Dark-spot segmentation for oil spill detection based on multifeature fusion classification in single-pol synthetic aperture radar imagery

Haitao Lang
Xingyao Zhang
Yuyang Xi
Xi Zhang
Wei Li

Dark-spot segmentation for oil spill detection based on multifeature fusion classification in single-pol synthetic aperture radar imagery

Haitao Lang,a,b,* Xingyao Zhang,a Yuyang Xi,a Xi Zhang,c and Wei Li,d

aBeijing University of Chemical Technology, Department of Physics and Electronics, No. 15 Beisanhuan East Road, Beijing 100029, China
bBeijing University of Chemical Technology, Beijing Key Laboratory of Environmentally Harmful Chemicals Analysis, No. 15 Beisanhuan East Road, Beijing 100029, China
cThe First Institute of Oceanography SOA, Laboratory of Remote Sensing, No. 6 Xianxialing Road, Qingdao 266061, China
dBeijing University of Chemical Technology, College of Information Science & Technology, Beijing 100029, China

Abstract. In recent years, oil spill surveillance with space-borne synthetic aperture radar (SAR) has received unprecedented attention and has been gradually developed into a common technique for maritime environment protection. A typical SAR-based oil spill detection process consists of three steps: (1) dark-spot segmentation, (2) feature extraction, and (3) oil spill and look-alike discrimination. As a preliminary task in the oil spill detection process chain, dark-spot segmentation is a critical and fundamental step prior to feature extraction and classification, since its output has a direct impact on the two subsequent stages. The balance between the detection probability and false alarm probability has a vital impact on the performance of the entire detection system. Unfortunately, this problem has not drawn as much attention as the other two stages. A specific effort has been placed on dark-spot segmentation in single-pol SAR imagery. A combination of fine designed features, including gray features, geometric features, and textural features, is proposed to characterize the oil spill and seawater for improving the performance of dark-spot segmentation. In the proposed process chain, a histogram stretching transform is incorporated before the gray feature extraction to enhance the contrast between possible oil spills and water. A simple but effective multiple-level thresholding algorithm is developed to conduct a binary classification before the geometric feature extraction to obtain more accurate area features. A local binary pattern code is computed and assigned as the textural feature for a pixel to characterize the physical difference between oil spills and water. The experimental result confirms that the proposed fine designed feature combination outperforms existing approaches in both aspects of overall segmentation accuracy and the capability to balance detection probability and false alarm probability. It is a promising alternative that can be incorporated into existing oil spill detection systems to further improve system performance. © 2017 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JRS.11.015006]

Keywords: oil spill surveillance; synthetic aperture radar; remote sensing; marine pollution.

Paper 16593 received Aug. 22, 2016; accepted for publication Dec. 20, 2016; published online Jan. 12, 2017.

1 Introduction

Oceans, covering approximately 71% of Earth’s surface and 90% of Earth’s biosphere, are crucial to human activity. Human health and well-being are tied to the vitality of the global ocean and coastal systems.1 With the expansion of global industrialization, there has been an explosive increase in the scope of incidents of marine pollution, resulting in more and more pollutants
being discharged into the sea. Oil spills are among the most lethal threats to the marine environment since they reduce sunlight penetration through the water, thereby limiting the photosynthesis of marine plants and phytoplankton and causing tremendous destruction in marine ecosystems. In April 2010, the explosion and sinking of British Petroleum’s Gulf of Mexico Deepwater Horizon oil platform resulted in an unprecedented environmental disaster, with 60,000 barrels (8200 metric tons) of oil per day issuing from the well for 87 days. The accident has had serious effects on vulnerable maritime species, wildlife habitats, fishing activity in the Gulf, coastal ecology, and the tourism industry. However, such dramatic accidents occur infrequently and represent only a small fraction of the pollution problem at sea. Actually, about half of all oil spills in the marine environment are operative discharges from ships, and in most of these cases, the discharge is intentional and illicit; this has become a more dangerous threat to maritime environments.

Efficient monitoring and early warning are essential for preventing widespread damage from intentional or unintentional oil release and for reducing its adverse impact on ecosystems. In recent years, oil spill pollution surveillance is being regularly performed on an increasing scale worldwide. Through the use of modern remote sensing instrumentation, oil spills on the open sea can be monitored on a 24-h basis. Compared with optical remote sensing (visual, infrared, and near-infrared) and laser fluorescent sensors, because of the all-weather and day-and-night operational capability, synthetic aperture radar (SAR) is now the most common means of oil spill remote sensing. The combined use of space-borne SAR images and aircraft surveillance flights is a cost-effective way to monitor oil spills and catch polluters in large ocean areas.

Radar backscatter energy from the ocean surface is governed mainly by Bragg scattering of gravity-capillary waves, producing a “bright” image. In the SAR image, since oil film dampens capillary waves, and hence decreases the backscatter, an oil spill appears as a “dark” spot surrounded by the spill-free surface. Thus, essentially speaking, oil spill detection can be regarded as a dark-spot detection problem in SAR imagery. A typical SAR-based oil spill detection process consists of three steps: (1) dark-spot segmentation, (2) feature extraction, and (3) oil spill and look-alike discrimination. The dark-spot segmentation step highlights all oil spill candidates over the detection area. Then, a vector of features that quantitatively describes the dark spot is generated for the subsequent oil spill/look-alike discrimination step. Finally, the oil spills are distinguished from the look-alikes (the look-alikes are essentially natural phenomena that can also dampen short waves and create dark patches on the sea surface, including low wind areas, areas sheltered by land, rain cells, organic films, grease ice, wind fronts, up-welling zones, oceanic fronts, algae blooms, and current shear zones) based on appropriate classification techniques.

