UWIT: Underwater Image Toolbox for Optical Image Processing and Mosaicking in MATLAB

Ryan Eustice, Oscar Pizarro, Hanumant Singh, Jonathan Howland
Department of Applied Ocean Physics and Engineering
Woods Hole Oceanographic Institution
MS 7, Woods Hole, MA 02543, USA
e-mail: ryan@whoi.edu

***Abstract*—This paper shows results from our development of an extended MATLAB image processing toolbox, which implements some useful optical image processing and mosaicking algorithms found in the literature. We surveyed and selected algorithms from the field which showed promise in application to the underwater environment. We then extended these algorithms to explicitly deal with the unique constraints of underwater imagery in the building of our toolbox. As such, the algorithms implemented include:**

- **contrast limited adaptive histogram specification (CLAHS) to deal with the inherent nonuniform lighting in underwater imagery.**
- **Fourier based methods for scale, rotation, and translation recovery which provide robustness against dissimilar image regions.**
- **local normalized correlation for image registration to handle the low-feature, unstructured environment of the seafloor.**
- **multiresolution pyramidal blending of images to form a composite seamless mosaic without blurring or loss of detail near image borders.**

In this paper we highlight the mathematical formulation behind each of these algorithms using consistent notation and a unified framework. We depict some of our MATLAB toolbox results with an assortment of underwater imagery.

***Index Terms*—Image processing, mosaicking, MATLAB.**

I. INTRODUCTION

ONE aspect of our research here at the Deep Submergence Laboratory of the Woods Hole Oceanographic Institution is optical image processing and mosaicking. Production of the high resolution, large area imagery of the seafloor required by our end user community, (geologists, biologists, archaeologists, ...) would be impossible were it not for the technique of photomosaicking whereby a large area image is assimilated from tens to hundreds of smaller-footprint, overlapping images. The propagation of visible light underwater suffers from rapid attenuation and scattering, which in combination with a limited camera-to-light separation available on most vehicle platforms, places strong constraints on the ability to image large areas of the seafloor optically. Special consideration must be given to the unique constraints associated with the underwater environment.

Our research in underwater image processing and photomosaicking deals explicitly with these constraints. In this paper we cover some successful techniques in the literature for image processing and mosaicking which have been extended to deal with unique peculiarities of the underwater environment. The algorithms and order of sections presented in this paper are:

- contrast limited adaptive histogram specification (CLAHS) to deal with the inherent nonuniform lighting in underwater imagery.
- Fourier based methods for scale, rotation, and translation recovery which provide robustness against dissimilar image regions.
- local normalized correlation for image registration to handle the low-feature, unstructured environment of the seafloor.
- multiresolution pyramidal blending of images to form a composite seamless mosaic without blurring or loss of detail near image borders.

As MATLAB has evolved into a very common and standard tool for rapidly prototyping and testing algorithms in the academic and scientific communities, we present results of our recently developed MATLAB toolbox. This toolbox provides extended image processing and mosaicking capabilities to the standard MATLAB imaging toolbox. This toolbox is freely available to the user community and can be obtained by contacting the authors.

II. HISTOGRAM SPECIFICATION

Deep sea imaging requires that vehicles carry their own light sources with them as ambient light is nonexistent. Different power requirements for different vehicle architectures (*i.e.* ROVs *vs.* AUVs) results in vastly different lighting patterns and large differences in the amount of energy expended in lighting the scenery. This, in conjunction with the rapid attenuation of light underwater, often results in low-contrast, non-uniformly lit imagery. The lighting artifacts present in the imagery pose an additional challenge to the image registration process as many conventional algorithms will "lock on" to these artifact intensity gradients created by the vehicle lighting.

To compensate for the low-contrast, nonuniform illumination present in the raw imagery, we employ the technique of contrast limited adaptive histogram specification (CLAHS) as a preprocessing step. The basis of the technique is described in [1] where it was first applied to low-contrast medical imagery. The idea behind the technique is to first subdivide the original image into equal area contextual regions. Each region is then histogram equalized whereby a monotonically nondecreasing gray level transform, $g = T[f]$, is determined for each region which maps the local gray level histogram to an ideal gray level distribution [2]. To limit the amount of contrast enhancement, specifically in homogeneous regions, the concept of a clip limit is used. The clip limit is defined as a multiple of the average histogram contents. Defining a clip limit results in limiting the contents of any individual histogram bin to contain at most the clip limit multiple of the average histogram contents, excess

pixels are then uniformly distributed to the remaining bins. The overall effect is to limit the slope of the cumulative histogram which is used in the calculation of the gray level transform.

