Underwater Imaging with a Moving Acoustic Lens

Behzad Kamgar-Parsi, *Member, IEEE,* Lawrence J. Rosenblum, *Senior Member, IEEE,* and Edward O. Belcher, *Member, IEEE*

***Abstract*— The acoustic lens is a high-resolution, forward-looking sonar for three-dimensional (3-D) underwater imaging. In this paper, we discuss processing the lens data for recreating and visualizing the scene. Acoustical imaging, compared to optical imaging, is sparse and low resolution. To achieve higher resolution, we obtain a denser sample by mounting the lens on a moving platform and passing over the scene. This introduces the problem of data fusion from multiple overlapping views for scene formation, which we discuss. We also discuss the improvements in object reconstruction by combining data from several passes over an object. We present algorithms for pass registration and show that this process can be done with enough accuracy to improve the image and provide greater detail about the object. The results of in-water experiments show the degree to which size and shape can be obtained under (nearly) ideal conditions.**

***Index Terms*— Acoustic imaging, acoustic lens, pass registration, 3-D underwater imaging.**

I. Introduction

THE CAPABILITY to image underwater objects with high resolution is important in many scientific and engineering applications. Optical cameras and lasers provide high-resolution images that can be easily interpreted, but their visibility is limited to distances of no more than tens of meters in clear water. They fail at centimeter ranges in turbid water, a common condition in coastal waters and in waters disturbed by people. Acoustic signals, however, propagate in turbid water with little degradation. Thus, acoustical imaging becomes the principal means of sight in this environment. In some cases, such as in studying ocean floor hydrothermal plumes, sonars are the preferred imaging devices, since acoustic wave propagation and scattering are much more sensitive to variations in temperature, salinity, etc.

Conventional sonars operate at longer ranges than optical instruments. Low-frequency mapping sonars survey wide swaths of ocean bottom, but only resolve features larger than tens of meters. As the acoustic frequency rises, resolution improves but range decreases. Sidescan sonars can produce inches resolution on a two-dimensional (2-D) projection, but do not directly resolve height information. The acoustic lens, described in Section II, is a new high-frequency, forward-looking sonar developed especially for three-dimensional (3-D) underwater imaging. It has high-resolution (for the ocean), and is best suited for close-up imaging (distances of several meters). The lens, which is a delay-and-sum beamformer, eliminates the need for beamforming electronics, an advantage when size, weight, and power consumption of the sensor are important.

The acoustic lens and other high-resolution imaging sonars are still low resolution compared to optical imaging devices. For example, the lens prototype discussed in this paper has a resolution of 25 cm at a range of 10 m. Thus, taking a single snapshot of the scene yields a sparse, low-resolution image. To obtain a denser data set, and improve the resolution, we mount the lens on a moving platform and make a pass (or several passes) over the scene. This data collection strategy introduces the problem of sensor position information, and the attendant difficulties of registering and combining data. In this work, the lens is mounted on a carriage in a tow tank and moves with a known velocity. Thus, highly accurate positional data for the lens is available, a luxury lacking in most underwater operations. These experiments serve to demonstrate the limits of what can be achieved by this sensor mounted on a moving platform.

The acoustic lens outputs 3-D gray-level data with nonuniform resolution. The grayness is the intensity of acoustic backscatter from surfaces. In this paper, we discuss the techniques for the various steps that are involved in obtaining clear 3-D images from the raw lens data. In Section III, we discuss scene reconstruction techniques. In Section IV, we discuss filtering and rendering algorithms, and present the results of controlled in-water experiments in a tow tank. The object sizes and their features in the experiments are of the order of one meter to a few centimeters. They are viewed from distances of several meters. We also discuss combining images taken in several passes to improve the resolution of the reconstructed objects. Sensor position data in a pass relative to other passes are not accurately known even in the controlled experiments conducted here. In Section V, we develop appropriate registration techniques to align and combine data from several passes.

II. The Acoustic Lens

Traditional sonar systems use mechanical or electrical beamforming techniques to scan a highly directional beam over a field of view. Acoustic lens technology, on the other hand, forms high-resolution, conical beams by focusing sound on small transducers populating a retina (see Fig. 1). Transmission time delays determine range, while transducer position yields bearing and elevation coordinates. Lens technology [4]

Manuscript received January 26, 1996; revised April 2, 1997. This work was supported by the Office of Naval Research. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Hua Lee.

B. Kamgar-Parsi and L. J. Rosenblum are with the Naval Research Laboratory, Washington, D.C. 20375 USA (e-mail: behzad@ait.nrl.navy.mil).

E. O. Belcher is with the Applied Physics Laboratory, University of Washington, Seattle, WA 98105 USA.

Publisher Item Identifier S 1057-7149(98)00336-4.

1057–7149/98$10.00 © 1998 IEEE

Authorized licensed use limited to: NRL. Downloaded on December 1, 2009 at 15:29 from IEEE Xplore. Restrictions apply.

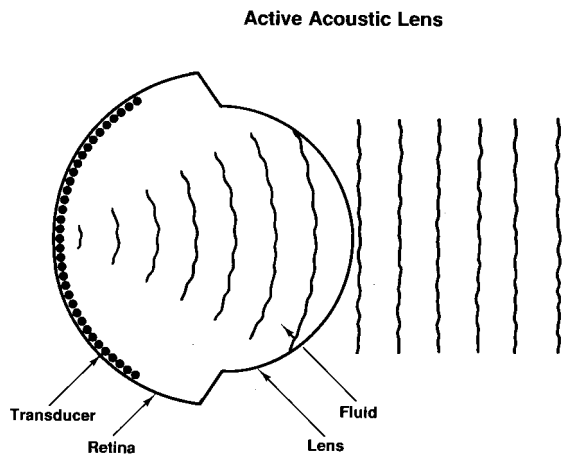

Fig. 1. Lens focusing diagram.

is not new, but recent advances in digital electronics now permit dealing with a large volume of data in real time. The acoustic lens prototype discussed in this paper is described in [2]. We include a brief description here.

