Mosaicing of acoustic camera images

K. Kim, N. Neretti and N. Intrator

Abstract: An algorithm for image registration and mosaicing on underwater sonar image sequences characterised by a high noise level, inhomogeneous illumination and low frame rate is presented. Imaging geometry of acoustic cameras is significantly different from that of pinhole cameras. For a planar surface viewed through a pinhole camera undergoing translational and rotational motion, registration can be obtained via a projective transformation. For an acoustic camera, it is shown that, under the same conditions, an affine transformation is a good approximation. A novel image fusion method, which maximises the signal-to-noise ratio of the mosaic image is proposed. The full procedure includes illumination correction, feature based transformation estimation, and image fusion for mosaicing.

1 Introduction

The acquisition of underwater images is performed in noisy environments with low visibility. For optical images in those environments, often natural light is not available, and even if artificial light is applied, the visible range is limited.

For this reason, sonar systems are widely used to obtain images of seabed or other underwater objects.

An acoustic camera is a novel device that can produce a real time underwater image sequence. Detailed imaging methods of acoustic cameras can be found in [1]. Acoustic cameras provide extremely high resolution (for a sonar) and rapid refresh rates [1]. Despite those 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 size of the transducers 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. This is the reason for the Rician distribution of the ultrasound image noise. In addition, there is often a background ultrasound noise in underwater environments. It follows that the SNR is significantly lower than in optical images.
(iii) Low resolution with respect to optical images: owing to the limitation in the transducer size, the number of transducers that can be packed in an array is physically restricted, and so is the number of pixels in the horizontal axis. For example, a mine reacquisition and identification sonar (MIRIS) has 64 transducers [1].
(iv) Inhomogeneous insonification: The unique geometry of an acoustic camera requires the sonar device to be aligned parallel to the surface of interest, so that the whole surface falls within the vertical field of view of the acoustic camera [1]. This alignment is not always trivial, and the misalignment often makes dark areas in acoustic camera images.

The above limitations can be addressed by image mosaicing, which is broadly used to build a wider view image [2–4], or to estimate the motion of a vehicle [5, 6]. For ordinary images, mosaicing is also used for image enhancement such as denoising, deblurring, or super-resolution [7, 8].

There has been extensive research on image mosaicing, and its applications [9–13]. However, standard methods for image registration [14, 15] are not directly applicable to acoustic camera images, because of the discrepancy of image quality, inhomogeneous insonification profile, and different geometry. Marks *et al.* have described a mosaicing algorithm of the ocean floor taken with an optical camera [2]. Rzhanov *et al.* have also described a mosaicing algorithm of underwater optical images resulting in high resolution seabed maps [3]. Both of them deal with a similar problem of illumination, but use different methods: image matching by edge detection and Fourier based matching, which are not directly related to our work. In addition, since their mosaicing algorithms are not intended for image quality enhancement, we need to come up with a different mosaicing algorithm.

In this paper, we describe a mosaicing algorithm for a sequence of acoustic camera images. We show that an affine transformation is appropriate for images taken from an acoustic camera undergoing translational and rotational motion. We propose a method to register acoustic camera images from a video sequence using a feature matching algorithm. Based on the parameters of image registration, a mosaic image is built. During the mosaicing, the image quality is enhanced in terms of SNR and resolution.

2 Properties of acoustic camera images

Sonar image acquisition includes several steps, insonification, scattering, and detection of the returning signal. In this Section, we describe physical aspects of images acquired from acoustic lens sonar systems, or acoustic cameras.

© IEE, 2005

IEE Proceedings online no. 20045015

doi: 10.1049/ip-rsn:20045015

Paper first received 21st May 2004 and in revised form 22nd April 2005

The authors are with the Institute for Brain and Neural Systems, Brown University, Box 1843 Providence RI 02912, USA

E-mail: kio@brown.edu

Downloaded 07 Jan 2006 to 128.148.206.57. Redistribution subject to IEE licence or copyright, see http://ieedl.org/copyright.jsp

The emission and reception of ultrasound pulses by an acoustic camera is restricted within the vertical field of view, which is $\pm 5^\circ$ from the plane of image acquisition. When the object is out of this vertical field of view, it appears dark in the image as it is poorly insonified. This property makes a typical insonification pattern in acoustic camera images, which consequently brings out the necessity of insonification correction for registration and mosaicing of the images.

The pixel size and angular coverage of acoustic cameras vary depending on the type of the camera. We have used a dual-frequency identification sonar (DIDSON) system, which has been developed for the purpose of underwater target exploration [1, 16].

