Advances in Large-Area Photomosaicking Underwater

Hanumant Singh, Jonathan Howland, and Oscar Pizarro, *Student Member, IEEE*

***Abstract*—The propagation of visible light underwater suffers rapid attenuation and extreme scattering. This, in combination with the limited camera-to-light separation available on most imaging platforms, places severe limitations on our ability to optically image large areas of the sea floor at high resolution. In this paper, we present a general framework for mosaicking large areas underwater with a specific emphasis on the issues that are unique to the underwater environment. At the individual image level, we examine the role of attenuation, scattering, and camera to light separation and present the tradeoffs involved in optimizing a particular imaging geometry. We also examine the arbitrary image-registration problem in the face of conditions prevalent underwater, namely a moving nonuniform lighting source and the effects of a featureless unstructured terrain. Our analysis is based on photomosaics encompassing several hundred images on archaeological, forensic, and geological expeditions from a diverse set of imaging platforms, including the *NR-1* nuclear submarine, the manned submersible *Alvin*, the *Argo* towed vehicle, the *Jason* remotely operated vehicle, and the *ABE* autonomous underwater vehicle.**

***Index Terms*—Mosaicking, optical imaging, underwater vehicles.**

I. Introduction

FROM ITS modest beginnings in the 1890s, with the work of Louis Baton in the French Riviera, underwater photography progressed to deeper depths prior to the Second World War with the work of Beebe [1] and others to eventually become an important tool in deep-sea research during and after the Second World War [2]. With the loss of the nuclear-powered submarine the *USS Thresher* in 1963, there was a surge of interest in deep-sea photographic equipment and techniques, which eventually led to the routine use of photomosaicking in an underwater context. The early techniques employed rectifying lenses in the laboratory that reprojected images to allow the merging of multiple images into a composite mosaic. Numerous examples of these techniques exist till today, including mosaics of the *Monitor* and the *Thresher* [3], and of geological sites of interest photographed during project Famous [4].

Even with the rapid growth of photogrammetric measurements for satellite imaging, the techniques for photomosaicking underwater continued to be manual, subjective, and based on film cameras and analog optics. In the late 1980s, with the advent of digital computers and charge-coupled devices (CCD) cameras, efforts at the Deep Submergence Laboratory, Woods Hole Oceanographic Institution, Woods Hole, MA, focused on using specialized image-processing hardware for manipulating and processing digital images of underwater scenes to manually construct photomosaics [5], [6]. Work in this context is also being pursued at other institutions. Marks *et al.* reported results in automated mosaicking from real underwater digital imagery assuming translation only motion with very small rotation [7]. Negahdaripour *et al.* have looked at different aspects of using continuous video acquired from test tank data for mosaicking and visual navigation [8]. Trucco *et al.* [9] have looked at robust methods for extracting features from continuous video (and sonar), while Gracias *et al.* [10] describe a scheme for building a mosaic from continuous video, which may then be used for visual feedback.

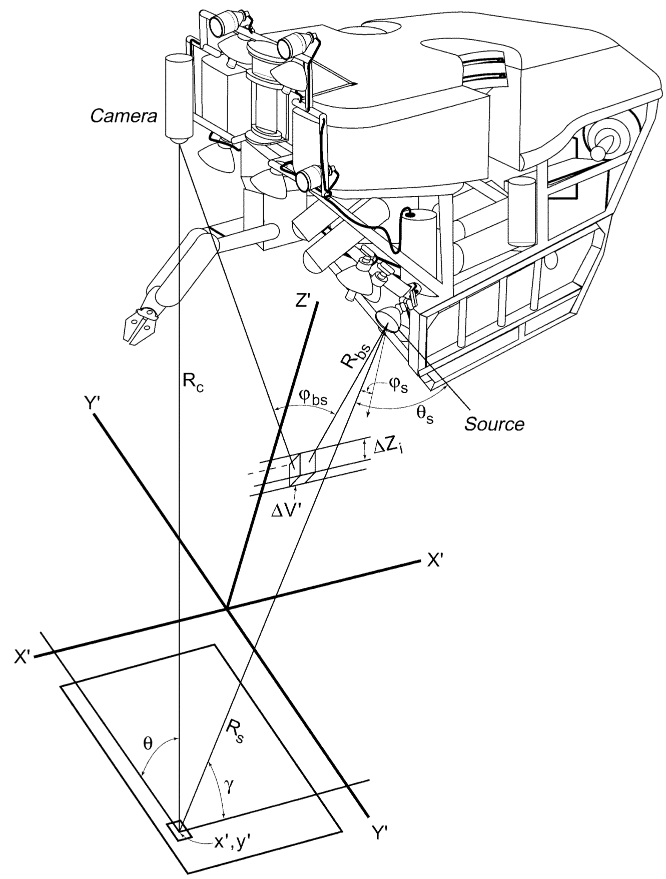


Fig. 1. Lighting geometry for imaging from underwater vehicles. The energy received at the camera is composed of direct reflected energy, backscatter from the water column, and forward-scattering processes.

Manuscript received November 2002; revised April 7, 2004. This work was funded by a grant from the National Science Foundation CenSSIS Engineering Research Center under Grant EEC-9986321. This is WHOI contribution.

The authors are with the Deep Submergence Laboratory, Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, MA 02543 USA (e-mail: hsingh@whoi.edu).

Digital Object Identifier 10.1109/JOE.2004.831619

TABLE I
LIGHTING CHARACTERISTICS OF REPRESENTATIVE IMAGING PLATFORMS

Platform	Camera-to-light Separation	Power	Relative Coverage	Limitations
NR-1 Submarine	Large O(several meters)	Essentially Unlimited O(10kW)	O(10m x 10m)	Direct Component limited by Lambertian Scattering
DSV Alvin	Large (meters)	Limited O(1kW)	O(5m x 5m)	Limited Power albeit with good camera-to-light separation
ROV Jason	Medium O(meters)	Essentially Unlimited O(10kW)	O(5m x 5m)	Backscatter Limited
AUV Abe	Limited O(1m)	Very limited O(0.1kW)	O(1m x 1m)	Non-uniform lighting / low overlap

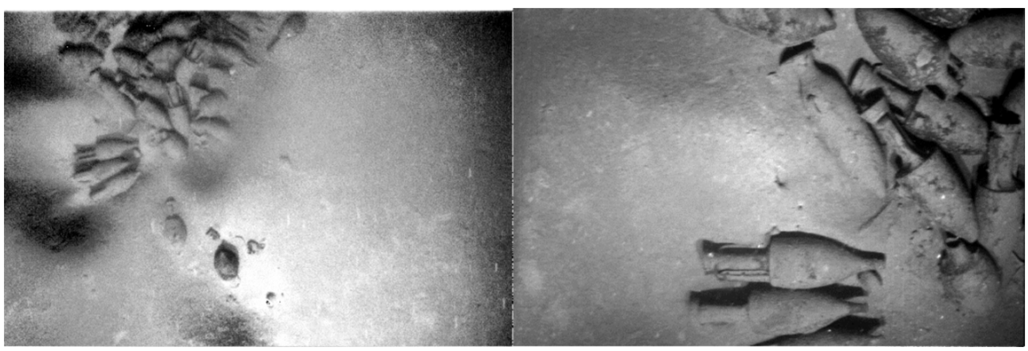


Fig. 2. Amphora pile imaged with the large camera-to-light separation associated with the *NR-1* nuclear submarine flying at an altitude of camera 8 m (left image). The smaller camera-to-light separation associated boxed in the left image is clearer, but would require an order of magnitude more images to mosaic with the *Jason* remotely operated vehicle (ROV) flying at an altitude of 4 m (right image). We see that the image on the right covers the area while the image on the left covers a larger area. Image quality is limited by the energy received at the camera. We note that the same camera was used in both cases (cf Figs. 9 and 10).

Independent of the work being carried out underwater, photomosaicking is actively being pursued for mobile land robots and consumer applications. While a comprehensive review of the work on land is outside the scope of this paper, some of the techniques being actively pursued include feature-based methods [11], spatial methods [12], [13], and linear differential geometry [14]. In general, although researchers have reported significant advances on land, there are fundamental differences between these applications and the underwater domain.

Contributions of this paper: In this paper, we focus primarily on the aspects of mosaicking that are unique to the underwater environment. We make the assumption that the large metacentric height associated with most underwater imaging platforms minimizes the pitch-and-roll effects in underwater imagery.

1) *Role of Lighting.* If we examine the role of mosaicking underwater, it is primarily a means to obtain a global perspective in a region where the rapid attenuation of electromagnetic radiation (and, specifically, visible light) limits the ability to encompass large areas within a single frame. Our focus is on strobed light sources and digital still cameras (as opposed to continuous frame video), because a number of underwater imaging platforms cannot afford the power associated with continuous lighting. Conventional wisdom dictates that higher power and larger camera-to-light separation would allow us to image much larger areas, but our analysis shows that this is not the case.
2) *Featureless Imagery and Terrain Effects.* We propose a significant modification to existing algorithms that allows us to make photomosaics in featureless areas that characterize large parts of the underwater terrain. We show the applicability of our algorithm with representative data and compare its performance with commercially available software. We also present a technique that utilizes insights into the nature of underwater imaging to allow us to build photomosaics in areas where the bathymetry significantly distorts the mosaics associated with high-resolution close-in imagery.
3) *Handling Arbitrarily Large Rotations* We look at Fourier-based methods for mosaicking imagery with very large rotational differences. Such imagery is characteristically collected with deep towed vehicles whose dynamics make structured surveys difficult. We show that our technique works well with large arbitrary rotations and also provides a measure for when the image-registration process is successful.

This paper is organized in the following manner. Section II looks at the role of lighting geometry in individual image acquisition. Section III presents two techniques for two-dimensional (2-D) image registration for constructing large photomosaics of real underwater imagery. Section IV presents some extensions to these techniques for handling large photomosaics in unstructured three-dimensional (3-D) underwater terrains, followed by some general conclusions in Section V.

