PennState
Applied Research Laboratory

Alternative Representations of Information for Acoustics (ARIA)

Final Report

J. Daniel Park

Technical Report
File No.: 22-006
28 October 2022

Applied Research Laboratory
P.O. Box 30
State College, PA 16804-0030

Sponsored by: U.S. Office of Naval Research
Grant No.: N00014-19-1-2221

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
28-10-2022	Final	March 01, 2019 - August 31, 2022

4. TITLE AND SUBTITLE
Alternative Representation of Information for Acoustics

5a. CONTRACT NUMBER

5b. GRANT NUMBER
N00014-19-1-2221

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Joonho Daniel Park

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The Pennsylvania State University Applied Research Labotatory
Office of Sponsored Programs
110 Technology Center Building
University Park, PA 16802-7000

8. PERFORMING ORGANIZATION REPORT NUMBER
TR-22-006

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
875 North Randolph Street
Arlington, VA 22203-1995

10. SPONSOR/MONITOR'S ACRONYM(S)
ONR

11. SPONSORING/MONITORING AGENCY REPORT NUMBER

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for Public Release; Distribution is Unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Classification of objects based on acoustic or electromagnetic measurements is performed in many remote sensing applications underwater and in-air. Active sonar insonification induces multiple types of acoustic scattering phenomena including direct reflection as well as structural resonance. It is observed that the choice of representation for the raw measurement affects the shape and strength of discriminatory features and the classification performance that utilize them. Using in-air acoustic measurements collected in a noise-controlled laboratory setting, this work develops a statistical model for discriminatory features and a framework to identify the discriminatory pixels in multiple representations as well as an approach to quantify their discriminatory power in the presence of additive noise of varying levels. This framework is used to compare the relative classification performance bounds under independent pixels assumption as well as conventional feature energy detector.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE	SAR		Joonho Park
U	U	U			19b. TELEPONE NUMBER (*Include area code*) (814) 865-7507

Applied Research Laboratory
State College, PA 16804

Technical Report
22-006
28 October 2022

Sponsored by: U.S. Office of Naval Research

Alternative Representations of Information for Acoustics (ARIA)

Final Report

J. Daniel Park

Abstract

Classification of objects based on acoustic or electromagnetic measurements is performed in many remote sensing applications underwater and in-air. Active sonar insonification induces multiple types of acoustic scattering phenomena including direct reflection as well as structural resonance. It is observed that the choice of representation for the raw measurement affects the shape and strength of discriminatory features and the classification performance that utilize them. Using in-air acoustic measurements collected in a noise-controlled laboratory setting, this work develops a statistical model for discriminatory features and a framework to identify the discriminatory pixels in multiple representations as well as an approach to quantify their discriminatory power in the presence of additive noise of varying levels. This framework is used to compare the relative classification performance bounds under independent pixels assumption as well as conventional feature energy detector.

Table of Contents

List of Figures

List of Tables

1 Introduction

remote sensing uses sound as the primary sensing modality due to its efficient propagation compared to light or other electromagnetic counterparts that are used in terrestrial environments. Both sonar (sound navigation and ranging) and radar (radio detection and ranging) utilize wave phenomenology for object detection, classification, and localization. Acoustic data collected with sonar systems are processed and represented in various forms of data products, such as spatial imagery or spectrograms that extract and organize the information embedded in the raw data, which are appropriate for human consumption or for further analysis and decision-making using automated target recognition (ATR) algorithms.

Model-based signal processing [1] in remote sensing applications achieves high quality data products by utilizing signal structures associated with the known acoustic phenomena, sound interactions with the environment, and its propagation through the medium. For example, synthetic aperture sonar (SAS) processing reconstructs spatial imagery of the scattered intensity of the imaging scene with isotropic scatterer assumptions [2,3]. The optics-like behavior of high frequency sound (above 100kHz) is exploited to achieve high-contrast picture-like imagery. However, when the signal structure is different from the expected, or when systematic errors are introduced the quality of image degrades [4], which renders the information extracted from it less reliable.

Alternative representations with more appropriate assumptions on the signal structure may be considered in order to improve the quality of data product or to extract additional information not accessible with conventional processing [5,6], at the expense of increased computational cost and potentially transforming the raw data into a less familiar domain such as Fourier spectrum or spatial wavenumber domain [7,8]. However, some of the expected benefits are more interpretable organization of features and improved robustness against noise. Noise is an unavoidable element of real world data, whose structure may not conform to any of the assumed signal models, and obscures the signal. Therefore, the presence of noise can affect the quality of information extracted from alternative representations, therefore, it is useful to understand how the decision performance degrades with increasing level of noise.

Decisions are made based on the analysis of the features accessible in the data products, for applications including environmental characterization, object detection, or fishing survey. The reliability of the decision depends on the quality of the processed data, which is often affected by the post-processed signal-to-noise (SNR) level. One of the main objectives in underwater acoustic signal processing is to increase the reliability of decisions by improving the post-processed SNR [9], through re-organization of the raw data into a more suitable representation.

Given that multiple types of acoustic phenomena are associated the overall active sonar response, we investigate a set of representations to assess their relative discriminatory power, under a range of noise levels. While the set of representations explored in this work may not be optimal, they provide insight into the trade-off involved with multiple representation-based decision-making approach. Using noise-controlled data collected on canonical objects designed for representation analysis, we discuss the detection of discriminatory features and

statistical separability of classes using the detected features. We demonstrate that utilizing multiple representations can improve the reliability of decisions, especially for noisy data.

The remainder of the paper is organized as follows; Section 2 introduces the two main representations SAS imagery and k-space. Section 3 derives the pixel statistics and how they change with additive noise, followed by Section 4 where a summary statistic of the discriminatory pixels is derived. In Sections 5 and 6 the data collected with an in-air measurement system and the experimental results are presented, followed by conclusions.

2 Sonar Data and Representation

The primary sensor for underwater remote sensing applications is acoustic due to its relative efficiency compared to electromagnetic wave propagation, which is much more prevalent in terrestrial applications. Similar to synthetic aperture radar (SAR) used by aerial vehicles, synthetic aperture sonar (SAS) employed by underwater vehicles allow for aperture lengths beyond the physical constraints of real aperture arrays. With the aid of accurate motion estimation and compensation of underwater sensor platforms, forming a large synthetic aperture leads to finer resolution of SAS imagery, from which shape features of objects and sea floor textures can be extracted [10, 11].

Human perception relies heavily on its visual sensor, and utilizes basic shape features [12] and their relative organization for object recognition and classification [13, 14]. This justifies the use of high-frequency sonar systems and the associated acoustic wavelengths that exhibit similar scattering phenomena as optical scattering, and the transformation of acoustic raw data into a visual representation for information processing by the human perception system [15, 16]. This is also advantageous for signal processing as mature image analysis tools can be applied almost directly to acoustic image data products, and the information extraction methods for further decision-making.

However, it is empirically observed that structural acoustic phenomenology at longer wavelengths, or lower frequencies, contain discriminatory information [17] beyond just the object shape. For example, a pair of objects with the same exterior, but made with different materials would exhibit a different overall acoustic response due to the differences in structural resonances. While their geometric shape features would appear nearly identical in a SAS image, their resonant responses would not be clearly represented in the imagery.

A different type of representation may be better suited for this discriminatory information, or the resonant response. The spatial wavenumber domain, or k-space, is the 2D Fourier transform of the complex-valued SAS image. The magnitude of this representation, or k-space, in particular shows the distribution of resonance energy over a range of insonification angles of the sonar.

One of the advantages of the wavenumber domain representation over the spatial image is in the level of concentration and the organization of acoustic energy associated with structural acoustics of the object. The underlying assumption of image formation SAS processing is the isotropic nature of the geometric scattering response, which allows for large coherent

processing gain over a wide range of scattering angles. However, the structural response of the same object is not independent of scattering angle. Instead, it is the aspect-dependent response that helps distinguish different objects with the same exterior.

Fig. 1 shows the time series data for solid copper cylinder and hollow copper cylinder with 0.82mm wall thickness, whose exterior dimensions are identical, but their differences in the structural responses evident in the late-time resonance response after 6ms. While it is difficult to recognize structured organization of features associated with the resonances in the image domain shown in the comparison between Fig. 2 (a) and (b), the aspect-dependent resonance structure of the hollow object is clear in the late-time wavenumber domain image in the comparison between Fig. 2 (e) and (f), which allows for a more visually intuitive discrimination between the two objects. Note that (c) and (d) are the wavenumber images of the full time series, in which the specular scattering response have dominated the relatively weaker resonances.

