

Article

Ghost Removal from Forward-Scan Sonar Views near the Sea Surface for Image Enhancement and 3-D Object Modeling

Yuhan Liu and Shahriar Negahdaripour * 

ECE Department, University of Miami, Coral Gables, FL 33146, USA; liuyuhan9172@foxmail.com

* Correspondence: nshahriar@miami.edu

Abstract: Underwater sonar is the primary remote sensing and imaging modality within turbid environments with poor visibility. The two-dimensional (2-D) images of a target near the air–sea interface (or resting on a hard seabed), acquired by forward-scan sonar (FSS), are generally corrupted by the ghost and sometimes mirror components, formed by the multipath propagation of transmitted acoustic beams. In the processing of the 2-D FSS views to generate an accurate three-dimensional (3-D) object model, the corrupted regions have to be discarded. The sonar tilt angle and distance from the sea surface are two important parameters for the accurate localization of the ghost and mirror components. We propose a unified optimization technique for improving both the measurements of these two parameters from inexpensive sensors and the accuracy of a 3-D object model using 2-D FSS images at known poses. The solution is obtained by the recursive updating of sonar parameters and 3-D object model. Utilizing the 3-D object model, we can enhance the original images and generate synthetic views for arbitrary sonar poses. We demonstrate the performance of our method in experiments with the synthetic and real images of three targets: two dominantly convex coral rocks and a highly concave toy wood table.

Keywords: 2-D imaging sonar; 3-D object modeling; multipath; sea-surface reflection; ghost object; mirror image; image enhancement

**Citation:** Liu, Y.;Negahdaripour, S. Ghost Removal from Forward-Scan Sonar Views near the Sea Surface for Image Enhancement and 3-D Object Modeling. *Remote Sens.* **2024**, *16*, 3814. <https://doi.org/10.3390/rs16203814>

Academic Editors: Andrzej Stateczny, Martin Gade, Silvia Liberata Ullo and Li Zhang

Received: 24 May 2024

Revised: 2 October 2024

Accepted: 6 October 2024

Published: 14 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In addition to enabling the generation of environmental maps, building 3-D object models from 2-D images provides information for automatic target recognition, establishing the types, sizes, and relative spatial arrangement of various scene objects, collision-free navigation, landmark-based positioning, reallocating objects within mapped sites, and change detection.

For underwater applications, 2-D optical imaging has restricted range in relatively clear waters and quickly becomes ineffective under poor visibility and turbidity—conditions that are commonly encountered in polluted ports, shallow turbulent surf zones, rivers, marine sanctuaries, lakes, ponds, and other still bodies of water. Here, sonar is the preferred and primary visual sensing modality, by virtue of acoustic signal penetrability through silt, mud, and other similar sources of water turbidity.

The imaging sonars have been commonly utilized in subsea remote sensing applications to generate 2-D visual and 3-D topographical maps of the ocean floor [1,2]. In one deployment scenario, a side-scan sonar (SSS) operating at frequencies of 100–500 KHz can map narrow strips of the terrain that extend up to several hundred meters on either side of a ship [3,4]. The typical (across-track) range resolutions of several centimeters can be improved to 1–2 [cm] at operating frequencies of about 1 MHz. This is achieved at the cost of decreased maximum ranges of only 10's of meters because the acoustic waves at these high frequencies attenuate faster than those at lower frequencies that are commonly employed for long-range applications.

The Synthetic Aperture Sonar (SAS) technology [2,5,6] overcomes the decreasing SSS resolution (when operating at the lower frequencies required for long-distance applications) by integrating data collected over multiple sonar transmissions with a synthetic array. Here, constant resolution with range can be achieved by scaling of the array; increasing the array length for longer ranges.

The 2-D multi-beam FSS systems, originally developed in the early 2000s at operating frequencies of 100's KHz, have been commonly deployed near the front end of the hull for obstacle-free navigation in shallow, poorly-charted waters [7–9]. The advanced units operating at 2–3 MHz produce high-resolution video images of small objects with sub-centimeter resolution at 10–30 frames/s but a limited maximum range of about a dozen meters; e.g., [10–12]. The mapping of relatively large areas of the sea bottom can be achieved by image mosaicing [13–17]. The increasing use of these high-resolution FSS systems has sparked interest in developing algorithms for automated 2-D sonar image processing, interpretation, and 3-D object reconstruction.

In 2-D ranging systems, it is common to employ the spherical coordinate $(\mathcal{R}, \theta, \phi)^T$ for the 3-D position of an object/scene surface points. Here, the elevation angle ϕ is lost in the projection of these 3-D points onto the 2-D sonar image. The 3-D reconstruction from FSS data involves the estimation of the unknown ϕ by various “shape from X” techniques [18]. For example, several 2-D images at a multitude of known poses can be used to generate high-resolution 3-D object models [19–25], with key applications in target recognition, environmental mapping, and marine scientific studies that rely on the distribution of object types and sizes. However, satisfactory performance often requires certain implicit/explicit restrictive assumptions and conditions about the environment, sonar trajectory, as well as target shape. For example, nearby reflecting surfaces in cluttered scenes and (or) surface patches within object concavities generate multipath that can corrupt and degrade the target image. Thus, the key conditions to minimize data corruption for accurate 3-D model reconstruction include the following: (1) dominantly convex targets (with only mild concavities producing no more than negligible multipath reverberation); and (2) no nearby surfaces with strong acoustic reflectance, e.g., air–sea interface, hard bottoms, and clutter.

The sonar images of targets resting on hard bottoms or in shallow water (floating) near the air–water interface can be corrupted by the contributions from the multipath propagation of acoustic beams [19,20,26,27]. These generate ghost regions that often overlap with the target image (the multipath phenomenon also arises in deploying a high-resolution radar on an autonomous road vehicle for environmental perception [25,28–31], where the key objective is to locate the “ghost” vehicles in the scene images). Referring to each of the four views of a coral rock in Figure 1, the indistinguishable overlapping ghost object corrupts both the shape and intensity values of (some parts within) the coral image region. Additionally, a mirror region is formed by the “virtual mirror object” above the water surface (or below the seabed). As in (a), the mirror component (nearly) coincides with the ghost region when the image plane is (roughly) parallel to the sea surface (or bottom). It separates into a distinct blob upon rotating the sonar about the viewing axis; here, from 0° to 67.5° in steps of 22.5° .

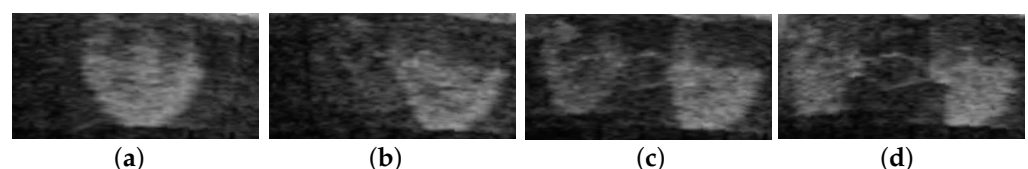


Figure 1. Ghost component overlaps with and is indistinguishable from the object region in every view. Mirror component at reference position (elevation axis pointing upward) overlaps with both object and ghost regions (a). As the sonar rotates about the viewing direction (from 0° to 67.5° in increments of 22.5° , here), it separating from the object (b), and forms a distinct blob (c,d).

In this work, we investigate the generation of a 3-D model for an object floating near the sea surface, imaged by an FSS from a multitude of known poses. To this end, improved 3-D modeling performance directly depends on the accurate knowledge of sonar pose relative to the air–water interface, allowing us to circumvent the shape distortion in the 3-D object modeling process. To be precise, the air–water interface distance and the sonar tilt angle are two key pose parameters in localizing the ghost and mirror components. Both parameters can be determined using relatively inexpensive sensors [32,33], but the measurement inaccuracies are generally unacceptable. We propose to estimate these with reduced uncertainty and (or) improved precision by devising an integrated recursive optimization technique that builds on an earlier 3-D modeling framework [19,20]. Moreover, we exploit the mirror contours, where distinct in certain views, as a regularizer to improve the reconstruction accuracy.

Additionally, our method enables the generation of ghost-free object views from the model-based synthetic images [26,27,34–36], using a look-up table that is constructed from the intensity distributions within the uncorrupted object regions in the data and synthetic images. Experimental results confirm that our method performs consistently with real and computer-generated data under similar imaging conditions. The real data includes the images at known poses of two dominantly convex coral rocks and a concave toy wood table, recorded by a lens-based Dual-Frequency IDentification SONar (DIDSON) [10].

The remaining sections are organized as follows: Section 2 covers (1) some notations and relevant technical background, including 3-D coordinate transformation, 3-D-to-2-D sonar projection, representation of sonar data as a beam-bin array, and transformation to generate a 2-D FSS image; (2) a unified optimization framework for 3-D target modeling and sonar pose estimation, error metrics applied for the quantitative assessment of the results, and the generation of ghost-free object images. In Section 3, we first present the results of experiments with synthetic data, aimed at assessing the accuracy and convergence rate of our optimization scheme for 3-D modeling and sonar pose estimation. Here, we utilize the 3-D models of our three targets generated by a Kinect camera under the same imaging conditions as the real data. We then give the results of experiments with real data. We summarize our contributions in Section 4.

2. Materials and Methods

2.1. Notation, Coordinate Transformation and FSS Image Formation

The Cartesian $\mathbf{P} = (X, Y, Z)$ and spherical $\mathbf{\Pi} = (\mathfrak{R}, \theta, \phi)$ coordinates of a 3-D point transform according to the following equations:

$$\mathbf{P} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \mathfrak{R} \begin{bmatrix} \cos \phi \sin \theta \\ \cos \phi \cos \theta \\ \sin \phi \end{bmatrix} \longrightarrow \mathbf{\Pi} = \begin{bmatrix} \mathfrak{R} \\ \theta \\ \phi \end{bmatrix} = \begin{bmatrix} \sqrt{X^2 + Y^2 + Z^2} \\ \tan^{-1}(X/Y) \\ \tan^{-1}(Z/\sqrt{X^2 + Y^2}) \end{bmatrix} \quad (1)$$

Figure 2a depicts a single sonar beam in the azimuth direction θ . In a multi-beam FSS, the horizontal field of view (FoV) $|\theta| \leq W_\theta$ is covered by N_b beams, each with a narrow horizontal width $\delta\theta$. The vertical FoV is fixed by the half-angle beam width W_ϕ . For most existing FSS systems, we have $\delta\theta = 0.25^\circ - 1^\circ$, $6^\circ \leq W_\phi \leq 10^\circ$ and $15^\circ \leq W_\theta \leq 65^\circ$.

At various sample times, the round-trip times of flight T_{tof} of target echos in a direction θ yield the range measurements $\mathfrak{R} = \frac{1}{2}T_{tof}v$, where v is the sound speed within the medium. Ideally, we calculate $(\mathfrak{R}, \theta)$ from the so-called “beam-bin” (b, B) coordinates by the following linear equations:

$$\begin{aligned} \mathfrak{R} &= \mathfrak{R}_{\min} + (B - 1)\delta\mathfrak{R} \quad (B = 1, 2, \dots, N_B) \\ \theta &= -W_\theta + (b - 1)\delta\theta \quad (b = 1, 2, \dots, N_b) \end{aligned} \quad (2)$$

where N_B range bins cover the minimum range $\mathfrak{R}_{\min}$ to maximum range $\mathfrak{R}_{\max}$ with resolution $\delta\mathfrak{R} = (\mathfrak{R}_{\max} - \mathfrak{R}_{\min})/(N_B - 1)$. In practice, deviations from these models are

determined and rectified by calibration; e.g., a “mild” cubic equation models the b -to- θ transformation in a lens-based DIDSON with lens distortion [37].

