CIHSPS 2005 - IEEE International Conference on
Computational Intelligence for Homeland Security and Personal Safety
Orlando, FL, USA, 31 March – 1 April 2005

Non-iterative Construction of Super-Resolution Image from an Acoustic Camera Video Sequence

K. Kim[1], N. Neretti[1], N. Intrator[1]

[1]Institute for Brain and Neural Systems,
Brown University,
Providence RI 02912, USA
Phone: +1(401)863-3920, Fax: +1(401)863-3494, E–mail: kio@brown.edu.

Abstract – *The low visibility condition of underwater environments is often a limiting factor in the exploration of rivers or coastal areas. In such situations, relatively high resolution images produced by an acoustic camera can be an informative measure for the detection of terrain and objects. In this paper, we present a non-iterative super-resolution algorithm for the construction of a high resolution image from multiple frames of acoustic camera images. We discuss the geometry of insonification and image acquisition of an acoustic camera, and describe the detailed procedures for the registration, insonification correction, and fusion.*

Keywords – *acoustic camera, mosaicing, registration, video enhancement, super-resolution.*

I. INTRODUCTION

An acoustic camera is a high resolution ultrasonic imaging device that can produce underwater video sequences, very useful for the purpose of the surveillance of coastal areas [1]. It can be mounted on an autonomous underwater vehicle (AUV) or carried by a diver, and can be sent out to explore potentially harmful objects.

Despite the merits of acoustic cameras over other sonar systems, it still has shortcomings compared to normal optical cameras:

(i) Limitation of sight range: Unlike optical cameras which have a 2-D array of photosensors, acoustic cameras have a 1-D transducer array. 2-D representation is obtained from the temporal sequence of the transducer array. For this reason, it can collect information from a limited range.

(ii) Low signal-to-noise ratio (SNR): The transducer size is comparable to the wavelength of ultrasonic waves, so the intensity of a pixel depends not only on the amplitude, but also on the phase difference of the reflected signal. For this reason, the presence of background noise in underwater environments results in a Rician distribution of noise observed in ultrasound images [2]. The SNR is also significantly lower than in optical images because of the transducer size.

(iii) Low resolution with respect to optical images: Due to the large scale of the wavelength of ultrasound compared to the wavelength of light, the number of pixels in the horizontal axis is limited.

(iv) Inhomogeneous insonification: Since one dimension of the image is acquired merely based on the relative signal arrival times, a single insonifier was used. Consequently, due to the anisotropy of ultrasound radiation from the insonifier, the insonification is inhomogeneous in acoustic camera images.

The above limitations can be addressed by image mosaicing, which is broadly used to build a wider view image [3], [4], [5], or to estimate the motion of a vehicle [6], [7]. For ordinary images, mosaicing is also used for image enhancement such as denoising, deblurring, or super-resolution [8], [9].

In this paper, we describe the mosaicing of a sequence of acoustic camera images, and present a new non-iterative algorithm for the construction of a high resolution image from the image sequence.

II. IMAGING GEOMETRY

The transformation between two acoustic camera images can be calculated by putting one image into the coordinate system where the image is on the xy-plane with the positive y-axis along the center line of the image and the center of the arc at the origin (Fig. 1).

During the imaging process, a point denoted by a position vector $\boldsymbol{x} = (x, y, z)^\top$ is projected to the polar coordinates (r, a) as follows:

$$r = |\boldsymbol{x}|, \qquad \alpha = \sin^{-1}\frac{x}{r_{xy}}, \tag{1}$$

where $r_{xy} \equiv (x^2 + y^2)^{1/2}$, or to the Cartesian coordinates $\mathbf{u} = (u, v)$,

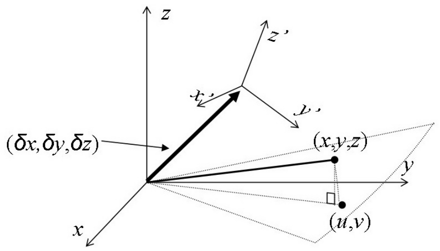


Fig. 1. The imaging geometry of an acoustic camera. The camera is located at the origin of the xyz-coordinate system with the pitch, yaw, and roll 0. In the next frame ($x'y'z'$-coordinate), the camera is displaced by $\delta\boldsymbol{x} = (\delta x, \delta y, \delta z)^\top$ and rotated by (ϕ, θ, ψ).

$$u = r\sin\alpha = x\sqrt{1+\tan^2\beta} \tag{2}$$

$$v = r\cos\alpha = y\sqrt{1+\tan^2\beta}, \tag{3}$$

where β is the angle between $\boldsymbol{x}$ and the imaging plane. When the camera is translated by $\delta\boldsymbol{x} = (\delta x, \delta y, \delta z)^\top$ and rotated by (ϕ, θ, ψ), the new coordinates of $\boldsymbol{x}$ are

$$\boldsymbol{x}' = (x', y', z')^\top = \boldsymbol{R}_{\phi\theta\psi}(\boldsymbol{x} - \delta\boldsymbol{x}) \tag{4}$$

where $\boldsymbol{R}_{\phi\theta\psi}$ is a 3×3 rotation matrix[1].

