Segmentation, Classification and Modeling of Two-Dimensional Forward-Scan Sonar Imagery for Efficient Coding and Synthesis

Mohammad Haghighat*, Xiuying Li*, Zicheng Fang*, Yang Zhang† and Shahriar Negahdaripour*
*Department of Electrical and Computer Engineering, University of Miami, Coral Gables, FL, USA
†Department of Ocean Engineering, Ocean University of China, Qingdao, China

***Abstract*—In this paper, we present methods for segmenting noisy two-dimensional forward-scan sonar images and classify and model their background. The segmentation approach differentiates the highlight blobs, cast shadows, and the background of sonar images. There is usually little information within relatively large background regions corresponding to the flat sea bottom and (or) water column, as they are often corrupted with speckle noise. Our experiments show that the background texture is dominated by the speckle noise which has the appearance of a pseudo-random texture. We show that the background texture of the underwater sonar images can be categorized by a small number of classes. The statistical features work better than the texture-based features in categorizing the pseudo-random background, which further strengthen our hypothesis of the dominance of noise over the background texture. As a result, we can model the noisy background with a few parameters. This has an application in coding the sonar images in which highlight blob regions and cast shadows are coded at the encoder side while the speckle noise-corrupted background can be synthesized at the decoder side. Since the background regions occupy a large fraction of the FS sonar image, we expect higher compression rates than most current image or video coding standards and other custom-designed sonar image compression techniques.**

***Index Terms*—forward-scan sonar imagery, sonar image segmentation, sonar background classification, speckle noise modeling and synthesis.**

I. Introduction

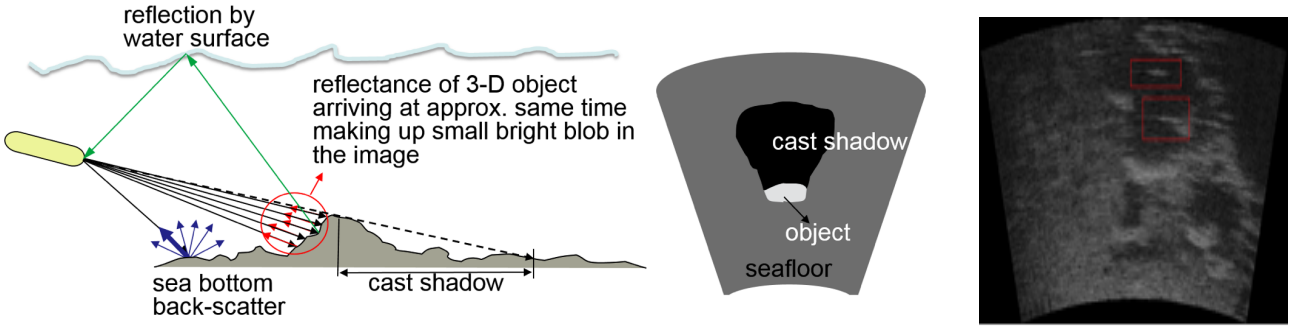
Acoustic signals can penetrate through silt and other sources of turbidity that prohibit the deployment of optical systems, the most common imaging modality in the terrestrial domain. This has motivated the development and improvement of high-frequency 2-D forward-scan (FS) video sonar systems over the past decade, to meet the critical need of scene imaging at improved resolution under poor visibility [1]. Automated processing of FS sonar video imagery enables significant capabilities for a wide variety of underwater task and operations, *e.g.*, fish stock assessment, seafloor and habitat mapping, and the inspection of pipelines and other structures.

Automated sonar image processing is rather complex due to presence of and interactions among visual cues and artifacts. To elaborate, we first note that FS sonar systems are typically deployed at large grazing angles (relative to the sea bottom) in order to 1) image a larger region of the sea floor within a single image; 2) improve image quality and contrast by increasing the diffuse backscatter returns relative to specular reflections. Consequently, referring to the cartoon drawing in Fig. 1(a), the 3-D targets on the sea bottom can often be detected by two visual cues with distinct characteristics: 1) A thin but horizontally elongated bright image blob, generated by the backscattered signal from visible object surfaces; 2) shadows cast by the object on neighboring background surfaces. In addition to target detection, the cast shadows provide useful visual cues for 3-D object shape reconstruction and sonar motion estimation [1], [2]. Additionally, artifacts within these regions can arise due to multiple reflections.