The beam-bin intensity array $I(b, B)$ encodes the strength of (collective) echoes from potentially several scene surface patches within the elevation arc W_ϕ , leading to the inherent many-to-one projection ambiguity of FSS imaging. We construct the 2-D FSS image $I(x, y)$ from the beam-bin data using the transformation

$$\begin{bmatrix} x \\ y \end{bmatrix} = \Re \begin{bmatrix} \sin \theta \\ \cos \theta \end{bmatrix} \quad (3)$$

The elevation angles ϕ of points on an object or scene surfaces (over some region of interest) are lost in the 3-D-to-2-D projection process. The application of a 3-D reconstruction technique involves the recovery of unknown ϕ from the radiometric and (or) geometric cues in one or more FSS images. In this work, we utilize the data captured with a DIDSON unit, for which an image formation model has been derived [26]; this enables synthesizing images of a given object at known poses relative to the sonar.

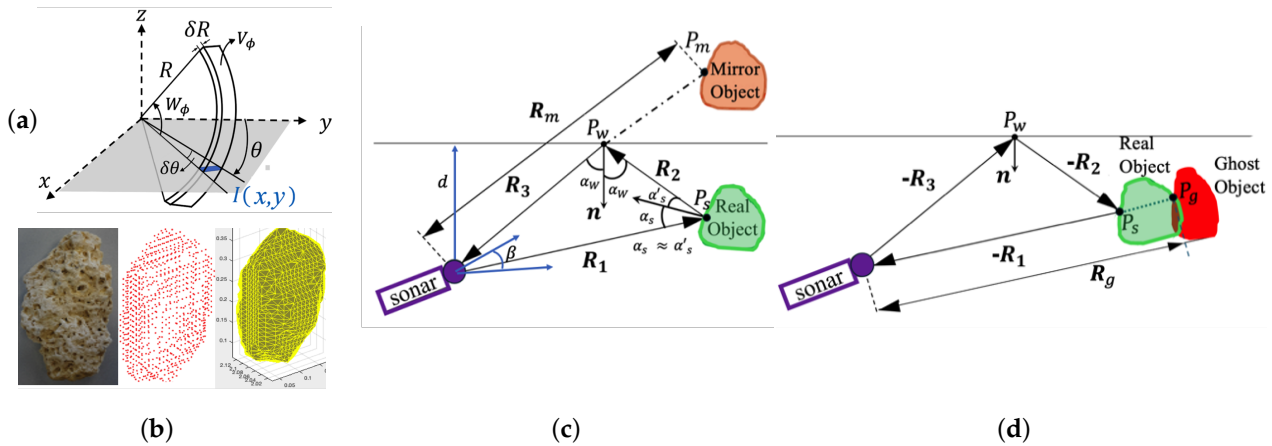


Figure 2. (a) For a sonar beam in θ direction, image intensity I of pixel (x, y) depends on cumulative echoes from unknown number of surface patches within volume V_ϕ arriving at sonar receiver simultaneously; V_ϕ covers elevation-angle interval $[-W_\phi, W_\phi]$, range interval $[\Re, \Re + \delta\Re]$ along the beam covering azimuthal-angle interval $[\theta, \theta + \delta\theta]$. (b) A coral rock with voxelated volume and triangular surface mesh of SC solution. (c) Virtual mirror object geometry: transmitted sound waves in direction $\mathbf{R}_1$ are scattered by surface at P_s . Reflected portion along “unique direction” $\mathbf{R}_2$ towards P_w on water surface (with surface normal $\mathbf{n}$) is specularly reflected towards the sonar along $\mathbf{R}_3$, leading to the appearance of a virtual mirror object point at P_m . (d) Virtual ghost object geometry: considering the reverse direction of the mirror-point pathway, sound waves traveling along $-\mathbf{R}_3$ are specularly reflected towards the object along $-\mathbf{R}_2$, and are scattered at P_s , of which components along $-\mathbf{R}_1$ are captured by the sonar. This leads to the appearance of ghost point P_g along the sonar beam directed at P_s (at a longer range $\mathbf{R}_g$).

2.2. 3-D Object Modeling

Applying the space carving (SC) paradigm [38], a 3-D reconstruction method has been proposed [39] to generate an object surface model S^o from M images I_m at distinct sonar poses $\mathbf{m}_m$ ($m = 0, 1, \dots, M-1$). The $M = M_p M_r$ poses are represented by 3-D translation $\mathbf{t}_i$ ($i = 0, 1, \dots, M_p - 1$) and rotation $\mathbf{r}_j$ ($j = 0, 1, \dots, M_r - 1$) vectors relative to reference pose $\mathbf{m}_0 = [\mathbf{t}_0^T, \mathbf{r}_0^T]^T$. The carving efficiency is improved by acquiring M_r rotated images at each of M_p positions around the object; through incremental roll rotations $d\omega_r = \pi / M_r$ [rad] (about the viewing axis) from $-\pi$ to $-\pi + (M_r - 1)d\omega_r$ (The shape complexity, namely, convexity or concavity are two key factors in setting M_p and M_r).

The SC method exploits geometric cues solely. First, the 3-D space visible in all sonar poses is discretized into volume elements (voxels). Each is examined to label as an

object voxel if it projects onto the feasible region in every view; otherwise it is labeled as a non-object voxel and discarded. The feasible region of each view is defined by the object highlight and dark/shadow region behind it, onto which the object surface patches occluded by the visible surfaces may project. The surface model S^o is the outer boundary of the volume formed by all object voxels, as shown in Figure 2b for the coral-two rock in our experiments. Referring to Figure 2b, the SC model S^o consists of a closed mesh of N_T triangles formed by N_p vertices $\mathbf{P}_l^o$ ($l = 1, \dots, N_p$); the triangulation connectivity list T_s fixes the N_T triangles in S^o , formed by the triplet sets $\{\mathbf{P}_i^o, \mathbf{P}_j^o, \mathbf{P}_k^o\}$.

The 3-D model refinement scheme in [19,20] is an efficient implementation of the optimization framework in [40] that utilizes the intensity data within object regions in various views by minimizing the discrepancy between the data and synthetic target images, generated by the sonar image formation model in [26]. Starting with the SC mesh model S^o , the sought-after optimal model $\tilde{S}$ is the discrete approximation of the true continuous surface S^c , defined by the N_p displaced vertices $\tilde{\mathbf{P}}_l = \mathbf{P}_l^o + \tilde{\mathbf{V}}_l$ ($l = 1, \dots, N_p$) of the N_T triangles in the triangular connectivity list T_s of S^o . The optimal displacements $\tilde{\mathbf{V}} = [\tilde{\mathbf{V}}_1, \tilde{\mathbf{V}}_2, \dots, \tilde{\mathbf{V}}_{N_p}]$ are calculated iteratively by minimizing the sum-squared difference between the data I_m and synthetic images $\tilde{I}_m$ over all M poses:

$$\tilde{\mathbf{V}} = \arg \min_{\tilde{\mathbf{S}}} E(S) = \sum_{m=1}^M |I_m - \tilde{I}_m(x, y; S)|^2 \quad (4)$$

At iteration t , the incremental 3-D vertex displacement field $\tilde{\mathbf{V}}^t$ (to revise the 3-D model $\tilde{S}^t$) is estimated from the 2-D motion fields $\mathbf{v}_m^t$ that align the data and synthetic images at “relevant” sonar poses $\mathbf{m}_m$ ($m = 1, 2, \dots, M^- \leq M$). Here, $(M - M^-)$ views may be excluded if the vertices/patches are invisible or the 2-D motion vectors have large errors.

For targets imaged near the sea surface, we need to account for the impact of multipath propagation, namely, the object region corrupted by the ghost component. To this end, multipath modeling serves two purposes: (1) to localize and discard the ghost-corrupted object region to avoid the distortion of the 3-D model; and (2) to exploit the geometric constraints imposed by the mirror image boundaries in relevant views to improve the 3-D object modeling. It is noted that the multipath arising from reverberation within object concavities and ground reflection for objects on the seabed have been incorporated in sonar simulators [26,27,34–36,41–43]. However, these investigations offer no model validation with real data and calibrated objects. In our work, we rely solely on geometric cues, i.e., the ghost and mirror locations and the mirror contour, for 3-D model refinement.

2.3. Unified Optimization Framework

Inexpensive depth sensors with an accuracy of 0.1% have negligible error at depths of a few meters [32]. However, a low-cost angle sensor with an uncertainty of $\pm 0.6^\circ$ [33] can introduce large biases in localizing the ghost, mirror, and thus the uncorrupted object regions. We next describe a unified optimization scheme by incorporating the estimation of the two pose parameters within the 3-D target modeling framework.

Figure 3a depicts the block diagram of the process: recursively updating the pose parameters and the 3-D shape. Here, we employ the 3-D object model estimate to calculate the new values of sonar depth d and tilt angle β . These are adopted only if they yield a smaller error, subsequently employed in the next iteration of 3-D shape refinement. The iterative updating of the 3-D target model, detailed in [19,20] and depicted in Figure 3b, consists of three key steps: (1) computation of 2-D motion vectors in various views for the alignment of object regions in the data and synthetic images (constructed from the 3-D model); (2) estimation of the 3-D motions of corresponding triangular-patch centers from the 2-D image motion fields for all views; (3) transformation of 3-D motions from the patch centers to the 3-D model vertices. Finally, we update the 3-D shape by displacing the model vertices. The execution of the two-mode optimization continues as long as the objective error function diminishes.

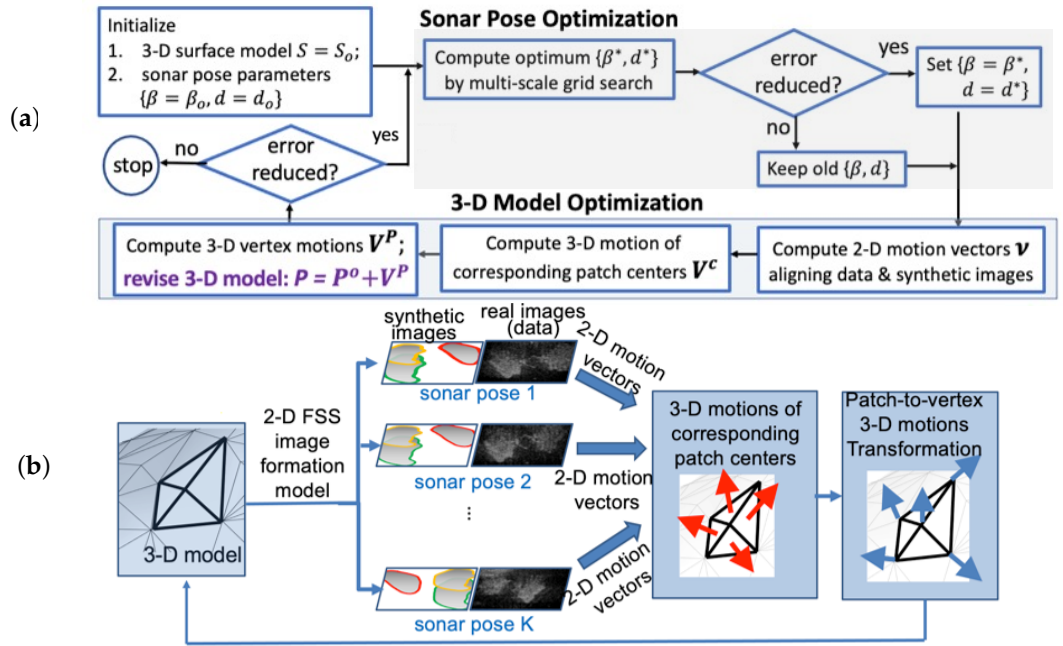


Figure 3. (a) Block diagram of entire algorithm; (b) steps in 3-D shape optimization by displacement of model vertices, computed from 3-D vertex motions that are estimated from the 2-D image motions aligning the object regions in the data and synthetic views.

2.4. Sonar Pose Optimization

Referring to Figure 2d, the ghost image, generally overlapping with the more distant region of the object image, is formed by the sound propagation along the longer multipath routes. The paths for all visible surface patches are readily calculated based on the sonar pose relative to the target and to the air–water interface [19,20]. However, inaccuracies in the initial 3-D SC model (note that the 3-D model incorporates the relative sonar-to-target pose) and the depth and tilt sensor measurements can lead to the imprecise location of the ghost region, thus diminishing the effectiveness of utilizing the image intensity and contour information for various views. We minimize both types of error through the optimization method described next.

