Small Sample Set Inshore Ship Detection From VHR Optical Remote Sensing Images Based on Structured Sparse Representation

Yin Zhuang, *Member, IEEE*, Lianlin Li, *Senior Member, IEEE*, and He Chen

***Abstract*—Inshore ship detection from very high resolution (VHR) optical remote sensing images has been playing a critical role in various civil and military applications. However, it brings up an important challenge, which is difficult to complete effective and robust feature extraction when valid inshore ship training sample acquired is limited, and the severe imbalance problem exists of positive and negative samples. In order to tackle the abovementioned difficulties, the structured sparse representation model (SSRM) is proposed to achieve inshore ship detection in more effectively and robustly way by circumstances of the small sample set. Here, SSRM has two steps that include inshore ship region proposal (RP) and orientation prediction (OP). Related to the RP process, the error matrix embedded in SSRM not only prevents to build the high-dimension background subdictionary and imbalance problem of positive and negative samples, but also achieves an effective intraclass robustness description of inshore ships and background. For the OP stage, the low-rank constraint of common sharing atoms in SSRM can make inshore ship direction be extracted by their sparse coding. In addition, based on RP and OP guidance, the proposed comprehensive structure voting can achieve an accurate contour detection of inshore ships. Finally, several experimental results employ that Google Earth service, HRSC 2016, and DOTA datasets proved the effectiveness of the proposed method. The results show that proposed inshore ship detection method can provide approximately 83.7% *Recall* and 72.3% *Precision* by using only over 100 positive training samples, which outperforms the state of the art methods.**

***Index Terms*—Inshore ship detection, optical remote sensing, sparse representation (SR), small sample set, very high resolution (VHR).**

I. Introduction

With the development of very high resolution (VHR) remote sensing technology, it has been widely used for the ship salvage, port traffic trade control, and ship oil spill monitoring applications [1]–[5]. Related to these applications, automatic ship detection technology has been attracted more attention. In particular, inshore ship detection is a critical technology for civil and military investigations or monitoring systems [2], [4]. However, it is a challenging task, because sometimes the valid inshore ship training sample acquired is limited, and there is a severe imbalance problem of training samples when building a robust detection model. First, it is hard or impossible to annotate massive inshore ship samples from large view scale VHR optical remote sensing images, as shown in Fig. 1. Therefore, we have to explore the small sample set learning way to achieve effective and robust inshore ship detection. On the other hand, some unpredicted interferences (i.e., jetties, convex banks, rectangular roofs, etc.) from complex background contain more categories, and they all have a great probability to be false alarms. Thus, considering these issues, various automatic inshore ship detection methods have been proposed [6]–[30], and these methods can roughly divided into three categories: manually designed feature (MDF)-based methods [6]–[12], parameter space transformation (PST)-based methods [13]–[17], and automatically feature learning based methods [18]–[30].

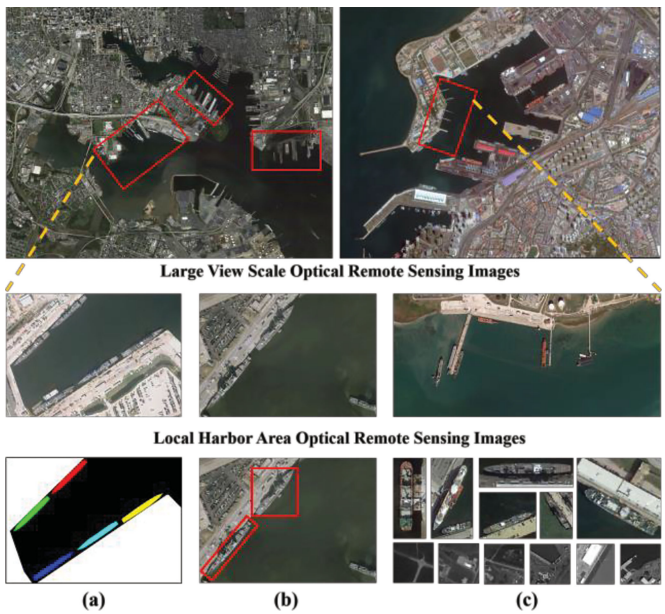


Fig. 1. Optical remote sensing images annotation. (a) HRSC 2016 image with a refined pixel-level annotation. (b) DOTA image with a bounding box-level annotation. (c) Google Earth image with image-level annotation.