As depicted in Fig. 1(a), the object is also insonified indirectly by the acoustic waves that are reflected from the sea floor and the sea surface when operating within shallow waters. The fraction of this indirect incident energy – reflected by the visible object surfaces towards the sonar receiver along various beams – travels longer distances than those due to the direct insonification. The multipath component due to ground reflection generally distorts the object highlight, while component due to surface reflection often appears as bright streaks within the shadow regions. Additionally, the multi-path components can be generated by strong nearby reflectors, *e.g.*, metallic objects. Due to unknown number, location and pose of scene objects, these distortions of object highlights and the cast show due to multi-path components are generally unpredictable. Fig. 1(b) depicts a sample FS sonar image captured in very shallow water, where the highlights regions generated by the surface reflections are identified by the red squares.

For the scene interpretation from FS sonar image, informative image regions of interest (ROI) comprise of the object blobs, cast shadows, and highlights from multi-path reflections. Thus, these regions, offering useful visual cues about shape, positions and sizes of various objects, distance from the sea surface, etc., may be treated as the signal components.

The treatment of speckle noise due to coherent interference of acoustic waves is one of the serious complexities in automated sonar image processing. The speckle noise often has the appearance of, and may be indistinguishable from pseudo-random texture of the background scene surfaces. This can be noted within the background regions of three sample FS sonar images in Fig. 2, recorded in a lake (a,b) and a marina (c). In some cases, the speckle noise may also distort the signal component, overshadowing certain visual cues for image interpretation. There is generally little information



(a) (b)

Fig. 1. (a) Schematic diagram of FS sonar imaging. (b) Highlights within object shadows generated by water-surface reflection.

within relatively large background regions corresponding to the flat sea bottom and (or) water column, particularly when corrupted with speckle noise. Thus, it may be desirable to discriminate between the signal component containing critical visual cues and the background, prior to scene interpretation through pre-processing. The varying composition of the noise-corrupted background regions (e.g., water column, soft to hard bottoms) leads to different texture characteristics, as shown in Fig. 3. In order to reconstruct a sonar image, it is important to classify and model these noisy background areas.

This study is aimed at devising a solution for effective compact representation of a sonar image by: 1) segmenting the signal (ROI) from the noisy background; 2) deriving an efficient representation of the latter based on the speckle noise and its statistics. To elaborate, we extract the foreground ROI (highlight and shadow regions) and model the noisy background. While our goals is solely the segmentation, classification and modeling, one should take note of some key applications: efficient FS sonar video coding and synthesis, as well as robot localization based of ROI and background classification.

For the coding application, highlight blob regions and cast shadows are coded at the encoder side while the speckle noise-corrupted background is synthesized at the decoder side. Because the background regions occupy a large fraction of the FS sonar image, we expect higher compression rates than most current image or video coding standards and other custom-designed sonar image compression techniques that do not fully exploit the unique texture characteristics in sonar imagery [3]–[6]. For operations involving autonomous underwater vehicles (AUVs), the reduced bit rate for video transmission could enable transmission of real-time FS sonar video through underwater acoustic channels. Moreover, for image/video synthesis applications, background classification and modeling enable improving the subjective visual appearance based on ray-casting [7], [8]. Moreover, this work can motivate more applications for the realization of key robotics capabilities in turbid waters. For certain applications, high volume of watermark information can be incorporated within the noisy background region, enabling integrity check/verification of the decoded data.

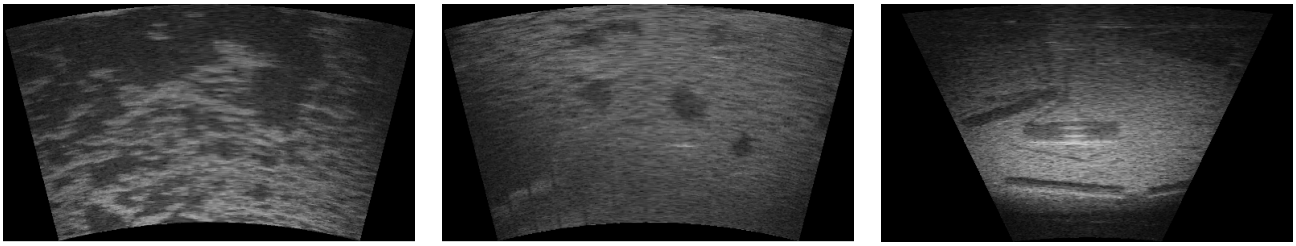
In this paper, we apply the k-means segmentation technique to differentiate the highlight blobs, cast shadows, and the background. We have noted that the texture segmentation techniques are not effective for this purpose. The reason is mostly due to the dominance of the noise over the available texture features. However, since the three major regions differ in the average intensity, intensity-based multi-level thresholding has proven to be effective.