The linear transformation $\boldsymbol{T}$ between two images should satisfy

$$\begin{aligned}\begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix} &= \begin{pmatrix} x'\sqrt{1+\tan^2\beta'} \\ y'\sqrt{1+\tan^2\beta'} \\ 1 \end{pmatrix} \\ &= \boldsymbol{T}\begin{pmatrix} x\sqrt{1+\tan^2\beta} \\ y\sqrt{1+\tan^2\beta} \\ 1 \end{pmatrix} \\ &= \boldsymbol{T}\begin{pmatrix} u \\ v \\ 1 \end{pmatrix},\end{aligned}$$

where $\tan\beta' = z'/(x'^2 + y'^2)^{1/2}$. When the reflecting points of the target object are located roughly on a plane such as the sea floor, z can be approximated by

$$z = ax + by + z_0. \tag{5}$$

u' and v' can then be rewritten as

$$\begin{aligned} u' = &\left(1 + \frac{z'^2}{x'^2+y'^2}\right)^{\frac{1}{2}} \times \\ &\big\{ (R_{11} + R_{13}a)\,x + (R_{12} + R_{13}b)\,y \\ &- (R_{11}\delta x + R_{12}\delta y + R_{13}(\delta z - z_0)) \big\}, \\ v' = &\left(1 + \frac{z'^2}{x'^2+y'^2}\right)^{\frac{1}{2}} \times \\ &\big\{ (R_{21} + R_{23}a)\,x + (R_{22} + R_{23}b)\,y \\ &- (R_{21}\delta x + R_{22}\delta y + R_{23}(\delta z - z_0)) \big\}, \end{aligned}$$

where a, b, and $z'/(x'^2 + y'^2)^{1/2}$ are sufficiently small that their squares are negligible. For example, the maximum deviation angle $\beta_{\max}$ of a DIDSON system is $5°$, thus $(1 + \tan^2\beta_{\max})^{1/2} = 1.0038$.

Up to a first order of approximation, we have,

$$\boldsymbol{T} = \begin{pmatrix} R_{11} + R_{13}a & R_{12} + R_{13}b & \begin{matrix} -(R_{11}\delta x + R_{12}\delta y \\ +R_{13}(\delta z - z_0)) \end{matrix} \\ R_{21} + R_{23}a & R_{22} + R_{23}b & \begin{matrix} -(R_{21}\delta x + R_{22}\delta y \\ +R_{23}(\delta z - z_0)) \end{matrix} \\ 0 & 0 & 1 \end{pmatrix}. \tag{6}$$

This serves as a first order approximation of the transform between two acoustic camera images. Further approximation will be studied in subsequent work by segmenting the image into local planes depending on the levels of elevation.

The six unknown parameters of the affine transform can be obtained by matching features in two images. However, other parameters such as R_{ij}, a, b, or $\delta\boldsymbol{x}$ in (6) cannot be figured out separately because those parameters are coupled and underconstrained. Consequently, under the above approximation, it is impossible to reconstruct the precise egomotion of the acoustic camera merely based on image registration parameters.

III. METHODOLOGY

The typical four steps of image registration are: feature detection, feature matching, transformation estimation, and image resampling and transformation [10]. Feature detection is the process of finding objects such as corners, edges, line intersections, etc. from images, manually or automatically. The features from the sensed image are paired with the corresponding features in the reference image in the second step. In the third step, the transformation is estimated based on the displacement vector of each feature. Once the mapping is established, the multiple images are combined to generate a fusion image.

[1]

$$\boldsymbol{R}_{\phi\theta\psi} = \begin{pmatrix} \begin{matrix}\cos\phi\cos\psi \\ -\sin\phi\sin\theta\sin\psi\end{matrix} & -\sin\phi\cos\theta & \begin{matrix}\cos\phi\sin\psi \\ -\sin\phi\sin\theta\cos\psi\end{matrix} \\ \begin{matrix}\sin\phi\cos\psi \\ +\cos\phi\sin\theta\sin\psi\end{matrix} & \cos\phi\cos\theta & \begin{matrix}\sin\phi\sin\psi \\ -\cos\phi\sin\theta\cos\psi\end{matrix} \\ -\cos\theta\sin\psi & \sin\theta & \cos\theta\cos\psi. \end{pmatrix}.$$

In our work, we have found that high curvature points can be useful as features of interest in acoustic camera images. The sum of squared difference is used to measure the dissimilarity between two images in the second step. Transform parameters are estimated via a random sampling based method. After the parameters of the affine transform are obtained, all images are combined by weighted average.