The lens retina is populated with transducers in eight rows of 16 elements each. The transducers in each row are separated and the rows staggered to produce 1.5° resolution beams at the 3 dB points in both azimuth and elevation. The hemispherical retina and the acoustically transparent hemispherical lens window together form a cavity that is filled with a fluid. The fluid has a speed approximately 60% of the sound speed in water, and refracts the sound entering the window to focus on the retina. This lens has a field of view of 48° in azimuth and 12° in elevation. Its operating frequency is 300 kHz, which corresponds to a wavelength of 0.5 cm in water. Each beam has a conical beam pattern for transmit and receive. The range bins, set by the sampling rate, are about 10 cm for ranges up to about 100 m. The sensor is designed to focus in an ambient temperature range that spans 13° Celsius without having to change the lens fluid.

In the active mode, a row of transducers on the lens retina simultaneously emit a pulse and then receive and record the energy of the backscattered waves. There are two sources of scattering; the first is volume scattering caused by the inhomogeneities in the water due to air bubbles, density fluctuations of colloidal particles, etc. The second source is scattering from interfaces such as water-seafloor, or object surfaces, where there is a significant change in the index of refraction. Volume scattering is usually much weaker than surface scattering and in processing of the lens data can be filtered out with thresholding. The signals received by each transducer are first corrected for beam spreading (spherical spreading in deep water, or cylindrical spreading in shallow water), so that the recorded backscatter signal (in dB) [21] is proportional to

$$E = 10 \log \left(S \int dA \, B_T(\theta, \phi) \, B_R(\theta, \phi) \right).$$

S is the coefficient of surface scattering, dA is the area element on the scattering surface, B_T and B_R are transmit and receive beam patterns, and θ and ϕ are the polar and azimuth angles of the area element dA with respect to the transducer axis. B_T has a fanlike pattern composed of 16 narrow cones, while B_R is a single narrow cone. As we see from the expression for E, the backscatter signal depends on the surface material, as well as the scattering angle, so that a plastic surface ensonified at near normal incidence angles appears like a metallic surface ensonified at a slant.

A sonar beam is strongly backscattered whenever it crosses an interface separating two media with significantly different indices of refraction. For example, if a metallic object is ensonified by a sonar, the beam is backscattered both as it enters and exits the object. The backscatter from the front surface is obviously stronger, but the return signal from the rear surface can also be significant and contain useful information. The acoustic lens senses and retains both reflections. This is unlike threshold sonars often used for robot navigation, where the information after the first strong return is discarded. The acoustic lens also uniquely localizes the backscatter signal, unlike 3-D interferometric sonars, which may yield ambiguous elevations [19]. These advantages make the lens a unique sonar for viewing 3-D underwater scenes.

III. Scene Reconstruction from a Single Pass

For scene reconstruction, we partition the space into a 3-D grid of cubic voxels. This regular voxel-based volumetric representation of the environment is well suited for transmission, image processing, and manipulations by most graphics algorithms. We wish to estimate the return signal strength from within each voxel. High voxel values indicate the presence of objects or interfaces at those locations.

The acoustic lens has a constant resolution, Δ, along range which is set by the sampling rate. Its lateral resolution, however, depends on range, ρ, as $\rho \tan \gamma$, where $\gamma = 1.5°$ is the angle of the conical receive beam pattern. For example, at 1 m range the lateral resolution is 2.5 cm, and at 10 m range it degrades to 25 cm. Hence, keeping the lens stationary and taking a snapshot produces data with inadequate resolution. Since we wish to image small objects and resolve small features, then the scenario for imaging a scene is constantly to move the lens as it continually pings and collects backscatter signals and, if possible, to make several passes over the scene. In this manner, one obtains a dense set of data with much greater resolution. The issue of correcting for navigational errors and misregistration of passes is a difficult problem in its own [10]. In this section, we assume the exact position and orientation of the lens are known at all times with respect to a global coordinate system fixed to the scene. We discuss registering and combining data from multiple passes in Section V.

Estimating voxel values involves two stages. First, since each range bin intersects several voxels its backscatter must be distributed among them. Second, a voxel is typically ensonified by several beams, resulting in different estimates that must be combined to yield a single value. Algorithms for distributing returns involve modeling the beam pattern and computing range bin intersections with voxels. This is conceptually straightforward, but it can be computationally intensive.

Authorized licensed use limited to: NRL. Downloaded on December 1, 2009 at 15:29 from IEEE Xplore. Restrictions apply.

Algorithms for combining estimates can be nontrivial because we may have to deal with occlusions, partial ensonifications, and differences in angle of incidence. In this section, we discuss several techniques for distributing and combining the data.

The voxel size is also an important consideration. As we noted, because of its conical beams, the acoustic lens produces data with varying spatial resolutions. In the near range, it has high resolution, but as the range increases the volume of range bins increase as range-squared. In multiple-pass scenarios the same part of the scene may be sampled with different resolutions. Since the scene is not sampled with uniform resolution, one may have to choose different voxel sizes for reconstructing different regions. Experience guides us to choose voxels with volumes comparable to the volume of the range bins in the region(s) of interest; also the number of voxels spanning a given region should be roughly comparable to the number of data points within that region.

For gridding the data, the range bins must be mapped onto the scene coordinate frame. This is done by first transforming the range data (ρ, θ, ϕ) in the polar coordinate system to the Cartesian coordinate system (X, Y, Z) attached to the lens. The elevation angle, θ, and the azimuth, ϕ, are known from the row and column of the transducer on the lens retina. Then we transform (X, Y, Z) to the global Cartesian coordinates (x, y, z) fixed to the scene, which requires sensor position data.

A. Distributing Single Returns

The backscatter energy received in a range bin is the sum of all returns from everywhere in that bin. Since a range bin typically overlaps with several voxels, the energy must be distributed among those voxels. Denote the backscatter energy of range bin α by f_α, and the contribution of voxel i to this backscatter energy by $g_{i\alpha}$. The general scheme for estimating $g_{i\alpha}$ is

$$g_{i\alpha} = f_\alpha w_{i\alpha} \Big/ \sum_j w_{j\alpha} \tag{1}$$

where the weight normalization sum is over all voxels that get a share of the backscatter signal of range bin α. The problem is to assign appropriate weights to voxels.

