To appear in the *International Journal of Remote Sensing*
Vol. 00, No. 00, Month 20XX, 1–20

Multi-feature Sea-land Segmentation based on Pixel-wise Learning for Optical Remote Sensing Imagery

Dan Wang[a], Xinrui Cui[b], Fengying Xie[a], Zhiguo Jiang[a] and Zhenwei Shi[a]*

[a]*School of Astronautics, Beihang University, Beijing, China*; [b]*Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, Canada*

(*Received 00 Month 20XX; accepted 00 Month 20XX*)

Robust sea-land segmentation in optical remote sensing images is challenging because of the complex sea-land environment and scene diversity. Here, we propose a novel multi-feature sea-land segmentation method via pixel-wise learning for optical remote sensing images. Multiple features like grayscale, local statistical information, edge, texture and structure, are first extracted from each pixel in training images and then used to learn a multi-feature sea-land classifier, which transforms the segmentation issue into pixel-wise binary classification problem. In our approach, a new multi-feature sea-land segmentation algorithm is put forward based on the approximation of Newton method. Experiments on Google-Earth, VRSS-1 and Gaofen-1 images demonstrate that the proposed approach yields more robust and accurate sea-land segmentation results.

Keywords: sea-land segmentation; multiple features; pixel-wise learning; optical remote sensing images

1. Introduction

With the development of optical remote sensing techniques, marine target detection based on optical remote sensing images has gradually obtained more attention (Liu et al. 2014; Yang et al. 2014; Tang et al. 2015). However, complex sea-land environment usually poses difficulty in fields like ship detection (Corbane et al. 2010) and oil leakage detection by introducing massive false targets in land area. In this context, sea-land segmentation in optical remote sensing images becomes a key step to marine target detection since good segmentation between sea and land would enhance the detection accuracy and efficiency by removing false targets in land area.

Currently, sea-land segmentation methods can be broadly classified into two categories, i.e. geographic information based methods and non-geographic information based methods. Researches in the first category mainly rely on geographic information database of global harbors and coastlines as standard reference. After the determination of several geographic coordinates in one particular optical remote sensing image, image matching between geographic information database and this image is used to conduct sea-land segmentation. However, the accuracy of geographic coordinates is highly demanding in these above methods. Besides, new

*Corresponding author. Email: shizhenwei@buaa.edu.cn

artificial coasts and islands are difficult to be identified correctly because they are usually not included in the above database.

For sea-land segmentation methods in the second category, geographic information database is not required. To date, several non-geographic information based methods have been proposed based on features like grayscale and texture, etc. Grayscale represents only intensity information which varies from black at the weakest intensity to white at the strongest. According to grayscale values, pixels in optical remote sensing images are classified into two categories representing sea and land separately. In grayscale-based methods, You et al. (You and Li 2011) proposed a sea-land segmentation method based on OTSU (Otsu 1979) and Gaussian statistical sea model, in which the threshold was determined by statistical relationships among pixels. Combining two thresholds, Zhang (Zhang and Wang 2010) put forward an improved minimum filter method to reduce the negative influence of sea surface texture. However, thresholds usually cannot be determined appropriately especially when applied to images with evenly distributed grayscale. Besides, grayscale could not reflect the spatial distribution information in images. As another description of sea-land segmentation, texture mainly gives us regional arrangement information to reflect surface property in images. In relevant researches, Zhu et al. (Zhu et al. 2010) applied local multiple patterns, a texture operator, to protrude edge and thus enhanced the representation ability of feature set. C. Sagiv et al. (Sagiv, Sochen, and Zeevi 2006) proposed a texture segmentation algorithm based on geodesic and edgeless active contours model. But one drawback of the texture feature is that it can be easily interfered by external factors such as sunlight, reflection and resolution, etc.

In addition, the above traditional feature-based sea-land segmentation methods also have the following limits:

(1) Traditional methods usually need human intervention to set special segmentation criteria, to establish the statistical models or tune parameters, which lacks flexibility in complex environments.
(2) Those above methods are mainly based on simple features, and thus unable to discriminate features of sea and land in complex scenario.

To tackle the above limits, more robust and adaptive methods for non-geographic information based sea-land segmentation need to be further explored. Supervised learning methods have been widely applied to object detection and classification (Huang, Wang, and Wang 2015; Han et al. 2015; An and Shi 2015). Supervised learning benefits in automatically learning rules and crucial information from large amount of training data, which could create models that are more robust than that determined by special artificial criteria (Li et al. 2014). Therefore, sea-land segmentation methods based on supervised learning would be superior to traditional methods. As another advantage of supervised learning methods, using multiple features like grayscale, local statistical information, edge, texture and structure, can depict complex sea-land information more efficiently than using simple features. It is noticeable that in most of these supervised learning methods patches were set as learning units. These patches are usually rectangular for targets like pedestrians and faces, and designed according to the object size (Zhao et al. 2016; Man, Dong, and Guo 2015). However, in sea-land segmentation problem, the size of sea and land is not fixed in different geographic environments. Meanwhile, the contour of sea and land is amorphous, which does not match with traditional rectangular patches. Therefore, it is necessary to design a new approach for sea-land segmentation based on pixel-wise learning. In this paper, considering all the above factors,

we propose a new supervised learning based sea-land segmentation method. Pixel-wise multiple features are first extracted from the whole image and are then used to learn a robust sea-land classifier which finally transform the segment problem into a binary classification problem. Concretely, the sea-land segmentation problem is converted to optimization problem and is finally solved by the approximation of Newton method which is of low computational complexity and fast convergence.

The rest of this paper is structured as follows. Section 2 elaborates multi-feature sea-land segmentation approach based on pixel-wise learning, including feature extraction, the designing and application of multi-feature sea-land classifier. Section 3 presents an experimental validation of multi-feature sea-land segmentation approach. Finally, conclusions are shown in Section 4.

2. multi-feature sea-land segmentation approach based on pixel-wise learning

Here, a supervised learning based method successfully converts sea-land segmentation issue into a binary classification problem. The proposed multi-feature sea-land segmentation approach based on pixel-wise learning contains three steps (as shown in Figure 1). First, pixel-wise features are extracted from training images. In this step, it is pixel rather than patch that is regarded as the learning unit, because there is hardly any patches with appropriate sizes and shapes that could match various sea-land boundaries. Pixel-wise features including grayscale, local statistical information, edge, texture, and structure, are extracted to form a feature data cubic.

Second, multi-feature sea-land classifier is trained based on the whitened feature data. Here, the designing of multi-feature sea-land classifier is transformed to an optimization problem. In designing process, the feature data cubic is processed by centering and whitening. Whitening is essentially decorrelation followed by scaling, making the whitened feature variables have unit variance and be mutual uncorrelated. Then, Logistic cost function is selected as measure function to minimize the classification error, because it is conceptually simple and has appealing statistical properties, especially robustness. Meanwhile, a multi-feature sea-land segmentation algorithm based on the approximation of Newton method is proposed to solve the above optimization problem. It is noticeable that the challenging calculation of inverse of Hessian is simplified by centering and whitening. Therefore, the proposed algorithm does not need Hessian matrix inversion and has roughly the same convergence speed with the real Newton method.

Third, the classifier is used to conduct sea-land segmentation in testing images. Here, we need to extract the same feature data cubics from testing images and conduct centering and whitening process at first. Then we apply the trained multi-feature sea-land classifier to the whitened data cubics and get candidate maps. Finally some post-processing operations are used to modify previous results to obtain the final segmentation maps.