A. Coordinate mapping and inhomogeneous insonification equalization

In order to restore the spatial homogeneity of the image, a transformation to the Cartesian coordinates has to be performed. Due to the fact that the field of view in the angular coordinate of different sensors does not overlap, the resulting pixel size in the Cartesian coordinates is not homogeneous. Therefore, nearest neighbor interpolation was applied to fill the gaps in the image in the Cartesian coordinate system.

Due to the acoustic acquisition of images which was performed by insonifying the area with a single source, an inhomogeneous intensity profile is obtained. This has to be corrected for efficient mosaicing.

Previous work on separation of illumination from reflectance has been based on Retinex theory [11], [12]. Using a homomorphic filtering method [13], one estimates the insonification of an image, and reconstructs the image under uniform insonification. Modeling the insonification as a homomorphic filter, the image produced by the sonar system can be written as

$$\tilde{\mathbf{I}}_i = \mathrm{L}_i \mathrm{M}_i \hat{\mathbf{I}} + \boldsymbol{\eta} \tag{7}$$

where $\tilde{\mathbf{I}}_i$ is the i-th observed image, L_i the i-th insonification operation, M_i the i-th down-sampling operation, $\hat{\mathbf{I}}$ the ground truth image under uniform insonification, and $\boldsymbol{\eta}$ a Gaussian noise. The estimated insonification intensity $\tilde{\mathrm{L}}_i$ is calculated by applying a Gaussian filter to the original image, $\tilde{\mathrm{L}}_i = \mathrm{diag}\left(\mathrm{H}\tilde{\mathbf{I}}_i\right)$, where H is a Gaussian blurring operation. The estimated uniform insonification image is

$$\tilde{\mathbf{I}}_i^{(u)} = \tilde{\mathrm{L}}_i^{-1}(\mathrm{L}_i \mathrm{M}_i \hat{\mathbf{I}} + \boldsymbol{\eta}) \simeq \mathrm{M}_i \hat{\mathbf{I}} + \tilde{\mathrm{L}}_i^{-1} \boldsymbol{\eta}. \tag{8}$$

The sum of squared difference between the i-th and j-th uniform insonification images is

$$SSD_{i,j} = ||\tilde{\mathbf{I}}_i^{(u)} - \tilde{\mathbf{I}}_j^{(u)}||^2 + ||\tilde{\mathrm{L}}_i^{-1}\boldsymbol{\eta}_i - \tilde{\mathrm{L}}_i^{-1}\boldsymbol{\eta}_i||^2 \tag{9}$$

and can be used as a dis-similarity measure. The second term in (9) is independent of the true image, and may be regarded as a constant, provided the noise is uniform.

A regularization factor may be added to $\tilde{\mathrm{L}}_i$ to prevent erroneously excessive intensity from the speckles in low insonification regions.

B. Feature detection and putative matching

Feature detection and matching are computationally demanding. A Gaussian pyramid has been proposed as a multiscale approach for efficient feature detection and matching [14], [10]. A pixel in the polar coordinates image corresponds to 1 or several pixels in the mapped image. For example, in an image with the range 8.25∼44.25m, the number of pixels that correspond to a single pixel in the polar coordinates image varied from 1 to 28. Magnified pixels result in jagged edges in the mapped image. In our images, feature detection at the third level of the Gaussian pyramid reduces false detection of corners at the jagged edges.

Feature detection and putative matching is initialized by translational displacement detection. Translational displacement between the sensed image and the reference image is calculated by an exhaustive search on the fourth level of the Gaussian pyramid. This process drastically reduces the area of exhaustive search.

After translation is estimated, high curvature points of the sensed image are detected using the Harris corner detector [15]. The second moment matrix M is computed using the following relationship:

$$\mathrm{M} = e^{-\mathbf{x}^\top \mathbf{x}/2\sigma_s^2} \otimes \left((\nabla I)(\nabla I)^\top\right), \tag{10}$$

where σ_s is the scale factor of corners, and ∇I is the gradient vector of the image. The response after the Harris corner detection is,

$$R = \det \mathrm{M} - k\mathrm{Tr}(\mathrm{M})^2, \tag{11}$$

where k is set to 0.04. The local maxima of R correspond to corners. The corners are matched to the corresponding points in the reference image by another exhaustive search on the third level of the Gaussian pyramid.

C. Transform estimation

Image changes due to the sonar system movement are modeled by an affine transformation as derived in the section II. The affine transformation describes the image changes by yaw, small pitch and roll and translational movement of the sonar system. This is valid when multiple objects are not present at the same range and angle, which is the case with the great majority of images in our dataset [1].

