in 1994, and the M.S. and Ph.D. degrees from the National Institute for Scientific Research (Institut National de la Recherche Scientique), Montreal, QC, Canada, in 2003 and 2005, respectively.

He held a postdoctoral appointment at the Underwater Vision and Imaging Lab, University of Miami, Coral Gables, FL, from November 2005 to May 2007. His research interests cover image processing and computer vision.

Hamed Pirsiavash received the B.S. degree in electrical engineering from the Iran University of Science and Technology, Tehran, Iran, in 2003, and the M.S. degree in electrical engineering from the Sharif University of Technology, Tehran, in 2006. He is currently pursuing the Ph.D. degree in computer science at the University of California, Irvine.

He joined Image Processing and Underwater Vision Lab, University of Miami, Coral Gables, FL, as a research assistant. His research interests are multimedia processing, computer vision, and machine learning.