Sonar Image Feature Detection and Matching for Acoustic Structure from Motion

Tianxiang Lin

`tianxian@andrew.cmu.edu`

1 Introduction

In underwater scenarios, optical sensors such as monocular and stereo cameras face limitations in perceiving the surrounding environment due to water turbidity, as well as the reflection and attenuation of light energy. Typically, their empirical perception range is confined to 1-2 meters [1]. However, acoustic signals, which can effectively propagate through turbid water with comparatively lower energy loss than light, empower acoustic sensors, especially imaging sonars, to overcome these drawbacks.

Feature-based Structure-from-Motion (SfM) Algorithm is widely used for state estimation from camera images. But few attempts were made to restore poses from raw sonar images, namely Acoustic Structure-from-Motion (ASfM). The challenges are as follows. Sonar images are more informative than camera images because of different sensor models. Each pixel of the one-channel sonar images contains not only the reflected acoustic intensity but also the spatial information regarding the sonar itself. Various and complex noise patterns, including speckle noise, and unexpected artifacts resulting from complicated environments and sensor itself exist as well. Meanwhile, due to the elevation ambiguity of imaging sonar, even subtle movement of the sensor results in enormous differences in underwater object contours. These factors make conventional image filtering and feature detection and matching ineffective. Also, a different triangulation should be designed.

This paper mainly focuses on the prerequisite of feature-based ASfM algorithm. To overcome the challenges brought by imaging sonars, firstly, the idea of constant false alarm rate (CFAR) algorithm is adapted from radar image processing for object segmentation and denoising for sonar images. Secondly, a self-defined loss is designed and introduced to remove the outlier matches of two sets of feature points from A-KAZE detector.

2 Problem Formation

2.1 Imaging Sonar Geometry

Consider a point $\boldsymbol{P}(\theta, r, \phi)$ in the field of view of the imaging sonar, parameterized in the local spherical sonar coordinate system as seen in Fig. 1. Here, θ, r, and ϕ represent the bearing, range, and elevation angle of the point respectively. The conversion of $\boldsymbol{P}$ to the Cartesian frame to point $\boldsymbol{C}(x, y, z)$ and vice versa is

$$\boldsymbol{C} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} r \cos\phi \cos\theta \\ r \cos\phi \sin\theta \\ r \sin\phi \end{bmatrix} \tag{1}$$

$$\boldsymbol{P} = \begin{bmatrix} \theta \\ r \\ \phi \end{bmatrix} = \begin{bmatrix} \arctan 2(y, x) \\ \sqrt{x^2 + y^2 + z^2} \\ \arctan 2(z, \sqrt{x^2 + y^2}) \end{bmatrix} \tag{2}$$

Imaging sonars generate partial spherical measurements by sending out acoustic signals into a frustum. Time of flight measured from the reflected signals observed by the transceivers provide the range r

and bearing θ, of the reflecting surface. However, these measurements are unable to disambiguate the elevation ϕ of the reflected signal. Due to this, all detected returns from a single elevation arc project onto the same pixel of a range image $I(\theta, r)$. For a pixel corresponding to a certain bearing and range in I, the pixel's intensity corresponds to the intensity of the reflected signal from all returns along the elevation.
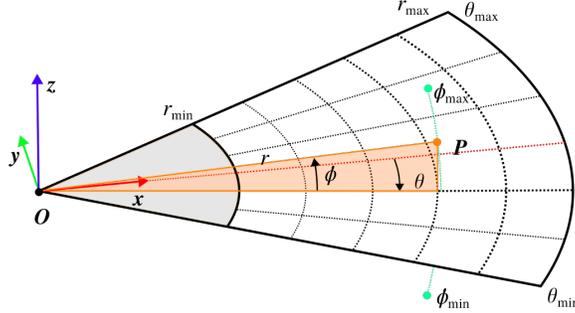

Figure 1: Geometry of a single sonar image. Point $\mathbf{P}$ is represented by range r, elevation ϕ, and bearing θ. $r_{\max}$, $r_{\min}$, $\phi_{\max}$, $\phi_{\min}$, $\theta_{\max}$, and $\theta_{\min}$ are respectively the maximum and minimum ranges, elevation angles, and azimuth angles of the imaging sonar.

Along with elevation ambiguity, shadow zones also appear when objects closer to the sonar obstruct the view of obstacles behind them. Hence, low pixel intensity values in the sonar images do not necessarily mean the absence of obstacles.