The detailed procedure of the algorithm is as follows:

(i) Feature points estimation: Using the Harris corner detector, compute 50 interest points in a preprocessed acoustic camera image.

(ii) Corresponding points search: For a square patch around each feature point in the sensed image, find the sub-pixel-wise displacement in the next image, using a cross-correlation based matching.

(iii) Transform parameter estimation: Repeat the following (1)-(3) for 1000 samples.

(1) Select 3 putative matching pairs.

(2) Using the matching pairs, estimate the parameters of the affine transform.

(3) Find the inliers of the estimated transform, and repeat (2) with the inliers until the estimated inliers are stabilized.

(iv) Set a certain k percentile to define a threshold n of feature points. Then, find the n pairs of points that are closest to each other. The least mean squared error of the pairs is used as the criterion.

We use the criterion of least square error of k% of samples, where k is empirically determined. It is similar to the least-median of squares (LMS) [16], but it differs in that it can have a lower breakdown point (k instead of 0.5 of LMS), and it uses the mean squared error instead of the k-th percentile as the measure of error. It works well with only a few feature point pairs with high percentage of outliers. In addition, it yields a measure of goodness of the transformation, which helps to decide whether to continue mosaicing or to stop, for example when the risk of mismatch is high.

D. Mosaicing

After registration, a mosaic image is constructed. Since the noise is present regardless of the insonification condition, it can deteriorate the mosaic image if not treated properly. For example, if we average well-insonified images and poorly-insonified images, the SNR will become lower because noise will be accumulated.

A uniformly insonified image is, when additive noise $\boldsymbol{\eta}$ of a standard deviation σ is added,

$$\hat{\mathbf{I}} + \boldsymbol{\eta}_\sigma \tag{12}$$

and the SNR is simply $1/\sigma$, and when it is insonified with inhomogeneous insonification L_i,

$$\mathrm{L}_i\hat{\mathbf{I}} + \boldsymbol{\eta}_\sigma \tag{13}$$

the SNR is $\frac{1}{\sigma}\mathrm{diag}(\mathrm{L}_i)$.

Mosaicing via uniform averaging can be described as the following relationship:

$$\tilde{\mathbf{I}}^{(u)}_{mosaic} = \frac{1}{N}\sum_{i=1}^{N}\tilde{\mathbf{I}}_i, \tag{14}$$

the SNR of the mosaic image is

$$SNR^{(u)}_{mosaic} = \frac{1}{\sigma\sqrt{N}}\sum_i \mathrm{L}_i. \tag{15}$$

Note that the SNR is a column vector of the same size as $\hat{\mathrm{L}}_i$ because the insonification intensity varies within the image.

A desirable situation is when poorly insonified regions receive lower weight in the averaging. Denote the weight of the i-th image by α_i, where $\sum_i \alpha_i = \mathbf{I}$, and $\mathbf{I}$ is the identity matrix. Then, the mosaic image produced via inhomogeneous weighting is,

$$\tilde{\mathbf{I}}^{(w)}_{mosaic} = \sum \alpha_i \tilde{\mathbf{I}}_i \tag{16}$$

of which the SNR, $SNR^{(w)}_{mosaic}$ is,

$$SNR^{(w)}_{mosaic}(\alpha_1, \ldots, \alpha_N) = \frac{1}{\sigma}\left(\sum \alpha_i^2\right)^{1/2}\sum \alpha_i \mathrm{L}_i \tag{17}$$

$$\leq \frac{1}{\sigma}\left(\sum \mathrm{L}_i^2\right)^{1/2}. \tag{18}$$

Equality holds in (18) when $\alpha_k = (\sum \mathrm{L}_i)^{-1}\mathrm{L}_k$. Thus, the maximum SNR of the mosaic image is achieved when the transformed images are combined as follows:

$$\tilde{\mathbf{I}}^{(w)}_{mosaic} = \left(\sum \mathrm{L}_i\right)^{-1}\sum \mathrm{L}_i\mathbf{I}_i. \tag{19}$$

E. Maximum-Likelihood estimation of equalized image

Together with the weighted mosaic image, the image quality can be enhanced by an additional process. According to the inhomogeneous insonification model, the intensity of a reflected signal varies depending on the insonification condition. Given the observed pixel values $\tilde{\mathbf{I}}_1, \tilde{\mathbf{I}}_2, ...\tilde{\mathbf{I}}_N$ under the insonification intensity $\mathrm{L}_1, \mathrm{L}_2, ...\mathrm{L}_N$, the estimated true reflectivity image is $\tilde{\mathbf{I}}$ that maximizes

