Toward Large-Area Mosaicing for Underwater Scientific Applications

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

***Abstract*—Severe attenuation and backscatter of light fundamentally limits our ability to image extended underwater scenes. Generating a composite view or mosaic from multiple overlapping images is usually the most practical and flexible way around this limitation. In this paper, we look at the general constraints associated with imaging from underwater vehicles for scientific applications—low overlap, nonuniform lighting, and unstructured motion—and present a methodology for dealing with these constraints toward a solution of the problem of large-area global mosaicing. Our approach assumes that the extended scene is planar and determines the homographies for each image by estimating and compensating for radial distortion, topology estimation through feature-based pairwise image registration using a multiscale Harris interest point detector coupled with a feature descriptor based on Zernike moments, and global registration across all images based on the initial registration derived from the pairwise estimates. This approach is purely image based and does not assume that navigation data is available. We demonstrate the utility of our techniques using real data obtained using the Jason remotely operated vehicle (ROV) at an archaeological site covering a hundreds of square meters.**

***Index Terms*—Feature-based registration, global alignment, global registration, multiscale Harris interest-point detector, radial distortion compensation, underwater mosaicing, Zernike moments.**

I. Introduction

OPTICAL imaging underwater offers a high level of detail and presents information in a way that can be naturally interpreted by humans. Unfortunately, the underwater environment imposes severe limitations on optical imaging. For many underwater applications, light attenuation and backscatter limit the coverage of a single image to only a fraction of the area of interest [1] (Fig. 2) and, in spite of improvements in imaging hardware and processing techniques, the practical coverage of a single image is on the order of 10 m^2.

This limited coverage is a serious impediment to underwater exploration and science as sites of interest usually encompass much larger scales—hydrothermal vents and spreading ridges in geology [2], ancient shipwrecks and settlements in archaeology [3], [4], forensic studies of modern shipwrecks and airplane accidents [5], [6], and surveys of benthic ecosystems and species in biology [7]–[9].

Generating a composite view from multiple overlapping images is usually the most practical and flexible way around this limitation. Recent years have seen significant advances in mosaicing [10]–[14] and full three-dimensional (3-D) reconstruction [15]–[18], although these results are land based and do not address issues specific to underwater imaging.

Due to the rapid attenuation of electromagnetic radiation underwater, ambient lighting is practically nonexistent after the first few tens of meters of depth. This implies that most optical imaging in the deep ocean is done by vehicles carrying their own light source or sources. Power and/or size limitations result in lighting patterns that are far from uniform (Figs. 1 and 3) and that produce shadows that move across the scene as the vehicle moves. Variable lighting may be partially compensated for by using high-dynamic-range cameras and adaptive histogram specification [19]. However, another serious concern is the advent of battery-powered autonomous underwater vehicles (AUVs) for imaging surveys [2], which impose additional constraints given their limited energy budgets. AUV surveys are typically performed with strobes rather than continuous lighting and acquire low-overlap imagery in order to preserve power and to cover greater distances. Optical imaging is also performed from towed sleds, where cable dynamics can induce significant changes in camera position and orientation from frame to frame [5]. In its most general form, generating composite views underwater requires dealing with imagery acquired with low overlap, terrain relief, nonuniform lighting, and unstructured surveys.

A. Overview of Related Work

Underwater mosaicing has been studied mostly in the context of vision-based navigation and station keeping close to the sea floor [20]–[23]. Global alignment is considered in [20], [21] through the two-dimensional (2-D) topology of image placements and a "smoothing" stage to adjust the placement of images in the mosaic. Real-time constraints force the homographies to be pure translations, good enough for local navigation but not for an accurate rendition of the sea floor. Registration is based on matching borders of zones with different textures (the sign of the Laplacian of Gaussian). Variable lighting destroys the brightness constancy constraint (BCC), which is the key assumption in most direct (intensity)-based methods [24]. In [22], a modified BCC is used to account for light attenuation underwater, but this method has not been proven for low overlap imagery and for unstructured terrain.

Recently, Gracias and Victor-Santos [23] presented a global-alignment solution for underwater mosaicing with excellent results for video-based imagery over an area of approximately 50 m^2. Given the relatively slow speed of underwater vehicles, consecutive frames at video rates provide narrow baseline imagery. This allows them to simplify the matching stage by

Manuscript received May 29, 2003; revised October 7, 2003.

The authors are with the Woods Hole Oceanographic Institution, Woods Hole, MA 02543 USA (e-mail: opizarro@whoi.edu; hsingh@whoi.edu).

Digital Object Identifier 10.1109/JOE.2003.819154

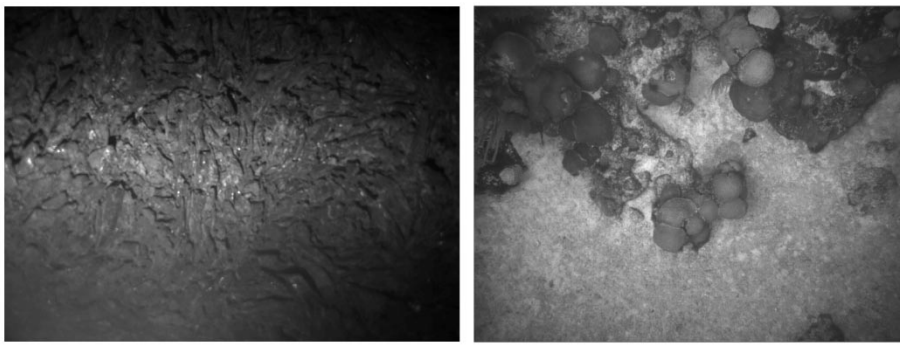


Fig. 1. Sample images from a lava flow and a coral reef survey performed by AUVs. The strong falloff in lighting is typical of energy-limited vehicles.

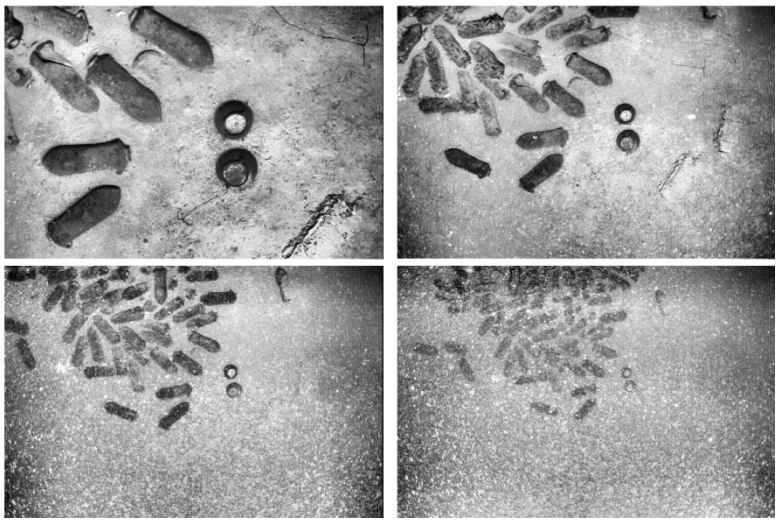


Fig. 2. Set of images taken at different altitudes (from 3.5 to 12.5 m in 3-m increments). Light attenuation and backscatter limit the altitude from which images can be acquired. Covering an area of interest may require several hundreds of images.

assuming that translation is the dominant motion between consecutive frames (correlation is used to match feature points). Even though their global mosaic is constructed with a subset of images with significant interimage motion, the feature matching is performed with high overlap (basically, the homography between two images with low overlap is calculated as the concatenation of video-rate homographies). It is not clear how this technique would fare when only low-overlap imagery is available. In addition, their method does not account for lens distortion, which can have a significant impact at larger scales. Note that the main concern of all these approaches is vision-based navigation rather than building large mosaics.

Mosaicing makes assumptions about camera motion and scenery to merge several images into a single view, effectively increasing the field of view of the camera without sacrificing image resolution. The key assumption for a distortion-free mosaic is that the images either come from a purely rotating camera or that the scene is planar [25], [15]. In either case, the camera motion does not induce any parallax and, therefore, no 3-D effects are involved. These assumptions often do not hold in underwater applications and obvious distortions in the mosaic are usually present (note that light attenuation and backscatter in underwater imagery rules out the traditional land-based approach of acquiring distant, nearly orthographic imagery). In contrast to mosaicing, one could use information from multiple views to attempt full 3-D reconstruction, as in structure-from-motion (SFM) [15] and photogrammetry [26]. Even though recovering a 3-D structure is the proper approach to composite image generation when dealing with a translating camera over nonplanar surfaces, these techniques are considerably more complex, even for land-based applications (with high overlap, structured motion, and good lighting). However, full 3-D reconstruction is not always possible for very low overlap surveys (where many points are imaged only once). The emphasis in this paper is on 2-D mosaicing, based on the belief that there are fundamental challenges in underwater imaging (large-scale relating images) that can be addressed in the simpler context of mosaicing approximately planar scenes to provide a global perspective of the area of interest.

B. Paper Overview

In summary, there currently is no practical, robust, and repeatable way to generate an underwater mosaic that combines hundreds to thousands of images acquired with low overlap, poor lighting, and possibly in an unstructured fashion. This paper demonstrates large-area underwater mosaicing by addressing all these issues with an effective image-registration technique in a global-mosaicing framework.

Our approach for relating images is based on a multiscale feature detector with matching based on descriptors invariant

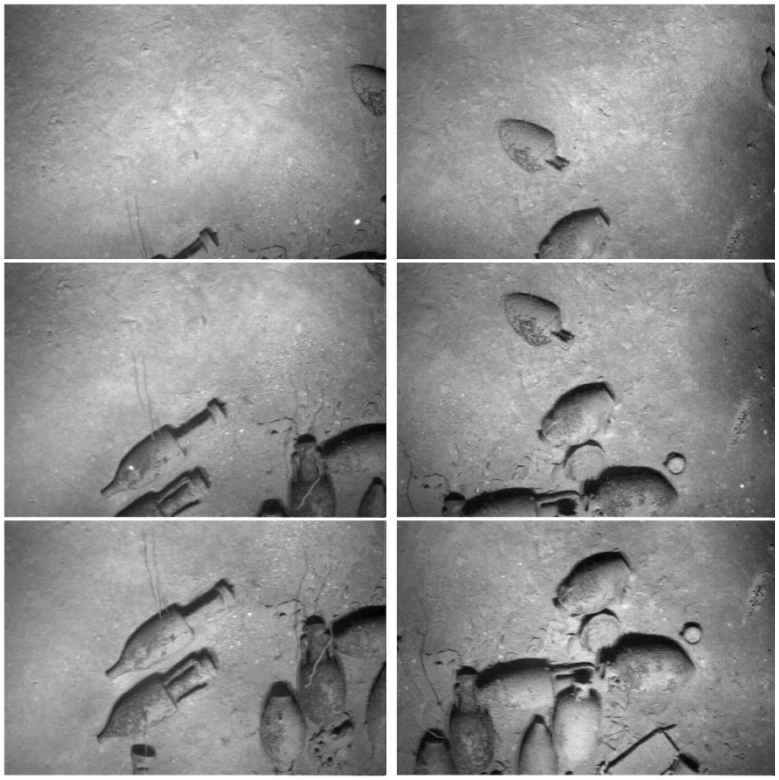


Fig. 3. Sample of the set of 29 images used as a working example to illustrate some of the stages of the mosaicing process. Each column represents a section of adjacent survey tracklines. Images in a column represent a temporal sequence (the vehicle moved down the left column and up the right column with approximately constant heading). Note the low spatial overlap and the failure of the BCC, due to underwater lighting.

to rotation, scaling, and affine changes in intensity. Global registration is performed by considering all known overlaps between images when solving for the transforms that map images onto the mosaic frame. New overlaps become apparent as the topology of the mosaic is refined. Note that this approach is purely image based and does not assume that navigation data is available.

Results in recent years [11], [27], [14], [28], [29] have clearly established the need for global considerations when generating a mosaic that consists of multiple tracklines. Our approach breaks down the problem into three distinct phases, as follows:

- radial-distortion compensation;
- topology estimation through pairwise registration;
- global registration.

The second and third phases (Fig. 4) are similar to other global-mosaicing approaches [11], [23], the difference lying in our emphasis in using low-overlap, feature-based registration. The formulation of the global-registration problem is kept simple by only explicitly solving for the homographies from image to mosaic and not the positions of features. In addition, the topology-estimation stage uses projective models that are linear in their parameters, which allows for a linear least-squares solution of homographies. The global-registration stage uses planar projective transforms that require solving a nonlinear least-squares problem.

The first phase arises from the realization that when dealing with large image sets and low overlap, effective mosaicing underwater requires compensating for radial distortion. This is implemented as a critical preprocessing step whereby images are prewarped to compensate for radial distortion before attempting to mosaic them. To accomplish this, a global mosaic is created from a small subset of underwater images of an approximately planar surface. The solution for the mosaic considers a nonlinear projection model that accounts for radial distortion. The result is that the images are registered onto the mosaic in a globally consistent fashion and are corrected for radial distortion. This correction is then applied to the full image set before attempting to mosaic it, allowing the use of simpler transforms in the topology-estimation and global-mosaicing stages.

The rest of the paper details our approach. Section II describes how feature detection and matching is performed. Section III addresses topology estimation, while Section IV deals with global mosaicing. Radial-distortion compensation can be seen as an extension of the global-mosaicing technique and is treated as such in Section V. Results for a large mosaic are presented in Section VI. Final remarks are offered in Section VII.

II. Feature Detection and Matching

A. Relating Overlapping Imagery

A pair of overlapping images of the same scene can be related by a *motion model* [30], [31]. A 2-D motion model imposes on an image a single global 2-D transformation that defines the displacement of every pixel. This type of transform is the basis of 2-D registration for mosaicing. A 3-D motion model has a global-motion component (related to the camera motion), but also a set of local parameters representing the 3-D structure.