II. UNDERWATER LIGHTING

There has been considerable work done in understanding the propagation of light in the ocean following the pioneering work

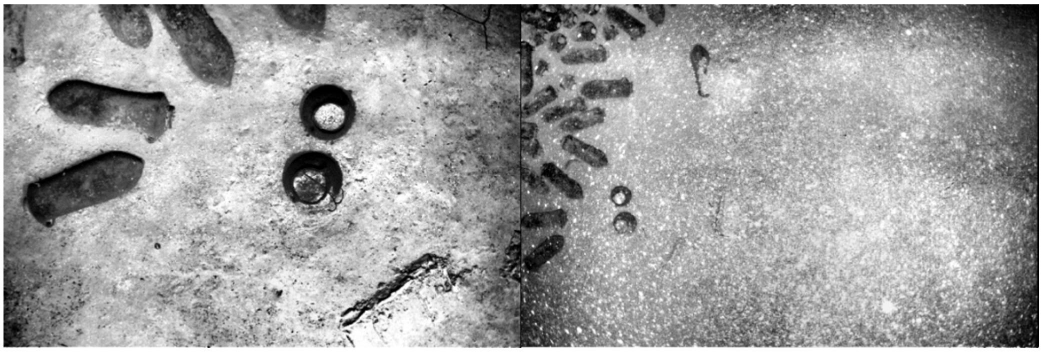


Fig. 3. Effects of backscatter for geometries with small camera-to-light separation. The images of the two cooking pots and amphorae, which are part of a Phoenician shipwreck taken by the *Jason* ROV from altitudes of 4 m (left image) and 8 m (right image). The area corresponding to the left image is boxed in the right image. Backscatter effects severely limit image quality at the higher altitude.

of Duntley [15]. A number of important papers in this area are collated in Caimi [16]. These include discussions on the inherent and apparent optical properties of seawater—attenuation, absorption, refractive index, scattering, and radiative transfer. For the purposes of large-area photomosaicking, the primary issues are the camera-to-light separation and power available for lighting. Jaffe [17] was the first to simulate the effectiveness of side-lighting for underwater imaging systems, but in this section we wish to build on this analysis for real imaging platforms. To examine these issues, we follow an approach similar to that of Jaffe [17], which was based on the approach by Glamery [18].

The light energy appearing on the image plane E_{total} consists of several components—a direct component E_d, a backscattered component E_{bs}, and a forward-scattered component E_{fs}.

$$E_{\mathrm{total}} = E_d + E_{\mathrm{fs}} + E_{\mathrm{bs}}. \tag{1}$$

For the purposes of image acquisition for large-area photomosaicking, we focus on the direct and backscattered components only. We can ignore the forward-scattered component because, for standard camera geometries, the forward scattering simply adds and subtracts small contributions to the dominant terms.

The expression for the direct component can be arrived at by considering the incident irradiance on the object, the reflectance function associated with the object, and by taking into account factors such as the geometric optics of the camera, attenuation in the water column, and spherical spreading losses.

If we consider the lighting geometry depicted in Fig. 1, the incident irradiance pattern E_I is given by

$$E_I = BP(\Theta_s, \Phi_s)\frac{e^{-cR_s}}{R_s^2}\cos\gamma \tag{2}$$

where

subscript s refers to quantities measured from the light source;
subscript c refers to quantities measured from the camera;
subscript bs refers to quantities measured with respect to a backscattered volume element;
$BP(\cdot)$ is the beam pattern associated with the light source and is the angle between the object and source;
c is an attenuation constant associated with the propagation of light in sea water;
R is the distance along the ray from the object to the source.

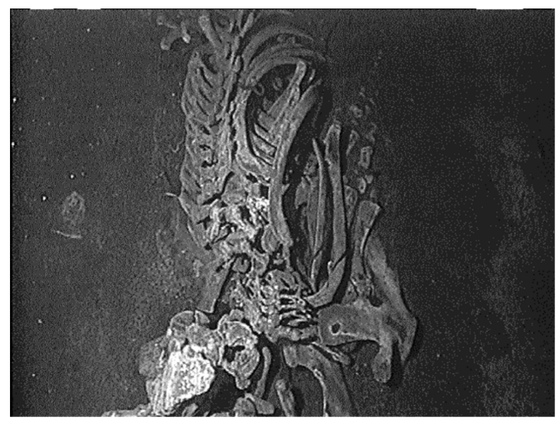


Fig. 4. Whale remains imaged by the *DSV Alvin* off San Diego, CA, at a depth of 1700 m. Limitations in power led to overall intensity levels that are quite low even though the image is uniformly illuminated.

Fig. 5. Geological site of interest imaged by the *ABE* autonomous underwater vehicle (AUV). Severe power limitations have resulted in a low-intensity nonuniformly illuminated image.

The direct component E_d can then be expressed as

$$E_d = E_1(x', y', 0)e^{-cR}\frac{cL(x', y')}{4f_n}(\cos\Theta)^4 T_1 \times \left[\frac{R_c - F_1}{R_c}\right]^2 \tag{3}$$

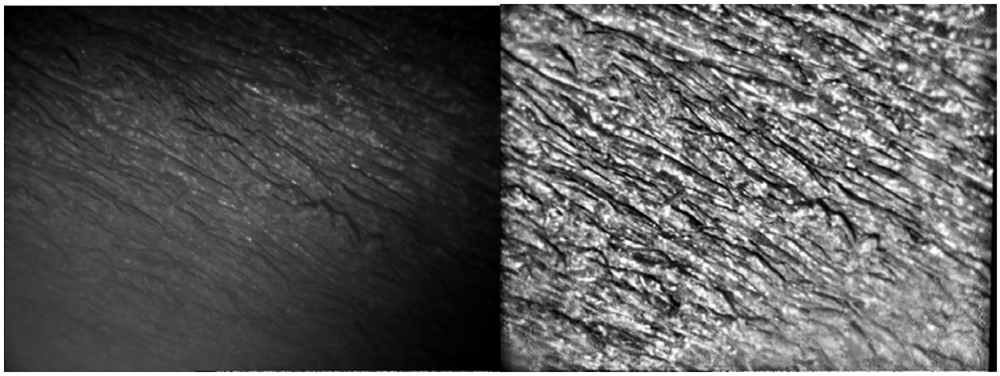


Fig. 6. (left) Original *ABE* imagery and (right) an adaptive histogram equalized version. While this technique compensates very well for the nonuniform lighting pattern, it cannot (as seen in the lower right corner) compensate for parts of the image where the light intensity levels are of the order of the sensitivity of the camera.

Fig. 7. Low-contrast *Jason* imagery (left) before and (right) after adaptive histogram equalization using a Rayleigh probability distribution function (pdf). An adaptive histogram equalization brings out a level of detail not evident in the original imagery.

where $(x', y', 0)$ refers to the fixed-coordinate system with respect to the planar reflectance map located at $z' = 0$.

R distance between the object and the camera;
$L(x', y')$ reflectance function;
f f number of the camera of focal length F_1 is the angle between the object and the aperture of the camera;
T_1 transmittance of the lens.

The expression for the backscattered component E_{bs} can be similarly arrived at by considering the summation of energy backscattered by volume elements along the path of illumination. The irradiance E_s incident on a volume element V' is

$$E_s(x', y', z') = BP(\Theta, \Phi)\frac{e^{-cR_{\mathrm{bs}}}}{R_{\mathrm{bs}}^2} \quad (4)$$

where R_{bs} is the distance between the volume element and the light source.

The radiant energy that is scattered toward the camera due to this volume element V' is

$$H_{\mathrm{bs}}(\Phi) = \beta(\Phi_{\mathrm{bs}})E_s(x', y', z')\Delta V' \quad (5)$$

where β is the volume-scattering function.

By integrating over the volume of water that contributes to the energy at the camera, we get

$$E_{\mathrm{bs}} = \sum_{i=1}^{N} e^{-cZ_{ci}}\beta(\Phi_{\mathrm{bs}})E_s(x', y', z') \times \Pi\frac{\Delta Z_i}{4f_n^2}(\cos\Theta)^3 T_l\left[\frac{Z_{ci} - f_l}{Z_{ci}}\right]^2 \quad (6)$$

where Z_{ci} is the distance between the volume element and the camera and Z_i is the thickness of the backscattering volume V'.

In combination with Table I, which shows the lighting and camera geometries of various imaging platforms, these equations provide a context within which we can discuss the relationship between lighting and image acquisition. Examining (3) and (6), we note the following.

1) The direct component of energy arriving at the camera, which we would like to maximize, falls off as the fourth power of the cosine of the angle of incidence between the object being imaged and the light source, exponentially with distance.
2) The backscattered component arriving at the camera, which we would like to minimize, is a volume function that increases with increasing volume and falls off as the third power of the cosine of the angle of incidence. Specifically, we can now examine how power and camera-to-light separation affect large-area photomosaicking.

Fundamentally, we would like to acquire individual images that

1) maximize the field of view;
2) minimize the effects of backscatter;
3) minimize the energy required to illuminate the scene.