We employ superscript $(\cdot)^O$ and $(\cdot)^M$ for entities associated with the object and mirror components, respectively; e.g., 3-D models S^O and S^M explicitly differentiate between the true model and the virtual 3-D mirror object, respectively. For a sonar pose $\mathbf{m}_m$ ($m = 1, 2, \dots, M$), a *frontal patch* refers to the visible surface patch at shortest range along each sonar beam, projecting onto a *frontal contour* point in the corresponding view.

Let $\mathcal{C}_m^O = \{c_{mk}^O \mid k = 1, 2, \dots, N_m^O\}$ and $\mathcal{C}_m^M = \{c_{ml}^M \mid l = 1, 2, \dots, N_m^M\}$ denote the sets of frontal contour points of the object and mirror regions in image I_m . Likewise, $\tilde{\mathcal{C}}_m^O = \{\tilde{c}_{mi}^O \mid i = 1, 2, \dots, \tilde{N}_m^O\}$ and $\tilde{\mathcal{C}}_m^M = \{\tilde{c}_{mj}^M \mid j = 1, 2, \dots, \tilde{N}_m^M\}$ denote the same in the synthetic image $\tilde{I}_m$. Note that $N_m^O \neq \tilde{N}_m^O$ and $N_m^M \neq \tilde{N}_m^M$ generally, for real quantized data. We perform contour alignment by the Iterative Closest Point (ICP) algorithm [44,45], namely, the iteratively reweighted least squares (IRLS) variant [46,47]. A key advantage of the ICP methods is the applicability in registering two unequal-size unmatched point sets.

Figure 4 depicts these contours superimposed on a sample real image. The registration of contours $\tilde{\mathcal{C}}_m^O$ and $\mathcal{C}_m^O$ with $\tilde{N}_m^O \leq \tilde{N}_m^O$ inlier matching points $\{\tilde{c}_{mi}^O, c_{mk}^O\}$ (with indices $\{i, k(i)\}$) yields the planar transformation $\tilde{\mathcal{C}}_m^O = T^O(\tilde{\mathcal{C}}_m^O)$ ($\tilde{\mathcal{C}}_m^O = \{\tilde{c}_{mi}^O \mid i = 1, 2, \dots, \tilde{N}_m^O\}$) ($T^O(\cdot)$ involves 2-D translation and in-plane rotation).

The 2-D motion of contour $\tilde{\mathcal{C}}_m^O$ is given by $\tilde{\mathbf{V}}_m^O = \{\tilde{\mathbf{v}}_{mi}^O \mid i = 1, 2, \dots, \tilde{N}_m^O\}$, where $\tilde{\mathbf{v}}_{mi}^O = \tilde{c}_{mi}^O - c_{mi}^O$.

For each synthetic image $\tilde{I}_m$, the contour error is computed based on the discrepancy between the contour pixels $c_{mk(i)}^O$ in the data I_m and the transformed contour pixels $\tilde{c}_{mi}^{'O}$ in the synthetic view $\tilde{I}_m$:

$$CE(\tilde{I}_m) = \frac{1}{\tilde{N}_m^{'O}} \sum_{i=1}^{\tilde{N}_m^{'O}} |c_{mk(i)}^O - \tilde{c}_{mi}^{'O}| \quad (5)$$

Following the same steps for the mirror image, $\tilde{N}_m^{'M}$ pairs $\{\tilde{c}_{mj}^{'M}, c_{ml}^M\}$ of inlier mirror-contour correspondences with indices $\{j, l(j)\}$ are established by the transformation $\tilde{C}_m^{'M} = T^M(\tilde{C}_m^M)$. This yields the 2-D mirror-contour motions $\tilde{V}_m^M = \{\tilde{v}_{mj}^M \mid j = 1, 2, \dots, \tilde{N}_m^{'M} \leq \tilde{N}_m^M\}$, where $\tilde{v}_{mj}^M = \tilde{c}_{mj}^{'M} - \tilde{c}_{mj}^M$.

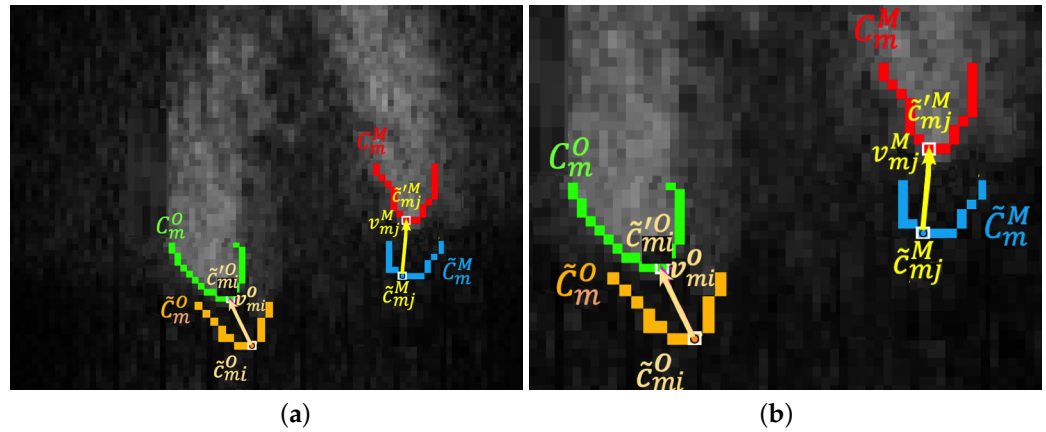


Figure 4. (a) The 2-D vectors $\{v_{mi}^O, v_{mj}^M\}$ align the frontal contours $\{C_m^O, C_m^M\}$ of the object and mirror regions in the real images with counterparts $\{\tilde{C}_m^O, \tilde{C}_m^M\}$ in the synthetic views; (b) magnified view of relevant regions.

With accurate sonar pose parameters D and β , the 2-D motions $\tilde{V}_m^O$ and $\tilde{V}_m^M$ computed from images I_m and $\tilde{I}_m$ (for pose $\mathbf{m}_m$) incorporate the inaccuracies of estimated 3-D models $\tilde{S}^O$ and $\tilde{S}^M$ in a consistent manner.

To explain, we utilize the simplified notation $\{z \mid z = 1, 2, \dots, Z_m\}$ for the set of Z_m frontal patch indices in view $\mathbf{m}_m$ for both the true and estimated object-mirror pairs $\{S^O, S^M\}$ and $\{\tilde{S}^O, \tilde{S}^M\}$, respectively. Moreover, the 2-D motion vectors v_{mz}^O and v_{mz}^M align the frontal contours of objects and mirror regions in the real and synthetic images, respectively. Finally, the erroneous 2-D motion vectors based on imperfect knowledge of sonar depth and tilt angle are denoted $\tilde{v}_{mz}^O$ and $\tilde{v}_{mz}^M$.

Next, let the pairs $\{\mathbf{P}_z^c, \tilde{\mathbf{P}}_z^c\}$ represent the true and estimated 3-D frontal patch centers on $\{S^O, \tilde{S}^O\}$, projecting onto matching 2-D object-contour points $\{c_{mz}^O, \tilde{c}_{mz}^O\}$ in the data I_m and synthetic view $\tilde{I}_m$, respectively. In [19,20], we show how to calculate the 3-D motion of the patch center $\tilde{\mathbf{V}}_z^O = \tilde{\mathbf{P}}_z^c - \mathbf{P}_z^c$ from the redundant 2-D object-contour motions $\tilde{v}_{mz}^O$ in visible views $m = 1, 2, \dots, M^- \leq M$.

In the same manner, (1) the patch-center pair $\{\mathbf{P}_z^c, \tilde{\mathbf{P}}_z^c\}$ of true and estimated 3-D mirror model pair $\{S^M, \tilde{S}^M\}$ project onto matching 2-D mirror-contour pairs $\{c_{mz}^M, \tilde{c}_{mz}^M\}$ in the data I_m and synthetic view $\tilde{I}_m$, respectively; and (2) the 3-D mirror patch-center motion $\tilde{\mathbf{V}}_z^M = \tilde{\mathbf{P}}_z^c - \mathbf{P}_z^c$ is calculated from the redundant 2-D mirror-contour motions $\tilde{v}_{mz}^M$ over the visible mirror patches.

The *consistent manner* means that mapping of the pair $\{\mathbf{P}_z^c, \tilde{\mathbf{P}}_z^c\}$ to two corresponding 3-D mirror pairs on $\{S^M, \tilde{S}^M\}$ yields the same terminal points $\{\mathbf{P}_z^c, \tilde{\mathbf{P}}_z^c\}$ that are derived from the 2-D mirror-contour motions $\tilde{v}_{mz}^M$ (calculated from all views where the mirror surface patch centered at $\tilde{\mathbf{P}}_z^c$ is visible).

This so-called consistency does not hold when sonar pose parameters are imprecise: deviations arise in $\tilde{\mathbf{V}}_z^O$ and $\tilde{\mathbf{V}}_z^M$ due to the inaccurate 2-D motions calculated from the dislocated object and mirror contours. First, we expect the 3-D motion vector $\tilde{\mathbf{V}}_z^O$ to be larger than the estimate based on precise sonar depth and tilt measurements. Accordingly, we would seek sonar pose parameters that minimize $\|\tilde{\mathbf{V}}_z^O\|$. Next, inconsistency exists between $\tilde{\mathbf{V}}_z^M$, calculated from the 2-D mirror contour motions, and $\mathbf{V}_z^M$ based on the mapping of 3-D patch-center motion from the object $\tilde{S}^O$ to the virtual mirror $\tilde{S}^M$. We can directly calculate the discrepancy $\|\mathbf{V}_z^M - \tilde{\mathbf{V}}_z^M\|$, which is also to be minimized by the optimal sonar pose parameters. Collectively, we define and minimize the error metric $E_\mu(D, \beta)$ that integrates the errors μ^m for all sonar poses:

$$\mu^m = \frac{1}{Z_m} \sum_{z=1}^{Z_m} \|\mathbf{V}_z^O\| + \|\mathbf{V}_z^M - \tilde{\mathbf{V}}_z^M\|; \quad E_\mu(D, \beta) = \frac{1}{M} \sum_{m=1}^M \mu^m \quad (6)$$

The error metric $E_\mu(D, \beta)$ is a highly nonlinear function of parameters D and β , requiring a computationally expensive non-linear optimization scheme (e.g., some variation of iterative gradient descent [48]). Moreover, the convergence to the global minimum cannot be guaranteed. Instead, we determine the optimal values D^* and β^* more efficiently by multi-resolution grid search in a 2-D parameter space. That is, representing the uncertainty bounds of sensor measurements D_0 (of depth) and β_0 (of tilt angle) by $\pm e_D$ and $\pm e_\beta$, we calculate $E_\mu(D, \beta)$ over the 2-D grid $\mathcal{U}_g = [D_0 - e_D, D_0 + e_D; \beta_0 - e_\beta, \beta_0 + e_\beta]$ with (low to high) grid resolution δD and $\delta \beta$:

$$(D^*, \beta^*) = \arg \min_{(D, \beta) \in \mathcal{U}_g} E_\mu(D, \beta) \quad (7)$$

We update D and β in a “greedy” manner: (D^*, β^*) is accepted if it gives a smaller error E_μ compared with the latest value; see the “Sonar Orientation Optimization” block in Figure 3a. Next, we carry out the computational steps in the “3-D Model Optimization” block as described in [20].