Our experiments show that the background texture of the underwater sonar images, some examples of which are shown in Fig. 3, can be categorized by a small number of classes. Here, we use an unsupervised technique to cluster all the background images in our training data into different background classes. Then, using these classes, we train a supervised system to label the class of the background in the test images. This will reduce the complexity of the background modeling and number of parameters to represent them. Our experiments showed that statistical features were discriminating the different background classes much better than the well-known texture-based features. These results strengthen our hypothesis of the dominance of noise over texture features in the background regions. Comprehensive experiments are conducted by comparing our scheme with other sonar texture synthesis methods [9]–[11].

The rest of the paper is organized as follows: Section II describes the methods used for pre-processing and segmentation of FS sonar images. Section III presents the feature extraction and clustering method used for background classification. Background modeling and synthesis methods are explained in Section IV. The implementation details and experimental results are presented in Section V, and finally, Section VI concludes the paper.

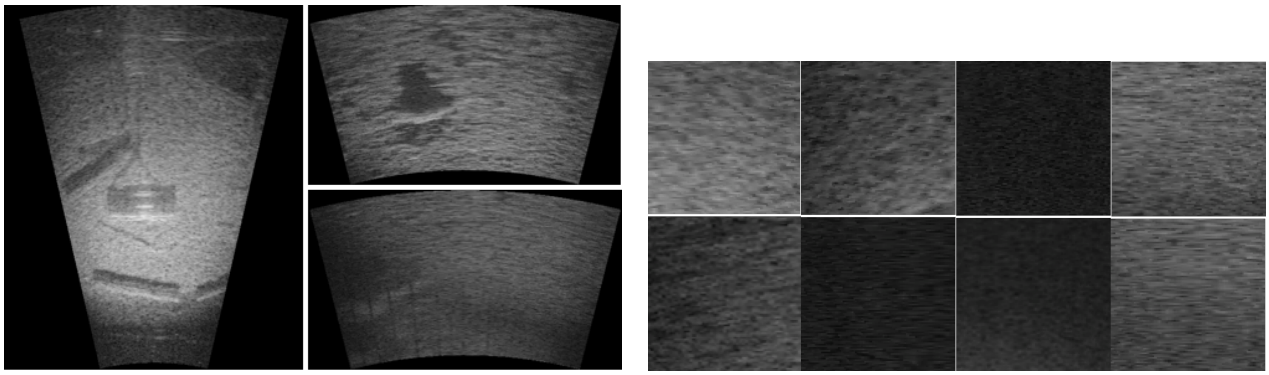
II. Pre-processing and Segmentation

In FS sonar images, the three object, shadow and background regions often have significantly different intensity levels. Hence, we may use the k-means clustering technique to differentiate among them. However, the segmentation result is only roughly accurate. In particular, a sonar image is generally



(a) (b) (c)

Fig. 2. Sample FS sonar images captured in a lake (a, b) and a marina (c).



(a) (b)

Fig. 3. (a) Three sample FS sonar images with different background types. (b) Background from these and other types of images.

corrupted by certain patterns due to the imprefections of sonar hardware. For example, the central region of the image in Fig. 3(a) is brighter within the near-field central than within the margin and far field. An effective pre-processing technique to estimate the stationary FS sonar image pattern is to compute the average over a large number of images with varying scene content. Fig. 4(a), (b) and (c) show a sample image, the stationary FS sonar image pattern, and the image after illumination normalization, respectively.

Another factor that affects the segmentation results is the speckle noise. To overcome noise issue, it is generally sufficient to average a small number of consecutive frames, only two in this work. As it will be discussed in the next section, in order to have better results, we apply the averaging in the log domain in which the noise is additive. Finally, we can apply a median filter as the final step after returning the image to the spatial domain. The result of normalized image after noise reduction is depicted in Fig. 4(d), which can be compared with the original frame in Fig. 4(a).

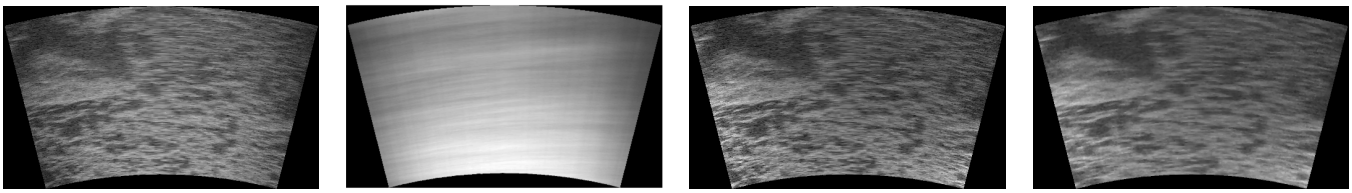
Applying the k-means clustering method to the pre-proccessed FS sonar image yields the result given in Fig. 5(a). As stated, this is accurate only roughly due to misclassification of isolated outlier pixels. The result is improved by applying consistency verification with a majority filter [12]. If the majority of pixels surrounding a point are from a different cluster, we change the label of the point to that of the majority of the neighboring points. Figs. 5(b-d) show the result of consistency verification using majority filter of size 3×3 once and twice, and size 5×5 once, respectively. In these images, gray, black and white regions represent the object, shadow and background regions, respectively. In experiments with sonar images of varying types, applying the 3×3 window twice has proven to produce consistent segmentation results.