Among the above process chains, extensive efforts have been placed on extracting features and discriminating the oil spills from look-alikes, with encouraging results. In contrast, the dark-spot segmentation step, often regarded as an initial process, has not drawn as much attention as the subsequent two stages in previous studies. Actually, as a preliminary task in the oil spill detection process chain, dark-spot segmentation is a critical and fundamental step prior to the subsequent steps since its output has a direct impact on feature extraction and discrimination. If the dark-spot segmentation has relatively low detection probability, it will lead to too many missed possible oil spills, causing poor performance of the oil spill detection system. In contrast, if the dark-spot segmentation yields too many false alarms, i.e., too much seawater is segmented as oil spill candidates, it will significantly increase the workload of subsequent feature extraction and discrimination. Thus, balancing the detection probability and false alarm probability is a critical problem to be carefully addressed in the dark-spot segmentation step. Unfortunately, there has been little effort to specifically address this problem.

In the existing literature, the threshold-based algorithm is a common approach used to carry out dark-spot segmentation because of its computational efficiency. Chang et al. applied a simple global threshold calculated by the moment-preserving method to conduct a binary classification of image pixels. Solberg et al. applied a local adaptive threshold algorithm calculated using the ratio of the standard deviation to the mean value in the local window for dark-spot segmentation. This technique has proven to be both robust and simple to implement. Ganta et al. proposed an adaptive level set threshold based on the illumination and reflectance
components to address the problem that the presence of glitter induces speckles in the SAR imagery. Shu et al.\textsuperscript{16} made use of a spatial density threshold to differentiate the dark spot from the ocean background. Traditional thresholding-based methods of dark-spot segmentation are mostly dependent on gray information. Recently, new effort is being made to further improve dark-spot segmentation performance by combining multiple features; it has been widely acknowledged in the remote sensing and image processing community that properly combining multiple features always results in good classification/segmentation/recognition performance. Significant research effort has focused on various fusion strategies, including feature-level fusion, decision-level fusion, etc.\textsuperscript{17–22} In the dark-spot segmentation task, Feng et al.\textsuperscript{23} jointly used gray and edge-based segmentation to extract oil spills from the complex sea scene. Singha et al.\textsuperscript{10} combined the pixel intensity value, the mean intensity, and the standard deviation of a $3 \times 3$ pixel-centered neighbor area to generate a new vector and used it as the input neurons to an artificial neural networks (ANN). Those multiple features fusion-based methods presented a positive impact on dark-spot segmentation.

In this paper, we focus on fusing multiple features to improve the performance of dark-spot segmentation in single-pol SAR images. A new combination of fine designed features, including gray features, geometric features, and textural features, is proposed to characterize the oil spill and seawater for improving the performance of dark-spot segmentation. The experimental result confirms that the proposed fine designed feature combination outperforms existing approaches in both aspects of overall segmentation accuracy and the capability to balance detection probability and false alarm probability. It is a promising alternative that can be incorporated into existing oil spill detection systems to further improve system performance.

The major contribution of this study is twofold:

- A specific effort has been placed on dark-spot segmentation by balancing the detection probability and false alarm probability in single-pol SAR imagery. Dark-spot segmentation is a critical and fundamental step prior to feature extraction and classification. The balance between the detection probability and false alarm probability has a vital impact on the performance of the entire detection system. Unfortunately, this problem has not drawn as much attention so far. In addition, although the full-pol SAR imagery can provide more information about oil spills and seawater to improve the segmentation of dark spots, this method has restrictive system requirements, i.e., full-pol SAR imagery is needed, and is not operationally available with most of the satellite sensors. For example, TerraSAR-X and COSMO-SkyMed can only provide dual-pol images for oil spill monitoring. Hence, it is necessary to develop the dark-spot segmentation method based on the single-pol SAR image for real-world applications.

- We develop three types of fine designed image features to characterize oil spills and seawater, including: (1) gray features ($f_1$), which significantly enhance the contrast between oil spills and seawater by a histogram stretching transform; (2) geometric features ($f_2$), extracted from the binary image classified by the proposed multilevel thresholding algorithm, which reflect the difference in covered area between oil spills and seawater; and (3) textural features ($f_3$), which characterize the different appearance of oil spills and seawater in the SAR image.

The rest of this paper is organized as follows. Section 2 introduces the study area and the dataset used to validate the proposed method. Section 3 provides a detailed description of the proposed feature extraction and fusion method and corresponding process chain. Section 4 presents the comparative experimental results. Section 5 concludes the paper.