Manuscript received September 19, 2019; revised November 29, 2019, January 8, 2020, March 13, 2020, and April 5, 2020; accepted April 6, 2020. Date of publication April 23, 2020; date of current version May 29, 2020. This work was supported in part by the Chang Jiang Scholars Program under Grant T2012122, in part by the Hundred Leading Talent Project of Beijing Science and Technology under Grant Z141101001514005, and in part by the China Postdoctoral Science Foundation under Grant 2019M650345. *(Corresponding author: He Chen.)*

Yin Zhuang and Lianlin Li are with the School of Electronic Engineering and Computer Science, Peking University, Beijing 100871, China (e-mail: zhuangyin640829@163.com; lianlin.li@pku.edu.cn).

He Chen is with the Beijing Key Laboratory of Embedded Real-time Information Processing Technology, Beijing 100081, China (e-mail: chenhe@bit.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2020.2987827

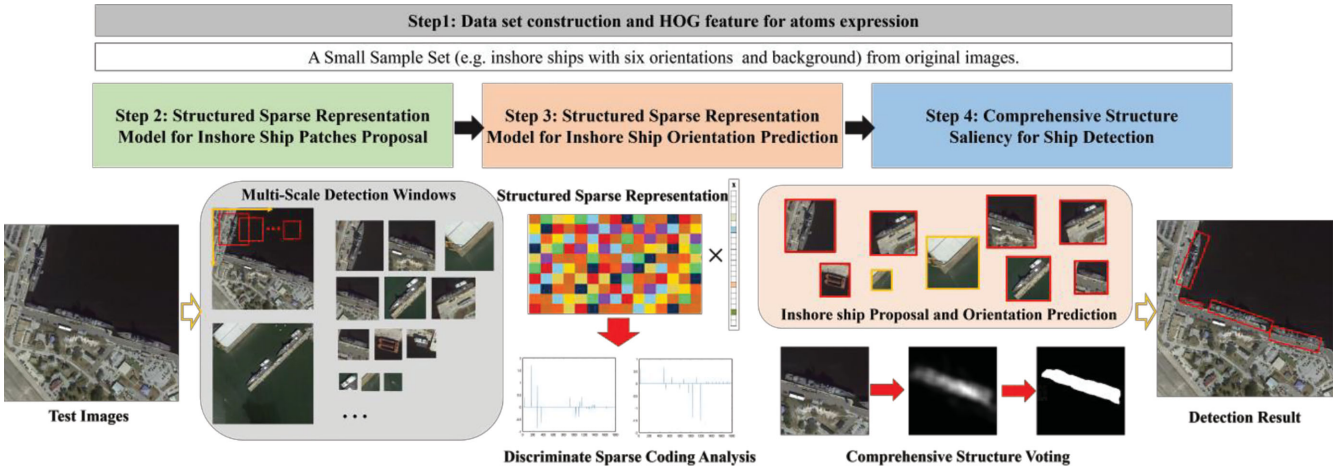


Fig. 2. Proposed structured SR inshore ship detection framework.