In the absence of other information, it is reasonable to assume that the backscatter signal coming from a voxel is proportional to its volume overlap with the range bin, convolved with the receive beam pattern profile. Thus we must find all voxels that overlap with range bin α; compute the volume overlap of each of these voxels with the range bin; convolve it with the beam pattern to obtain the effective overlap volume $V_{i\alpha}$; and set $w_{i\alpha} = V_{i\alpha}$. However, computing the exact volume overlap of two solid objects when they are positioned arbitrarily is a nontrivial task, even for simple geometries such as a cube (the voxel) and a spherical cone slice (the range bin). We used the following approach: divide the cubic voxel into a large number of small cubic cells, say N^3 cells; determine if the center of each cell falls inside the conical slice, then the volume overlap is proportional to the number of *in cells* to N^3. This is computationally expensive, since the cost goes up as N^3 with the desired accuracy. Yet the results are not much different from other schemes commonly used. Below we discuss two such algorithms.

Algorithm 1: Consider the grid made up of voxel centers. Find the *cell* in which the center of range bin α, $\vec{r}_\alpha$, falls. Distribute the backscatter energy f_α among the eight grid points (or voxels) forming the cell according to their distances from the range bin center, i.e., $w_{i\alpha} = 1/\|\vec{r}_i - \vec{r}_\alpha\|^\nu$, where $\vec{r}_i$ is the center of voxel i, and ν is a positive integer (we used $\nu = 2$).

Algorithm 2: Find the voxel inside which the range bin center falls, and assign all of the backscatter energy to that voxel, i.e., $g_{i\alpha} = f_\alpha$.

Algorithm 1 is only slightly slower than Algorithm 2, but the results are more satisfactory. In the experiments we have analyzed, the scenes reconstructed using Algorithm 1 agrees better with the ground truth, and appear less noisy because the scheme performs a degree of smoothing. Furthermore, Algorithm 2 is sensitive to the positioning of the voxel grid. For example, if we displace the grid $s/2$ along any direction (s is the grid size), the reconstructed scene becomes noticeably distorted. This effect is present also in Algorithm 1 but to a far lesser degree, especially if the voxel size is chosen appropriately, i.e., comparable to the size of the range bins. Note that both algorithms implicitly assume that range bin sizes are constant. If we wish to take into account the variable range bin size we may use the Gaussian weighting function

$$w_{i\alpha} \propto \exp\left[-(\vec{r}_i - \vec{r}_\alpha)^2/2\sigma^2\right]$$

with σ^3 chosen proportional to the range bin volume. The differences between these schemes become less pronounced as the scene is sampled more densely. The figures presented in this paper have been produced using Algorithm 1.

B. Combining Multiple Estimates

Each voxel typically receives contributions from several range bins, and the question is how to estimate the backscatter energy g_i from voxel i from the set of returns $\{g_{i\alpha}\}$. Unlike evaluating $g_{i\alpha}$, estimating g_i is not straightforward, since the voxel values from individual range bins may have been obtained under very different conditions. The following two extremes illustrate this point.

Consider the empty voxel i, which is near the surface of an object. Suppose a particular range bin in beam 1 that covers this voxel has no overlaps with the object and has zero return. One would estimate that $g_{i1} = 0$. Suppose a given range bin in beam 2 covers voxel i and overlaps with the object. Since the return is significant one would estimate that the voxel is nonempty, i.e., $g_{i2} \neq 0$, and so on, with other beams that ensonify voxel i. In this case

$$g_i = \min_\alpha \{g_{i\alpha}\}$$

yields the correct estimate, namely that voxel i is empty. Now consider the case where voxel i is on the surface of an object. Suppose a given range bin in beam 1, which covers the voxel, views the object at normal angle. Since the return is strong, we would assign a high value to g_{i1}. Suppose a given range bin

Authorized licensed use limited to: NRL. Downloaded on December 1, 2009 at 15:29 from IEEE Xplore. Restrictions apply.

in beam 2 also covers the voxel but views the object surface at a near tangential angle. Since the return is weak we would assign a near zero value to g_{i2}, and so on. In this case

$$g_i = \max_{\alpha}\{g_{i\alpha}\}$$

yields the best estimate, that is voxel i is nonempty.

As a simple compromise, we treat all gridded backscatter values $g_{i\alpha}$ identically and take their unbiased average

$$g_i = \frac{1}{n_i} \sum_{\alpha=\{n_i\}} g_{i\alpha} \tag{2}$$

to be the voxel estimate. Here n_i is the number of *hits* received by voxel i. For the experiments we have analyzed, this scheme yielded results that agreed well with the ground truths, and performed significantly better than the above-mentioned two extreme cases. For cases where the scene is cluttered with many objects and occlusion is frequent, one may have to develop more elaborate algorithms based on Bayesian evidence accumulation techniques, or Kalman filter for dynamic scenes or when the sensor position information is uncertain. Note that the gridding algorithms presented here can process data sequentially and update the voxel estimates as new data stream in.

IV. Experiments

In this section, we present results of controlled experiments performed with the acoustic lens that illustrate how well underwater objects can be reconstructed. In these examples the lens is mounted on a carriage in a tow tank and moves on a straight line at constant speed. Objects are placed on or near the tank bottom and the lens passes over them. The carriage speed is fairly low, about 0.3 knots, and a pass is completed within 1 min. The bottom is made of concrete, and the backscatter from it is so strong that we are able to deduce its position accurately. Since our purpose is to find out how objects appear to the acoustic lens, the bottom has been replaced in the reconstructed scene with a constant intensity plane. Two experimental data sets are examined. In the first, the lens passes over five metallic spheres with diameters ranging from 25 to 45 cm that are suspended 1 m above the bottom of the tow tank. In the second case, which is of greater interest because we pass over a larger object with different scale features, the lens passes over a 1.6-m long remotely operated vehicle (ROV). For both data sets, the scenes shown in this section are reconstructed from single passes over the objects.

After obtaining a 3-D gray-level image, we must use a technique to get a clear view of the objects of interest through ambient noise. There are two main approaches to visualizing 3-D scalar data: volume rendering and surface rendering [5], [11]. In volume rendering we assign to each voxel a color and a partial transparency, and then form images by blending together colored, semitransparent voxels that project to the same pixel on the image plane. In surface rendering, we first apply a surface detector to the volume data, then tile the surface with polygons and finally render the surface. The main drawback in surface rendering is that we have to classify each voxel as belonging to a surface or not. Its advantage is in rendering speed.