2 Study Area and Dataset

In this study, the oil spill accident in the Gulf of Mexico that occurred on April 20, 2010, and the Prestige oil spill accident that occurred on November 13, 2002, were used to validate the proposed method. The Gulf of Mexico oil spill began on April 20, 2010, and flowed for 87 days, until it was capped on July 15, 2010. It is considered the largest accidental marine oil spill in the history of the petroleum industry, with a total estimated discharge of 4.9 million barrels. The Prestige oil spill off Galicia in northwestern Spain was caused by the sinking of the oil tanker...
The MV Prestige spill in 2002. The spill polluted thousands of kilometers of coastline and more than one thousand beaches on the Spanish, French, and Portuguese coast, as well as causing great harm to the local fishing industry. The spill is the largest environmental disaster in the history of both Spain and Portugal.

Both test datasets were acquired by ENVISAT operating in C band ScanSAR Wide Swath Mode with vertical transmit-vertical receive (VV) polarization at a resolution of 150 m, on May 2, 2010, over the Gulf of Mexico and on November 17, 2002, over Galicia in northwestern Spain. Generally speaking, VV polarization can yield better results and C band radar can yield good oil spill imagery. Figures 1(a) and 1(b) illustrate the corresponding SAR images, and the white rectangles in each image show the studied regions of interest (ROI), with a size of 1200 × 1200 for both. Figure 2 shows the enlarged ROI regions for our study, which are termed the “Mexico” data and “Prestige” data, respectively. The ability of radar to detect oil is limited by sea state. Too low sea state will not produce enough sea clutter in the surrounding to contrast with an oil spill, and too high sea state will also block detection. Accepted wind speed limits are 1.5...
to 10 m/s. In our test, the wind speed for Mexico and Prestige was 6 to 8 m/s and 3 to 7 m/s (rough estimates), respectively, which make the oil spill detectable, as shown in Fig. 2.

3 Methodology

As illustrated in Fig. 3, the proposed dark-spot segmentation algorithm is comprised of four main steps: (1) the preprocessing step aims to conduct radiometric calibration and geometric correction and remove the speckle and noise in the original SAR image by successively using a median filter and mean filter. (2) In the feature extraction step, three types of fine designed features, including gray features, geometric features, and textural features, are extracted from a single-pol SAR image. A histogram stretching algorithm is applied to enhance the contrast between oil spill candidates and seawater before gray feature extraction. An adaptive multilevel thresholding algorithm is proposed to conduct a binary classification before the geometric feature extraction. The textural feature is characterized by a local binary pattern (LBP) code.

(3) In the segmentation step, three types of features are fused to a representation vector and used as input to a three-layer perceptron neural network classifier to conduct dark-spot detection under a supervised classification framework. (4) Finally, morphology processing is used to reduce the unreasonable results of the preceding binary classification.

3.1 Synthetic Aperture Radar Image Preprocessing

The preprocessing is designed to compensate for the trend of radar backscatter from the sea induced by the increase in incidence angle at long range and remove speckle and noise. In this paper, the NEST toolbox is applied to conduct the radiometric calibration and geometric correction to the original SAR data. After that, a simple but effective hierarchical filtering method is applied. First, a median filter with a $3 \times 3$ window according to the image resolution is used to reduce the high-valued noise and speckle pixels while preserving the edge information of objects. Then, a mean filter with the same window size is used to give appropriate smoothing to the SAR image to reduce the amount of intensity variation intraclasses.

3.2 Feature Extraction

3.2.1 Gray features

The intensity of the pixel is the most commonly used gray feature to distinguish oil spills from the free sea surface. Because oil spills decrease the sea surface gravity-capillary waves and hence decrease the backscatter of related areas, the oil spills appear as dark spots, while the surrounding sea surface remains relatively bright in the SAR image. This physical characteristic makes it
possible to segment oil spills from the sea surface, based on the intensity of pixels in the SAR image. However, because the absolute value of backscattering from both the sea surface and oil spill is relatively low compared with other maritime targets on the sea, the dynamic range of pixel values is relatively narrow; thus, the brightness contrast between oil spills and seawater is not enough to determine a simple segmentation threshold. To address this problem, in this paper, a histogram stretching transformation is incorporated to extend the dynamic range of intensity values before the gray feature extraction. As shown in Fig. 4(a), a histogram of an image can tell us about the data distribution with respect to image gray levels. The idea of contrast enhancement using histogram stretching transformation is the mapping of the range of gray-level values in the original image to a new range. For example, if the image displayed using the histogram in Fig. 4(a) appears dark due to the majority of pixel gray-level values are between $x_1$ and $x_2$, we can linearly stretch the histogram to transform the gray-level range $(x_1, x_2)$ in Fig. 4(a) to a new gray-level range $(x'_1, x'_2)$ in Fig. 4(c) by a piece-wise linear transformation, as shown in Fig. 4(b). The mathematical form is

$$x' = \begin{cases} 
    a_1 \times x, & 0 \leq x \leq x_1 \\
    a_2 \times (x - x_1) + x'_1, & x_1 < x \leq x_2, \\
    a_3 \times (x - x_2) + x'_2, & x_2 < x \leq x_3.
\end{cases} \quad (1)$$

![Fig. 4](image.png) **Fig. 4** Image contrast enhancement by histogram stretching transformation. (a) the original histogram, (b) piece-wise linear transformation, and (c) histogram after adjustment.