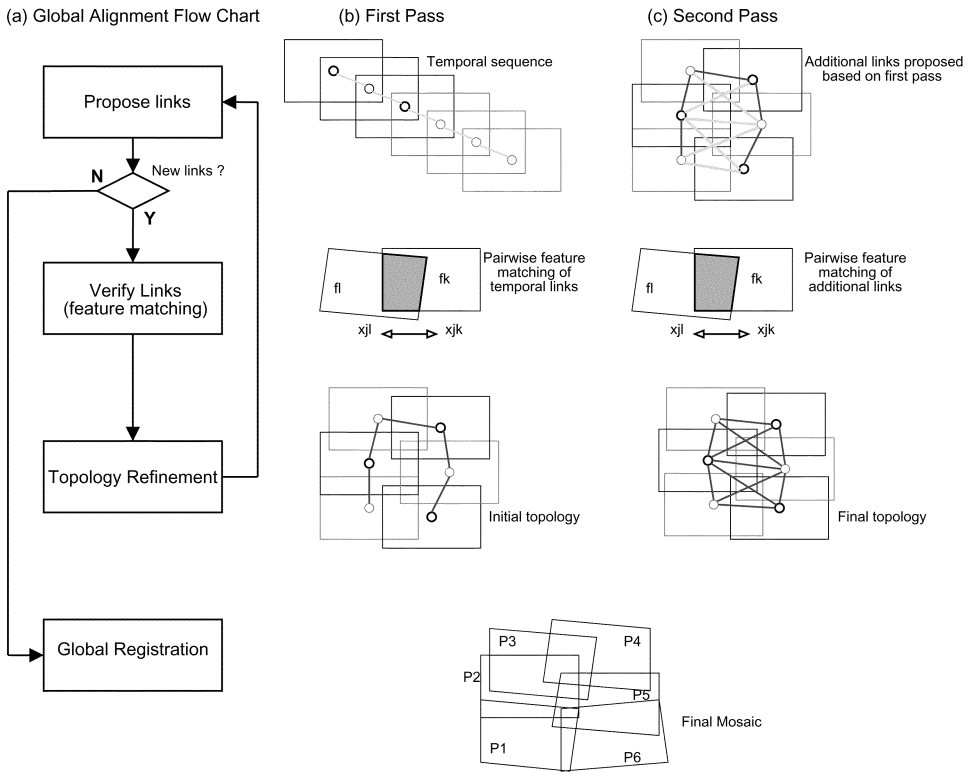


Fig. 4. (a) Flowchart for mosaicing with global alignment through topology estimation. (b) and (c) illustrate the process for a set of six images. Proposed links (based on possible overlap given the current topology estimate) are depicted in light gray, verified links (based on a successful feature matching) in black. Feature matching is represented for the jth feature x_j between images f_l and f_k as $x_{jl} \longleftrightarrow x_{jk}$.

Techniques that recover motion models have traditionally been divided into feature-based (indirect) methods and intensity-based (direct) methods. Feature-based methods extract a relatively small set of distinct features from each image and then recover their correspondence. These matching features are then used to model motion. The Harris corner detector [32] forms the basis of some successful feature-based underwater mosaicking work [23]. Indirect methods must address the *correspondence problem*, i.e., they need to determine which features match between images before solving for the transform that relates an image pair. Traditional implementations pair up features based on a similarity measure such as correlation [17]. Since mismatches occur, robustness is increased by adding an outlier rejection stage [33].

Direct methods, on the other hand, estimate motion or shape from measurable quantities at each pixel of the images. Most direct methods rely on the BCC [24] and are implemented in a coarse-to-fine manner within a multiresolution pyramid. This allows initial misalignments of up to 10%–15% of image size. The BCC is not valid underwater because vehicles carry their own light source. As the camera moves, so does the light source, resulting in significant changes in intensity as well as moving shadows.

Our approach is largely dictated by the particular requirements of underwater imaging. Low-overlap situations with unknown motion between images are not effectively dealt with by registration methods that assume pure translation or that use whole image information to determine similarity. Direct methods, even in coarse-to-fine implementations [30], assume that the motion is mainly translational between frames. These methods have shown to be extremely effective with short baselines, i.e., high overlap [34]. However, rotations of more than 10–15° require other methods to prealign images for reliable registration. Our own extensive testing with underwater imagery indicates that direct methods alone, even those that use local normalized correlation as a similarity measure [35], are not suitable as a general registration technique. Imagery with significant rotation, scaling, or even affine transforms can be registered using frequency-based techniques [36], [37]. These methods implicitly rely on large overlap in order to effectively relate the spectra. Again, the type of imagery obtained in underwater surveys does not satisfy these assumptions. Our approach, inspired by recent results in wide baseline matching [38], [39], uses feature detection and matching as a first stage in image registration. The key is using feature detectors, feature descriptors, and similarity measure that are robust to changes in rotation, scaling, and lighting.

B. Scale-Space Representation

The process by which information is extracted from images can be seen as the interaction of the image data with operators. The design as well as the interpretation of the response of these operators is one of the key issues in machine vision. Scale-space theory offers a working framework to deal with these problems based on the observation that image data presents structures at different scales and that one cannot expect to know beforehand the scale of relevant structure. The natural approach is then to

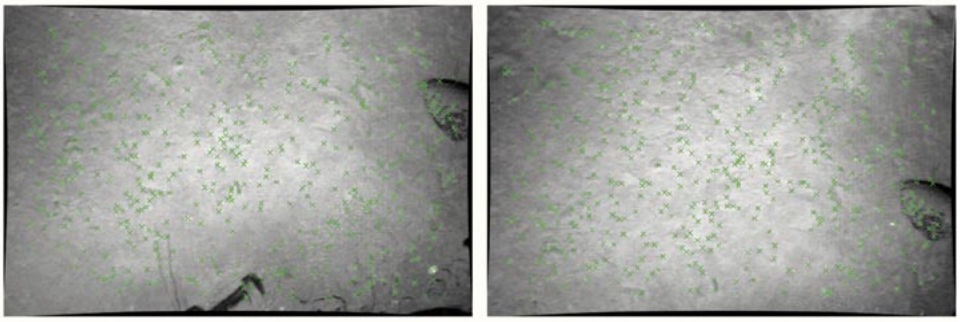


Fig. 5. Independently derived feature points for a pair of overlapping images. The multiscale Harris interest-point detector extracted the feature pionts. Notice the significant changes in lighting.

consider the image at all scales simultaneously. In the most general case, convolution with Gaussian kernels and their derivatives form a class of low-level operators for scale-space representations [40] and have shown great promise in feature detection and matching [41]. Being able to describe a feature point not only by its location but also by its characteristic scale improves the ability to deal with imagery that contains objects with different scales and/or imagery of the same object taken at different scales. Both situations can be expected to occur in underwater imagery.

C. Interest Points

Describing features in a way that is invariant to expected geometric and photometric transformations is important for succesful matching. Perhaps even more imporatant is to have confidence that we will actually identify the same interest point in two overlapping images. We choose to detect features with a multiscale version of the Harris interest-point detector [32], [42] since it is designed to identify the same interest point in the presence of rotation and scale changes. Fig. 5 shows a typical set of Harris feature points for an overlapping pair of images.

The Harris measure is the second-moment matrix, describing the curvature of the autocorrelation function in the neighborhood of a point $\mathbf{x}$ as

$$\mu(\mathbf{x}, \sigma_I, \sigma_D) = \sigma_D^2 g(\sigma_I) * \begin{bmatrix} L_x^2(\mathbf{x};\sigma_D) & L_x L_y(\mathbf{x};\sigma_D) \\ L_x L_y(\mathbf{x};\sigma_D) & L_y^2(\mathbf{x};\sigma_D) \end{bmatrix} \tag{1}$$

where $*$ is convolution operator, σ_I is the integration scale, σ_D the derivation scale, $g(\sigma)$ the Gaussian of standard deviation σ, and L the image I smoothed by a Gaussian, such that L_x represents a smoothed derivative

$$L_x(\mathbf{x};\sigma_D) = \frac{\partial}{\partial x} g(\sigma_D) * I(\mathbf{x}). \tag{2}$$

The second moment matrix offers a description of local image structure at a given scale. An interest point or a corner point will have a μ with two positive eigenvalues (significant changes in intensity in any direction). An edge will present one positive and one zero eigenvalue and a uniform area will have two zero eigenvalues.

In the multiscale framework, a feature point will have a location and a characteristic scale that is determined after representing the image at different scales. Rather than convolve the image with a Gaussian of increasing variance, the image can be downsampled so as to reuse the same Gaussian mask. This will prove particularly useful in describing the neighborhood of the feature point at that given scale. Our actual implementation uses a pyramid representation of the image to change scales (σ_D and σ_I) and increases computational efficiency [43].

The characteristic scale of a feature is considered to be the scale at which the scale-normalized Laplacian $|\sigma^2(L_{xx}(\mathbf{x};\sigma) + L_{yy}(\mathbf{x};\sigma))|$ is maximum [44]. A variable threshold on the strength of the interest point (the smaller eigenvalue of μ at the characteristic scale) limits the number of feature points to a manageable size (400–600 points are usually needed to ensure matching in a low-overlap situation).

D. Description of Features

Matching in narrow baseline applications is traditionally performed with an intensity-based similarity measure between image patches centered around feature points. In its simplest and most common form, the description of the feature point is the image patch around it and the similarity measure is usually some variant of the sum of squared differences or cross correlation. For wide-baseline (low-overlap) mosaicing applications where the terrain is approximately planar, the camera has a narrow field of view and its optical axis is roughly perpendicular to the terrain, the appearance of a particular image patch does not undergo significant perspective transformation. This is the case for most underwater imaging vehicles, which are designed for passive stability in pitch and roll and operate deep enough to be immune to wave action. Matching features in the presence of heading and altitude changes [larger one-dimensional (1-D) rotations and scale changes] requires compensating for the motion (with prior knowledge) or using a description and similarity measure that are invariant to these transformations. The latter approach is used in this paper. Instead of directly comparing intensities of the image patches, the patches are transformed into a descriptor vector that is invariant to 1-D rotation and scaling. For some degree of robustness to lighting changes, the descriptor is modified to also be invariant to affine changes in intensity. Matching is then based on a distance measure between descriptor vectors.

Image patches have been described by differential [41], [45] and moment invariants [46], [47]. Differential invariants are constructed from combinations of intensity derivatives that are constant to some geometric and radiometric transformations

such as translation, rotation, scaling, and affine brightness changes. Moment invariants can be constructed from nonlinear combinations of geometric moments (the projection of the image patch $f(x,y)$ onto the basis set of monomials $x^p y^q$). Since the basis set for this projection is not orthogonal, these invariant moments contain redundant information and have a limited ability to represent an image in the presence of noise. Orthogonal moments based on orthogonal polynomials such as Zernike moments have been shown to be invariant to some linear operations, have superior reconstruction capabilities in the presence of noise, and low redundancy as compared to other moment representations [46], [48], [49]. Extensive testing with underwater imagery has led us to describe image patches around feature points using a formulation based on Zernike moments.

1) Zernike Moments: Zernike moments are derived from Zernike polynomials, which form an orthogonal basis over the interior of the unit circle, i.e., $x^2 + y^2 = 1$ [48]. If we denote the set of polynomials of order n and repetition m by $V_{nm}(x,y)$, then since these polynomials are complex, their form is usually expressed as

$$V_{nm}(x,y) = V_{nm}(\rho,\theta) = R_{nm}(\rho)e^{jm\theta} \tag{3}$$

with n a positive or zero integer, m an integer such that $n - |m|$ is even, and $|m| \leq n$. We've also defined polar coordinates $\rho = \sqrt{x^2+y^2}, \theta = \arctan(y/x)$. Note $V^*_{nm}(\rho,\theta) = V_{n,-m}(\rho,\theta)$.

The radial polynomial $R_{nm}(\rho)$ is real and of degree $n \geq 0$, with no power of ρ less than $|m|$.

$$R_{nm}(\rho) = \sum_{s=0}^{\frac{n-|m|}{2}} \frac{(-1)^s (n-s)!}{s!\left(\frac{n+|m|}{2}-s\right)!\left(\frac{n-|m|}{2}-s\right)!}\rho^{n-2s}. \tag{4}$$

The Zernike moment of order n with repetition m corresponding to the projection of an image function $f(x,y)$ (in the unit circle) is given by

$$A_{nm} = \frac{n+1}{\pi}\int\int_{x^2+y^2\leq 1} f(x,y)V^*_{nm}(x,y)\,dx\,dy. \tag{5}$$

Note that A_{nm} is complex and $A^*_{nm} = A_{n,-m}$. In the case of a discrete image $f[x,y]$, the moments are approximated as

$$A_{nm} = \frac{n+1}{\pi}\sum_x\sum_y f[x,y]V^*_{nm}(x,y), \quad x^2+y^2 \leq 1. \tag{6}$$

The magnitude of Zernike moments are rotationally invariant, i.e., corresponding Zernike coefficients of two image patches that differ only by a rotation have the same magnitude, and their phase difference is related to the angle of rotation. For two images that differ by a rotation ϕ

$$g(r,\theta) = f(r,\theta+\phi) \tag{7}$$

their Zernike moments are related by

$$A_{nm}(g) = A_{nm}(f)e^{jm\phi}. \tag{8}$$

Note that the recovery of the rotation angle using moments with $|m| \neq 1$ is nontrivial [50] because any rotation $\alpha, (g(r,\theta) = f(r,\theta+\phi))$ of the form

$$\alpha = \phi + \frac{2\pi k}{m}, \quad k = 0,\ldots,m-1; \quad m > 0 \tag{9}$$

will produce the phase difference $m\phi$ between $A_{nm}(f)$ and $A_{nm}(g)$.

Zernike moments are not truly invariant to changes in scale, but the magnitude ratio of corresponding moments does approximate the relative scale of the image patches. Given

$$g(sr,\theta) = f(r,\theta) \tag{10}$$

the relative scale s between images f and g can be approximated by

$$s \approx \frac{|A_{nm}(g)|}{|A_{nm}(f)|} \tag{11}$$

where $A_{nm}(g)$ and $A_{nm}(f)$ are the Zernike moments for $g(r,\theta)$ and $f(r,\theta)$, respectively.

In [51], this relationship is demonstrated empirically and used successfully to mosaic images based on one matching point of interest. Their approach calculates Zernike moments for a disk around the matching point. The rotation and scaling factors that relate the images are extracted directly from the relationship between Zernike moments. Translation is dealt with by using phase correlation once images are corrected for rotation and scaling.