2.2 Constant False Alarm Rate

Constant False Alarm Rate (CFAR) is an adaptive technique that determines the detection threshold by assessing the interference power in the vicinity of the test cell [2], [3]. The following equation is to compute the dynamic thresholds t for each pixel:

$$t = \alpha (\sum_{(x,y) \in N} I(x,y) - \sum_{(x,y) \in G} I(x,y))/n,$$

where $I(x, y)$ is the pixel in the image I at column x and row y, N is the reference cell, G is the guard cell, n is the number of pixels in the test cell, and the multiplier α is related to the value of the probability of false alarm P_{fa}:

$$\alpha = n(P_{fa}^{-1/n} - 1).$$

2.3 Structural-and-spatial Loss

The self-designed structural-and-spatial loss (SSL) is used to eliminate the outliers during feature matching. The first part of the loss is related to the structure of the area adjacent to the feature point. Given a pair of matched feature points and a radius r, two squares with the widths of $(2r + 1)$ are cropped from the surroundings of their corresponding feature points. Then the structural similarity index (SSIM) is computed from these two squares and normalized between 0 and 1.

The second part of the loss is the normalized Euclidean distance between two spatial vectors v:

$$v = (\mu_x, \mu_y, \mu_r)^\top,$$

where μ_x is the average intensity of the feature point's column, μ_y is the average intensity of the feature point's row, and μ_r is the average intensity of the square with width $2r + 1$ where the feature point is the center. The vector is also normalized between 0 and 1.

3 Experiments

In this section, the filtering and matching results are presented for both simulation and real-world sonar images. For our simulated data, we use the HoloOcean simulator [4], [5] and for our real-world

data, we use a Bluefin Hovering Autonomous Underwater Vehicle (HAUV) [6] equipped with a SoundMetrics DIDSON 300m sonar.

The code base can be accessed through the following link: `https://drive.google.com/drive/folders/12h3520V7aijaEweapXAIyEKiRYmx4F7t?usp=sharing`.

3.1 Sonar Image Filtering

For simulation and real-world data, CFAR is computed for each pixel within the sonar image for further thresholding. The label mat is computed and then denoised by DBSCAN [7] clustering algorithm.

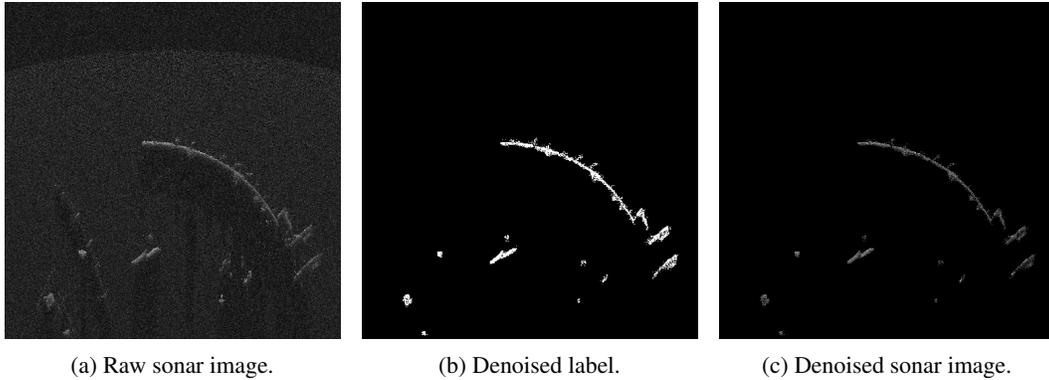(a) Raw sonar image.   (b) Denoised label.   (c) Denoised sonar image.

Figure 2: Denoising results for sonar images collected from HoloOcean Simulator.

For simulation data, a qualitative visualization result is shown in Fig. 2. From the result, while maintaining most details of objects in the scene, the proposed baseline removes all noises.

(a) Raw.   (b) Denoised label.   (c) Ground truth.   (d) Wavelet.   (e) Ours.
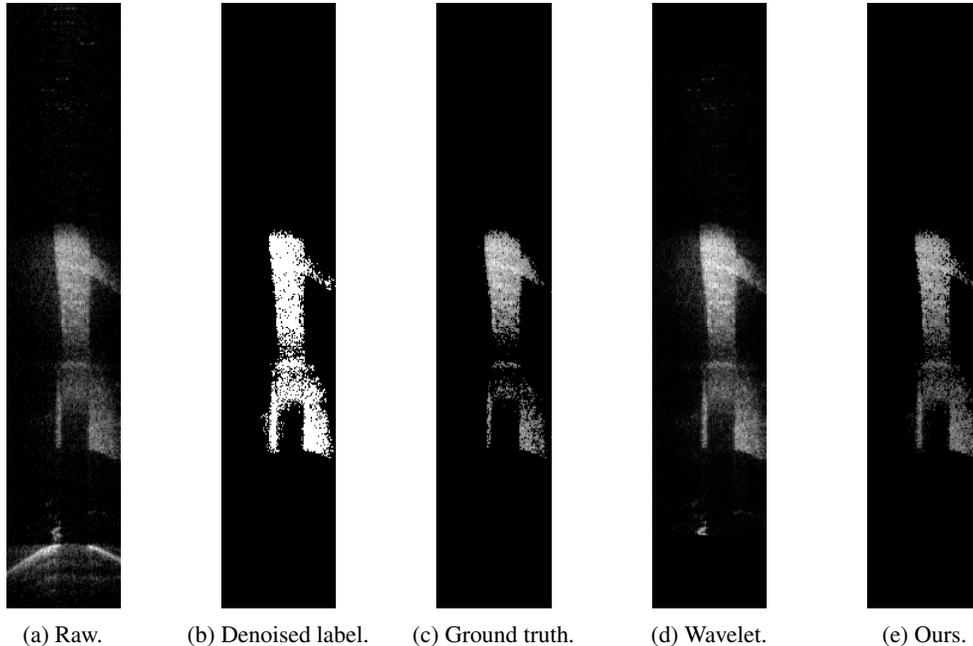
Figure 3: Denoising results for sonar images collected from DIDSON sonar.

