

OCTOBER 14 2020

Sonar target representation using two-dimensional Gabor wavelet features

Bernice Kubicek  ; Ananya Sen Gupta  ; Ivars Kirsteins



J. Acoust. Soc. Am. 148, 2061–2072 (2020)

<https://doi.org/10.1121/10.0002168>



Sonar target representation using two-dimensional Gabor wavelet features

Bernice Kubicek,^{1,a)} Ananya Sen Gupta,^{1,b)} and Ivars Kirsteins²

¹Electrical and Computer Engineering, University of Iowa, Iowa City, Iowa 52240, USA

²Naval Undersea Warfare Center, Newport, Rhode Island, 02841, USA

ABSTRACT:

This paper introduces a feature extraction technique that identifies highly informative features from sonar magnitude spectra for automated target classification. The approach involves creating feature representations through convolution of a two-dimensional Gabor wavelet and acoustic color magnitudes to capture elastic waves. This feature representation contains extracted localized features in the form of Gabor stripes, which are representative of unique targets and are invariant of target aspect angle. Further processing removes non-informative features through a threshold-based culling. This paper presents an approach that begins connecting model-based domain knowledge with machine learning techniques to allow interpretation of the extracted features while simultaneously enabling robust target classification. The relative performance of three supervised machine learning classifiers, specifically a support vector machine, random forest, and feed-forward neural network are used to quantitatively demonstrate the representations' informationally rich extracted features. Classifiers are trained and tested with acoustic color spectrograms and features extracted using the algorithm, interpreted as stripes, from two public domain field datasets. An increase in classification performance is generally seen, with the largest being a 47% increase from the random forest tree trained on the 1–31 kHz PondEx10 data, suggesting relatively small datasets can achieve high classification accuracy if model-cognizant feature extraction is utilized. © 2020 Acoustical Society of America.

<https://doi.org/10.1121/10.0002168>

(Received 5 March 2020; revised 14 September 2020; accepted 19 September 2020; published online 14 October 2020)

[Editor: Zoi-Heleni Michalopoulou]

Pages: 2061–2072

I. INTRODUCTION

Sonar target classification continues to be an ongoing challenge due to unpredictable, varying parameters present in underwater ocean acoustics. As a sonar signal travels through the ocean, it changes and weakens due to spreading, scattering, and absorption—due to shear viscosity, density fluctuations, sound speed fluctuations, and frequency changes caused by the Doppler effect.¹ Sound speed in the ocean is dependent on temperature, salinity, and pressure,² while the sound speed profile changes with latitude, season, time of day, and weather.³ Generalized equations have been created to provide a rough estimate of the speed, but parameters such as temperature and salinity must also be estimated.⁴ The sea has irregular boundaries with motion fluctuating constantly,⁵ and reflection and scattering from the sea surface and sea bottom introduce multipath effects dependent on various sediment properties.⁶ Furthermore, target properties and location such as geometry, material, orientation, size, depth, and proud/buried status creates a nonlinear combination of the listed parameters making sonar target classification difficult.⁷

We propose a two-dimensional (2-D) Gabor wavelet-based algorithm for post-processing sonar acoustic color

images to extract highly informative features specific to a target class. The algorithm is applied to experimental field data from the Pond Experiment in March 2010⁸ (PondEx10) and the Target and Reverberation Experiment in April 2013 (TREX13).⁹ There are three main advantages to employ the 2-D Gabor wavelet. First, in image processing, it has been shown to provide data compression through separation and localization of features; this is shown true when the 2-D Gabor wavelet is convolved across the acoustic color magnitude domain. Second, in sonar reverberation modeling, two main types of models exist: (i) the point-scatterer model and (ii) the Rayleigh reverberation-envelope model (cell-scattering model).^{10,11} The latter is caused by clutter in the sea environment. If the number of scatterers in a cell is large enough, the amplitude distribution tends to be Rayleigh distributed resulting in a Gaussian distributed reverberation assuming the central limit theorem holds.¹¹ Third, the Gabor wavelets can be easily shifted in frequency and contracted or dilated making them reasonable representations for elastic target resonances. The 2-D Gabor wavelet is a Gaussian modulated with a plane wave. By tuning the 2-D Gabor bandwidth, important frequencies are captured while others are filtered out. By tuning the standard deviation of the Gaussian envelope the assumed Rayleigh distribution can also be captured. When the 2-D Gabor wavelet is properly matched to the acoustic color magnitude, highly informative features are extracted which provide an increase in classification accuracy. The remainder of the algorithm identifies

^{a)}Electronic mail: bernice-kubicek@uiowa.edu, ORCID: 0000-0001-7943-8619.

^{b)}ORCID: 0000-0001-5255-8493.

and removes small non-informative features, culminating in the creation of features we refer to as “Gabor stripes” or simply “stripes” since they manifest as striped features representing a target. Target classification has been performed using a support vector machine (SVM), random forest tree (RFT), and a feed-forward neural network (NN) to quantitatively show the increased accuracy of our feature extraction techniques across the different classifiers as well as ensure focus on our feature extraction rather than the development of novel machine learning algorithms. We show with the application of this algorithm the overall classification accuracy increases as features are more representative of their respective targets.

The organization of this paper is as follows: Sec. I provides the background motivation for this work and key contributions are described in the following paragraphs. Related research can be found in Sec. II. Data descriptions are presented in Sec. III. A mathematical description of the Gabor wavelet and optimal parameter selection is provided in Sec. IV, the weighting of the Gabor stripes is also discussed. The developed algorithm is described in Sec. V. Section VI presents the results and related discussion. Last, Sec. VII provides concluding remarks and future considerations. A quantitative validation of an algorithm parameter is shown in Appendix A and a list of acronyms used throughout are provided in Appendix B.

The novelty and archival value of this work is in our effort to combine domain knowledge and machine learning to achieve quantifiable gains in classification performance using model-cognizant feature extraction. We demonstrate using a wavelet in the form of a plane wave restricted by a Gaussian envelope leads to superior feature extraction. Our results, validated over two separate public domain field datasets, suggest relatively small datasets can achieve high classification accuracy through these informationally-rich features.

The output of the feature extraction algorithm is a compact representation that embodies the frequency-cross range signature of various targets as stripes. The feature stripes are generated by convolving an optimal 2-D Gabor wavelet across a target’s acoustic color image. The extracted feature stripes will be referred to as Gabor stripes with either a binary weighting or unique weighting. Results indicate some of the Gabor feature stripes are target aspect angle invariant, as they change shape but not their overlap relationship.

Although we present machine learning results, the focus of this work is not machine learning, but rather feature extraction and interpretation of the feature representations. Hence, we present machine learning results as a quantitative comparison between unfiltered and filtered data and present qualitative results of the Gabor stripe representation.

II. RELATED WORK

Due to the complicated nature of robust sonar target identification and classification of unexploded ordinances

(UXOs), there have been various feature extraction algorithms and machine learning techniques developed and tested on both simulated data and experimental field data. We now describe recent research^{12–17} in the field. Acoustic color images were simulated for UXOs via finite-element analysis,¹² used to train a relevance vector machine (RVM), and then tested on experimental data. In another instance, acoustic color images were processed differently,¹³ where features were generated with ellipsoid manifolds from the acoustic color images of experimental field data and used to distinguish and characterize targets using an SVM. Fischell and Schmidt¹⁴ used an SVM to classify spherical versus cylindrical bistatic scattering data from both simulated data and experimental data collected by an autonomous underwater vehicle. Robinson *et al.*¹⁵ generated a persistence diagram from synthetic aperture sonar and target-to-target distance metrics as features before hierarchical classification of experimental field data. Hall *et al.*¹⁶ demonstrated the use of fast ray model simulations to generate acoustic color images, these were used to train a variation of the matched subspace classifier (MSC) and then tested on experimental data to discriminate between UXOs and non-UXOs. Last, improved resolution for synthetic aperture sonar through compressive sampling was demonstrated by Xenaki and Pailhas.¹⁷

Gabor wavelets are a common choice for processing sonar data due to the ability to capture localized target features in the acoustic color domain. For example, Gabor-like signals were used to model the echolocation clicks of marine life prior to automated characterization.¹⁸ Others^{19–21} have designed a filter-bank of Gabor wavelets for edge detection before application of their developed algorithms due to the wavelet’s variable frequency and orientations. More specifically, a filter-bank of Gabor wavelets was used prior to an amplitude based dominant component analysis for channel selection and segmentation of a sonar image.¹⁹ Song *et al.*²⁰ employed a filter-bank of Gabor wavelets for edge feature extraction in a method to register images from a forward-looking sonar. Similarly, Zhang *et al.*²¹ extracted feature points via a filter-bank of Gabor wavelets coupled with a polar transform to register forward-looking sonar images. Other works have used a single Gabor wavelet to localize features.^{22,23} A Gabor wavelet was applied to acoustic color images of experimental field data prior to classification via a linear SVM.²² We develop a similar approach as the previous reference,²² but differ in the extraction and weighting of features as Gabor stripes. Additionally, classification is performed across all ranges of the processed images rather than a small section. Many modifications can be made when extracting the Gabor stripe representation, one specifically is discussed here, but the overarching algorithm can be tailored to other experimental field data as well.