We've experimented with different ideal gray level distributions such as uniform, exponential, and Rayleigh distributions. Our results seem to suggest that the Rayleigh distribution is most suited to underwater imagery. Figures 1(a) and 1(b) depict typical low-contrast, nonuniformly illuminated AUV imagery and the processed CLAHS results, respectively. Global gray level histograms of the pre and post processed images are also shown highlighting the increased effective use of the image's dynamic range.

III. Fourier Based Image Registration

One of the fundamental tasks in photomosaicking is image registration. Many methods exist with some of the most prevalent being: correlation based methods which use pixel values directly, fast Fourier methods which use frequency domain information, and feature based methods which use low-level features such as edges and corners.

Frequency domain methods approach the problem of image registration from a signal processing perspective. The major advantages of their approach are: (1) processing speed which is gained through exploitation of the Fast-Fourier-Transform's computational efficiency, and (2) robustness to narrowband frequency dependent noise such as low frequency shifts in mean image intensity [3]. By transforming the original image coordinates to a coordinate system where rotations and scaling appear as shifts, we can make use of the Fourier phase shifting property to recover image translations, scaling, and rotation [4]. Our particular implementation for transforming coordinate systems is based upon [5], but has been modified to use normalized correlation for determination of image translations [6].

Translation Only

The underlying property that's being exploited is the phase shifting property of Fourier transforms. Consider two images f_1 and f_2, which differ only by a translational offset

$$f_2(x, y) = f_1(x - x_0, y - y_0) \tag{1}$$

Via the shifting property, their Fourier transforms are related by

$$F_2(\omega_1, \omega_2) = F_1(\omega_1, \omega_2) e^{-j(\omega_1 x_0 + \omega_2 y_0)} \tag{2}$$

The translational offset can be recovered by locating the impulse associated with the inverse transform of the cross-power spectrum of the two images

$$\frac{F_2(\omega_1, \omega_2) \cdot F_1(\omega_1, \omega_2)^*}{\|F_2(\omega_1, \omega_2) \cdot F_1(\omega_1, \omega_2)^*\|} = e^{-j(\omega_1 x_0 + \omega_2 y_0)} \tag{3}$$

$$\stackrel{\mathfrak{F}^{-1}}{=} \delta(x - x_0, y - y_0) \tag{4}$$

Rotation Only

This same property can be exploited for images which are rotated and scaled by representing them in a coordinate system where scale and rotations appear as shifts. For example, when f_2 is a rotated version of f_1

$$f_2(x, y) = f_1(x \cos\theta_0 + y \sin\theta_0, -x \sin\theta_0 + y \cos\theta_0) \tag{5}$$

their Fourier transforms are related by

$$F_2(\omega_1, \omega_2) = F_1(\omega_1 \cos\theta_0 + \omega_2 \sin\theta_0, -\omega_1 \sin\theta_0 + \omega_2 \cos\theta_0) \tag{6}$$

Using only the magnitudes of the Fourier transforms and converting to polar coordinates, we see that the rotation can be represented as a shift

$$M_2(\rho, \theta) = M_1(\rho, \theta - \theta_0) \tag{7}$$

Scaling Only

Similarly, when two images are related by a scale factor, a, then their Fourier transforms are related by

$$F_2(\omega_1, \omega_2) = \frac{1}{a^2} F_1(\omega_1/a, \omega_2/a) \tag{8}$$

Taking the logarithm of the frequency axes results in the scale appearing as a shift (ignoring the $1/a^2$ scale factor)

$$F_2(\log\omega_1, \log\omega_2) = F_1(\log\omega_1 - \log a, \log\omega_2 - \log a) \tag{9}$$

Translation, Rotation, and Scaling

When translation, rotation, and scaling are all present between the two images, we see that representing the magnitudes in a log-polar coordinate system results in

$$M_2(\rho, \theta) = M_1(\rho/a, \theta - \theta_0) \tag{10}$$

$$M_2(\log\rho, \theta) = M_1(\log\rho - \log a, \theta - \theta_0) \tag{11}$$

Rotation and scaling can now both be recovered using the Fourier phase shifting property. After recovery of those parameters, image f_2 can be warped to compensate for the rotation and scaling. Finally, the standard phase correlation technique can be applied to recover the remaining translational offset between f_1 and f_2.