2.1. *The construction of multi-feature data*

Feature extraction plays an important role in sea-land segmentation problem. In general, informative features greatly facilitate the training process. To obtain a robust multi-feature sea-land classifier, five different features (grayscale, local statistical information, edge, texture and structure) are used to represent different yet

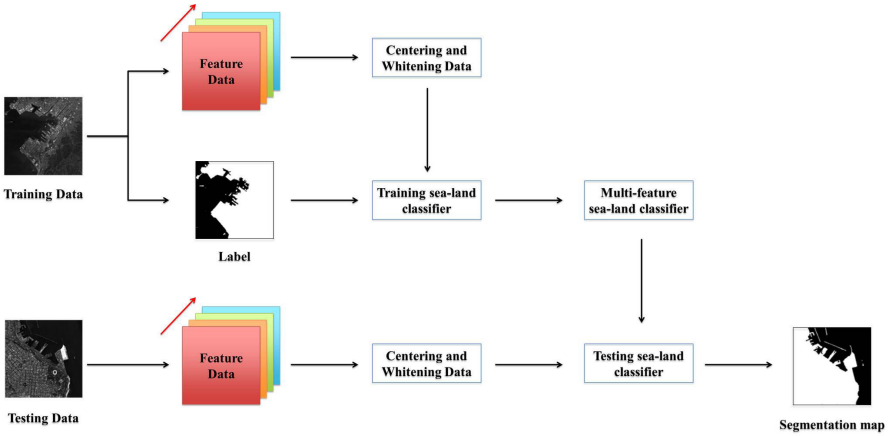

Figure 1. Workflow of multi-feature sea-land segmentation based on pixel-wise learning.

complementary characteristics of the whole image. Details of the feature extraction are introduced as follows.

2.1.1. Grayscale

Grayscale is a commonly used and powerful feature in sea-land segmentation since there is usually big grayscale difference between sea and land. In this paper, grayscale feature can be extracted from each pixel and form one feature plane. Afterwards, the feature plane is normalized by subtracting its mean value over the entire image.

2.1.2. Local statistical information

For sea-land segmentation, one can observe two facts: 1. land area usually has higher intensity compared with sea because of its larger reflectivity. 2. sea feature has generally smoother surface and lower intensity variance than those of land area. Here, we use the local mean value and local variance value to describe the local statistical information. They can be written as

$$\mathbf{M}(i) = \frac{1}{W} \sum_{j \in \Re(i)}^{W} \mathbf{I}_j \tag{1}$$

$$\mathbf{V}(i) = \sqrt{\frac{1}{W} \sum_{j \in \Re(i)}^{W} (\mathbf{I}_j - \bar{\mathbf{I}})^2} \tag{2}$$

where $\mathbf{I}$ is the input image, $\mathbf{M}(i)$ and $\mathbf{V}(i)$ represent the mean and variance value of pixel i, $\Re(i)$ is a local window, W is the total pixel number in $\Re(i)$. In practical application, $\Re(i)$ is usually a square window with its center at pixel i. In our experiments, the window width of is $\Re(i)$ set to 3, 7, 11, yielding 6 filter responses at each pixel.

2.1.3. Edge

Edge information is introduced to help determine sea-land boundaries. Edge can be defined as any pixels within images whose intensities are higher or lower than

those of surrounding pixels. Since sea-land boundaries are usually of high intensity difference, edge information should be used for sea-land segmentation. To obtain edges, sobel operator (Kanopoulos, Vasanthavada, and Baker 1988) is applied to detect edges of four directions {0°, 45°, 90°, 135°}, yielding 4 filter responses at each pixel.

2.1.4. Texture

In general concept, images can be regarded as the combination of textural part and structural part (Aujol et al. 2006). Textural part, with some periodicity and oscillatory nature, represents fine-scale spatial arrangement information in images. For sea-land segmentation, multiple land cover types exist in land area usually contain more complex texture information than sea area in optical remote sensing images.

The texture information is obtained through grayscale distribution of pixel and neighborhoods around. In this paper, we choose Gabor filter (Mehrotra, Namuduri, and Ranganathan 1992) to extract texture feature. Since Gabor filter is sensitive to the edge of images and has wide adaptation to illumination variation, it has been widely applied in texture extraction. Gabor filter could be written as:

$$g(x, y; \lambda, \theta, \sigma) = \exp(-\frac{x'^2 + y'^2}{2\sigma^2}) \exp(i(2\pi \frac{x'}{\lambda})) \tag{3}$$

where $x' = x\cos\theta + y\sin\theta$, $y' = -x\sin\theta + y\cos\theta$. Here, λ represents the wavelength of the sinusoidal function, θ represents the orientation, σ is the standard deviation of the Gaussian envelope.

In this paper, the wavelength of the sinusoidal function is defined as 1, the standard deviation of Gaussian envelope is set as 1, and the orientation is set as $\{0°, 45°, 90°, 135°\}$. Therefore, we could obtain 4 responses for each pixel.

2.1.5. Structure

Regardless of fine scale-details, structural part corresponds to the main large objects in images (Aujol et al. 2006). As a holistic description, it can provide information about main large boundaries in sea-land images. Here, we use a joint static and dynamic guidance model (Ham, Cho, and Ponce 2015) to exploit the structure information in an input image. The model takes advantages of both static and dynamic guidance, which can effectively control image structures at different scales and can handle a variety of types of data from different sensors (Ham, Cho, and Ponce 2015). In structure feature map, textures are penalized much more than informative structures, making main structure information salient. Here, we obtain 6 structure maps with different scales by empirically using different regularization parameters which adjust the scale feature.

All the above 21 feature maps are stacked to form original feature data. In next two sections, details of multi-feature sea-land classifier designing and application are provided.

2.2. *The design of multi-feature sea-land classifier*

Based on the above feature data cubics, we design a multi-feature sea-land classifier by minimizing the classification error between the data and their labels. Here,

multi-feature sea-land segmentation algorithm is proposed and details of the proposed algorithm could be mathematically analyzed as follows.

2.2.1. Robust estimation

Based on features extracted above, we could get multi-feature vector $\mathbf{x} \in \mathbb{R}^{L\times 1}$ for each pixel (L represents the total number of features).

$$\mathbf{x} = \begin{bmatrix} x_1 \; x_2 \; \cdots \; x_{(L-1)} \; x_L \end{bmatrix}^T \tag{4}$$

Given a feature data set $\mathbf{X} \in \mathbb{R}^{L\times N}$ (N is the number of pixels), the training data set D can be written as

$$D = \{(\mathbf{x}_1, z_1), (\mathbf{x}_2, z_2), ..., (\mathbf{x}_N, z_N)\} \tag{5}$$

where $\mathbf{x}_i \in \mathbb{R}^{L\times 1}$ denotes multi-feature vector for i-th pixel in training data and $z_i \in \{0, 1\}_{i=1}^{N}$ is label for the i-th pixel. $\mathbf{x}_i$ belongs to land or sea if the label $z_i = 1$ or 0.

In the proposed method, weight vector $\mathbf{w} \in \mathbb{R}^{L\times 1}$ is designed to categorize pixels as sea or land. For feature vector of each pixel, the residual error turns out to be

$$e_i = \mathbf{w}^T\mathbf{x}_i - z_i, \quad i = 1, ..., N \tag{6}$$

In such circumstance, the cost function $G(e)$ could be defined as

$$G(e) \equiv G(\mathbf{w}^T\mathbf{x} - z) \tag{7}$$

Through the minimum of cost function for training data set, the optimal weight vector $\mathbf{w}^*$ can be calculated.

$$\mathbf{w}^* = \arg\min J(\mathbf{w}) \tag{8}$$

where $J(\mathbf{w}) \equiv \frac{1}{N}\sum_{i=1}^{N} G(\mathbf{w}^T\mathbf{x}_i - z_i)$. Here, $J(\mathbf{w})$ could also be formulated using an expectation form.

$$J(\mathbf{w}) \equiv \mathbb{E}[G(\mathbf{w}^T\mathbf{x} - z)] \tag{9}$$

In traditional method, squared Euclidean distance is frequently regarded as cost function. However, squared Euclidean distance has some drawbacks in practice, when its value has to be estimated from a measured sample. The main problem is its high sensitivity to outliers. Therefore, the squared Euclidean distance may not be a robust measure of cost function. To solve the above problem, a robust cost function need to be selected. It is noteworthy that robust cost function should satisfy the following properties:

- nonnegative valued, $G(e) \geq 0, \forall e$;
- even symmetric, $G(e) = G(-e)$;
- nondecreasing for $e \geq 0$;
- has unique minimum at the origin, $G(0) = 0$;
- has first two derivatives piecewise continuous.