MDF-based methods employ low-level feature descriptions (e.g., length–width ratio, parallel lines, hull texture, fore "V" shaped structure, contour edge, and HOG) for local or global feature description to achieve inshore ship detection. Then, due to the description capability limitation of low-level feature, these methods cannot perform well for inshore ship detection. To provide accurate inshore ship detection results, the PST-based methods are often considered for refined inshore ship detection against complex harbor background. Although these method has a certain ability for inshore ship detection when the valid training sample acquired is limited. However, when they face the complex and cluster harbor background interferences, these methods are difficult to balance the over- or underfitting feature description in the parameter space domain. Therefore, these PST-based methods often lead to inshore ship leakage or produce a lot of false alarms. In order to improve the inshore ship detection performance, several convolution neural network (CNN) and sparse representation (SR)-based methods have been proposed. Because of their outstanding feature description ability, they are becoming the main object detection framework. Zhang *et al*. [19] proposed an S-CNN-based inshore ship detection method, which used the "V" structure and parallel line characteristics to achieve inshore ship candidate proposals. Lin *et al.* [21] also proposed a fully convolutional network (FCN) with a task partitioning strategy for inshore ship detection. Li *et al.* [23] proposed the multiscale deep feature embedding way in a faster-RCNN framework to achieve the inshore ship detection from optical remote sensing images. In addition to CNN-based methods, Yokoya and Iwasaki [27] proposed the SR-Hough object detection method consisting of the inshore ship and background subdictionaries. Here, the inshore ship patches were extracted by sparse coding feature, and Hough voting was employed for inshore ship refined locations. Chen *et al.* [29] proposed the label constraint SR model combining with a semisupervised superpixel clustering training strategy of multiple-category backgrounds to achieve object detection. However, including CNN- and SR-based inshore ship detection methods, they all require a large number of training samples to train the CNN architecture and high-dimensional SR dictionary. Especially for the SR-based method, they have to use the abundant inshore ship and background training samples to update atoms and their sparse coding coefficients alternately and achieve an intraclass completeness description in high-dimensional subdictionary. Since the lower dimensional subdictionary, which needs less training samples to setup, has a certain lack of target and background feature description ability. Therefore, for SR-based inshore ship detection methods, the challenge problem is to use the small sample set to train a powerful lower dimensional SR dictionary and achieve the inshore ship detection from optical remote sensing complex harbor background.

In this article, related to the problems mentioned before, the small sample set structured SR model (SSRM) for inshore ship detection is proposed, as shown in Fig. 2. Here, the SSRM is built for inshore ship region proposal (RP) and orientation prediction (OP). First, for RP, the error matrix embedded lower dimensional structured SR dictionary can adapt the intraclass variation descriptions of inshore ships and backgrounds. Then, related to OP stage, the low-rank constraint of common sharing atom structure for lower dimensional SR dictionary can provide the accurate OP results from RP results. Finally, based on RP and OP guidance, the comprehensive structure voting (CSV) method is proposed to achieve the refined inshore ship contour detection. Finally, we employed inshore ship datasets of Google Earth service, HRSC 2016, and DOTA to demonstrate that the proposed SSRM and CSV methods can achieve the refined inshore ship contour detection and reach approximately 83.7% *Recall* and 72.3% *Precision* by using only over 100 positive training samples than the state-of-the-art methods. In general, the major contributions of this article can be summarized as follows.

1) The proposed SSRM and CSV can achieve better refined inshore ship contour detection by the challenging small sample set than the state-of-the-art methods.
2) We addressed the imbalance problem of training samples between inshore ships and backgrounds by the constructed special structure of low-dimensional SSRM than the traditional SR dictionary designing, which is high dimensional and hardly be trained by a large number of various training samples.

3) We also proposed the rapid and effective CSV method for refined inshore ship contour detection based on the RP and OP guidance without complex parameter space calculation.

The rest of this article is organized as follows. In Section II, we briefly introduce the basic principle of SR. Then, in Section III, we describe the proposed SSRM and CSV inshore ship detection framework in detail. Next, extensive experiments of the proposed method are presented in Section IV. Finally, in Section V, the comparative performance of small sample set are discussed.

II. Basic Concepts of SR

This research study aims to achieve and demonstrate the refined inshore ship detection by a small sample set. In this article, the SR principle is employed. Recently, the SR-based method is widely used for object detection [27]–[29], [33], [34], image enhancement [35], [36], and classification [37]–[39]. The SR principle can be explained as a given set of input data, and it can be combined with a few linear elements from a set of representation patterns. Here, these representation patterns are called atoms, the set of all atoms is called a dictionary, and the coefficients of the linear combinations are called sparse coding. A classification example can be provided by considering a set of input signals$\{y_1, y_2, \ldots, y_p\} \subset \mathbb{R}^N$ and then determining a set of atoms $\{d_1, d_2, \ldots, d_M\} \subset \mathbb{R}^N$ together with a set of sparse coding coefficient vectors $\{x_1, x_2, \ldots, x_p\} \subset \mathbb{R}^M$ [40]. Therefore, each input vector y_i can be approximated by a linear combination $y_i \approx \sum_{l=1}^{M} x_j(l) d_l$, where most entries of $x_j$are zeroes or close to zeroes. Here, $\|\cdot\|_1$ denote the L_1-norm constraint of sparse coding. Then, the classic sparse coding problem can be expressed as follows:

$$\min_{D,C} \|Y - DX\|_F^2, \text{ s.t. } \forall i, \|x_i\|_0 \leq T \tag{1}$$

where $D = [d_1, d_2, \ldots, d_M] \in \mathbb{R}^{N \times M}$denotes the learned dictionary, and $Y = [y_1, y_2, \ldots, y_P] \in R^{N \times P}$denotes input samples as column vectors in a matrix. $X = [x_1, x_2, \ldots x_P] \in \mathbb{R}^{M \times P}$denotes the matrix containing the corresponding sparse coding vectors. Then, the threshold T determines the sparsity degree on each sparse coding vector. Here, the normalization constraint can prevent possible unbounded solutions, which states that $\|d_j\|_2 = 1$ for all indexes j. Equation (1) is the dictionary learning regulation rule, which considers only the approximate error between the input data and their resultant succinct expression. In recent years, for the sparse coding process, an abundance of literature has focus on discriminating sparse coding generation to improve the sparse coding discriminative ability [38], [39], [41]. Here, the basic idea of discriminative sparse coding is generated by some supervised learning strategies in the sparse coding process. Then, the most existing approaches for discriminative sparse coding are based on the following expression:

$$\min_{D,C} \|Y - DX\|_F^2 + \gamma J(X; L). \tag{2}$$

Subject to $\|x_i\|_0 \leq T$,$\|d_j\|_2 \leq 1$, for all i and j, where γ is the impact weight. L is a matrix that encodes the label information of training sample, and $J(\cdot; L)$ denotes a penalty function that measures the discriminative error between the codes and related to labels. A lot of research studies are developing $J(\cdot; L)$ to achieve many works [33]–[39]. In this article, we utilize the basic SR principle to build the structured SR and generate the discriminative sparse coding to achieve inshore ship detection based on a small sample set.

III. Proposed Methodology

As shown in Fig. 2, this article proposed the small sample set inshore ship detection framework that is consist of three major components.

1) SSRM for inshore ship RP;
2) SSRM for inshore ship OP;
3) CSV for refined inshore ship contour detection.

These three components are detailed as follows.

A. SSRM for Inshore Ship RP

For inshore ship RP, the proposed SSRM is shown in Fig. 3. Here, it contains three subdictionaries, which represent the categories of inshore ships, background, and error matrix. Then, in SSRM, the inshore ship and background subdictionaries can jointly ensure their interclass discriminative description of sparse coding by sparsity and priori labeled sparse coding constraints. Next, the error matrix embedded in SSRM can reflect intraclass variances of inshore ship and background, which can leverage inshore ship or background completeness description and small sample set learning. In order to train the special SSRM for inshore ship RP, the small training sample set has to be setup. First, similar to Fig. 2 step 1, we manually select 350 image-level inshore ship samples with identical direction and different types. Then, we rotate these samples in six directions (i.e., 30°, 60°, 90°, 120°, 150°, and 180°) to adapt the orientation changes. Next, based on selected positive inshore ship training samples, we randomly select enough background samples and utilize the zero-mean normalized cross correlation [27] method to automatically find an equal number of background samples that are more similar to the 350 rotated inshore ships. When we set up the balanced image-level small sample set, an 8×8 cell HOG feature descriptor [29] and the principle component analysis algorithm are employed for basic atom expression and then regulated all atoms into 576×1. Similar to Fig. 2 step 2, in Fig. 3, the blue matrix D_t is the inshore ship subdictionary, and the gray matrix D_b is the background subdictionary. Where the white matrix E_r is the error matrix that is the important part of SSRM to leverage the intraclass completeness description and lower dimensional dictionary training by a small sample set. Therefore, the SSRM can be expressed in three parts as follows:

$$D_E = [D_t, D_b, E_r]. \tag{3}$$

where D_t and D_b are the trainable dictionary, and E_r is the fixed matrix of diagonal unit array. The dimension of E_r is according to the atom vector's. Then, Let $D_E = [D'_E, E_r]$; next, $D'_E = [D_t, D_b]$. From the SR principle introduced before, if given an