III. Classification of FS Sonar Background: Statistical vs Texture-Based Features

After segmentation, we can extract samples from background region. We categorize these samples into several groups by k-means clustering based on different feature types. Some examples of popular texture-based feature types are Gabor wavelet features [13], [14], Gray Level Co-occurrence Matrix (GLCM) features [15], and first-order statistical features.

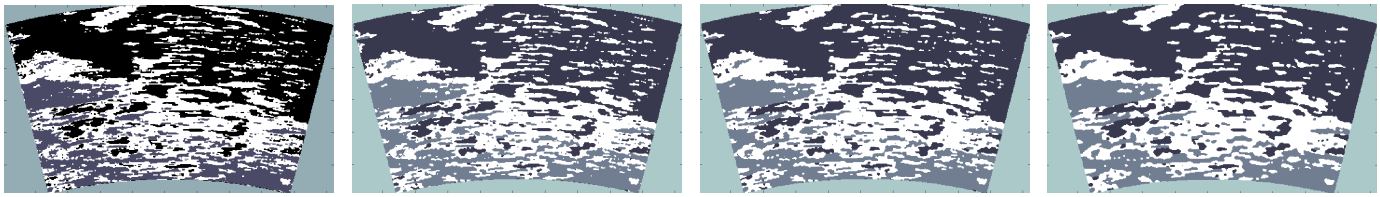
We tried the above-mentioned features individually and also their fusion. In this work, we use silhouette plots to compare the effectiveness of clustering results using different features [16]. For each sample, the silhouette values, ranging from -1 to 1, shows how well this sample matches its cluster [17]. A good clustering solution results in high silhouette values for the samples. Negative silhouette values, however, indicate that the clusters are not well separated. The silhouette plots of our k-means classification results using different sets of features are shown in Fig. 6(a-c).

Based the silhouette plots, the clustering result of the first-order statistical features has highest silhouette values and fewest negative values. This suggests that the first-order statistical features perform better than Gabor and GLCM features



(a) (b) (c) (d)

Fig. 4. (a) A sample sonar image. (b) The stationary FS sonar image pattern. (c) Image after illumination normalization. (d) Image after normalization and noise reduction.



(a) (b) (c) (d)

Fig. 5. (a) Result of k-means segmentation. (b-d) Result after consistency verification using majority filter of size 3×3 once (b), 3×3 twice (c), 5×5 once (d).

for the k-means clustering of the sonar backgrounds. The reason might be the lack of any particular shape or texel elements in the the sonar background samples. The first-order statistical features which work well include the mean, variance, skewness, third and fourth central momentum of the samples. Therefore, if we model the background as a uniform region corrupted with noise, we can use the first-order statistical features to represent each background cluster.

The result shown in Fig. 6(c) shows that even using the statistical features, some negative silhouette values still exist, and two of the clusters contain only few samples. Since the number of the samples in the two smallest clusters and those with negative silhouette values is very low, we consider them as outliers. In order to improve the classification accuracy, we remove these outlier samples, and apply the k-means for four clusters. For the final classification result, the silhouette plot is shown in Fig. 6(d).

As mentioned above, we use the first-order statistical features and the average feature vector of each cluster is used to represent that cluster. At the testing time, having a query image, for every pixel in the background region, we extract the statistical features in a neighborhood (window) around that pixel and classify the region using a simple minimum distance classifier. Each pixel in the background area is labeled with a background class using a sliding window, which sweeps through the whole background area. As a result, we determine the background class of each region. Please note that the consistency verification is also used here to enhance the results.

IV. Background Modeling and Synthesis

In the previous section, we concluded that the statistical features are most suitable in describing the background classes. K-mean clustering showed that the background samples extracted from the training images can be categorized into four classes. We represent each of these classes with the average feature vector of all samples in the class. However, since the statistical features extracted from the background were dominated by the speckle noise, these features actually represent the noise within these regions.