In general, one obtains more robust estimators when choosing a cost function that does not grow too fast with increasing residual error. Proper cost function can

give a very good compromise between the properties of computational complexity and robustness, which is conceptually simple, fast to compute, and have appealing statistical properties, especially robustness.

Under careful consideration, we select Logistic cost function as the substitute for squared Euclidean distance. The Logistic cost function is written as follows and has proved very useful:

$$G(e) = \frac{1}{\beta} ln(cosh(\beta e)) \tag{10}$$

where β is a constant (usually set as 1).

The first order derivative g and second order derivative g' of the Logistic cost function are

$$g(e) = tanh(\beta e) \tag{11}$$

$$g'(e) = \beta(1 - tanh^2(\beta e)) \tag{12}$$

After the selection of robust cost function, the design of multi-feature sea-land segmentation is converted to a convex optimization problem.

2.2.2. The multi-feature sea-land segmentation algorithm

Gradient descent is one of the most common approaches to the unconstrained problem of minimizing cost function. However, the convergence rate of gradient descent is sensitive to learning speed. Besides, when the optimization problem is ill-conditioned, the gradient descent will have slow convergence. Newton method is an efficient way to solve above problems, since it provides an optimal quadratic way to converge in a small number of steps.

For the minimization problem of cost function

$$\mathbf{w}^* = \arg\min J(\mathbf{w}) = \arg\min \frac{1}{N} \sum_{i=1}^{N} G(\mathbf{w}^T \mathbf{x}_i - z_i) \tag{13}$$

Newton method can be expressed as

$$\mathbf{w} \leftarrow \mathbf{w} - [\frac{\partial^2 J(\mathbf{w})}{\partial \mathbf{w}^2}]^{-1} \frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} \tag{14}$$

However, it is obvious that Newton method is computationally much more demanding per iteration than the gradient descent. The inverse of the Hessian has to be calculated at each step, which is very difficult for high dimensional problems. Besides, the Hessian matrix may also happen to be ill-conditioned or close to a singular matrix at some step of the algorithm, which may introduce numerical errors into the iteration.

To solve the above problems, we use diagonal matrix as the approximation of Hessian matrix to increase the computational speed and stability. Concretely, multi-feature data is processed by centering and whitening without affecting the estimation of multi-feature sea-land classifier. After the above processing, the feature data become statistically uncorrelated and the covariance converts to unit matrix.

Centering means making multi-feature vector has zero mean. For data $\mathbf{x} \in \mathbb{R}^{L \times 1}$,

centered data $\hat{\mathbf{x}} \in \mathbb{R}^{L\times 1}$ is obtained from the following equation:

$$\hat{\mathbf{x}} = \mathbf{x} - \mathbb{E}[\mathbf{x}] \tag{15}$$

where $\mathbb{E}[\cdot]$ calculates the expectation.

Whitening is essentially decorrelation followed by scaling, which implies that it can be achieved by a linear operation mathematically. For a zero-mean random vector $\mathbf{c} = [c_1, \cdots, c_n]^T$, it is whitened vector when its covariance is the unit matrix ($\mathbb{E}[\mathbf{c}\mathbf{c}^T] = \mathbf{I}$), which means its elements c_i are uncorrelated. For multi-feature vector $\hat{\mathbf{x}}$, its whitened vector $\tilde{\mathbf{x}} \in \mathbb{R}^{L\times 1}$ could be obtained with the following equations:

$$\tilde{\mathbf{x}} = \mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T\hat{\mathbf{x}} \tag{16}$$

where $\mathbf{E} = (\mathbf{e}_1, \cdots, \mathbf{e}_n)$ is the orthogonal matrix of eigenvectors of $\mathbb{E}[\hat{\mathbf{x}}\hat{\mathbf{x}}^T]$ and $\mathbf{D} = diag(d_1 \cdots d_n)$ represents the diagonal matrix of its eigenvalues.

In such circumstance, the covariance of whitened vector is the unit matrix:

$$\begin{aligned}
\mathbb{E}[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T] &= \mathbb{E}[(\mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T\hat{\mathbf{x}})(\mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T\hat{\mathbf{x}})^T] \\
&= \mathbb{E}[\mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T\hat{\mathbf{x}}\hat{\mathbf{x}}^T\mathbf{E}(\mathbf{D}^{-1/2})^T\mathbf{E}^T] \\
&= \mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T\mathbb{E}[\hat{\mathbf{x}}\hat{\mathbf{x}}^T]\mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T \\
&= \mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T\mathbf{E}\mathbf{D}\mathbf{E}^T\mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T \\
&= \mathbf{I}
\end{aligned} \tag{17}$$

After the processing, whitened feature data cubic for each image is obtained and then used to train classifier. In this context, the cost function is rewritten as

$$G(\tilde{e}) \equiv G(\mathbf{w}^T\tilde{\mathbf{x}} - z) \tag{18}$$

Then the optimization problem turns out to be:

$$\mathbf{w}^* = \arg\min J(\mathbf{w}) = \arg\min \frac{1}{N}\sum_{i=1}^{N} G(\mathbf{w}^T\tilde{\mathbf{x}}_i - z_i) \tag{19}$$

where $J(\mathbf{w}) \equiv \mathbb{E}[G(\mathbf{w}^T\tilde{\mathbf{x}} - z)]$.

The first and second order derivatives of G in (18) are

$$\frac{\partial G(\mathbf{w}^T\tilde{\mathbf{x}} - z)}{\partial \mathbf{w}} = g(\mathbf{w}^T\tilde{\mathbf{x}} - z)\tilde{\mathbf{x}} \tag{20}$$

$$\frac{\partial^2 G(\mathbf{w}^T\tilde{\mathbf{x}} - z)}{\partial \mathbf{w}^2} = g'(\mathbf{w}^T\tilde{\mathbf{x}} - z)\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T \tag{21}$$

Then the first and second order derivatives of $J(\mathbf{w})$ could be written as

$$\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} = \mathbb{E}[g(\mathbf{w}^T\tilde{\mathbf{x}} - z)\tilde{\mathbf{x}}] \tag{22}$$

$$\frac{\partial^2 J(\mathbf{w})}{\partial \mathbf{w}^2} = \mathbb{E}[g'(\mathbf{w}^T\tilde{\mathbf{x}} - z)\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T] \tag{23}$$

Algorithm 1 Pseudocode of multi-feature sea-land segmentation Algorithm

Initialization:

1: Center the multiple feature data $\mathbf{x}$ to make its mean zero. $\hat{\mathbf{x}} = \mathbf{x} - \mathbb{E}[\mathbf{x}]$

2: Whiten the data and make the covariance of $\tilde{\mathbf{x}}$ the unit matrix. $\tilde{\mathbf{x}} = \mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^T\hat{\mathbf{x}}$

3: Choose an random initial vector $\mathbf{w}$ of unit norm.

Repeat:

4: $\mathbf{w} \leftarrow \mathbf{w} - \mathbb{E}[g(\mathbf{w}^T\tilde{\mathbf{x}} - z)\tilde{\mathbf{x}}]/\mathbb{E}[g'(\mathbf{w}^T\tilde{\mathbf{x}} - z)]$

5: If not converged, go back to step (4).

Output: the optimal multi-feature sea-land classifier $\mathbf{w}$.