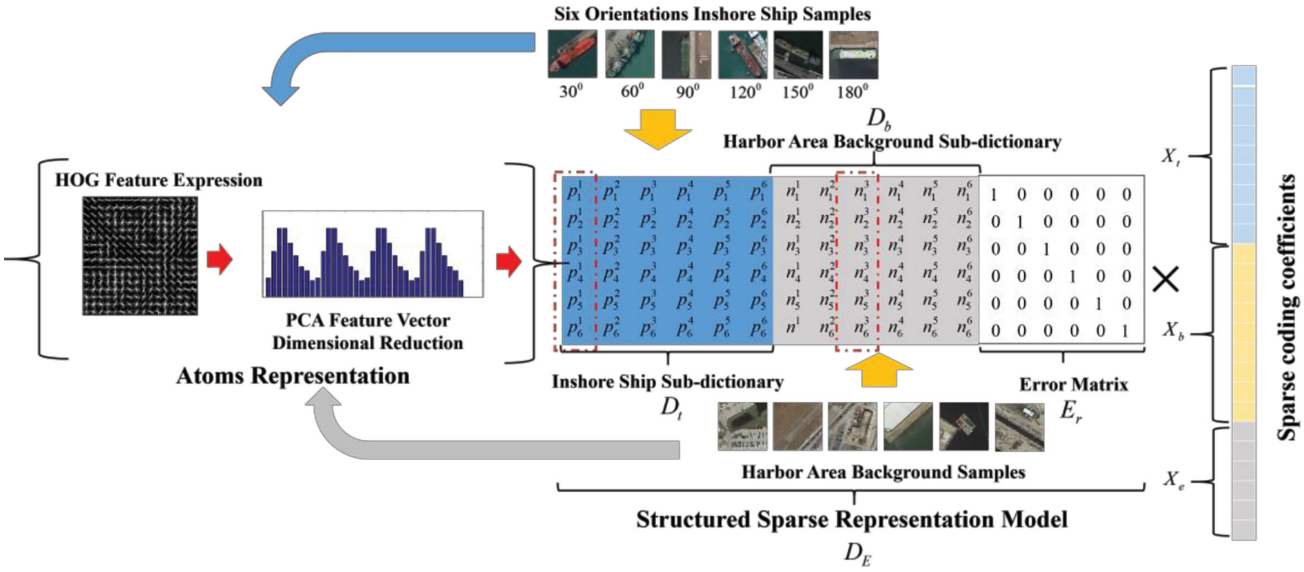


Fig. 3. Structured SR for inshore ship RP.

observation sample Y, we can reconstruct Y by using D_E and D'_E, as calculated in the following equations:

$$\begin{aligned} Y_1 \approx D_E X = &\ (d_{1,t}\alpha_{1,t} + \cdots + d_{n,t}\alpha_{n,t}) \\ &+ (d_{1,b}\alpha_{1,b} + \cdots + d_{n,b}\alpha_{n,b}) \\ &+ (e_{1,r}\alpha_{1,r} + \cdots + e_{n,r}\alpha_{n,r}). \end{aligned} \tag{4}$$

$$\begin{aligned} Y_2 \approx D'_E X = &\ (d_{1,t}\alpha_{1,t} + \cdots + d_{n,t}\alpha_{n,t}) \\ &+ (d_{1,b}\alpha_{1,b} + \cdots + d_{n,b}\alpha_{n,b}). \end{aligned} \tag{5}$$