There are various techniques for surface reconstruction. One approach that we have used is to first identify voxels that are surface candidates, i.e., voxels that indicate discontinuity in the backscatter intensity, and then reconstruct a surface from these voxels. For discontinuity detection one may use, for example, the technique of [12]. For surface reconstruction there are a number of techniques; see [3] and the references therein. Another approach is to perform the detection and reconstruction simultaneously; see [20]. The marching cubes algorithm [14] is also a valuable tool for quickly obtaining a surface, by approximating the *true* surface with an *iso-value* surface. Surfaces reconstructed with these techniques and rendered using Phong shading produce realistic looking objects. Nevertheless, we find that (for the data obtained by this sensor) volume rendering produces images that look more similar to the actual objects. We believe it is because of the relatively uncertain position of surface voxels due to low resolution as well as sensor motion blur. (For higher resolution, sensors on stationary platforms surface rendering produces equally good images [8].) The scenes are typically 100^3 cubic voxels, which can be relatively slow to render. However, more efficient ways of volume rendering are being developed [1]. The computer images presented in this paper are generated by volume rendering with an adaptive subvoxel trilinear interpolation of the gray-level values.

The backscatter intensity from objects' surfaces exceed the sensor/ocean noise levels at these ranges, yet the partial opacity of noisy voxels prevents us from getting a clear view of the objects of interest. Thus, we have to filter the scene before volume rendering. We have found that thresholding alone is unsatisfactory, because certain voxels belonging to objects may be eliminated. To prevent this, each voxel below the threshold is tested. It is kept only if at least n of its 26 neighbors exceed the threshold. At the same time, we wish to eliminate stray voxels that happen to have high values, possibly because of secondary backscatter. Hence, each voxel above the threshold is also tested to verify that some number of neighboring voxels are also above the threshold. Images thus obtained with $2 \leq n \leq 7$ are fairly similar, which gives us a measure of confidence as to the validity of these filtering operations. The effects of these filters are shown for the ROV below. Similar filters for 2-D images are discussed in [17].

A. Spheres

In this experiment, a constellation of five metallic spheres are suspended at a height of about 1 m above the bottom of the tow tank. Fig. 2 shows the sketch of the experiment. The three small spheres lined up along the lens track are 25 cm in diameter, while the diameter of the two large spheres is 45 cm. The tow carriage moves the lens over the spheres at a known rate of speed. In Fig. 3, we display the reconstructed spheres viewed from the top and the side of the tow tank.

The recreated scene, in spite of distorted appearance of some of the spheres, is in good agreement with the ground truth. The spheres stand out clearly at the correct locations in the scene;

Authorized licensed use limited to: NRL. Downloaded on December 1, 2009 at 15:29 from IEEE Xplore. Restrictions apply.

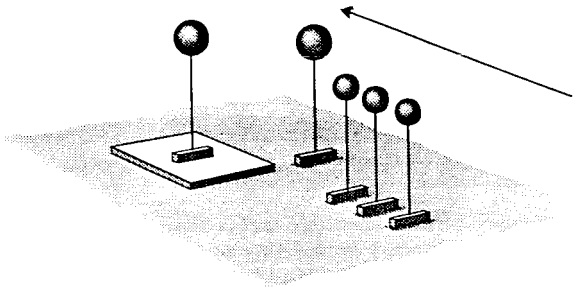

Fig. 2. Sketch of the experiment with five spheres. The arrow indicates the lens's track.

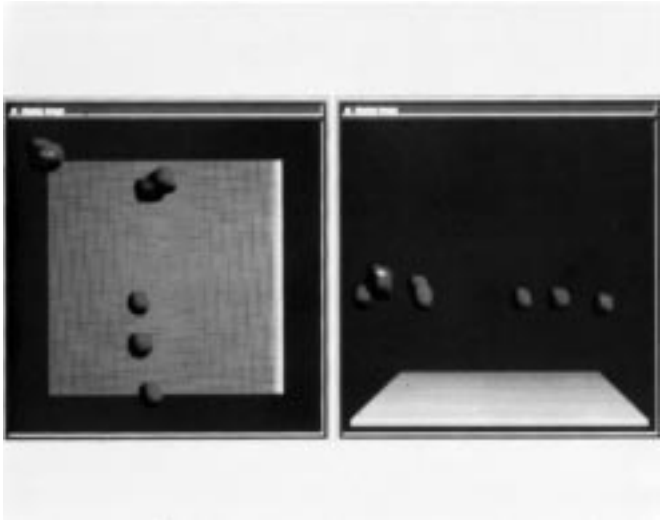

Fig. 3. Top and side views of a volume rendered scene with five spheres.

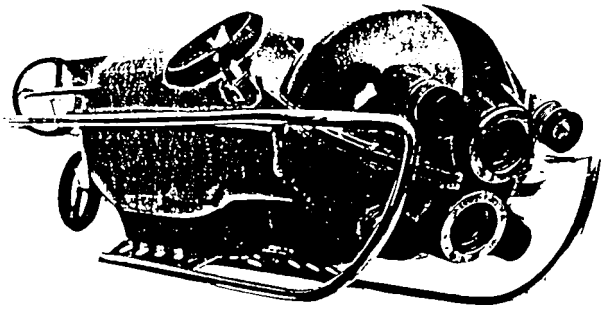

Fig. 4. Picture of the 1.6-m-long ROV. The circle shows the lateral resolution of a single transducer at the object.