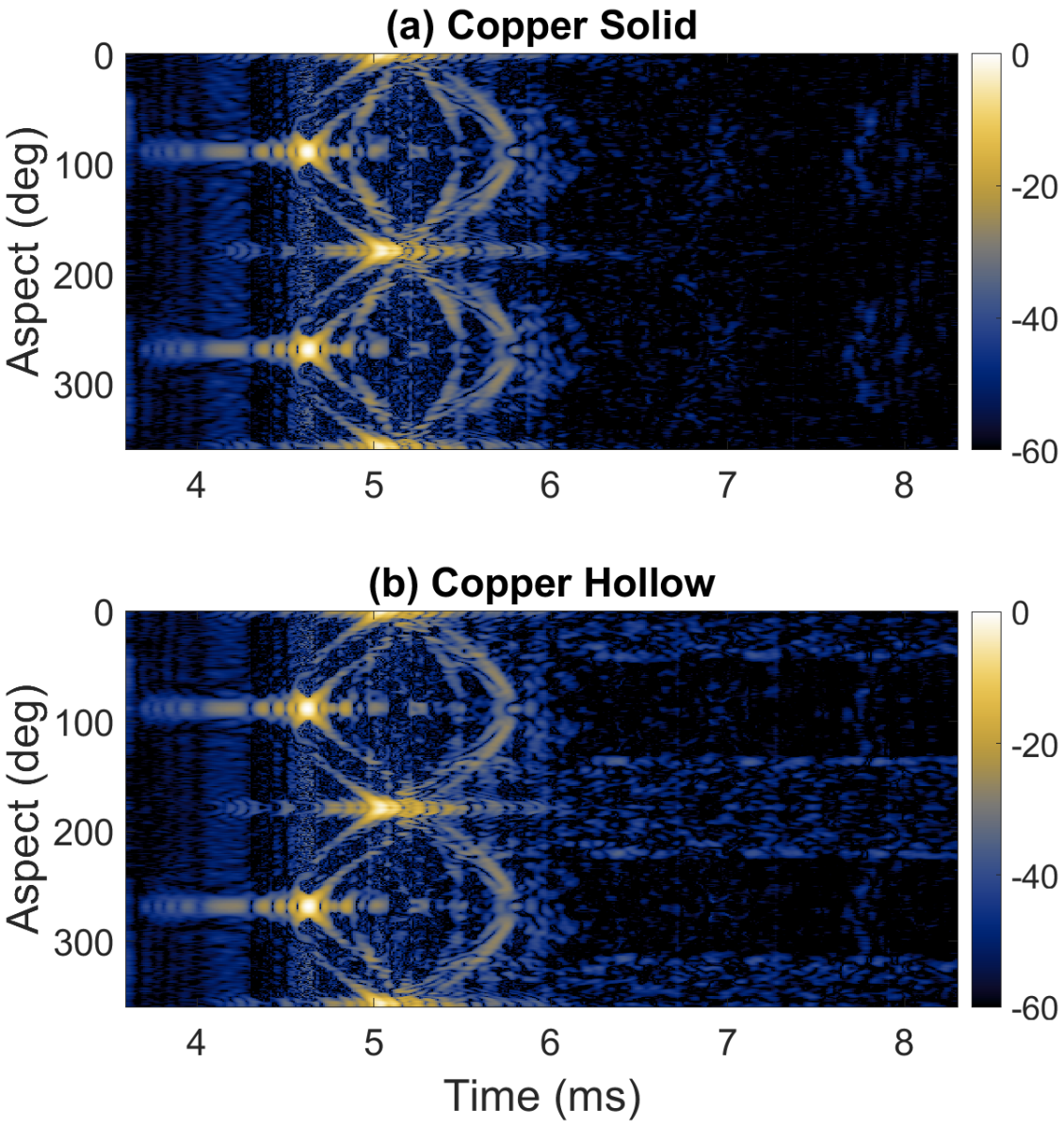


Figure 1: Time series data of solid copper cylinder and hollow copper cylinder with 0.82mm wall thickness. Aspect-dependent late-time resonances are visible from 6ms

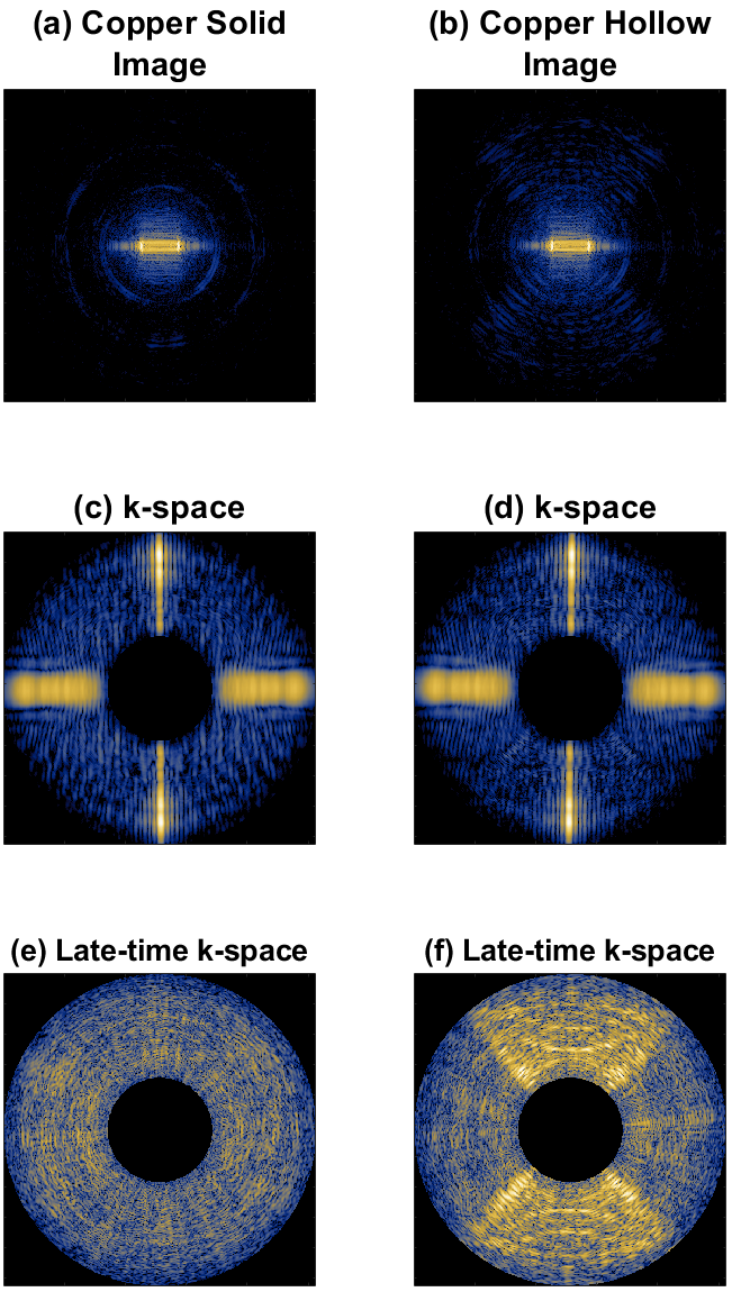


Figure 2: Synthetic aperture sonar image and wavenumber domain representations of acoustic data collected on solid and hollow cylinders. Last row are k-space of the late-time resonances.

2.1 Synthetic Aperture Sonar Imagery

The use of SAS imagery in underwater remote sensing applications has been increasing for its fine image resolution that does not degrade with range, while it does degrade with range in the case of conventional real aperture sonar (RAS) imagery. Each pixel in the SAS image is formed by aligning and adding the scattering responses of an object from multiple pings. It is also referred to as delay-and-sum beamforming [18, 19]. Constructively adding the object responses to reconstruct high-contrast imagery not only requires accurately estimating and compensating for the sensor motion, but also relies on the object responses from different scattering directions to be similar to each other so that they constructively interfere. An ideal elementary target for SAS processing that yields the theoretical resolution limit is a point scatterer, whose response is independent of direction of the sonar. The resolution of the SAS image is determined by the width of the point spread function (PSF), which is determined by the bandwidth of the pulse used by the sonar, and by the length of the synthetic aperture [20]. Wider band pulses and larger synthetic aperture achieve narrower PSF [21]. The specular response of an object is well-approximated as a specific spatial arrangement of point scatterers that form the shape of the object.

The overall response of objects, however, include components that are not well-approximated as a collection of point scatterers. For example, while much of the transmitted pulse it reflected off of the object surface back to the sonar, part of the pulse energy is acoustically coupled onto the object and re-radiated back to the sensor at a later time. This process is frequency-selective and determined by the structural properties of the object [17]. The spectral characteristics of these components are also dependent on the direction of scattering. As a result, the acoustic energy associated with these components do not focus as tightly as the ideal PSF of the point scatterer. The SAS image signatures associated with these acoustic phenomena are out-of-focus and low-contrast compared to the specular component.

2.2 Spatial Wavenumber Imagery

Fine-resolution high-contrast SAS imagery is a result of spatio-temporal pulse compression over time and the synthetic aperture [20]. The specular component of object scattering suffers little distortion from the broadband transmitted pulse, and therefore can achieve high level of pulse compression. On the other hand, structurally excited resonant responses are spectrally narrow due to the frequency-selective mechanisms associated with it, and late in time with indeterminable timing.

In order to re-organize the shape of the feature and to improve the concentration level of acoustic energy associated with the late-time resonant responses, an alternative representation such as the Fourier domain may be considered in lieu of SAS imagery. In its simplest form of description, the 2-D spatial Fourier transform of the complex-valued SAS image is the representation of interest in this work. Other names for this representation are spatial wavenumber domain, or k-space. This alternative representation is also a complex-valued image, but only the magnitude of k-space will be considered in this work. While the notion of an elementary point-like target in this domain is physically elusive, compared to the point scatterer in the spatial domain, a spatially-limited version of a point target in k-space could

be conceived. The k-space signatures that make up the late-time resonant response, seen as traces of point clusters in Fig. 2 (f), is considered a specific arrangement of this elementary target response, similar to how the object shape is a specific arrangement of point scatterers in the spatial domain.

It is worth noting that energy is conserved when transforming between image and k-space, but its distribution and feature-level concentrations vary significantly. In practice, signal is mixed with additive noise, and the concentration level of acoustic signature determines how much noise the target signatures can withstand the noise before they are dominated by noise and become indistinguishable.

2.3 Time Segmented Representations

While the specular component of the overall acoustic scattering response from the object's exterior surface is well-focused by the image formation algorithm to yield high-contrast shape features in the spatial image, the late-time resonant responses do not conform to the image formation model. Their resonant characteristics imply Fourier bases may be better suited for representing these components. However, the resonant response is significantly weak compared to the specular component, and direct Fourier transform of the entire data results in the resonance features being obscured by the specular features. The overlapping of features in the k-space hinders the detection of discriminatory features. This phenomenon can be mitigated by applying a preprocessing step to separate the early-time response and the late-time response by time-gating the overall data relative to the first arrival of the specular response.

As discussed in Section 2, Figs. 2 (e) and (f) are a comparison of the late-time only k-space of for solid cylinder and hollow cylinder. It is clear the late-time resonances are discriminatory features between the two objects, given the noise level is sufficiently low. While the late-time resonant response that discriminate between solid cylinders and hollow cylinders appear as defocused blurry feature in the spatial image, in the k-space they are organized in a more structured pattern tracing out a curved trajectory. The feature concentration level can be further improved beyond k-space if the shape of the spectral features were known a priori and incorporated in the processing algorithm.

3 Image Feature Statistics

In this section, we first derive statistical models for each pixel, followed by estimating the parameters for the model. Then, we discuss their role in discriminatory feature identification for classification.