2.5. Error Metric

We define two error measures to assess the performance of our method [19,20]: to quantify the 3-D object model error and the discrepancy between the data and synthetic images generated from the 3-D model. Denoting $\tilde{\mathcal{I}}^t = \{\tilde{I}_m^t \mid m = 1, 2, \dots, M\}$ the set of synthetic beam-bin images at iteration t , we employ the average intensity error (IE) comprising of pixel set Ω_m^t within the uncorrupted object region of $\tilde{I}_m^t$, in calculating the average intensity error (AIE) over the whole synthetic image set:

$$\text{IE}(\tilde{I}_m^t) = \frac{1}{|\Omega_m^t|} \sum_{(b, B) \in \Omega_m^t} |\tilde{I}_m^t(b, B) - I^t(b, B)| \quad \text{AIE}(\tilde{\mathcal{I}}^t) = \frac{1}{M} \sum_{m=1}^M \text{IE}(\tilde{I}_m^t) \quad (8)$$

where $|\Omega_m^t|$ denotes the cardinality of set Ω_m^t . We also compute the average contour error (ACE) from (5):

$$\text{ACE}(\tilde{\mathcal{I}}^t) = \frac{1}{M} \sum_{m=1}^M \text{CE}(\tilde{I}_m^t) \quad (9)$$

Finally, these two average errors (over the entire data set) are normalized with respect to their initial values (using the synthetic images of the 3-D SC solution):

$$\text{NAIE}(\tilde{\mathcal{I}}^t) = \text{AIE}(\tilde{\mathcal{I}}^t) / \text{AIE}(\tilde{\mathcal{I}}^0) \quad \text{NACE}(\tilde{\mathcal{I}}^t) = \text{ACE}(\tilde{\mathcal{I}}^t) / \text{ACE}(\tilde{\mathcal{I}}^0) \quad (10)$$

and combined to establish the image error at iteration t :

$$E_I(t) = \frac{1}{2} (\text{NAIE}(\tilde{\mathcal{I}}^t) + \text{NACE}(\tilde{\mathcal{I}}^t)) \quad (11)$$

The 3-D modeling error can be quantified in terms of the volumes $V^c = V(S^c)$ and $\tilde{V} = V(\tilde{S})$ of the true and estimated 3-D object models. With no knowledge of S^c , we instead utilize the model generated by a Kinect camera [49], albeit not perfect. Accordingly, we define the normalized volumetric error at iteration t [39]:

$$E_V^t = \frac{\tilde{V}^t + V^c - 2V_{\tilde{S}^t \cap S^c}}{\tilde{V}^t + V^c - V_{\tilde{S}^t \cap S^c}}; \quad V_{\tilde{S}^t \cap S^c} = V(\tilde{S}^t \cap S^c) \quad (12)$$

The metric $0 \leq E_V \leq 1$ is the ratio of non-common volumes of the two 3-D models to their union. Thus, zero error indicates two identical volumes, and the maximum unity error corresponds to no common volume. We utilize E_V primarily to determine the correlation with E_I and whether E_I (which is computable) can serve as an indicator of E_V (which we cannot determine in practice).

2.6. Enhanced Target Images

As discussed, almost all object views are generally corrupted by the ghost image, and only some by the mirror image component. For any arbitrary sonar pose, we can synthesize a noise-free, uncorrupted image from the 3-D target model. In addition, it is highly desired to generate enhanced real object views by removing the ghost contributions.

Let $I_o = \cup(I_o^N, I_o^O)$ and $I_g = \cup(I_g^N, I_g^O)$ denote the intensity values of object and ghost components, respectively, formed by the union of their respective overlapping and non-overlapping regions (steps 2a and 2b in Figure 5b,c). Ignoring the impact of noise, I_o^N and I_g^N are measured directly from the image. The object image enhancement involves the generation of I_o by the estimation of I_o^O . In the process given below to generate the object image, we assume the transformation to map the synthetic image to the real data within the non-overlap object region can be reliably applied to estimate the intensity values of the object within the overlap region. Referring to Figure 5, the complete process involves the following steps:

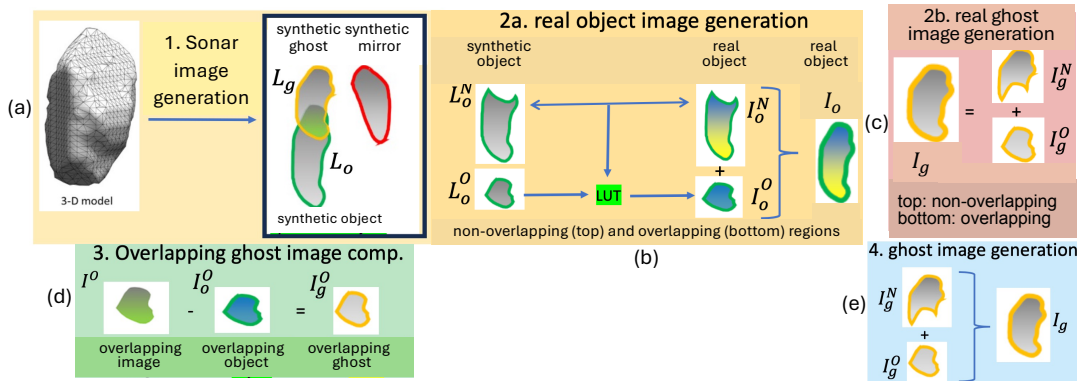


Figure 5. Processing steps in the decomposition of sonar data into object and ghost components. (a) Generation of synthetic object image from image formation model, and localizing the ghost and mirror components to identify regions overlapping with object image. (b) Segmentation of real and synthetic object regions into overlapping and non-overlapping parts, using non-overlapping region in generating the LUT for synthetic-to-real object transformation, and apply the LUT to reconstruct overlapping object region to complete the object image by fusing with non-overlapping part. (c) Segmentation of ghost area into overlapping and non-overlapping regions, producing the non-overlapping part. (d) Discounting for the object image within overlap area to generate the ghost component. (e) Generation of ghost image from overlapping and non-overlapping components.

1. Generation of synthetic logarithmic object image L_o based on the model in [26], applied to the 3-D object model $\tilde{S}$ (step 1);
2. Localization of ghost region by employing the estimated sonar pose parameters [19,20] (step 1);

3. Segmentation of non-overlapping and overlapping object regions L_o^N from L_o^O , respectively (step 2a), and the generation of normalized intensity values $\hat{L}_o^N$ with zero mean and unit variance;
4. Calculation of mean-variance pair $\{\bar{I}, \sigma_I\}$ of intensity values I_o^N , and applying the transformation $L_o^N = \sigma_I \hat{L}_o^N + \bar{I}$;
5. Construction of look-up table (LUT) to transform L_o^N to I_o^N by matching their histograms (step 2a);
6. Computation of scaled intensity values $L_o^O = \sigma_I \hat{L}_o^O + \bar{I}$ within overlapping object region, and applying the LUT to map L_o^O values to I_o^O values (step 2a).

In some applications, including the development and verification of a ghost image formation model, it may be desired to isolate the ghost component from the real data. Figure 5c–e depicts the process, once the object image has been computed as described above.

3. Results

Experiments with both synthetic and real images of three targets are presented to demonstrate different aspects of our method. These include two “dominantly-convex” coral rocks with mild local concavities and a concave “toy” wood table; see Figure 6.

The simulations with synthetic data allow us to assess the performance and convergence of the unified optimization scheme, namely, errors in converged sonar depth and tilt angle values, the discrepancy between the optimized 3-D model (initialized by the SC solution) and the Kinect 3-D model, as well as the consistency of results from experiments with synthetic and real data.

For ground truth, we employ the 3-D surface mesh models of these objects generated by a Kinect camera, utilizing Kinect video while slowly rotating the target in air at a distance of about 1 [m]. Although not perfect, various spot measurements confirm an accuracy of better than an order of magnitude in comparison to the average error of 3-D models generated from the 2-D sonar data. The real data for the two coral rocks comprises 16 total images at two opposite, north (N) and south (S) sides of each object. At each position, eight images are captured while the sonar rolls on a motorized rotator about the viewing axis from -90° to 67.5° , in increments of 22.5° ; the unrotated (zero-degree roll) at the S position is the reference view. For the wood table, we have a total of 32 images from four sides: east (E) and west (W), in addition to N and S.

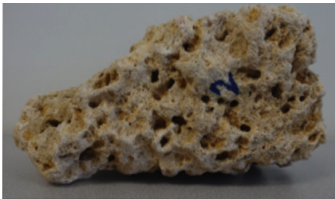
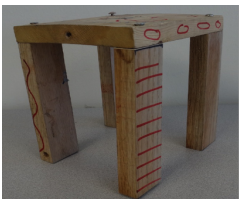
coral rock-one	coral rock-two	wood table
H=22.4 [cm], Max. W=11.4 [cm]	H=30 [cm], Max. W=16.4 [cm]	H=16.5 [cm], Max. W=15.0 [cm]
		
Sonar tilt: 10.1°	11.3°	11.0°
Sonar depth: 0.43 [m]	0.48 [m]	0.46 [m]
Target Range: 2.05–2.18 [m]	2.03–2.20 [m]	2.05–2.20 [m]

Figure 6. Three targets—two dominantly convex coral rocks with mild local concavities and a highly concave wood table—with height, maximum width, and imaging conditions.

We employ these sonar poses and the parameters of each experiment listed in Figure 6 to generate the synthetic images of the 3-D Kinect model by applying the DIDSON image formation model [26]. In assessing convergence, all experiments are initialized with the 3-D SC model [39] and imprecise depth and tilt-angle values.

Finally, the mirror component overlaps with the object region in three views: at the reference pose and two rotated views at roll angles of $\pm 22.5^\circ$. Discarding these,

the optimization is applied to 10 images of each coral rock and 20 images of the wood table (recall that we require a distinct mirror region boundary in the optimization process). Moreover, we treat the discarded views as the new ones to compare with the computer-generated images.

3.1. Experiments with Synthetic Data

Figure 7a–d depicts the results of the experiment with the coral-one data. The sonar depth and tilt parameters in (a,b) approach the assumed values of 43 [cm] and 10.3° (shown in red dashed lines) in seven iterations. In (d), we compare the initialized SC solution (top: blue mesh) and optimized 3-D model (bottom: blue mesh), superimposed on the Kinect model (black mesh). Decreasing in tandem, the image E_I and volumetric E_V errors in (c) confirm that the reduced image error leads to an improved 3-D model. Thus, the measurable E_I serves as an indicator of incalculable 3-D modeling accuracy. Initialized with the SC solution, the improved target model after seven iterations reduces the volumetric error by about 16% and the image error by better than 20%.

Similar results are obtained for the second coral rock. Referring to Figure 8a–d, the sonar depth in (a) and tilt angle in (b) approach the assumed values of 48 [cm] and 11.5° (red dashed lines), and the image and volumetric errors in (c) improve by better than 20% in four iterations.

A noteworthy observation is that the reconstructed 3-D models of both rocks preserve (some of) the mild local concavities of their surfaces. For example, these are pronounced in the XY and XZ views of coral-one; see columns 2–3 in Figures 7d and 8d.

The experimental results for the concave wood table are presented in Figure 9a–d. Here, the reverberation among the inner surfaces of the four legs leads to their inflated appearances in different views. Minimizing the discrepancy between the data and synthetic images yields a 3-D model with thickened legs. Consequently, the volumetric improvement (with respect to the initial SC solution) in (c) is relatively insignificant, despite an image error improvement of better than 20%. Still, the reconstructed 3-D model captures the structural and topological properties of the shape, including the top surface, four legs, and the deep concavity within these legs.

Moreover, the convergence of the sonar depth in (a) and tilt angle in (b) takes 10 iterations. A notable observation in (c) is the non-monotonic convergence of the image (red circles) and volumetric (blue squares) errors. These errors increase (before resuming the improvement in accuracy) at iterations where the sonar pose parameters diverge from their true values (red dashed lines).