Our implementation uses the first 25 $(n \leq 8, m \geq 0)$ coefficients in the Zernike expansion of a disk (of radius proportional to the characteristic scale) around all interest points. A second expansion is performed with a radius of 140% of the characteristic radius to increase the discrimination between similar patterns at the characteristic scale (we picked 140% as a compromise; smaller scales do not improve matching significantly, whereas scales larger than 160% appear to make the patches dissimilar in the presence of 3-D structure or strong lighting differences).

To obtain some robustness to changes in lighting, prior to calculating the Zernike moments, the image patch W of $f(x,y)$ is demeaned and normalized by the energy content of the patch

$$N(f(x,y)) = \frac{\sum_{i,j}(f(x+i,y+j) - \bar{f}_W)}{\sqrt{\sum_{i,j}(f(x+i,y+j) - \bar{f}_W)^2}} \tag{12}$$

where $\bar{f}_W$ is the mean of $f(x,y)$ over the patch W. Notice that

$$N(f(x,y)) = N(af(x,y)+b) \tag{13}$$

effectively providing invariance to affine changes in intensity.

We construct a vector of descriptors based on the Zernike moments by using the log of the magnitude of the Zernike coefficients and the phase of the A_{n1} moments. Identical (up to a rotation) image patches should have corresponding moments with the same magnitude, but in the case where there might be a difference in the selected scale we have

$$\log(s_{nm}) = \log(|A_{nm}(f)|) - \log(|A_{nm}(g)|) \tag{14}$$

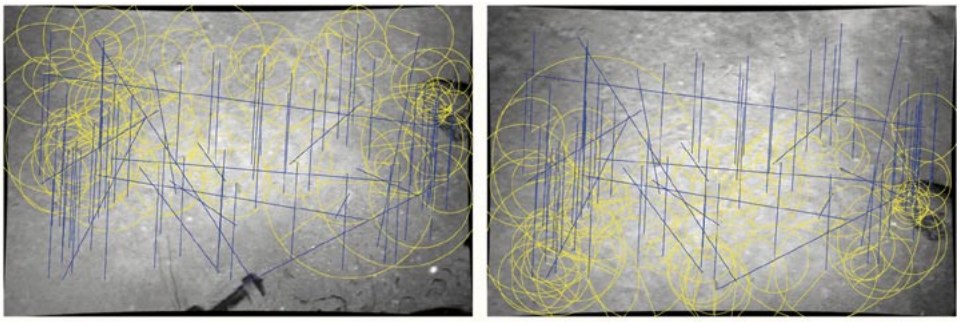


Fig. 6. Matched points according to minimum distance of feature vectors (74 matching pairs out of the 400 feature points in each image). The circles represent the characteristic scale of the feature and the blue lines indicate the interimage motion of the features. Notice that, along with a number of consistent matches (vertical), we also obtain numerous inconsistent matches across the two images.

and

$$\phi_{n1} = \angle A_{n1}(f) - \angle A_{n1}(g). \tag{15}$$

E. Feature Matching

We perform matching by finding the pair of features that minimize a distance measure between the associated descriptor vectors. The set of potential matches are then pared down by observing the distance to the next-most-similar feature. In cases where these distances are similar, the match is rejected. Finally, potential outliers are removed by using least median of squares (LMS), a robust motion-estimation technique [52], [53] similar to random sampling and consensus (RANSAC) [33]. Section II-5.2 describes the LMS approach in more detail.

1) Initial Match: The distance measure (based on Zernike moments of up to order n and repetition m) for the preliminary match is established by concatenating the s_{nm} for all considered n and m into a vector $\mathbf{s}$ and concatenating the ϕ_{n1} for all considered n into a vector $\mathbf{\Phi}$. We define the distance $d_{f,g}$ as

$$d_{f,g} = \mathrm{var}(\log(\mathbf{s})) + \mathrm{var}(\mathbf{\Phi}). \tag{16}$$

The particular distance used in this case is the sum of the variance of the log of magnitude ratios and the variance of the phase difference between corresponding coefficients. A perfect match yields corresponding coefficients with a constant magnitude ratio (scale factor) and with a constant phase difference (rotation) and their distance, as defined, is zero.

Since overlapping images will also have features in the areas that do not overlap (features that are not in common), the match is performed in both directions. That is, first from the image with fewer features to the one with more features. Those tentatively selected feature points in the second image are then matched back to the first image. If they still match to the same point, then the match is considered potentially valid.

The set of potential matches is further reduced by examining the ratio of distances between the second-best match and the best match. If this is below a given threshold (we use 1.1 since higher values tend to eliminate good matches), the match is considered of low confidence and rejected. The idea is to eliminate matches for which the minimum distance is not clearly distinguishable. Fig. 6 shows a typical correspondence set at this stage of feature matching. We note that mismatches are still possible because the descriptors are not truly invariant under general geometric and lighting variations. In addition, some repeating textures can be matched even if they do not arise from the same scene point. Thus, we see in Fig. 6 that while the predominant matches are vertical, there are a number of other matches that are considered valid between these two images.

2) Outlier Rejection (Least Median of Squares): To eliminate outliers as well as to expand the set of correct matches, an affine motion model between the two images is estimated using LMS [52]. The basic steps for outlier rejection based on LMS are as follows.

- Start with the set S of potential correspondences (based on simililarity) and set the LMS residuals M_{LS} to a large number.
- Repeat for N trials:
  - Randomly select a subset s from the potential correspondences.
  - Calculate $H(s)$, the homography implied by the subset.
  - Apply the homography to S and calculate the median of squared residuals M_s.
  - If $M_s < M_{LS}$
    - Set $H = H(s)$
    - Set $M_{LS} = M_s$.

The LMS algorithm produces a homography H that comes from the set that best explains most of the data (in the LMS sense). This homography can be used to define a set of inliers S_{in} based on a threshold on the residuals. As presented, the LMS technique will fail if there are more outliers than inliers. In such a case, a distance below the median value should be used. In addition, we consider the final homography $H(S_{\mathrm{in}})$ using all inliers. If there are not enough inliers to reliably estimate the motion, the overlap is considered nonexistent. Figs. 7, 9, and 11 show typical reduced correspondence sets, which show the strength of our technique. Figs. 8, 10, and 12 are the corresponding mosaics based on the reduced correspondence sets.

III. Topology Estimation

A. Assumptions and Approach

Our basic assumption in building a mosaic is that the images have been acquired in a temporal sequence. We do not require navigation data; we only assume that each image overlaps, at a

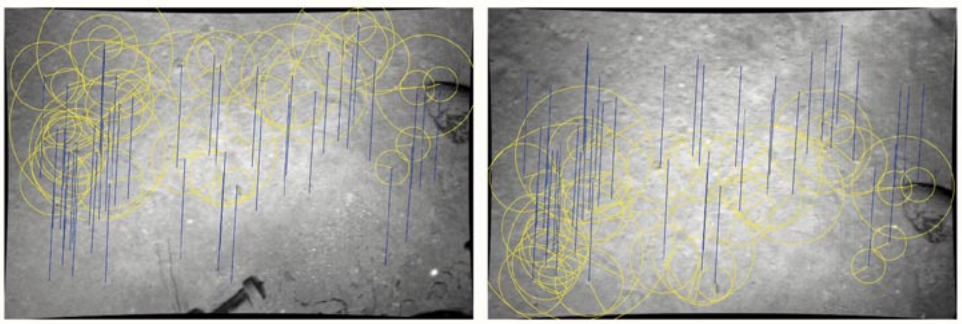


Fig. 7. Matching feature points after LMS for the set shown in Fig. 6. The LMS algorithm has reduced the set to 41 consistent matches. Circles represent the characteristic scale of the feature and blue lines indicate the interimage motion of the features. The direction of the yellow radii represents the phase of the first complex Zernike moment $(A_{1,1})$. If matching features are truly identical (up to a rotation), the difference in orientation of the yellow radii should correspond to the rotation of the feature. Changes in lighting and clipping near the image borders will introduce changes in the phase of $A_{1,1}$.

Fig. 8. Mosaic based on matching shown in Fig. 7, rendered as an average of intensities. The 41 corresponding feature points used to calculate the homographies to map onto the mosaic frame are shown in blue (top image) and green (bottom image).

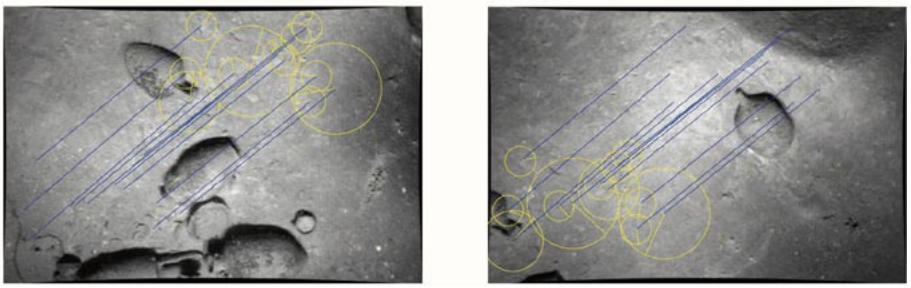


Fig. 9. Example of feature matching under low overlap.

minimum, with the preceding and succeeding images in the sequence. Early mosaicing approaches [54] constructed a mosaic by relying on the temporal-image sequence to perform pairwise registration of images, then concatenating these interimage homographies to build the mosaic. In contrast, a globally consistent mosaic uses all overlap information, including overlap from images that are not consecutive in time [11], [27], [14], [28], [29]. The overlaps between images can be considered as links that define the topology of the mosaic, with each image being a node. There are two approaches to building a mosaic using all available overlap information.

- *Incremental mosaic* is the process by which the mosaic is constructed is incremental. Each new image in the sequence is related to the previous image to find corresponding points. Based on that relationship, possible overlaps with other images are verified and the set of matching points is stored. The transforms of the new and all previous images are then calculated considering all

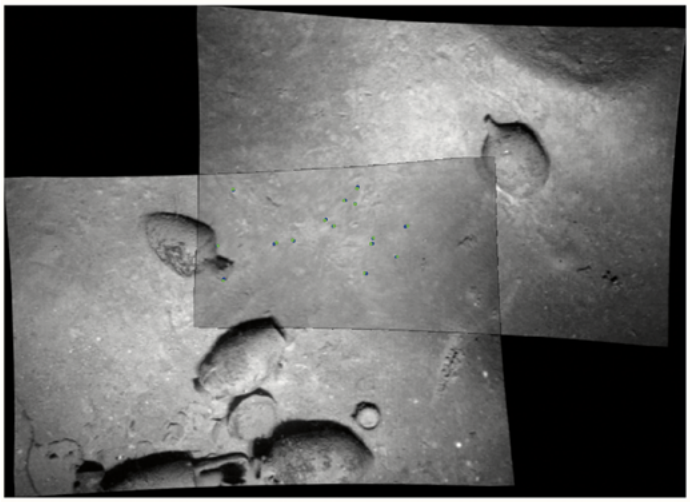


Fig. 10. Mosaic based on matching shown in Fig. 9, rendered as average of intensities. Note the misregistration on the top right corner of the overlap area. The homography does not map that corner well since there are no correspondences near it (probably due to the presence of 3-D structure).

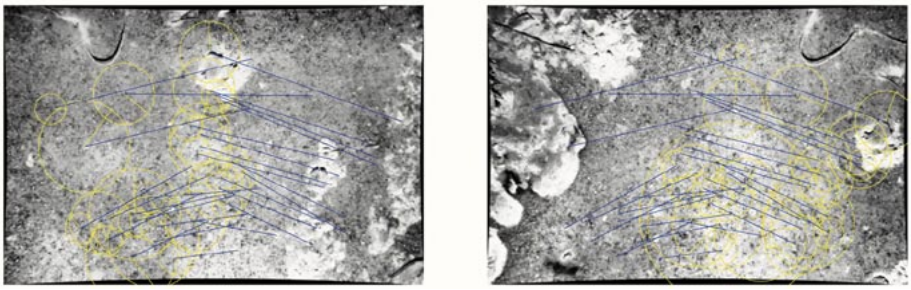


Fig. 11. Another example of feature matching under low overlap and significant rotation. This image pair belongs to a survey performed from a towed sled. Fig. 12 shows a mosaic of these images.

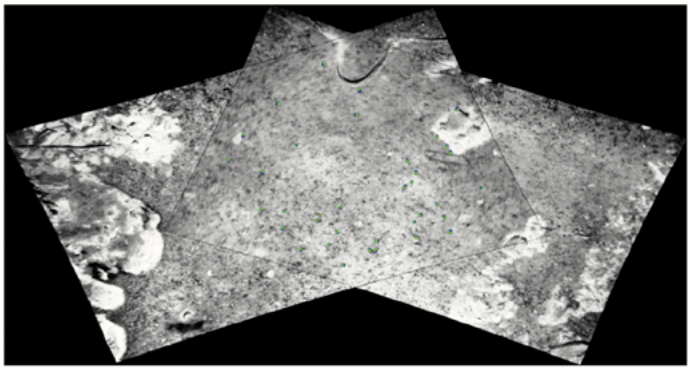


Fig. 12. Mosaic based on matching shown in Fig. 11. Rendered as average of intensities.

known overlaps. In essence, a growing global mosaic is solved after each new image is added. This could be the basis of a real-time approach.

- *Incremental links* is the approach that initially solves for the global mosaic using the overlaps of the temporal sequence. Notice that this is better than simply concatenating pairwise homographies, since the transform for each image is determined considering the overlap to the next and the previous image. Given the initial layout, all new possible overlaps (links in the topology) are verified; this information is incorporated and the transforms for all images are recalculated. This process is repeated until the topology stabilizes and no new links are added. In essence, the global mosaic is created and then refined by adding constraints as new overlaps become apparent.

The second approach, although solving a larger problem, will typically converge in just a few iterations (3–6), depending upon how aggressively we propose new links.