![Fig. 5](image.png) **Fig. 5** Example of SAR image after filtering and histogram stretching transform. Top lines, from left to right are the original Mexico data and the images after filtering and histogram stretching, respectively. Histogram statistics is shown in the lower right corner of each subimage. The bottom line shows the corresponding results for the Prestige data.
under the condition that \( \int_{0}^{x} pdf(x) = \int_{0}^{x'} pdf(x') = 1 \). The examples shown in Fig. 5 prove that the histogram stretching transform can efficiently enhance the contrast between the oil spill and seawater by enlarging the dynamic range. Therefore, before extracting the gray feature \( f_1 \), we apply a histogram stretching transform to the filtered image.

### 3.2.2 Geometric features

Previous studies have shown that oil spills can be clustered according to their geometric features, such as area, perimeter, and shape.\(^{26,27}\) Considering the complexity of the calculation, in this paper, the area feature \( f_2 \) of connected regions is applied to characterize the oil spills and seawater. To extract accurate area features, the SAR image after preprocessing is first segmented by the proposed multilevel thresholding algorithm to generate a binary image. Since the intensity distributions of both the oil spill and seawater are not homogeneous, it is impossible to obtain a proper segmentation by a single global threshold.\(^{14,16}\) The local Otsu algorithm\(^{28}\) can yield satisfactory binary classification results, but it involves much computation time since it conducts an exhaustive search to evaluate the criterion for maximizing the between-class variance. In this paper, we propose a simple and effective multilevel thresholding algorithm to carry out the binary classification before geometric feature extraction. The proposed method is efficient because its computation depends only on the average and standard derivation. The pseudocode is described as Algorithm 1.

The proposed multilevel thresholding algorithm is based on applying an adaptive hierarchical division strategy to the image. For the \( i \)'th division, named the \( i \)'th level in the proposed algorithm, the candidate threshold is calculated for each divided subimage first based on the simple average and standard derivation statistics. If the calculated threshold is smaller than the current threshold [calculated in the preceding \((i - 1)\)'th level], the current threshold is retained and division of this subimage is stopped. Otherwise, the current threshold is updated, and this subimage

### Algorithm 1 Proposed multilevel thresholding algorithm.

**Initialization:**

Set \( i = 0; \ i_1 = I; \ i_{\text{max}} = \text{MAX}; \ T_i = \text{avg}(I_i) + \beta \cdot \text{std}(I_i); \)

**Iteration:**

1: while \( i \leq i_{\text{max}} \) do

2: \( i = i + 1; \)

3: Divide \( I_{i-1} \) into \( 2 \times 2 \) subimages \( I_i; \)

4: for all \( I_i \) do

5: \( T_i = \text{avg}(I_i) + \beta \cdot \text{std}(I_i); \)

6: if \( T_i > T_{i-1} \) then

7: Set threshold for the corresponding subimage as \( T_i; \)

8: Record the corresponding subimage for dividing in next iteration;

9: else

10: Set threshold for the corresponding subimage as \( T_{i-1}; \)

11: Remove the corresponding subimage for dividing in next iteration;

12: end if

13: end for

14: end while
is divided again into the succeeding \((i + 1)\) th level. In this way, adaptive thresholds are obtained depending on the intensity distribution in different parts of the image. Although the image is divided into \(2 \times 2\) subimages recursively, it does not mean the size of SAR image should be fixed to power of 2. An example of the proposed multilevel thresholding algorithm is illustrated in Fig. 6.

After the adaptive multilevel thresholding segmentation, a binary image is obtained. Based on this binary image, the connected component analysis method is applied to extract geometric features (in this paper, area features) for each pixel. For a binary image, a region is said to be connected when all pixels in this region share the same class label +1 or 0. There are many highly efficient connected component labeling methods for binary images. We apply a MATLAB® (Ref. 30) function “bwconncomp” to find connected components in a binary image. After the connected regions are labeled, the area for each region is computed and is assigned to the pixels in the corresponding region as the geometric feature \(f_2\). As illustrated in Fig. 7, for connected “Region A,” the area is 29; thus, for all pixels in this region, the \(f_2\) is set to 29. Similarly, for “Region B,” the \(f_2\) of all pixels is set to 40, i.e., the area of this region.

### 3.2.3 Textural features

As mentioned above, the different physical characteristics of oil film and seawater result in different textural properties in the SAR image. An LBP descriptor can be used to describe the local textural feature of an image. The most important property of the LBP descriptor is its robustness to monotonic gray-scale changes and rotation variance. Because of its discriminative power...
and computational simplicity, the LBP texture descriptor has become a popular tool in various applications. In this paper, for a given pixel, the value of its LBP code is utilized as the textural feature \( f_3 \) for the given pixel. The LBP code is computed based on binary comparisons with its \( p \) surrounding neighborhoods with the radius \( r \), as defined in Eq. (2) and shown in Fig. 8.

\[
\text{LBPCODE}_{p,r} = \sum_{n=0}^{p-1} 2^n \cdot f(x_{r,n} - x_{0,0}), \quad \text{if } y \geq 0 \quad \text{then } f(y) = 1; \quad \text{else } (y) = 0. \tag{2}
\]