To simplify the inversion of $\frac{\partial^2 J(\mathbf{w})}{\partial \mathbf{w}^2}$ matrix, we decide to approximate $\frac{\partial^2 J(\mathbf{w})}{\partial \mathbf{w}^2}$. Since the data is sphered, a reasonable approximation seems to be

$$\begin{aligned} \frac{\partial^2 J(\mathbf{w})}{\partial \mathbf{w}^2} &= \mathbb{E}[g'(\mathbf{w}^T\tilde{\mathbf{x}} - z)\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T] \\ &\approx \mathbb{E}[g'(\mathbf{w}^T\tilde{\mathbf{x}} - z)]\mathbb{E}[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T] \\ &= \mathbb{E}[g'(\mathbf{w}^T\tilde{\mathbf{x}} - z)]\mathbf{I} \end{aligned} \tag{24}$$

Therefore, $\frac{\partial^2 J(\mathbf{w})}{\partial \mathbf{w}^2}$ is approximated to a diagonal matrix, and can be easily inverted. Therefore, using (19) and (21), we can obtain the following approximative Newton iteration:

$$\mathbf{w} \leftarrow \mathbf{w} - \mathbb{E}[g(\mathbf{w}^T\tilde{\mathbf{x}} - z)\tilde{\mathbf{x}}]/\mathbb{E}[g'(\mathbf{w}^T\tilde{\mathbf{x}} - z)] \tag{25}$$

Based on the special property of feature data after centering and whitening, one can find that multi-feature sea-land segmentation algorithm does not need a matrix inversion but still converges roughly with the same number of iterations as the real Newton method. Note that the convergence of our proposed algorithm means that the distance between the former and new values of $\mathbf{w}$ is very close, and here we set the distance as 0.00001.

2.3. *Application of multi-feature sea-land classifier*

To verify the efficiency of multi-feature sea-land classifier, same feature extraction and processing steps need to be conducted in testing images at first. Then we apply the optimal weight vector $\mathbf{w}$ to the feature data to distinguish sea and land.

$$y(\tilde{\mathbf{x}}) = \mathbf{w}^T\tilde{\mathbf{x}} \tag{26}$$

where $\tilde{\mathbf{x}}$ represents the feature data. For each pixel, an output score $y(\tilde{\mathbf{x}})$ could be obtained through equation (26). Theoretically, the output score should be close to pixel label, which means score values of land pixels should be larger than that of sea pixels (the label of sea and land pixels are set as 0 and 1 in the training data respectively).

To further obtain binary images for sea-land segmentation, we are expected to determine threshold to distinguish sea and land. Here, the histogram of candidate image is of bimodal distribution, in which there are two peaks, one appears in a low-level value and the other in a high-level value. In this context, OTSU method,

an image thresholding algorithm, appears to be one appropriate choice since it could divide images with bimodal distribution into two categories, and make their interclass variance maximized. Though many improved method based on OTSU have been widely used in thresholding, we still adopt OTSU method to determine the appropriate threshold in histogram with bimodal distribution due to its low calculation and rapid computing.

Assuming the threshold value in OTSU method is t, it should be expressed as

$$t = \arg\max_{t}\{\omega_0(\mu_0 - \overline{\mu})^2 + \omega_1(\mu_1 - \overline{\mu})^2\} \tag{27}$$

where $\omega_0 = N_0/N, \omega_1 = N_1/N, \mu_0 = \sum_{i=1}^{t} i \cdot p_i, \mu_1 = \sum_{i=t}^{256} i \cdot p_i, \overline{\mu} = \omega_0\mu_0 + \omega_1\mu_1$. Here, N_0, N_1 and N are respectively the number of sea pixels, land pixels and whole pixels, respectively. p_i is the frequency of gray level i. μ_0, μ_1 and $\overline{\mu}$ respectively represent mean gray values of sea, land and the whole image. For sea-land segmentation in RGB color image, t will be an value between 1 and 256. Traversal strategy is adopted to obtain the result. Therefore, we could obtain a segmentation map R^{seg} using the Otsu's threshold, and it is written as:

$$R_{(i)}^{seg} = \begin{cases} 1, \mathbf{G}_{(i)} \geq t \\ 0, otherwise \end{cases} \tag{28}$$

where t is the calculated threshold value using Otsu's method. Here, land and sea pixels are respectively labeled with value 1 and 0.

Here, Figure 2 shows the detailed application of the multi-feature sea-land classifier and Figure 2(a) is the original Google-Earth image. As illustrated in Figure 2(b), the whole output score results could form a candidate map, where pixels of sea area are darker than that of land area. To further obtain binary images for sea-land segmentation, the output score is then compared with a threshold t: if $y(\tilde{\mathbf{x}}) \geq t$, $\tilde{\mathbf{x}}$ belongs to the land area, else, $\tilde{\mathbf{x}}$ belongs to the sea. Here, the histogram of candidate image is of bimodal distribution. In this context, OTSU method, an image thresholding algorithm, is one appropriate choice to obtain the threshold value T since it could divide images with bimodal distribution into two categories, and make their interclass variance maximized (Otsu 1979).

As shown in Figure 2(c), binary image for sea-land segmentation is obtained, and the result is more obvious compared with Figure 2(b). However, as shown in Figure 2(c), the land area is distributed with tiny black holes. Besides, some tiny interferences still exist in sea area. These interferences on land and sea should be removed to obtain optimal sea-land segmentation.

To address the above problems, two image processing approaches will be implemented. First, morphological close operator (Soille 2004) is used on the binary map, which is comprised of two fundamental morphological operators: image dilation and corrosion. The close operator could fill small holes and smooth the boundaries without changing the shape much. Here, a mask of disk shape with radius 3 is applied. Second, tiny interferences on sea area are removed based on similar method in (An and Shi 2015). Concretely, the pixel number of each region is counted and the region will be recognized as tiny interference if the number is smaller than a given constant. After the implementation of above two procedures, the final sea-land segmentation result is obtained as shown in Figure 2(d).

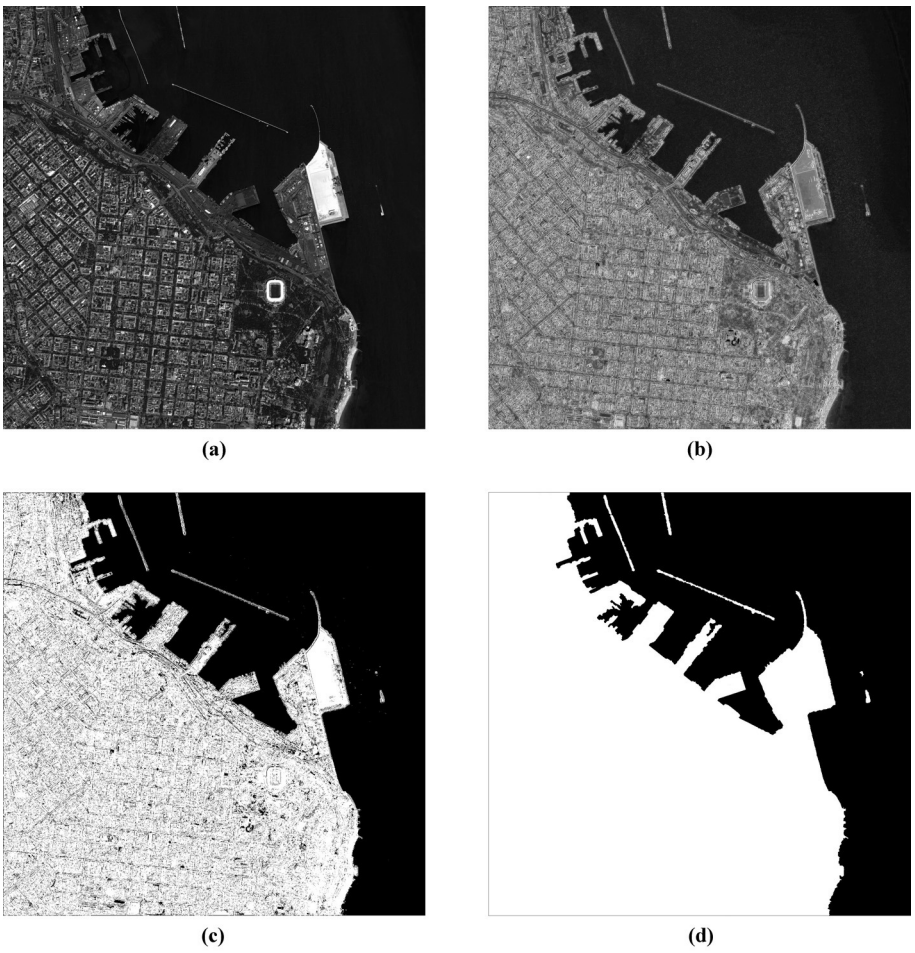

Figure 2. The detailed application of multi-feature sea-land classifier. (a) is the original Google-Earth image. (b) is the candidate map directly obtained from multi-feature sea-land classifier. (c) is the binary map obtained from OTSU method. (d) is the final sea-land segmentation map after using morphological close operator and removing tiny objects.