Let the measured and basebanded acoustic data vector be

$$c = \alpha s + n, \tag{1}$$

where α is a scalar, s is known complex-valued signal with unit norm, *i.e.,* $s^H s = 1$, and n is complex normal random vector, $n \sim \mathcal{CN}(0, \sigma_n^2 I_2)$, where I_2 is an identity matrix of

dimension 2. The zero-lag peak of the matched filtered data $Z = s^H c$ is a random variable, and is also a complex normal random variable. We consider a signal detection problem with two possible outcomes [22, 23]; H_0: signal is absent, H_1: signal is present. Therefore,

$$H_0(\alpha = 0) : Z \sim \mathcal{CN}(0, \sigma_n^2 I_2), \tag{2}$$

$$H_1(\alpha \neq 0) : Z \sim \mathcal{CN}(\alpha, \sigma_n^2 I_2). \tag{3}$$

The magnitude of Z is Rice-distributed, $|Z| \sim Rice(\alpha, \sigma_n)$, where $\alpha = 0$ is a special case, in which the distribution becomes Rayleigh. In a more general signal classification problem, two types of signals may be defined in (1) as s_0 and s_1, but the cross correlation may be significantly small, $s_1^H s_2 \approx 0$.

SAS image formation algorithms, combine multiple matched filtered measurements ($m = 1, 2, 3, \ldots, M$) to generate complex-valued spatial image, $I(x, y)$. The spatial wavenumber representation, or k-space, is denoted as $K(k_x, k_y)$. Each discrete pixel in the spatial image, $I(x(i), y(j))$ is computed as a coherent sum of M measurements, each of which is aligned so that they all correspond to the same scatterer position, $(x(i), y(j))$.

$$I(x(i), y(i)) = \frac{1}{M} \sum_{m=1}^{M} Z_m(t - \tau_m(i, j)) \tag{4}$$

$$= \frac{1}{M} \sum_{m=1}^{M} \alpha_m^2 R_{s,s}(t - \tau_m) + R_{s,n_m}(t - \tau_m) \tag{5}$$

where t is time since each transmission, $\tau_m(i, j)$ is round trip time of the m^{th} pulse from the transmitter to and from the scatterer at $(x(i), y(j))$ back to the receiver, $R_{s,s}$ is the auto-correlation function of $s(t)$ whose peak is 1, and R_{s,n_m} is the cross-correlation function of $s(t)$ and $n_m(t)$. The pixels $(k_x(i), k_y(j))$ in the wavenumber domain image are the 2-D spectral coefficients of the spatial image. The component wavenumbers k_x and k_y can be represented as $(\omega, \theta) = \left(\frac{c}{2} \sqrt{k_x^2 + k_y^2}), \tan^{-1} \left(\frac{k_y}{k_x} \right) \right)$, where c is the speed of sound.

In the absence of echo return, or when $\alpha_m = 0$, the formation of each pixel at position $(x(i), y(j))$ is computed as a sum of noise samples only. For simplicity of analysis, independence between the noise samples is assumed, which leads to independent pixels with complex normal distribution $I(x(i), y(j)) \sim \mathcal{CN}(0, \frac{\sigma_n^2}{M} I_2)$. Each pixel in k-space is a linear combination of each pixel in the image. Since each image pixel is independent, the linear combinations complex normal pixels are also distributed complex normal. A discussion on the dependence between pixels in k-space will follow in a later section.

For pixels containing objects, the pixel value is the integrated scattering response over all measurements as (5) with $\alpha_m \neq 0$. While the scattering response from different parts of the object and their orientation relative to sensors affect the received level, for simplicity of analysis the underlying true scattering strength is assumed to be constant, $\alpha_m = \alpha$.

The intensity of pixels in SAS image or in k-space image is the magnitude of the complex-

valued pixel. The pixel value itself can be re-formulated as a sum of the signal component and the noise component,

$$I = \mu \cdot e^{j\Theta} + U + jV \tag{6}$$

where U and V are the real and imaginary part of the complex normal with $\mathcal{CN} \sim N(0, \frac{1}{M}\sigma_n^2 I)$. The scalar μ is constant, and the phase is uniformly distributed, $\Theta \sim U(-\pi, \pi)$. When $\alpha = 0$, and in turn $\mu = 0$, I is complex normal with the joint PDF of U and V,

$$f_{UV}(u, v) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{u^2 + v^2}{2\sigma^2}\right) \tag{7}$$

with $\sigma = \sigma_n/\sqrt{M}$. In polar coordinates, the joint distribution of phase and magnitude is,

$$f_{\Theta R}(\theta, r) = f_\Theta(\theta) \cdot f_R(r) = \frac{1}{2\pi} \cdot \frac{r}{\beta^2} \exp\left(-\frac{r^2}{2\beta^2}\right), \tag{8}$$

where the phase distribution is uniform and the magnitude is Rayleigh-distributed, $|I| \sim Rayleigh(\beta)$, with $\beta = \sigma_n/\sqrt{M}$. For estimating the parameters using samples $r_k, k = 1, 2, 3, \ldots, K$, we use the method of moments. Using the second moment of Rayleigh variable, $E[|I|^2] = 2\beta^2$, the estimated Rayleigh parameter is $\hat{\beta} = (\frac{1}{2K}\sum_k r_k^2)^{1/2}$.

For non-zero μ case, the joint distribution between the real and the imaginary part of the complex-valued random variable is a circular non-zero mean bivariate normal distribution, described as

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2 + \mu^2}{2\sigma^2}\right) I_0\left(\frac{\mu\sqrt{x^2 + y^2}}{\sigma^2}\right) \tag{9}$$

where $I_0(\cdot)$ is the modified Bessel function with of the first kind, with constant μ and $\sigma = \sigma_n/\sqrt{M}$. The magnitude components is expressed as,

$$f_R(r) = \frac{r}{\beta^2} \exp\left(-\frac{r^2 + \mu^2}{2\beta^2}\right) I_0\left(\frac{\mu r}{\beta^2}\right) \tag{10}$$

The magnitude of I is Rice-distributed, $|I| \sim Rice(\mu, \beta)$, where $\mu = M \cdot \alpha^2 \cdot$, and $\beta = \sigma_n/\sqrt{M}$. It is known that Rice distribution is well-approximated as normal distribution for $\mu/\beta > 3$, $|I| \approx N(\mu, \beta^2)$. On the other hand, when $\mu/\beta < 1$, $|I| \approx Rayleigh(\beta)$.

Again using the method of moments for estimating the Rice parameters, we use the fact the second and fourth moments of Rice random variable are $E[|I|^2] = 2\sigma^2 + \mu^2$, and $E[|I|^4] = 8\beta^4 + 8\beta^2\mu^2 + \mu^4$. The estimated parameters are,

$$\hat{\mu} = \left(2\left(\frac{1}{K}\sum_k r_k^2\right)^2 - \frac{1}{K}\sum_k r_k^4\right)^{\frac{1}{4}} \tag{11}$$

$$\hat{\sigma} = \left(\frac{1}{2} \left(\frac{1}{K} \sum_{k} r_k^2 - \hat{\mu}^2 \right) \right)^{\frac{1}{2}}. \tag{12}$$

3.1 Feature Organization and Representation

In the formulation of a 2-D complex image pixel (6), the spatial intensity distribution of SAS image is determined by the spatial arrangement of $\mu(i,j)$, where i and j are horizontal and vertical indices, respectively. The μ values are determined by α_m values over all measurements via the image formation algorithm. The intensity distribution in k-space is dependent not just on $\mu(i,j)$, but also on $\Theta(i,j)$, since both the magnitude and phase contribute to the 2-D complex Fourier transform. It should be noted that both $\mu(i,j), \Theta(i,j)$ are not independent from $\mu(k,l), \Theta(k,l)$, for any 4-tuple i,j,k,l. However, the signal component and the noise component of each pixel are mutually independent.

Given a pair of objects, *e.g.*, solid cylinder and hollow cylinder, with identical exterior dimensions, represented in spatial image with N pixels, their $\mu(i,j)$ features around the perimeter of the object would be very similarly shaped, as seen in the middle sections of Fig. 2 (a) and (b). On the other hand, the late-time elastic response from the hollow cylinder would yield a different $\mu(i,j)$ pattern from those of a solid cylinder. Consider generating a set of magnitude images with the same measurement setup with independent noise realizations. The distinguishability between the two objects given a sample image is determined by the difference between the function $\mu(i,j)$, relative to the variance introduced by the noise component of (6). More specifically, given $|I| \subseteq \mathbb{R}^N$, and a pair of N-dimensional joint PDFs, each for solid cylinder and hollow cylinder, $P(I)$ and $Q(I)$, respectively, the statistical distance between the two joint PDFs determines how distinguishable the two classes are.

Not all of the pixels contribute equally to this decision, especially those with small pixel-wise mean difference $\Delta\mu_{SH}(i,j) = |\mu_S(i,j) - \mu_H(i,j)|$ relative to the pixel-wise variance. Furthermore, including the non-discriminatory pixels in the decision may even hurt the performance since they are more likely to increase the variance than to increase the mean difference. Therefore, as the noise level increases, the number of pixels that maintain their discriminatory contribution will decrease.

3.2 Organized Discriminatory Feature and Representation

Considering alternative representations where the pixel model (6) holds, one representation may maintain its discriminatory power over other representations up to higher levels of noise when the μ distribution is such that much of the energy is concentrated into fewer pixels, or if the shape of the feature is easier to recognize. As briefly discussed in Sec 2.2, it is known that Fourier transform is unitary, *i.e.*, the energy or sum of squares between image and k-space is conserved. We are interested in this property holding for the discriminatory information between the two objects of interest. However, as will be discussed in Section 4.2, the discriminatory power is not necessarily conserved between transforms. The Fourier

transform re-distributes the acoustic energy associated with a discriminatory phenomenon, which also affects the level of concentration of this energy. Whether this concentrated energy is above the noise floor determines if the discriminatory feature is detectable. Even at similar levels of concentration, more recognizable shape may facilitate easier feature detection, at least for human consumption or vision-based machine learning algorithms that are trained to find object edges [13].

As the level of additive noise increases, the variance of the noise component of pixel value, σ_n^2 increases, and the shape parameter, β, for the pixel intensity distribution (10) also increases as $\beta^2 = \sigma_n^2/M$. The pixel-level signal-to-noise ratio is defined as $SNR(i,j) = \mu(i,j)^2/\sigma_n^2$, or $SNR_{dB}(i,j) = 10\log_{10}(\mu^2(i,j)/\sigma_n^2)$ where σ_n is considered constant over all pixels. As the noise floor rises the features will be obscured, rendering classification based on the representation unreliable. Assuming the location of discriminatory pixels are known, the classification error will begin to grow as the most discriminatory pixel is dominated by noise, or $\max_{i,j} SNR_{dB}(i,j) \to 0$.