B. Formulation

The mosaicing problem is basically that of finding the homographies that map every image onto the mosaic frame. Stating this as a minimization of a cost function, we need to determine

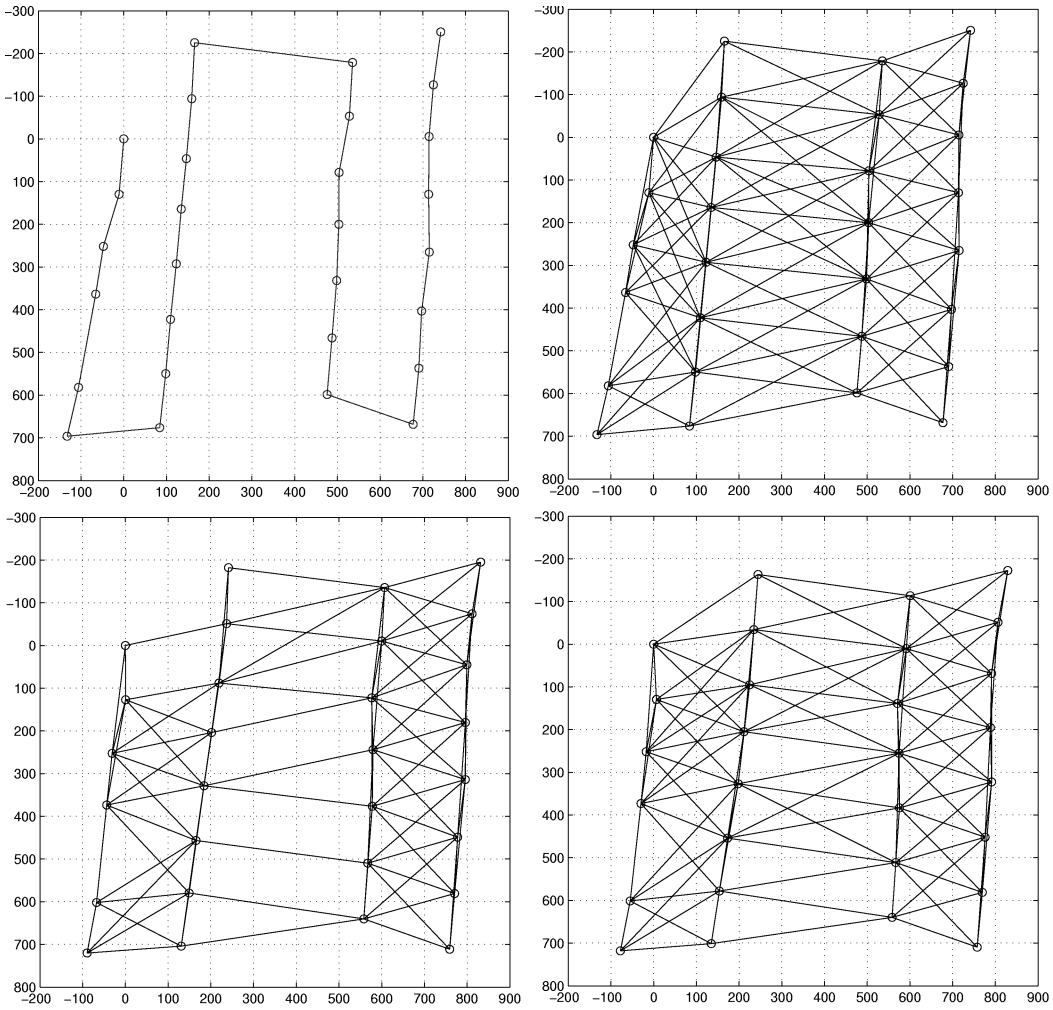


Fig. 13. (Top left) Image layout after initial topology estimate (using only temporal overlap information) and proposed links based on possible overlap given the topology estimate (top right). (Black is for verified links and gray is for proposed ones.) Axes correspond to mosaic coordinates (in pixels). Image layout after second (bottom left) and final (bottom right) topology refinements. (Only verified links are shown.)

the homographies that map every image onto the mosaic such that the distance between corresponding points is minimized in the mosaic frame.

1) Pairwise Registration: From the feature matching between image pairs k and l we have

$$\mathbf{x}_{jl} = \mathbf{x}_{jk} + \mathbf{u}(\mathbf{x}_{jk}) \tag{17}$$

where $\mathbf{x}_{jl}$ and $\mathbf{x}_{jk}$ are the 2-D positions of the jth feature point in images l and k, respectively, and $\mathbf{u}$ is the motion vector between the features.

We can register the image k onto image l by fitting a 2-D parametric transform to a set of corresponding points such that

$$\mathbf{x}_{jl} = P_{lk}(\mathbf{x}_{jk}) \tag{18}$$

where P_{lk} is the parametric transform that maps image k onto image l.

The parameters of P_{lk} are determined in a least-squares formulation by minimizing the distance between the features in l and the corresponding features in k, mapped onto l

$$\min_{P_{lk}} \sum_{j} |\mathbf{x}_{jl} - P_{lk}(\mathbf{x}_{jk})|. \tag{19}$$

2) Global Registration: To create a mosaic that is globally consistent, we solve for the transforms considering all overlaps. We generalize the pairwise registration to all images by requiring that all images of feature j should map to the same point in the mosaic. This is implemented as a least-squares problem, solving for the transform parameters that minimize the distance between corresponding feature points on the mosaic frame.

The cost function to minimize in this case is

$$J_t = \sum_{l,k} \sum_{j} |P_l(\mathbf{x}_{jl}) - P_k(\mathbf{x}_{jk})| \tag{20}$$

where l, k defines a link (images l and k overlap). The positions of feature j on images l and k are $\mathbf{x}_{jl}$ and $\mathbf{x}_{jk}$, respectively. They map onto the mosaic at positions $P_l(\mathbf{x}_{jl})$ and $P_k(\mathbf{x}_{jk})$, respectively.

This function is minimized over the parameters for the transforms for all images

$$\min_{P_1 \ldots P_N} \sum_{l,k} \sum_{j} |P_l(\mathbf{x}_{jl}) - P_k(\mathbf{x}_{jk})|. \tag{21}$$

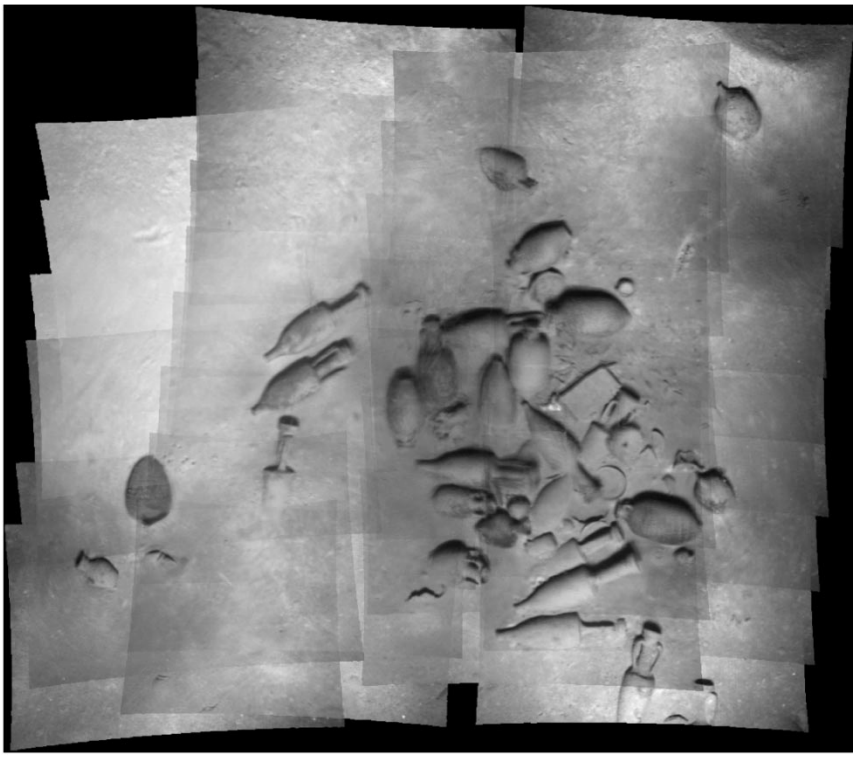


Fig. 14. Global mosaic of an example set of 29 images. The mosaic is rendered using the average of overlapping pixel intensities. Ghosting (in particular on the top right and bottom left), as well as the blurring of some of the objects, are evidence of misregistrations.

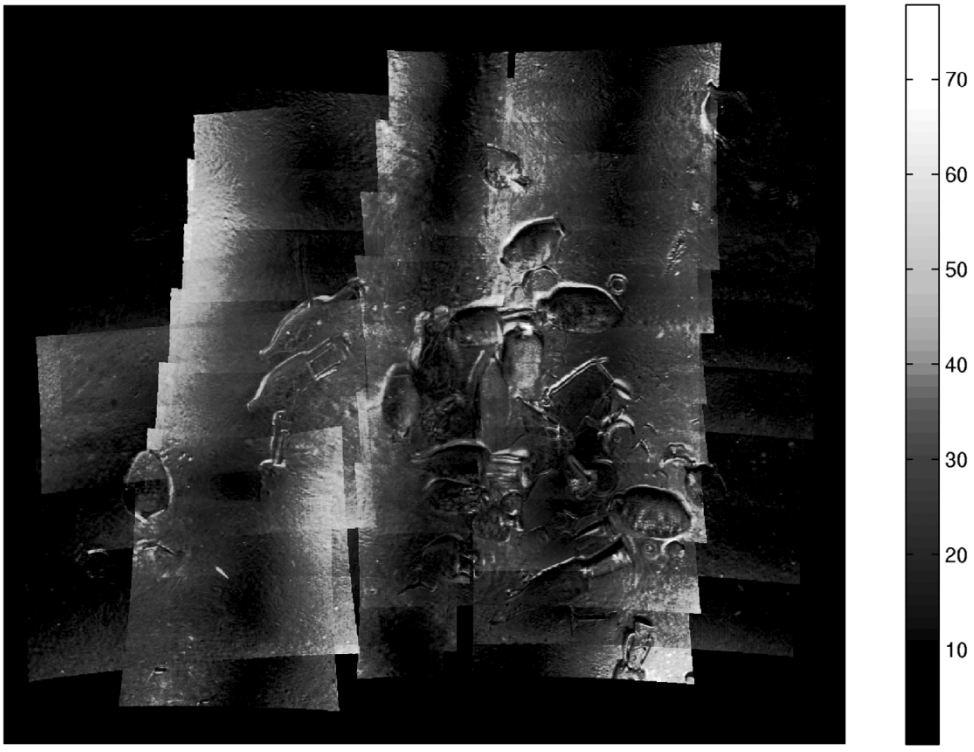


Fig. 15. Standard deviation of all source samples at each sample of the mosaic in Fig. 14. The standard deviation in areas with none or one image is set to zero by definition. High variability in pixel intensities is due to misregistrations (edge-like ghosting near objects) or changes in lighting (gradual changes that increase toward image edges).

Note that the cost function J_t, as proposed, underconstrains the problem. For example, given a solution set $P_1 \dots P_N$, another set of transforms that differs only by a translation and rotation of the whole mosaic would also minimize the cost function.

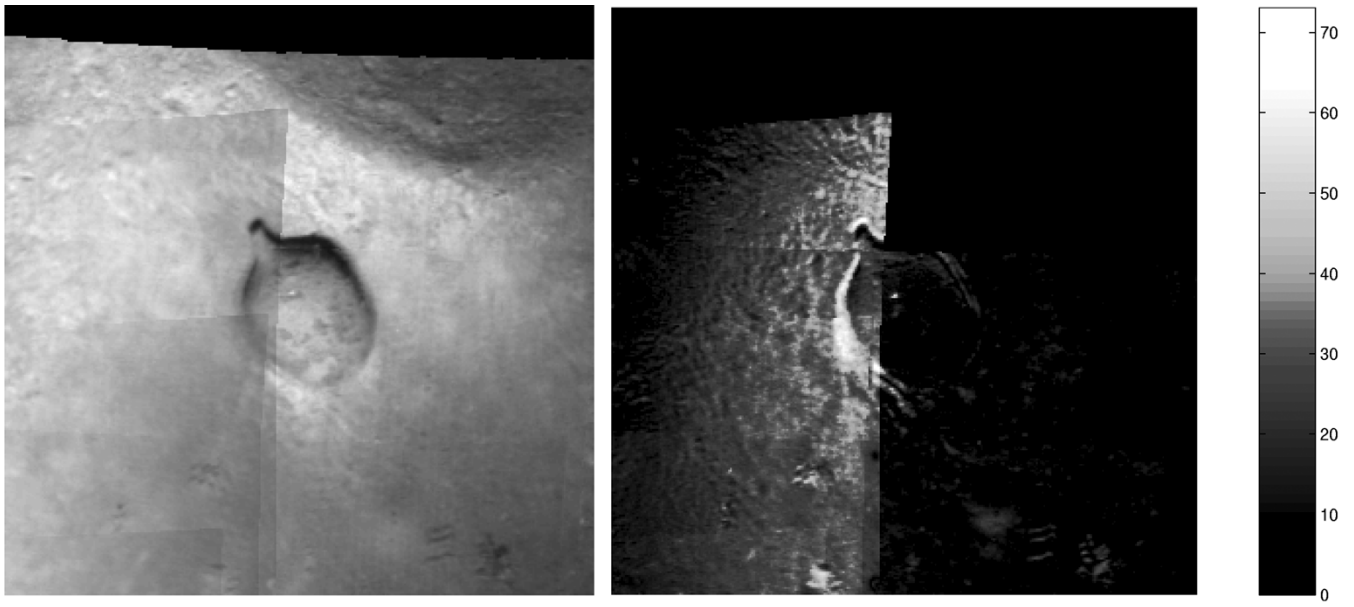


Fig. 16. Detail of top right corner of the mosaic in Fig. 14. Average (left) and standard deviation (right) of overlapping source samples. The misregistration is apparent in the ghosting around the amphora and the blurred terrain in the image of the average. In the standard-deviation image, repeated edge-like structures represent misregistration, while areas with high but approximately uniform variability suggest differences in lighting between the overlapping images.