From (4) and (5), $D_t = [d_{1,t}, d_{2,t}, \ldots d_{n,t}]$, $D_b = [d_{1,b}, d_{2,b}, \ldots d_{n,b}]$, and $E_r = [e_{1,r}, e_{2,r}, \ldots e_{n,r}]$. $X = [\alpha_{1,t}, \ldots \alpha_{n,t}, \alpha_{1,b}, \ldots \alpha_{n,b}, \alpha_{1,r}, \ldots \alpha_{n,r}]$ and $X = [\alpha_{1,t}, \ldots \alpha_{n,t}, \alpha_{1,b}, \ldots \alpha_{n,b}]$ represent the sparse coding coefficients of D_E and D'_E, respectively. There are two optimum solutions Y_1 and Y_2, both of which are similar to observation sample Y. Here, Y_1 is more similar to Y because of its higher dimensional SR dictionary. Then, assume that Y_1 is equal to Y. Therefore, the reconstruction error Err can be calculated as follows:

$$\text{Err} = Y - Y_2 = Y_1 - Y_2 = e_{1,r}\alpha_{1,r} + \cdots + e_{n,r}\alpha_{n,r}. \tag{6}$$

In (6), the reconstruction error can be expressed by the error matrix in D_E. Here, an example of E_r can be expressed as (7) when the atom vector is 6×1. The error matrix can be embedded into SSRM to measure the intraclass variance of inshore ships or background by its sparse coding coefficients distributed in error matrix part

$$E_r = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{7}$$

Then, before the intraclass variance investigation by error matrix, we have to train the discriminative structured dictionary of inshore ship and background parts to make inshore ships be detected from complex harbor background. Inspired by the label consistent K-SVD algorithm [38], [42], we fixed the error matrix E_r and trained $D'_E = [D_t, D_b]$ as

$$\begin{aligned} \langle D'_E, A, X \rangle = \arg \min_{D'_E, A, X} &\ \|Y - [D_t, D_b]\, X\|_2^2 \\ &+ \alpha \|Q - AX\|_2^2 + \gamma \|x\|_1. \end{aligned} \tag{8}$$

For the D'_E dictionary training process, several inshore ship and background samples are individually employed to initialize each subdictionary by using the K-SVD algorithm [42]. Then, following (8), K-SVD and LASSO algorithms [43] are used to train D'_E and generate discriminative sparse coding alternately. Here, first term in (8) is the reconstruction error of observation sample Y, and the second term can force the sparse code X set toward the expected sparse coding set of Q. The third term is the L_1-norm constraint for each sparse code x in X. The example of the expected sparse code Q can be expressed as follows:

$$Q = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}. \tag{9}$$

In (9), each column of Q is the expected sparse code of trained D'_E. Here, the example considers that the first three column elements in Q are expected sparse code coefficients of the inshore ship and that the last three column elements are background expected sparse code coefficients. The priori sparse coding Q expects that if the observation sample is an inshore ship, its nonzero sparse code coefficients have to occur in the inshore ship subdictionary. Otherwise, its nonzero sparse code coefficients

have to occur in the background subdictionary. Then, when the observation function is minimized in (8), D'_E could generate more discriminative sparse codes by their expected Q constraint. Related to D'_E updating training process, it would alternatively update D'_E and the relevant sparse code coefficients by K-SVD and LASSO algorithms. However, in the updating process, we also have to initialize A as follows:

$$A = \arg\min_{A} \|Q - AX\|^2 + \lambda \|A\|_2^2. \tag{10}$$

By using multivariable ridge regression [44], quadratic loss, and L_2-norm regularization, A can be initialized as follows:

$$A = QX^t\left(XX^t + \lambda I\right)^{-1} \tag{11}$$

when the dictionary D'_E is initialized, the sparse code X can be generated. In (11), X is used to initialize transformation matrix A. When A is initialized and meets the minimum reconstruction error with L_1-norm sparsity constraint of X, we can follow the observation function (12) to optimize (10). In (12), $L^i(D'_E, y_i, A, q_i) = \|Q - AX\|^2$

$$\begin{aligned}
&\min_{D,A} \sum_i L^i\left(D'_E, y_i, A, q_i\right) + \frac{v}{2}\|A\|_F^2 \\
&\text{s.t. } \mathrm{x}_i = \arg\min_x \|y_i - D'_E x\|_2^2 + \gamma\|x\|_1, i \in \{1, \ldots, N\} \\
&\|d_j\|_2^2 \le 1, j \in \{1, \ldots, K\}.
\end{aligned} \tag{12}$$