Various machine learning techniques have been employed for target classification and source localization throughout the years. Recently, an overview relevant to acoustic data was provided by Bianco *et al.*²⁴ Of specific interest to this research are the sections regarding SVMs and NNs. Niu *et al.*²⁵ describe the same supervised machine

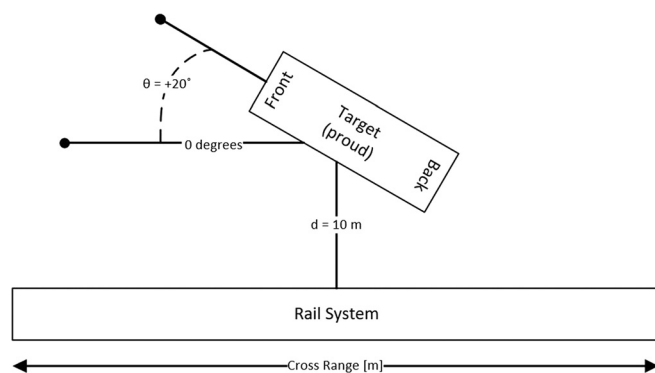


FIG. 1. Experiment geometry with respect to a target at a $+20^\circ$ aspect angle.

learning techniques explored in this paper, but for source localization rather than target classification. A Bayesian learning approach has also been used for source localization by many sonar researchers.^{26–31} Although not an extensive list, we include these to show the success of Bayesian learning in the field. Our deviation from the Bayesian approach was adopted to perform classification without prior knowledge. Online active learning techniques, in which human interaction is required to classify the decision with the highest uncertainty, was demonstrated to significantly decrease the chances of model divergence,³² specifically when paired with unsupervised clustering. We differentiate our work from these works through autonomous classification via model-based features.

III. EXPERIMENTAL SETUP AND DATA DESCRIPTION

Public domain field data from the Pond Experiment performed in 2010⁸ (PondEx10) and the Target and Reverberation Experiment (TREX13) performed in 2013⁹ were used throughout this project. In PondEx10, 11 targets were placed at different aspect angles in a fresh water pond at the Naval Surface Warfare Center in Panama City. The

University of Washington's Applied Physics Laboratory's rail system⁸ was used to collect the data. Of the 11 PondEx10 targets, we analyze four: a solid aluminum cylinder 2 ft (61 cm) long with a 1 ft (30.5 cm) inner diameter, an aluminum replica of a 4 in. artillery shell, a steel replica of a 4 in. artillery shell, and rock number 1. In TREX13, responses were collected for 27 targets, we analyze the solid aluminum cylinder, the aluminum replica of a 4 in. artillery shell, the steel replica of a 4 in. artillery shell, and a 155 mm Howitzer with collar due to insufficient rock data. All targets were proud at 10 meters with aspect angles of -80° to $+80^\circ$ in 20° increments; broadside to the rail system was considered 0° . An example of a target at a $+20^\circ$ aspect angle is shown in Fig. 1, where a cross range of 0 meters is at the midpoint of the rail system and the target is located 10 meters away from the rail system. For both datasets, the transmitted source was a 6 ms 1–31 kHz linear frequency modulated (LFM) chirp.^{8,9}

The raw field data were backscattered data from the proud targets and presented in the cross range-time domain.³³ These data were transformed into acoustic color magnitudes by taking a Blackman windowed 8192 point fast Fourier transform (FFT) of the time signal for each cross range position. An example of this transformation is shown in Fig. 2, where decibels are relative to the highest pixel value.

IV. ADAPTING THE GABOR WAVELET TO PRODUCE FEATURE STRIPES

The 2-D Gabor wavelet was generated through a 2-D Gaussian multiplied by a sinusoid. To reduce the number of variable parameters when finding the optimal Gabor wavelet, the 2-D Gaussian spatial aspect ratio and the sinusoid's orientation and phase offset were kept constant at 1, 1, and 0° , respectively. After this simplification, the 2-D Gabor wavelet reduces to Eq. (1). The dimensions of the Gabor wavelet are in pixels. A single pixel has physical units of $7.57\text{ Hz} \times 2.5\text{ cm}$ for both datasets.

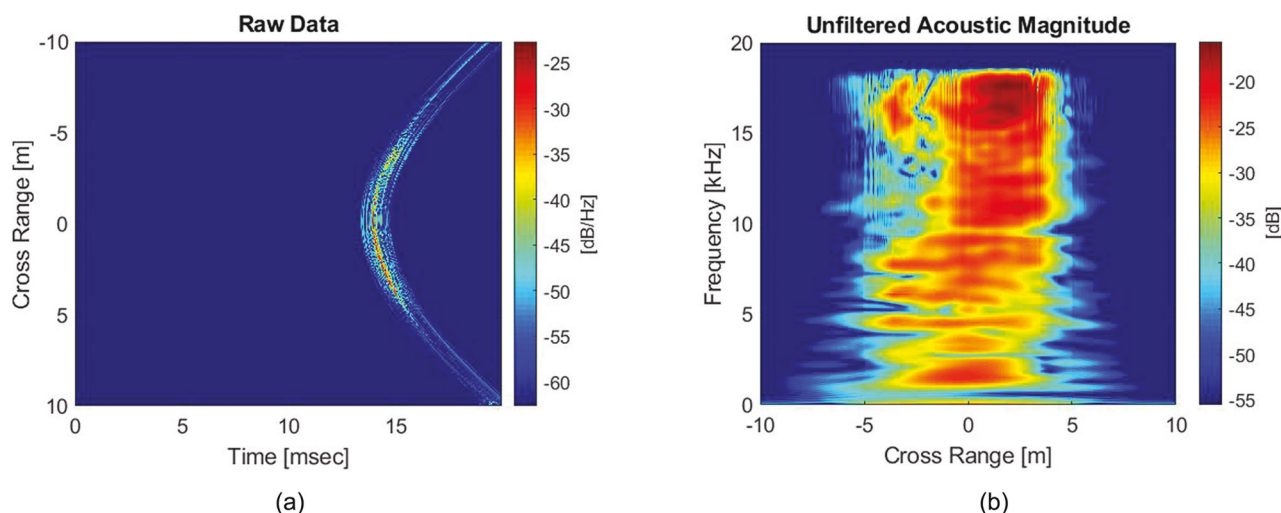


FIG. 2. Transformation from the (a) cross range-time domain to (b) an acoustic color image using a Blackman windowed 8192-FFT for the PondEx10 steel UXO target at a $+80^\circ$ aspect angle, units are with reference to the largest pixel value.

$$g(x, y; \lambda, \sigma) = \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) \exp\left(\frac{j2\pi x}{\lambda}\right), \quad (1)$$

where x denotes the spatial resolution of the cross-range axis in pixels, y denotes the spatial resolution of the frequency axis in pixels, λ represents the spatial wavelength (pixels/cycles), equivalent to the inverse of the preferred spatial frequency $1/f$, of the sinusoid, and σ represents the standard deviation of the Gaussian envelope.