Implementation

The raw images are all first preprocessed using CLAHS to minimize lighting effects present in the imagery. The image is cosine windowed and then edge enhanced using a Laplacian of Gaussian filter. Mapping of the Fourier magnitude from Cartesian to polar coordinates only requires the top half of the spectra due to complex conjugate symmetry. Recovery of the translations θ and $\log a$ is achieved through normalized correlation of the two log-polar magnitudes M_1 and M_2. The image M_2 is then rotated and scaled according to the results. Finally, the warped M_2 is normalized correlated with M_1 to recover the translational offsets.

The parameters used allow us to, theoretically, resolve rotations of 0.3° and 0.44% change in scaling. Figure 2 shows example underwater imagery from the Derbyshire data set and the resulting warped imagery after recovery of the parameters.

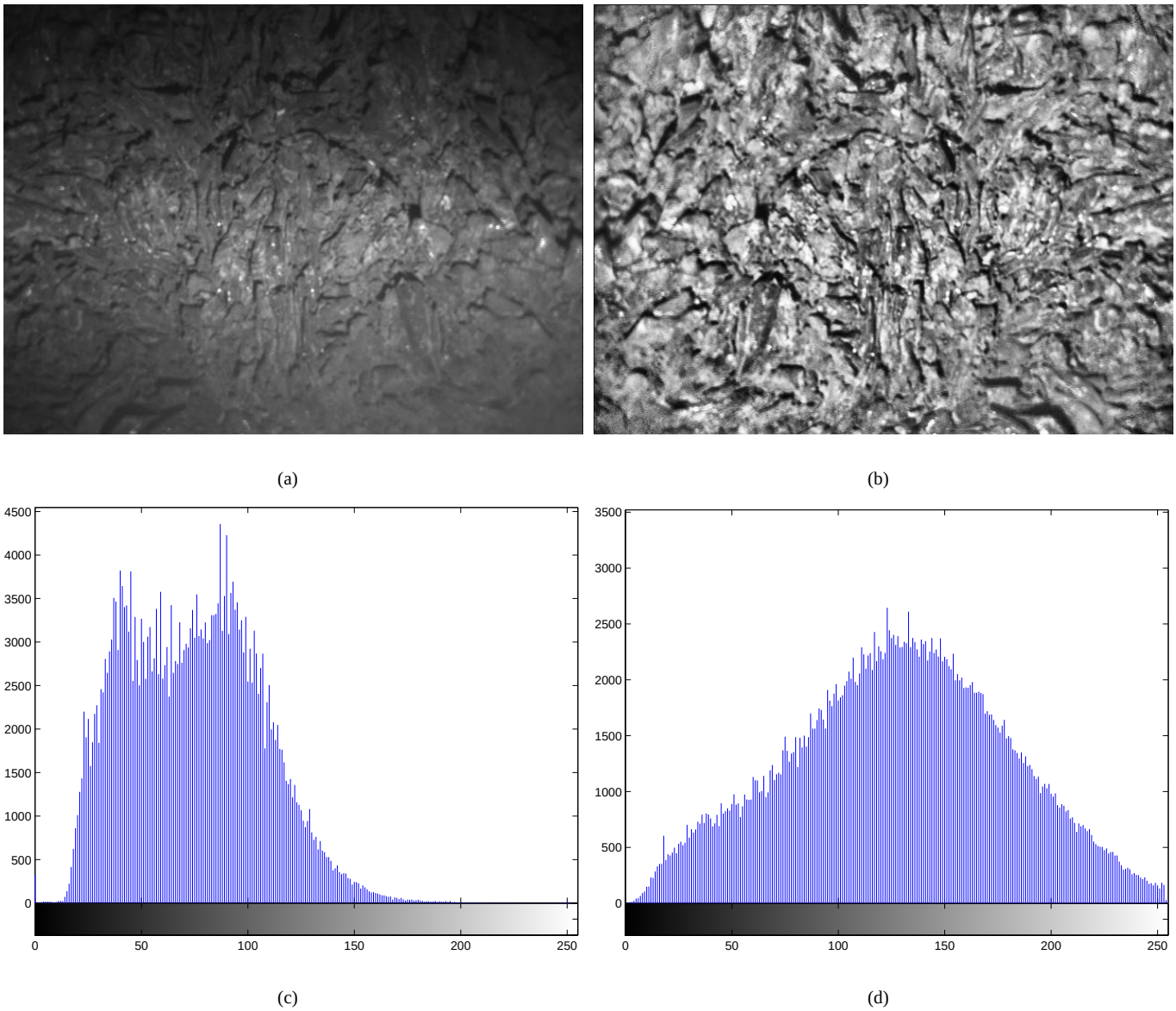


Fig. 1. *(a)* Original imagery collected by the AUV ABE acquired at a geological site of interest in the East Pacific Rise. *(b)* Adaptive histogram equalized imagery. While this technique compensates very well for the nonuniform lighting pattern, it cannot (as seen in the lower right corner of the image) compensate for parts of the image where the light intensity levels are of the order of the sensitivity of the camera. *(c)* Histogram of the original image. *(d)* Histogram of the CLAHS processed image.

IV. Local Normalized Correlation