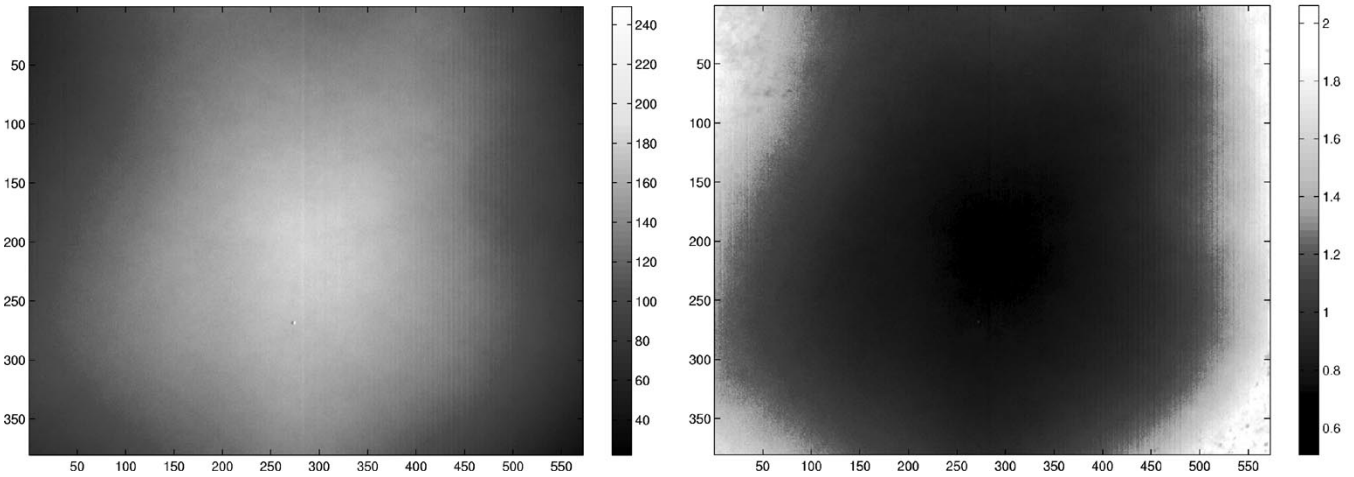


Fig. 17. Average of pixel values over all images of approximately flat sections of the Skerki D wreck (left) and the multiplicative radiometric correction factors (left).

To address this issue, we add a distortion penalty

$$J_d = \sum_{k=1}^{N} |(P_k(x_{\mathrm{tr}}) - P_k(x_{\mathrm{bl}})) - (x_{\mathrm{tr}} - x_{\mathrm{bl}})| + |(P_k(x_{\mathrm{tl}}) - P_k(x_{\mathrm{br}})) - (x_{\mathrm{tl}} - x_{\mathrm{br}})| \quad (22)$$

where $x_{\mathrm{tr}}, x_{\mathrm{tl}}, x_{\mathrm{br}}, x_{\mathrm{bl}}$ stand for top right, top left, bottom right, and bottom left corners of the image. J_d is then the sum (over all images) of the magnitude squared of the difference between the original diagonal vectors and the transformed diagonal vectors. This penalizes excessive transformation of the images, in particular avoiding the trivial solution (where all features are mapped onto the origin). In practice, this cost term guides the minimization to a solution where rotation and scale changes are "distributed" over all images. Notice, for example, that the images that form the mosaic in Fig. 12 were both rotated by roughly the same magnitude onto the mosaic.

An additional cost term is added to account for the overall translation of the mosaic

$$J_c = |P_1(x_0)| \quad (23)$$

which is the distance squared of the transformed origin of the first image. By minimizing this term, the origin of the transformed image will not translate.

The overall cost function has the form

$$J = J_t + w_d \cdot J_d + w_c \cdot J_c \quad (24)$$

where w_d, w_c are weighting factors for the different cost terms. We have observed that the solution is insensitive to w_c (which could actually be incorporated as a constraint in a Lagrange multiplier solution). There is also a broad range in which w_d has little effect on the final J_t; in practice, w_d is initially set in the range $[10^{-6}, 10^{-2}]$ and not changed afterward.

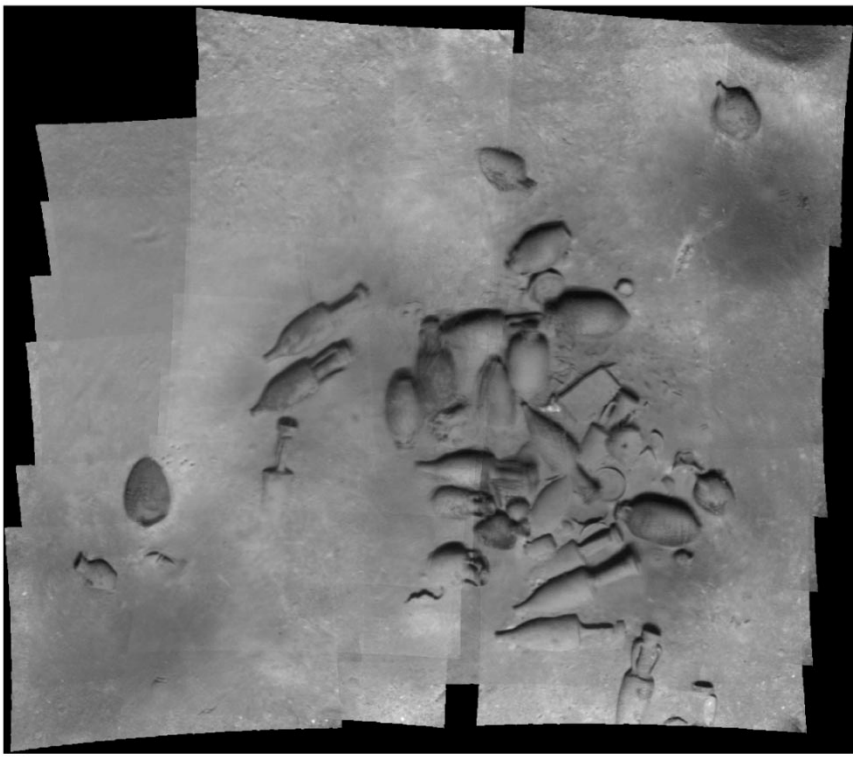


Fig. 18. Global mosaic of an example set of 29 images with radiometric correction. The mosaic is rendered using the average of overlapping pixel intensities.

Fig. 19. Standard deviation of all radiometrically corrected source samples at each mosaic sample (Fig. 18). High variability in pixel intensities is due to misregistrations (edge-like ghosting near objects). Notice that the variability due to lighting is significantly less than in Fig. 15.

By using parametrized transforms that are linear in the parameters (affine), J is quadratic and can easily be solved by linear least squares in one iteration.

C. *Refinement*

After solving for the homographies, new links (overlaps) can be hypothesized from the topology of the images. A simple

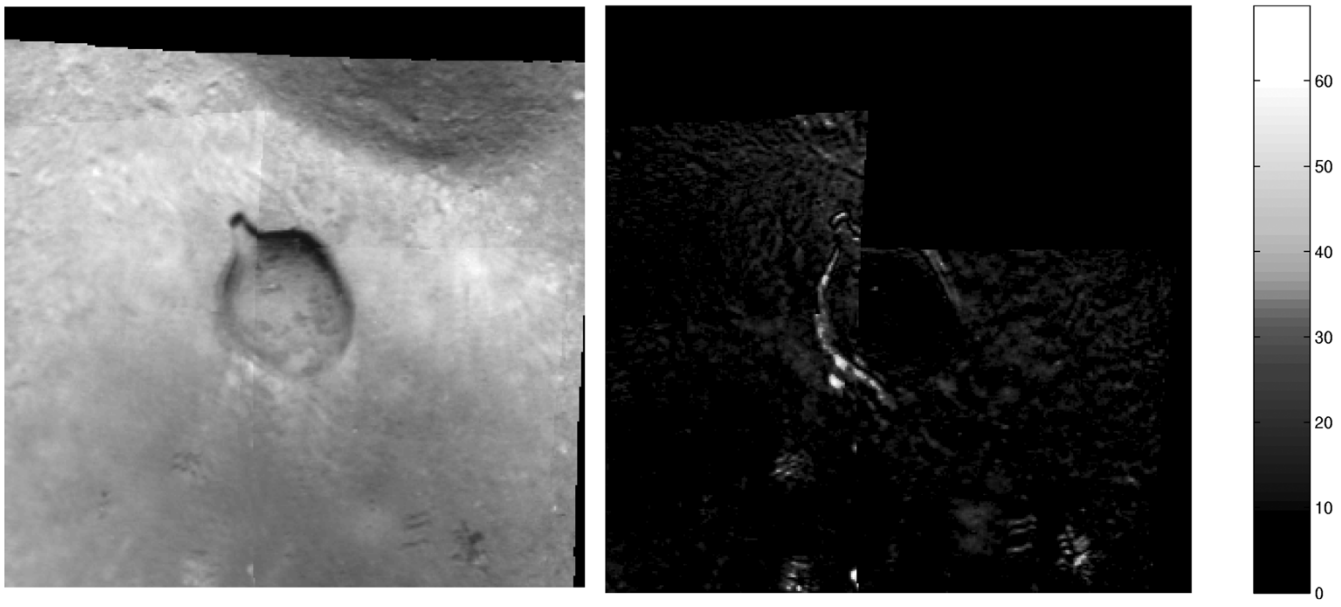


Fig. 20. Detail of top-right corner of the radiometrically corrected mosaic in Fig. 18. Average (left) and standard deviation (right) of overlapping source samples. Notice that most of the variability due to lighting differences between overlapping images has been removed (compare to Fig. 16).

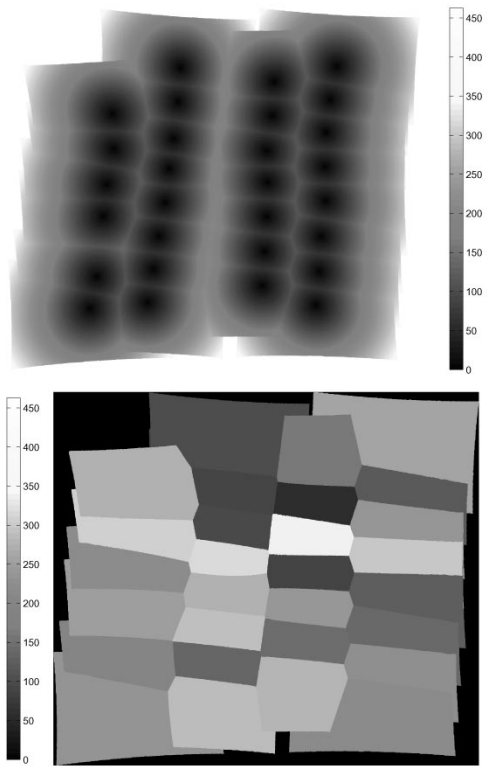


Fig. 21. (top) Minimum distance map (from center of each image) of the mosaic in Fig. 18. (bottom) Tessellation based on minimum distance map.

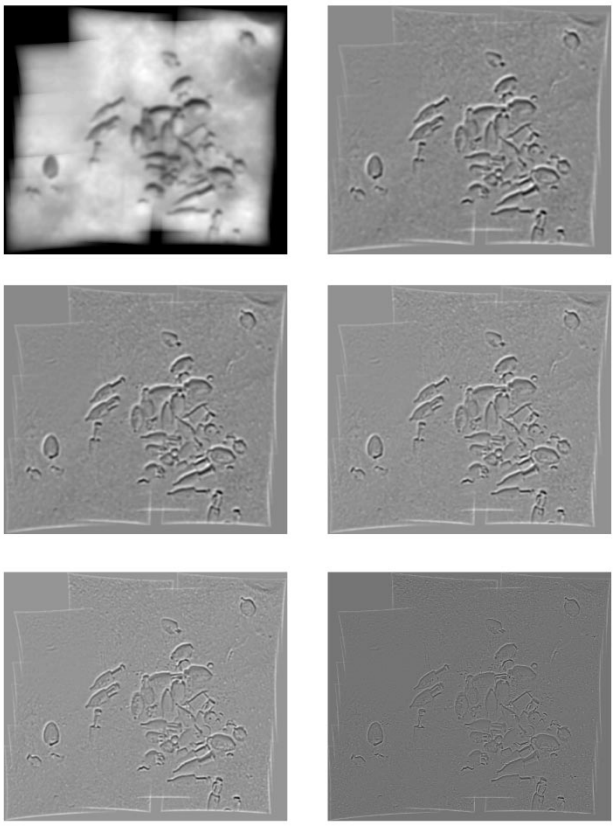


Fig. 22. Mosaic at each frequency band, constructed according to the tessellation of Fig. 21.

distance measure between centers is used for this. These possible links are verified by feature matching and any matching points are incorporated into the minimization. The process is repeated until no new links are proposed. Fig. 13 illustrates the topology-estimation process for the example set of 29 images. The purpose of this stage is to identify all significant overlaps, not to render the best possible mosaic. For this reason, only an affine transform (linear, six parameters) rather than a projective transform is solved for each image, making the matrix inverted at each refinement of the topology of size $6N \times 6N$, with N the number of images.

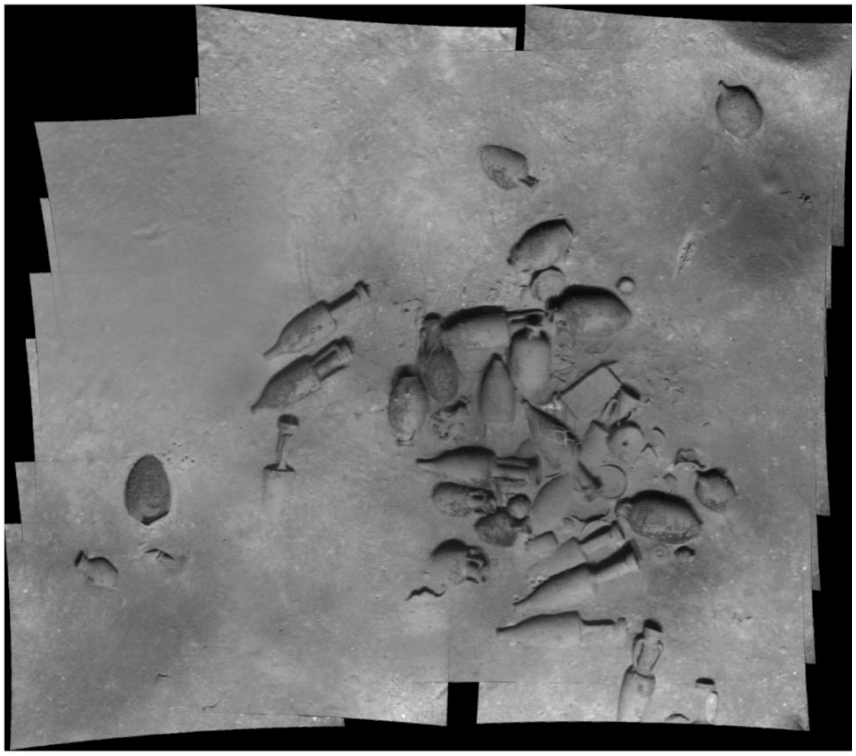


Fig. 23. Global mosaic of an example set of 29 images with radiometric correction. The mosaic is rendered by multifrequency blending.