With respect to (12), the dictionary D'_E and transform matrix A updating process can be formed as follows:

$$D'^{(t)}_E = D'^{(t)}_E - \rho_t \frac{\partial L^i}{\partial D'^{(t)}_E} \tag{13}$$

$$A^{(t)} = A^{(t)} - \rho_t \frac{\partial L^i}{\partial A^{(t)}} \tag{14}$$

where ρ_t is the learning rate for the updating process. Then, the stochastic gradient descent algorithm [45] is used for gradient optimization. However, sometimes the sparse coding coefficients perform all zero values, which lose the gradient information. Therefore, we have to lead the auxiliary variable ϕ and $\phi \in \mathbb{R}^K$. If the sparse code coefficients are all zeroes, then $\phi = 0$. Otherwise, $\phi = (D'^t_E D'_E)^{-1}\frac{\partial L^i}{\partial x_i}$ and $\frac{\partial L^i}{\partial x_i} = A^t(Ax_i - q_i)$. Next, (13) and (14) can be calculated via (15) and (16), respectively

$$\frac{\partial L^i}{\partial D'_E} = -D'_E \phi x_i^t + (y_i - D'_E x_i)\phi^t \tag{15}$$

$$\frac{\partial L^i}{\partial A} = (Ax_i - q_i)\, x_i^t + \nu A. \tag{16}$$

Following D'_E and A updating process by minimizing the observation function (8), we use trained D'_E to generate more discriminative sparse code according to observation sample of inshore ship or background. Next, the trained D'_E is combined with an error matrix, as shown in Fig. 3, and the LASSO algorithm [43] is employed to produce the sparse coding coefficients that include the error matrix part. Finally, related to setup SSRM and generated sparse coding coefficients, the proposed confidence value is used to extract the inshore ship patches from multiscale detection windows and achieve inshore ship RP. Here, the inshore ship confidence value can be calculated by the following equations:

$$\text{confidence} = \begin{cases} \dfrac{w_t \sum_{i=1}^{N_t} x_i}{\sum_{i=N_t+1}^{N_b} x_i + \sum_{i=N_b+1}^{N_{Er}} x_i + eps}, & \text{if } w_t > w_b. \\ \dfrac{\sum_{i=1}^{N_t} x_i}{w_b\left(\sum_{i=N_t+1}^{N_b} x_i + \sum_{i=N_b+1}^{N_{Er}} x_i + eps\right)}, & \text{if } w_t \le w_b. \end{cases} \tag{17}$$

$$S = \frac{1}{1 + \exp\left(-\text{confident}\right)}. \tag{18}$$

In (17), w_t or w_b is the maximum sparse coding coefficient in inshore ship or background part. N_t, N_b, and N_{Er} are the sizes of D_t, D_b, and E_r in SSRM, respectively. Then, the multiscale sliding window samples are all resized into 120 × 120 and expressed as a 576 × 1 HOG atom vector to generate their sparse coding coefficients. When the observation samples are inshore ships, most of sparse coding coefficients occur in inshore ship subdictionary, and fewer coefficients are in the background and error matrix subdictionaries. Otherwise, due to background containing the variance interferences with a larger intraclass difference, there are insufficient atoms to support completeness description of background subdictionary, and if the observation samples are background or unseen interference samples, it would produce many sparse coding coefficients in background or error matrix part. Consequently, we can utilize the inshore ship or background part combined with an error matrix part to produce discriminative sparse coding coefficients under small sample set training. Then, we can also calculate the *confidence* value by using (17) according to their different intraclass variances. In (17), *eps* is a very small number to prevent zero in the denominator. Next, following inshore ship *confidence* value calculation, (18) regulates the confidence value from "0" to "1." If multiscale windows provide the observation sample with an inshore ship, (18) provides an S closer to "1." Finally, we can utilize a confidence threshold value of 0.9 or more than 0.9 to achieve rapid and accurate inshore ship RP.