4 Discriminatory Feature Detection And Classification Performance Prediction

For remote sensing applications, we are concerned with the detection of discriminatory features in noisy data. Two classes of quantitative measures are utilized to assess the overall utility of the representations for object classification. First measure is the amount of structure the discriminatory feature shape possess that can be recognized, *i.e.*, how well the discriminatory features are organized, or structured, as they are depicted in the representation. Second measure is the amount of statistically discriminatory information that is present in the representation regardless of the shape of the discriminatory features.

4.1 Discriminatory Features

The consumers of images, whether in the spatial domain or in k-space, are human reviewers or automated algorithms trained on relevant images. At a low level, they rely on finding edges of objects that are both visually salient and potentially discriminatory. Structural similarity (SSIM), developed for predicting perceived quality of digital image and video, is a widely used measure to capture visual similarity between two images [24]. Unlike other image similarity measures, SSIM is less sensitive to the absolute difference. Instead, it correlates better with the visually perceived difference between images [24]. Either a scalar SSIM value can be measured from a pair of images, or a map of local SSIM can be generated. From this map between a pair of representations, $SSIM(i,j)$, a map of structural dissimilarity (SDIS) is defined as, $SDIS(i,j) = 1 - SSIM(i,j)$.

For each object class pair comparison, the SDIS map from all object pair comparisons are used to find the most commonly discriminatory pixels to use in (18) and (20). If a given SDIS value of a pixel is above the top 1% threshold of an exponential fit of SDIS-value distribution for more than one third of comparison object pairs, it is included in the reference

discriminatory pixel set, X_D. This is repeated for each time segmentation and for each domain, yielding $3 \times 2 = 6$ representations.

4.2 Log Likelihood Ratio and Chernoff Information

Consider a vectorized image, $\boldsymbol{x}$, and the index for each pixel is converted from (i, j), to $n = 1, 2, 3, \ldots, N$. In a binary classification problem between two objects with N-dimensional joint PDFs P and Q, and Bayesian priors π_0 and $\pi_1 = 1 - \pi_0$, respectively, the optimal decision rule in terms of the minimum error is based on the Neyman-Pearson lemma [25], which utilizes the likelihood ratio test,

$$\Lambda(\boldsymbol{x}) = \frac{\pi_0}{\pi_1} \frac{P(x_1, x_2, \cdots, x_N)}{Q(x_1, x_2, \cdots, x_N)} \lessgtr T. \tag{13}$$

Assuming independent distributions (i.d.) between pixels, the log-likelihood ratio function is,

$$\begin{aligned} l(\boldsymbol{x}) &= \ln \Lambda(\boldsymbol{x}) \\ &= \ln \frac{\pi_0}{\pi_1} + \sum_{n=1}^{N} \big(\ln P_n(x_n) - \ln Q_n(x_n) \big), \end{aligned} \tag{14}$$

where P_n and Q_n for each pixel magnitude are Rice distributions with $R_{P_n} \sim \mathit{Rice}(\nu_n, \beta_n)$ and $R_{Q_n} \sim \mathit{Rice}(\mu_n, \sigma_n)$. With equal priors, *i.e.*, $\pi_0 = \pi_1 = \frac{1}{2}$, and using the approximation $\ln\left(I_0(x)\right) \approx x - \frac{1}{2}\ln\left(2\pi x\right)$, (14) is expanded to,

$$\begin{aligned} l(\boldsymbol{x}) &= \sum_{n=1}^{N} \ln \frac{\sigma_n^2}{\beta_n^2} - \frac{\nu_n^2}{2\beta_n^2} + \frac{\mu_n^2}{2\sigma_n^2} - \frac{1}{2} \ln \frac{\nu_n \sigma_n^2}{\beta_n^2 \mu_n} \\ &\quad + \left(\frac{\nu_n}{\beta_n^2} - \frac{\mu_n}{\sigma_n^2} \right) x_n + \left(-\frac{1}{2\beta_n^2} + \frac{1}{2\sigma_n^2} \right) x_n^2 \\ &= \sum_{n=1}^{N} a_n + b_n x_n + c_n x_n^2, \end{aligned} \tag{15}$$

where a_n, b_n, c_n are the coefficients for the second order polynomial of x_n. The log-likelihood ratio is the sum of Rice random variables, x_n's, with different parameters, which converges to a Normal random variable, $Y \sim N(\mu_l, \sigma_l^2)$ when N is sufficiently large [26]. The Normal parameters are computed as

$$\mu_l = \sum_{n=1}^{N} a_n + b_n E[x_n] + c_n E[x_n^2] \tag{16}$$

$$\sigma_l^2 = \sum_{n=1}^{N} b_n^2 V[x_n] + c_n^2 V[x_n^2] + 2 b_n c_n (E[x_n^3] - E[x_n] E[x_n^2]) \tag{17}$$

where $E[\cdot]$ is the expected value and $V[\cdot]$ is the variance, respectively. These quantities are dependent on the object class, *e.g.*, $E_P[\cdot]$ or $E_Q[\cdot]$. Note that the limit of the sum in (14) or (15) reduces to N_D when only considering the discriminatory pixels, and the Normal approximation still holds with sufficiently large N_D.

With the summarized decision statistics of the discriminatory pixels modeled as Normal random variables, Y_P and Y_Q, we can predict the change in their parameters as the noise level increases via (10), (15), and (16),(17). Now, by characterizing how the statistical distance between the two PDFs changes with the noise level, we can predict how much noise the discriminatory features can withstand before the classification error begins to grow, for each representation.

Chernoff information is a statistical divergence measure between P and Q, defined as,

$$C(P,Q) = \max_{\lambda\in(0,1)} -\ln \int P^{\lambda}(x)Q^{1-\lambda}(x)dx. \tag{18}$$

As the integrand indicates, $C(P,Q)$ is also considered the geodesic bisection between a pair of distributions in terms of information geometry, and upper bounds the error of Bayesian decision rule [27] as,

$$E^* \leq \pi_0^{\lambda^*}\pi_1^{1-\lambda^*}e^{-C^*(P,Q)}, \tag{19}$$

where λ^* is the value that achieves the maximum in (18). The Chernoff information for a pair of Normal distributions $N(\mu_P,\sigma_P^2)$ and $N(\mu_Q,\sigma_Q^2)$ is [28]

$$\begin{aligned} C(P,Q) = &\frac{\lambda(1-\lambda)}{2}(\mu_Q-\mu_P)^T[\lambda\sigma_P^2+(1-\lambda)\sigma_Q^2]^{-1}(\mu_Q-\mu_P) \\ &+\frac{1}{2}\ln\frac{|\lambda\sigma_P^2+(1-\lambda)\sigma_Q^2|}{|\sigma_P^2|^{\lambda}|\sigma_Q^2|^{1-\lambda}} \end{aligned} \tag{20}$$

where Bhattacharrya distance is found with $\lambda = 1/2$, and the Chernoff information is found by maximizing (20) either analytically or numerically [28]. It is worth noting the SNR-like function akin to the pixel level SNR discussed in Section 3.2. A subtle distinction from a direct interpretation of feature SNR is that even if both μ's are large, which implies the pixel level SNR is large, the statistical divergence can still be small if their values are similar, as will be discussed with experimental data.

Another point worth noting with SAS imagery generated in practice is the post-processing step referred to as dynamic range compression (DRC) used for intuitive interpretation by human visual system. Linear scale pixel intensities are normalized by amplifying the low-amplitude pixels, or suppressing the high-amplitude pixels. While sophisticated DRC methods are intricate, often data-driven, normalization applied to the linear scale, applying a logarithm transformation is a good approximation of this mapping. It is known that Chernoff information is invariant to parameter transformation of random variables [25], and the image statistics on the logarithmic scale yields the same statistical divergence as the linear scale.

A simplified version of (20) with $\lambda = 1/2$ with equal variance, $\sigma_P = \sigma_Q = \sigma$ is

$$C(P,Q) = \frac{1}{8}\frac{(\mu_Q - \mu_P)^2}{\sigma^2} \tag{21}$$

It is worth noting the effective signal to noise ratio between the μ's and the σ is the main contributor to the statistical separability between the two classes. This is consistent with the typical detection problem, however, the formulation of these parameters, as described in (16) and (17), is a sophisticated combination of the pixel parameters, which are affected by the choice of representation as laid out in Section 2.

5 AirSAS Data Signal Processing

While acoustic data is primarily used for underwater remote sensing applications, complications due to the cost of data collection and processing challenges stemming from sound speed variability and sensor motion instability, in-air laboratory environment measurement system such as one used for this work, AirSAS [29], serves as an apt alternative for controlled experiments. AirSAS is used to collect synthetic aperture sonar data in a circular collection geometry with stationary sensors and objects placed on a rotating table [29]. The main objective is to classify between three classes of cylindrical objects from their SAS data representations and characterize the performance with increasing noise levels. The three classes are Solid, Hollow, and Pipe. Table 1 summarizes the properties of the objects used for this analysis.

Table 1: List of cylindrical objects. All objects are 50.8mm in diameter and 203.2mm in length.

ID	Object	Matrerial	Wall Thickness	Note
1	Solid	Steel	N/A	
2	Solid	Copper	N/A	
3	Solid	Wood	N/A	
4	Hollow	Steel	1.65mm	
5	Hollow	Steel	1.65mm	2mm Hole
6	Hollow	Copper	1.65mm	
7	Hollow	Copper	1.65mm	2mm Hole
8	Hollow	Copper	0.82mm	
9	Hollow	Copper	0.82mm	2mm Hole
10	Pipe	Steel	1.65mm	
11	Pipe	Copper	1.65mm	
12	Pipe	Copper	0.82mm	
13	Pipe	Aluminum	1.65mm	

Both spatial imagery and k-space imagery were used for the analysis, with an interest in comparing the strength of discriminatory features, and their robustness against additive noise in

the acoustic measurements. The dimension of both spatial and spatial wavenumber (k-space) imagery are 180 x 180 pixels, spanning $1.45m \times 1.45m$, and $-550m^{-1} \times 550m^{-1}$, respectively. Since the choice of representation domain affects the level of signature concentration, leading to different levels of pixel-level SNR, we expected the classification performance will remain superior on the representation which has a higher concentration of discriminatory signature. In the case of similar concentration level, the domain with more structured features will be easier to interpret.