The DIDSON system has 96 transducers and the horizontal field of view is 28°. A set of acoustic lenses focuses the returning signal such that each sensor has a receiving beamwidth 0.3° in the horizontal axis, and 10° in the vertical axis. Each transducer produces an intensity profile that corresponds to a specific angle where the range information is obtained from the focal length of the acoustic lens array. The result is either a 96×512 or a 48×512 polar coordinates image, which has to be mapped to Cartesian coordinates in order to recover the original geometry.

Since the shortest range a DIDSON system can scan is 0.75 m and the maximum window length is 36 m, the ratio of the largest pixel size and the smallest pixel size can be up to $(36.75/0.75) = 49$. This means, a pixel in the polar coordinates image can occupy from one to 49 pixels in the Cartesian coordinates image.

Like other B-scan ultrasonic devices, acoustic cameras obtain pixel values by calculating the intensity of the returning signal. Owing to the diverse sources of background noise, the actual water pressure observed at a transducer is the sum of multiple waves with different phase. This is often approached through a random walk problem in the phase space, and brings a different noise structure called Rayleigh distribution when a signal is not present in the image, and in general the noise is modelled by Rician distribution [17]. A Rician probability density function (PDF) is in many cases approximated by a Gaussian PDF with the justification that when the SNR is high, their probability density functions are almost the same [18].

3 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 centre line of the image and the centre 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

$$r = |x| \tag{1}$$

$$\alpha = \sin^{-1} \frac{x}{r_{xy}} \tag{2}$$

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

$$u = r \sin \alpha = x\sqrt{1 + \tan^2 \beta} \tag{3}$$

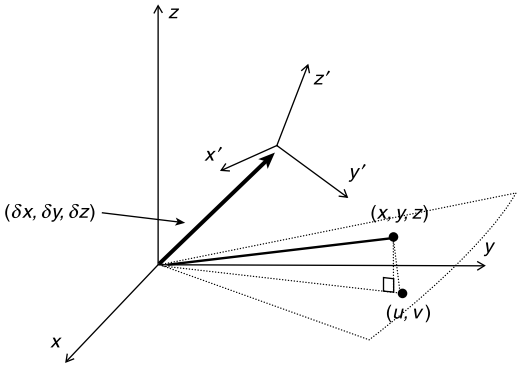


Fig. 1 *Imaging geometry of an acoustic camera*

The camera is located at the origin of the xyz-coordinate system with the pitch, yaw, and roll each set at 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)

$$v = r \cos \alpha = y\sqrt{1 + \tan^2 \beta} \tag{4}$$

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

$$x' = (x', y', z')^\top = R_{\phi\theta\psi}(x - \delta x) \tag{5}$$

where the rotation matrix $\boldsymbol{R}_{\phi\theta\psi}$ is a 3×3 matrix

$$\boldsymbol{R}_{\phi\theta\psi} = \begin{pmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{pmatrix} \tag{6}$$

$$\begin{aligned}
R_{11} &= \cos\phi\cos\psi - \sin\phi\sin\theta\sin\psi \\
R_{12} &= -\sin\phi\cos\theta \\
R_{13} &= \cos\phi\sin\psi - \sin\phi\sin\theta\cos\psi \\
R_{21} &= \sin\phi\cos\psi + \cos\phi\sin\theta\sin\psi \\
R_{22} &= \cos\phi\cos\theta \\
R_{23} &= \sin\phi\sin\psi - \cos\phi\sin\theta\cos\psi \\
R_{31} &= -\cos\theta\sin\psi \\
R_{32} &= \sin\theta \\
R_{33} &= \cos\theta\cos\psi.
\end{aligned}$$

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

$$\begin{aligned}
\begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix} &= \begin{pmatrix} x'\sqrt{1 + \tan^2 \beta'} \\ y'\sqrt{1 + \tan^2 \beta'} \\ 1 \end{pmatrix} \\
&= \boldsymbol{T} \begin{pmatrix} x\sqrt{1 + \tan^2 \beta} \\ y\sqrt{1 + \tan^2 \beta} \\ 1 \end{pmatrix} \\
&= \boldsymbol{T} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix}
\end{aligned} \tag{7}$$

where

$$\beta' = \tan^{-1} \frac{z'}{(x'^2 + y'^2)^{1/2}}$$

When the reflecting points of the target object are located roughly on a plane such as the sea floor, z can be approximated by

Downloaded 07 Jan 2006 to 128.148.206.57. Redistribution subject to IEE licence or copyright, see http://ieedl.org/copyright.jsp

$$z = ax + by + z_0 \tag{8}$$

$u'$and v' can then be rewritten as

$$\begin{aligned} u' =& \left(1 + \frac{z'^2}{x'^2 + y'^2}\right)^{1over2} \\ & \times \{(R_{11} + R_{13}a)x + (R_{12} + R_{13}b)y \\ & - (R_{11}\delta x + R_{12}\delta y + R_{13}(\delta z - z_0))\} \\ v' =& \left(1 + \frac{z'^2}{x'^2 + y'^2}\right)^{\frac{1}{2}} \\ & \times \{(R_{21} + R_{23}a)x + (R_{22} + R_{23}b)y \\ & - (R_{21}\delta x + R_{22}\delta y + R_{23}(\delta z - z_0))\} \end{aligned} \tag{9}$$

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 $\tan^2 \beta_{\max} = 0.0038$.