Normalized correlation is a practical measure of similarity [3]. Normalized correlation of two signals is invariant to local changes in mean and contrast. When two signals are linearly related, their normalized correlation is 1, as seen in Equation 12 below. When two signals are not linearly related, but do contain similar spatial variations, normalized correlation will still yield a value close to unity [7].

$$S[x,y;u,v] = \frac{\sum_{i,j}(f[x+i,y+j]-\overline{f_W})\cdot(g[x+u+i,y+v+j)-\overline{g_W})}{\sqrt{\sum_{i,j}(f[x+i,y+j]-\overline{f_W})^2}\sqrt{\sum_{i,j}(g[x+u+i,y+v+j)-\overline{g_W})^2}} \quad (12)$$

The lack of rich features in underwater imagery poses difficult challenges for indirect feature based methods, and experimental evidence suggests that direct correlation based methods yield good results. We use local normalized correlation surfaces calculated for each pixel to determine a dense set of correspondences between images. This set of dense correspondences is then pruned by only considering pixels which have a concave correlation surface as reliable matches (We fit a quadratic surface near the correlation surface peak and analytically check for concavity as a method of outlier rejection [8]).

Figure 3 displays mosaic results when images are mosaicked together in a globally consistent manner utilizing all available cross-linked image pair correspondences. Local normalized correlation was used as the similarity measure for determining point correspondences.

V. Multiresolution Pyramidal Based Blending

Due to the rapid attenuation of light underwater, the only way to get a large area view of the seafloor is to build up a photo-