### 3.3 Dark-Spot Segmentation Using Artificial Neural Network

Previous studies have shown that an ANN is suitable for dark-spot segmentation. We also apply a multilayer perceptron (MLP) classifier, specifically, a three-layer perceptron neural network, to conduct the dark-spot segmentation task. In our implementation, the input layer consists of three neurons, corresponding to the gray feature \( f_1 \), geometric feature \( f_2 \), and textural feature \( f_3 \). The output layer consists of two neurons corresponding to the two classes, i.e., dark-spots and seawater. The hidden layer of six neurons is selected, which is determined based on a trial experiment on the training data, by adding hidden neurons to the neural network architecture and evaluating the resulting effect on the accuracy of the image segmentation. Another reason for setting the number of intermediate neurons as six is based on an ad-hoc rule that this number should not be larger than double the neurons of the input layer. An error backpropagation algorithm is utilized to train the network.
3.4 Morphology Postprocessing

After the previous dark-spot detection process, fragmentation of dark-spot regions and tiny blank holes in some dark-spot regions often occur. Thus, we first apply a “closing” operator of mathematical morphology to fill the tiny blank holes and connect the separate dark-spot regions. A 3 × 3 unit matrix is used as the structure matrix to conduct the basic erosion and dilation operations. Then, an empirical area threshold $T_A$ and a contrast threshold $T_C$ are used to further remove the incorrectly detected dark spots. The average contrast of a detected dark-spot region $i$ is defined by

$$C_i = \frac{\mu_{i,S} - \mu_i}{\sigma_{i,S}},$$  \hspace{1cm} (3)$$

where $\mu_i$ is the average intensity of the detected dark-spot region $i$, $\mu_{i,S}$ is the average intensity of the background surrounding the dark-spot region $i$ with the extended edge of two pixels, instead of the average intensity of the whole background as used by Shu et al.,$^{16}$ and $\sigma_{i,S}$ is the standard deviation of the intensity of the surrounding background. In this way, we take the local contrast under consideration to improve the performance. Only regions with an area greater than $T_A$ and an average contrast greater than $T_C$ are regarded as the real dark-spot regions.

4 Experiments

4.1 Comparison Algorithm

As mentioned earlier, few previous studies have intentionally focused on the dark-spot segmentation task. Among them, Shu et al.$^{16}$ made use of a thresholding algorithm to conduct dark-spot segmentation. Feng et al.$^{23}$ jointly made use of a thresholding algorithm and edge detection to detect dark-spot regions. Recently, Topouzelis et al.$^{33}$ and Singha et al.$^{10}$ have proven that ANN trained by proper feature combination can achieve high dark-spot detection performance, which is better than that of adaptive thresholding segmentation and edge detection methods.

To present a comprehensive evaluation on the proposed method (hereafter abbreviated as PROP), we compare it with the method proposed by Singha et al.$^{10}$ which combined one gray feature and two statistical textural features to generate a new vector and used it as the input to an ANN classifier to conduct dark-spot segmentation. In their implementation, the input layer consists of three neurons corresponding to the intensity value of the pixel, the mean intensity, and the standard deviation of a 3 × 3 pixel-centered neighbor area. The output layer consists of two neurons, corresponding to two classes (i.e., dark spots and seawater), and the hidden layer of six neurons is selected as optimal and determined by trial and error. In this paper, we reimplemented Singha’s method (hereafter abbreviated as SINGHA) and tested it on our study areas.

Furthermore, we also implemented the proposed method with a support vector machine (SVM) classifier$^{35,36}$ to test the performance of different classification methods (abbreviated as SVM).

4.2 Experimental Protocol

For clarity, we describe the basic experimental protocol in this section.

- In the preprocessing step, a 3 × 3 window was selected both for median and mean filtering based on the principle that on the one hand, the filter can remove the speckle and noise effectively, and on the other hand, it can degrade the image quality as little as possible. From Table 1, it is shown that a 3 × 3 window caused relatively small image quality degradation, which is assessed by the peak signal-to-noise ratio (PSNR).$^{37}$

- In the feature extraction step, to extract gray features, a histogram stretching transformation is applied. Generally speaking, the selection of parameters for stretching transformation is data dependent. In our implementation, the SAR image is 8 bit per pixel; thus, the $x_3$ and $x_3'$ are set to 255, i.e., the maximum gray-level value of data. Based on the statistical analysis
to original SAR image, the $x_1$ and $x_2$ are set to 25 and 125, respectively, while the $x'_1$ and $x'_2$ are set to 25 and 225, respectively. The corresponding $a_1$, $a_2$, $a_3$ can be calculated and set to 1, 2, and 3/13, respectively.

- In the feature extraction step, to extract geometric features, in the proposed multilevel thresholding algorithm, the maximum number of divisions ($i_{\text{MAX}}$) was set to 8, considering that after 8 divisions, the size of an image with an original size of 1024 × 1024 would be reduced to $4 \times 4$. The balance parameter $\beta$ was set to 0.3 based on empirical tests, as shown in Fig. 9. Then, an eight-connected component analysis method was used to obtain the connected region for area calculation. To determine the parameter $\beta$, we tested the area segmentation performance on three subimages, as shown in Figs. 9(a)–9(c). Figure 9(d) plots the coverage rate between the segmented area using the proposed method and the ground truth (GT). It was shown that the area coverage rate (i.e., the accuracy) varied with the parameter $\beta$. Totally speaking, $\beta = 0.3$ is an optimal selection.