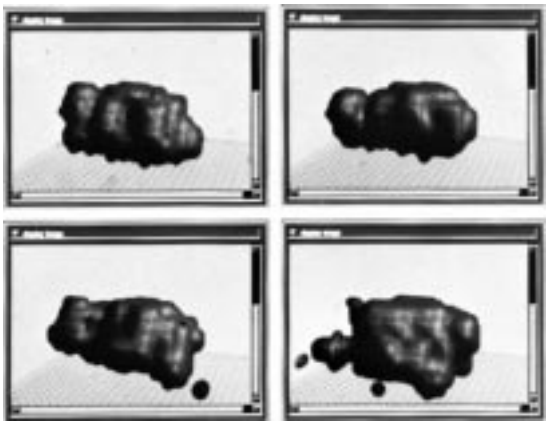

Fig. 5. Acoustic images of the ROV made from single passes: run one, top left; run two, top right; run three bottom left; run four, bottom right. Voxel size is 12 cm.

note that one of the spheres is mounted on the pallet, and even this detail is correctly recreated in the acoustic snapshot—that sphere appears higher in the side view of Fig. 3. The sizes of the objects are also recovered correctly to within voxel dimensions. Although the spatial digitization relative to the object size is too coarse to yield much shape information (the diameters of the spheres span only three to five voxels), nevertheless the objects have a spherical appearance. In this experiment, the closest distance for viewing the spheres is 8 m. That is the single beam lateral resolution at target is at best, $8 \times \tan 1.5^\circ \approx 0.2$ m, or 20 cm, which is comparable to the diameter of the spheres. Thus, imaging the spheres with a stationary lens would not have yielded their correct sizes, much less their shapes. Remarkably, by moving the lens, the resolution of the reconstructed spheres is increased to the degree that even their shapes may be inferred.

B. The ROV

In this experiment, the 1.6-m-long ROV (see Fig. 4) used in underwater explorations is placed on a small pallet on the bottom of the tow tank. Several passes were made over the ROV, changing the ROV's angle with the lens track direction by 30° each time. In Fig. 5, we show acoustic images of the ROV, reproduced from single passes. The orientation angle of the ROV's long axis with respect to the lens track is (approximately) 90, 60, 30, and 0°, in runs one through four, respectively. For comparison with the ROV's picture, all four runs displayed in Fig. 5 are viewed from angles that roughly correspond to the same viewing angle as in Fig. 4. The scenes are processed and volume rendered as described earlier in this section. The dimension of the cubic voxels in Fig. 5 is 12 cm. The recreations of the ROV have the correct size, dimensions, and orientations with respect to the tow tank. One can also discern the outlines of the head and the body of the ROV in the volume rendered images of runs one and two. Finer features with sizes less than or comparable to a voxel dimension (12 cm) are, of course, not discernible because of insufficient resolution.

Fig. 6 is the same as Fig. 5, except here the voxel dimension is 8 cm. These figures show the effects of voxel size. The data is obviously too sparse to support the higher resolution reconstructions, as the recreated objects (with the possible exception of run one) have lost their resemblance to the ROV. Our main purpose, however, in showing the single passes at this higher resolution is for comparison with the results we obtain in the next section when we register and combine these four passes. When the data from four passes are combined, there are roughly four times as much data, and the conjecture is that the number of voxels may be increased by roughly a factor of four. This suggests a voxel size of 8 cm for the combined four-pass data; the ratio of voxel volumes is

Authorized licensed use limited to: NRL. Downloaded on December 1, 2009 at 15:29 from IEEE Xplore. Restrictions apply.

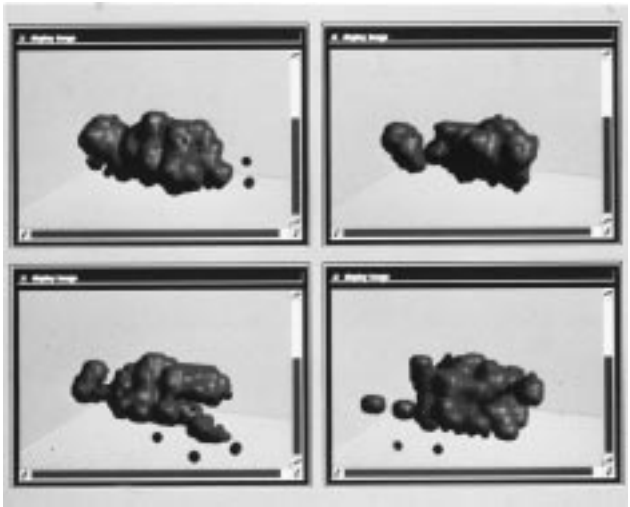

Fig. 6. Same as Fig. 5. Voxel size is 8 cm.

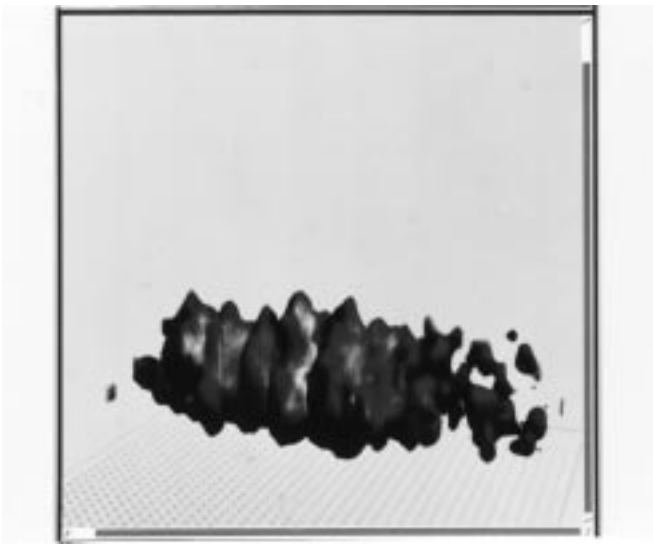

Fig. 7. Reconstruction of the ROV from run one without filtering the scene. Voxel size is 8 cm.

$(12/8)^3 = 3.4$. The results of Section V agree with this conjecture (see Fig. 9).

In Fig. 7, we show the scene recreated from run one data without the use of the n-neighbor filtering described above. Comparing this result with the filtered scene given in Fig. 6, shows the effects of this filtering in cleaning the noise as well as preserving certain features.