The half magnitude spatial bandwidth (BW) of the Gabor wavelet is defined in Eq. (2),

$$BW = \log_2\left(\frac{\frac{\sigma\pi + \ln(2)}{\lambda} - \frac{\ln(2)}{2}}{\frac{\sigma\pi - \ln(2)}{\lambda} - \frac{\ln(2)}{2}}\right) = \log_2\left(\frac{r\pi + \frac{\ln(2)}{2}}{r\pi - \frac{\ln(2)}{2}}\right), \quad (2)$$

where the previous notation holds and r is the ratio of the standard deviation of the Gaussian envelope to the spatial wavelength, $r = \sigma/\lambda$. The half magnitude spatial bandwidth of the Gabor wavelet is measured in units of octaves apart from the preferred spatial frequency, $f = 1/\lambda$, described in Eq. (3),

$$Oct = \log_2\left(\frac{f_2}{f_1}\right) = \log_2\left(\frac{\lambda_1}{\lambda_2}\right), \quad (3)$$

where f_1 corresponds to the lower spatial frequency, f_2 corresponds to the higher spatial frequency, λ_1 corresponds to the larger spatial wavelength, and λ_2 corresponds to the smaller spatial wavelength. As the Gabor wavelet is a bandpass filter, adjusting the spatial wavelength, λ , allows for capture of informative spatial frequencies. Adjusting the standard deviation of the Gaussian, σ allows for scaling of the bandwidth of the Gabor wavelet, either loosening or tightening the focus about the specified spatial frequency. By adjusting this quantity we match the defining/scaling parameter of the Rayleigh distribution.

A. Description of feature Gabor stripes

Mathematically, the representation of an n -stripe Gabor feature of a target is given as

$$\Gamma(x, y) = \cup_{i=1}^n \psi_i(x, y), \quad (4)$$

where $\{\psi_i(x, y)\}_{i=1, \dots, n}$ denotes the i th stripe of the n -stripe group in the spectral domain, $\cup$ denotes the set theory union operator, and x and y denote the spatial resolution of the cross-range and frequency axis in pixels, respectively.

For interpretation as Gabor stripes, computationally $\psi_i(x, y)$ is equivalent to the indicator function representing a target's spectral signature $\mathcal{S}$. This ensures every stripe is considered equally weighted,

$$\psi_i(x, y) = \mathcal{I}_i = \begin{cases} 1 & (x, y) \in \mathcal{S} \\ 0 & \text{else.} \end{cases} \quad (5)$$

To distinguish unique Gabor stripes, each $\psi_i(x, y)$ is assigned a linearly spaced number in the range from 0 to 10 to ensure different weights for separated feature stripes. In terms of a weighted indicator function,

$$\psi_i(x, y) = w\mathcal{I}_i = \begin{cases} \frac{10}{n}i & (x, y) \in \mathcal{S} \\ 0 & \text{else,} \end{cases} \quad (6)$$

where $w\mathcal{I}_i$ is a weighted indicator function and n is the total number of stripes. This allows classifiers to distinguish between separated stripes.

B. Optimal Gabor wavelet parameter selection

An SVM is used to determine optimal parameters for the Gabor wavelet by applying the algorithm, described in Sec. V, with varying λ and ratio $r = \sigma/\lambda$. Results from the PondEx10 dataset are included throughout this section as a visual explanation; similar analysis has also been performed on the TREX13 dataset. Experimental field data is randomly separated into two groups, one subset is used to determine the optimal Gabor wavelet parameters and the other subset is for classification results. This was done to remove bias from optimal parameter selection from the use of all experimental data.

The various wavelengths and ratios simulated for the PondEx10 data are enumerated as: $\lambda \in \{2, 2.5, 3, 3.5, 4, 4.5, 5, 10, 15\}$ and $r \in \{0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.5, 2, 2.5, 5, 7, 10, 15, 20, 25\}$; the binary weighting [Eq. (5)] was used. From the previously listed values, the choices for λ and r are chosen empirically through examination of Fig. 3, which depicts the PondEx10 dataset's SVM overall classification accuracy versus the ratio for different λ .

The lines accented by bold in Fig. 3 represent parameters that resulted in an overall classification accuracy greater than 99%. A wavelength of $\lambda = 5.0$ and ratio $r = \sigma/\lambda = 1.2$ ($\sigma = 6$) were chosen as the optimal Gabor parameters to classify the remaining images. Similar analysis was performed on the TREX13 data, where optimal parameters

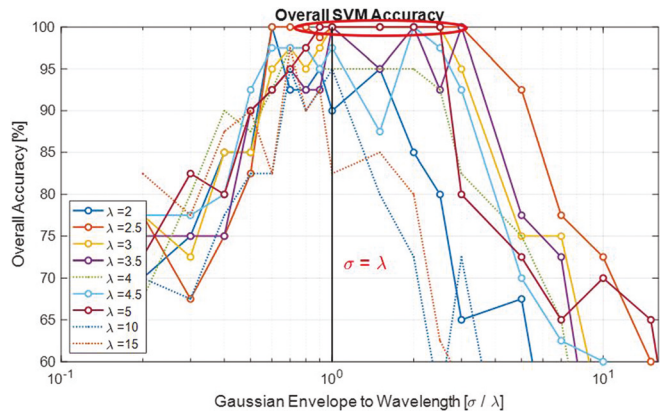


FIG. 3. Optimal PondEx10 Gabor parameter selection: SVM overall classification accuracy versus the ratio, r , as a function of λ . Lines in bold accent a classification accuracy over 99%, where $\lambda = 2, 2.5, 3, 3.5, 4.5$, and 5 meet the criteria. Parameters of $\lambda = 5$, $r = 1.2$, ($\sigma = 6$) were chosen as optimal.

were determined to be $\lambda = 10$ and $r = 0.5$ ($\sigma = 5.0$). Sub-optimal parameters of $\lambda = 5.0$ and ratio $r = \sigma/\lambda = 0.2$ ($\sigma = 1$) were used to generate a poor Gabor wavelet for the PondEx10 dataset, sub-optimal parameters of $\lambda = 10$ and $r = 0.2$ ($\sigma = 2$) were chosen for the TREX13 dataset. This was done to examine the impact of choosing an incorrectly matched wavelet for both datasets.

V. ALGORITHM DESCRIPTION

After Gabor parameter selection, we employ a three-stage technical approach described below to extract the Gabor feature stripes and perform classification.

- (i) Stage 1: Extract spectral features that incorporate domain knowledge of physical phenomena, described in Sec. V A. We achieve Gabor feature extraction through the employment of a seven-step algorithm.
- (ii) Stage 2: Project the extracted features into the stripe representation and assign weighting given in Eqs. (5) or (6).
- (iii) Stage 3: Interface with three machine learning techniques for classification of sonar targets, described in Sec. V C.

A. Stage 1—Extraction of spectral features

To achieve stage 1, we have adapted the Rayleigh reverberation-envelope model^{10,11} and developed a feature extraction technique that employs the 2-D Gabor wavelet filter to separate overlapped spectral acoustic color magnitudes. The feature extraction technique consists of a seven-step algorithm. Select steps are visually explained in Fig. 4 using the steel UXO with an $+80^\circ$ aspect angle from the PondEx10 dataset.

Step 1: Vertically augment the acoustic color magnitude images.

Step 2: Apply the 2-D Gabor wavelet as a filter to the acoustic color magnitude image via 2-D convolution, shown in Fig. 4(a).

Step 3: Generate a histogram from the magnitudes of each Gabor filtered acoustic color image. The bin width is calculated using Scott's rule described in Eq. (7). Calculate the magnitude of the 90th percentile of the histogram to be used as a threshold in Step 4. The 90th percentile was chosen empirically as it provided visually pleasing separated stripes, quantitative validation is provided in Appendix A, where an increase in overall accuracy (OA) is seen between the 85th and 90th percentile,

$$\text{bin width} = 3.5 \sigma n^{-1/3}. \quad (7)$$

Step 4: Use the magnitude of the 90th percentile as a threshold to binarize the acoustic color images.

Step 5: Label the connected pixels of the resulting image using both 4- and 8-connectivity to create objects. Here, 4-connectivity is described as pixels touching on their edges but not vertices, and 8-connectivity is described as pixels touching on their edges or vertices.³⁴ In Fig. 4(b), there are 39 connected objects.

Step 6: Cull non-informative features by removing connected objects made up of pixels with less than 10% of the maximum number of pixels per object per image. In Fig. 4(c), there are five connected objects compared to the 39 connected objects listed in step 5.