- In the feature extraction step, to extract textural features, the LBP code was computed based on binary comparisons with its $p = 8$ surrounding neighborhoods with a radius of $r = 1$.

- In the classification step, a three-layer perceptron neural network classifier was used. The number of neurons for the input layer, hidden layer, and output layer was 3, 6, and 2, respectively. The sigmoid activation function was used and backpropagation algorithm was used to train the network. The learning rate and the momentum backpropagation rate are set to 0.05 and 0.10, respectively, based on the trial experiment. The training stops when the mean square error reaches zero or a predefined maximum number of epochs (10,000 times in our implementation) is reached. For the comparative algorithm, SVM classifier, we utilized the LibSVM implementation. The radial basis function kernel was adopted. The optimal hyper-parameters were selected by grid search in a discretized two dimensional parameter space along $2^d$, where $d = 4, 3, \ldots, 3, 4$ for kernel bandwidth $\gamma$ and regularization parameter $C$.

Table 1 The determination of filtering window size based on PSNR.

<table>
<thead>
<tr>
<th>Method</th>
<th>Gulf of Mexico</th>
<th>Prestige oil spill</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3 × 3</td>
<td>5 × 5</td>
</tr>
<tr>
<td>Median (dB)</td>
<td>70.3445</td>
<td>67.0417</td>
</tr>
<tr>
<td>Mean (dB)</td>
<td>75.2559</td>
<td>73.3551</td>
</tr>
</tbody>
</table>

Fig. 9 The determination of parameter $\beta$ in Algorithm 1.
In the morphology postprocessing step, a $3 \times 3$ unit matrix is used as the structure matrix to conduct the basic erosion and dilation operations. The area threshold $T_A$ and contrast threshold $T_C$ were set to 80 and 2, respectively. These two values were selected by a grid search in a discretized two dimensional parameter space, specifically $T_A \in \{50, 60, \ldots, 140, 150\}$ and $T_C \in \{1.5, 2.0, 2.5, \ldots, 8\}$, to maximize the criterion $P_d/P_f$ as defined in Eq. (4).

All comparative methods were evaluated under a supervised classification framework. Limited to the small amount of data we collected, we randomly selected 20 areas from Mexico data and Prestige data, in which 10 areas are oil spills and 10 areas are sea water, and used them as the training data. The size of each sample area is $20 \times 20$. Thus, $20 \times 20 \times 20 = 8000$ pixels were used to train the classifier, which accounts for 0.28% [8000/(2 x 1200 x 1200)] of total pixels of the two datasets. Then the learned classifiers for different methods were tested on the two study areas for comparative evaluation. It should be noted that the training data were not separated intentionally from the two study areas and were reused in the test stage. Rigorously speaking, the reuse of training data in the test stage should be avoided. While considering that the amount of training data only accounts for 0.28% of the total pixels and thus has no undeservedly neglected impact on the quantitative assessment, we did not separate the training data from the study area in the test phase to present a visually complete output result, as shown in Fig. 10.

### 4.3 Results and Discussion

The dark-spot segmentation performance of three comparative methods was evaluated both qualitatively and quantitatively. To present a convincing assessment, we created a GT for each dataset by visual inspection of the image. To obtain a GT, several operators first individually labeled each SAR image pixel by pixel as dark-spots (black areas) or seawater (white areas). Then, all labeling results were merged into a robust GT by a majority vote. As a qualitative comparison, the outputs of the comparative methods were compared with the GT for each study area, as shown in Fig. 10. It is easy to see that for Mexico data, generally speaking, the output of PROP is more like the GT than SINGHA and SVM. For Prestige data, it is hard to say which output is more like the GT. More accurate comparison requires a quantitative assessment. To present a quantitative assessment, each output was compared with the GT pixel by pixel to obtain a confusion matrix. Then, three widely used indicators were derived from the confusion matrix to quantitatively evaluate the performance of comparative methods, including the detection probability ($P_d$) and false alarm probability ($P_f$) for dark spots and the overall accuracy ($P_{acc}$), which are defined as

![Fig. 10 Experimental results of comparative methods for Gulf of Mexico and Prestige oil spill.](image-url)
\[ P_d = \frac{TP}{TP + FN} \times 100\% \]
\[ P_f = \frac{FP}{TP + FP} \times 100\% \]
\[ P_{acc} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%, \quad (4) \]

where TP (true positive) and TN (true negative) denote the number of correctly detected dark spots and seawater, respectively, while FP (false positive) and FN (false negative) denote the number of false alarms and missing dark spots, respectively. The results for the two study areas are listed in Table 2.