IV. Global Registration

A. Assumptions and Approach

The global-registration problem attempts to find the transforms for each image that belongs to the mosaic by simultaneously considering all overlaps. The only difference between this and the topology-estimation stage is that we use a planar projective transform (nonlinear, eight parameters) to map images onto the mosaic. The function optimization largely dictates the complexity and accuracy of the solution. The formulation used here attempts to recover only the transforms and not the globally consistent mosaic coordinates of feature points. This differs from other feature-based approaches that also attempt to recover feature positions [13], [55]. Although these methods are potentially more accurate and also lend themselves to traditional bundle adjustment, they are computationally more demanding. For example, a global mosaic using planar projective transforms (eight parameters) requires eight N parameters, whereas a technique that also estimates the position of the features could have on the order of 50 features per image (100 parameters), requiring us to solve for on the order of $100N$ parameters. In addition, when solving only for the transforms, the structure of the covariance matrix allows for very simple and fast updates, since each parameter is associated with only one image. When estimating feature positions, the update of the covariance matrix is more complex since the parameters associated with features appear on multiple images.

B. Formulation

The global-registration stage consists of solving a planar—projective transform by using the known topology. This, again, is formulated as a least-squares problem but, since the transform is nonlinear, the solution is obtained iteratively using the Levenberg–Marquardt method [56]. The final estimate from the topology-estimation stage (affine transforms) is used as the starting point. The mosaic of the motivating set of images and the standard deviation of intensities is presented in Figs. 14 and 15. Note that as the mosaic is rendered as the average of intensities, ghosting indicates misregistration.

C. Radiometric Correction

The lighting pattern for this data set is far from ideal. Figs. 14 and 15 suggest that a considerable amount of variability is due to nonuniform lighting. The images used in this mosaic belong to the Skerki D wreck, a much larger data set (see Fig. 30) with several sections that are approximately flat and of the same material. Under these assumptions, a simple radiometric correction was calculated based on the inverse of the average of intensities for each pixel

$$R(x,y) = \overline{\bar{f}(x,y)} \cdot \min\left(\frac{1}{\bar{f}(x,y)}, \quad \frac{\text{Maxscale}}{\bar{f}(x,y) + 4 \cdot \sigma(x,y)}\right) \tag{25}$$

where $\bar{f}(x,y)$ and $\sigma(x,y)$ are, respectively, the average and standard deviation of intensity over all images f at pixel (x,y), $\overline{\bar{f}(x,y)}$ is the intensity average over all pixels and all images, and Maxscale is the maximum scale value. The corrected image is then $f_c(x,y) = f(x,y) \cdot R(x,y)$. The correction $R(x,y)$ can be interpreted as the gain that will bring the average at $f_c(x,y)$ to the overall average (making $\bar{f}_c(x,y)$ equal to $\overline{\bar{f}(x,y)}$). This

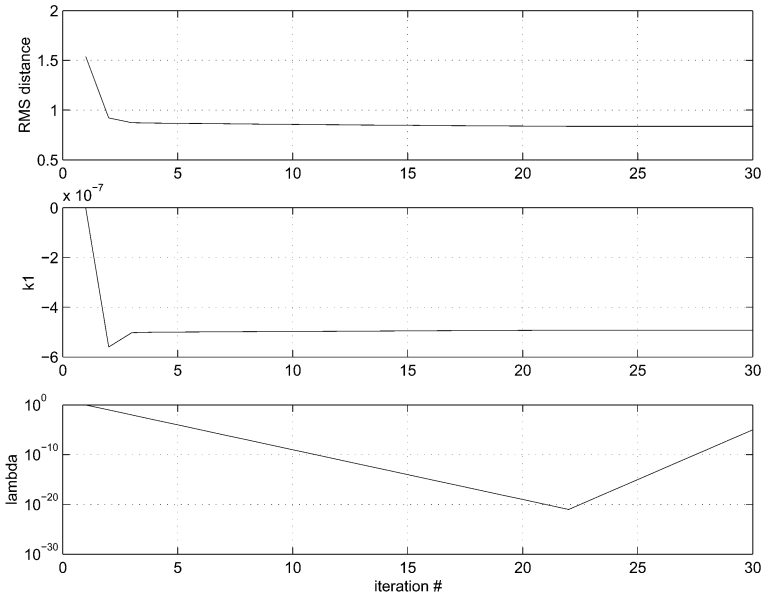


Fig. 24. Evolution of (top) RMS distance of corresponding points on the mosaic, (middle) the radial distortion coefficient, and (bottom) the weighting factor used in the Levenberg–Marquardt algorithm to switch from Newton–Gauss to gradient descent. Note that the radial distortion parameter reduces the RMS distance by more than 50%.

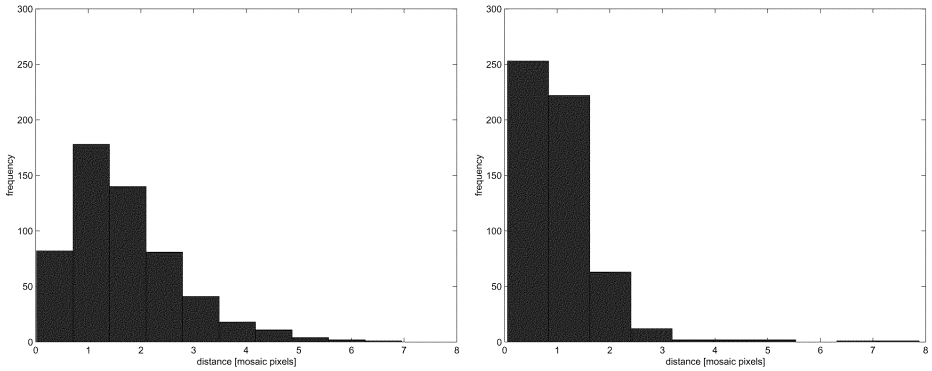


Fig. 25. Distribution of distances between corresponding points on the mosaic, (left) without radial distortion compensation and (right) with distortion compensation.

is limited, however, to a maximum gain such that an intensity 4σ above the mean at (x, y) will map on to the maximum of the intensity scale, Maxscale (255 for this data set). Fig. 17 show the average and radiometric correction for images of the Skerki D data set. The mosaic of Fig. 14, rendered with radiometrically corrected images, is presented in Fig. 18. Note that the correction was not used to calculate the image transforms. The standard deviation and details of the misregistered area for the radiometrically corrected mosaic are in Figs. 19 and 20.

D. Multiresolution Blending

Representing the final mosaic as the average of intensities gives a good sense of the quality of registration. Aesthetically, however, a mosaic based on averaging is unsatisfactory in the presence of ghosting or significant lighting changes. One option to assign intensities in the mosaic is to select the one corresponding to the pixel closest to the center of an image. This approach basically creates a voronoi tessellation on the mosaic. The sharp edge between cells can be smoothed by using multifrequency blending [57]. A mosaic rendering based on this idea is presented in [58] and [29]. To implement this, a distance map for each pixel to the center of the image is mapped onto the mosaic using the same homographies. When multiple distance maps overlap, the minimum distance is kept (Fig. 21). Concurrently, another mosaic keeps track of to which image a minimum distance belongs. This becomes the tessellation illustrated in Fig. 21, used to render the mosaic. Each cell is then mapped back onto the image coordinate frame. The multiresolution masks for the cell are constructed, mosaics for each frequency band are then built as in Fig. 22, and are finally added up to generate the blended mosaic (Fig. 23). Note that the radiometrically corrected images were used at this stage.

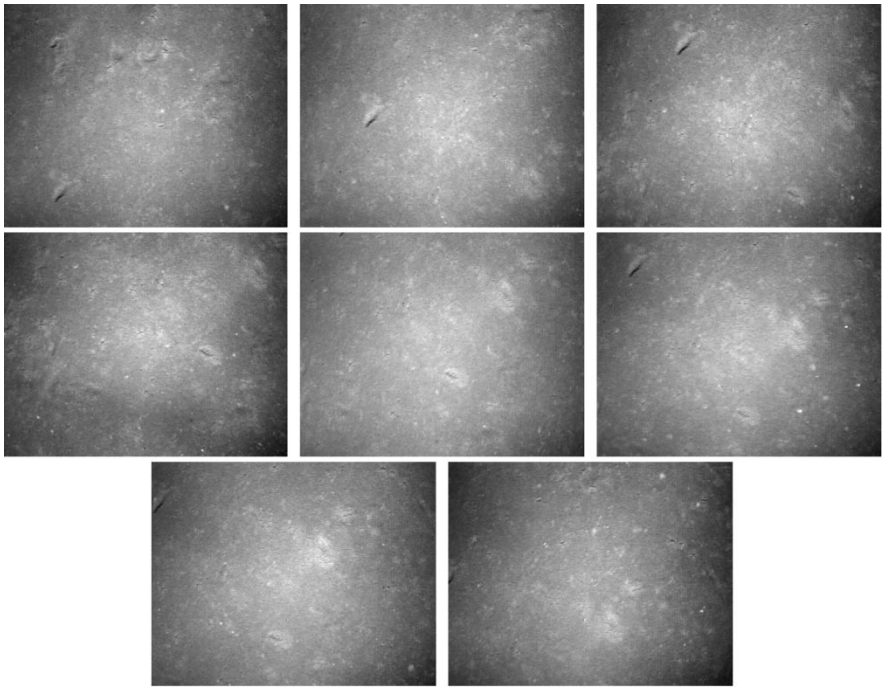


Fig. 26. Images used to estimate the radial-distortion coefficient.

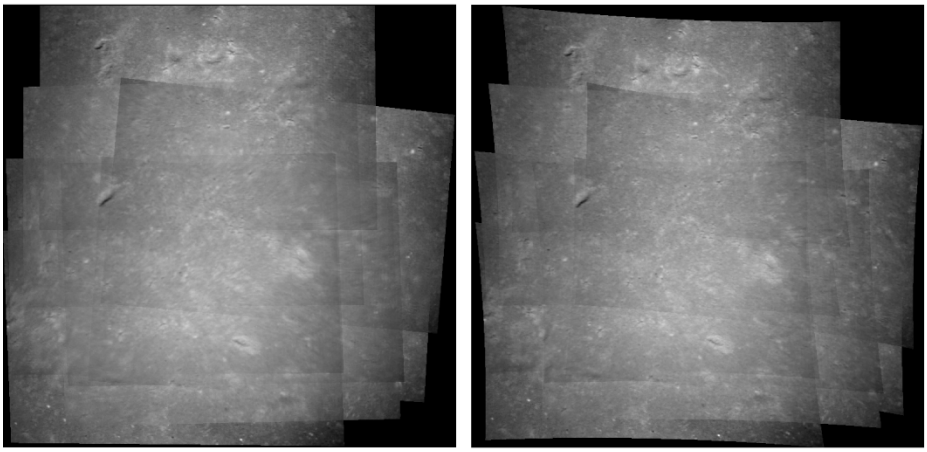


Fig. 27. Global mosaic before and after radial-distortion compensation.

V. Radial Distortion Compensation

A. Assumptions and Approach

Radial distortion can be significant for underwater imaging systems given the difference in refractive index between water and air. In areas of high overlap typical of video imaging and for mosaics consisting of a few tens of images covering small areas, the effects of radial distortion might not be readily apparent. However, low overlap and large mosaics require the correction of radial distortion for accurate rendering. For mosaicing purposes, it is sufficient to extract a simple cubic model [59] from a set of images of overlapping images from an approximately planar surface. Our approach is similar to [10], but we minimize the distance between mapped corresponding features rather than the intensity variance between transformed images due to the concerns about nonuniform lighting and the failure of the BCC.

B. Formulation

We can model radial distortion as

$$\mathbf{x}^d = d(\mathbf{x}^i; k_1) = \mathbf{x}^i + k_1 \left({\mathbf{x}^i}^T \mathbf{x}^i \right) \mathbf{x}^i$$
$$\mathbf{x}^m = P(\mathbf{x}^i) = P(d^{-1}(\mathbf{x}^d; k_1))$$

where $\mathbf{x}^i$ is the feature position in an ideal (corrected) frame, k_1 is the pincushion/barrel distortion coefficient, $\mathbf{x}^d$ is the distorted coordinates of the feature (coordinates on the acquired image), and $\mathbf{x}^m$ is the coordinate on the mosaic. This is a simplified

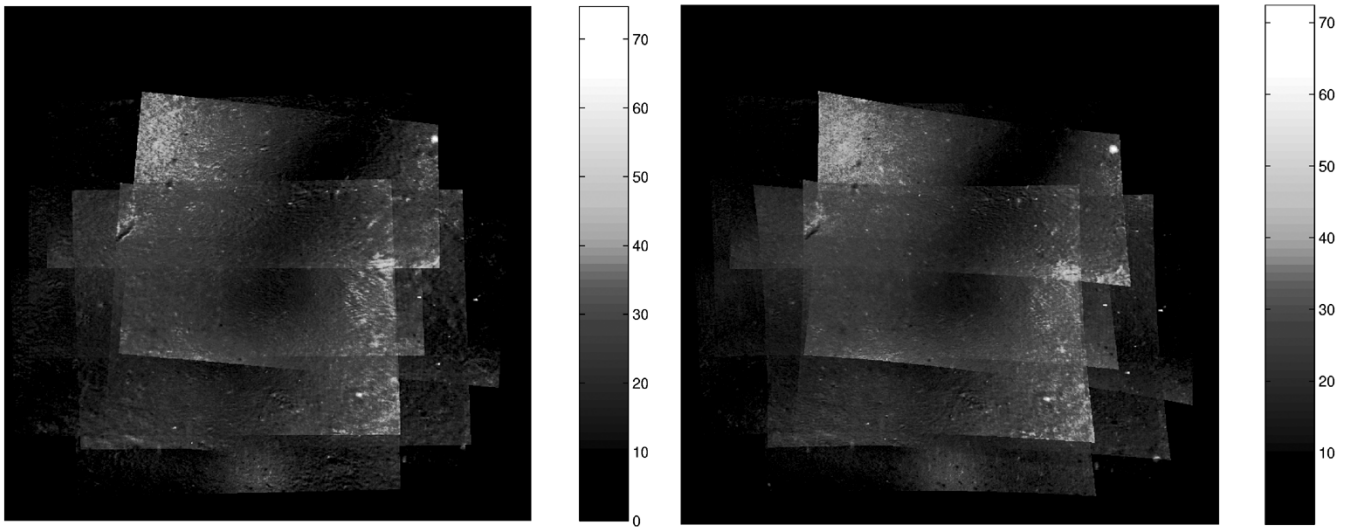


Fig. 28. Standard deviation of source samples at each mosaic sample before and after radial-distortion compensation. Note the reduction of the high-frequency content in the radially compensated mosaic. This indicates improved registration, in particular near image corners.