3.2. Experiments with Real Data

Figures 7d', 8d', and 9d' are the 3-D Kinect (red. mesh) and the optimized model based on initialization with the SC solution (blue). The reconstruction accuracy of the optimized model is consistent with those from the computer simulations, albeit at a slower convergence rate; compare (c) and second row in (d) with (c',d') in each figure. The same conclusion is drawn from the wood table reconstruction, despite inferior precision relative to the two dominantly convex coral rocks. Interestingly, the convergence is achieved monotonically and faster, directly tied to the same behavior for the sonar pose parameters.

Here again, some of the mild concavities of the two coral rocks are recovered, as in the experiments with the computer-generated data. Moreover, the reconstructed model captures the structural and topological properties of the wood table. The largest discrepancy arises in the coral-two reconstruction over the SW–W and E–NE regions (e.g., see XY projection). Here, the thickness variations are not replicated in the estimated model, mainly due to the unavailability of data from the E and W directions (this has been confirmed in an experiment with real-synthetic mixed data, using additional computer-generated views from these two positions [19,20]).

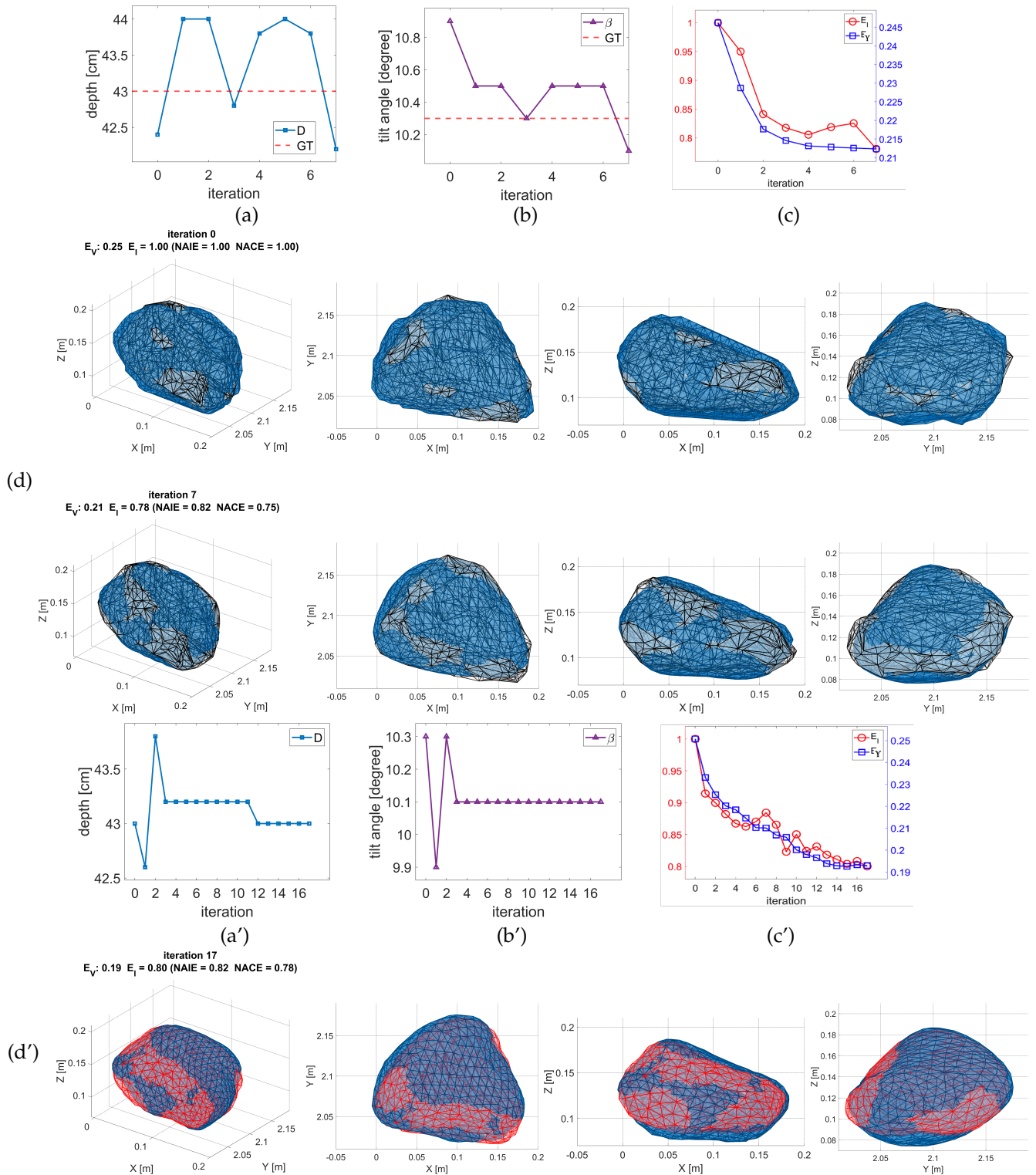


Figure 7. Coral-one experiment—(a–d) synthetic and (a'–d') real data. (a,b,a',b') Optimization of sonar depth and tilt parameters. (c,c') Image E_I and volumetric E_V errors moving in tandem confirm 3-D model improvement with reduced image error. (d) Initialized SC solution (top) and optimized 3-D model (bottom), shown by blue surface mesh, are superimposed on Kinect model (black mesh); (d') optimized SC (blue mesh) and Kinect (red mesh) models.

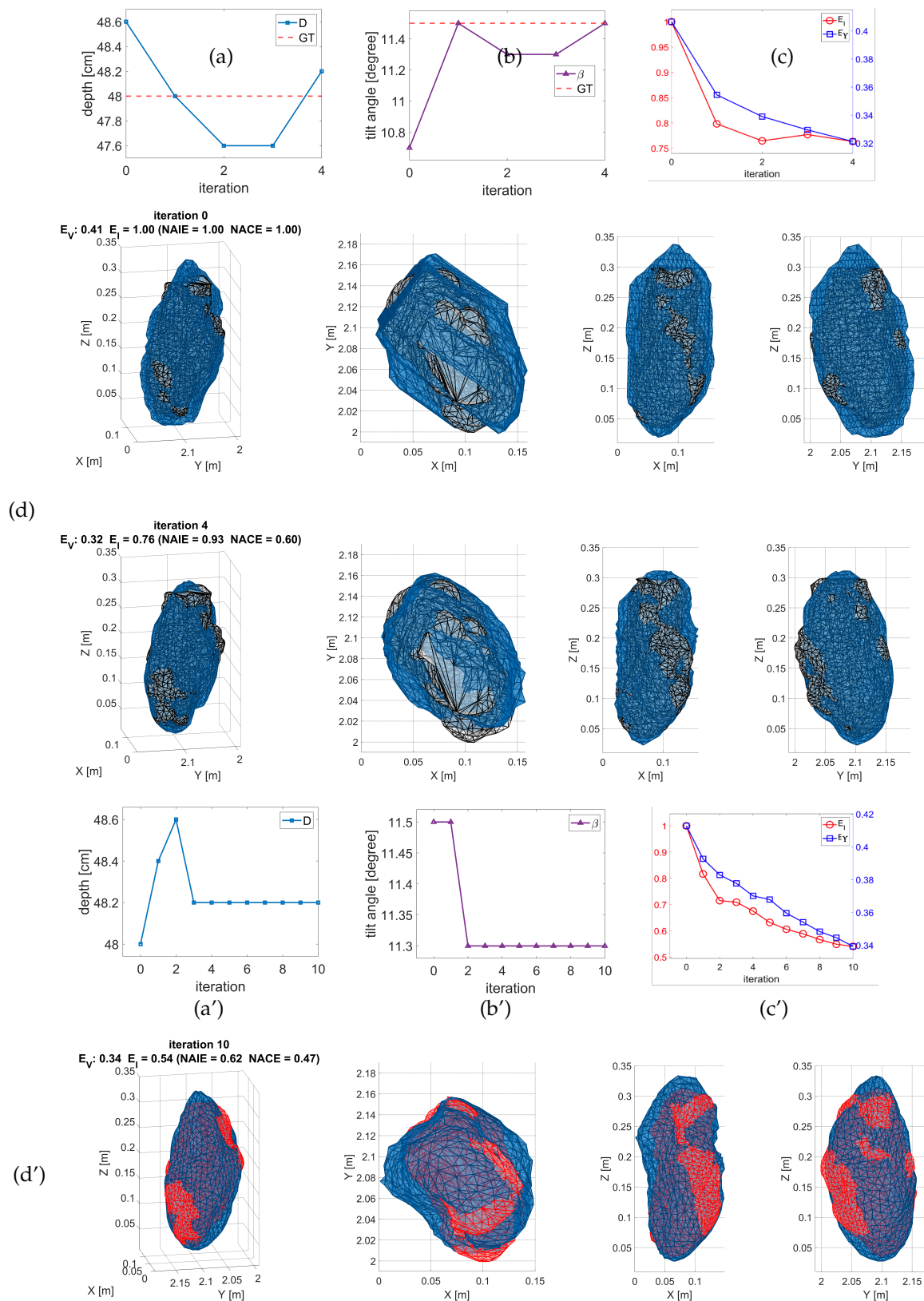


Figure 8. Coral-two experiment—(a–d) synthetic and (a'–d') real data. (a,b,a',b') Optimization of sonar depth and tilt parameters. (c,c') Improving 3-D model leads to smaller volumetric E_V and image E_I errors. (d) Kinect model (black mesh) superimposed on initialized SC solution (top) and optimized 3-D model (bottom), shown by blue surface meshes. (d') Optimized SC (blue mesh) and Kinect (red mesh) models.

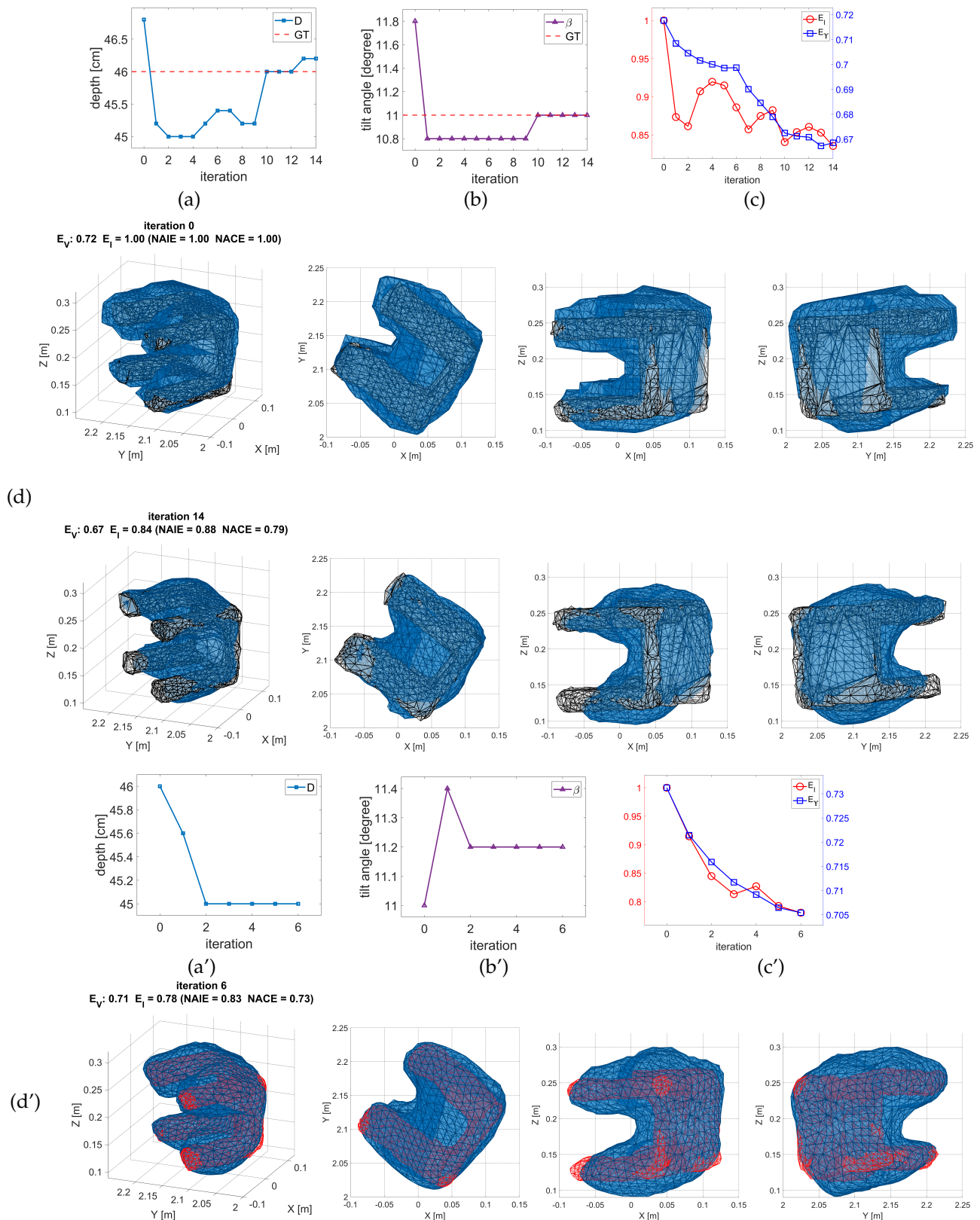


Figure 9. Wood table experiment—(a–d) synthetic and (a'–d') real data. (a,b,a',b') Optimization of sonar depth and tilt parameters. (c,c') Improving 3-D model reduces both volumetric E_V and image E_I errors. (d) Kinect model (black mesh) superimposed on initialized SC solution (top) and optimized 3-D model (bottom), shown by blue surface meshes. (d') Optimized SC (blue mesh) and Kinect (red mesh) models.