3. Experiments

In this section, experiments are conducted to evaluate the performance of multi-feature sea-land segmentation approach. The proposed segmentation approach is compared with local binary pattern (LBP) based segementation(Xia et al. 2014) and maximum entropy (ME) based segmentation (Kapur, Sahoo, and Wong 1985) from visual and quantitative aspects based on images obtained from Google-Earth, VRSS-1 and Gaofen-1 satellites. We also compare with OTSU based segmentation method (Otsu 1979) as a baseline method. The above methods served as comparison are all common non-geographic information based sea-land segmentation methods. To the best of our knowledge, currently there is no supervised learning based sea-land segmentation algorithm, therefore, we propose our pixel-wise learning based algorithm and compare it with the above traditional sea-land segmentation algorithms. It should be noted that the two image processing approaches (morphological close operator and tiny interferences removement) are applied to all the compared methods. Each method has been adjusted to reach their optimal parameter for sea-land segmentation.

The size of Google-Earth images is 5000 $\times$ 5000 pixels with spatial resolution of 2.15 m. For panchromatic images obtained from VRSS-1, the size is 14078 $\times$

14150 pixels with spatial resolution of 2.5 m. For panchromatic images obtained from Gaofen-1, the size is 18192 × 18000 pixels with spatial resolution of 2 m.

Here, we choose two metrics to evaluate the performance of different algorithms, i.e. the right rate of land area (LR) and the error rate (ER).

LR is used to measure the proportion of correctly classified land pixel numbers to the real land pixel numbers, and it could be written as

$$LR = \frac{CL}{LG} \tag{29}$$

where CL is the number of correctly recognized land pixels, and LG is the number of land pixels in ground truth.

ER measures the percentage of wrong classified pixel numbers in the whole image, and it could be expressed as

$$ER = \frac{SL + LS}{TN} \tag{30}$$

where SL indicates the number of real sea pixels which are misjudged as land pixels, LS denotes the number of real land pixels which are misjudged as sea pixels, and TN represents the whole number of pixels in the input image.

3.1. *Experiments based on Google-Earth images*

3.1.1. Visual analysis

In our experiments, three Google-Earth images are used to train multi-feature sea-land classifier, and the rest twelve images are used as testing images. To reduce the requirement for the data resource and make our approach a general application for different image models, we calculate the mean value of three channels (RGB) of images to get their grayscale model and just use the grayscale images in our approach.

Here, we choose six images (three of them are simple, the rest are complex) from testing data set to evaluate sea-land segmentation performance of different methods. In the six testing images, three are easy for segmentation (simple); while the rest are of complex land environment, making them difficult for correct segmentation. The ground truth of Google-Earth images are obtained by manual annotation. Corresponding ground truth images are shown in Figure 3. Comparison of sea-land segmentation results based on three simple images is shown in Figure 4. Rows from up to down are the original Google-Earth images, the corresponding sea-land segmentation images obtained from OTSU based segmentation, LBP based segmentation, ME based segmentation and the proposed method, respectively. Columns from left to right show the comparison of results obtained from different segmentation methods for each Google-Earth image.

As shown in Figure 4, there are big differences among results of four sea-land segmentation methods. Compared with the ground truth shown in Figure 3(a-c), our segmentation results outperform traditional results. First, for complex coastlines in Figure 4(a), sea-land boundaries shown in our segmentation images are more distinct and complete compared with blurred ones in traditional segmentation images. In addition, segmentation images of three traditional methods contain misjudgments to the ground truth land area. Comparing the right-most column in Figure 4, it is noteworthy that the forest in bottom-left corner of image is misjudged

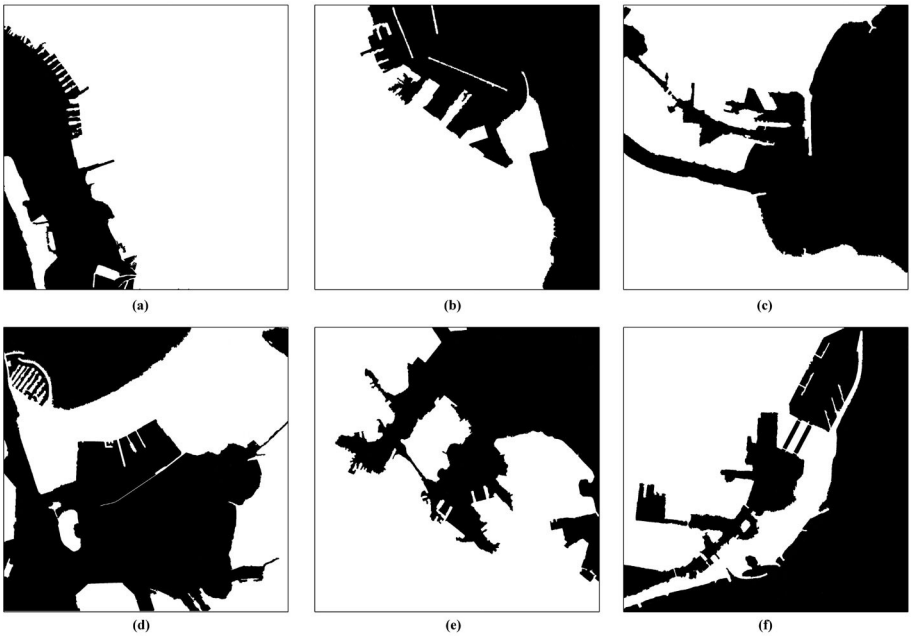

Figure 3. The ground truth data of six Google-Earth images. (a-c) simple Google-Earth images; (d-f) complex Google-Earth images.

in traditional methods. However, misjudgments barely exist in segmentation images obtained from our proposed method. From Figure 4(b), one can observe large amount of texture information (streets and buildings) in land area, which brings multiple disturbances to traditional segmentation results. However, segmentation image obtained from the proposed method seldom contains such wrong segmentation. It is noticeable that Figure 4(c) contains large area of trees which has similar grayscale with sea area. The corresponding results indicate that traditional methods misjudge this area as sea at varying degrees, but the proposed method classify this area as land correctly.

Then three more complicated Google-Earth images are used to test further performance of the proposed method. As shown in Figure 5(a), the existence of shoal makes segmentation of near coastline-regions difficult. In Figure 5(d, g, j), coastlines near shoal is blurred and their adjacent regions contain lots of misjudgments. Besides, large stretch of trees can hardly be classified correctly via traditional methods. For Figure 5(b), widespread forests and lame light intensity increase the difficulty of sea-land segmentation. The corresponding traditional segmentation images contain numerous misjudgments, especially for ME based result in this situation. In Figure 5(c), the grayscale of inner harbor is darker than that of open sea. Besides, farmland has similar grayscale to that of open sea. Traditional segmentation images shown in Figure 5(f, i, l) roughly classify both inner harbor and open sea correctly, yet misjudge farmland as sea area. On the contrary, as shown in Figure 5(m, n, o), the proposed method show more accurate sea-land segmentation results .

In short, results obtained from traditional segmentation methods (shown in Figure 5) have more misjudgments compared with that in Figure 3. However, the proposed method could still obtain good sea-land segmentation results in complex scenario, which indicates its self-adaptivity and robustness.