In order to separate the late-time resonant response from the specular response, as discussed in Section 2.3, the data is time-gated in three ways; early- (3.7-6.1ms), late-(6.1-8.5ms), and full-time. The purpose of the full-time case is to serve as a reference, especially for k-space, in which the weak resonant features are obscured by the strong specular response so that we can observe the impact on classification. Time-gating is done by taking the product of the time-gate mask with the raw data before image formation algorithm is applied.

In Fig. 2 (f), it is evident that hollow cylinders exhibit resonant signatures in late time segment, more conspicuously concentrated in the wavenumber domain. In the spatial image, although the corresponding acoustic energy is contained in the same angular range as that of the wavenumber domain, they are radially dispersed with little visual structure, as depicted in Fig. 2 (b). We posit the underlying signal model of the late-time acoustic signal is more consistent with the underlying signal bases of the wavenumber representation, compared to that of the image representation. On the other hand, the noise part of the data does not conform to either underlying signal models of the two representations, failing to coherently accumulate. Instead, as the noise level increases it does begin to fill in the low signal-level pixels of the images, and also increases the variance of all pixels. Consequently, the distribution of the pixel magnitude converges to Rayleigh with a larger shape parameter, for both the low- and high-signal levels.

Taking advantage of the precise control of the AirSAS framework, we aligned the measurements from different objects, which allows for using SSIM and in turn SDIS discussed in 3.2 to identify discriminatory pixels between object classes used for analysis. Fig. 3 (a) and (b) show the SDIS map that highlights the discriminatory features for hollow copper vs. copper pipe, and (c) and (d) for solid copper vs. hollow copper, respectively, both in image and in k-space.

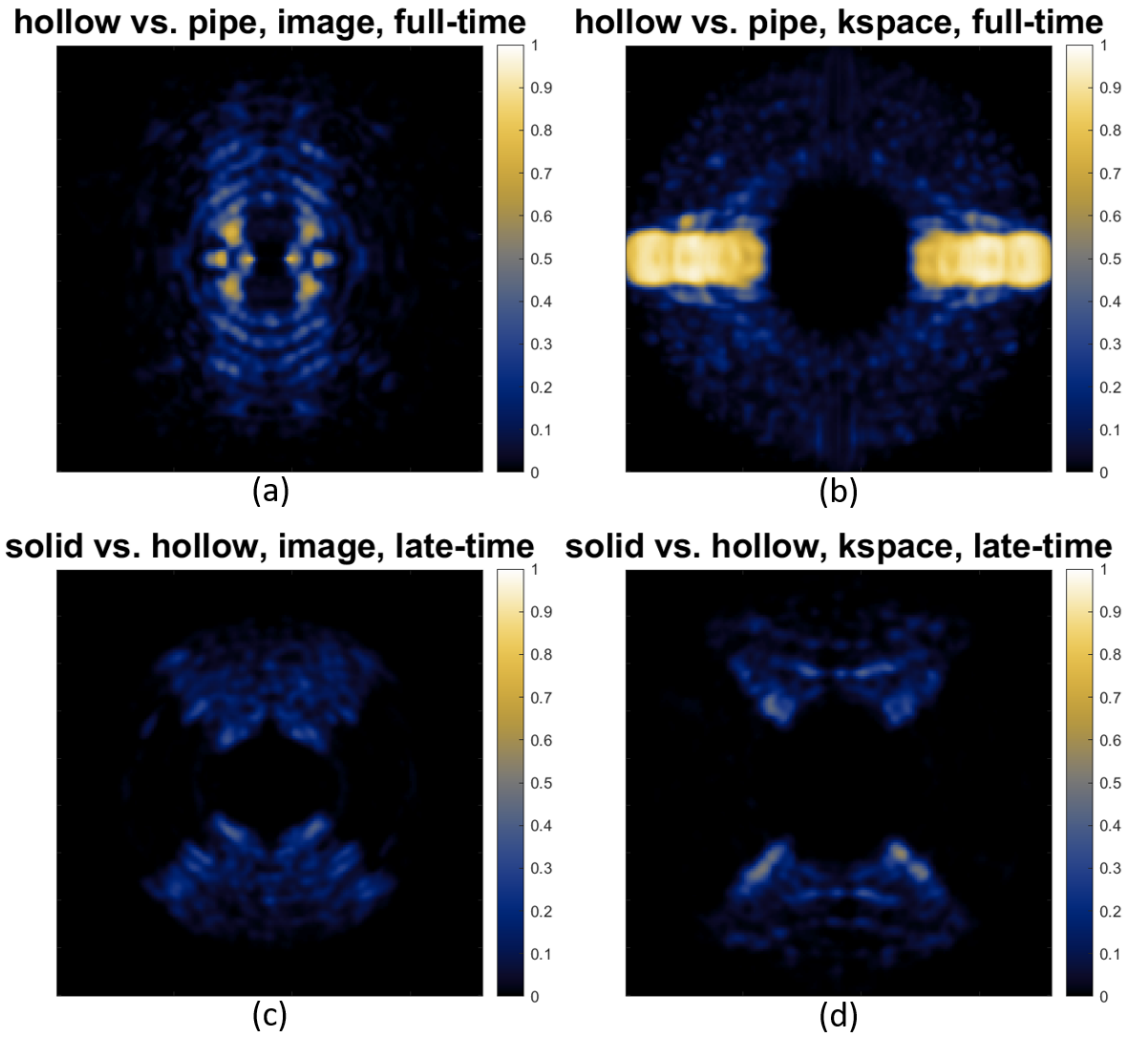


Figure 3: Structural Dissimilarity (SDIS) maps between (a) (b) hollow copper cylinder vs. copper pipe, (c) (d) solid copper vs. hollow copper cylinders, in image and k-space, respectively.

5.1 Noisy Data

For the noise analysis, randomly-generated additive noise samples were added to the measured data collected in a noise-controlled environment. This additional noise is independent from the measured signal and noise, and the resulting variance of noise pixels should be the sum of the initial variance and the additional variance, which is accounted for in the performance prediction model discussed in Section 3. Assuming the signal contribution to the pixel variance is known, the increased variance of a pixel variance is due to the additive noise. As the additive noise level becomes high, this variance should dominate the overall variance. The range of noise level was determined so that at the highest noise level even the strongest discriminatory features would be dominated by noise, and at the weakest level it would not affect the pixel statistics significantly, with 5dB increments. Referenced to

the background noise level σ_{ref}, the noise-to-noise level was chosen to be NL$= 20\log_{10}\frac{\sigma_{AN}}{\sigma_{ref}}$, ranging from -50 up to 20 at 5dB increments, and 600 samples per each noise level.

5.2 Feature Energy Detectors

The feature summary statistics discussed in Section 4.2 is considered a well-informed optimal feature detector, *i.e.*, discriminatory pixels are identified and their parameters are known, with the independent distribution assumption. In this section, we also discuss the performance of an empirical feature detectors designed with less information assumed to be known, applied to the noisy data.

First, the discriminatory pixels are known, each of which are assumed to be independently distributed (i.d.). However, their parameters or distribution functions are not known. Under these assumptions, the test statistic is derived in a similar manner as (16) and ()17) for the pair of normal distribution, but the coefficients are $a_n = 0, b_n = 0, c_n = 1$. We call this case the independent pixel energy detector (IPED).

Second, we remove the i.d. assumption, and apply an energy detector on the same set of pixels that may include dependent pixels energy detector (DPED).

$$E_{DP} = \sum_{n=1}^{N_D} x_n^2. \tag{22}$$

For the two empirically designed features, we estimate the mean and variance of the detector statistics, then compute the Chernoff information in the same manner as the well-informed feature detector.

5.3 Discriminatory Acoustic Phenomena

There are three binary classification pairs among the three object classes Solid cylinder, Hollow cylinder, and Pipe. Table 2 summarizes the expected discriminatory structural acoustic phenomena in each time segment. For example, the Solid-Hollow comparison is expected to be distinguishable in the late-time, but less so in early-time due to their similarity specular response. This is by design, to serve as a reference comparison. In Hollow-Pipe comparison, the end caps are likely the strongest discriminating feature, but there will also be differences in the late-time resonances. Note that although the type of discriminatory acoustic phenomena are the same between Hollow-Pipe and Pipe-Solid, how their features are represented in image and k-space are different.

Table 2: Object classification pairs and their discriminatory acoustic phenomena.

Object 1	Object 2	Pairs	Early	Late	Full
Solid(3)	Hollow(6)	18	None	Resonance	Resonance
Hollow(6)	Pipe(4)	24	End caps	Resonance	Both
Pipe(4)	Solid(3)	12	End caps	Resonance	Both

6 Results

For each representation, the Bhattacharyya distance, computed as a special case of the Chernoff information with $\lambda = \frac{1}{2}$, is computed for its simplicity of analytical expression for normal distributions (21). The Bhattacharyya bound is less tight than the Chernoff bound, but it is sufficient for characterizing the relative performance on noisy data.

The top row of Fig. 4 shows the Bhattacharyya distance as a function of noise level (NL), for the all six representations on the same plot for direct comparison. Since this distance was computed with the natural logarithm, its unit is nats. The three columns correspond to the comparison pairs Solid vs. Hollow, Hollow vs. Pipe, and Pipe vs. Solid. The bottom row of Fig. 4 is the corresponding accuracy performance, computed as $Acc = 1 - E^*$ in (19), with equal priors. Note that the transition NL from high- to low-accuracy coincides with where the Bhattacharyya distance crosses 0 nats. As expected, the discriminatory power of each representation degrades with increasing noise. However, it is worth noting the relative discriminatory power of each representation over the whole range of noise level. In low NL region, or high SNR region, late-time representations show greater discriminatory power in Solid vs. Hollow and Pipe vs. Solid comparisons. However, their relatively weak resonances are dominated by noise quickly.

It is also worth noting the accuracy curves of multiple representations stay near 1 as long as the Bhattacharyya distance is sufficiently greater than 0 nats. As a result, the benefit a larger buffer of Bhattacharyya distance above 0 nats is not immediately apparent from the accuracy curves. However, the robustness of classification performance against other factors such as imperfect signal processing or external interference may be improved due to the redundancy of pixels within the representation, *i.e.*, even when some of the pixels are dominated by noise other remaining pixels help maintain the overall discriminatory power. In fact, this behavior is observed in the Fig. 4 (b) full-time image curve, where it tracks the late-time image curve up to NL= -30dB, then tracks early-time image curve for the remainder of NL values. This is due to the full-time image containing both early-time and late-time features in the image as separated pixels, and the late-time feature being more discriminatory in low NL region. On the other hand, this behavior is not observed in corresponding k-space since the early-time and late-time signatures are overlapped in the full-time case.