Up to a first order of approximation, we have

$$\begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix} = \boldsymbol{T} \begin{pmatrix} u \\ u \\ 1 \end{pmatrix} \tag{10}$$

where

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

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

The six unknown parameters of the affine transformation can be obtained by matching features in two images. However, other parameters such as R_{ij}, a, b, or $\delta\boldsymbol{x}$ in (11) cannot be figured out separately because those parameters are coupled and under-constrained. Consequently, under the above approximation, it is impossible to reconstruct the precise motion of the acoustic camera merely based on image registration parameters.

4 Methodology

The typical four steps of image registration are: feature detection, feature matching, transformation estimation, and image resampling and transformation [14]. Feature detection is the process of finding objects such as corners, edges, line intersections, etc., 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 between images is established, the multiple images are combined to generate a mosaic image.

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. Transformation parameters are estimated via a random sampling based method. After the parameters of the affine transformation are obtained, all images are combined by weighted average.

4.1 Coordinate mapping and inhomogeneous insonification equalisation

In order to restore the spatial homogeneity of the image, a transformation to the Cartesian coordinates has to be performed. Owing 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 neighbour interpolation was applied to fill the gaps in the image in the Cartesian coordinate system.

Owing 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 image registration and mosaicing. For example, Rzhanov *et al.* have subtracted a 2-D polynomial spline of the image from the original image [3]. Previous work on separation of illumination from reflectance was based on the Retinex theory [19]; The Retinex theory was designed for optical images with low noise. Using a homomorphic filtering method with a Gaussian retinex surround [20], Jobson *et al.* estimated the illumination of an image, and reconstructed the image under uniform illumination. While noise is stronger with an acoustic camera, we demonstrate that, when including the noise term in the model, the sum of squared difference is still a good dissimilarity measure after the retinex rendition.

The noisy image is modeled by

$$I(\boldsymbol{u}) = L(\boldsymbol{u})\hat{I}(\boldsymbol{u}) + \eta_{\sigma_G}(\boldsymbol{u}) \tag{12}$$

where $I(\boldsymbol{u})$ is the observed image, $L(\boldsymbol{u})$ the insonification intensity, $\hat{I}(\boldsymbol{u})$ the normalised image under uniform insonification, and $\eta_{\sigma_G}(\boldsymbol{u})$ a Gaussian noise with standard deviation σ_G at $\boldsymbol{u}$. The estimated insonification intensity $\tilde{L}$ is calculated by applying a Gaussian filter to the original image, $\tilde{L}(\boldsymbol{u}) = I(\boldsymbol{u}) \otimes e^{-|\boldsymbol{u}|^2/2\sigma^2}$ and the estimated uniform insonification image is

$$\begin{aligned} \tilde{I}(\boldsymbol{u}) =& \frac{L(\boldsymbol{u})}{\tilde{L}(\boldsymbol{u})}\hat{I}(\boldsymbol{u}) + \frac{\eta(\boldsymbol{u})}{\tilde{L}(\boldsymbol{u})} \\ \simeq& \hat{I}(\boldsymbol{u}) + \frac{\eta(\boldsymbol{u})}{\tilde{L}(\boldsymbol{u})} \end{aligned} \tag{13}$$

The sum of squared difference between two uniform insonification images is

$$\begin{aligned} SSD_{1,2} =& \iint (\tilde{I}_1(\boldsymbol{u}) - \tilde{I}_2(\boldsymbol{u}))^2 d^2u \\ \simeq& \iint (\hat{I}_1(\boldsymbol{u}) - \hat{I}_2(\boldsymbol{u}))^2 d^2u \\ & + \iint \left(\frac{\eta_1(\boldsymbol{u})}{\tilde{L}_1(\boldsymbol{u})} - \frac{\eta_2(\boldsymbol{u})}{\tilde{L}_2(\boldsymbol{u})}\right)^2 d^2u \end{aligned} \tag{14}$$

The second integral in (14) is independent of the true image, and may be regarded as a constant, provided the noise is uniform.

A regularisation factor that is added to $\tilde{L}$ prevents erroneously excessive intensity in the equalised image from the speckles in low insonification regions. The computation speed is improved by calculating the convolution in the frequency domain.