Unfortunately, these requirements all are at odds with each other. Thus, for example, while flying at higher altitudes increases the field of view, it does so at the cost of higher backscatter and larger energy requirements. These interacting

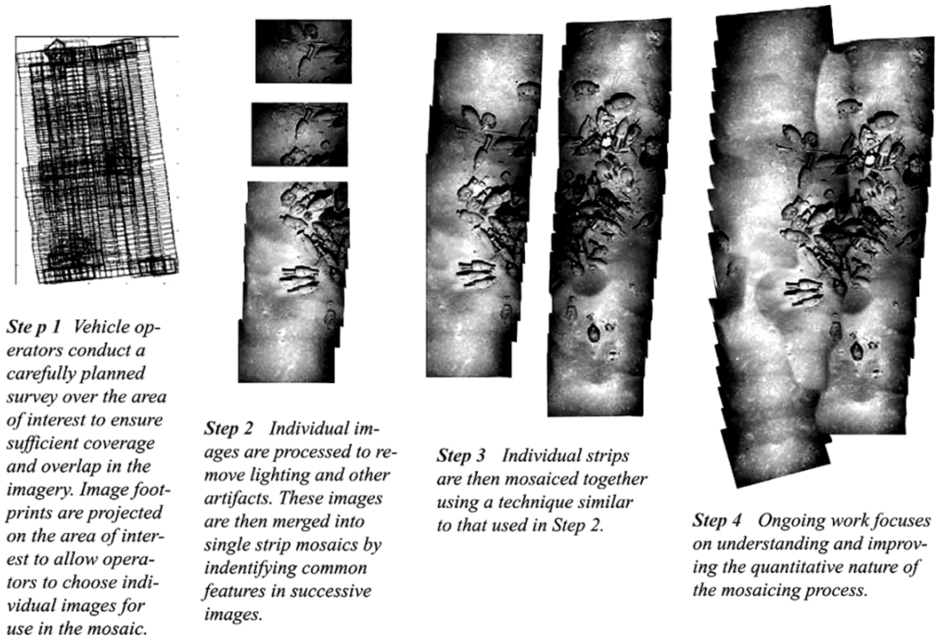


Fig. 8. Process of photomosaicking.

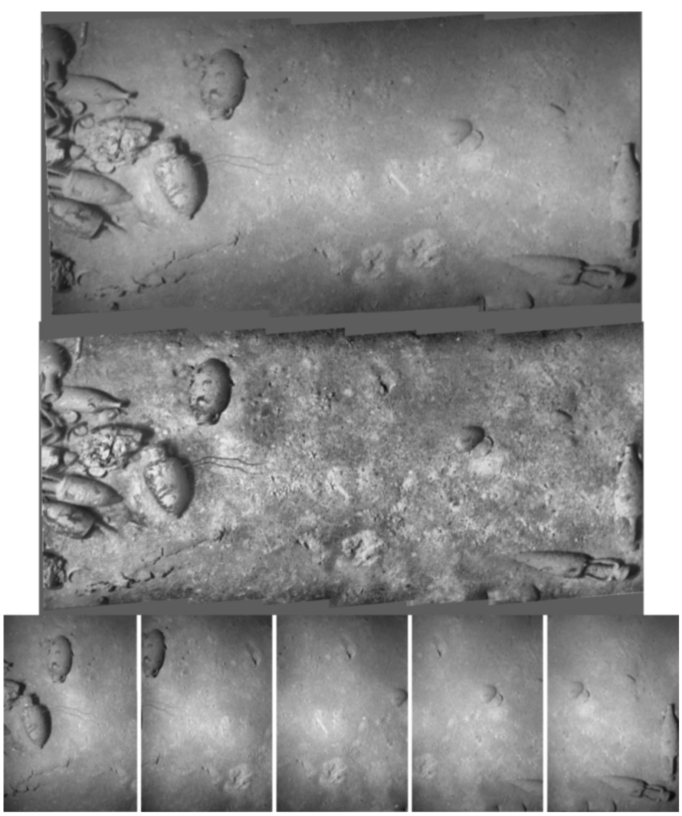


Fig. 9. Mosaicking featureless low-texture imagery. The top mosaic was constructed with commercial software. A close inspection shows that a number of features are duplicated and some areas are not represented in this mosaic. This is in contrast to our technique, which is shown in the middle image. The significant individual images that compose these mosaics are shown later for reference.

effects manifest themselves differently on different imaging platforms. Example imagery acquired from the different platforms is shown in Figs. 2–5.

Fig. 2 illustrates the effect of flying at different altitudes. Fig. 2 (left) shows one image of an amphora pile taken from the *NR-1* submarine. Such a platform has essentially very large camera-to-light separation and almost unlimited power. This amphora forms part of an ancient Roman shipwreck discovered in 800 m of water in the Mediterranean and covering 30×15 m [23]. Fig. 2 (right) shows part of the same pile imaged from the *Jason* remotely operated vehicle (ROV). In both cases, an identical camera was used. The tradeoff is clear to see—while the image taken from the *NR-1* nuclear submarine covers a much larger area, the image taken from *Jason*, while of high resolution, covers a fairly small area in comparison. However, the cosine term in (3) limits our ability to look inside the "shadowed" depressions while using the *NR-1* from a high altitude. On the flip side, the clarity in the *Jason* ROV imagery comes at the cost of the increased complexity that is associated with the larger number of images required to photomosaic the same scene (cf. Figs. 10 and 11). Finally, we note that the *NR-1* imagery is not limited by backscatter due to the large camera-to-light separation that this platform can accommodate.

Fig. 3 highlights the importance of backscatter in scenarios with limited camera-to-light separation. The images are of a Phoenician shipwreck located off the coast of Israel in the Mediterranean. The *Jason* ROV (with its limited camera-to-light separation) was used to image this site from a height of 8 m (right image) as well as from a height of 4 m (left image). The right image is characterized by saturated splotches corresponding to backscatter from the water column. In comparison, the left image, acquired at a height of 4 m, is remarkably clear and devoid of these effects. Comparing Fig. 3 with Fig. 2 highlights the importance of camera-to-light separation as predicted by (6).

Figs. 4 and 5 illustrate the role of power in acquiring imagery. Fig. 4 shows an image of the skeletal remains of a whale taken from the manned submersible *Alvin*. The image was acquired

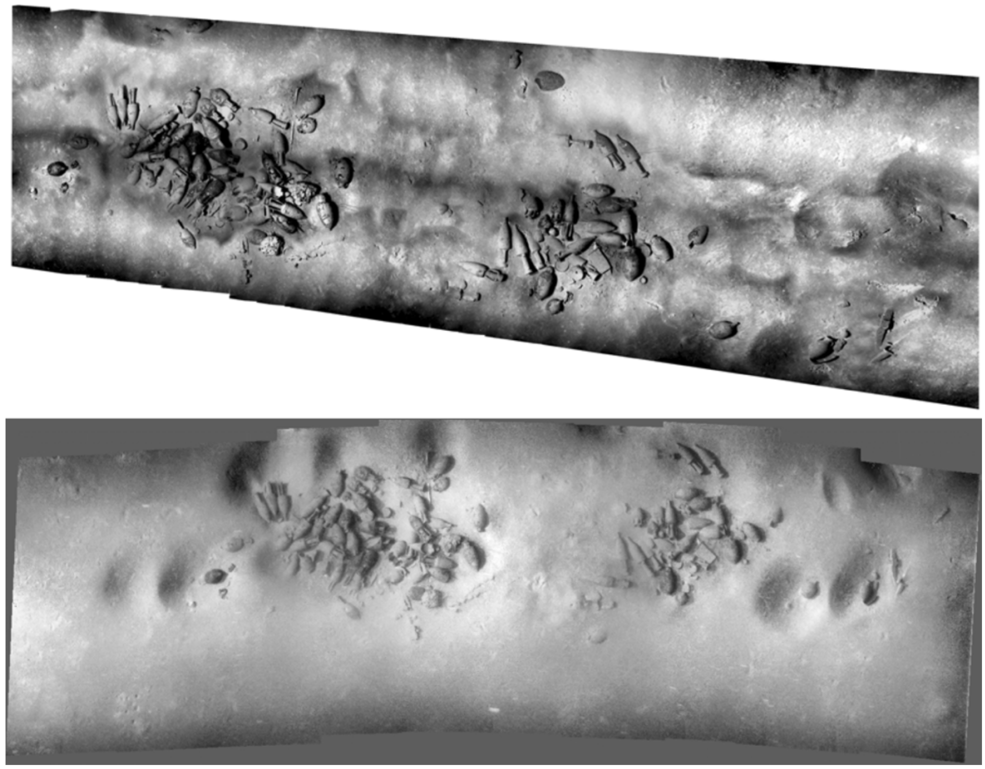


Fig. 10. Photomosaics of a Fourth Century BC Roman shipwreck located in 800 m of water in the Mediterranean off of the coast of Sicily, Italy. The top image was constructed using 180 images taken in four overlapping strips using the *Jason* ROV. The right image was constructed using ten images from the *NR-1* nuclear submarine. Our techniques work for both sets though the differences between the mosaics—resolution, complexity in constructing the mosaic—represent fundamental tradeoffs between the two imaging platforms.

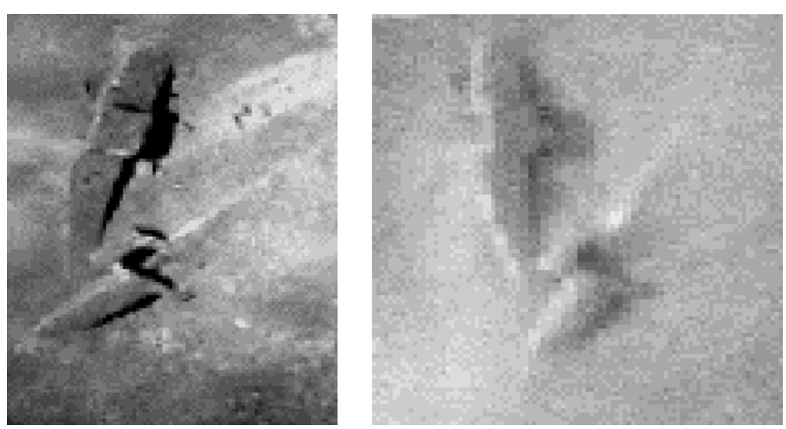


Fig. 11. Details from (left) the *Jason* image and (right) the *NR-1* image derived photomosaics of the Fourth Century BC Roman shipwreck. As expected, the imagery from the *Jason* photomosaic has much higher resolution.

at a height of about 5 m over the site in an effort to minimize the perspective distortions associated with the 3-D structure. While the lighting pattern achieved is quite uniform, the overall image intensity is quite low. Fig. 5 is an image taken from the *ABE* autonomous underwater vehicle (AUV) being flown 5 m above volcanic terrain in the East Pacific Rise. The nonuniform lighting pattern is clearly visible in this image.