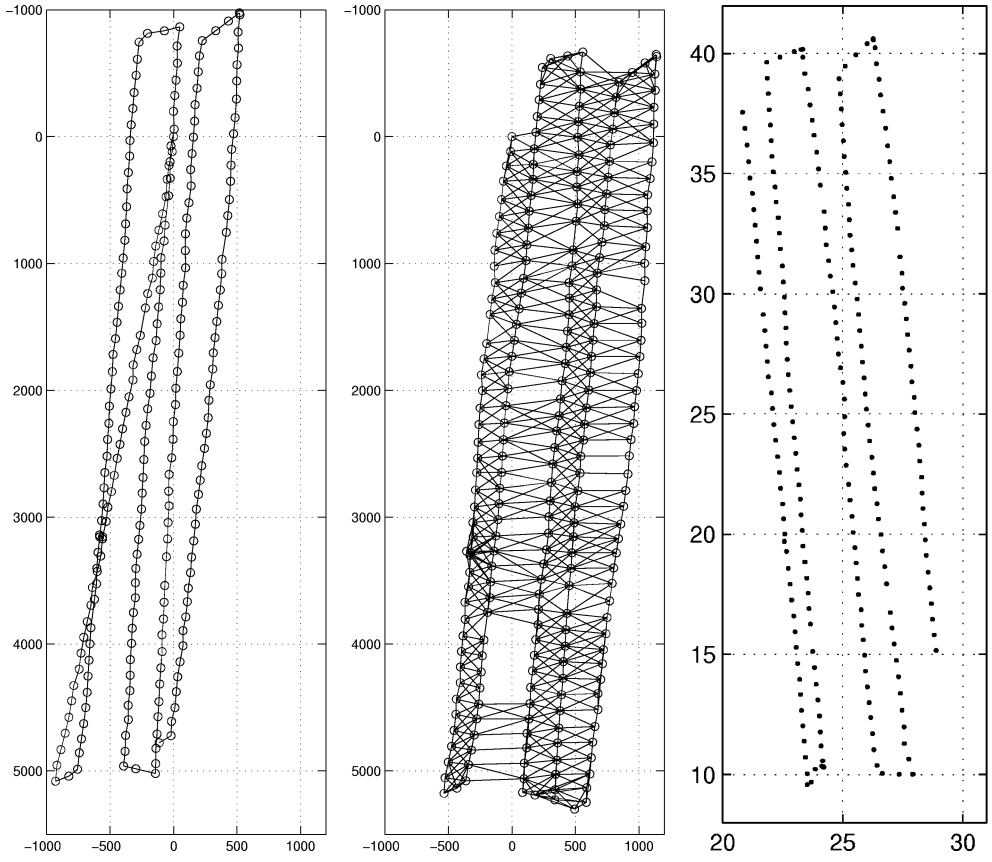


Fig. 29. (left) Initial and (center) final topology for the mosaic of Skerki D wreck site (axes in pixels). The origin for the topology is the center of the first image. Note that the relative position of the first trackline (depicted in gray) shifts dramatically from the initial to the final topology. The resulting mosaic is shown in Fig. 30. The vehicle nav-based position (right), in meters, at the instants the images were taken (with the exception of the last images where nav records are missing). The position was estimated using a 300-kHz-long baseline acoustic net and a Doppler velocity log, with centimeter-level accuracies [64]. The origin for position is the survey origin. The vehicle-navigation data is oriented with north in the $+\mathrm{Y}$ direction, whereas the mosaic orientation simply minimizes the overall rotation of the images. The first image of the mosaic corresponds to the westernmost image. The overall shape of the mosaic corresponds to the shape of the survey; notice the relative spacing of the tracklines and instances with larger spacing between consecutive images.

model, as we are assuming that the center of radial distortion is at the image center and we are ignoring higher order terms. Our results show that this simple model improves registration and can be calculated from survey imagery. A more complete distortion compensation is possible with proper calibration procedures [60], but these require access to the camera and a known underwater target (which is not always possible for older data sets). Distortion compensation based on point correspondences

Fig. 30. Mosaic of the complete Skerki D wreck (233 images), covering an area of approximately 350 m^2. The mosaic is rendered as the average of pixel intensities.

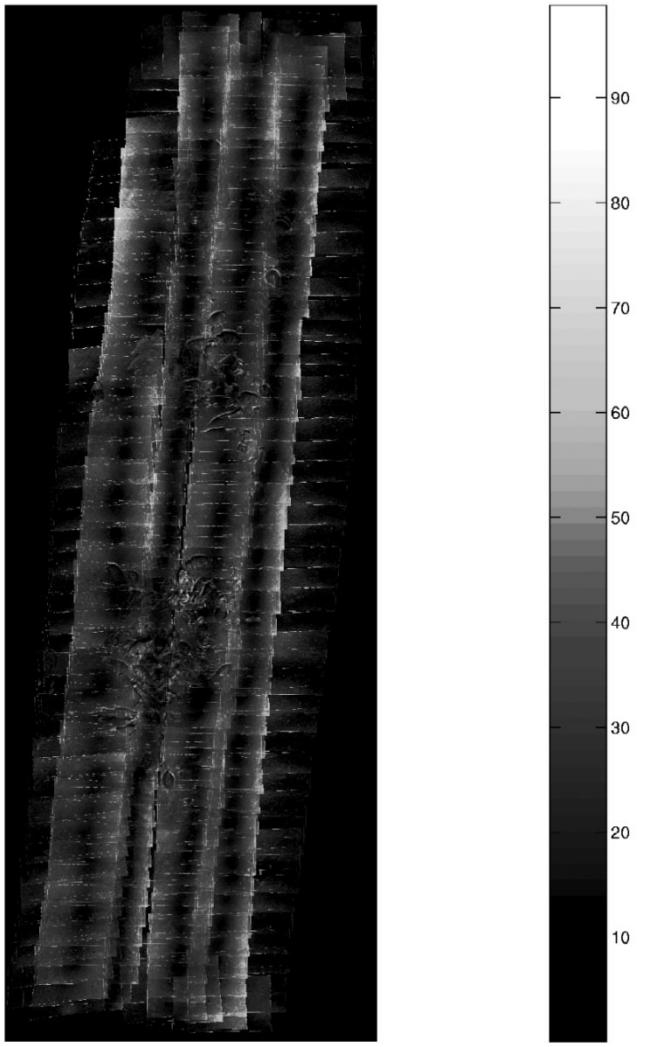


Fig. 31. Standard deviation of all source samples for each sample of the mosaic in Fig. 30.

using "curved" uncalibrated multiview geometry has been proposed in [61]–[63], but these methods either assume the presence of significant 3-D structure [61], [62] or solve for one radial-distortion parameter [63].

The cost function to minimize for global registration is now nonlinear as

$$J_t = \sum_{l,k} \sum_{j} \left| P_l\left(\mathbf{x}_{jl}^i\right) - P_k\left(\mathbf{x}_{jk}^i\right) \right| \tag{26}$$

and in terms of known quantities

$$J_t = \sum_{l,k} \sum_{j} \left| P_l\left(d^{-1}\left(\mathbf{x}_{jl}^d, k_1\right)\right) - P_k\left(d^{-1}\left(\mathbf{x}_{jk}^d, k_1\right)\right) \right|. \tag{27}$$

The solution minimizes J over all transform parameters and the radial-distortion coefficient

$$\min_{P_1 \ldots P_N, k_1} J \tag{28}$$

where J is given by J_t and the cost terms for geometric distortion and mosaic origin.

C. Implementation

The minimization problem is solved iteratively by using the Levenberg–Marquardt algorithm. Note that $d(x^i)$ does not have an analytic inverse, so it is estimated using the Gauss–Newton method, starting with the initial guess $x^i = x^d$. Fig. 26 shows a set of images used to estimate k_1. For this image set, the estimate of the pincushion coefficient is $k_1 = -4.93 \cdot 10^{-7}$. Estimates from other sets of images from the same deployment are in close accordance. The optimization improves registration by reducing the distance between corresponding points in the mosaic frame (Figs. 24 and 25). The mosaic with and without radial-distortion compensation is presented in Fig. 27, where the pixels in the mosaic are displayed as averages of overlapping intensities. The improvement by accounting for radial distortion is noticeable

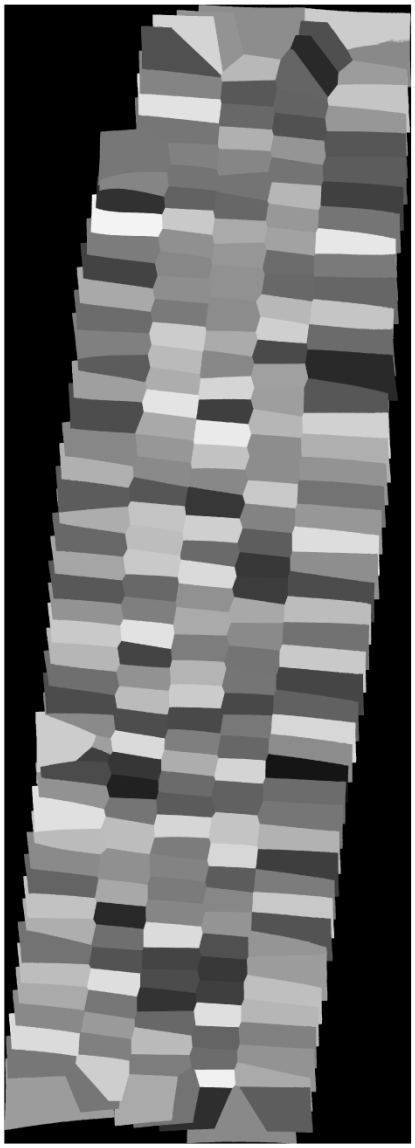


Fig. 32. Tessellation used in the multiresolution blended mosaic of the Skerki D wreck.

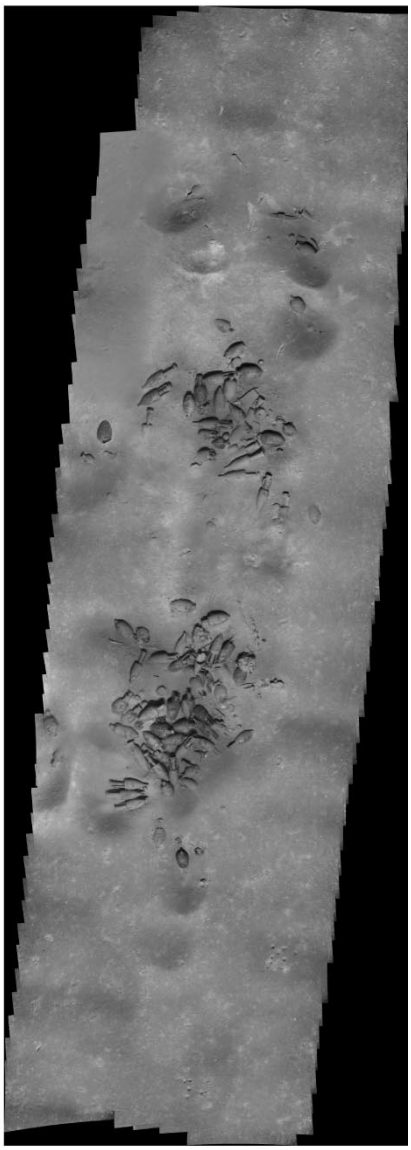


Fig. 33. Multiresolution blended mosaic of the complete Skerki D wreck (233 images) covering an area of approximately 350 m^2, based on the tessellation of Fig. 32. Images were radiometrically corrected as described in Section IV-3 and corrected for radial distortion using the approach described in Section V.

in Fig. 28 as a reduction of the high-frequency content in the standard deviation of intensities, related to the ghosting due to misalignment. Notice that the low-frequency component of the variance does not change, as this is due mainly to the differences in lighting conditions.

The pincushion parameter is assumed to be constant for a given deployment. Once k_1 is determined from a close to planar set of images, all images that are to be used on a mosaic are radially compensated before feature detection and matching. The mosaics in Figs. 12, 14, and 30 used images precompensated for radial distortion.

VI. Results

A. Underwater Surveys

Fig. 30 is the mosaic of the complete survey of the D wreck of Skerki bank [3], performed by the Jason remotely operated vehicles (ROVs). It consists of 233 images, covering an area of approximately 350 m^2. All images are compensated for radial distortion using the results of Section V. Fig. 31 illustrates the quality of the global registration. Fig. 29 shows the initial estimate of the position of images based on their temporal sequence and the final estimate based on their temporal sequence and the final estimate based on all verified overlaps. Fig. 33 is a rendering of the mosaic using multifrequency blending, as described in Section IV-4 and the radiometric correction presented in Section IV-3. Fig. 32 shows the tessallation used to blend the mosaic.

VII. Conclusion and Future Work

We have presented an approach for large-area mosaicing underwater. It is based on the assumption of a planar scene and

addresses the demanding constraints of underwater imaging and underwater surveys from robotic vehicles, such as

- low overlap;
- unstructured motion;
- poor, variable lighting.