Downloaded 07 Jan 2006 to 128.148.206.57. Redistribution subject to IEE licence or copyright, see http://ieedl.org/copyright.jsp

4.2 Feature detection and putative matching

Feature detection and matching are computationally demanding. A Gaussian pyramid algorithm has been proposed as a multiscale approach for efficient feature detection and matching [21, 14]. As mentioned in Section 2, 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 8.25–44.25 m, 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 initialised 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 [22]. The second moment matrix $\boldsymbol{M}$ is computed using the following relationship

$$\boldsymbol{M} = e^{-x^T x/2\sigma_s^2} \otimes ((\nabla I)(\nabla I)^T) \tag{15}$$

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

$$R = \det \boldsymbol{M} - k\mathrm{Tr}(\boldsymbol{M})^2 \tag{16}$$

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

4.3 Transformation estimation

Image changes due to the sonar system movement are modelled by an affine transformation as derived in the previous Section. 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. This 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 from an equalised 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) Transformation 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 stabilised.
(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.

In general, we can get better registration if we find and match more feature points from images. However, the structure in an acoustic camera image is not sophisticated owing to the resolution and noise. In step (i), 50 turned out to be a reasonable number of feature points that we can reliably find from most of acoustic camera images.

We use the criterion of least square error of $k\%$ of samples, where k is determined empirically. It is similar to the least-median of squares (LMS) method [23] in addition to the random sample consensus (RANSAC) algorithm [24], 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 percentile as the measure of error. It works well with a small number of feature point pairs with a 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, because the risk of mismatch is high.

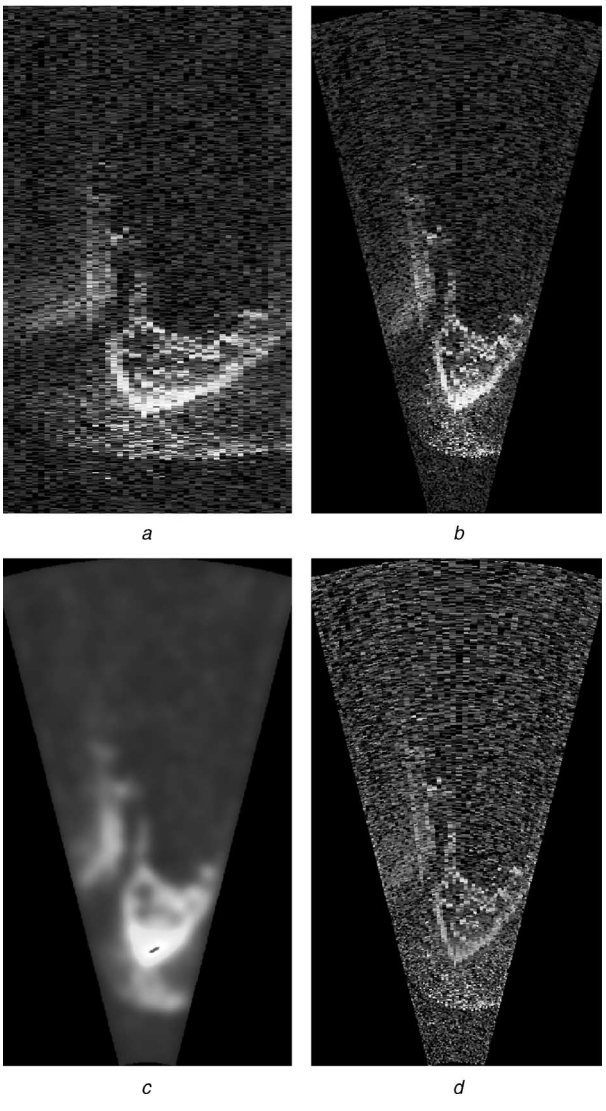


Fig. 2 *Original and transformed images, and estimated and corrected insonification images*

a An original polar coordinates image from a DIDSON system. The resolution is 48 × 512. The range coverage is 8.25 m to 44.25 m and the angle coverage is 28°
b Panel ***a*** mapped to the Cartesian coordinates. The resolution is 512 × 844
c The estimated insonification of panel *b*. This image was produced by convolving panel *b* with a 2-D isotropic Gaussian kernel with $\sigma_G = 50$ pixels. The regularisation constant was set to 15
d The estimated uniform insonification image of *b*

Downloaded 07 Jan 2006 to 128.148.206.57. Redistribution subject to IEE licence or copyright, see http://ieedl.org/copyright.jsp

Provided that there are about 40% of inlier feature point pairs, the probability that three inlier pairs are drawn is 0.48 with 10 samples, 0.9987 with 100 samples, and $1 - 10^{-27}$ with 1000 samples. The repetition time in step (iii) may vary depending on the quality of the images.