A. Modeling the Noise Distribution

it is known that the image with speckle noise can be represented by a multiplicative model:

$$\tilde{S} = S \times N \,. \tag{1}$$

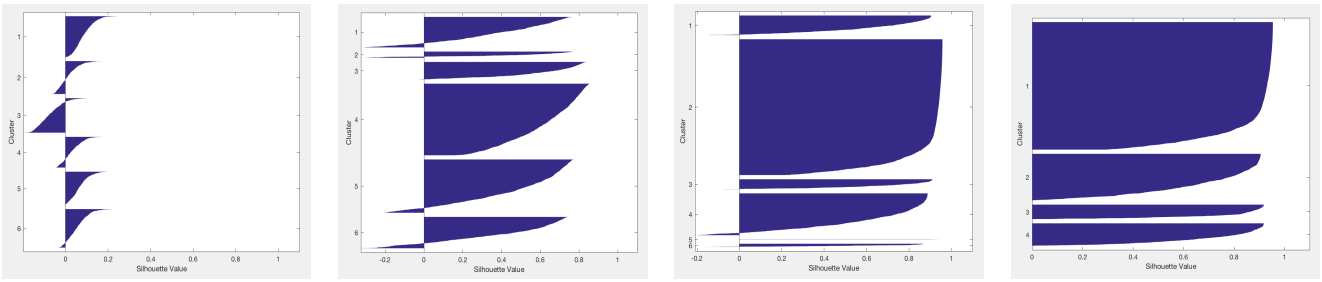
where $\tilde{S}$ denotes the noisy signal, and the signal (S) is the perfect image without the multiplicative noise (N). Taking the logarithm of the above equation, we have:

$$log(\tilde{S}) = log(S) + log(N) \,. \tag{2}$$

where the noise impact is now additive. We may assume that an estimation of the noiseless image can be obtained by averaging the consecutive frames in the log domain as described in Section II. We can achieve the distribution of the noise in the log domain by simply subtracting the logarithm of the noiseless signal (S) from the logarithm of the noisy signal. Our experimental results on different estimation models show that the distribution of the noise in log domain resembles a Gaussian (normal) distribution. This means that the noise in the image spatial domain has a *lognormal* distribution[1].

We can readily calculate the mean (m) and the variance (ν) of the noise distribution in the log domain. The mean (μ) and

[1]The lognormal distribution is a probability distribution whose logarithm has a normal distribution



(a) (b) (c) (d)

Fig. 6. Silhouette plot of background classification in six clusters using (a) Gabor features, (b) GLCM features, and (c) the first-order statistical features. (d) Silhouette plot of revised samples in four clusters using the statistical features.

the standard deviation (σ) of the distribution can be calculated from the corresponding values for the Gaussian distribution:

$$\mu = log(\frac{m}{\sqrt{1+\nu/m^2}}) \tag{3}$$

$$\sigma = \sqrt{log(1+\nu/m^2)} \tag{4}$$

Likewise, the mean m and the variance ν of the lognormal random variable are functions of μ and σ:

$$m = e^{\mu+\frac{\sigma^2}{2}} \tag{5}$$

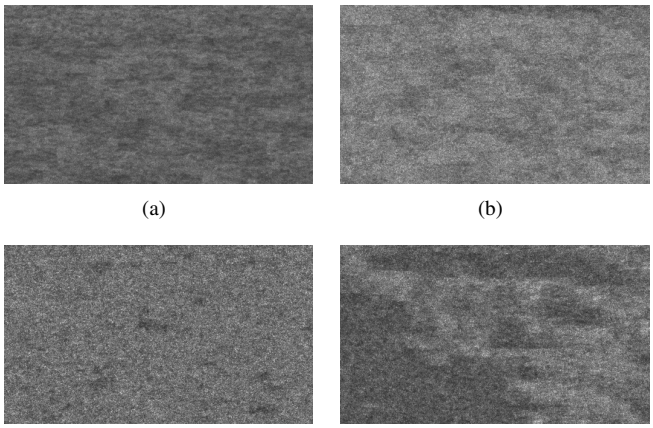
$$\nu = e^{2\mu+\sigma^2}(e^{\sigma^2}-1) \tag{6}$$

No that we have established the background classes, we can transform all the samples of each class to the log domain and calculate the average mean and average standard deviation of each class over all its samples. This will present the noise parameters of each class, which can later be used in the process of synthesizing the background classes.

B. Background Synthesis

In order to synthesize the background image of each class we use an image quilting [18] technique to create the background texture and then add its corresponding speckle noise. Image quilting is a fast and very simple texture synthesis algorithms that generates visual appearance in which a new image is synthesized by stitching together small patches of existing texture images. It takes a sample of texture and generate an unlimited amount of image data which, while not exactly like the original, will be perceived by humans to be the same texture.

Relying on psychophysical and computational models of human texture discrimination, it is shown that two texture images will be perceived by human observers to be the same if some appropriate statistics of these images match [19]–[21]. If we simply tile the patches randomly taken from the input texture to create the synthesized background, the resulting image will suffer from blocking artifact. To address this problem, the image quilting algorithm [18] lets the patches have ragged edges and allows having overlaps in the placement of patches onto the new image. Before placing a selected patch into the texture, we calculate the discrepancy in the overlap region between the patch and other patches. The minimum cost path through that discrepancy surface is chosen to be the boundary of the new patch.