Next, we present sample results of target image enhancement. For the first coral rock, Figure 10 depicts three samples, out of five rotated views used in the optimization at S (pos 1) and N (pos 2) sonar positions. Columns (a,b) include the original data and the processed ghost-free image, respectively. Columns (c,d) are the synthetic images constructed from the initial SC and optimized 3-D models, respectively. We can readily confirm that the consistency within the object region in the data and synthetic images improves in almost all views after optimization. The same conclusions are drawn from the results for the second coral rock in Figure 11 and the wood table images in Figure 12 (despite the lower 3-D reconstruction accuracy of the optimized solution).

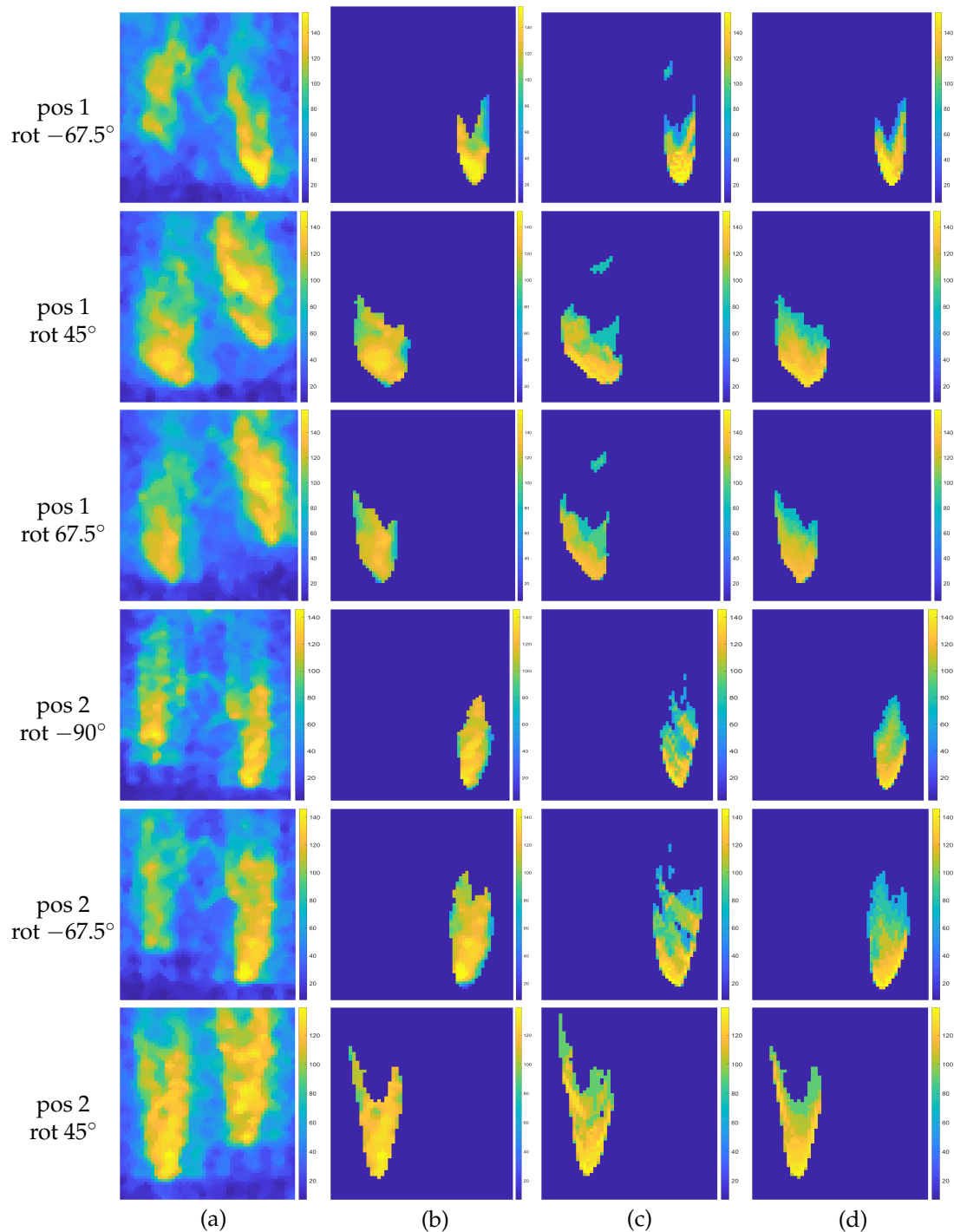


Figure 10. Coral-one experiment—(a) data; (b) data over image region only; (c) initial and (d) optimized synthetic view generated by the 3-D model.

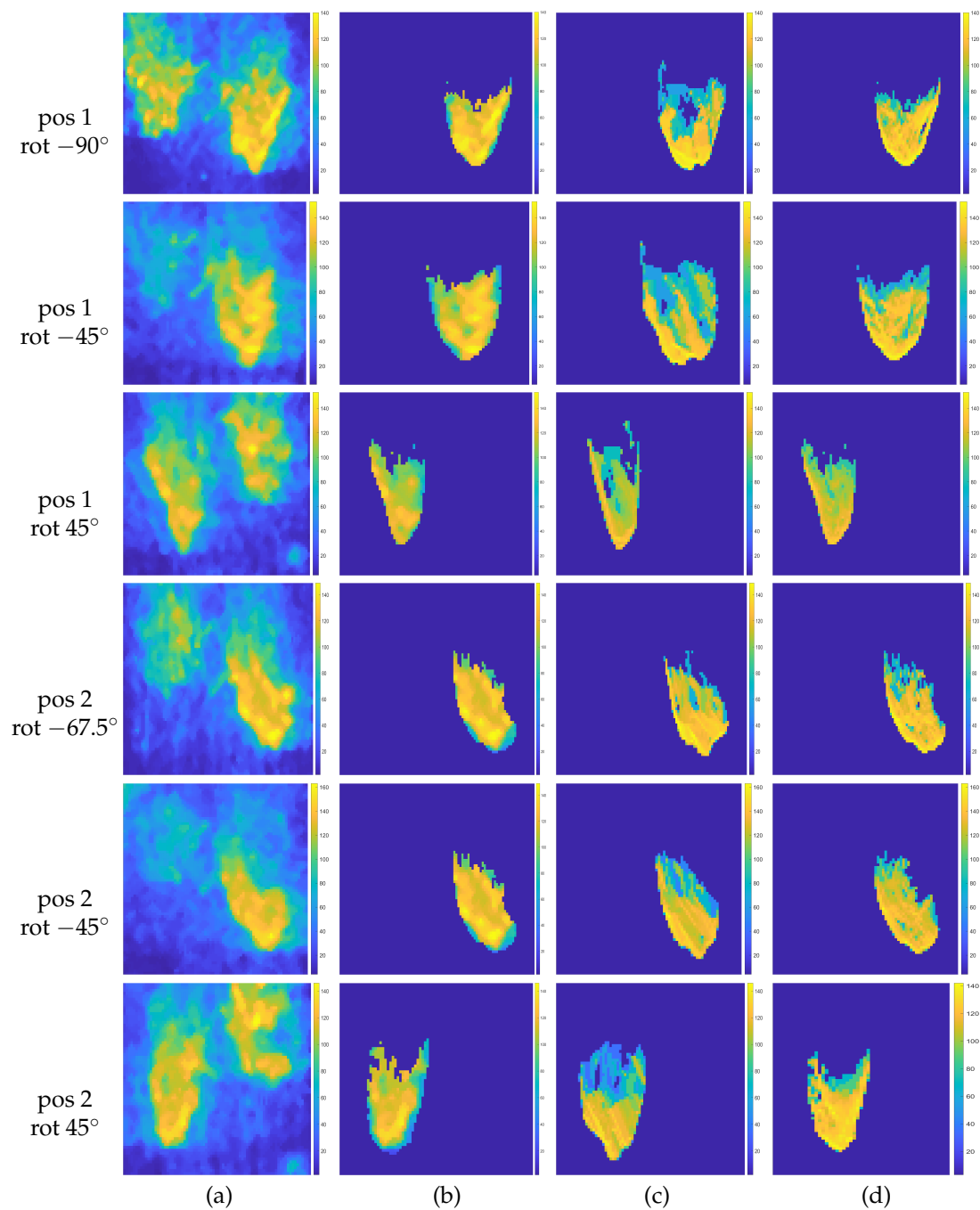


Figure 11. Coral-two experiment—(a) data; (b) data over image region only; (c) initial and (d) optimized synthetic view generated by the 3-D model.