Step 7: Dimensional reduction by reducing the image size prior to feature stripe weighting. The highest and lowest frequency feature locations and the leftmost and rightmost cross range feature locations of all images were determined and used as boundaries, leaving the most amount of information in the smallest form. In Fig. 4(d), the pixels outside the red dotted lines were removed.

B. Stage 2—Feature representations/data standardization

The outputs of the algorithm were feature representations in the form of matrices, visualized throughout this manuscript as images for convenient interpretation. The feature extraction algorithm treats the data as an $n \times p$ matrix where each cell represents a pixel of the image. In stage 2, the extracted features are weighted as a binary Gabor stripes [Eq. (5)] or as unique Gabor stripes [Eq. (6)].

C. Stage 3—Classifiers

To determine the impact of the Gabor wavelet and stripe representations, classifiers are trained on the (i) original unfiltered binarized images, (ii) images produced through sub-optimal Gabor wavelet parameters, and (iii) images produced through optimal Gabor wavelet parameters. After Gabor wavelet parameter selection, the second subset of data is divided into 80% for training, and 20% for testing. The classifiers are iteratively trained and tested with results tabulated in Table I.

All classifiers are trained and tested on the same processed data but in different forms. The four-class linear SVM and RFT are trained and tested on the $n \times p$ matrices reshaped into a feature vector of a $1 \times (np)$ array. Each pixel from an image is considered a feature such that $x_i \in \mathbb{R}^{(np)}$, where x_i represents a single observation. The out of bag prediction error is used to determine the number of trees grown for the RFT, illustrated in Fig. 5. An out of bag observation refers to the remaining observations after bagging has been used to create trees (not seen during training). An elbow emerges near 80–90 trees in Fig. 5, resulting in 85 trees grown for the RFT.

The NN is trained and tested on the matrices saved as a JPEG image and is designed as a feed-forward, three-layer network for classification. This network consists of an input layer of the same size as the feature image, $x_i \in \mathbb{R}^{(np)}$. A fully connected layer follows $z_i \in \mathbb{R}^{k \times np}$. The output layer contains a softmax activation function, Eq. (8), with $y_i \in \mathbb{R}^k$,

$$y_k(x, w) = \frac{\exp(b_k)}{\sum_k \exp(b_k)}, \quad (8)$$

where y is the output, k is the number of classes (in this case 4), z is the output vector from the fully connected layer

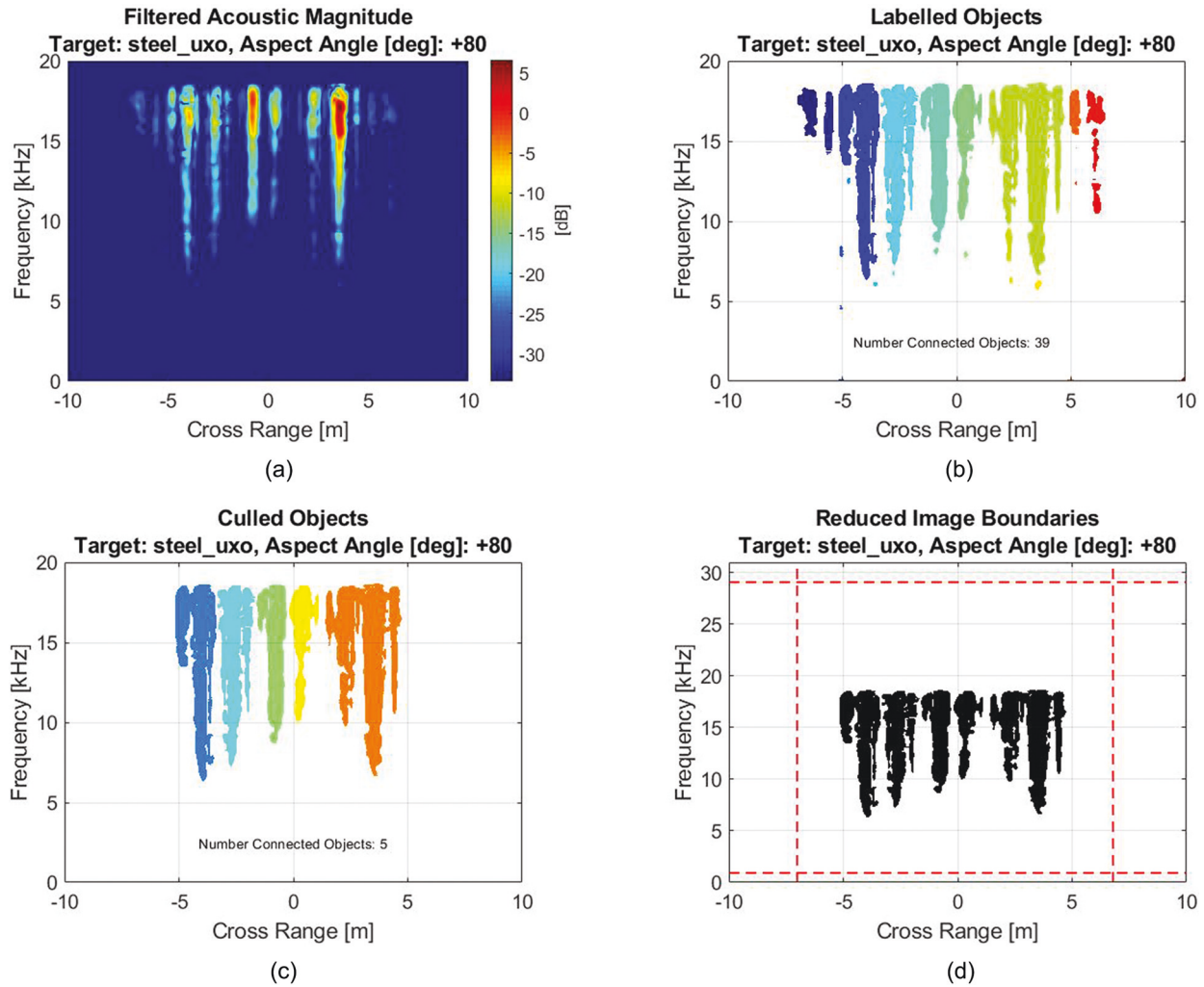


FIG. 4. Illustration of select algorithm steps: (a) Gabor filtered data, a decibel is relative to the highest pixel value, (b) labeled connected pixel objects, each color represents a different connected object. (c) Non-informative feature culling: remove connected objects with a pixel count of less than 10% of the maximum pixel count per object for an image. (d) Dimensional reduction by removing pixels that do not contain information. The red dashed line represents the boundary, where the inside square is input to the classifiers. (a)–(c) have been zoomed in to the area of interest.

input to the softmax function, and b is the bias. The stochastic gradient descent with momentum solver is used, the default value of 0.9 is used for momentum, and the learning rate is initialized at 0.01.

VI. RESULTS

There are four subsections of results, presented in such a way to complement each other, by first providing the quantitative classifier results then qualitatively explaining the impact of our feature extraction techniques for both PondEx10 and TREX13 datasets.

Section VIA provides the OA [Eq. (9)] of the classifiers when compared amongst each other for both weightings of the Gabor stripe feature representations. Here we quantitatively validate claims that the Gabor stripe feature representation increases classifier accuracy and investigate the impact of the two feature weightings [Eqs. (5) and (6)]. Section VIB visually explores the impact of the Gabor wavelet spatial bandwidth on the extracted feature stripes.

Section VIC visually investigates the separation and connections of the Gabor feature stripes. Last, the Gabor feature stripes of a target at all aspect angles for both datasets are visually compared to unfiltered extracted features in Sec. VID.