Real-world data are also processed with the same baseline. The qualitative visualization result is shown in Fig. 3. The proposed baseline results in denoised sonar images close to ground truth images. A quantitative comparison between the proposed denoising method and wavelet transform, which is a frequently used filtering method for sonar images, is also measured by computing the average peak

3

signal-to-noise ratios (PSNR) of the sonar image sequence. Table 1 shows that the proposed method gives better filtering results because of a higher PSNR value.

Table 1: Average PSNR for filtering methods.

Filters	Raw	Wavelet	Ours
PSNRs	20.09	23.76	25.66

3.2 Feature Detection and Matching

With predicted masks from CFAR, an A-KAZE detector detects the feature points from the labeled areas and computes the descriptors altogether. After matching two sets of feature points, SSL is computed for the removal of outlier matches. Fig 4 and 5 show the matching between 2 different frames of sonar images from both simulation and real-world datasets.
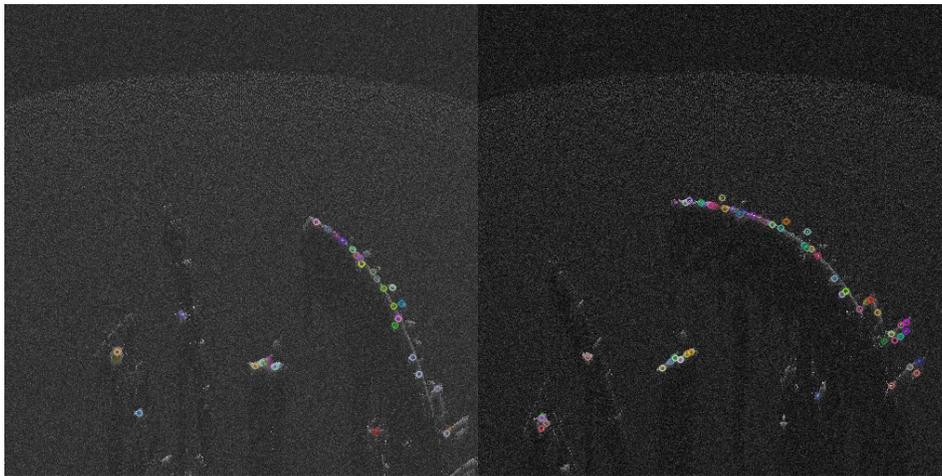
(a) A-KAZE feature points.
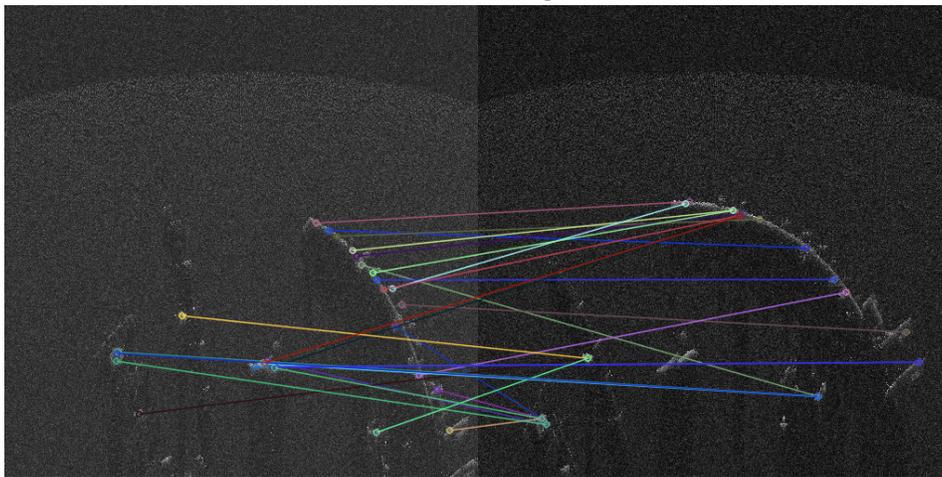
(b) Inlier matches after outlier removal.