V. Registration of Multiple-Pass Data

Combining data from several passes should result in reduced random noise and increased resolution. This would be straightforward if sensor position data were always accurate. However, in practice, ships and underwater vehicles in the ocean are affected by a variety of factors, such as currents; thus, position information contain significant inaccuracies. Accordingly, algorithms are required to correctly align the multiple views of the scene. In bottom mapping, misregistrations of hundreds of meters are common, and complex

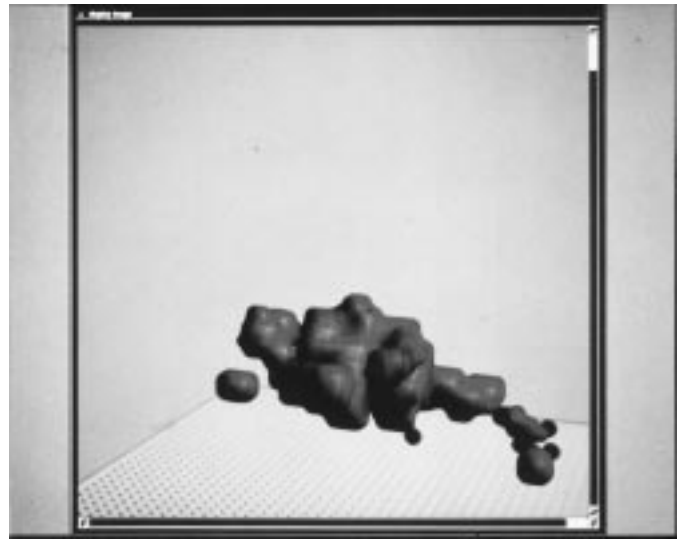

Fig. 8. The ROV as reconstructed from all four passes without registration. Voxel size is 8 cm.

Fig. 9. The ROV as reconstructed from four passes after registration. Voxel size is 8 cm.

techniques are needed to register intersecting swaths of bathymetric data [10]. Even controlled conditions in the tow tank left much to be desired. To produce data equivalent to making multiple passes over a fixed object from different angles, several runs were made over the ROV with its orientation in the x-y plane (the cross-track-along-track plane) changed by a nominal value of 30° for each run. For moving the ROV between runs, upon the completion of a run, divers lifted the ROV and replaced it on a pallet along lines at 30° intervals about a center point.

We have processed the four runs, described in the previous section, to determine what can be gained by using multiple passes. Not surprisingly, the recorded positions of the ROV's proved to be inexact. Fig. 8 shows the ROV using data from all four runs without performing any registration corrections—clearly, correction for the misalignment of passes is needed!

Authorized licensed use limited to: NRL. Downloaded on December 1, 2009 at 15:29 from IEEE Xplore. Restrictions apply.

Suppose, in two different passes, we obtain two views of the same scene (views A and B), which partially overlap. View B is shifted and rotated with respect to A by an unknown amount because of the lack of precise position information for the sensor platform (or for the object, in our experiment). The registration process requires computing translational and rotational adjustments so that the overlapping portions of A and B properly align. Note that these are 3-D range images, hence the scale is known.

A common approach in scene matching is to detect corresponding features and then match them. However, the low data resolutions do not provide distinguishing features. Another approach, perhaps better suited to this data, is to enclose the object in B in a box; use the box as the template and find its best match in scene A by minimizing a distance measure such as

$$D = \int d\vec{r}\left[Sg_B(\vec{r}) - g_A(\vec{r})\right]^2$$

where S is a translation followed by a rotation, and the integral is over the template volume. Here $g_A(\vec{r})$ and $g_B(\vec{r})$ denote the gray-level values of the two scenes. The action of S on the template is nonlinear, hence, minimizing D with respect to transformation parameters has to be carried out through an iterative downhill search, which may fail to find the correct answer. Or, one may attempt an exhaustive search by trying all translations and rotations. Both techniques are computationally expensive. We have found that matching moments instead to register the scenes yields satisfactory results. Below we give a description.

A. Matching Moments

The use of moments for 2-D image representation and pattern recognition was first discussed by Hu [6]. Moment invariants were extended to 3-D by Sadjadi and Hall [18] and have been further refined by Lo and Don [13]. The moments of a 3-D scene are defined as

$$m_{ijk} = \int d\vec{r}\, x^i y^j z^k\, g(\vec{r}). \tag{3}$$

In these experiments, the ROV is the dominant object and appears unoccluded in every pass. By thresholding, we may segment it from the background with reasonable accuracy. The scene after thresholding becomes a binary image. The zeroth moment m_{000} is thus a normalizing factor and is a measure of the volume of the object. The normalized first moments

$$x_c = \frac{m_{100}}{m_{000}}, \quad y_c = \frac{m_{010}}{m_{000}}, \quad z_c = \frac{m_{001}}{m_{000}} \tag{4}$$

are the coordinates of the object centroid. The difference in the centroids of A and B uniquely determines the translation of scene B.

We also need the normalized central moments defined as

$$\mu_{ijk} = \frac{1}{m_{000}} \int d\vec{r}\,(x - x_c)^i (y - y_c)^j (z - z_c)^k\, g(\vec{r}). \tag{5}$$

The eigenvectors of the matrix of second order central moments

$$Q = \begin{pmatrix} \mu_{200} & \mu_{110} & \mu_{101} \\ \mu_{110} & \mu_{020} & \mu_{011} \\ \mu_{101} & \mu_{011} & \mu_{002} \end{pmatrix} \tag{6}$$

are the orientations of the principal axes of the object. The largest eigenvalue corresponds to the longest axis, etc. By matching the corresponding principal axes of the object in scenes A and B, we can determine how much scene B must be rotated. The rotation matrix, R, may be constructed from

$$R = \sum_{i=1}^{3} a_i\, b_i^T = \begin{pmatrix} a_{1x} & a_{2x} & a_{3x} \\ a_{1y} & a_{2y} & a_{3y} \\ a_{1z} & a_{2z} & a_{3z} \end{pmatrix} \begin{pmatrix} b_{1x} & b_{1y} & b_{1z} \\ b_{2x} & b_{2y} & b_{2z} \\ b_{3x} & b_{3y} & a_{3z} \end{pmatrix} \tag{7}$$

where $a_i = (a_{ix}, a_{iy}, a_{iz})^T$ and b_i (superscript T stands for transpose) are the eigenvectors of the matrices Q^A and Q^B with corresponding eigenvalues, i.e.,

$$Q^A\, a_i = \lambda_i^A\, a_i, \qquad Q^B\, b_i = \lambda_i^B\, b_i. \tag{8}$$

Under ideal conditions, the eigenvalues $\lambda_i^A = \lambda_i^B$.