A. Overall classification accuracy for the SVM, RFT, and NN

The OA is a metric that describes, on average, how well the classifier performed. Overall classification accuracy is calculated using Eq. (9),

$$OA = \frac{\text{Total Images Correctly Classified}}{\text{Total Images Attempted}} \cdot 100\%. \quad (9)$$

The OA for the SVM, RFT, and NN trained and tested on the binary Gabor feature stripes and the uniquely weighted Gabor feature stripes are listed in Table I. An increase in OA is seen after the application of the optimal

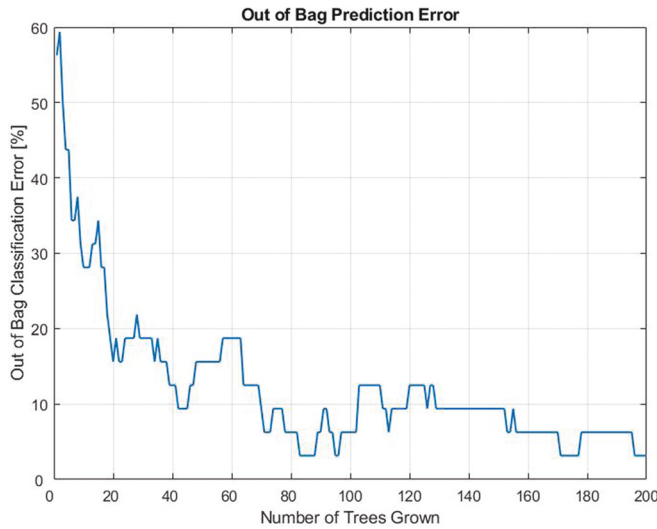


FIG. 5. Out of bag prediction error used to determine the number of trees to grow for RFT. Eighty-five trees were chosen to be grown, with an out of bag prediction error of 3.125%.

Gabor wavelet, supporting our claim of an increase in classification accuracy through the application of a physics-conscious wavelet and feature extraction technique. The increased OA when comparing the SVM to the RFT could be due to bias when selecting the optimal Gabor wavelet parameters, as the Gabor parameters were selected through using the SVM. In most cases, the largest increase in OA is seen when the features are equally weighted binary as Gabor stripes rather than uniquely weighted Gabor stripes due to the features not fully separating or connecting (see Fig. 8 and Sec. VIC). This lack of separation propagates through the uniquely weighted Gabor stripes by assigning incorrect values to Gabor stripes that should be either separated or connected, resulting in a worse OA. The binary Gabor stripe weighting inherently treats all features the same, thereby reducing the effect of incorrectly labeled Gabor stripes.

B. Impact of Gabor wavelet parameters on feature extraction

When originally determining the Gabor wavelet parameters to use throughout this investigation, an increase in the overall classification accuracy occurred where the ratio was near one (Fig. 3, p. 4), where r is the ratio of the standard deviation of the Gaussian distribution to the spatial

wavelength, $r = \sigma/\lambda$, introduced in Sec. IV B. The ratio is inversely proportional to the half magnitude spatial bandwidth of the Gabor filter. Optimal ratios were converted to half magnitude spatial bandwidths using Eq. (2), with units of octaves away from the preferred frequency, $1/\lambda$; refer to Eq. (3) for a mathematical definition of an octave. Ratios of $r=0.2$, 5 are included as extremity cases. The spatial frequency band of the Gabor wavelet was calculated through Eq. (3); both the half magnitude spatial bandwidth and spatial frequency bands are displayed in Table II.

Figures 6 and 7 show extracted features of the PondEx10 steel UXO at a $+80^\circ$ aspect angle and the TREX13 aluminum UXO with a $+60^\circ$ aspect angle, respectively. Figures 6(a) and 7(a) show the original images, Figs. 6(b) and 7(b) show the filtered images with $BW=4.941$ ($r=0.2$), and Figs. 6(c) and 7(c) filtered images with $BW=0.108$ ($r=5$). Each color represents a different Gabor stripe described by Eq. (6). At large bandwidths (small ratios), the Gaussian envelope is too large and allowed nearly all information through, shown in Table II by the large spatial frequency pass-band. This resulted in Gabor stripes seemingly “unfiltered” and mimics the original unfiltered images, thereby not disentangling the spectral overlap. At small bandwidths (large ratios), the Gaussian envelope tightens and allows little information due to the small spatial frequency pass-band shown in Table II. This has the effect of smoothing and combining Gabor stripes. At the optimal ratio or bandwidth, the Gaussian envelope allows enough information to pass through and creates distinct, separated stripes, shown for the PondEx10 data in Fig. 4(c).

C. Gabor feature stripe separation and connection

The binarized pixel groups generated through 4-connectivity were compared to those generated through 8-connectivity, Sec. VA step 5. In all cases, it is found that 99.9% of the pixels contained the same value after binarization. Since there was less than a 0.1% difference between the type of connectivity used, all results previously shown are generated using 8-connectivity.

It is found that the algorithm did not completely disentangle spectral overlap and generate individual stripes in all cases. Ideal feature separation is shown in the case of the PondEx10 steel UXO at a $+80^\circ$ aspect angle in Fig. 4(c), where distinct Gabor stripes are formed. Two non-ideal cases are seen in the PondEx10: in Fig. 8(a) the aluminum UXO at a -20° aspect angle, and in Fig. 8(b), the rock at a -20° aspect angle. For

TABLE I. Overall accuracy (OA) and for the SVM, RFT, and NN when trained and tested on the equally weighted binary Gabor feature stripes described by Eq. (5) and uniquely weighted Gabor stripes described by Eq. (6) for both the PondEx10 and TREX13 datasets. For the definition of OA, refer to Eq. (9).

Data/classifier	Original images			Sub-optimal images			Optimal images		
	SVM	RFT	NN	SVM	RFT	NN	SVM	RFT	NN
PondEx10/equal weights OA (%)	72.5	66.2	50	75	63.3	72.5	97.1	98.5	94.1
PondEx10/unique weights OA (%)	71.3	56.6	48.5	69.1	54.4	64	89	86	95.6
TREX13/equal weights OA (%)	30.9	31.7	36.8	39	36	50.7	72.8	52.2	88.9
TREX13/unique weights OA (%)	30.2	32.4	46.3	31.7	36	41.2	57.4	36.8	55.9

TABLE II. Ratios converted to spatial bandwidths with Eq. (2) and converted to lower and upper bounds of the spatial frequency band with Eq. (3).

	PondEx10 ($\lambda = 5$)			TREX13 ($\lambda = 10$)		
Ratio	0.2	1.2	5	0.2	0.5	5
Bandwidth (Oct)	4.941	0.454	0.108	4.941	1.137	0.108
Lower bounds spatial freq. (cycles/pixel)	0.036	0.171	0.193	0.018	0.067	0.096
Upper bounds spatial freq. (cycles/pixel)	1.109	0.234	0.208	0.554	0.148	0.104

the TREX13 data, two non-ideal cases are seen in Fig. 8(c) of the aluminum UXO at a -40° aspect angle and in Fig. 8(d) of the 155 mm Howitzer at a $+20^\circ$ aspect angle.

In the case of the aluminum UXO [Figs. 8(a) and 8(c)], the stripes did not properly separate and created two large Gabor stripes. After visual inspection, the circled feature stripes in Figs. 8(a) and 8(c) should be separated into three or four stripes rather than one. In the case of the rock [Fig. 8(b)], the stripes did not fully connect. The circled stripes should be one or two connected stripes rather than three. This is also seen in Fig. 8(d), where the four features should be connected.

D. Gabor feature stripes across all target aspect angles

We have claimed that the Gabor stripe representation is independent of a target's aspect angle, providing a higher classification rate when compared to the unfiltered magnitude spectra. To qualitatively validate this claim, all target aspect angles have been vertically augmented and sent through the algorithm for the unfiltered and optimally filtered data. All resulting binary features have been summed across the feature domain to create a complete response of unfiltered and optimally filtered features. These are displayed in Fig. 9, where Figs. 9(a) and 9(b) show the PondEx10 steel UXO and Figs. 9(c) and 9(d) show the TREX13 aluminum UXO.

A clear invariance is seen in the PondEx10 dataset, where the unfiltered features [Fig. 9(a)] are spread out across all frequencies and cross ranges and the filtered Gabor feature stripes [Fig. 9(b)] form distinct localized stripes. When examining the TREX13 dataset, the Gabor feature stripes

are not completely aspect angle invariant. However, localization still occurs when comparing the unfiltered features [Fig. 9(c)] to the filtered features [Fig. 9(d)].