4.4 Mosaicing and resolution enhancement via image fusion

After the 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 be deteriorated because noise may accumulate. In this case, mosaicing via averaging can be described as the following relationship

$$I_{mosaic}(\boldsymbol{u}) = \frac{1}{N}\sum_{i=1}^{N} I_i(\boldsymbol{T}_i\boldsymbol{u}) \tag{17}$$

where $\boldsymbol{T}_i$ is the transformation matrix from the perspective of the mosaic image to the perspective of the ith image. The SNR of the mosaic image is

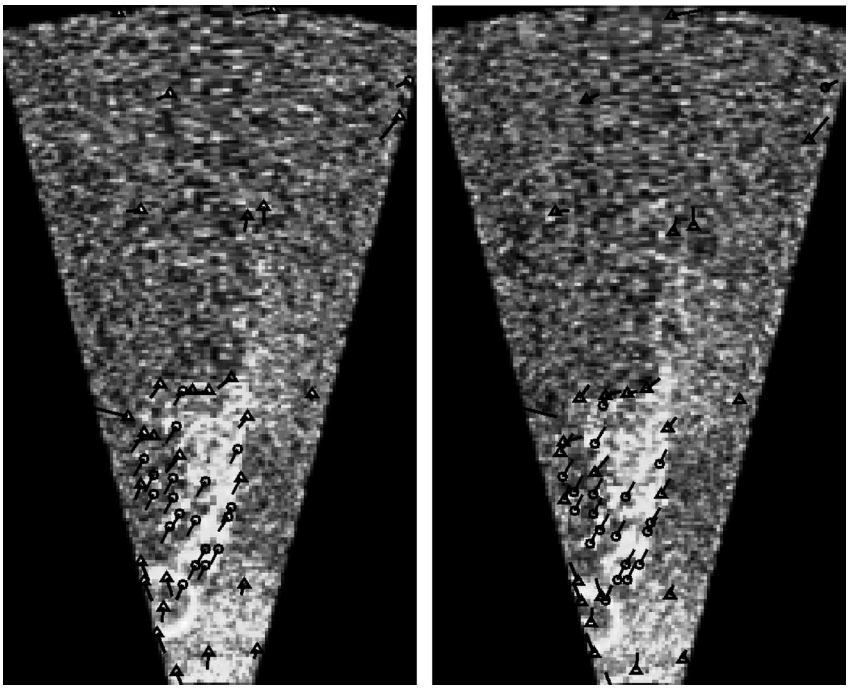


Fig. 3 *Matched (circle) and non-matched (triangle) feature points obtained from the third level of the Gaussian pyramid using the Harris corner detection*