The rotation matrix has an eightfold ambiguity, due to the ambiguity in the positive directions of the principal axes a_i and b_i relative to each other. If we require (a_1, a_2, a_3) and (b_1, b_2, b_3) to form right-handed coordinate systems such that $a_1 \times a_2 = a_3$ and $b_1 \times b_2 = b_3$, then the ambiguity is only fourfold. This ambiguity can be resolved by matching third-order moments in the following manner. First, we form vectors $\vec{v}^A$ and $\vec{v}^B$ from the third-order moments

$$\begin{aligned} v_x &= \mu_{300} + \mu_{120} + \mu_{102}, \\ v_y &= \mu_{030} + \mu_{210} + \mu_{012}, \\ v_z &= \mu_{003} + \mu_{201} + \mu_{021}. \end{aligned} \tag{9}$$

Then we calculate $\vec{u}^B$ by rotating $\vec{v}^B$, i.e., $\vec{u}^B = R\vec{v}^B$ where R is the rotation matrix candidate transforming B to A, and compare it to $\vec{v}^A$. That $\vec{u}^B$ which is identical (or closest) to $\vec{v}^A$ is obtained with the correct rotation matrix.

We note that if the scene is composed of several objects, registration by moment matching can still be solved in closed form by using the 3-D line matching technique presented in [9].

B. Results

In practice, the data is discrete, so the integrals become sums over voxels above a selected threshold. For a wide range of thresholds, both the centroids and principal axes were stable. As an example, in Table I we present the centroid location and the axes orientations as a function of the intensity threshold for run two. The centroid varies by about 1 cm and the principal axes by less than 3°, for thresholds ranging from 300 to 900. Other runs yield similar results. This demonstrates the reliability of the process.

Table II contains translational and rotational corrections for the four runs. In this table we assume run four position is correct, and list the required adjustments to register the other runs. We see that cross-track positional estimates were in

Authorized licensed use limited to: NRL. Downloaded on December 1, 2009 at 15:29 from IEEE Xplore. Restrictions apply.

TABLE I
ROV's Estimated Position in Run Two for Different Intensity Thresholds; the Maximum Intensity Is 4200. (x_c, y_c, z_c) Are the Centroid Coordinates. ϕ Is the Angle of the ROV's Long Axis with the Lens Track, and θ Is Its Angle with the Tank Bottom Plane

	x-shift cross-track	y-shift along-track	rotation angle
Run 1	6 cm	39 cm	88.5°
Run 2	3 cm	82 cm	68.1°
Run 3	6 cm	16 cm	38.5°
Run 4	0 cm	0 cm	0.0°

TABLE II
Translation and Angular Adjustments Found Using the Registration Algorithms. Rotation Angles Were Nominally 90, 60, 30, and 0°. Corrections Are Shown Relative to Run Four

Threshold	Points	x_c m	y_c m	z_c m	ϕ degree	θ degree
100	3672	2.97	14.28	0.86	70.1	4.0
200	1724	2.99	14.31	0.80	73.5	5.1
300	1189	2.99	14.30	0.80	71.9	3.5
400	856	2.99	14.29	0.79	69.4	2.2
500	647	2.98	14.29	0.79	68.2	1.9
600	478	2.98	14.28	0.78	68.1	2.8
700	381	2.98	14.28	0.79	67.7	3.1
800	320	2.99	14.27	0.78	68.5	3.0
900	259	3.01	14.25	0.78	68.7	2.3
1000	218	3.03	14.25	0.78	68.5	2.0
1500	89	3.05	14.24	0.76	71.8	0.5

error by a few cm, along track estimates by nearly 1 m, and rotational estimates by nearly 10°. The rather large along-track corrections are primarily due to inaccuracies in recording the time the moving lens started collecting data. Rotational angles involving depth (i.e., roll and tilt angles), as well as vertical translations, were very small (as one would expect since the ROV was resting on the pallet). We do not list these very small ($<1°$ and $\simeq 1$ cm) angles and shifts in Table II. The small depth rotational angles and translations provided further evidence of the stability of the process.

Fig. 9 shows the ROV using registered data from the four runs. A comparison with Fig. 4 shows that the appearance of coarser features has improved, and that small features such as the strobe lights and cameras in the head and the large thruster on the side now appear in the reconstructed figure as *bumps* in the appropriate locations. Note that the closest viewing distance of the ROV by the lens is about 8 m. This means that a conical receive beam has a circular footprint with diameter of 20 cm on the ROV, which is as large as the radius of the ROV's head. Hence, it is remarkable that the ROV can be reconstructed with the degree of detail seen in Fig. 9.

VI. Concluding Remarks

We have presented techniques for scene reconstruction, filtering, pass registration, and visualization of backscatter data obtained by an acoustic lens mounted on a moving platform. The resulting 3-D underwater images show a remarkable degree of detail, even though the lens resolution is fairly low (25 cm at a distance of 10 m) compared to the imaged objects. The images presented here are taken under controlled conditions where the lens position and pose is known accurately at all times during a pass, thus these images show the limits of what can be achieved with a moving lens.

The fundamental limitation of underwater imaging with the acoustic lens, and other 3-D imaging sonars, is the sparse, low-resolution data, compared to TV images. The low sensor resolution can be partially overcome by obtaining a denser data set. This is done by moving the sensor over the scene to obtain many overlapping views, as well as by combining data from several passes over the object. In most practical situations, however, where the imaging system is held by a diver or is subject to unknown motions, the sensor position may not be known accurately enough. Also, pass registration is a nontrivial problem in a controlled experiment in a tow tank and will be even more difficult in the ocean. These factors can introduce noise, distort objects, and further blur small scale features. Clearly, higher resolution imaging sonars are needed. Lens-based acoustic imaging system prototypes with centimeter resolutions are currently being developed [7], [8]. The image acquisition of these systems, however, are not yet real time. The image processing and visualization techniques discussed here are applicable to newer lens prototypes.

An effect that is present in acoustic imaging (more so than in optical imaging) is interreflection of acoustic waves between objects, or between objects and the bottom. These interreflections produce noise and false surfaces. Detection and removal of these surfaces with standard image processing techniques does not appear to be possible. We are developing techniques, similar to the recent works in optical imaging [15], for detection of false surfaces due to interreflection. This will allow a cleaner reconstruction of scenes from acoustic backscatter.