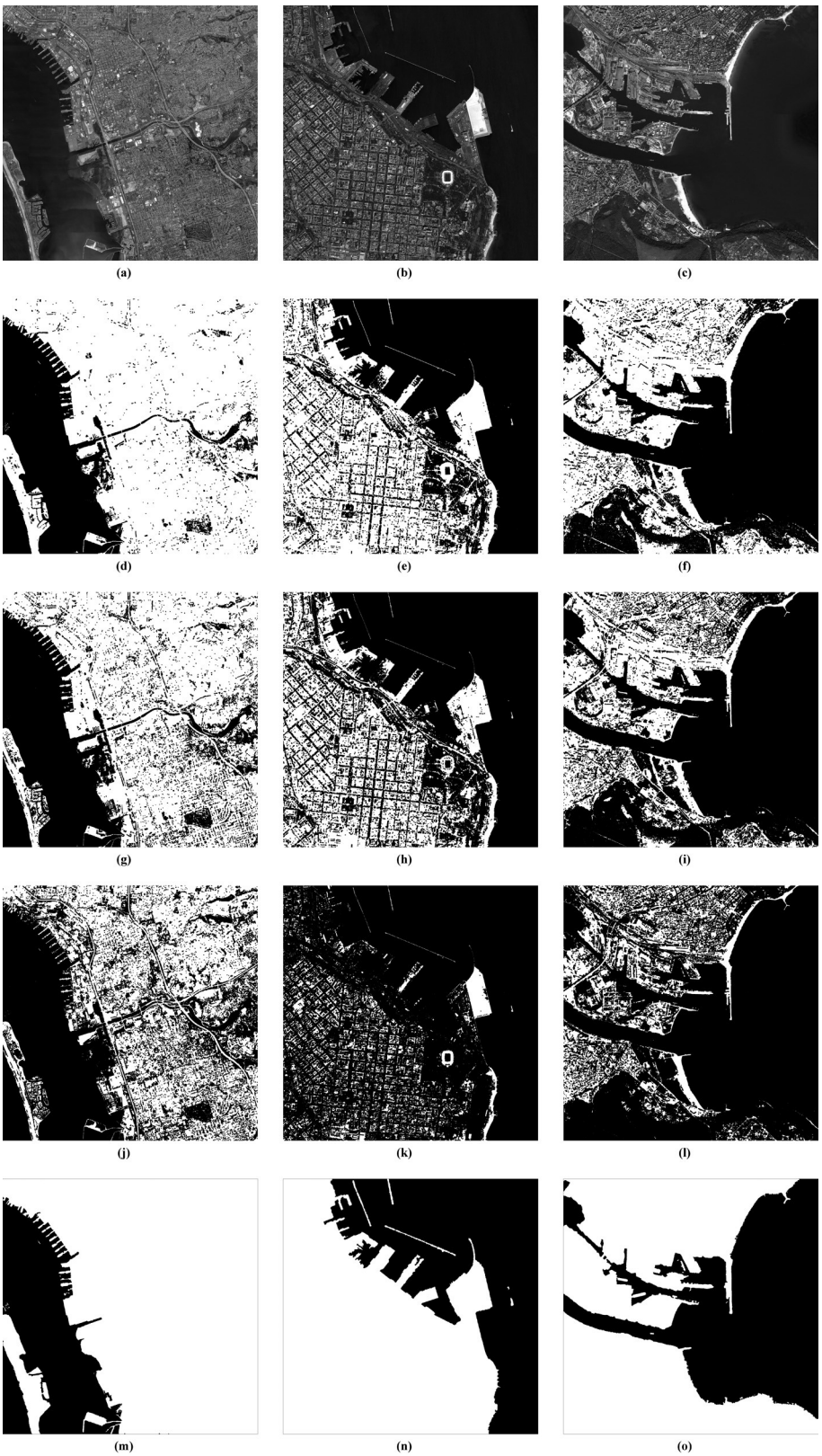

Figure 4. Visual comparison of sea-land segmentation results based on simple Google-Earth images: (a-c) Original testing images, segmentation results based on (d-f) OTSU, (g-i) LBP, (j-l) ME, and (m-o) the proposed method.

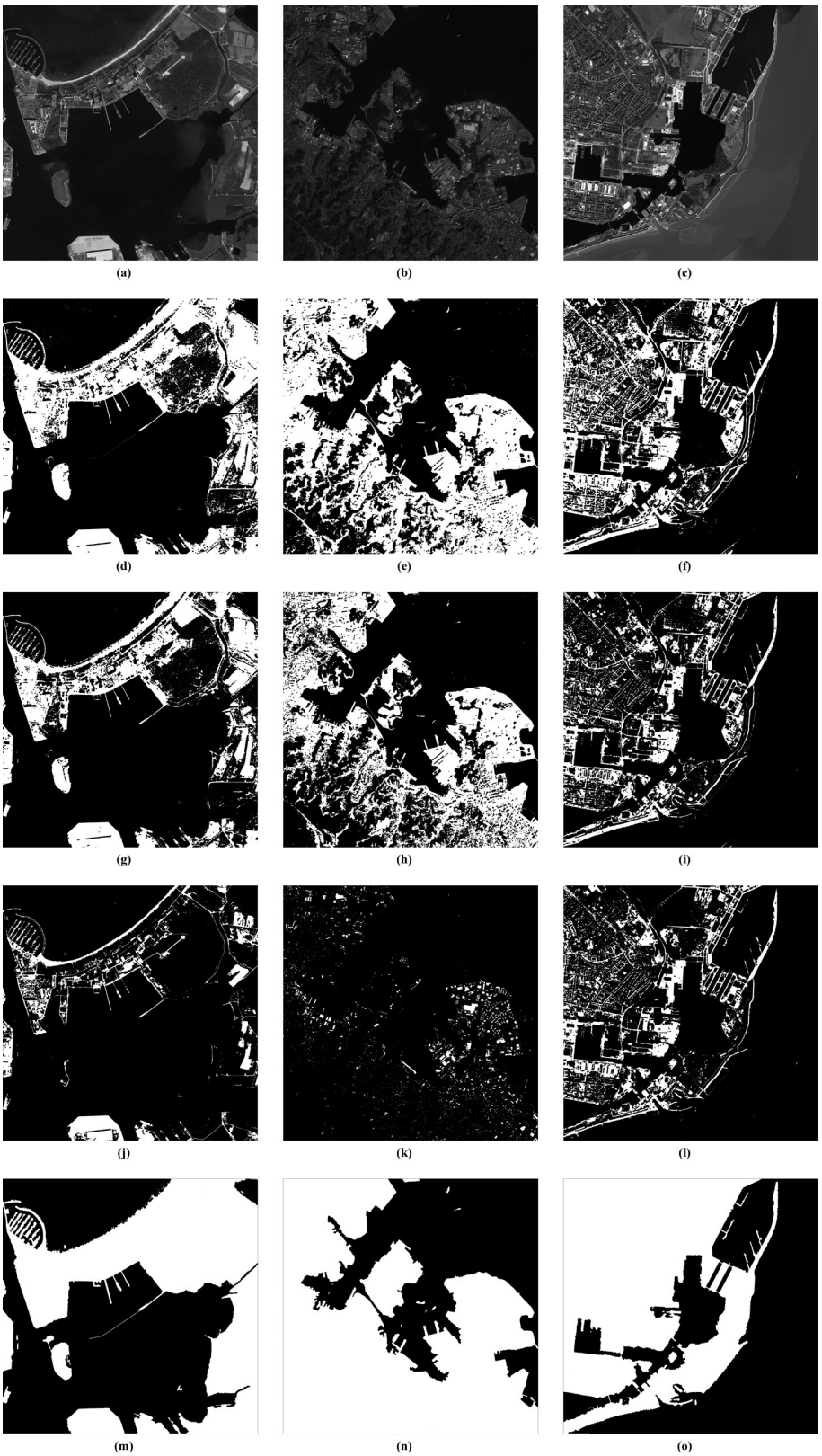

Figure 5. Visual comparison of sea-land segmentation results based on complex Google-Earth images. (a-c) Original testing images; segmentation results based on (d-f) OTSU, (g-i) LBP, (j-l) ME, and (m-o) the proposed method.