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

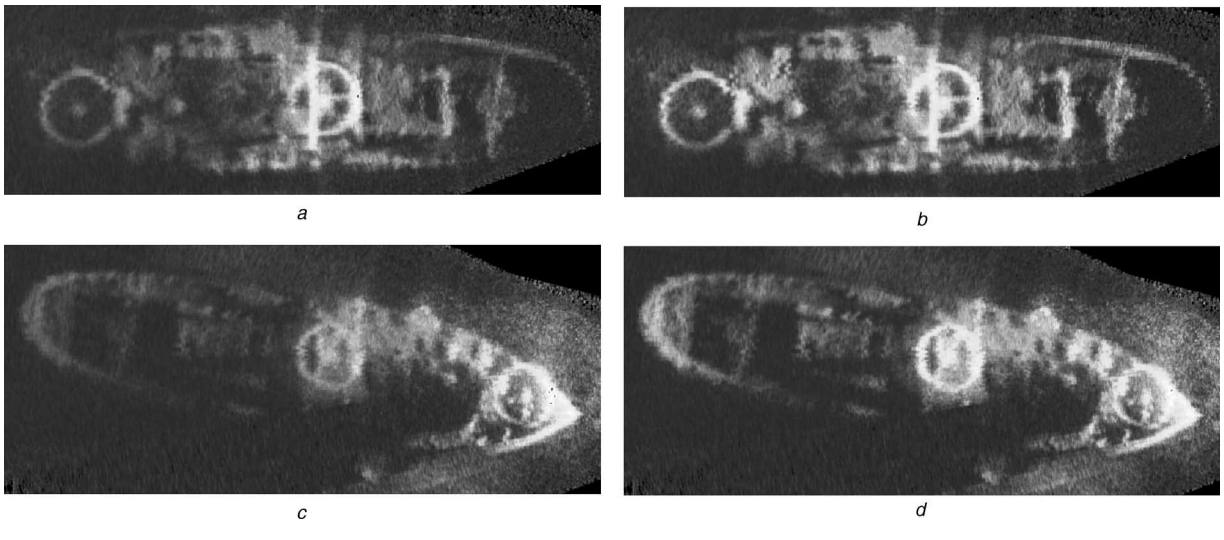
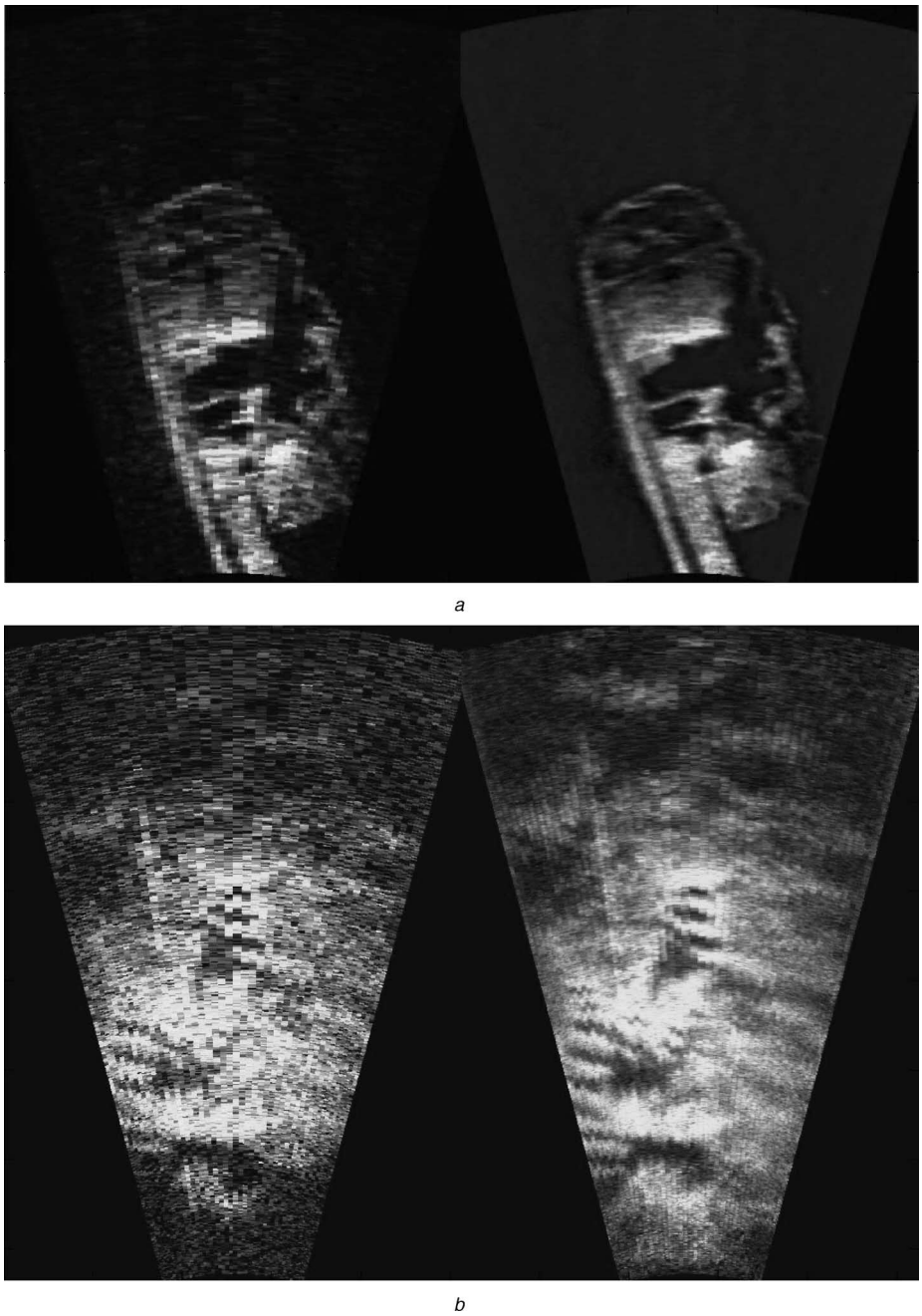


Fig. 4 *Demonstration of weighted averaging effect*

Mosaicing was performed with 38 images which were averaged after the corresponding motion compensation transformation was applied to each of them
a Uniform average of the whole images
b Weighted average, in which insonification profile was utilised during averaging (see Section 4.4 for details)
c Same target from different image sequence
d Same target from different image sequence utilising the insonification profile during averaging

Downloaded 07 Jan 2006 to 128.148.206.57. Redistribution subject to IEE licence or copyright, see http://ieedl.org/copyright.jsp



a

b

Fig. 5 *Resolution enhancement by averaging images*

a Ship wreckage image juxtaposed with a mosaic image of 5 consecutive frames followed by a geometric transformation (see Section 4.3)
b Coral image with a mosaic image of 7 consecutive frames

$$SNR_{mosaic}(\boldsymbol{u}) = \frac{\sum_i L_i(\boldsymbol{T}_i\boldsymbol{u})}{\sigma} \tag{18}$$

where $L_i(\boldsymbol{u})$ is the insonification intensity of the ith image at $\boldsymbol{u}$. Note that the SNR is a function of $\boldsymbol{u}$ because the insonification intensity varies within the image.