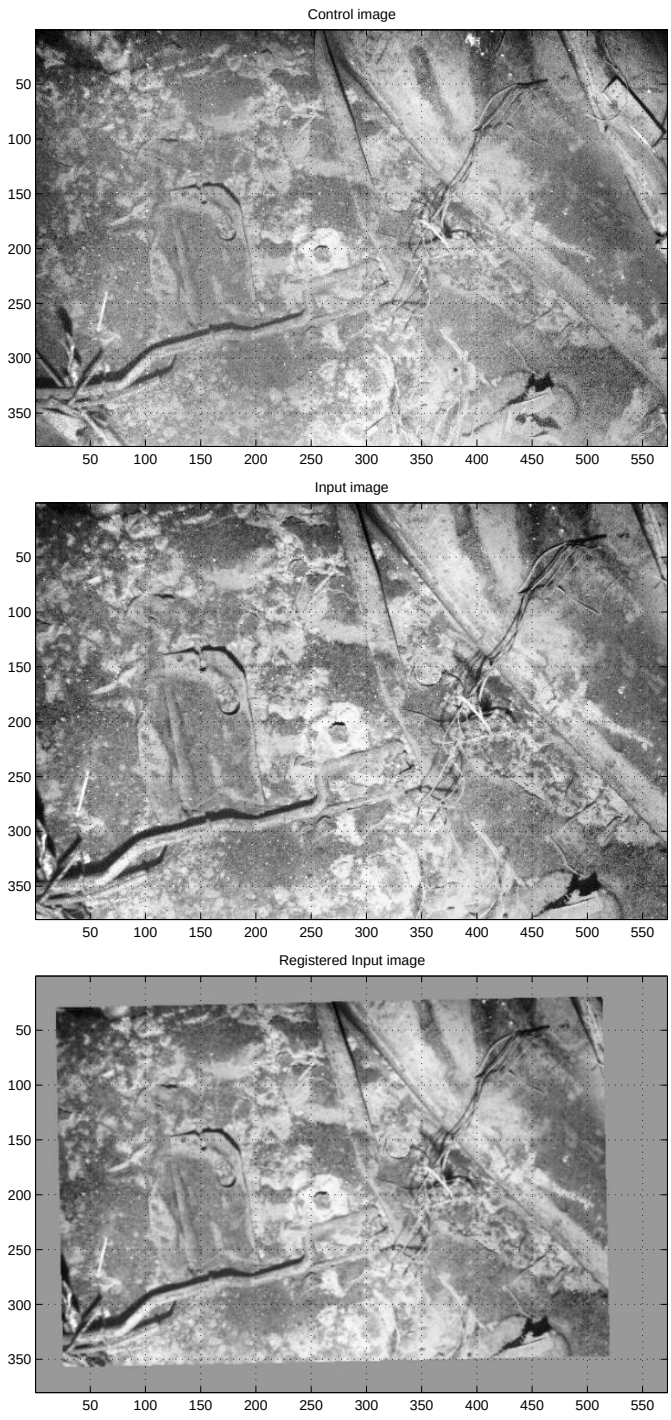


Fig. 2. Registration of underwater imagery from the Derbyshire data set using Fourier based methods to recover scale, rotation, and translation. The recovered parameters are $\theta_0 = 1.2°$, $scaling = 1.1567$, $x_0 = 18$, $y_0 = 18$.

mosaic from smaller local images as seen in Figure 3. The mosaic technique is used to construct an image with a far larger field of view and greater resolution than could be obtained with a single photograph. However, once the mosaic is generated, differences in image intensities due to image processing or acquisition can lead to clearly visible borders between images in the mosaic. A technical problem in image representation then, is how to join image borders so that the boundary between them is not visible?

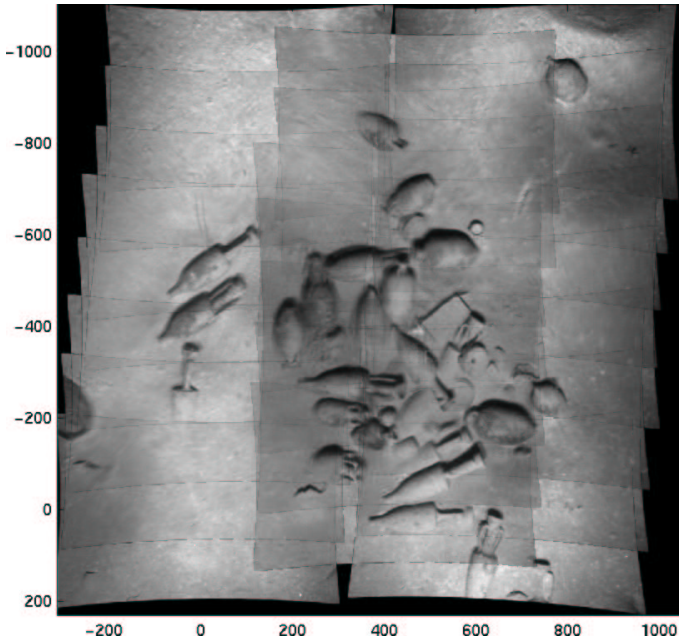


Fig. 3. A sequence of images which was mosaicked together in a globally consistent manner utilizing all available cross-linked image pair correspondences. To obtain good registration, especially along edges, we must compensate for lens radial distortion. The similarity metric used for point correspondence was local normalized correlation. The mosaic is rendered as the average of the intensities of overlapping pixels. To preserve mismatches, the results are presented without blending.