Table 1. The right rate of four sea-land segmentation methods on six Google-Earth images

	No.1	No.2	No.3	No.4	No.5	No.6	Ave
OTSU	0.9094	0.6292	0.6611	0.4085	0.6565	0.6396	0.6507
LBP	0.7615	0.4699	0.5157	0.2543	0.5137	0.4825	0.4996
ME	0.5450	0.1864	0.0291	0.2360	0.1220	0.2825	0.2335
Ours	**0.9989**	**0.9686**	**0.9913**	**0.9882**	**0.9966**	**0.9977**	**0.9902**

Table 2. The error rate of four sea-land segmentation methods on six Google-Earth images

	No.1	No.2	No.3	No.4	No.5	No.6	Ave
OTSU	0.0700	0.1667	0.2167	0.3045	0.2041	0.2079	0.1950
LBP	0.1832	0.2352	0.3094	0.3788	0.2889	0.2971	0.2821
ME	0.3493	0.3565	0.6199	0.3858	0.5215	0.4100	0.4405
Ours	**0.0029**	**0.0207**	**0.0072**	**0.0209**	**0.0025**	**0.0041**	**0.0097**

3.1.2. Quantitative analysis

Two quantitative estimations, namely, LR and ER, are applied to evaluate four different methods. Here, we provided two Tables indicating LR and ER of four sea-land segmentation methods on six Google-Earth images. Indices with bold types represent the best ones in four sea-land segmentation methods.

Table 1 shows the whole LR values and the average value of each segmentation method on six Google-Earth images. Table 2 shows the whole ER values and the average value of each segmentation method on six Google-Earth images. It can be seen from Table 1 and Table 2 that the proposed sea-land segmentation method achieves the highest LR of 0.9902 and the lowest ER of 0.0097, outperforming traditional methods by large margins. These results are in accordance with visual comparison (Figures 4 and 5) where plenty of land regions in testing images are misjudged as sea in traditional methods while classified efficiently in the proposed method. Since the proposed segmentation approach based on supervised learning could self-adaptively learn more robust sea-land segmentation rules from multi-feature training data, the results of our proposed method perform better than traditional methods.

3.2. *Experiments based on VRSS-1 and Gaofen-1 images*

3.2.1. Visual analysis

Here, two panchromatic images (one VRSS-1 image, one Gaofen-1 image) are used to train multi-feature sea-land classifier, and the rest three images are applied as testing images. In the experiments, all of VRSS-1 and Gaofen-1 images are first downsampled by a factor of 2/5 and then used in different sea-land segmentation methods. The ground truth of VRSS-1 and Gaofen-1 images are obtained by manual annotation and shown in Figure 6.

Figure 7 presents one original VRSS-1 testing image, two original Gaofen-1 testing images and their corresponding sea-land segmentation images obtained from four methods. From traditional segmentation results, one can observe that land regions as dark as sea are wrong classified as sea area. With similar grayscale value to sea area, large forest regions (Figure 7(a)) and broad farmlands (Figure 7(b, c)) are misjudged as sea in traditional methods. For traditional segmentation results shown in Figure 7(d-l), these misjudgments make land area interfered by distributed black holes to different extents, and it is difficult to recognize sea-land boundaries. In contrast, the above problems barely exist in segmentation images

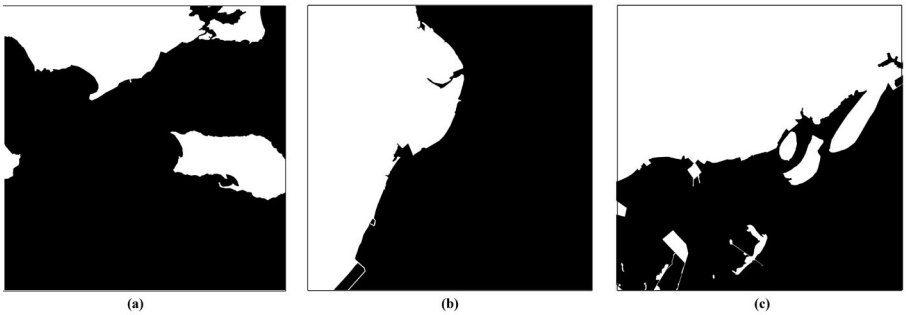

Figure 6. The ground truth data of VRSS-1 and Gaofen-1 images. (a) Original VRSS-1 testing image, (b-c) original Gaofen-1 testing images.

Table 3. The right rate and average value of four sea-land segmentation methods on VRSS-1 and Gaofen-1 images

	No.1	No.2	No.3	Ave
OTSU	0.7529	0.6766	0.7997	0.7431
LBP	0.6774	0.6430	0.7828	0.7011
ME	0.8689	0.6995	0.9079	0.8254
Ours	**0.9893**	**0.9860**	**0.9934**	**0.9896**

Table 4. The error rate and average value of four sea-land segmentation methods on VRSS-1 and Gaofen-1 images

	No.1	No.2	No.3	Ave
OTSU	0.0573	0.1671	0.0714	0.0986
LBP	0.0746	0.1824	0.0775	0.1115
ME	0.0331	0.1569	0.0334	0.0745
Ours	**0.0075**	**0.0141**	**0.0033**	**0.0083**

obtained from our proposed method.

3.2.2. Quantitative analysis

Same quantitative estimations, LR and ER, are applied to evaluate four different methods. Here, we provided two tables indicating LR and ER of four sea-land segmentation methods on VRSS-1 and Gaofen-1 images. Indices with bold types represent the best ones in four sea-land segmentation methods. Table 3 shows the whole LR values and the average value of each segmentation method on VRSS-1 and Gaofen-1 images. Table 4 shows the whole ER values and the average value of each segmentation method on VRSS-1 and Gaofen-1 images.

Table 3 and Table 4 show that the proposed sea-land segmentation method outperforms traditional segmentation methods in large degrees, with the highest LR of 0.9896 and the lowest ER of 0.0083. These results are in accordance with visual comparison (Figure 7) where many land regions in testing images are misjudged as sea in traditional methods while identified correctly in the proposed method.

From the above experiments on Google-Earth, VRSS-1 and Gaofen-1 images, one can observe the efficiency of the proposed method compared with traditional sea-land segmentation methods. In land area, large-scale vegetation (forests, farmlands) is difficult to segment because it usually involves similar feature information with sea area. Land regions like residential area and industrial zone, contain complex feature information and thus become a difficult area for sea-land segmentation. It is usually hard to solve these problems by traditional segmentation methods based on single feature and prior model. In contrast, good performances in the above