In our algorithm, poorly insonified regions receive lower weight in the averaging. Denote the weight of the ith image by $\alpha_i(\boldsymbol{u})$, where $\sum_i \alpha_i(\boldsymbol{T}_i\boldsymbol{u}) = 1$. Then, the mosaic image is

$$I_{mosaic}(\boldsymbol{u}) = \sum \alpha_i(\boldsymbol{T}_i\boldsymbol{u}) I_i(\boldsymbol{T}_i\boldsymbol{u}) \tag{19}$$

of which the SNR_{mosaic} is

$$SNR_{mosaic}(\boldsymbol{u}; \alpha_1, \ldots, \alpha_N) = \frac{(\sum \alpha_i L_i)}{\sigma\sqrt{\sum \alpha_i^2}} \leq \frac{\sqrt{\sum L_i^2}}{\sigma} \tag{20}$$

Equality holds when $\alpha_k(\boldsymbol{u}) = L_k(\boldsymbol{u}) / \sum L_i(\boldsymbol{T}_k^{-1}\boldsymbol{T}_i\boldsymbol{u})$. Thus, the maximum SNR of the mosaic image is achieved when the transformed images are combined as follows

$$I_{mosaic}(\boldsymbol{u}) = \frac{\sum L_i(\boldsymbol{T}_i\boldsymbol{u}) I_i(\boldsymbol{T}_i\boldsymbol{u})}{\sum L_i(\boldsymbol{T}_i\boldsymbol{u})} \tag{21}$$

This weighted averaging method reduces the influence of noise in poorly insonified regions.

5 Results

The algorithm was tested on a boat wreckage sequence [Note 1]. A DIDSON system scanned a shipwreck at

Note 1: The data has been provided by E. O. Belcher from Applied Physics Laboratory, University of Washington under ONR support.

Downloaded 07 Jan 2006 to 128.148.206.57. Redistribution subject to IEE licence or copyright, see http://ieedl.org/copyright.jsp

approximately 30 metres 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. The body of the ship exposed in each frame is less than 20% in each frame. The algorithm was applied to those two sub-sequences to build two mosaic images.

Figures 2*a* and *b* depict the same acoustic image in the original polar coordinates and the transformed Cartesian coordinates, respectively. A collection of pixels with the same pixel intensity can be seen in the Cartesian coordinates image.

Estimated insonification based on the method described in Section 4.1 is depicted in Fig. 2*c*. The insonification corrected image is depicted in Fig. 2*d*. The insonification correction, which equalises the image, increases the dynamic range of the averaged (mosaiced) image.

Figure 3 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.

Cross-correlation was found to be more robust than a conventional approach [25] 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 features is not well defined. Figure 4 represents the main result of the paper, 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. Left panels of Figs. 5*a* and *b* show the original single frames detail of the 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 multiple pixels, so that after averaging, a subpixel resolution is achieved (see right panels of Figs. 5*a* and *b*).

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. 4 on the whole target, and in particular in the comparison of two frames of the targets in Fig. 5. The top two panels of Fig. 4 depict a mosaiced image from the same sequence of acoustic images. In panel *b*, the insonification profile was utilised during averaging. Panels *c* and *d* represent the same target from a different acoustic image sequence with panel *d* utilising the insonification profile during averaging.

6 Conclusion

Acoustic camera technology is becoming essential for underwater exploration in noisy environments with low visibility. 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 modelling 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 the Harris corner detector are matched locally to the reference image via cross-correlation. This method was found to be more robust than a conventional approach in which features are found independently and matched between two images. In particular, this is more pronounced when the number of pixels available for feature comparison is limited.

7 Acknowledgments

This work was partly supported by ONR grant N00014-02-C-0296. The authors thank E. O. Belcher for providing full details about the data. Leon N. Cooper and other members of IBNS have provided valuable comments.