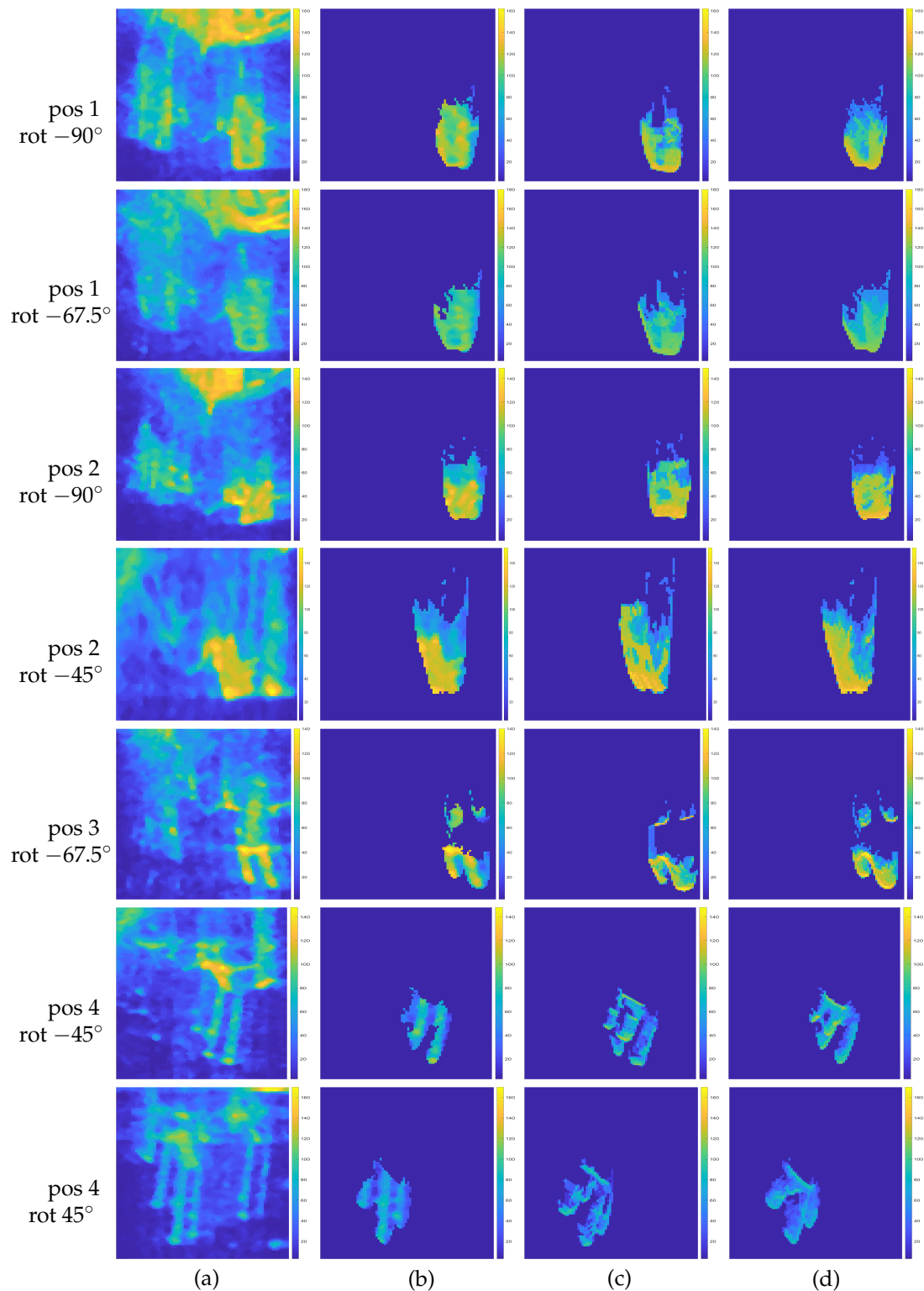


Figure 12. Wood table experiment—(a) data; (b) data within object region only; (c) initial and (d) optimized synthetic view generated by the 3-D model.

Sample sonar images at zero and $\pm 22.5^\circ$ roll angles (which are not used in the optimization) are found in Figures 13 and 14. Here, results in the last column can be treated as synthetic views at new arbitrary sonar poses and can be compared with the ghost-free data in column (b) to assess accuracy in the presence of unmodeled environmental factors (of the original data).

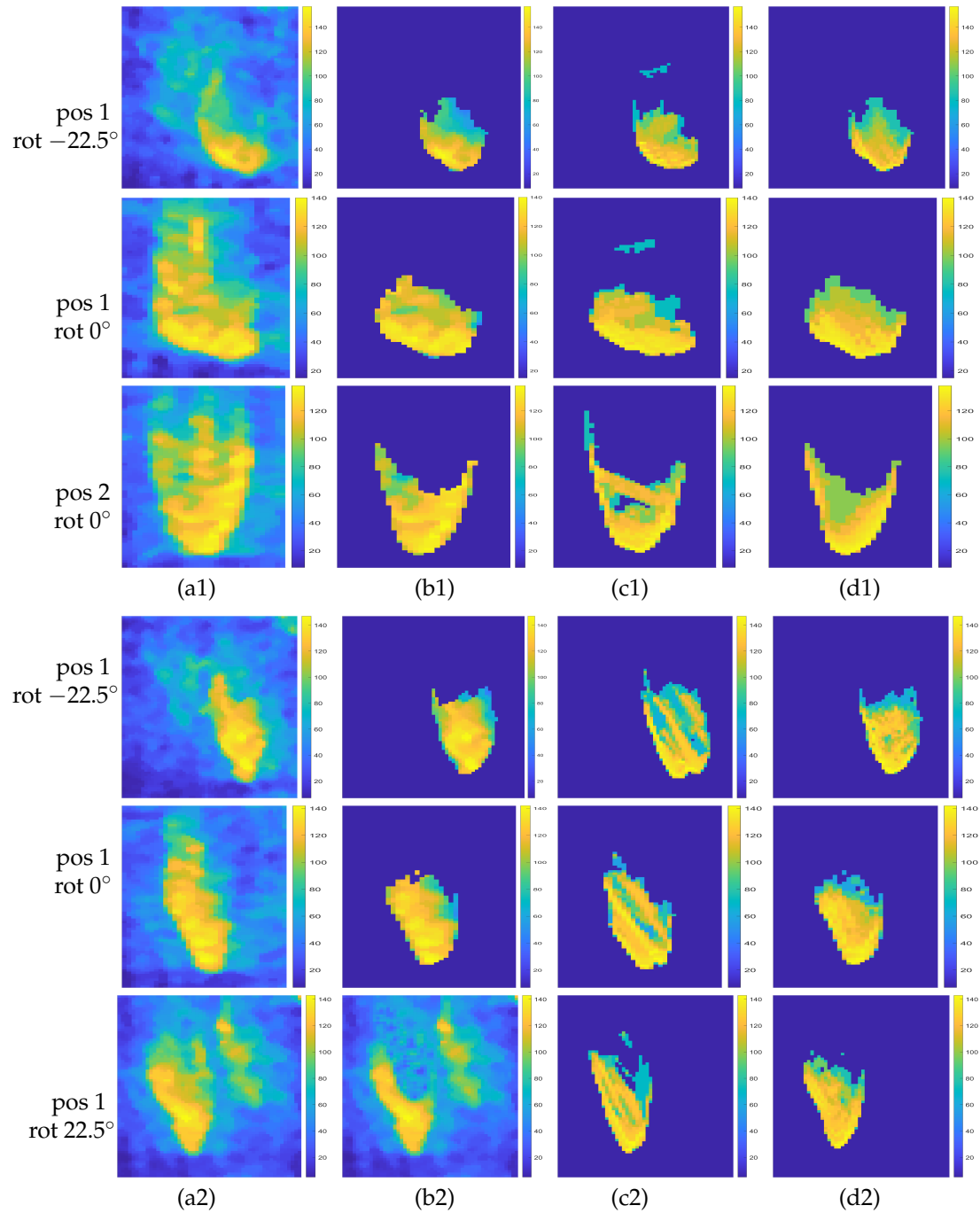


Figure 13. Sets of images as in previous experiments for (a1–d1) coral-one and (a2–d2) coral-two views, in which object, ghost, and mirror components overlap (not used in the optimization). (a1,a2) data; (b1,b2) data within object region only; (c1,c2) initial and (d1,d2) optimized synthetic view generated by the 3-D model.

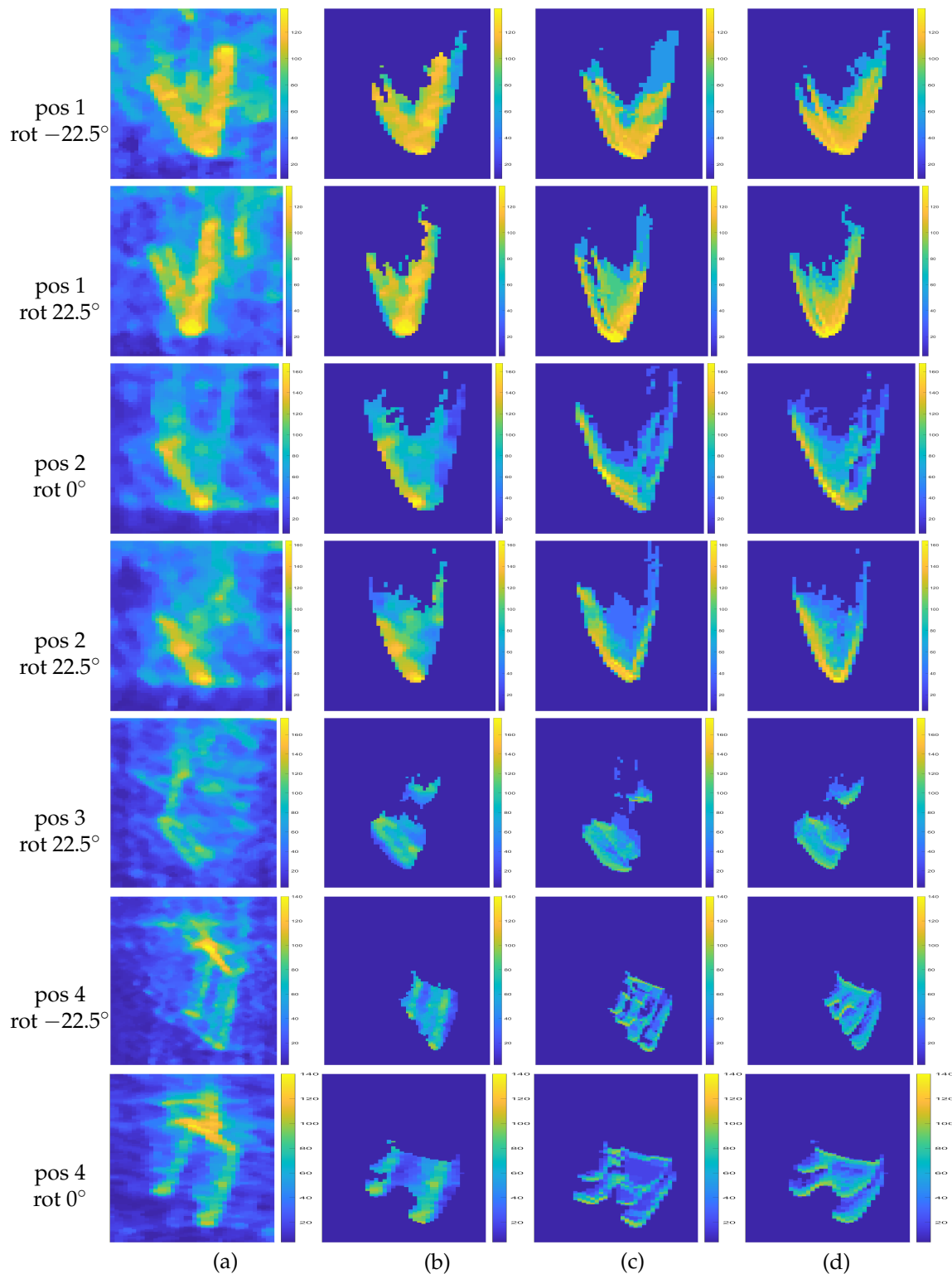


Figure 14. Sets of wood table images as in previous figures for views in which object, ghost, and mirror components overlap (not used in optimization process). (a) data; (b) data within object region only; (c) initial and (d) optimized synthetic view generated by the 3-D model.

4. Conclusions

In this work, we address the reconstruction of a 3-D object model from 2-D FSS images at known poses, with multipath propagation due to sea surface reflectance. To this end, we propose a unified optimization framework to improve the accuracy of the two key sonar pose parameters (measured by low-cost sensors) and the 3-D model. Embedded in the

approach to localize and remove the ghost images formed by multipath is the generation of synthetic views from the estimated 3-D object model based on the environmental conditions. This has a number of applications, including the generation of a large volume of training data under a variety of environmental conditions for object recognition. Moreover, a LUT is generated from the distributions of object intensities within the non-overlap region in the real and synthetic images and utilized in generating the object within the ghost-corrupted overlap region.

The targets in three experiments with synthetic and real data are two dominantly convex coral rocks and a concave toy wood table. Assuming the same setup as for the real data, computer simulations with synthetic data highlight the accuracy and convergence in 10 or fewer iterations of the estimated pose parameters and the 3-D target model. Comparable performances are achieved in experiments with real data.

The results of the wood table experiment demonstrate the applicability of our method in constructing shapes with deep concavities, albeit at lower accuracy. Here, the primary complexity is the inability to account for the shape-dependent multipath without a priori knowledge of the target shape. The multipath reverberations from object surfaces within concavities lead to the thickened appearances of object parts. To decouple the direct and multipath components, we require knowledge of the unknown 3-D object model. This poses a chicken-and-egg problem, which is faced by most other methods for 3-D object reconstruction from FSS imagery. Aside from collecting a much larger volume of data from different poses looking into concavities at close ranges, further investigation is needed for devising a suitable mechanism to accurately account for the multipath generated by the surface patches within concavities.