$$P(\tilde{\mathbf{I}}|\tilde{\mathbf{I}}_1, \tilde{\mathbf{I}}_2, ...\tilde{\mathbf{I}}_N) = \frac{P(\tilde{\mathbf{I}}_1, \tilde{\mathbf{I}}_2, ...\tilde{\mathbf{I}}_N|\tilde{\mathbf{I}})P(\tilde{\mathbf{I}})}{P(\tilde{\mathbf{I}}_1, \tilde{\mathbf{I}}_2, ...\tilde{\mathbf{I}}_N)} \tag{20}$$

and, by approximating the Rician noise distribution by a Gaussian distribution [17], we have the maximum likelihood image

$$\begin{aligned}\tilde{\mathbf{I}}_{mle} &= \arg\max_{\tilde{\mathbf{I}}} \ln P(\tilde{\mathbf{I}}_1, \tilde{\mathbf{I}}_2, ...\tilde{\mathbf{I}}_N|\tilde{\mathbf{I}}) \\ &= \arg\max_{\tilde{\mathbf{I}}} -\frac{1}{2\sigma^2}\Big\{\mathrm{L}_1^2\left(\mathrm{L}_1^{-1}\tilde{\mathbf{I}}_1 - \tilde{\mathbf{I}}\right)^2 + ... \\ &\qquad + \mathrm{L}_N^2\left(\mathrm{L}_N^{-1}\tilde{\mathbf{I}}_N - \tilde{\mathbf{I}}\right)^2\Big\}\end{aligned}$$

where the denominator of (20) is neglected. By combining the exponents, finally,

$$\tilde{\mathbf{I}}_{mle} = \arg\max_{\tilde{\mathbf{I}}} -\frac{1}{2\sigma_0^2}\left(\tilde{\mathbf{I}}_0 - \tilde{\mathbf{I}}\right)^2 \tag{21}$$

where

$$\frac{1}{\sigma_0^2} = \frac{1}{\sigma^2}\sum_i \mathrm{L}_i^2, \tag{22}$$

$$\tilde{\mathbf{I}}_0 = \left(\sum_i \mathrm{L}_i^2\right)^{-1}\sum_i \mathrm{L}_i\tilde{\mathbf{I}}_i. \tag{23}$$

This result is useful since we can choose an appropriate prior $P(\tilde{\mathbf{I}})$ to get the best a posteriori estimation of the equalized image. Note that the maximum-likelihood estimation result has the same weight as in (19), and the only difference is that the sum of the insonification intensity in (19) is replaced by the square sum in (23). Therefore, this maximum likelihood estimation image has the same maximum SNR property as the weighted mosaic image in (19).

F. Estimation of the Prior from Point Spread Function

A more detailed procedure of image formation from an imaging device is typically modelled by the following equation:

$$\tilde{\mathbf{I}}_i = \mathrm{L}_i \mathrm{M}_i \mathrm{P}_i \hat{\mathbf{I}} + \boldsymbol{\eta}$$

where $\hat{\mathbf{I}}$ represents the true or high-resolution image, P_i the i-th point spread function operation, M_i the down-sampling operation, L_i the illumination (or insonification for sonar systems), $\boldsymbol{\eta}$ the additive noise, and $\tilde{\mathbf{I}}_i$ the i-th observed image.

The process point spread function (PSF) and down-sampling operation fuse multiple pixels in the high resolution image into one pixel in the low resolution image. Usually, the fusion is performed with different spatial weights given to the high resolution pixels, which is in most of the cases modeled by Gaussian distributions. The fusion process can be represented as the following:

$$\boldsymbol{\theta}^\top \mathbf{w} = \bar{\theta} \tag{24}$$

$$(1 \;\; 1 \;\; \ldots \;\; 1) \cdot \mathbf{w} = 1, \tag{25}$$

where $\bar{\theta}$ is a low resolution pixel value, $\mathbf{w}$ the weight vector, $\boldsymbol{\theta}$ a vector of the pixels in the high resolution image that correspond to the pixel in the low resolution image.

On the other hand, we have another constraint to estimate the probability density function $\boldsymbol{\theta}$, which is that a pixel value is not independent of its adjacent pixel values. A Gaussian Markov Random Field is one of the popular models for this property, which gives us the following probability density function of $\boldsymbol{\theta}$:

$$P(\boldsymbol{\theta}) = N \exp\left(-\boldsymbol{\theta}^\top (\mathrm{D}^\top \mathrm{D}) \boldsymbol{\theta}\right), \tag{26}$$

where D is an operation of first derivative, and N a normalization constant.

Combining those two constraints, we obtain the prior, and the final maximum a posteriori (MAP) image can be obtained:

$$\tilde{\mathbf{I}}_{map} = \left(\sum_i \mathrm{L}_i^2 \mathrm{P}_i\right)^{-1} \sum_i \mathrm{L}_i \mathrm{P}_i \tilde{\mathbf{I}}_i, \tag{27}$$

where P_i is the point spread function operated onto the ground truth image.