Effective feature detection and matching stages under these constraints serve as the basis of a global-mosaicing technique. Our results include the largest known published underwater mosaic generated entirely without manual intervention. The overall shape of the mosaic is in accordance with the navigation data for that survey.

We have also extended the global-mosaicing technique to allow for radial-distortion estimation. We note that this technique as formulated will only find a local optimum and, furthermore, that our model is the simplest one possible to address lens distortion (higher order terms and a center of radial distortion could be incorporated). As such, it should only be considered as an improvement for mosaicing, not as an attempt at camera calibration.

Future work will extend these results in several directions, as follows.

- Discard the planarity assumption by modeling structure and motion.
- Extend feature matching to scenes with 3-D structure and wide baseline
- Incorporate full-camera calibration and distortion compensation.
- Incorporate vehicle navigation (if available).

By accepting and accounting for the presence of 3-D structures, we open the possibility of generating reconstructions of more complex scenes. We would be using a model that is rich enough to allow meaningful interpretation of errors and uncertainties. On the other hand, establishing correspondence and, more specifically, feature description and matching, is significantly harder in the presence of 3-D structures and relatively wide baselines. However, by incorporating accurate camera calibration and distortion compensation, interimage motion can be described by a rotation and baseline direction. It is then possible to use navigation data (position and orientation) to propose overlapping imagery and guide matching.

ACKNOWLEDGMENT

The authors would like to thank the Deep Submergence Laboratory Operations Group; without them, this paper would not be possible. Also, special thanks to S. Teller of the Massachusetts Institute of Technology Laboratory for Computer Science (MIT LCS) for his thoughtful suggestions, which helped to give final form to this work.

REFERENCES

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

[2] D. Yoerger, A. Bradley, M.-H. Cormier, W. Ryan, and B. Walden, "Fine-scale seafloor survey in rugged deep-ocean terrain with an autonomous robot," *Proc. IEEE Int. Conf. Robotics and Automation*, pp. 1767–1774, 2000.

[3] R. Ballard, A. McCann, D. Yoerger, L. Whitcomb, D. Mindell, J. Oleson, H. Singh, B. Foley, and J. Adams, "The discovery of ancient history in the deep sea using advanced deep submergence technology,," *Deep-Sea Research I*, vol. 47, no. 9, pp. 1591–1620, 2000.

[4] R. Ballard, L. Stager, D. Master, D. Yoerger, D. Mindell, L. Whitcomb, H. Singh, and D. Piechota, "Iron age shipwrecks in deep water off Ashkelon, Israel," *Amer. J. Archaeology*, no. 2, pp. 151–168, Oct. 2001.

[5] J. Howland, "Digital Data Logging and Processing, Derbyshire Survey, 1997," Woods Hole Oceanographic Inst., Woods Hole, MA, Tech. Rep., Dec. 1999.

[6] National Transportation Safety Board, EgyptAir Flight 990, Boeing 767-366ER, SU-GAP, 60 Miles South of Nantucket, Massachuesetts, October 31, 1999, Washington, DC, 2002. Aircraft Accident Brief NSTB/AAB-02/01.

[7] C. Smith, "Whale falls: Chemosynthesis at the deep-sea floor," *Oceanus*, vol. 35, no. 3, pp. 74–78, 1992.

[8] K. Foote, "Censusing marine living resources in the gulf of maine: A proposal," *Proc. MTS/IEEE OCEANS'01*, pp. 1611–1614, 2001.

[9] J. Reynolds, R. Highsmith, B. Konar, C. Wheat, and D. Doudna, "Fisheries and fisheries habitat investigations using undersea technology," *Proc. MTS/IEEE OCEANS'01*, pp. 812–819, 2001.

[10] H. 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, 1999.

[11] H. Sawhney, S. Hsu, and R. Kumar, "Robust video mosaicing through topology inference and local to global alignment," in *Proc. Eur. Conf. Computer Vision*, Freiburg, Germany, 1998, pp. 103–119.

[12] H.-Y. Shum and R. Szeliski, "Panoramic Image Mosaics," Microsoft Research, Redmond, WA, Tech. Rep. MSR-TR-97-23, 1997.

[13] P. McLauchlan and A. Jaenicke, "Image mosaicing using sequential bundle adjustment," in *Proc. British Machine Vision Conf.*, Bristol, U.K., 2000, pp. 616–625.

[14] S. Coorg and S. Teller. (2000) Spherical mosaics with quaternions and dense correlation. *Int. J. Comput. Vision* [Online], vol (3), pp. 259–273. Available: citeseer.nj.nec.com/coorg00spherical.html

[15] O. Faugeras and O.-T. Luong, *The Geometry of Multiple Images: The Laws That Govern the Formation of Multiple Images of a Scene and Some of Their Applications*. Cambridge, MA: MIT Press, 2001.

[16] S. Maybank and O. Faugeras, "A theory of self-calibration of a moving camera," *Int. J.Comput. Vision*, vol. 8, no. 2, pp. 123–152, 1992.

[17] M. Pollefeys, R. Koch, M. Vergauwen, and L. Van Gool, "Hand-held acquisition of 3d models with a video camera," in *Proc. 2nd Int. Conf. 3-D Digital Imaging and Modeling*. Los Alamitos, CA, 1999, pp. 14–23.

[18] B. Triggs, "Autocalibration and the absolute quadric," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, 1997, pp. 609–614.

[19] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics Gems (IV)*, P. Heckbert, Ed. Boston, MA: Academic, 1994, pp. 474–485.

[20] S. Fleischer, R. Marks, S. Rock, and M. Lee, "Improved real-time video mosaicking of the ocean floor," in *Proc. MTS/IEEE Conf. OCEANS'95*, vol. 3, San Diego, CA, Oct. 1995, pp. 1935–1944.

[21] S. Fleischer, H. Wang, S. Rock, and M. Lee, "Video mosaicking along arbitrary vehicle paths," in *Proc. 1996 Symp. Autonomous Underwater Vehicle Technology*, Monterey, CA, June 1996, pp. 293–299.

[22] S. Negahdaripour, X. Xu, and L. Jin, "Direct estimation of motion from sea floor images for automatic station-keeping of submersible platforms," *IEEE J. Oceanic Eng.*, vol. 24, pp. 370–382, July 1999.

[23] N. Gracias and J. Santos-Victor, "Underwater mosaicing and trajectory reconstruction using global alignment," in *Proc. IEEE OCEANS'01*, Honolulu, HI, 2001, pp. 2557–2563.

[24] B. Horn, *Robot Vision*. Cambridge, MA: McGraw-Hill, 1986.

[25] R. Szeliski, "Image Mosaicing for Tele-Reality Applications," Cambridge Research Laboratory, Cambridge, MA, Tech. Rep. CRL 94/2, May 1994.

[26] C. Slama, Ed., *Manual of Photogrammetry*, 4th ed. Bethesda, MD: American Society of Photogrammetry, 1980.

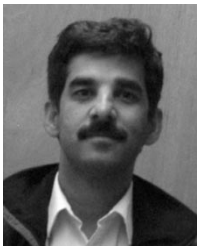
[27] H.-Y. Shum and R. Szeliski, "Construction and refinement of panoramic mosaics with global and local alignment," in *Proc. IEEE 6th Int. Conf. Computer Vision*, Bombay, India, Jan. 1998.

[28] A. Can, C. Stewart, B. Roysam, and H. Tanenbaum, "A feature-based technique for joint, linear estimation of high-order image-to-mosaic transformations: Application to mosaicing the curved human retina," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 2, Hilton Head, SC, 2000, pp. 585–591.

[29] S. Hsu, H. Sawhney, and R. Kumar, "Automated mosaics via topology inference," *IEEE Comput. Graph. Appl.*, vol. 22, no. 2, pp. 44–54, 2002.

[30] J. Bergen, P. Anandan, K. Hanna, and R. Hingorani, "Hierarchical model-based motion estimation," in *Proc. Eur. Conf. Computer Vision*, Santa Margarita Ligure, Italy, May 1992, pp. 237–252.
[31] R. Kumar, P. Anandan, M. Irani, J. Bergen, and K. Hanna, "Representation of scenes from collections of images," in *IEEE Workshop Representation of Visual Scenes*, Cambridge, MA, 1995, pp. 10–17.
[32] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. Alvey Conf.*, Manchester, U.K., Aug. 1988, pp. 189–192.
[33] M. A. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
[34] M. Irani, P. Anandan, J. Bergen, R. Kumar, and S. Hsu, "Mosaic representations of video sequences and their applications," *Signal Process.*, vol. 8, no. 4, 1995.
[35] M. Irani and P. Anandan, "Robust multi-sensor image alignment," in *Proc. 6th Int. Conf. Computer Vision*, Bombay, India, Jan. 1996, pp. 959–966.
[36] B. Reddy and B. Chatterji, "An FFT-based technique for translation, rotation, and scale-invariant image registration," *IEEE Trans. Image Process.*, vol. 5, pp. 1266–1271, Aug. 1996.
[37] S. Kruger and A. Calway, "Image registration using multiresolution frequency domain correlation," in *Proc. British Machine Vision Conf.*. Southampton, U.K., Sept. 1998, pp. 316–325.
[38] F. Schaffalitzky and A. Zisserman. Viewpoint invariant texture matching and wide baseline stereo. presented at *Proc. 8th Int. Conf. Computer Vision*. [Online]. Available: citeseer.nj.nec.com/schaffalitzky01viewpoint.html
[39] A. Baumberg, "Reliable feature matching across widely separated views," *Proc. Computer Vision and Pattern Recognition 2000*, pp. 774–781, 2000.
[40] T. Lindeberg. (1994) Scale-space theory: A basic tool for analysing structures at different scales. *J. Applied Stat.* [Online], vol (2), pp. 224–270. Available: citeseer.nj.nec.com/article/lindeberg94scalespace.html
[41] C. Schmid and R. Mohr. (1997, May) Local grayvalue invariants for image retrieval. *IEEE Trans. Pattern Anal. Machine Intell.* [Online], pp. 530–535. Available: citeseer.nj.nec.com/schmid97local.html
[42] C. Schmid, R. Mohr, and C. Bauckhage. (2000) Evaluation of interest point detectors. *Int. J. Comput. Vision* [Online], vol (2), pp. 151–172. Available: citeseer.nj.nec.com/schmid00evaluation.html
[43] D. Lowe. (1999) Object recognition from local scale-invariant features. *Int. Conf. Computer Vision (2)* [Online], pp. 1150–1157. Available: citeseer.nj.nec.com/lowe99object.html
[44] T. Lindeberg, "Feature detection with automatic scale selection," *Int. J. Comput. Vision*, vol. 30, no. 2, pp. 79–116, 1998.
[45] C. Schmid and R. Mohr. (1995, Aug.) Matching by Local Invariants. [Online]. Available: http://www.inrialpes.fr/movi/publi/Publications/1995/SM95
[46] C.-H. Teh and R. Chin, "On image analysis by the methods of moments," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 10, pp. 496–513, Apr. 1998.
[47] M. Hu, "Pattern recognition by moment invariants," *Inst. Radio Eng. Trans. Inform. Theory*, vol. 8, no. 2, pp. 179–187, 1962.
[48] A. Khotanzad and Y. Hong, "Invariant image recognition by Zernike moments," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 489–497, May 1990.
[49] W. Kim and Y. Kim, "A region-based shape descriptor using Zernike moments," *Signal Process.*, vol. 16, no. 1/2, pp. 95–102, 2000.
[50] ——, "Robust rotation angle estimator," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 21, pp. 768–773, Aug. 1999.
[51] F. Badra, A. Qumsieh, and G. Dudek, "Rotation and zooming in image mosaicing," in *Proc. 4th IEEE Workshop Applications of Computer Vision*, Princeton, NJ, 1998, pp. 50–55.
[52] P. Rousseeuw and A. Leroy, *Robust Regression and Outlier Detection*. New York: Wiley, 1987.
[53] Z. Zhang, "Determining the epipolar geometry and its uncertainty: A review," *Int. J. Comput. Vision*, vol. 27, no. 2, pp. 161–195, 1998.
[54] R. Szeliski, "Video mosaics for virtual environments," *IEEE Comput. Graph. Appl.*, vol. 16, pp. 22–30, Mar. 1996.
[55] D. Capel and A. Zisserman, "Automated mosaicing with super-resolution zoom," in *Proc. Int. Conf. Computer Vision and Pattern Recognition 1998*, Santa Barbara, CA, 1998, pp. 885–891.
[56] D. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena , 1995.
[57] P. Burt and E. Adelson, "A multiresolution spline with application to image mosaics," *ACM Trans. Graph.*, vol. 2, no. 4, pp. 217–236, 1983.
[58] H. Sawhney, R. Kumar, G. Gendel, J. Bergen, D. Dixon, and V. Paragano, "Videobrush: Experiences with consumer video mosaicing," in *Proc. 4th IEEE Workshop on Applications of Computer Vision*, Princeton, NJ, 1998.
[59] K. Atkinson, *Close Range Photogrammetry and Machine Vision*: Whittles, 1996.
[60] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, pp. 1330–1334, Nov. 2000.
[61] ——, On the epipolar geometry between two images with lens distortion. presented at *Proc. Int. Conf. Pattern Recognition (ICPR)*. [Online]. Available: citeseer.nj.nec.com/zhang96epipolar.html
[62] G. Stein, "Lens distortion calibration using point correspondences," in *Proc. IEEE Comput. Soc. Conf. Computer Vision and Pattern Recognition*, San Francisco, CA, June 1997, pp. 143–148.
[63] A. W. Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. presented at *Proc. Int. Conf. Computer Vision Pattern Recognition 2001*. [Online]. Available: citeseer.nj.nec.com/fitzgibbon01simultaneous.html
[64] L. Whitcomb, D. Yoerger, and H. Singh, "Advances in doppler-based navigation of underwater robotic vehicles," in *Proc. 1999 Int. Conf. Robotics and Automation*, vol. 1, Detroit, MI, 1999, pp. 399–406.

Oscar Pizarro (S'92) received the Engineer's 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, MA/Woods Hole Oceanographic Institute, Woods Hole, MA, joint program, where he is involved with underwater imaging and robotic underwater vehicles.



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

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