After examining the two major limitations—area of coverage and lighting artefacts—outlined in this section, we note that while the area of coverage is a tradeoff that is dictated by the platform and application, we can, to a certain extent, compensate for lighting artefacts. We can do so through the use of standard image-processing techniques, as outlined in Section II-A.

A. Compensating for Underwater Lighting

To deal with lighting artefacts due to nonuniform illumination and the low contrast associated with underwater imagery, we utilize the classic techniques associated with histogram equalization [19], adaptive histogram equalization and specification [20], and model-based adaptive histogram equalization [21]. With these techniques, the image is broken up into subregions. The optimal gray-scale distribution is calculated for each of these subregions, based upon its histogram and a previously determined transfer function, which is based upon the desired histogram of the subregion. Then, each pixel of the image is adjusted based on interpolation between the manipulated histograms of the surrounding subregions.

Fig. 12. Photomosaic of the skeletal remains of a whale imaged using the *Alvin* manned submersible. Although one obtains a very good composite view of the entire subject, the 3-D perspective effects show up as slight discontinuities between images across some parts of the mosaic.

Fig. 6 shows the application of adaptive histogram equalization on *ABE* imagery acquired at a geological site of interest in the East Pacific Rise. Our implementation is seen to correct quite well for the nonuniform lighting associated with low power requirements on AUVs. Image intensities are well distributed, as we would expect in a geologically uniform region. Fig. 7 shows the effect of histogram-equalizing low-contrast imagery acquired with the *Jason* ROV. We also plot out the histograms of the images before and after equalization. In this example, we have used a Rayleigh pdf to specify the equalization process. From an aesthetic standpoint, our user community prefers this over exponential or Gaussian distributions. As we demonstrate later (Fig. 9), the nonlinear transformation associated with histogram equalization is one important step in mosaicking featureless regions. While this process does bring out the detail in the imagery, image quality is a subjective measure and some individual users may prefer to use the imagery as it was originally acquired. We should also stress that while these techniques are powerful in compensating for lighting artifacts, they do not substitute for uniform lighting across the field of view. These subjective techniques, while enhancing detail, also tend to obfuscate lighting cues that human operators rely on for depth perception. Moreover, they may highlight lighting artifacts [Fig. 10 (left)] of the survey process that tend to distract from the whole image, as opposed to uniformly lit imagery where such artifacts are not noticeable (Fig. 16).

III. Image Registration and Photomosaicking

The process of photomosaicking involves aligning multiple images to one another and then merging these images into a seamless composite representation, as shown in Fig. 8. The alignment of a pair of images may be accomplished by a variety of techniques, which include manual tie point location, spatial techniques, and Fourier-transform-based methods. An extensive survey of image registration techniques is that by Brown [22]. These techniques vary from being very labor intensive to being completely automated in specific regimes of application with attendant tradeoffs in ease of use and accuracy. We look at each of these techniques in turn and, in particular, their roles in the context of underwater applications.

A. Manual Location of Tie Points

We should note that several packages exist that provide an easy interface to allow manual photomosaicking. Operators are presented with views of the images to be mosaicked and select common features across these images. A linear least-squares solver package is typically employed to solve the overconstrained set of equations that correspond to the manually selected tie points to estimate the transformation between the pair of images.

The residuals associated with the process of solving the overconstrained set of equations provides a level of quality control over the process. Typically, operators can fine-tune their choice of tie points to reject those with high residuals to produce an aesthetically pleasing mosaic. The techniques associated with manual mosaicking have additional significance in areas with unstructured hard-to-model relief, as discussed in Section III-B.

B. Spatial Methods

Our spatial approach to photomosaicking is based on that of Sawhney and Kumar [12], although it differs in two important ways to deal with the peculiarities of the underwater environment.

Let $I_1, \ldots I_N$ be the N images that we are trying to register. In our formulation, each of the images is related to an ideal 2-D coordinate system through an interior camera transformation (3-D to 2-D) and through an exterior view transformation (2-D to 2-D). The reason for choosing such a formulation is that we do not rely on a single reference image to be ideal or distortion-free (from 3-D terrain effects, for instance). We define a point p in the ideal coordinate system such that is the representation of that point in image I_i. The points p and p_i are related by the mapping

$$p_i = P(p; A_i) \tag{7}$$

where A_i is the 3-D to 2-D projection transformation that maps p to p_i.

A_i may be a Helmert, affine, projective, or higher order transformation. Given the transformation, intensities at points p_m in

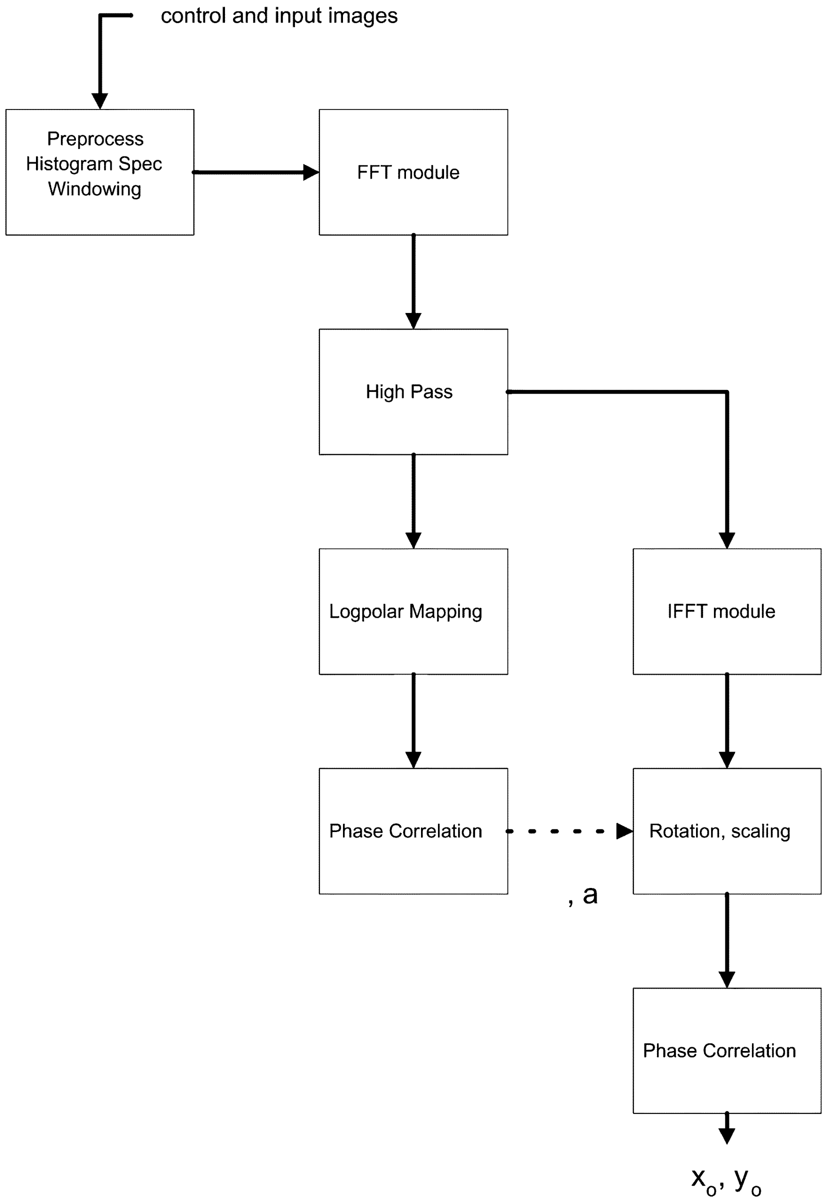


Fig. 13. Data flow diagram for our implementation of a Fourier-based registration technique.

image I_m and at point p_n in image I_n, which transform to the same reference coordinate p, are related through

$$I_m\big(p_m; (p, A_m)\big) = I_n\big(p_n; \big(p, A_n\big)\big). \tag{8}$$

This is the so-called *brightness constancy constraint*. However, while this may serve as a good-enough approximation in a number of land-based applications, it is not generally applicable for underwater imagery. As shown in Section II, the light source moves with the camera on underwater imaging platforms and, in the case for nonuniformly illuminated images, this equation does not hold true. Thus, we deviate from the formulation proposed by Sawhney and Kumar in two important ways.