VII. CONCLUDING REMARKS

We developed a 3-stage algorithm to represent and classify sonar acoustic color images, centered around a 2-D Gabor wavelet representative of the Rayleigh reverberation-envelope model. We determined the optimal Gabor wavelet parameters through an SVM for two sets of public domain field data. After Gabor parameter selection, three machine learning techniques were utilized for classification. Classification was performed on original binarized acoustic color images, sub-optimal Gabor parameter filtered images, and optimal Gabor parameter filtered images. All classifiers yielded higher than a 94% overall classification accuracy when PondEx10 images were generated with optimal Gabor parameters and Gabor stripes were binary weighted. Classifiers yielded higher than an 86% OA when filtered with a sub-optimal Gabor wavelet. We demonstrated improvements in machine classification accuracy due to model-informed feature extraction using the Gabor wavelet (domain-interpretable features) when compared to the original image (domain-agnostic) classification overall accuracies. We also provide a theoretic interpretation of the extracted Gabor features in terms of stripes of the targets' representation. As a target aspect angle changes, the feature Gabor stripes may change their shape, but not the overlap relationships between the stripes, providing robust target classification that is aspect angle invariant. This allows

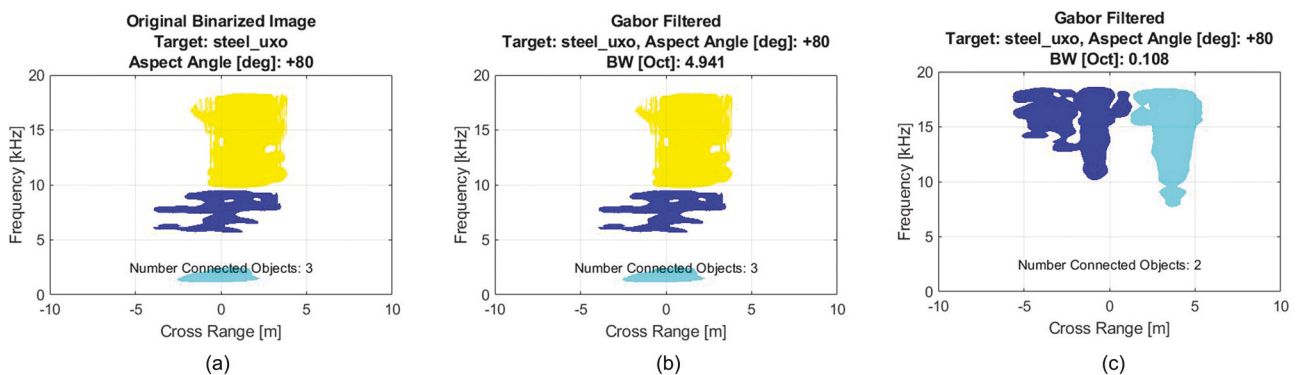


FIG. 6. Extracted features for the PondEx10 steel UXO at a $+80^\circ$ aspect angle for the (a) unfiltered, (b) filtered with $BW = 4.941$, and (c) filtered with $BW = 0.108$. Poor feature extraction is shown in (b), where the BW was too large to generate a meaningful representation. All spatial frequency components were allowed to pass through the Gabor wavelet. Poor feature extraction is also shown in (c), where the BW was too small and the spatial frequency pass-band did not allow enough information to pass through the Gabor wavelet. Ideal feature extraction is shown in Fig. 4(c) where separated Gabor stripes are formed.

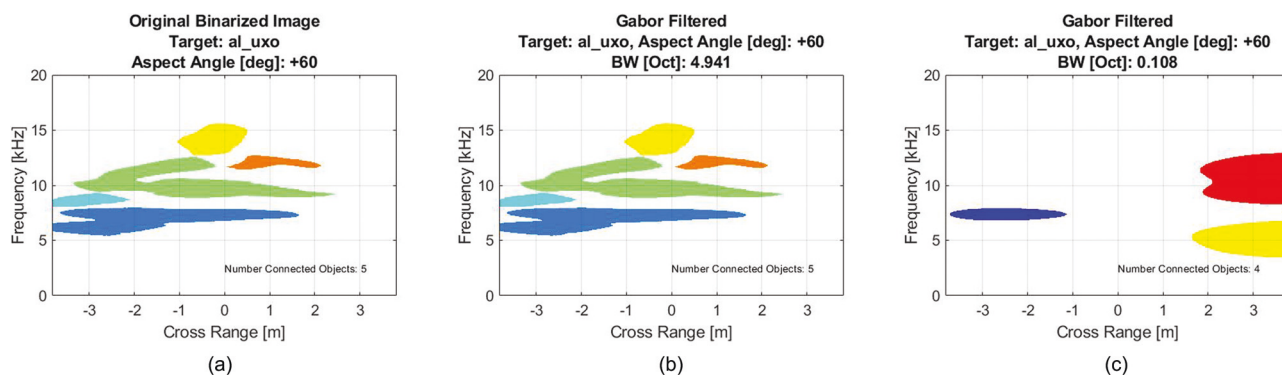


FIG. 7. Extracted features for the TREX13 aluminum UXO at a $+60^\circ$ aspect angle for the (a) unfiltered, (b) filtered with $BW = 4.941$, and (c) filtered with $BW = 0.108$. Poor feature extraction is shown in (b), where the BW was too large to generate a meaningful representation. All spatial frequency components were allowed to pass through the Gabor wavelet, resulting in the filtered image mimicking the original image, (a) and (b). Poor feature extraction is also shown in (c), where the BW was too small and the spatial frequency pass-band did not allow enough information to pass through the Gabor wavelet.

potentially overlapping sonar acoustic features with non-linear spectral morphology to be represented as binary pixel maps. We observed that taking the stripe representation at the granularity of individual stripes was counter-productive as the classification was vulnerable to incorrectly weighted

stripes. Future directions include expanding upon the stripe representation through assigning weights, studying the impact of changing additional Gabor parameters, analyzing additional public domain field data from targets with different geometries, and interfacing the Gabor stripe features