The problem can be viewed as joining two surfaces at their border, where the gray level values of the image $I(x, y)$ represent the surface's height above the (x, y) plane. The goal is to perturb each surface near the border so that they join smoothly without distorting the original surface features too grossly. Many methods are based upon a weighted sum technique where the size of the transition zone is an important parameter; too small of a transition zone relative to feature sizes results in the image border still being visible, albeit blurry, while too large of a transition zone results in a double exposure effect.

We implement a multiresolution pyramidal blending approach where the images to be blended are first decomposed into different band-pass frequency components, merged separately in each frequency band, and then reassembled into a single seamless composite mosaic [9]. The idea behind this technique is that the transition zone is optimally matched for feature sizes within each frequency band of the pyramid.

First, a Gaussian pyramid is constructed for each image where the base level in the pyramid, G_0, is the original image. Each successive level in the pyramid is a low-pass filtered and down sampled by factor of 2 version of the previous level, *i.e*

$$G_l[i, j] = \sum_{m=-2}^{2} \sum_{n=-2}^{2} w[m, n] G_{l-1}[2i + m, 2j + n] \quad (13)$$

where the 5×5 generating kernel, $w[m, n]$, is subject to the following four constraints:

1) For computational convenience, the kernel is separable, *i.e.* $w[m, n] = \hat{w}[m]\hat{w}[n]$.
2) The one-dimensional function $\hat{w}[\cdot]$ is symmetric.

3) $\hat{w}[\cdot]$ is normalized, *i.e.* $\sum_{i=-2}^{2} \hat{w}[i] = 1$.
4) Finally, each level l node must contribute the same total weight to level $l+1$ nodes resulting in the constraint: $\hat{w}[0] + 2\hat{w}[2] = 2\hat{w}[1]$.

Next, the different band-pass components are formed by generating the Laplacian pyramid. The Laplacian pyramid is generated from the Gaussian pyramid by expanding the image at the next higher level in the pyramid to the resolution of the current level and then subtracting them.

$$L_l[i,j] = G_l[i,j] - k \sum_{m=-2}^{2} \sum_{n=-2}^{2} G_{l+1}[\frac{i+m}{2}, \frac{j+n}{2}] \quad (14)$$