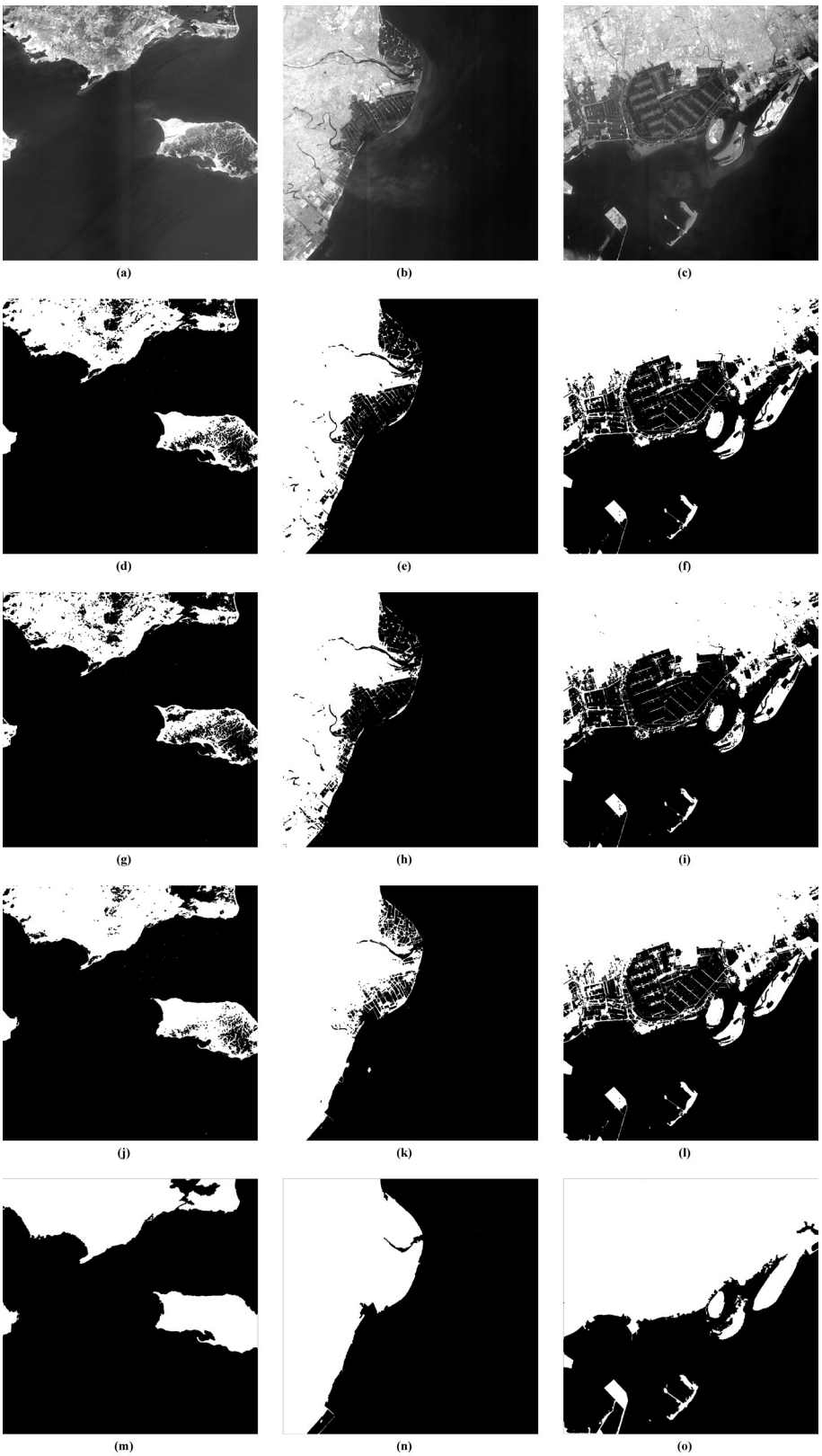

Figure 7. Visual comparison of sea-land segmentation results based on VRSS-1 and Gaofen-1 images. (a) Original VRSS-1 testing images, (b-c) original Gaofen-1 testing images; segmentation results based on (d-f) OTSU, (g-i) LBP, (j-l) ME, and (m-o) the proposed method.

testing images illustrates that our multi-feature sea-land segmentation approach is more self-adaptive and robust.

4. Conclusions

In this paper, sea-land segmentation issue is converted into a pixel-wise binary classification problem based on our proposed supervised learning based method. We then designed a multi-feature sea-land classifier with superior adaptability and robustness. The proposed multi-feature sea-land classifier is trained through each pixel in images instead of traditional rectangular patch because nearly no appropriate patch size could match amorphous boundaries between sea and land. Meanwhile, multi-feature sea-land segmentation algorithm is proposed based on the approximation of Newton method to design sea-land classifier. Experimental results and discussion on Google-Earth, VRSS-1 and Gaofen-1 images demonstrate that our proposed sea-land segmentation method yields more robust and accurate sea-land segmentation results.

Acknowledgment(s)

The work was supported by the National Natural Science Foundation of China under the Grants 61671037 and 61273245, the Beijing Natural Science Foundation under the Grant 4152031, the funding project of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University under the Grant BUAA-VR-16ZZ-03, and the Fundamental Research Funds for the Central Universities under the Grant YWF-16-BJ-J-30.

References

An, Z., and Z. Shi. 2015. "Scene Learning for Cloud Detection on Remote-Sensing Images." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8 (8): 4206–4222.

Aujol, Jean-François, Guy Gilboa, Tony Chan, and Stanley Osher. 2006. "Structure-Texture Image Decomposition—Modeling, Algorithms, and Parameter Selection." *International Journal of Computer Vision* 67 (1): 111–136.

Corbane, Christina, Laurent Najman, Emilien Pecoul, Laurent Demagistri, and Michel Petit. 2010. "A complete processing chain for ship detection using optical satellite imagery." *International Journal of Remote Sensing* 31 (22): 5837–5854.

Ham, B., M. Cho, and J. Ponce. 2015. "Robust image filtering using joint static and dynamic guidance." In *IEEE Conference on Computer Vision and Pattern Recognition,* 4823–4831.

Han, J., D. Zhang, G. Cheng, L. Guo, and J. Ren. 2015. "Object Detection in Optical Remote Sensing Images Based on Weakly Supervised Learning and High-Level Feature Learning." *IEEE Transactions on Geoscience and Remote Sensing* 53 (6): 3325–3337.

Huang, Y., W. Wang, and L. Wang. 2015. "Conditional High-Order Boltzmann Machine: A Supervised Learning Model for Relation Learning." In *IEEE International Conference on Computer Vision,* 4265–4273.

Kanopoulos, Nick, Nagesh Vasanthavada, and Robert L Baker. 1988. "Design of an image edge detection filter using the Sobel operator." *IEEE Journal of Solid-State Circuits* 23 (2): 358–367.

Kapur, J.N., P.K. Sahoo, and A.K.C. Wong. 1985. "A new method for gray-level pic-

ture thresholding using the entropy of the histogram." *Computer Vision, Graphics, and Image Processing* 29 (3): 273–285.

Li, W., L. Duan, D. Xu, and I. W. Tsang. 2014. "Learning With Augmented Features for Supervised and Semi-Supervised Heterogeneous Domain Adaptation." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36 (6): 1134–1148.

Liu, G., Y. Zhang, X. Zheng, X. Sun, K. Fu, and H. Wang. 2014. "A New Method on Inshore Ship Detection in High-Resolution Satellite Images Using Shape and Context Information." *IEEE Geoscience and Remote Sensing Letters* 11 (3): 617–621.

Man, Qixia, Pinliang Dong, and Huadong Guo. 2015. "Pixel- and feature-level fusion of hyperspectral and lidar data for urban land-use classification." *International Journal of Remote Sensing* 36 (6): 1618–1644.

Mehrotra, R., K. R. Namuduri, and N. Ranganathan. 1992. "Gabor filter-based edge detection." *Pattern Recognition* 25 (12): 1479–1494.

Otsu, N. 1979. "A Threshold Selection Method from Gray-Level Histograms." *IEEE Transactions on Systems, Man, and Cybernetics* 9 (1): 62–66.

Sagiv, Chen, N. A. Sochen, and Y. Y. Zeevi. 2006. "Integrated active contours for texture segmentation." *IEEE Transactions on Image Processing* 15 (6): 1633–1646.

Soille, P. 2004. *Morphological image analysis: principles and applications*. Springer-Verlag New York, Inc.

Tang, J., C. Deng, G. B. Huang, and B. Zhao. 2015. "Compressed-Domain Ship Detection on Spaceborne Optical Image Using Deep Neural Network and Extreme Learning Machine." *IEEE Geoscience and Remote Sensing Letters* 53 (3): 1174–1185.

Xia, Y., S. Wan, P. Jin, and L. Yue. 2014. "A Novel Sea-Land Segmentation Algorithm Based on Local Binary Patterns for Ship Detection." *International Journal of Signal Processing, Image Processing and Pattern Recognition* 7 (3): 237–246.

Yang, G., B. Li, S. Ji, F. Gao, and Q. Xu. 2014. "Ship Detection From Optical Satellite Images Based on Sea Surface Analysis." *IEEE Geoscience and Remote Sensing Letters* 11 (3): 641–645.

You, X., and W. Li. 2011. "A sea-land segmentation scheme based on statistical model of sea." In *4th International Congress on Image and Signal Processing,* Vol. 31155–1159.

Zhang, X., and Z. Wang. 2010. "Coastline extraction from remote sensing image based on improved minimum filter." *2nd IITA International Conference on Geoscience and Remote Sensing* 2: 44–47.

Zhao, K., W. S. Chu, F. De la Torre, J. F. Cohn, and H. Zhang. 2016. "Joint Patch and Multi-label Learning for Facial Action Unit and Holistic Expression Recognition." *IEEE Transactions on Image Processing* 25 (8): 3931–3946.

Zhu, C., H. Zhou, R. Wang, and J. Guo. 2010. "A Novel Hierarchical Method of Ship Detection from Spaceborne Optical Image Based on Shape and Texture Features." *IEEE Transactions on Geoscience and Remote Sensing* 48 (9): 3446–3456.