The robustness may be further improved by utilizing multiple representations. In Pipe vs. Solid, the most discriminatory representation switches from late-time image to full-time k-space, then to full-time image, over the full range of NL. A fused classifier utilizing all six

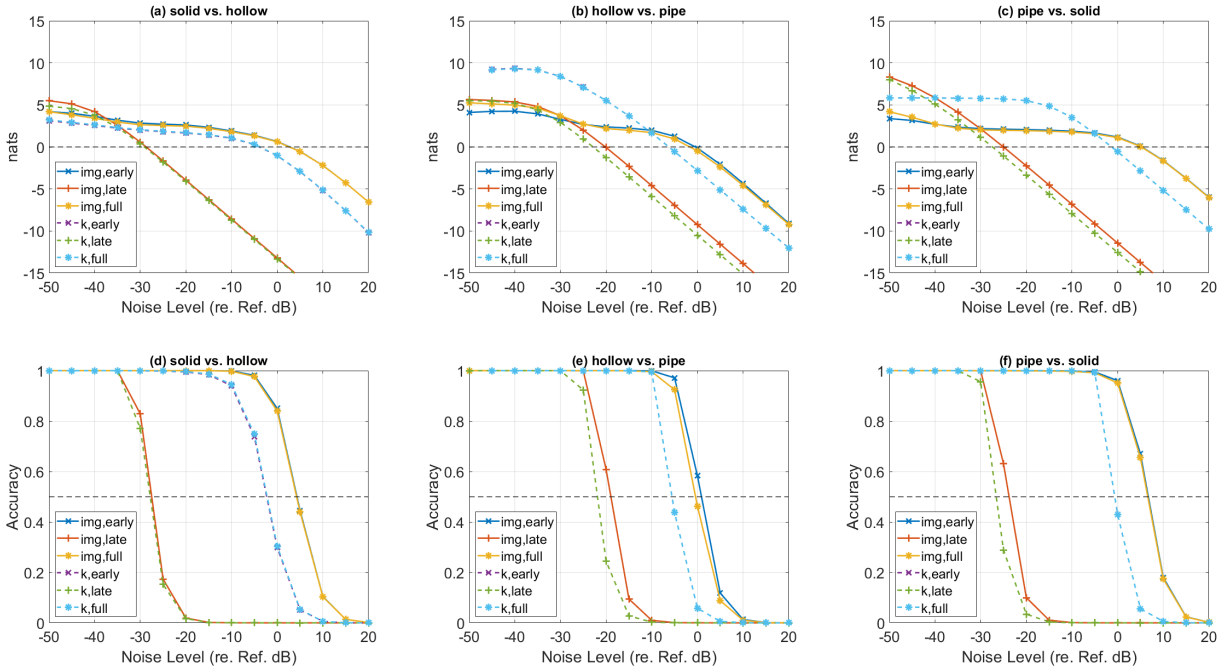


Figure 4: Rician distribution-based predicted Bhattacharyya distance of discriminatory pixels for representations [image,k-space]×[early,late,full]. Top row (a),(b),(c) corresponds to Solid vs. Hollow, Hollow vs. Pipe, Pipe vs. Solid, respectively. Bottom row (d), (e), (f) are the predicted accuracy for object classification of the corresponding column.

representations for this comparison would track the top curve of Fig. 4 (c) at all NL values. The best performing representations would be late-time image up to −40dB, full-time k-space between −40dB and −5dB, then full-time and early-time image after −5dB.

To demonstrate the visual discriminatory power differences of representations at different noise levels, let us consider the Hollow vs. Pipe comparison. At NL=−20dB, shown in Fig. 5, the early- and full-time image as well as the early- and full-time k-space are clearly discriminatory. The difference in feature shapes in the full-time k-space representation, while nuanced, are strong enough to make it more discriminatory than the image counterpart, also confirmed in the plots Fig. 4 (b) and (d).

At NL= 0dB, shown in Fig. 6, while the weaker ripple pattern in the full-time Pipe image is no longer visible, the end caps are still clearly discriminatory in (a). The acoustic energy associated with the end caps in (a) are more concentrated as bright pixels than those of the full-time k-space signatures shown in (c), and the late-time energy in (e) is not visible, fully buried under the noise floor. These observations are consistent with the plots in Fig. 4, in which the full-time image is still above 0 nats, full-time k-space is in the transition region, and the late-time k-space is well below 0 nats.

The set of discriminatory pixels was denoted as X_D in Section 4.1. The robustness of these pixels is quantified by comparing the discriminatory pixels identified using the noisy SDIS map between the classes to the reference discriminatory pixels, X_D. We measure how many of the discriminatory pixels in the noisy SDIS map, $X_D(NL)$, overlap with the reference X_D. Robustness of discriminatory pixels at a specified NL, $\rho(NL)$, is defined as the ratio of the number of pixels that are commonly identified as the discriminatory using the method discussed in Section 4.1 for the noisy SDIS map and the reference SDIS map, to $N_D = |\{x \in X_D\}|$.

$$\rho(NL) = \frac{|\{x \in X_D(NL)\} \cap \{x \in X_D\}|}{|\{x \in X_D\}|} \tag{23}$$

At higher NL values the detection of any object feature may be difficult, especially for the late-time signatures in (e) and (f) of Fig. 6. Fig. 7 shows how ρ changes with NL for each object class comparison and for all representations. Fewer of the pixels in X_D are consistently detectable as the noise level increases, and the degradation occurs more quickly for the relatively weaker late-time resonant responses. On the other hand, early-time image feature maintains detectability relatively well up to the highest NL, although their discriminatory power degrades as seen in Fig. 4. Although accuracy and ρ both degrade at higher noise levels, the transition points are not necessarily consistent with Fig. 7.

In practice, the position of discriminatory pixels are not known and they have to be detected before the class likelihood can be quantified. Therefore, in real world remote sensing applications, the classification performance is often modulated by the detectability of these features, but not necessarily proportional to ρ. For example, there may be fewer weak but highly discriminatory pixels than strong less discriminatory pixels. As the noise level increases, the small number of weaker pixels become undetectable and ρ decreases slightly while most of the discriminatory power is lost (late-time k-space Solid vs. Hollow). The opposite is also possible when there are many strong and discriminatory pixels such that even when ρ is halved the accuracy is still high (full-time image Hollow vs. Pipe). The early-time image ρ only being halved as accuracy approaches the bottom is in part due to the SDIS map using a Guassian kernel whose width was too wide resulting in including strong but non-discriminatory pixels nearby.

The predicted performance of the independent pixel energy detector (IPED) is shown in Fig. 8 in the same order as Fig. 4. Overall, the performance is lower than the optimal case, but the overall characteristics and the relative performance is generally consistent. This approximated model is useful and numerically more stable, however, not able to capture some of the subtle behaviors of the optimal model. In particular, the discriminatory energy in Hollow vs. Pipe in early-time and full-time appear not as discriminatory as the optimal case. This is in part due to the early-time segment containing some of the resonant energy from the hollow cylinder, contributing to the total energy and added variance. Furthermore, the tracking of upper limit across late-time and early-time image by the full-time image seen in Fig. 4 (b) is not observed in Fig. 8. The heterogeneity of the pixels associated with different acoustic phenomena is not fully captured by this approximated performance

model.

In the development of the predicted performances, the pixels were assumed to be i.d. While it is suspected that most represented features are not independent, it is not a simple task to characterize the complicated dependence structure between many pixels identified as discriminatory. This is particularly true in k-space, in which a small number of pixels in the image domain correspond to a large number of dependent pixels in its Fourier domain. The test statistic (22), which can be seen a sum of dependent random variables, can become more separable due to such dependent structure, as shown in Fig. 9.

In Fig. 9 instead of predicting the impact of additive noise with the i.d. assumption, we measure the feature energy via sum of squared pixel intensities, using the same set of pixels in X_D. As seen in most cases, the dependence between pixels causes the performance to increase. A notable exception is (a) Solid vs. Hollow early- and full-time image. The divergence of performance could be explained by the dependence structure of the pixels. It also hints at the potential to exploit the dependence in order improve the feature detection performance. This will be a topics of future work.

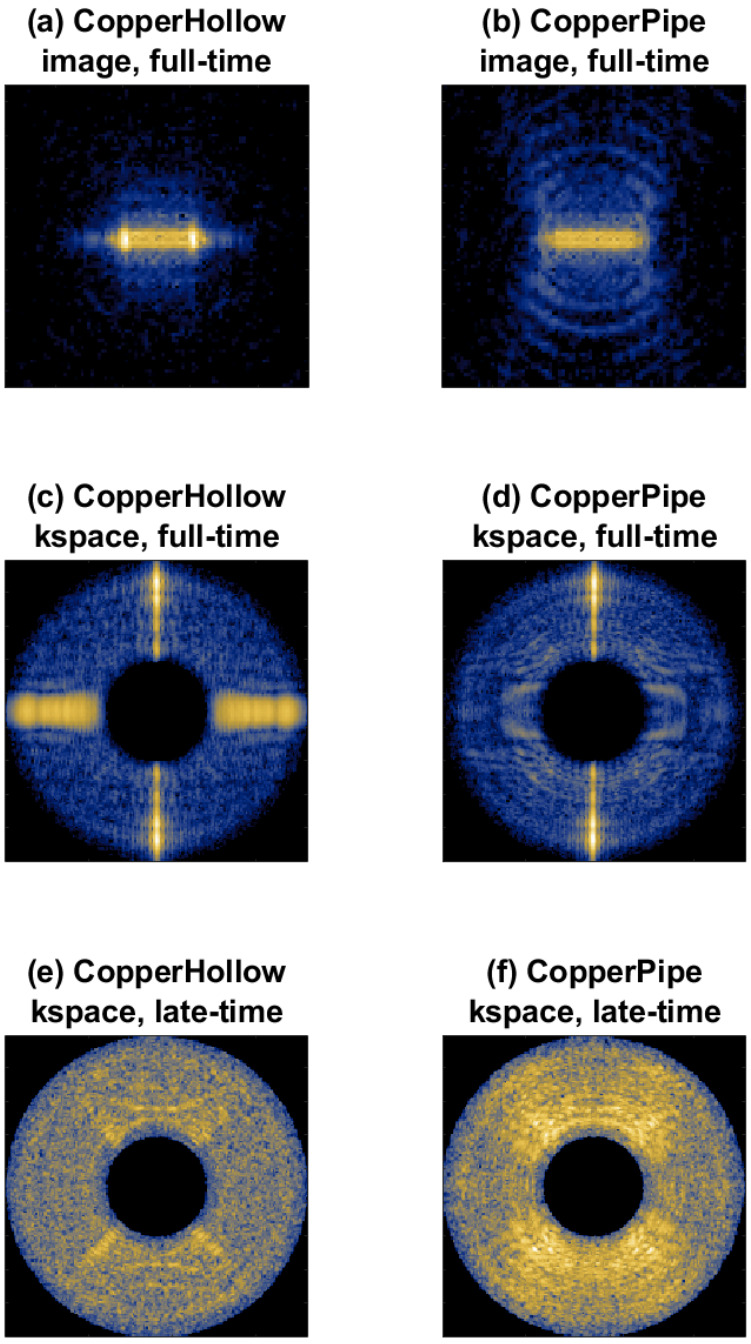


Figure 5: Hollow vs. Pipe representations at NL=−20dB, from top to bottom row; full-time image (zoomed in), full-time k-space, late-time k-space. The prominence of discriminatory features are visually consistent with the plots in Fig. 4 at NL=−20dB.