IV. RESULTS

The algorithm was tested on a boat wreckage sequence[2]. A DIDSON system [1] scanned a shipwreck at approximately 30 meters depth for 285 seconds and took 446 frames of images. About 40 frames among them show the vessel from head to stern, and another 40 frames show it from stern to head. We applied the algorithm to those two sub-sequences, and built two mosaic images.

Fig. 2 depicts two consecutive acoustic images together with a set of matched (circle) and non-matched (triangle) feature points. These matched feature points in the reference image, which were found using the cross-correlation of patches around the feature points in the sensed image, are used to estimate the geometric transformation between the two images.

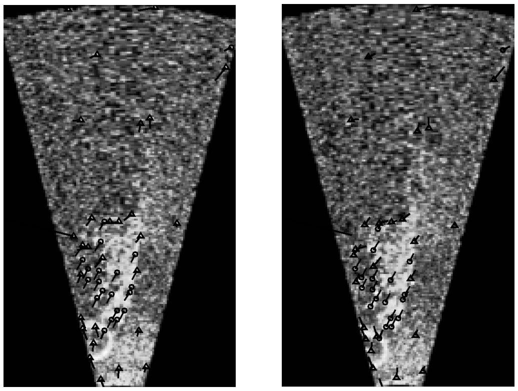


Fig. 2. Matched (circle) and non-matched (triangle) feature points which were obtained from the third level of the Gaussian pyramid using the Harris corner detector. Features from a sensed image are paired with corresponding points in the reference image. A 15×15 patch around each feature point in the sensed image is matched with the same sized patch from the corresponding 21×21 area in the reference image. The outliers (features with weaker matching) are defined by those pairs with higher matching error after the estimated transformation.

Cross-correlation was found to be more robust than a more conventional approach in which features are independently found and matched between the two images. This is a consequence of the high noise in the image and the fact that the exact location of the feature is not so well defined. Fig. 3 represents the main result of the paper, i.e., a mosaic image of multiple acoustic images. The mosaiced image contains information which spans multiple frames, each frame corresponding to a small portion of the insonified object. The combination images which have been transformed to be in the same coordinate system provide subpixel image resolution enhancement. Fig. 4(a) shows a frame of another shipwreck target before mosaicing. The resolution enhancement follows from the fact that one pixel in the original polar coordinate system is mapped to multiple pixels with the same intensity in the Cartesian coordinate system. Different frames lead to partial overlap of these

[2] The data has been provided by E. O. Belcher from Applied Physics Laboratory, University of Washington under ONR support.

multiple pixels, so that after averaging a subpixel resolution is achieved (See Fig. 4(a)).

Averaging of different acoustic images after bringing them to the same coordinate system (same viewpoint) leads to the classical effect of denoising. This is clearly seen in Fig. 3 on the whole target, and in particular in the comparison of a detailed structure of the other target in Fig. 4.

The top two panels of Fig. 3 depict a mosaiced image from the same sequence of acoustic images. In panel (b), the insonification profile was utilized during averaging. Panel (c) and (d) represent the same target from a different acoustic image sequence with panel (d) utilizing the insonification profile during averaging.

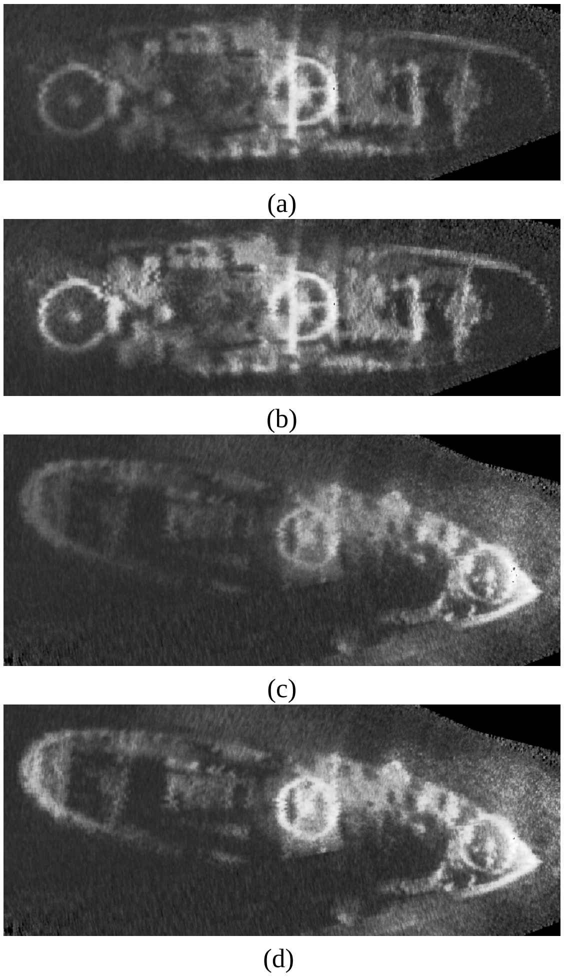
V. CONCLUSION

Acoustic camera technology is becoming essential for underwater exploration, and has the potential for the application to other areas for visual investigations. The acoustic camera, with its specific sensor design, poses some challenges in terms of image resolution, noise removal and area coverage.