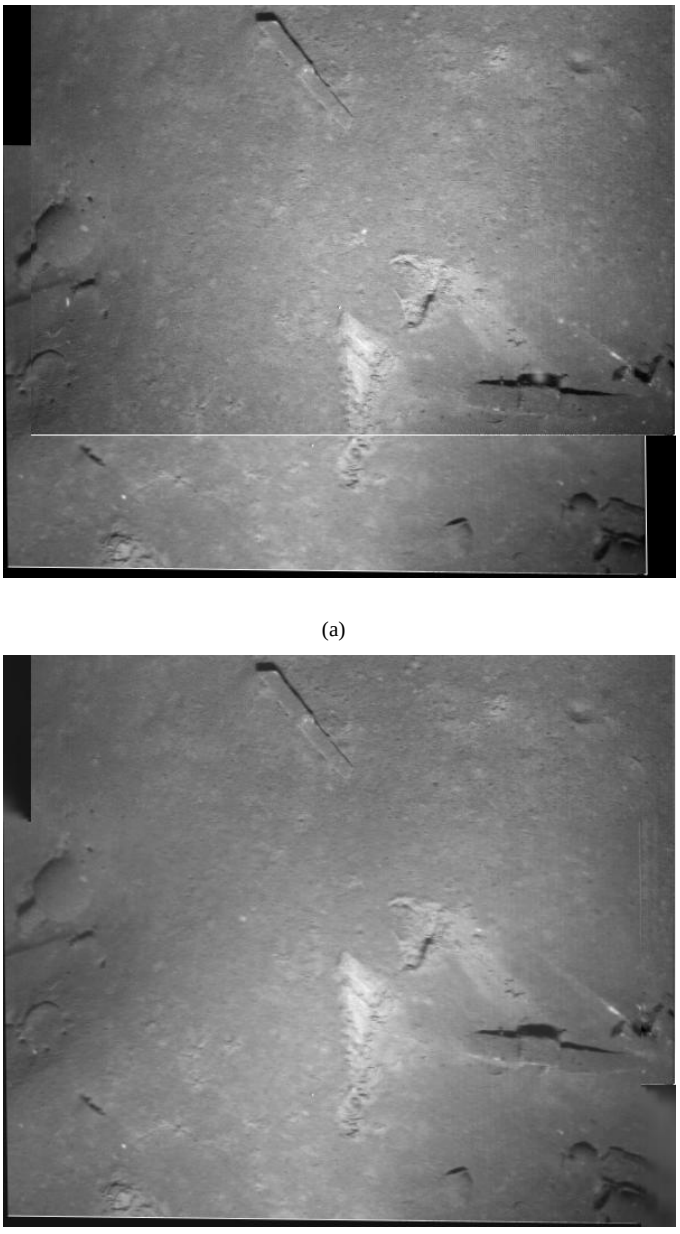
This results in each level of the Laplacian pyramid containing a separate one-octave, band-pass component of the original image. The two Laplacian pyramids, one for each image, are then merged at each level of the pyramid. The resulting seamless mosaic is then constructed by compressing the merged Laplacian pyramid via

$$I_{merged} = \sum_{l=0}^{N} L_{l,l_{merged}} \quad (15)$$

where N is the number of levels in the pyramid and the notation $L_{l,l}$ implies expansion of the level L_l, l times, up to the resolution of the base level, L_0. Figure 4 shows before and after results of the blending of a two image mosaic.

VI. Conclusions

This paper has presented results from our efforts to develop an extended MATLAB image processing and mosaicking toolbox. Proven algorithms from the land-based literature have been adapted and applied to the unique challenges of the underwater environment. The collection of algorithms presented in this paper group nicely into a unified framework for underwater imaging and mosaicking work. Our hierarchical framework of contrast limited adaptive histogram specification, Fourier based methods for image registration, local normalized correlation for a similarity measure, and multiresolution pyramidal based blending compose a core suite of functionality in a unified toolbox.



(a)

(b)

Fig. 4. *(a)* A two image mosaic with seam. The top image is overlaid over the bottom image. *(b)* The final blended result.

References

[1] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics Gems Iv*, Paul Heckbert, Ed., vol. IV, pp. 474–485. Academic Press, Boston, Date 1994.

[2] J. Lim, *Two-Dimensional Signal and Image Processing*, Prentice Hall, Englewood Cliffs, N.J., Date 1990.

[3] L.G. Brown, "A survey of image registration techniques," *ACM Computing Surveys*, vol. 24, no. 4, pp. 325–376, December 1992.

[4] E. De Castro and C. Morandi, "Registration of translated and rotated images using finite fourier transforms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, no. 5, pp. 700–703, September 1987.

[5] B.S. Reddy and B.N. Chatterji, "An fft-based technique for translation, rotation, and scale-invariant image registration," *IEEE Transactions on Image Processing*, vol. 5, no. 8, pp. 1266–1271, August 1996.

[6] O. Pizarro, H. Singh, and S. Lerner, "Towards image-based characterization of acoustic navigation," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, October 2000, vol. 3, pp. 1519–1524.

[7] M. Irani and P. Anandan, "Robust multi-sensor image alignment," in *Sixth International Conference on Computer Vision, 1998*, January 1996, pp. 959–966.

[8] R. Mandelbaum, G. Salgian, and H. Sawhney, "Correlation-based estimation of ego-motion and structure from motion and stereo," in *Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999*, Kerkyra, Greece, September 1999, vol. 1, pp. 544–550.

[9] P.J. Burt and E.H. Adelson, "A multiresolution spline with application to image mosaics," *ACM Transactions of Graphics*, vol. 2, no. 4, pp. 217–236, October 1983.