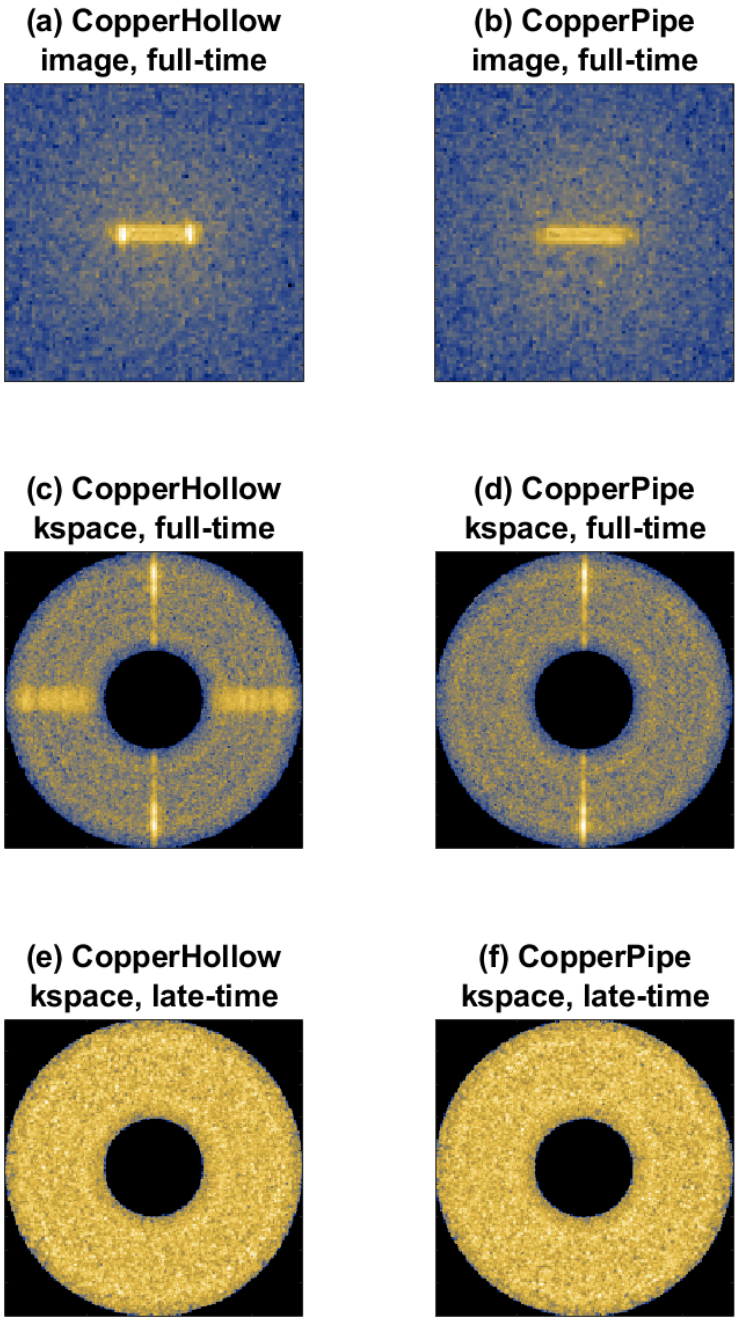
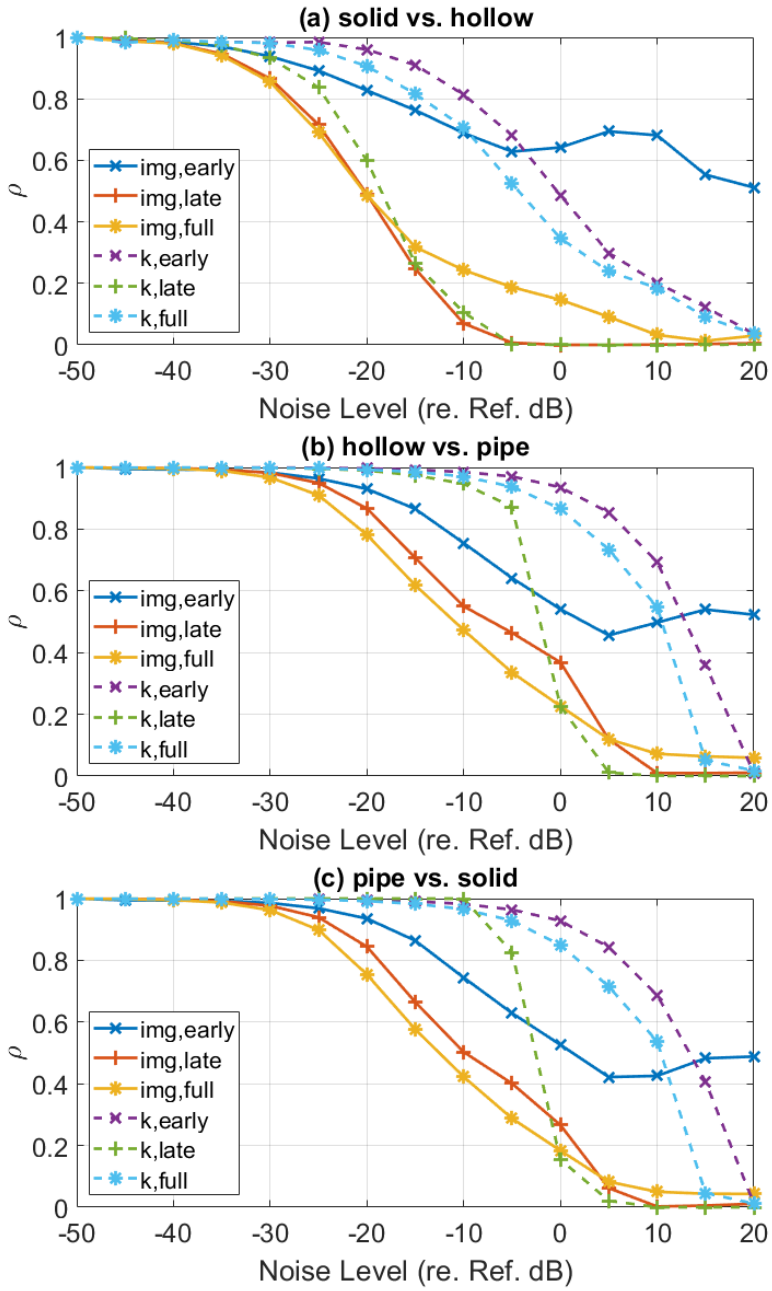


Figure 6: Hollow vs. Pipe representations at NL=0dB, from top to bottom row; full-time image (zoomed in), full-time k-space, late-time k-space. Some of the discriminatory features from Fig. 5 are no longer visible due to stronger noise. The late-time k-space representation is no longer discriminatory, consistent with the plots in Fig. 4 at NL=0dB.



.

Figure 7: Robustness of discriminatory pixels, ρ, over a full range of NL values. Each row corresponds to (a) Solid vs. Hollow, (b) Hollow vs. Pipe, and (c) Pipe vs. Solid. The detectability of discriminatory feature correlates with the accuracy curves in Fig. 4

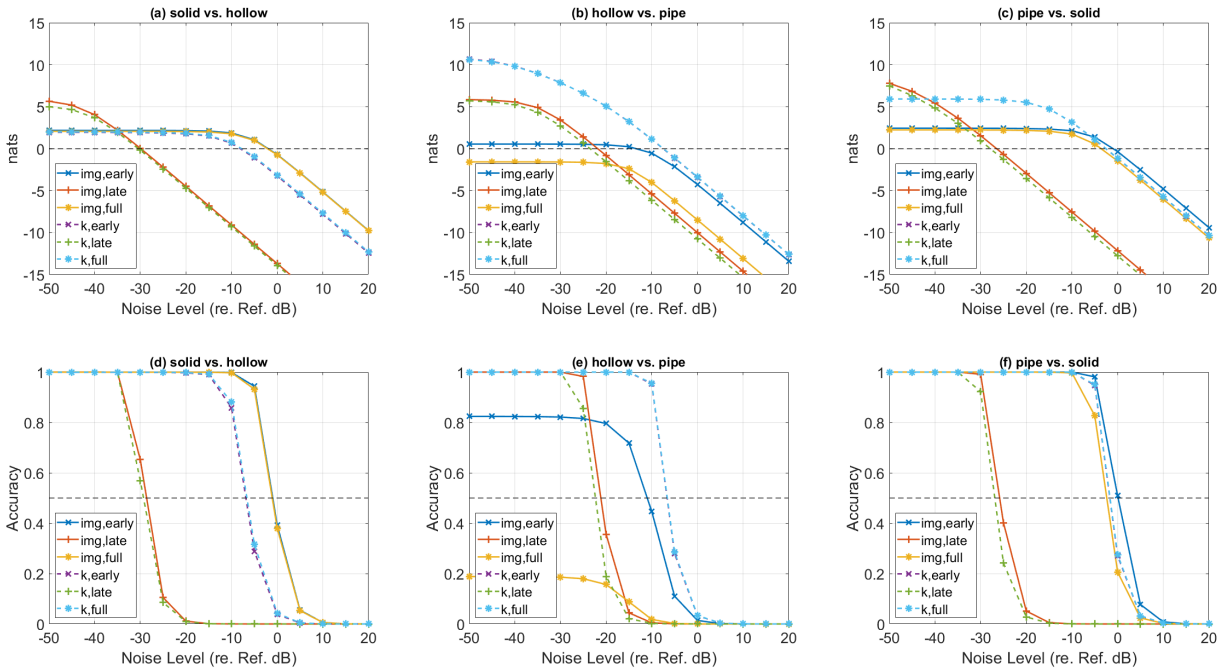


Figure 8: Independent pixel feature energy detector-based (IPED) predicted Bhattacharyya distance of discriminatory pixels for representations [image,k-space]×[early,late,full]. Top row (a),(b),(c) corresponds to Solid vs. Hollow, Hollow vs. Pipe, Pipe vs. Solid, respectively. Bottom row (d), (e), (f) are the predicted accuracy for object classification of the corresponding column.