1) We substitute adaptively histogram equalized imagery in place of the original images even though, due to aesthetic considerations, we may use the unequalized images for the mosaic. The histogram equalized *does not* in itself guarantee brightness constancy over the images (for example, it does not handle the movement of shadows with a moving light source—see below). However, it does succeed in removing the dominant effect of lighting patterns from the imagery. In the absence of shadows, it sets up an intensity constancy constraint (ICC) that we exploit.
2) We detect and discount the contribution due to shadows. As the light source moves, the shadows tend to also shift. Edge-based operators often pick up the transitions between shadows and objects, much more preferentially than edges that occur within the objects due to the low-contrast nature of underwater imagery. Thus, techniques that rely on edge-based operators often distort the mosaicking process, as they are fooled by this apparent motion of shadows across successive images. We detect and discount the contributions due to shadows in the following manner. Shadows are easy to detect by simply

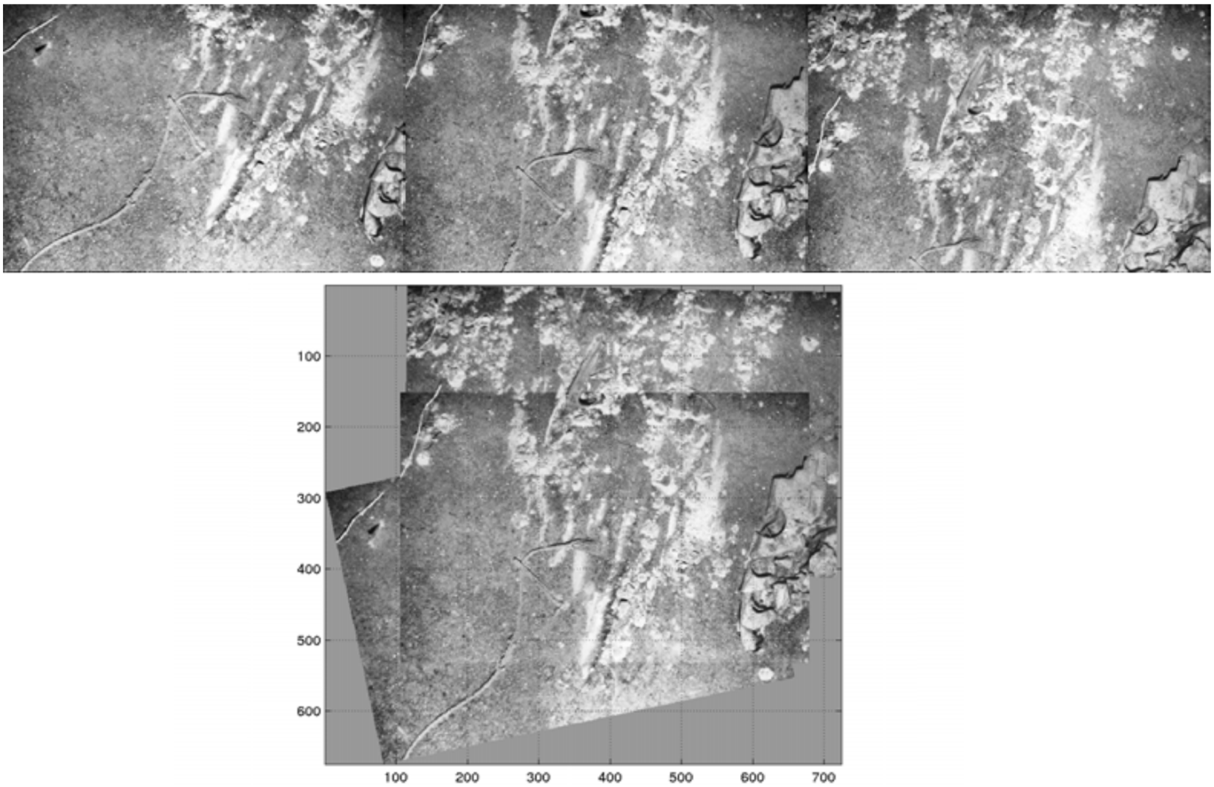


Fig. 14. Results of mosaicking three images (top row) acquired with an arbitrary survey strategy using our implementation of a Fourier-transform-based technique.

thresholding at a low intensity level. We can then grow the region around the shadows by dilating the image.

We note that, in order to align frames with displacements in the range of tens of pixels, we utilize a Laplacian of Gaussian pyramid. Rather than simply use the Laplacian of Gaussian of the histogram-equalized image, we mask this image with the results of the dilated shadows.

For ease of understanding and notation, we continue to use the term I_i to represent the histogram equalized, shadow masked, Laplacian of Gaussian filtered image.

The projection transformations for the N frames, $A_1 \ldots A_N$ are unknown, as is the correspondence between the points of the images. The correspondence between points of the images can be established only through the transformation of the reference coordinates in (7). In order to compute the correspondences and unknown parameters simultaneously, we formulate an error function that minimizes the variance in intensities of a set of corresponding points in the images that map to the same ideal reference coordinate. This translates to a minimization problem of the form

$$\min_{A_1 \cdots A_n} \Sigma_p \frac{1}{M(p)} \sum_i \left(I_i(p_i) - I^{-}(p)\right)^2 \tag{9}$$

where the point p_i in the frame i is a transformation of the point p in the reference frame and is the mean intensity value of all the p_i that map to p and $M(p)$ is the count of all p_i that map to p.

We can now develop the multi-image formulation using a plane projective transformation. For the purposes of this paper, we assume that the A_i represents an eight-parameter projective plane transformation

$$x_i = \frac{a_{11}^i x + a_{12}^i y + a_{13}^i}{a_{31}^i x + a_{32}^i y + a_{33}^i} \tag{10}$$

$$y_i = \frac{a_{21}^i + a_{22}^i y + a_{23}^i}{a_{31}^i + a_{32}^i y + a_{33}^i} \tag{11}$$

where a_{11}^i–a_{33}^i are the plane projective parameters that correspond to image I_i that we are trying to estimate. a_{33}^i can be set to 1 without a loss of generality.

It is evident from the optimization function in (9) that the transformations in (10) and (11) cannot be obtained in closed form. We utilize the Levenberg–Marquadt (LM) optimization for minimizing the sum of squares. In order to apply the LM technique, each term in (9) is linearized.

$$E\big((p_i;p);A\big) = \frac{1}{\sqrt{M(p)}}\big(I_i(p_i) - \bar{I}(p)\big) \tag{12}$$

where A (with no subscripts) refers to the set of all N unknown A_i's. Given a solution at the kth step for the unknown parameters A_k, each $E\big((p_i;p);A\big)$ is linearized around this solution as

$$E\big((p_i;p);A\big) = E\big((p_i;p);A^k\big) + \nabla E\,|_{A^k}\,[\delta A] \tag{13}$$

where

$$E\big((p_i;p);A^k\big) = \frac{1}{\sqrt{M(p)}}\big(I_i(p_i(p;A_i^k)) - \bar{I}(p^i(p;A_i^k))\big). \tag{14}$$

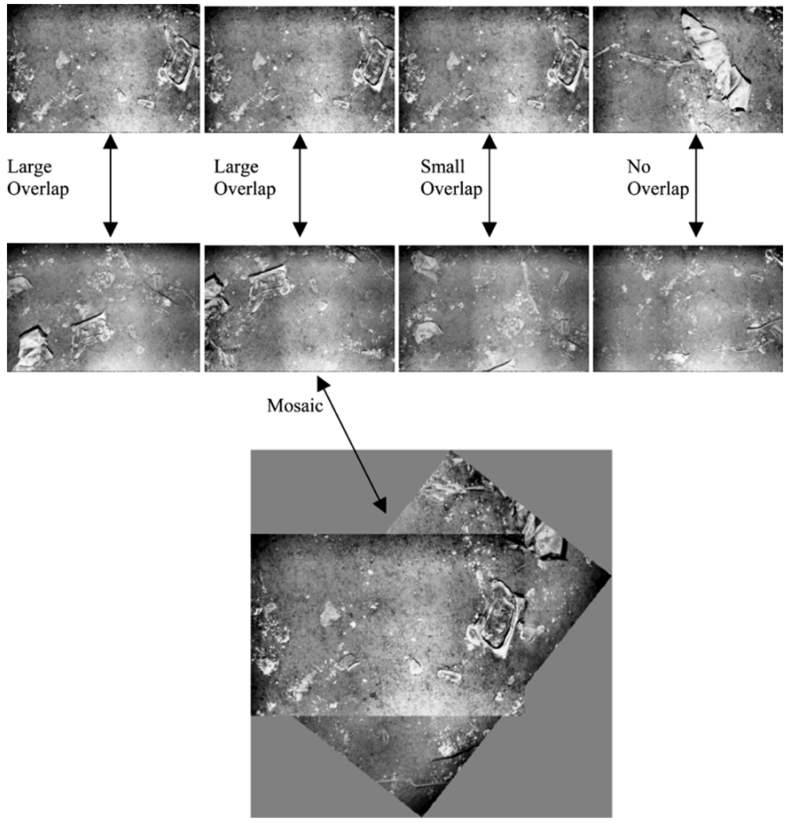


Fig. 15. Mosaicking images with large, small, and no overlap acquired over the course of an unstructured survey. Our algorithm succeeds in mosaicking images with large overlap, but just as importantly provides a measure to quantify the validity of the registration process.

The first term in (14) is the intensity value for image i sampled at location p_i, which is a forward mapping of the corresponding p in the reference image with mapping parameters A_i^k. Recall that we do not know the correspondences (p_i, p); these are known only through the mapping parameters [(10) and (11)]. Given that only the forward mapping from p to p_i is known, $I_i(p_i(p_i; A_i^k)$ can be written in terms of a warped image represented in the reference p coordinates; that is, $I_i(p_i(p_i; A_i^k) = I_i^w(p)$. The warped image is created by computing p_i for image I using p and the parameters A_i^k and interpolating the known values of the image I_i at integer pixel locations.

Therefore, $I_i^w(p)$ represents the current estimate of image i in the reference coordinates. Also, $E((p_i(;p; A_i^k) = 1/M(p)\sum_i I_i^w(p)$. Thus, we can write

$$E\big((p_i; p); A^k\big) = \frac{1}{\sqrt{M(p)}}\big(I_i^w(p) - \bar{I}^w(p)\big). \tag{15}$$

Given this form of the error term and the basic linearization

$$\sum_p \sum_i \big(E(p_i; p) + \nabla E(p_i; p)\,[\delta A]\big) \tag{16}$$

we can use a standard implementation for the LM algorithm (we use the Matlab implementation, which runs for a pair of images in under a minute on a 500-MHz Linux workstation) to estimate A and, thus, build up a mosaic.