Acknowledgment

The authors thank colleagues at the Naval Research Laboratories, E. Carey and C. Jones for organizing the experiments, and Behrooz Kamgar-Parsi for many helpful discussions.

References

[1] R. Avila *et al.*, "VolVis: A diversified volume visualization system," in *Proc. IEEE Visualization'94 Conf.*, Washington, DC, Oct. 1994, pp. 31–38.
[2] E. O. Belcher, D. Steiger, and L. J. Rosenblum, "A forward looking active acoustic lens," Tech. Rep. APL-UW TR 9113, Appl. Phys. Lab., Univ. Washington, Seattle, WA, May 1991.
[3] R. M. Bolle and B. C. Vemuri, "On three-dimensional surface reconstruction methods," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, pp. 1–13, 1991.
[4] D. L. Folds, "Status of ultrasonic lens development," in *Underwater Acoustic and Signal Processing, Proc. NATO Adv. Study Inst.*, Copenhagen, Denmark, Aug. 1980, pp. 263–279.
[5] H. Fuchs, M. Levoy, and S. M. Pizer, "Interactive visualization of 3D medical data," *Computer*, pp. 46–51, Aug. 1989.
[6] M. K. Hu, "Visual pattern recognition by moment invariants," *IEEE Trans. Inform. Theory*, vol. IT-8, pp. 179–187, 1962.
[7] B. Johnson *et al.*, "3-D acoustic imaging with a thin lens," in *Proc. IEEE Oceans'93 Conf.*, Victoria, B.C., vol. 3, pp. 444–449.
[8] B. Kamgar-Parsi, B. Johnson, D. L. Folds, and E. O. Belcher, "High-resolution underwater acoustic imaging with lens-based systems," *Int. J. Imag. Syst. Technol.*, vol. 8, pp 377–385, 1997
[9] B. Kamgar-Parsi and B. Kamgar-Parsi, "Matching sets of 3-D line segments," Tech. Rep. AIT-95-007, Inform. Technol. Div., Naval Res.

Authorized licensed use limited to: NRL. Downloaded on December 1, 2009 at 15:29 from IEEE Xplore. Restrictions apply.

Lab., Washington, DC, 1995. Also, *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 1090–1099, 1997.
[10] B. Kamgar-Parsi *et al.*, "Toward an automated system for a correctly registered bathymetric chart," *IEEE J. Oceanic Eng.*, vol. 14, pp. 314–325, 1989.
[11] A. Kaufman, *Volume Visualization*. Los Alamitos, CA: IEEE Comput. Soc., 1990.
[12] D. Lee, "Coping with discontinuities in computer vision: Their detection, classification, and measurement," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 321–344, 1990.
[13] C.-H. Lo and H.-S. Don, "3-D moment forms: Their construction and application to object identification and positioning," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 1053–1064, 1989.
[14] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3D surface construction algorithm," *Comput. Graph.*, vol. 21, pp. 163–169, 1987.
[15] S. Nayar, K. Ikeuchi, and T. Kanade, "Shape from interreflections," *Int. J. Comput. Vis.*, vol. 6, pp. 173–195, 1991.
[16] L. Rosenblum and B. Kamgar-Parsi, "3D reconstruction of small underwater objects using high-resolution sonar data," in *Proc. IEEE Symp. Autonomous Underwater Vehicle Technology*, Washington, DC, June 1992, pp. 228–235.
[17] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*. New York: Academic, 1982.
[18] F. A. Sadjadi and E. L. Hall, "Three-dimensional moment invariants," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 2, pp. 127–136, 1980.
[19] W. K. Stewart, "Three-dimensional modeling of seafloor backscatter from sidescan sonar for autonomous classification and navigation," Tech. Rep. 7120, Woods Hole Oceanograph. Inst., Woods Hole, MA, Jan. 1990.
[20] R. Szeliski, D. Tonnesen, and D. Terzopoulos, "Modeling surfaces of arbitrary topology with dynamic particles," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, New York, June 1993, pp. 82–87.
[21] R. J. Urick, *Principles of Underwater Sound*. New York: McGraw-Hill, 1983.

Behzad Kamgar-Parsi (M'89) received the Ph.D. degree in theoretical physics from the University of Maryland, College Park.

He joined the Naval Research Laboratory (NRL), Washington, DC, in 1990, where he is currently a Research Scientist in the Information Technology Division. Before joining NRL, he was a Research Scientist at the Computer Vision Laboratory, University of Maryland, and a Post-doctoral Fellow at Rockefeller University, New York, NY. His areas of interest include image processing, computer vision, and neural networks.

Dr. Kamgar-Parsi has served on program committees of conferences on neural networks, visualization, and computer vision. He is a member of the IEEE Computer Society, APS, and Sigma Xi.

Lawrence J. Rosenblum (M'90–SM'94) received the Ph.D. degree in number theory from The Ohio State University, Columbus.

He is Director of VR Systems and Research in the Information Technology Division, Naval Research Laboratory, Washington, DC, and Program Officer for Visualization and Computer Graphics at the Office of Naval Research. His research interests include VR, scientific visualization, and human-computer interfaces.

Dr. Rosenblum is on the editorial board of IEEE COMPUTER GRAPHICS AND APPLICATIONS, the IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, and the *Journal of the Virtual Reality Society*. He is a previous Chairman and current Director of the IEEE Technical Committee on Computer Graphics. He is a member of the IEEE Computer Society, ACM, Siggraph, and the American Geophysical Society.

Edward O. Belcher (M'90) received the Ph.D. degree in electrical engineering from the University of Washington, Seattle.

He is a Principal Engineer at the Applied Physics Laboratory, University of Washington. His current research is in the development of both liquid-filled and solid, thin, acoustic lens technology that can be applied to a variety of multibeam sonar systems. He recently lead a team that developed a 36-lb diver-held sonar with a sector scan containing 64 0.5° beams. Earlier projects have included underwater telemetry, enhancement of speech from deep-see divers in an oxygen-helium environment, quantification of fish in rivers, and characterization of bubbles in the bloodstreams of divers as they decompress. He has also taught in the Departments of Electrical Engineering and Speech and Hearing Sciences, University of Washington, Seattle.

Authorized licensed use limited to: NRL. Downloaded on December 1, 2009 at 15:29 from IEEE Xplore. Restrictions apply.