(a) (b) (c) (d)

Fig. 7. Sample synthesized background images. (a) Class 1. (b) Class 2. (c) Class 3. (d) Class 4.

In order to synthesize the background of a each class, we feed the image quilting algorithm with the texture patches cropped from random background samples of that class. Finally, we generate the speckle lognormal noise sample corresponding to the class, as a multiplicative field for the synthesized texture image. Fig. 7 shows sample synthesized background images for the four background classes defined on our sonar data.

V. Experiments and Analysis

In this section, we present the results of applying the method described in the previous sections to segment, classify, and synthesize a sonar image. First, the pre-processing and segmentation algorithm described in Section II is applied on input FS sonar image to differentiate among the highlight blobs, cast shadows, and the background regions. Then, the sliding window technique described in Section III is applied on the background regions to label the background classes. A binary mask is created for each class which is used to crop the synthesized background region and put together the new synthesized FS sonar image. It is noted that the highlight and

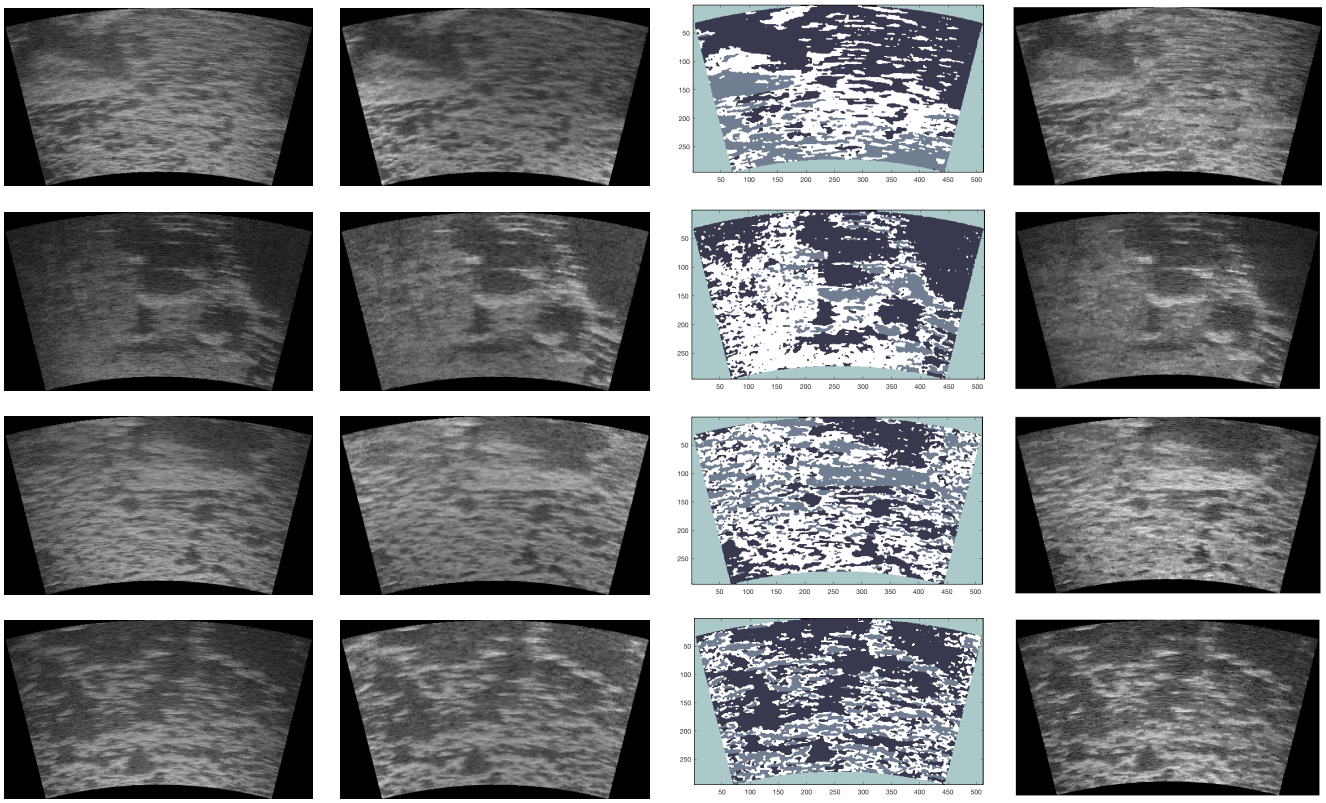


Fig. 8. Results of different steps on sonar images collected at University of Miami. Column 1: original images. Column 2: pre-processed images. Column 3: segmentation results. Column 4: images with synthesized background.

cast shadow regions are kept intact and we only synthesize the background region.