Author Contributions: Conceptualization, Y.L. and S.N.; methodology, Y.L. and S.N.; software, Y.L. and S.N.; Validation, Y.L. and S.N.; formal analysis, Y.L. and S.N.; investigation, Y.L. and S.N.; data curation, Y.L.; writing—original draft preparation, Y.L.; writing—review and editing, Y.L. and S.N.; visualization, Y.L. and S.N.; funding acquisition, S.N. All authors have read and agreed to the published version of the manuscript.

Funding: The research is based on work initially supported by the Office of Naval Research under grant N00014-15-1-2089.

Data Availability Statement: The data for this study were collected with a DIDSON unit purchased using funds from the ONR DURIP award N00014-06-1-0765. The distribution of the data is subject to the University of Miami policies for public sharing of the data.

Acknowledgments: The authors acknowledge support from the Herbert Wellness Center at the University of Miami for the use of the indoor pool facility for data acquisition.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Available online: <https://oceanexplorer.noaa.gov/edu/materials/multibeam-sonar-fact-sheet.pdf> (accessed on 5 October 2024).
2. Available online: https://www.uio.no/studier/emner/matnat/ifi/INF-GEO4310/h12/undervisningsmateriale/sonar_introduction_2012_compressed.pdf (accessed on 5 October 2024).
3. Available online: https://en.wikipedia.org/wiki/Side-scan_sonar (accessed on 5 October 2024).
4. Burguera, A.; Oliver, G. High-resolution underwater mapping using side-scan sonar. *PLoS ONE* **2016**, *11*, e0146396. [CrossRef] [PubMed]
5. Hansen, R. E. Synthetic aperture sonar technology review. *Marine Technology Society J.* **2013**, *47*, 117–127. [CrossRef]
6. Hayes, M.P.; Gough, P.T. Synthetic aperture sonar: A review of current status. *IEEE J. Ocean. Eng.* **2009**, *34*, 207–224. [CrossRef]
7. Available online: <https://www.sonardyne.com/applications/obstacle-avoidance/#:~:text=When%20your%20marine%20operation%20takes,warn%20of%20potential%20collision%20hazards> (accessed on 5 October 2024).
8. Available online: <https://www.simrad-yachting.com/sonar-and-transducers/forwardscan-sonar/> (accessed on 5 October 2024).
9. Available online: <https://www.teledynemarine.com/en-us/products/product-line/Pages/forward-looking-sonars.aspx> (accessed on 5 October 2024).
10. Available online: <http://www.soundmetrics.com/Products/DIDSON-Sonars> (accessed on 5 October 2024).

11. Available online: <https://www.seascapesubsea.com/product/oculus-m750d/#:~:{}%5Cprotect%5Cprotect%5Cleavevmode@ifvmode%5Ckern+.1667em%5Crelax:text=Description,horizontal%20fields%20of%20view%20%5Crespectively> (accessed on 5 October 2024).
12. Available online: <https://www.tritech.co.uk/products/gemini-720is> (accessed on 5 October 2024).
13. Ferreira, F.; Djapic, V.; Micheli, M.; Caccia, M. Forward looking sonar mosaicing for mine countermeasures. *Annu. Rev. Control.* **2015**, *40*, 212–226. [\[CrossRef\]](#)
14. Hurtos, N.; Petillot, Y.; Salvi, J. Fourier-based registrations for two-dimensional forward-looking sonar image mosaicing. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'12), Vilamoura-Algarve, Portugal, 7–12 October 2012; pp. 5298–5305.
15. Hurtos, N.; Nagappa, S.; Palomeras, N.; Salvi, J. Real-time mosaicing with two-dimensional forward-looking sonar. In Proceedings of the IEEE International Conference Robotics and Automation (ICRA'14), Hong Kong, China, 31 May–7 June 2014; pp. 601–606.
16. Kim, K.; Intrator, N.; Neretti, N. Image registration and mosaicing of noisy acoustic camera images. In Proceedings of the IEEE International Conference on Electronics, Circuits and Systems, (ICECS'04), Tel Aviv, Israel, 15 December 2004; pp. 527–530.
17. Negahdaripour, S.; Aykin, M.D.; Sinnarajah, S. Dynamic scene analysis and mosaicing of benthic habitats by FS sonar imaging—Issues and complexities. In Proceedings of the OCEANS'11 MTS/IEEE, Waikoloa, HI, USA, 19–22 September 2011.
18. Geiger, A. Computer Vision - Lecture 8.1 (Shape-from-X: Shape-from-Shading). Available online: <https://www.youtube.com/watch?v=YQ5QOiyF9U> (accessed on 5 October 2024).
19. Liu, Y. 3-D Object Modeling from 2-D Underwater Forward-Scan Sonar Imagery in the Presence of Multipath near Sea Surface. Master's Thesis, University of Miami, Coral Gables, FL, USA, 2022.
20. Liu, Y.; Negahdaripour, S. Object modeling from underwater forward-scan sonar imagery with sea-surface multipath. *IEEE J. Ocean. Eng.* **2024**, 1–14. [\[CrossRef\]](#)
21. Cho, H.; Kim, B.; Yu, S. Auv-based underwater 3-D point cloud generation using acoustic lens-based multibeam sonar. *IEEE J. Ocean. Eng.* **2018**, *43*, 856–872. [\[CrossRef\]](#)
22. Guerneve, T.; Subr, K.; Petillot, Y. Three-dimensional reconstruction of underwater objects using wide-aperture imaging sonar. *J. Field Robot.* **2018**, *35*, 890–905. [\[CrossRef\]](#)
23. Westman, E.; Gkioulekas, I.; Kaess, M. A volumetric albedo framework for 3d imaging sonar reconstruction. In Proceedings of the IEEE International Conference on Robotics and Automation, Paris, France, 31 May–31 August 2020; pp. 9645–9651.
24. Zerr, B.; Stage, B. Three-dimensional reconstruction of underwater objects from a sequence of sonar images. In Proceedings of the 3rd IEEE International Conference on Image Processing, Lausanne, Switzerland, 19 September 1996; Volume 3, pp. 927–930.
25. Wang, Y.; Ji, Y.; Liu, D.; Tsuchiya, H.; Yamashita, A.; Asama, H. Elevation angle estimation in 2d acoustic images using pseudo front view. *IEEE Robot. Autom. Lett.* **2021**, *6*, 1535–1542. [\[CrossRef\]](#)
26. Aykin, M.D.; Negahdaripour, S. Modeling 2-D lens-based forward-scan sonar imagery for targets with diffuse reflectance. *IEEE J. Ocean. Eng.* **2016**, *41*, 569–582. [\[CrossRef\]](#)
27. Wang, Y.; Wu, C.; Ji, Y.; Tsuchiya, H.; Asama, H.; Yamashita, A. 2d forward looking sonar simulation with ground echo modeling. *arXiv* **2023**, arXiv:2304.0814.
28. Kraus, F.; Scheiner, N.; Dietmayer, K. Using machine learning to detect ghost images in automotive radar. In Proceedings of the IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC'20), Rhodes, Greece, 20–23 September 2020.
29. Kraus, F.; Scheiner, N.; Ritter, W.; Dietmayer, K. The radar ghost dataset—An evaluation of ghost objects in automotive radar data. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, 27 September–1 October 2021; pp. 8570–8577.
30. Liu, C.; Liu, S.; Zhang, C.; Huang, Y.; Wang, H. Multipath propagation analysis and ghost target removal for FMCW automotive radars. In Proceedings of the IET International Radar Conference, Online, 4–6 November 2020; pp. 330–334.
31. Longman, O.; Villeval, S.; Bilik, I. Multipath ghost targets mitigation in automotive environments. In Proceedings of the IEEE Radar Conference, Atlanta, GA, USA, 7–14 May 2021; pp. 1–5.
32. Swl Marine Bronze Submersible Level Sensor. Available online: <https://www.sensorsone.com/swl-marine-bronze-submersible-level-sensor/> (accessed on 5 October 2024).
33. Tdk-TmrSensor. Available online: <https://product.tdk.com/en/techlibrary/productoverview/tmr-angle-sensors.html/> (accessed on 5 October 2024).
34. Barrault, G. Modeling the forward Look Sonar. Master's Thesis, Florida Atlantic University, Boca Raton, FL, USA, 2001.
35. Choi, W.S.; Olson, D.; Davis, D.; Zhang, M.; Racson, A.; Bingham, B.; McCarrin, M.; Vogt, C.; Herman, J. Physics-based modeling and simulation of multibeam echosounder perception for autonomous underwater manipulation. *Front. Robot. AI* **2021**, *8*, 706646. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Liu, D.; Wang, Y.; Ji, Y.; Tsuchiya, H.; Yamashita, A.; Asama, H. Cyclegan-based realistic image dataset generation for forward-looking sonar. *Adv. Robot.* **2021**, *35*, 242–254. [\[CrossRef\]](#)
37. Negahdaripour, S. Calibration of DIDSON forward-scan acoustic video camera. In Proceedings of the OCEANS 2005 MTS/IEEE, Washington, DC, USA, 17–23 September 2005.
38. Kutulakos, K.N.; Seitz, S.M. A theory of shape by space carving. *Int. J. Comput. Vis.* **2000**, *38*, 199–218. [\[CrossRef\]](#)
39. Aykin, M.D.; Negahdaripour, S. Three-dimensional target reconstruction from multiple 2-D forward-scan sonar views by space carving. *IEEE J. Ocean. Eng.* **2017**, *42*, 574–589. [\[CrossRef\]](#)

40. Negahdaripour, S.; Milenkovic, V.M.; Salarieh, N.; Mirzargar, M. Refining 3-D object models constructed from multiple fs sonar images by space carving. In Proceedings of the IEEE/MTS Oceans Conference-Anchorage, Anchorage, AK, USA, 18–21 September 2017; pp. 1–9.
41. Gu, J.-H.; Joe, H.-G.; Yu, S.-C. Development of image sonar simulator for underwater object recognition. In Proceedings of the OCEANS'13, San Diego, CA, USA, 23–27 September 2013.
42. Kwak, S.; Ji, Y.; Yamashita, A.; Asama, H. Development of acoustic camera-imaging simulator based on novel model. In Proceedings of the IEEE 15th International Conference on Environment and Electrical Engineering (EEEIC'15), Rome, Italy, 10–13 June 2015; pp. 1719–1724.
43. Potokar, E.; Lay, K.; Norman, K.; Benham, D.; Neilsen, T.B.; Kaess, M.; Mangelson, J.G. HoloOcean: Realistic sonar simulation. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Kyoto, Japan, 23–27 October 2022.
44. Besl, P.J.; McKay, N.D. A method for registration of 3-D shapes. *IEEE T. Pattern Anal. Mach. Intell.* **1992**, *14*, 239–256. [[CrossRef](#)]
45. Chen, Y.; Medioni, G. Object modeling by registration of multiple range images. *Image Vis. Comput.* **1992**, *10*, 145–155. [[CrossRef](#)]
46. Bergström, P.; Edlund, O. Robust registration of point sets using iteratively reweighted least squares. *Comput. Optim. Appl.* **2014**, *58*, 543–561. [[CrossRef](#)]
47. Bergström, P.; Edlund, O. Robust registration of surfaces using a refined iterative closest point algorithm with a trust region approach. *Numer. Algorithms* **2017**, *74*, 755–779. [[CrossRef](#)]
48. Ruder, S. An overview of gradient descent optimization algorithms. *arXiv* **2017**, arXiv:1609.04747v2.
49. Azure Kinect DK. Available online: <https://www.microsoft.com/en-us/d/azure-kinect-dk/8pp5vxmd9nhq?activetab=pivot:overviewtab> (accessed on 5 October 2024).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.