The results of our techniques are illustrated in Figs. 9–12. Fig. 9 highlights the advantage of our technique. The images correspond to the midsection (between the two amphora piles) of the Roman shipwreck shown in Fig. 10 [23]. We compare our technique with results obtained using the automated mosaicking package based on the work of Sawhney [12] and Peleg [13]. If we closely examine Fig. 9, we see that the highly effective blending associated with this package tends to mask the fact that features A, B, and C are mirrored in the mosaic while feature X has disappeared altogether. Our algorithm, on the other hand, handles such imagery quite effectively. Fig. 10 shows photomosaics of a Fourth Century BC Roman shipwreck located in 800 m of water in the Mediterranean. The image on the left was constructed from 180 images taken from the *Jason* ROV. While the individual strips were constructed automatically, the strips were mosaicked together by hand. The image on the left was constructed automatically using ten images acquired with the *NR-1* nuclear submarine. The tradeoffs between the two mosaics are interesting to note. The ten-image *NR-1* photomosaic obviously lacks the resolution of the 180-image mosaic, as can be seen from examining the detailed imagery of anchors from both mosaics, as shown in Fig. 11. The somewhat nonuniform lighting in the *Jason* imagery was compensated for by adaptive histogram equalization, which tends to negate the lighting cues that are so effective in providing a sense of the underlying relief in the mosaic derived from the *NR-1* imagery.

Fig. 12 is an eight-image photomosaic of the skeletal remains if a whale off the coast of San Diego, CA, imaged using the *Alvin* manned submersible. The intent in showing this imagery is to highlight our ability to handle distortions due to parallex. While the overall perspective that this mosaic provides is quite good,

TABLE II
TEN REPRESENTATIVE IMAGE PAIRS WITH ASSOCIATED NORMALIZED ROTATIONAL AND TRANSLATIONAL CORRELATION VALUES

Group	Image Pair	Rot-scale Correlation	Translational Correlation
Large Overlap	267.95, 266.95	0.357	0.331*
Large Overlap	267.95, 268.95	0.431	0.088
Large Overlap	267.95, 882.95	0.381	0.544*
Large Overlap	267.95, 883.6	0.374	0.097
Small Overlap	267.95, 265.95	0.293	0.096
Small Overlap	267.95, 269.95	0.472	0.096
Small Overlap	267.95, 881.6	0.255	0.082
Small Overlap	267.95, 884.6	0.299	0.095
No Overlap	265.95, 269.95	0.346	0.095
No Overlap	880.6, 883.6	0.269	0.093

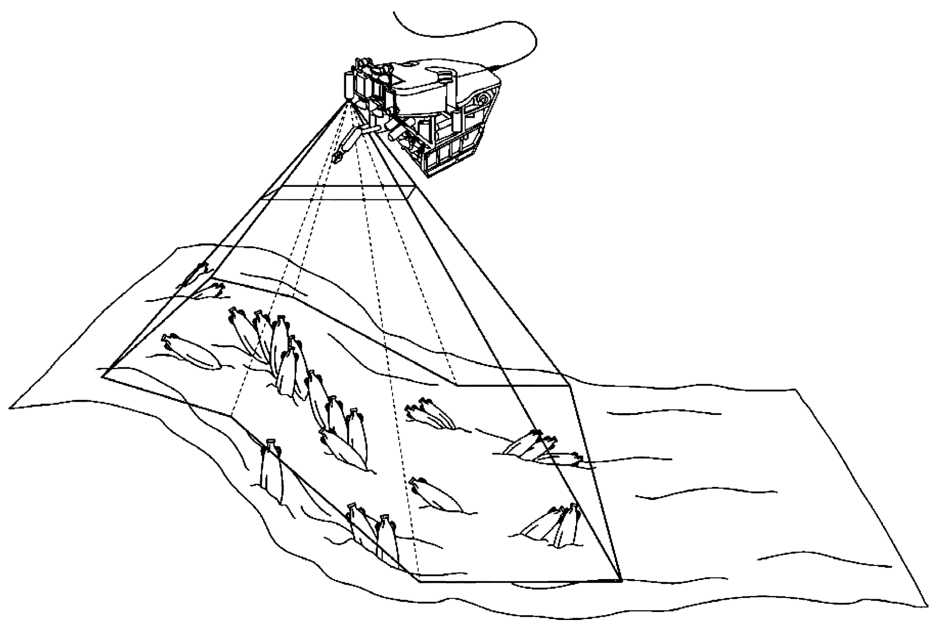


Fig. 16. Geometry of image formation in an unstructured terrain. Equal sections in the image plane do not represent equal areas on the sea floor. Projective transformations do not accurately capture the image-to-image geometry, which can lead to significant distortions, as shown in Fig. 14.

one can notice slight discontinuities across a few of the image boundaries.

C. Fourier Methods

The spatial image registration method outlined in Section III-B works well with imagery that has been collected in a structured survey. A large class of imaging platforms—operator-driven ROVs, towed camera sleds–cannot carry very structured closed-loop surveys, such as those used for acquiring the imagery in Section III-B. These surveys typically may contain large rotational and scale changes between successive images. To investigate photomosaicking from such platforms, we present a Fourier-based registration technique. Our basic approach relies on exploiting the translation property of the Fourier transform to determine rotation, scaling, and translation between to images [24], [25]. Consider I_1 and I_2 , two images that differ only by a translation (x_0, y_0)

$$I_2 = I_1(x - x_0, y - y_0). \tag{17}$$

Their Fourier transforms are related by the translation property

$$F_2(\omega_1, \omega_2) = e^{-j(\omega_1, x_0 + \omega_2 y_0)} F_1(\omega_1, \omega_2). \tag{18}$$

The basic phase-correlation technique obtains the phase through

$$\frac{F_1(\omega_1, \omega_2) \cdot F_2^*(\omega_1, \omega_2)}{F_1(\omega_1, \omega_2) \cdot F_2^*(\omega_1, \omega_2)} = e^{j(\omega_1, x_0 + \omega_2 y_0)} \tag{19}$$

(where the $*$ operator refers to the complex conjugate), which may be inverse transformed to get an impulse at the translation coordinates.

Rotation and scaling can be determined in a similar fashion if the images are represented in a domain where rotation and scaling appear as translations. Consider that images I_1 and I_2 now differ by a rotation and a translation (x_0, y_0)

$$I_2(x, y) = I_1(x \cos\Theta_0 + y \sin\Theta_0 - x_0, \\ -x \sin\Theta_0 + y \cos\Theta_0 - y_0). \tag{20}$$

Their Fourier transforms are related by the rotation and translation properties by

$$F_2(\omega_1, \omega_2) = e^{-j(\omega_1, x_0 + \omega_2 y_0)} \\ \times F_1(\omega_1 \cos\Theta_0 + \omega_2 \sin\Theta_0, -\omega_1 \sin\Theta_0 + \omega_2 \cos\Theta_0). \tag{21}$$

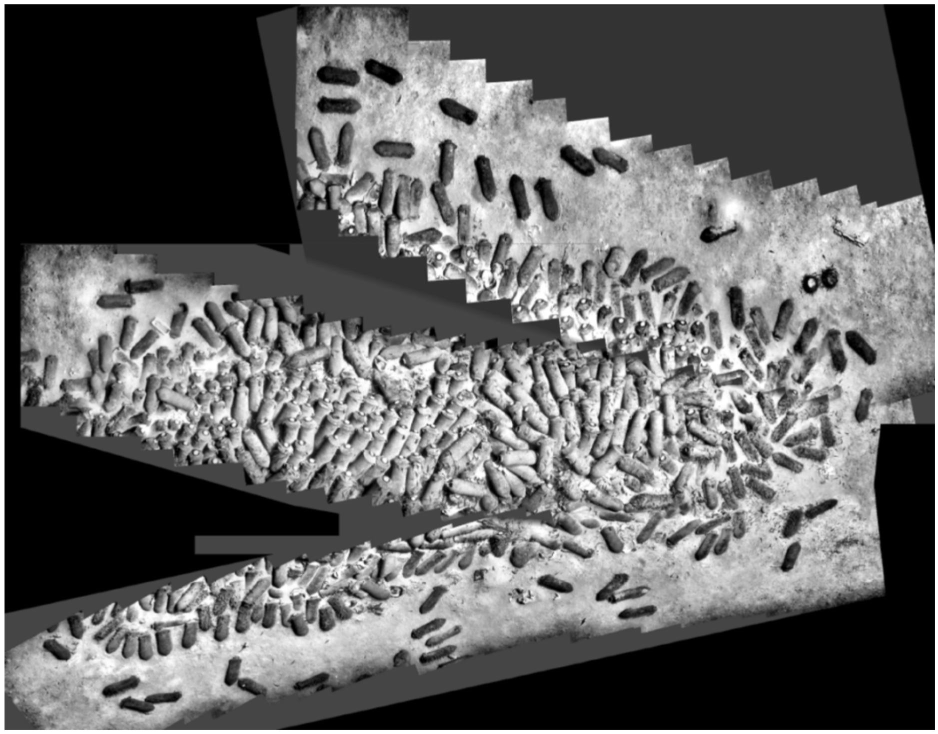


Fig. 17. Effects of terrain on photomosaicking. Differences in depth of the order of 1 m, compared to the 4-m altitude from which this imagery was obtained, result in distortions of the individual strips. The problem is essentially of representing a 3-D unstructured surface on a 2-D plane.