Fig. 8 and Fig. 9 show the above-mentioned steps on several sample FS sonar images. The samples in Fig. 8 are from the dataset collected at University of Miami, while Fig. 9 shows samples from publicly available Sound Metrics dataset available from [22]. The results show that our synthesized images are very similar to the input images subjectively. Since the background regions often occupy a large fraction of the FS sonar image, and we can represent them with just a small number of features, we expect our method to have applications in sonar image compression.

VI. Conclusions

In this paper, we proposed methods for segmentation , classification, and modeling of noisy two-dimensional forward-scan sonar images. We showed that the background region of the sonar images can be represented by a small number of parameters and it can be synthesized with no need to keep all the information about it. This is very important for the coding application in which the highlight blob regions and cast shadows are coded at the encoder side while the speckle noise-corrupted background is synthesized at the decoder side. Because the background regions occupy a large fraction of the FS sonar image, we expect higher compression rates than most current image or video coding standards and other custom-designed sonar image compression techniques. In future, we will investigate the results of the compression using the current technique.

References

[1] S. Negahdaripour, "On 3-D scene interpretation from FS sonar imagery," in *MTS/IEEE Oceans'12*, 2012, pp. 1–9.

[2] S. Negahdaripour, "On 3-D motion estimation from feature tracks in 2-D FS sonar video," *IEEE Transactions on Robotics*, vol. 29, no. 4, pp. 1016–1030, 2013.

[3] X. Wen, W. Yuling, and Z. Weiqing, "Sonar image processing system for an autonomous underwater vehicle (AUV)," in *MTS/IEEE Oceans'95*, vol. 3, 1995, pp. 1883–1886.

[4] J. Impagliazzo, W. Greene, and Q. Q. Huynh, "Wavelet image compression algorithm for side-scan sonar and teleradiology," in *SPIE's 1995 Symposium on OE/Aerospace Sensing and Dual Use Photonics*. International Society for Optics and Photonics, 1995, pp. 162–172.

[5] R. Cunha, M. Figueiredo, and C. Silvestre, "Simultaneous compression and denoising of side scan sonar images using the discrete wavelet transform," in *MTS/IEEE Oceans'00*, 2000, pp. 195–199.

[6] T. Higdon, "The compression of synthetic aperture sonar images," 2008.

[7] J.-H. Gu, H.-G. Joe, and S.-C. Yu, "Development of image sonar simulator for underwater object recognition," in *MTS/IEEE Oceans'13*, 2013, pp. 1–6.

[8] J. M. Bell, "Application of optical ray tracing techniques to the simulation of sonar images," *Optical Engineering*, vol. 36, no. 6, pp. 1806–1813, 1997.

[9] G. R. Elston and J. M. Bell, "Pseudospectral time-domain modeling of non-rayleigh reverberation: synthesis and statistical analysis of a side-scan sonar image of sand ripples," *IEEE Journal of Oceanic Engineering*, vol. 29, no. 2, pp. 317–329, 2004.

[10] J. Tegowski and A. Zielinski, "Synthesis and wavelet analysis of side-scan sonar sea bottom imagery," *Hydroacoustics*, vol. 9, pp. 199–208, 2006.