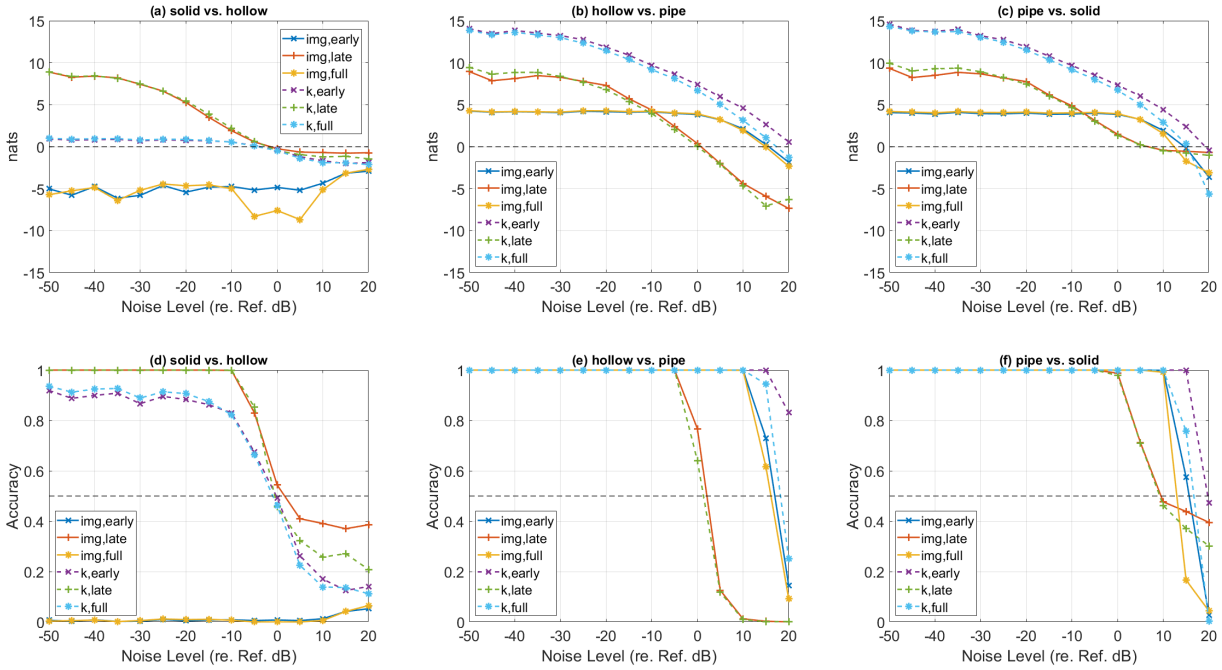


Figure 9: Dependent pixel feature energy detector-based (DPED) estimated Bhattacharyya distance of discriminatory pixels for representations [image,k-space]$\times$[early,late,full], measured at each noise level. Top row (a),(b),(c) corresponds to Solid vs. Hollow, Hollow vs. Pipe, Pipe vs. Solid, respectively. Bottom row (d), (e), (f) are the predicted accuracy for object classification of the corresponding column.

7 Conclusions and Future Work

In current underwater remote sensing applications using acoustic data, SAS imagery is the primary source of decision-making, without many alternative representations of information. In this paper, we demonstrated that utilizing additional alternative representations can help improve the overall decision performance by making the discriminatory information associated with different acoustic phenomena more accessible via isolation and reorganization. The detectability of discriminatory features is also dependent on the choice of representation, and the real-world object classification performance depends on the interplay between feature detectability and statistical divergence of such features. The relative utility of these alternative representations depends on the task and noise level, and the overall robustness of decision based on multiple representations can exploit this fact by characterizing the best representation across a range of noise level.

We developed a statistical model for the pixels in spatial image domain and spatial wavenumber domain, or k-space, and how additive noise affects the statistical model. This model is used to predict the inherent statistical separability between a pair of distributions, whose low-noise parameters can be measured from a data set collected through a series of noise-controlled experiments conducted in air. Using this model with an i.d. pixel assumption, the accuracy performance of the object classification between three object classes is predicted, and compared with an approximated model that utilizes just the feature energy. Statistics of the measured feature energy detector indicate the i.d. assumption is not generally true, but shows potential for improved feature detection performance by exploiting the inter-pixel dependence structure.

Future work topics include further investigation on, and exploitation of, the pixel dependence structure for improved robustness against noise, and additional alternative preprocessing and representations for more recognizable feature shapes for better concentration of acoustic energy and improved feature detectability. Utilization of complex-valued pixels for the phase information in addition to the magnitude of the pixels will also be investigated.

References

[1] J. V. Candy, *Model-based signal processing*. John Wiley & Sons, 2005.

[2] M. P. Hayes and P. T. Gough, "Broad-band synthetic aperture sonar," *IEEE Journal of Oceanic engineering*, vol. 17, no. 1, pp. 80–94, 1992.

[3] W. G. Carrara, "Soptlight synthetic aperture radar," *Signal Processing Algorithms*, 1995.

[4] D. A. Cook and D. C. Brown, "Synthetic aperture sonar image contrast prediction," *IEEE Journal of Oceanic Engineering*, vol. 43, no. 2, pp. 523–535, 2017.

[5] J. D. Park, T. E. Blanford, and D. C. Brown, "Late return focusing algorithm for circular synthetic aperture sonar data," *JASA Express Letters*, vol. 1, no. 1, p. 014801, 2021.

[6] B. Cowen, J. D. Park, T. E. Blanford, G. Goehle, and D. C. Brown, "Airsas: Controlled dataset generation for physics-informed machine learning," in *NeurIPS Data-Centric AI Workshop*, 2021.

[7] I. Gerg and D. Williams, "Additional representations for improving synthetic aperture sonar classification using con-volutional neural networks," in *4th International Conference on Synthetic Aperture Sonar and Synthetic Aperture Radar 2018*. Institute of Acoustics, 2018, pp. 11–22.

[8] D. P. Williams, "Acoustic-color-based convolutional neural networks for uxo classification with low-frequency sonar," in *John S. Papadakis (Hg.): UACE2019-Conference Proceedings. 5th Underwater Acoustics Conference and Exhibition. Hersonissos*, vol. 30, no. 05.07, 2019, pp. 421–428.

[9] D. A. Abraham, *Underwater Acoustic Signal Processing: Modeling, Detection, and Estimation*. Springer, 2019.

[10] M. P. Hayes and P. T. Gough, "Synthetic aperture sonar: A review of current status," *IEEE journal of oceanic engineering*, vol. 34, no. 3, pp. 207–224, 2009.

[11] R. E. Hansen, "Synthetic aperture sonar technology review," *Marine Technology Society Journal*, vol. 47, no. 5, 2013.

[12] U. Rajashekar, L. K. Cormack, and A. C. Bovik, "Image features that draw fixations," in *Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429)*, vol. 3. IEEE, 2003, pp. III–313.

[13] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[14] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[15] A. D. Matthews, T. C. Montgomery, D. A. Cook, J. W. Oeschger, and J. S. Stroud, "12.75" synthetic aperture sonar (sas), high resolution and automatic target recognition," in *OCEANS 2006*. IEEE, 2006, pp. 1–7.

[16] R. E. Hansen, H. J. Callow, T. O. Saebo, P. E. Hagen, and B. Langli, "High fidelity synthetic aperture sonar products for target analysis," in *OCEANS 2008*. IEEE, 2008, pp. 1–7.

[17] M. Zampolli, A. L. Espana, K. L. Williams, S. G. Kargl, E. I. Thorsos, J. L. Lopes, J. L. Kennedy, and P. L. Marston, "Low-to mid-frequency scattering from elastic objects on a sand sea floor: Simulation of frequency and aspect dependent structural echoes," *Journal of Computational Acoustics*, vol. 20, no. 02, p. 1240007, 2012.

[18] S. Haykin, "Array signal processing," *Englewood Cliffs*, 1985.

[19] D. H. Johnson and D. E. Dudgeon, *Array signal processing: concepts and techniques*. Simon & Schuster, Inc., 1992.

[20] M. Soumekh, *Synthetic aperture radar signal processing*. New York: Wiley, 1999, vol. 7.

[21] Y. Pailhas, Y. Petillot, and B. Mulgrew, "Increasing circular synthetic aperture sonar resolution via adapted wave atoms deconvolution," *The Journal of the Acoustical Society of America*, vol. 141, no. 4, pp. 2623–2632, 2017.

[22] W. S. Burdic, "Underwater acoustic systems analysis," *The Journal of the Acoustical Society of America*, vol. 89, no. 6, pp. 3020–3021, 1991.

[23] R. J. Urick, *Principles of underwater sound*. McGraw-Hill, New York. US, 1983.

[24] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[25] T. M. Cover, *Elements of information theory*. John Wiley & Sons, 1999.

[26] A. Papoulis, *Random Variables and Stochastic Processes*. McGraw Hill, 1965.

[27] F. Nielsen, "An information-geometric characterization of chernoff information," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 269–272, 2013.

[28] P. E. Hart, D. G. Stork, and R. O. Duda, *Pattern classification*. Wiley Hoboken, 2000.

[29] T. E. Blanford, J. D. McKay, D. C. Brown, J. D. Park, and S. F. Johnson, "Development of an in-air circular synthetic aperture sonar system as an educational tool," in *Proceedings of Meetings on Acoustics 177ASA*, vol. 36, no. 1. Acoustical Society of America, 2019, p. 070002.