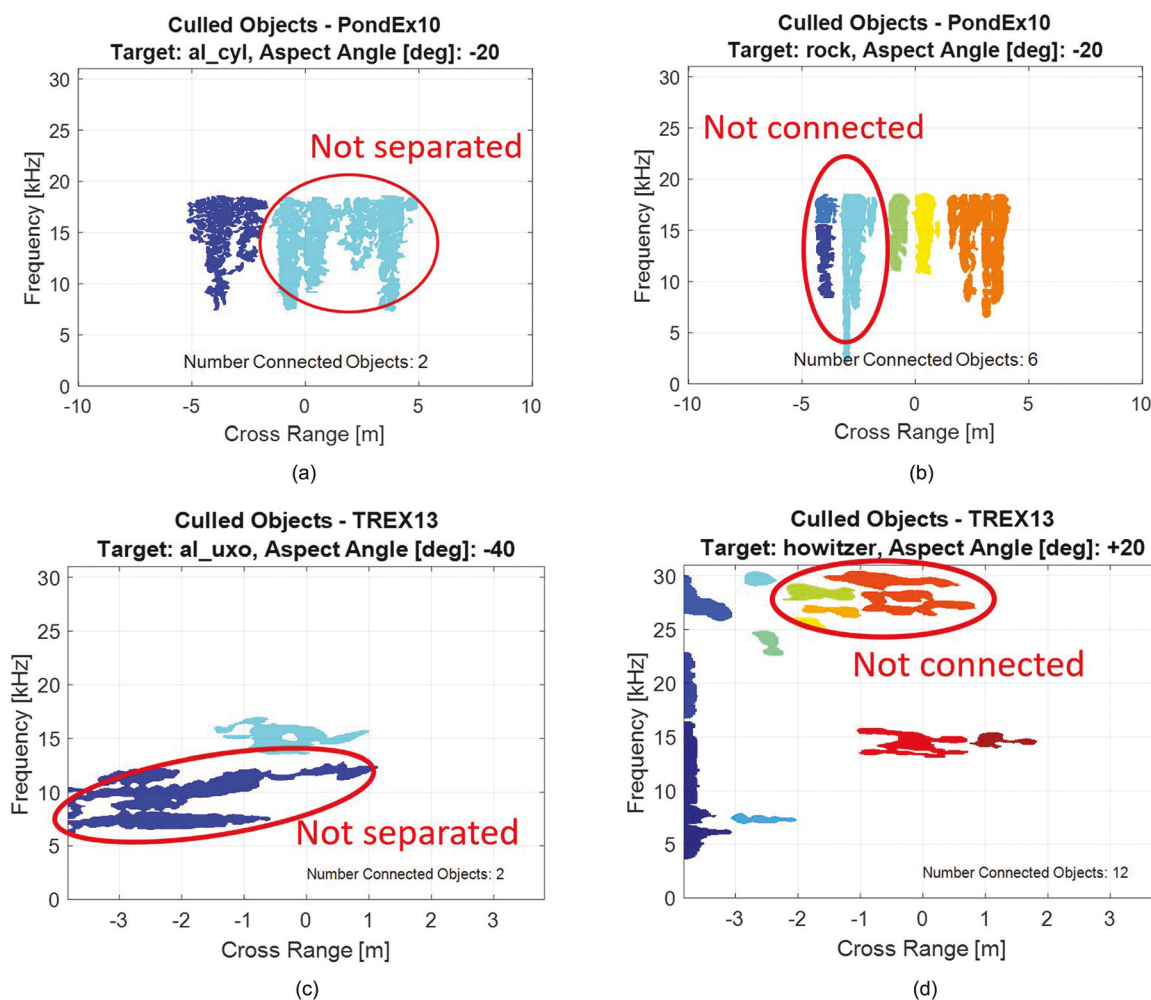


FIG. 8. Features not fully separated in the (a) PondEx10 aluminum UXO at a -20° aspect angle, and features not fully connected in the (b) PondEx10 rock at a -20° aspect angle. The red circles enclose features incorrectly formed. TREX13 non-ideal features are seen in the not separated Gabor feature stripes in the (c) aluminum UXO at a -40° aspect angle, and the features not connected in the (d) 155 mm Howitzer at a $+20^\circ$ aspect angle. Features properly connected and separated in the PondEx10 steel UXO at a $+80^\circ$ aspect angle are shown in Fig. 4(c).

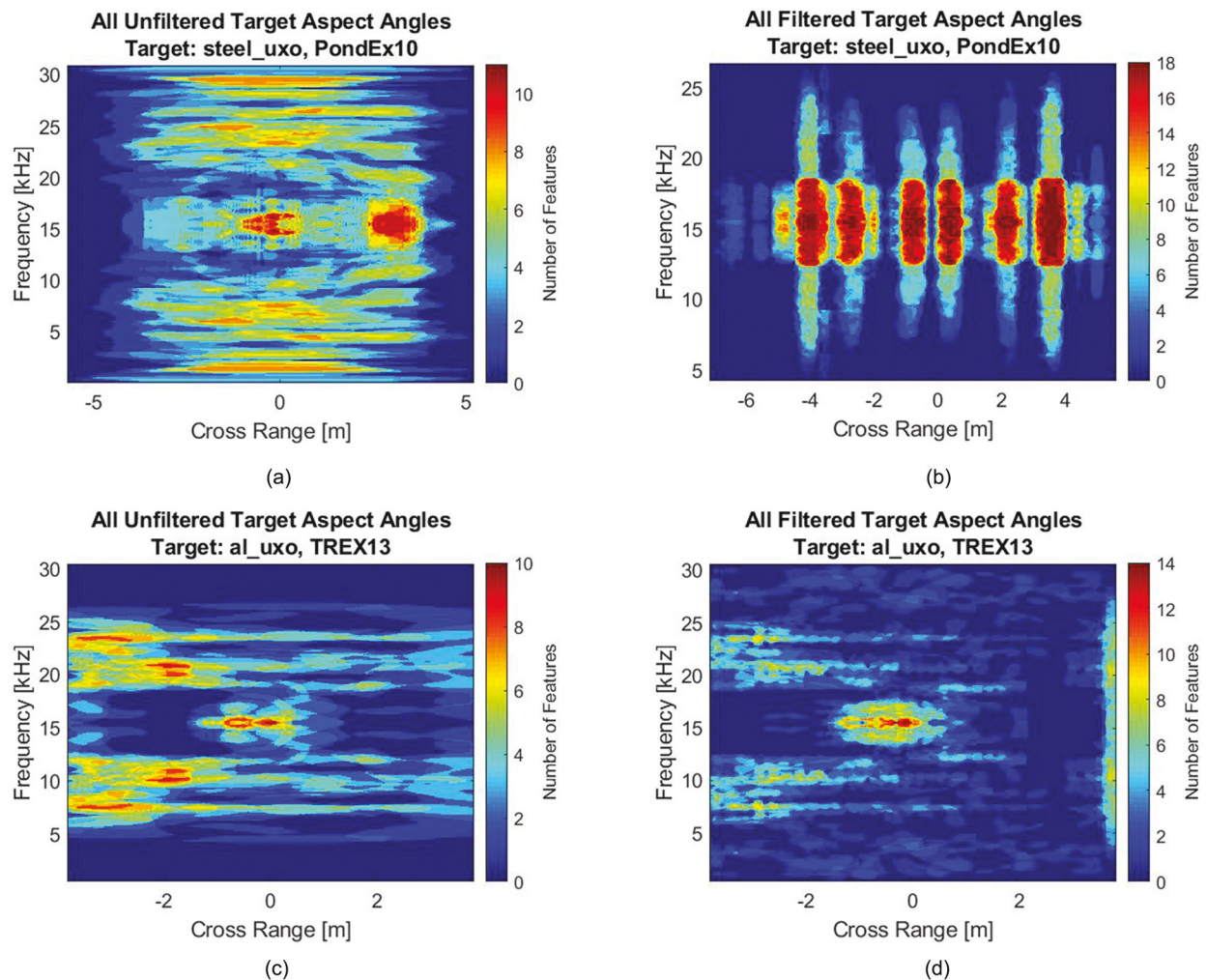


FIG. 9. Illustration of all target aspect angles after application of the feature extraction algorithm. Panels (a) and (b) show the unfiltered and filtered PondEx10 steel UXO, respectively. Panels (c) and (d) show the unfiltered and filtered TREX13 aluminum UXO. The colorbar associated with each panel shows the number of similar features present per pixel across all aspect angles where 18 is the maximum number of features possible. The horizontal line of symmetry seen near 15 kHz is due to the vertical augmentation of the unfiltered response in step 1 of the feature extraction algorithm.

with more sophisticated machine learning techniques, such as active learning generative adversarial networks.

ACKNOWLEDGMENTS

The authors would like to thank the Office of Naval Research for funding under Grant No. N00014-19-1-2436 and the National Science Foundation for funding under Grant No. 1808463. We would also like to thank Ryan McCarthy of the University of Iowa for his invaluable help and thoughtful insight throughout this research.

APPENDIX A: VALIDATION OF HISTOGRAM PERCENTILE

Throughout the optimal parameter selection in Sec. IV B, the 90th percentile of the Gabor acoustic color magnitude histogram was used as a threshold to binarize the image prior to feature culling and classification. This percentile was empirically chosen through visually examining the Gabor stripes to ensure they were fully separated and

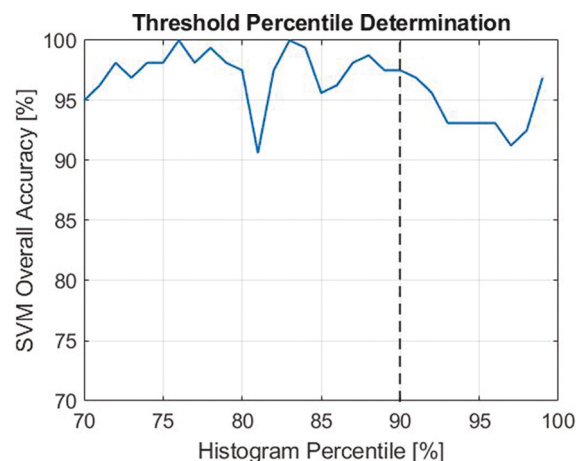


FIG. 10. SVM overall classification accuracy as a function of the histogram percentile used as a threshold to binarize the image. The black dotted line indicates the 90th percentile used in determining the optimal Gabor parameters.

connected (discussed in Sec. VIC). To quantitatively validate this choice, the PondEx10 optimal Gabor parameters ($\lambda = 5$, $r = \sigma/\lambda = 0.8$, $\sigma = 4$) were kept constant while the histogram percentile, p , was varied such that $p \in \{70, 71, \dots, 98, 99\}$ on the first subset of data. The resulting data was sent through the SVM classifier, where the OA was used to validate the percentile, shown in Fig. 10.

APPENDIX B: ACRONYMS

TABLE III. List of acronyms and their meanings.