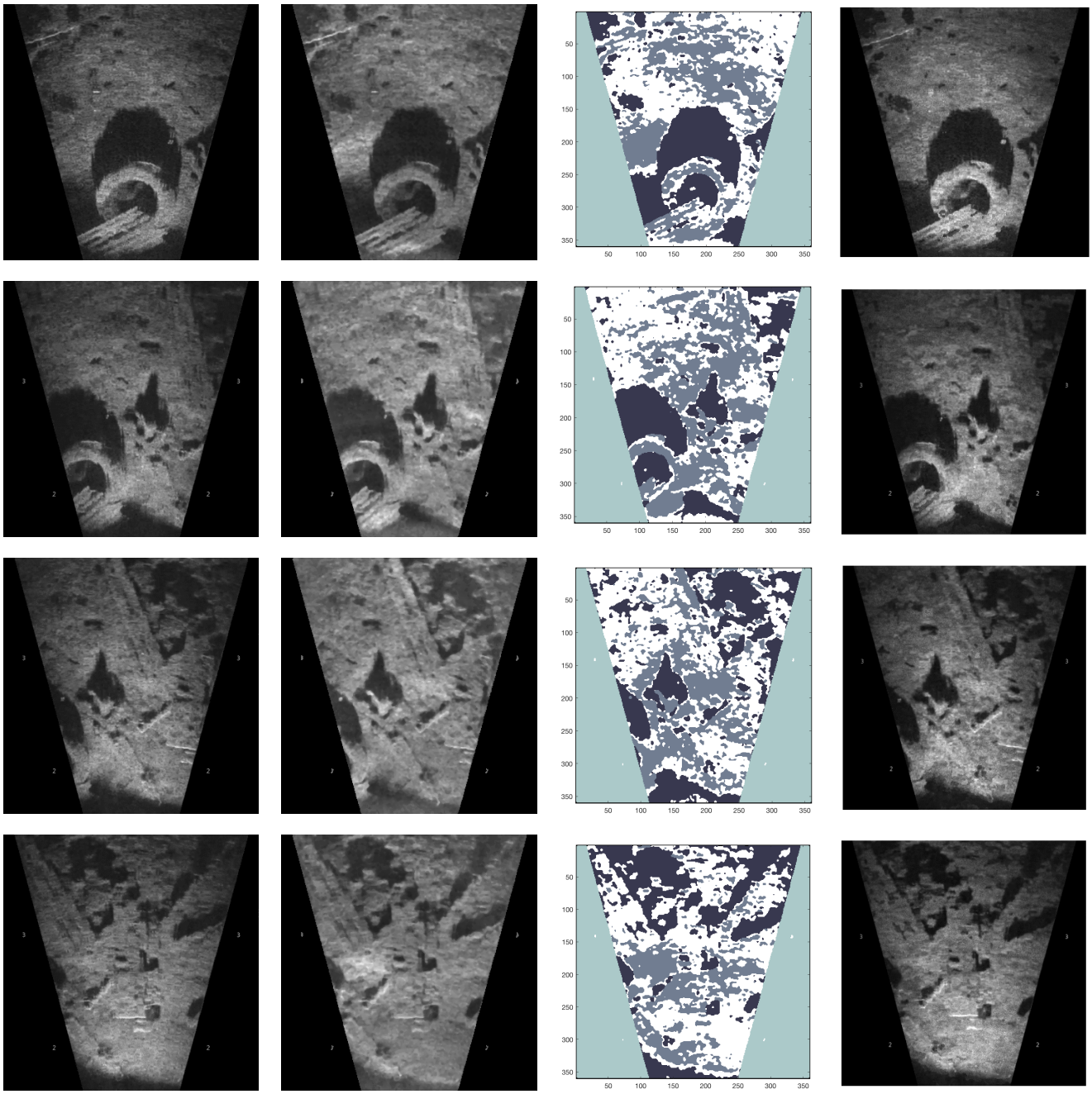


Fig. 9. Results of different steps on Barrel-Roll FS sonar images [22]. Column 1: original images. Column 2: pre-processed images. Column 3: segmentation results. Column 4: images with synthesized background.

[11] P. Blondel and O. G. Sichi, "Textural analyses of multibeam sonar imagery from stanton banks, northern ireland continental shelf," *Applied Acoustics*, vol. 70, no. 10, pp. 1288–1297, 2009.

[12] M. Haghighat, A. Aghagolzadeha, and H. Seyedarabia, "Multi-focus image fusion for visual sensor networks in dct domain," *Computers and Electrical Engineering*, vol. 37, no. 5, pp. 789–797, 2011.

[13] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Transactions on Image processing*, vol. 11, no. 4, pp. 467–476, 2002.

[14] M. Haghighat, S. Zonouz, and M. Abdel-Mottaleb, "CloudID: trustworthy cloud-based and cross-enterprise biometric identification," *Expert Systems with Applications*, vol. 42, no. 21, pp. 7905–7916, 2015.

[15] R. M. Haralick, K. Shanmugam *et al.*, "Textural features for image classification," *IEEE Transactions on systems, man, and cybernetics*, no. 6, pp. 610–621, 1973.

[16] L. Kaufman and P. J. Rousseeuw, *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons, 2009, vol. 344.

[17] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987.

[18] A. A. Efros and W. T. Freeman, "Image quilting for texture synthesis and transfer," in *Proc. of the 28th Annual Conference on Computer Graphics and Interactive Techniques*. ACM, 2001, pp. 341–346.

[19] B. Julesz, "Visual pattern discrimination," *IRE Transactions on Information Theory*, vol. 8, no. 2, pp. 84–92, 1962.

[20] J. R. Bergen and E. H. Adelson, "Early vision and texture perception,"

Nature, vol. 333, no. 6171, pp. 363–364, 1988.
[21] J. Malik and P. Perona, "Preattentive texture discrimination with early vision mechanisms," *Journal of the Optical Society of America A*, vol. 7, no. 5, pp. 923–932, 1990.
[22] "Sound metrics sonars," accessed: 2016-07-18. [Online]. Available: http://soundmetrics.com/Image-Gallery