By concentrating only in the magnitudes M_1 and M_2 of the transforms

$$M_2(\omega_1,\omega_2) = M_1(\omega_1 \cos\Theta_0 + \omega_2 \sin\Theta_0, -\omega_1 \sin\Theta_0 + \omega_2 \cos\Theta_0) \quad (22)$$

and by transforming to polar coordinates, the rotation appears as a shift

$$M_2(\rho,\Theta) = M_1(\rho, \Theta - \Theta_0). \quad (23)$$

Now, consider that the images I_1 and I_2 differ by a scaling factor a. Then, their transforms will be given by

$$F_2(\omega_1,\omega_2) = {}_a{}^{1/2} F_1\left(\frac{\omega_1}{a}, \frac{\omega_2}{a}\right). \quad (24)$$

By taking the log of the frequencies, we see that the log of the scaling also appears as a shift

$$F_2(\log\omega_1, \log\omega_2) = {}_a{}^{1/2} F_1(\log\omega_1 - \log a, \log\omega_2 - \log a) \quad (25)$$

if both scaling and rotation are present, as defined by

$$\rho_1 = \sqrt{\omega_1^2 + \omega_2^2} \quad (26)$$

$$\rho_2 = \frac{1}{a}\sqrt{\omega_1^2 + \omega_2^2} \quad (27)$$

$$\Theta_2 = a\tan\left(\frac{\left(\frac{\omega_1}{a}\right)}{\left(\frac{\omega_2}{a}\right)}\right) = a\tan\left(\frac{\omega_1}{\omega_2}\right) = \Theta_1. \quad (28)$$

The magnitudes of the transforms in polar coordinates would be

$$M_1(\rho,\Theta) = M_2\left(\frac{\rho,}{a}, \Theta - \Theta_0\right) \quad (29)$$

and, in the log of the radius of the frequency, we have

$$M_1(\log\rho, \Theta) = M_2(\log\rho - \log a, \Theta - \Theta_0) \quad (30)$$

which can be considered to be a translation. The phase-correlation technique can then by applied to the logpolar versions of M_1 and M_2 (considered now as images) to find $\log a$ and θ_0.

D. Implementation Issues

As with our process for image registration with spatial methods, we preprocess all images using adaptive histogram specification to minimize lighting effects. Edge enhancement is also carried out with a Laplacian of the Gaussian filter. Mapping from the magnitude of Fourier transforms in Cartesian coordinates to logpolar coordinates only requires the top half of the spectra, due to complex conjugate symmetry. Only the frequency annulus from approximately 14–200 samples is used, since most of the frequency content of the Laplacian of Gaussian (bandpass) filtered images lies in this band. Standard phase-correlation techniques recover θ and $\log a$ by inverting the phase and selecting the coordinates of the peak. The parameters used allow us to, theoretically, resolve rotations of 0.3° and 0.44% change in scaling.

The filtered input image is rotated and scaled (using bilinear interpolation) according to the results from the correlation of the logpolar image. Standard phase correlation identifies the translation offsets between the filtered control and the filtered, rotated, and scaled input. A data-flow diagram illustrating our implementation is shown in Fig. 13.

The results of applying this technique to three images is shown in Fig. 14. The site of interest is part of the debris field

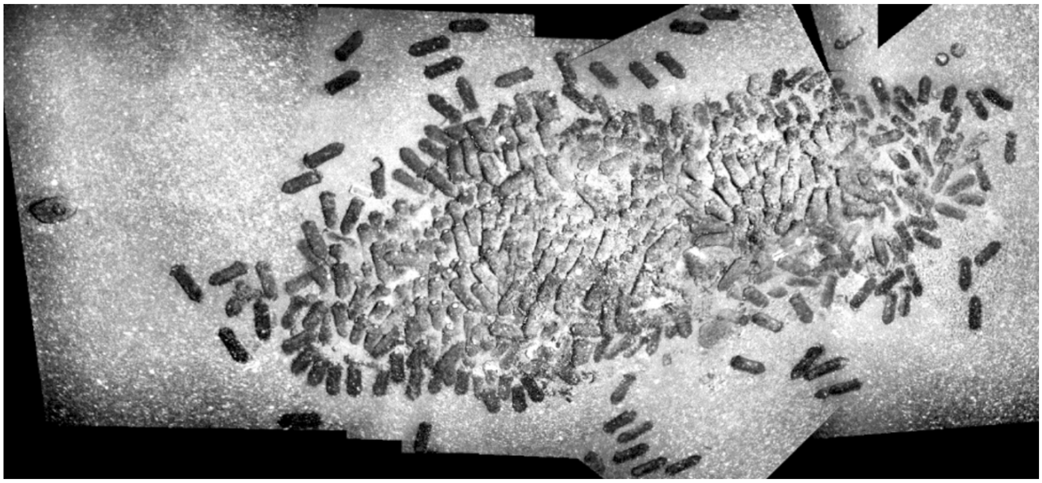


Fig. 18. High-altitude and, consequently, low-resolution pass over a site can provide the basis for a high resolution composite mosaic that deals satisfactorily with terrain effects.

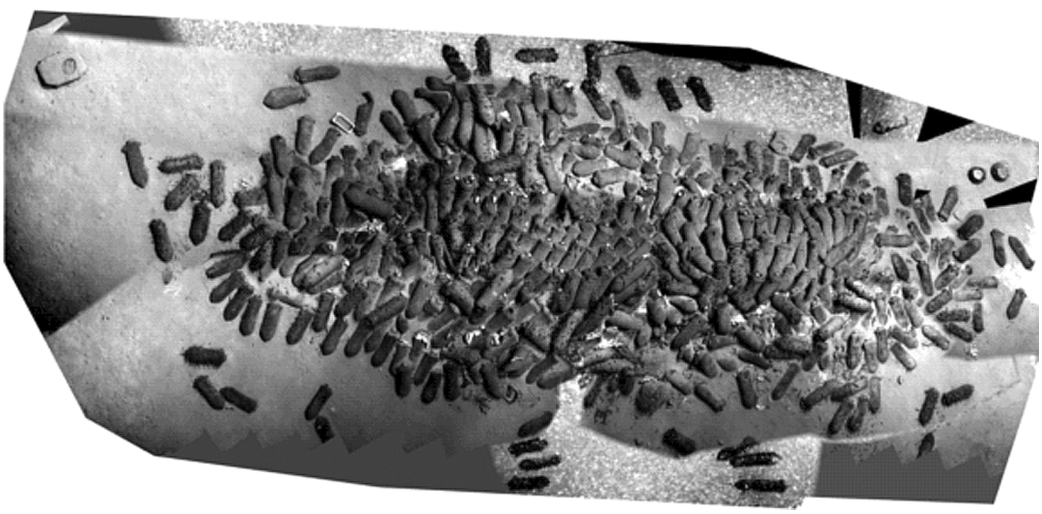


Fig. 19. Final mosaic based on finite element warping of individual strips to a high-altitude low-resolution mosaic of the site. This mosaic highlights the effectiveness of our technique for dealing with terrain effects.

that is associated with the wreck of the bulk carrier *M/V Derbyshire*, imaged by the *Argo* towed vehicle.

While the ability to mosaic images with arbitrary rotations is important in itself, our implementation is particularly suited for the characterization and enhancement of traditional navigation techniques underwater [26], because it has an implicit measure for rejecting false matches. This property is important, as it allows us to automatically use or discount information based on image mosaicking to complement our traditional acoustic navigation systems.

Table II shows a grouping of ten pairs of images with large, small, and no overlap, while Fig. 15 shows representative pairs of images from those groups. We also tabulate the values of the computed normalized correlations after: 1) calculating the best-fit rotation and scaling and 2) computing the best-fit translation associated with the rotation and scaling. The starred values in the column for translational correlation were judged to be successfully registered image pairs by human operators. The normalized translational correlation is, thus, seen to be a robust predictor of the validity of the registration process.

IV. Terrain Effects and Very-Low-Overlap Applications

So far, we have concentrated on applications where the terrain was relatively benign. For the Roman shipwreck (Fig. 10), the vessel is sitting on a sedimentary and, thus, relatively flat bottom. The whale photomosaic (Fig. 12) was generated with imagery acquired at a high altitude to minimize distortions due to 3-D effects. We now look at one example where the 3-D effects cannot be ignored. The site of interest is an Eighth Century BC Phoenician shipwreck located at a depth of 400 m off the coast of Israel. The surface expression of this site is a pile of approximately 300 closely packed amphorae. Physical oceanographic or possibly ship impact effects have scoured out a depression around the wreck, so that the difference in height from the central portion of the wreck to the depression is greater than 1 m. This difference in altitude is comparable to the 4-m altitude at which the *Jason* ROV was flown to obtain high-resolution imagery. The general geometry of such a site is shown in Fig. 16. What makes this data set particularly interesting and challenging is that:

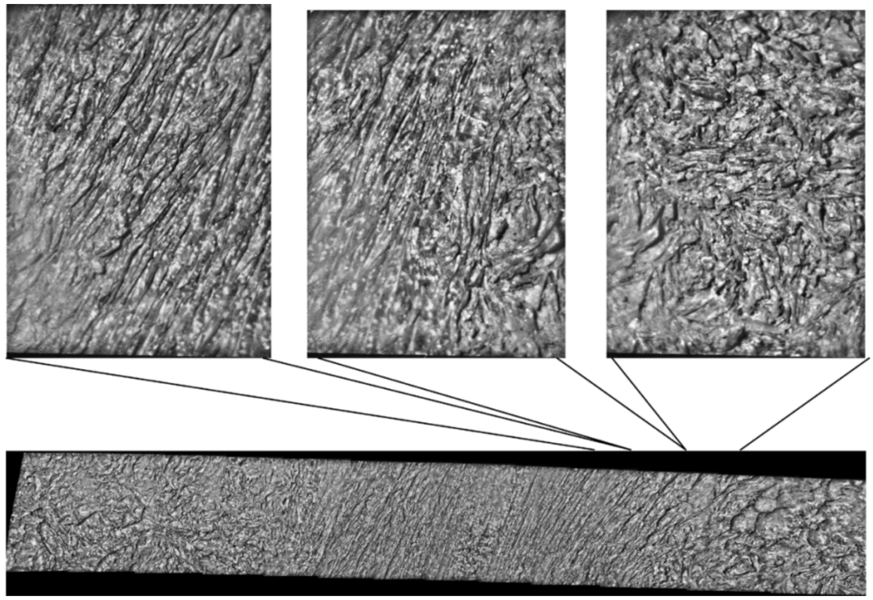


Fig. 20. Manually mosaicked *ABE* imagery of a geological channel in the East Pacific Rise. The very low overlap, corresponding to poorly lit boundaries (cf Fig. 6), is a significant challenge to the next generation of automated photomosaicking algorithms for underwater use.