Acronym	Full description
2-D	Two-dimensional
al_cyl	Aluminum cylinder target (from PondEx10 and TREX13 public domain field data)
al_uxo	Aluminum UXO target (from PondEx10 and TREX13 public domain field data)
dB	Decibel
NN	Neural network
OA	Overall accuracy [see Eq. (9) for a mathematical definition]
Opt.	Optimal (refers to optimal Gabor parameters used to generate images prior to training and testing of classifiers)
Orig.	Original (refers to the original images used to train and test classifiers)
PDF	Probability density function
PondEx10	Pond experiment performed March 2010
RFT	Random forest tree
rock	Rock 1 target (from PondEx10 public domain field data)
steel_uxo	Steel UXO target (from PondEx10 and TREX13 public domain field data)
Sub-opt.	Sub-optimal (refers to sub-optimal Gabor parameters used to generate images prior to training and testing of classifiers)
SVM	Support vector machine
UXO	Un-exploded ordnance

¹X. Lurton, *An Introduction to Underwater Acoustics: Principles and Applications* (Springer Science & Business Media, New York, 2002).

²C. C. Leroy, "Development of simple equations for accurate and more realistic calculation of the speed of sound in seawater," *J. Acoust. Soc. Am.* **46**(1B), 216–226 (1969).

³S. M. Flatté and F. D. Tappert, "Calculation of the effect of internal waves on oceanic sound transmission," *J. Acoust. Soc. Am.* **58**(6), 1151–1159 (1975).

⁴V. A. Del Grosso, "New equation for the speed of sound in natural waters (with comparisons to other equations)," *J. Acoust. Soc. Am.* **56**(4), 1084–1091 (1974).

⁵W. I. Roderick and B. F. Cron, "Frequency spectra of forward-scattered sound from the ocean surface," *J. Acoust. Soc. Am.* **48**(3B), 759–799 (1970).

⁶K. L. Williams and D. R. Jackson, "Bistatic bottom scattering: Model, experiments, and model/data comparison," *J. Acoust. Soc. Am.* **103**(1), 169–181 (1998).

⁷R. J. Urlick, *Principles of Underwater Sound for Engineers* (Tata McGraw-Hill Education, New York, 1967).

⁸S. G. Kargl and K. L. Williams, "Serdp project mr-1665" (2012).

⁹S. G. Kargl, "Serdp project mr-2231-fr" (2015).

¹⁰P. C. Etter, *Underwater Acoustic Modeling and Simulation* (CRC Press, Boca Raton, FL, 2018).

¹¹D. A. Abraham and A. P. Lyons, "Simulation of non-rayleigh reverberation and clutter," *IEEE J. Oceanic Eng.* **29**(2), 347–362 (2004).

¹²J. A. Bucaro, Z. J. Waters, B. H. Houston, H. J. Simplotson, A. Sarkissian, S. Dey, and T. J. Yoder, "Acoustic identification of buried underwater unexploded ordnance using a numerically trained classifier (1)," *J. Acoust. Soc. Am.* **132**(6), 3614–3617 (2012).

¹³D. Schupp, A. S. Gupta, and I. Kirsteins, "Characterization and classification of sonar targets using ellipsoid features," in *Proceedings of the 2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Orlando, FL (December 14–16, 2015), pp. 1352–1356.

¹⁴E. M. Fischell and H. Schmidt, "Classification of underwater targets from autonomous underwater vehicle sampled bistatic acoustic scattered fields," *J. Acoust. Soc. Am.* **138**(6), 3773–3784 (2015).

¹⁵M. Robinson, S. Fennell, B. DiZio, and J. Dumiak, "Geometry and topology of the space of sonar target echos," *J. Acoust. Soc. Am.* **143**(3), 1630–1645 (2018).

¹⁶J. J. Hall, M. R. Azimi-Sadjadi, S. G. Kargl, Y. Zhao, and K. L. Williams, "Underwater unexploded ordnance (UXO) classification using a matched subspace classifier with adaptive dictionaries," *IEEE J. Oceanic Eng.* **44**(3), 739–752 (2019).

¹⁷A. Xenaki and Y. Pailhas, "Compressive synthetic aperture sonar imaging with distributed optimization," *J. Acoust. Soc. Am.* **146**(3), 1839–1850 (2019).

¹⁸S. Madhusudhana, A. Gavrilov, and C. Erbe, "Automatic detection of echolocation clicks based on a Gabor model of their waveform," *J. Acoust. Soc. Am.* **137**(6), 3077–3086 (2015).

¹⁹Y. Attaf, A. O. Boudraa, and C. Ray, "Amplitude-based dominant component analysis for underwater mines extraction in side scans sonar," in *Proceedings of OCEANS 2016*, Shanghai, China (April 10–13, 2016), pp. 1–4.

²⁰S. Song, J. M. Herrmann, K. Liu, S. Li, and X. Feng, "Forward-looking sonar image mosaicking by feature tracking," in *2016 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, Qingdao, China (December 3–7, 2016), pp. 1613–1618.

²¹J. Zhang, F. Soheli, H. Bian, M. Bennamoun, and S. An, "Forward-looking sonar image registration using polar transform," in *Proceedings of OCEANS 2016 MTS/IEEE*, Monterey, CA (September 19–23, 2016), pp. 1–6.

²²A. S. Gupta and D. Schupp, "Characterization of sonar target data using Gabor wavelet features," in *2015 49th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA (November 8–11, 2015), pp. 1723–1726.

²³A. S. Gupta and I. Kirsteins, "Disentangling sonar target features using braided feature graphs," in *Proceedings of OCEANS 2017*, Anchorage, AK (September 18–21, 2017), pp. 1–5.

²⁴M. J. Bianco, P. Gerstoft, J. Traer, E. Ozanich, M. A. Roch, S. Gannot, and C.-A. Deledalle, "Machine learning in acoustics: Theory and applications," *J. Acoust. Soc. Am.* **146**(5), 3590–3628 (2019).

²⁵H. Niu, E. Reeves, and P. Gerstoft, "Source localization in an ocean waveguide using supervised machine learning," *J. Acoust. Soc. Am.* **142**(3), 1176–1188 (2017).

²⁶L. Guillon, S. E. Dosso, N. R. Chapman, and A. Drira, "Bayesian geoaoustic inversion with the image source method," *IEEE J. Oceanic Eng.* **41**(4), 1035–1044 (2016).

²⁷A. Xenaki, E. Fernandez-Grande, and P. Gerstoft, "Block-sparse beamforming for spatially extended sources in a bayesian formulation," *J. Acoust. Soc. Am.* **140**(3), 1828–1838 (2016).

²⁸C. Chen, A. Zare, H. N. Trinh, G. O. Omotara, J. T. Cobb, and T. A. Lagaunne, "Partial membership latent dirichlet allocation for soft image segmentation," *IEEE Trans. Image Process.* **26**(12), 5590–5602 (2017).

²⁹D. Bush and N. Xiang, "A model-based bayesian framework for sound source enumeration and direction of arrival estimation using a coprime microphone array," *J. Acoust. Soc. Am.* **143**(6), 3934–3945 (2018).

³⁰A. C. Barros and P. J. Gendron, "A computational bayesian approach for localizing an acoustic scatterer in a stratified ocean environment," *J. Acoust. Soc. Am.* **146**(3), EL245–EL250 (2019).

³¹Z.-H. Michalopoulou, A. Pole, and A. Abdi, "Bayesian coherent and incoherent matched-field localization and detection in the ocean," *J. Acoust. Soc. Am.* **146**(6), 4812–4820 (2019).

- ³²E. Kriminger, J. T. Cobb, and J. C. Príncipe, “Online active learning for automatic target recognition,” *IEEE J. Oceanic Eng.* **40**(3), 583–591 (2015).
- ³³K. L. Williams, S. G. Kargl, E. I. Thorsos, D. S. Burnett, J. L. Lopes, M. Zampolli, and P. L. Marston, “Acoustic scattering from a solid aluminum

cylinder in contact with a sand sediment: Measurements, modeling, and interpretation,” *J. Acoust. Soc. Am.* **127**(6), 3356–3371 (2010).

- ³⁴R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*, Vol. 1 (Addison-Wesley, Reading, UK, 1992), pp. 28–30.