In this paper, we have presented a complete algorithm to achieve image mosaicing, denoising and resolution enhancement from a sequence of acoustic camera images. We described the steps that were required to achieve this mosaicing. This included modeling the specific geometry of acoustic camera images which sharply differs from pinhole camera geometry.

The different geometry, and in particular, the fact that the images are acquired in a polar coordinate system, complicates the search and matching of feature points in consecutive images. Moreover, in this particular geometry, pixels in the polar coordinate system are mapped to a collection of pixels with the same intensity in the Cartesian coordinate system. Since consecutive images were taken from different viewpoints, a subpixel enhancement effect was achieved in the process of averaging in addition to the denoising effect. We have presented a novel method in which features extracted by a certain algorithm are locally matched to the reference image via cross-correlation. This method was found to be more robust than a conventional approach in which features are independently found and matched between two images. In particular, this is more pronounced when the number of pixels available for feature comparison is limited.

The different geometry, and in particular the fact that the images are acquired in a polar coordinate system complicates the search and matching of feature points in consecutive images. Moreover, in this particular geometry, pixels in the polar coordinate system are mapped to a collection of pixels with the same intensity in the Cartesian coordinate system. Since consecutive images were taken from different viewpoints, a subpixel enhancement effect was achieved in the process of averaging in addition to the denoising effect. We have presented a method in which features are found in a one image, and are then locally matched to the another image via cross-correlation.



(a)

(b)

(c)

(d)

Fig. 3. Demonstration of weighted averaging effect: Mosaicing was performed with 38 images which were averaged after the corresponding motion compensation transformation was applied to each of them. Panel (a) demonstrates the uniform average of the whole images, while panel (b) demonstrates the weighted average, in which insonification profile was utilized during averaging. (See section III. *D* for details.) Panel (c) and (d) depict the same target from a different image sequence.

ACKNOWLEDGMENT

This work was partly supported by ONR grant N00014-02-C-0296. We thank E. O. Belcher for answering all our questions about the data. We would like to send Professor Leon N Cooper and other members of IBNS for providing valuable comments.