As shown in Table 2, for Mexico data, the average accuracy \( P_{acc} \) obtained by PROP is 89.87\%, which is slightly better than the accuracy of 88.88\% obtained by SINGHA and slightly lower than that obtained using SVM (90.67\%). When considering the balance between the detection probability \( P_d \) and false alarm probability \( P_f \), SINGHA can achieve 80.72\% detection probability, but at a cost of up to 32.43\% false alarm probability. SVM can only achieve 70.85\% detection probability at a relatively low cost of 21.62\% false alarm probability, whereas PROP can achieve a higher detection probability, up to 81.34\% at a cost of 29.60\% false alarm probability. As we know, SINGHA and PROP apply the same MLP classifier but different feature combinations. From both aspects of average accuracy \( P_{acc} \) and balance between \( P_d \) and \( P_f \), the proposed method outperformed SINGHA. It demonstrates that the proposed feature combination is more suitable for dark-spot segmentation. On the other hand, when comparing SVM and PROP, i.e., the same feature combination but different classifier, the result showed that the proposed method can provide 10.49\% more correct dark spots than SVM, with less than 8\% more false alarms. In terms of oil spill disaster prevention and mitigation, PROP is more suitable for practical applications. This result confirms why so many researchers have chosen ANN as an oil spill detection and recognition classifier. For Prestige data, in terms of average accuracy \( P_{acc} \), the three compared methods yielded considerably good results of up to 98\%. In terms of detection probability \( P_d \) and false alarm probability \( P_f \), the proposed PROP beats SVM and SINGHA again, with the highest detection probability (81.45\%), highest total accuracy (98.07\%), and lowest false alarm probability (20.28\%).

In addition, to present a deep insight into the role of three types of features, we also conducted a classification test using different combinations of feature to form the feature vector. In this test, the gray feature was used as the basis of the features. It was tested separately (termed as GRAY) or combined with the geometric feature (termed as GRAY+GEOMETRIC) or texture feature (termed as GRAY+TEXTURAL). The classification performance obtained using geometric feature or texture feature individually was very poor; thus, it was not used for comparison. As aforementioned, visualized segmentation results are shown in Fig. 11, and a quantitative evaluation is listed in Table 3. It was observed that for Mexico data, gray feature and textural feature play very important roles for classification. The detection probability is up to 78.25\% using gray feature individually and is further improved to 80.43\% by adding textural feature. At the same time, the false alarm probability is maintained at a relatively low value (29.51\% and 29.45\% for two methods, respectively). These results are even close to the performance obtained by SINGHA and PROP (i.e., all three types of features are combined together). Moreover, compared with SINGHA, they obtained lower false alarm probability. It was also observed that for Mexico

Table 2  Quantitative assessment for comparative methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>Gulf of Mexico</th>
<th></th>
<th>Prestige oil spill</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( P_d ) (%)</td>
<td>( P_f ) (%)</td>
<td>( P_{acc} ) (%)</td>
<td>( P_d ) (%)</td>
</tr>
<tr>
<td>SINGHA</td>
<td>80.72</td>
<td>32.14</td>
<td>89.09</td>
<td>78.03</td>
</tr>
<tr>
<td>SVM</td>
<td>70.85</td>
<td>21.62</td>
<td>90.67</td>
<td>79.19</td>
</tr>
<tr>
<td>PROP</td>
<td>81.34</td>
<td>29.60</td>
<td>89.87</td>
<td>81.45</td>
</tr>
</tbody>
</table>
data, the geometric feature seems to play a negative role. When adding the geometric feature to
the gray feature, the detection probability dropped to 67.39% while the false alarm probability
jumped to 33.39%. Interestingly, the situation was completely reversed in the other Prestige data.
In this case, by adding the geometric feature to the gray feature, the detection probability
improved about 15% from 58.77% (GRAY) to 73.34% (GRAY+GEOMETRIC) and the false
alarm probability reduced to 14.36% from 16.33%. By analyzing the data, we think this contra-
diction results from the differences between the two data. In Mexico data, the oil spill appears as
a large, connected block area, while in Prestige data, the oil spill appears as a small, intermittent
thin belt. Thus, gray and textural features play a more important role in the former and the
geometric feature plays a relatively important role in the latter. This test also proves that the
positive/negative role of different features is data dependent. For three types of features,
gray, geometric, and textural, the use of a single one or a combination of two of the three cannot
address a variety of situations (data) well. A combination of all three features is the best choice at
the present stage.

We also found that for both test datasets, using each of the compared methods (SINGHA,
SVM, and PROP), the highest detection probability reached approximately 80% at a cost of
above 20% false alarm probability. This shows that developing a dark-spot segmentation method
with good balance between detection probability and false alarm probability is not a trivial task.
It is a task that should be taken seriously since its output has a direct impact on the whole oil spill
detection system. In real-world applications, people are more inclined to sacrifice false alarm
probability (i.e., allow more false alarms) to maintain a higher detection probability, to avoid
missing detection of possible oil spills. However, this can lead to lower oil spill detection effi-
ciency since too many interfering targets need to be screened out individually. Actually, the
experimental result achieved relatively high average accuracy and relatively low detection prob-
ability because we trained the classification model (both for ANN and SVM), and the objective

![Images of various combinations of multiple features for Gulf of Mexico and Prestige oil spill.](image)