Figure 4: Feature detection and matching result for sonar images from HoloOcean simulator.

From the results, with predicted masks, A-KAZE keypoints are located within the critical areas. With the help of the SSL loss, few or no outlier matches remain in the resulting matching result. However, due to the lack of ground truth poses of the imaging sonar, there is no prerequisite for a quantitative analysis. More data with accurate poses from real-world scenarios need to be collected soon.
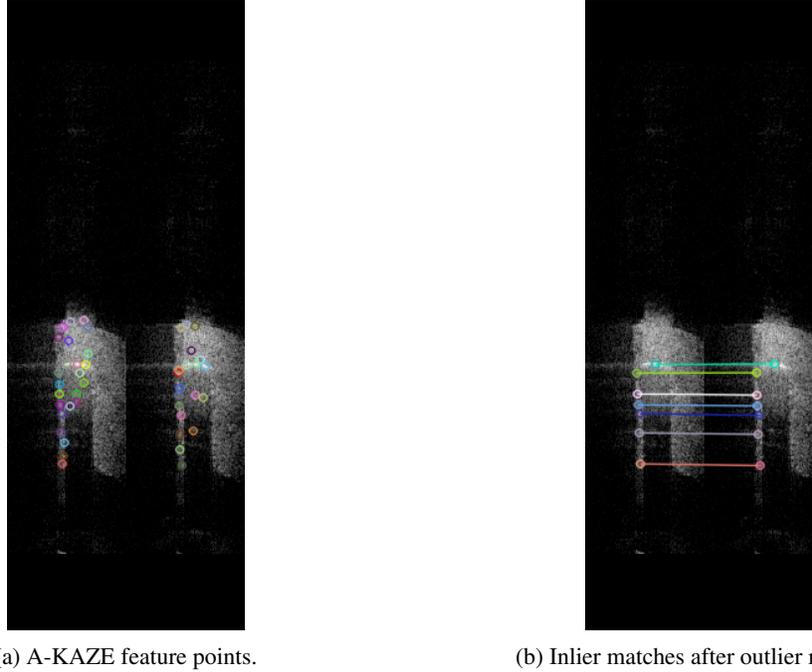
(a) A-KAZE feature points.   (b) Inlier matches after outlier removal.

Figure 5: Feature detection and matching result for sonar images collected from DIDSON sonar.

4 Conclusion and Future Work

In this paper, a novel baseline to filter noisy sonar images and outlier matching removal is presented. By adopting the CFAR algorithm from radar image denoising and designing new structural-and-spatial loss, the baseline demonstrates promising filtering and matching results.

For future directions of this work, we aim to build up an online acoustic structure-from-motion (ASfM) baseline using imaging sonar for general purposes to estimate the states of sensors and reconstruct the free spaces of the underwater environments for further applications like planning.

References

[1]   T. Lin, A. Hinduja, M. Qadri, and M. Kaess, "Conditional GANs for sonar image filtering with applications to underwater occupancy mapping," in *Proc. IEEE Intl. Conf. on Robotics and Automation (ICRA)*, London, UK, May 2023.

[2]   K. El-Darymli, P. McGuire, D. Power, and C. Moloney, "Target detection in synthetic aperture radar imagery: A state-of-the-art survey," *J. of Applied Remote Sensing*, vol. 7, no. 1, pp. 071 598–071 598, 2013.

[3]   G. G. Acosta and S. A. Villar, "Accumulated CA–CFAR process in 2-D for online object detection from sidescan sonar data," *IEEE J. of Oceanic Engineering*, vol. 40, pp. 558–569, 2015.

[4]   E. Potokar, S. Ashford, M. Kaess, and J. Mangelson, "HoloOcean: An underwater robotics simulator," in *Proc. IEEE Intl. Conf. on Robotics and Automation (ICRA)*, Philadelphia, PA, USA, May 2022, pp. 3040–3046.

[5]   E. Potokar, K. Lay, K. Norman, *et al.*, "HoloOcean: Realistic sonar simulation," in *Proc. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, Kyoto, Japan, Oct. 2022.

[6]   General Dynamics Mission Systems, *Bluefin HAUV*, `https://gdmissionsystems.com/products/underwater-vehicles/bluefin-hauv`.

[7]   M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. Intl. Conf. on Knowledge Discovery and Data Mining (KDD)*, Portland, Oregon, USA, Aug. 1996, pp. 226–231.