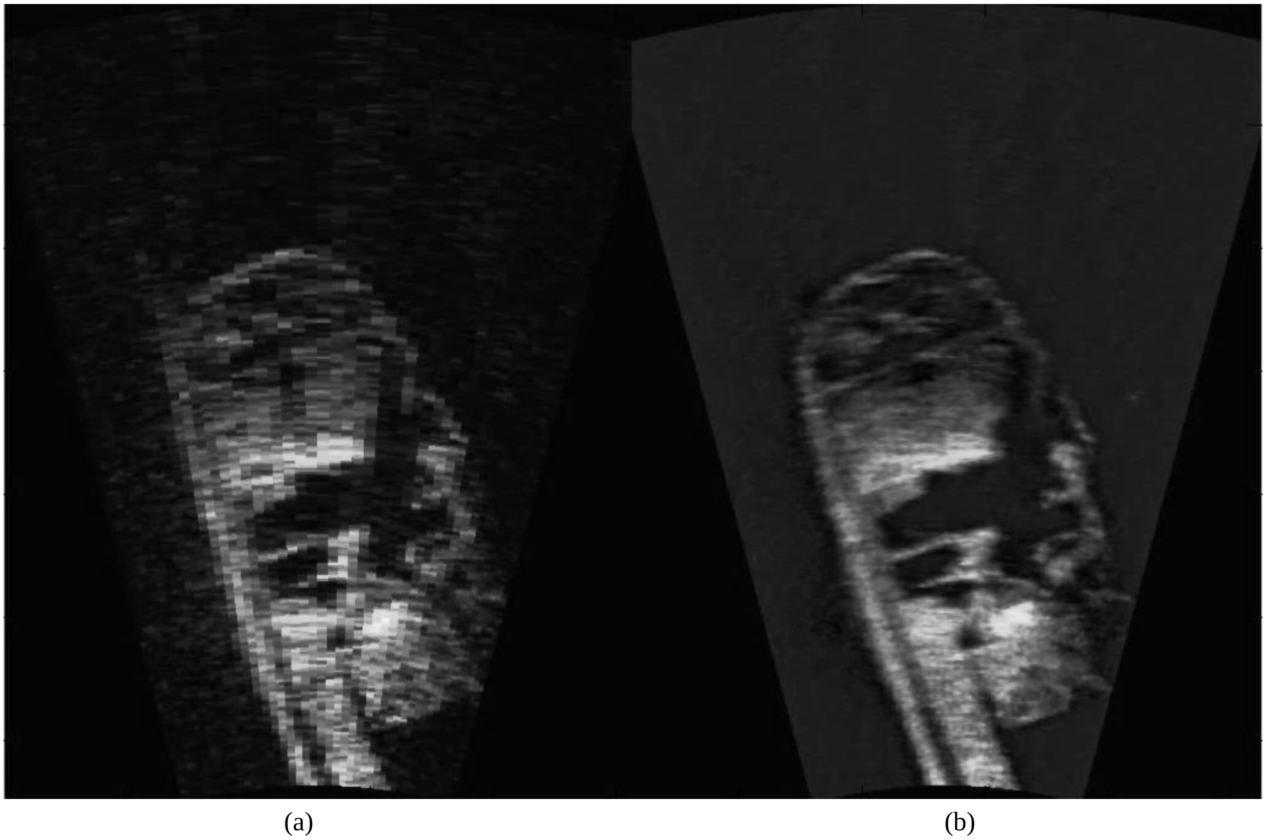


Fig. 4. Resolution enhancement by weighted averaging images. (a) A single frame image. (b) A mosaic image of 7 consecutive frames followed by a geometric transformation. (See Section III. *F*.)

REFERENCES

[1] E. O. Belcher, B. Matsuyama, and G. M. Trimble, “Object identification with acoustic lenses,” in *Proceedings of Oceans ’01 MTS/IEEE*, 2001, pp. 6–11.

[2] R. F. Wagner, S. W. Smith, J. M. Sandrik, and H. Lopez, “Statistics of speckle in ultrasound b-scans,” *IEEE Transactions on Sonics and Ultrasonics*, vol. 30, no. 3, pp. 156–163, 1983.

[3] R. L. Marks, S. M. Rock, and M. J. Lee, “Real-time video mosaicking of the ocean floor,” *IEEE Journal of Oceanic Engineering*, vol. 20, no. 3, pp. 229–241, 1995.

[4] Y. Rzhanov, L. M. Linnet, and R. Forbes, “Underwater video mosaicing for seabed mapping,” in *International Conference on Image processing*, 2000, vol. 2, pp. 224–227.

[5] U. Castellani, A. Fusiello, and V. Murino, “Registration of multiple acoustic range views for underwater scene reconstruction,” *Computer Vision and Image Understanding*, vol. 87, pp. 78–89, 2002.

[6] R. García, X. Cufí, and L. Pacheco, “Image mosaicking for estimating the motion of an underwater vehicle,” in *5th IFAC Conference on Manoeuvreing and Control of Marine Craft*, 2000.

[7] S. Negahdaripour, X. Xu, and A. Khamene, “A vision system for real-time positioning, navigation and video mosaicing of sea floor imagery in the application of rovs/auvs,” in *IEEE Workshop on Applications of Computer Vision*, 1998, pp. 248–249.

[8] D. Capel and A. Zisserman, “Computer vision applied to super-resolution,” *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 72–86, 2003.

[9] A. Smolic and T. Wiegand, “High-resolution video mosaicing,” in *International Conference on Image Processing*, 2001, vol. 3, pp. 872–875.

[10] B. Zitová and J. Flusser, “Image registration methods: a survey,” *Image and Vision Computing*, vol. 21, pp. 977–1000, 2003.

[11] E. H. Land, “Recent advances in the retinex theory and some implications for cortical computations: Color vision and the natural image,” *PNAS*, vol. 80, pp. 5163–5169, 1983.

[12] E. H. Land and J. J. McCann, “Lightness and the retinex theory,” *Journal of the Optical Society of America*, vol. 61, pp. 1–11, 1971.

[13] E. H. Land, “An alternative technique for the computation of the designator in the retinex theory of color vision,” *PNAS*, vol. 83, pp. 3078–3080, 1986.

[14] D. A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*, Pearson Education, Inc., Upper Saddle River, NJ, 2003.

[15] C. J. Harris and M. Stephens, “A combined corner and edge detector,” in *Proceedings on 4th Alvey Vision Conference*, 1988, pp. 189–192.

[16] P. J. Rousseeuw, “Least median of square regression,” *Journal of the American Statistical Association*, vol. 79, pp. 871–880, 1984.

[17] J. Sijbers, J. den Dekker, P. Scheunders, and Dirk Van Dyck, “Maximum-likelihood estimation of rician distribution parameters,” *IEEE Transactions on Medical Imaging*, vol. 18, no. 3, pp. 357–361, 1998.