1) There is a paucity of features in the depression off the main wreck in comparison to features on the wreck itself;
2) The imagery is very uniformly illuminated, allowing us to separate lighting effects from terrain effects;
3) The amphorae are all identical in size and serve as convenient markers, making it easy to spot and keep track of distortions in the imagery.

Thus, if consider three overlapping passes over the wreck (Fig. 17), the pass over the center is relatively undistorted, as all the amphorae within the imagery lie at roughly the same depth. The two passes on either side, however, span the central mound as well as the depression around the site. As we would expect from the geometry illustrated in Fig. 16, each of these passes gets skewed away from the amphorae pile as successive images are added to the strip. Also, as expected, the amphorae that have spilled over into the depression appear much bigger than those in the central pile even though, as noted earlier, they are all the same size. Essentially, the problem is of projecting a 3-D unstructured terrain on a 2-D planar surface. Several possible solutions suggest themselves. The overlapping imagery could be utilized to obtain a 3-D image reconstruction of the site under consideration. We could also utilize other sensor data to correct or compensate for the distortions. However, we present another alternative approach based solely on surveying with the camera itself, which builds on work already discussed in this paper.

We surveyed the wreck from a much higher altitude to minimize the effects of perspective distortion. The imagery acquired was of quite poor resolution due to backscatter effects, but was easy to mosaic. The results of this pass are shown in Fig. 18. The individual strips could then be overlaid over this mosaic, which served as an undistorted basemap for the individual strip mosaics distorted by the effects of terrain. We utilized a finite element-warping process for the overlay. Manually, we can identify common tie points between the individual strips and the high-altitude mosaic. If we consider tie points (we used close to 150 points) distributed across the mosaic, we can tesselate the convex hull associated with these points by the Delauney triangulation. A finite element warp with local support [27] was then used to warp the strip mosaics. The results of this process are shown in Fig. 19. The three strips have been merged to form a consistent high-resolution mosaic of the site, even in the face of significant terrain.

V. Conclusion

In this paper, we have outlined the issues that we believe make the task of mosaicking underwater challenging. We have looked specifically at the effects of lighting, power, and camera-to-light separation across a variety of imaging platforms typically in use underwater. We have shown how modifications to spatial algorithms developed for use on land can serve underwater applications. We have also presented a Fourier-based technique for image registration and mosaicking for unstructured surveys. Finally, we outlined a methodology for dealing with unstructured terrain underwater.

Our work in this area continues. In particular, we present a manually constructed photomosaic in Fig. 20, derived from imagery acquired by the *ABE* AUV. The challenge in AUV applications is to deal with very low overlap between successive frames due to severe power constraints. Further, the overlap lies on the boundary regions where the effects of low power lighting geometry are at their worst. Finally, we are also actively pursuing the fusion of photomosaicking techniques with high-resolution microbathymetric mapping to exploit the complementary nature of optical and acoustic imaging.

References

[1] W. Beebe, *Half Mile Down*. New York: Harcourt, Brace, 1934.
[2] A. C. Vine, "Early history of underwater photography," *Oceanus*, vol. 18, no. 3, pp. 2–10, 1975.
[3] R. B. Patterson, "Future developments in deep-sea imaging," *Oceanus*, vol. 18, no. 3, pp. 17–23, 1975.
[4] R. D. Ballard, "Photography from a submersible during project famous," *Oceanus*, vol. 18, no. 3, pp. 40–43, 1975.

[5] M. M. Sulanowska, S. E. Humphris, and J. C. Howland, "Detailed analysis of the surface morphology of the active TAG hydrothermal mound by mosaicking of digital images," *EOS, Trans. Amer. Geophys. Union*, vol. 77, pp. 768–768, 1996.

[6] J. C. Howland, H. Singh, M. Marra, and D. Potter, "Digital mosaicking of underwater imagery," *Sea Technol.*, pp. 65–69, 1999.

[7] R. Marks, S. Rock, and M. Lee, "Real-time video mosaicking of the Ocean floor," *IEEE J. Oceanic Eng.*, vol. 20, pp. 229–241, July 1995.

[8] S. Negahdaripour, X. Xu, and A. Khamene, "Applications of direct 3D motion estimation for underwater machine vision systems," in *Proc. IEEE OCEANS'98*, Nice, France, Sept. 1998.

[9] E. Trucco, Y. R. Petillot, T. I. Ruiz, K. Plakas, and D. M. Lane, "Feature tracking in video and sonar subsea sequences with applications," *Comput. Vision Image Understanding*, vol. 79, pp. 92–122, 2000.

[10] N. Gracias and J. Santos-Victor, "Underwater video mosaics as visual navigation maps," *Comput. Vision Image Understanding*, vol. 79, pp. 66–91, 2000.

[11] C. Tomasi and T. Kanade, "Detection and tracking of point features," Carnegie Mellon Univ., Pittsburgh, PA, Tech. Rep. CMU-CS-91-132, 1991.

[12] H. S. Sawhney and R. Kumar, "True multi-image alignment and its application to mosaicing and lens distortion correction," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, pp. 235–243, March 1999.

[13] S. Peleg and J. Herman, "Panoramic mosaics by manifold projection," in *Proc. Int. Conf. Computer Vision Pattern Recognition*, 1997, pp. 338–343.

[14] Y. Ma, J. Kosecka, and S. Satry, "Linear differential algorithm for motion recovery: A geometric approach," *Int. J. Comput. Vision*, vol. 36, pp. 71–89, January 2000.

[15] S. Q. Duntley, "Light in the sea," *J. Opt. Soc. Amer.*, vol. 53, pp. 214–233, 1963.

[16] F. M. Caimi, Ed., *Selected Papers on Underwater Optics*. ser. SPIE Milestone. Bellingham, WA: SPIE, 1996, vol. MS 118.

[17] J. S. Jaffe, "Computer modeling and the design of optimal underwater imaging systems," *IEEE J. Oceanic Eng.*, vol. 15, pp. 101–111, Apr. 1990.

[18] B. J. McGlamery, "A computer model for underwater camera systems," in *Ocean Opt. VI*, S. Q. Duntley, Ed. Bellingham, WA: SPIE, 1979, vol. 28.

[19] P. M. Narendra and R. C. Fitch, "Real-time adaptive contrast enhancement," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-3, pp. 655–661, Nov. 1981.

[20] S. Pizer *et al.*, "Adaptive histogram equalization and its variations," *Comput. Vision Graphics Image Process.*, vol. 39, no. 3, pp. 355–368, 1987.

[21] J. A. Stark and W. J. Fitzgerald, "Model-based adaptive histogram equalization," *Signal Process.*, vol. 37, pp. 193–200, 1994.

[22] L. G. Brown, "A survey of image registration techniques," *ACM Comput. Surv.*, vol. 24, no. 4, pp. 325–376, 1992.

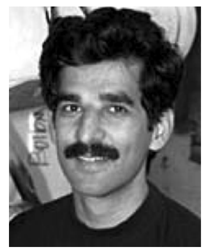
[23] R. D. Ballard, A. McCann, D. Yoerger, L. Whitcomb, D. Mindell, J. Oleson, H. Singh, B. Foley, J. Adams, and D. Piechota, "The discovery of ancient history in the deep sea using advanced deep submergence technology," *J. Deep Sea Res.* , to be published.

[24] E. DeCastro and C. Morandi, "Registration of translated and rotated images using finite Fourier transforms," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, pp. 700–703, Sept. 1987.

[25] B. S. Reddy and B. N. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE Trans. Image Processing*, vol. 5, pp. 1266–1271, Aug. 1996.

[26] L. L. Whitcomb, D. R. Yoerger, H. Singh, and J. Howland, "Advances in underwater robot vehicles for deep ocean exploration: Navigation, control, and survey operations," in *Robotics Research—The Ninth International Symposium*, London, U.K., 2000.

[27] G. Wolberg, *Digital Image Warping*. Los Alamitos, CA: IEEE Computer Soc. Press, 1990.



Hanumant Singh received the B.S. degrees as a distinguished graduate in computer science and electronics and computer engineering from George Mason University, Fairfax, VA, in 1989 and the Ph.D. degree from the Massachusetts Institute of Technology, Cambridge/Woods Hole Oceanographic Institution (WHOI), Woods Hole, MA, Joint Program in 1995.

He has been with WHOI since 1995, where his research interests include high-resolution imaging underwater and issues associated with docking, navigation, and the architecture and use of underwater vehicles.

Jonathan Howland received the B.S. degree in surveying engineering from the University of Maine, Orono.

He was with Western Geophysical, Automeric, Inc., and the Applied Physics Laboratory, Johns Hopkins University, Baltimore, MD, before joining the Woods Hole Oceanographic Institution, Woods Hole, MA, in 1990. He has worked extensively in the area of underwater imaging systems and has been one of the chief software architects for a variety of underwater platforms, including the *Jason II* and *Hercules* ROV systems.

Oscar Pizarro (S'92) received the Engineer degree in electronic engineering from the Universidad de Concepcion, Concepcion, Chile, in 1997. He is currently pursuing the Ph.D. degree at the Massachusetts Institute of Technology, Cambridge/Woods Hole Oceanographic Institute, Woods Hole, MA, joint program, where he is involved with underwater imaging and robotic underwater vehicles.