**Fig. 11** Experimental results of various combinations of multiple features for Gulf of Mexico and Prestige oil spill.

**Table 3** Quantitative assessment for different feature combinations.

<table>
<thead>
<tr>
<th>Method</th>
<th>Gulf of Mexico</th>
<th>Prestige oil spill</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P_d$ (%)</td>
<td>$P_f$ (%)</td>
</tr>
<tr>
<td>GRAY</td>
<td>78.25</td>
<td>29.51</td>
</tr>
<tr>
<td>GRAY+GEOMETRIC</td>
<td>67.39</td>
<td>33.39</td>
</tr>
<tr>
<td>GRAY+TEXTURAL</td>
<td>80.43</td>
<td>29.45</td>
</tr>
</tbody>
</table>

In the Mexico data, the oil spill appears as a large, connected block area, while in Prestige data, the oil spill appears as a small, intermittent thin belt. Thus, gray and textural features play a more important role in the former and the geometric feature plays a relatively important role in the latter. This test also proves that the positive/negative role of different features is data dependent. For three types of features, gray, geometric, and textural, the use of a single one or a combination of two of the three cannot address a variety of situations (data) well. A combination of all three features is the best choice at the present stage.

We also found that for both test datasets, using each of the compared methods (SINGHA, SVM, and PROP), the highest detection probability reached approximately 80% at a cost of above 20% false alarm probability. This shows that developing a dark-spot segmentation method with good balance between detection probability and false alarm probability is not a trivial task. It is a task that should be taken seriously since its output has a direct impact on the whole oil spill detection system. In real-world applications, people are more inclined to sacrifice false alarm probability (i.e., allow more false alarms) to maintain a higher detection probability, to avoid missing detection of possible oil spills. However, this can lead to lower oil spill detection efficiency since too many interfering targets need to be screened out individually. Actually, the experimental result achieved relatively high average accuracy and relatively low detection probability because we trained the classification model (both for ANN and SVM), and the objective...
The function to be optimized was the average accuracy. Theoretically, this problem can be addressed by modifying the objective function to maximize the detection probability.

5 Conclusion

Dark-spot segmentation is a critical and fundamental step in the oil spill detection process chain, prior to feature extraction and look-alike discrimination. In this paper, a specific effort has been placed on dark-spot segmentation in a single-pol SAR image. Aiming to improve the performance of dark-spot segmentation, a combination of fine designed features, including gray features, geometric features, and textural features, is proposed to characterize oil spills and seawater. Through a carefully designed feature extraction process, three discriminative features are extracted and fused to generate an oil spill representation vector and used as input to an ANN classifier. The experimental results validate that the proposed fine designed feature combination outperforms the existing approaches in both aspects of overall segmentation accuracy and the capability to balance detection probability and false alarm probability. It is a promising alternative that can be incorporated into existing oil spill detection systems to further improve the system performance.

In our future work, we will concentrate on further improving the detection probability and retaining a relatively low false alarm probability by modifying the objective function of the classifier by maximizing the detection probability. At the same time, we will also concentrate on widening the validation of the proposed method, in particular for different SAR sensors such as Sentinel-1, which is the European Space Agency space mission designed to take the place of the ENVISAT satellite.

Acknowledgments

The authors would like to thank the guest editors and the anonymous reviewers for their careful reading and constructive remarks that have contributed to improving the quality of this paper. The authors would like to thank Dr. Yanmin Zhang (Department of Physics, Ocean University of China) for her helpful suggestions. The authors would also like to thank Yukun Ge, Tian Gan, and Yongjie Xu for labeling the GT data. This work was supported in part by the National Natural Science Foundation of China under Grant No. 61471024, the National Marine Technology Program for Public Welfare under Grant No. 201505002-1, the R&D Special Foundation for Public Welfare Industry under Grant No. 201305025, and the Higher Education and High-Quality and World-Class Universities under Grant No. PY201619.

References


**Haitao Lang** is an associate professor at the University of Chemical Technology. He received his BS and MS degrees from the Ocean University of China in 2000 and 2003, respectively, and his PhD from Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, in 2006. His research interests include remote sensing image analysis, machine learning, and optical engineering. He is a member of SPIE and IEEE.

**Xingyao Zhang** is a graduate student at the Beijing University of Chemical Technology. Her research interests include digital image processing and pattern recognition.

**Yuyang Xi** is a graduate student at the Beijing University of Chemical Technology. He majors in physics, and his research interests include PolSAR data information extraction, targets detection, and recognition in SAR imagery.

**Xi Zhang** received his BS degree in information systems from Qingdao University of Science and Technology, Qingdao, China, in 2005, and his MS degree in signal and information processing, and his PhD in computer science from Ocean University of China, Qingdao, China, in 2008 and 2011, respectively. Since 2011, he has been a research associate with the First Institute of Oceanography, State Oceanic Administration, Qingdao, China. His research interests include microwave remote sensing and SAR data analysis.

**Wei Li** received his BE degree in telecommunications engineering from Xidian University, Xi’an, China, in 2007, his MS degree in information science and technology from Sun Yat-sen University, Guangzhou, China, in 2009, and his PhD in electrical and computer engineering from Mississippi State University, Starkville, Mississippi, USA, in 2012. He is currently with the College of Information Science and Technology, Beijing University of Chemical Technology, Beijing, China. His current research interests include statistical pattern recognition, hyperspectral image analysis, and data compression.