8 References

1 Belcher, E.O., Matsuyama, B., and Trimble, G.M.: 'Object identification with acoustic lenses'. Proc. Oceans '01 MTS/IEEE, 2001, pp. 6–11
2 Marks, R.L., Rock, S.M., and Lee, M.J.: 'Real-time video mosaicking of the ocean floor', *IEEE J. Ocean. Eng.*, 1995, **20**, (3), pp. 229–241
3 Rzhanov, Y., Linnet, L.M., and Forbes, R.: 'Underwater video mosaicing for seabed mapping'. Int. Conf. Image Process., 2000, **2**, pp. 224–227
4 Castellani, U., Fusiello, A., and Murino, V.: 'Registration of multiple acoustic range views for underwater scene reconstruction', *Comput. Vis. Image Underst.*, 2002, **87**, pp. 78–89
5 García, R., Cufí, X., and Pacheco, L.: 'Image mosaicking for estimating the motion of an underwater vehicle'. 5th IFAC Conf. on Manoeuvreing and Control of Marine Craft, 2000
6 Negahdaripour, S., Xu, X., and Khamene, A.: 'A vision system for real-time positioning, navigation and video mosaicing of sea floor imagery in the application of rovs/auvs'. IEEE Workshop on Appl. Comput. Vis., 1998, pp. 248–249
7 Capel, D., and Zisserman, A.: 'Computer vision applied to super-resolution', *IEEE Signal Process. Mag.*, 2003, **20**, (3), pp. 72–86
8 Smolic, A., and Wiegand, T.: 'High-resolution video mosaicing'. Int. Conf. on Image Process., 2001, **3**, pp. 872–875
9 Hansen, M., Anandan, P., Dana, K., Wal, G., and Burt, P.: 'Real-time scene stabilization and mosaic construction'. Proc. IEEE Workshop Appl. Comput. Vis., 1994, pp. 54–62
10 Mann, S., and Picard, R.W.: 'Virtual bellows: Constructing high quality stills from video'. Proc. IEEE Int. Conf. Image Process., 1994, pp. 363–367
11 Irani, M., Anandan, P., Bergen, J., Kumar, R., and Hsu, S.: 'Mosaic representations of video sequences and their applications', *Signal Process., Image Commun.*, 1996, **8**, (4), pp. 327–351
12 Irani, M., Hsu, S., and Anandan, P.: 'Video compression using mosaic representations', *Signal Process., Image Commun.*, 1995, **7**, pp. 529–552
13 Sawhney, H.S., and Ayer, S.: 'Compact representation of videos through dominant multiple motion estimation', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1996, **18**, (8), pp. 814–830
14 Zitová, B., and Flusser, J.: 'Image registration methods: a survey', *Image Vis. Comput.*, 2003, **21**, pp. 977–1000
15 Brown, L.G.: 'A survey of image registration techniques', *ACM Comput. Surv.*, 1992, **24**, (4), pp. 325–376
16 Belcher, E.O., Dinh, H.Q., Lynn, D.C., and Laughlin, T.J.: 'Beamforming and imaging with acoustic lenses in small, high-frequency sonars'. Proc. Oceans '99 MTS/IEEE, 1999, pp. 1495–1499
17 Wagner, R.F., Smith, S.W., Sandrik, J.M., and Lopez, H.: 'Statistics of speckle in ultrasound b-scans', *IEEE Trans. Sonics Ultrason.*, 1983, **30**, (3), pp. 156–163
18 Sijbers, J., den Dekker, J., Scheunders, P., and Van Dyck, D.: 'Maximum-likelihood estimation of rician distribution parameters', *IEEE Trans. Med. Imaging*, 1998, **18**, (3), pp. 357–361

Downloaded 07 Jan 2006 to 128.148.206.57. Redistribution subject to IEE licence or copyright, see http://ieedl.org/copyright.jsp

19 Land, E.H., and McCann, J.J.: 'Lightness and the retinex theory', *J. Opt. Soc. Am.*, 1971, **61**, pp. 1–11
20 Jobson, D.J., Rahman, Z., and Woodell, G.A.: 'Properties and performance of the center/surround retinex', *IEEE Trans. Image Process.*, 1997, **6**, pp. 451–462
21 Forsyth, D.A., and Ponce, J.: 'Computer vision: a modern approach' (Pearson Education, Inc., Upper Saddle River, NJ, 2003)
22 Harris, C.J., and Stephens, M.: 'A combined corner and edge detector'. Proc. 4th Alvey Vision Conf., 1988, pp. 189–192
23 Rousseeuw, P.J.: 'Least median of square regression', *J. Am. Stat. Assoc.*, 1984, **79**, pp. 871–880
24 Fischler, M.A., and Bolles, R.C.: 'Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography', *Commun. ACM*, 1981, **24**, (6), pp. 381–395
25 Zhang, Z., Deriche, R., Faugeras, O., and Luong, Q.-T.: 'A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry', *Artif. Intell.*, 1995, **78**, pp. 87–119

Downloaded 07 Jan 2006 to 128.148.206.57. Redistribution subject to IEE licence or copyright, see http://ieedl.org/copyright.jsp