B. SSRM for Inshore Ship OP

Regarding refined inshore ship detection, many methods [13]–[18] consider using inshore ship orientation to produce refined detection results. However, orientation information is sometimes difficult to extract from RP patches. Therefore, a novel SSRM for inshore ship OP is proposed, which describes the feature similarity of the proposed inshore ship patches and enhances their orientation difference over a small sample set, as shown in Fig. 4. The principle is also shown in Fig. 5, and it shows the feature similarity and differential expression.

There is an example of three classes to represent the different orientation inshore ship patches. These patches have more similar global structure and different orientations. Therefore, in feature representation space, we are aiming to sufficiently learn their similarity to identify the oriented difference. Based on this

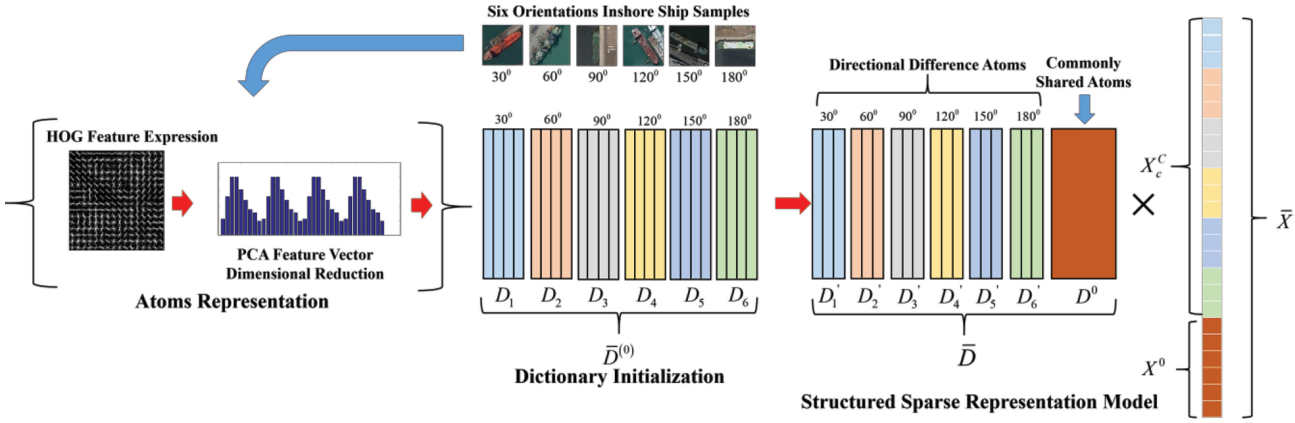


Fig. 4. Structured SR for inshore ship OP.

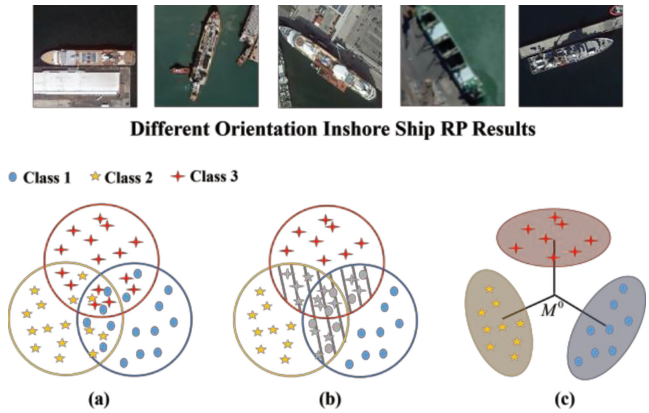


Fig. 5. Principle of the common sharing atom SSRM for abundant feature similarity expression.

principle, we set up the common sharing atom SSRM as follows:

$$L_{\overline{Y}}\left(\overline{D},\overline{X}\right) = \frac{1}{2}f_{\overline{Y}}\left(\overline{D},\overline{X}\right)+\lambda_1\left\|\overline{X}\right\|_1+\frac{\lambda_2}{2}g\left(\overline{X}\right)+\eta\left\|D^0\right\|_*. \quad (19)$$

$\overline{D}$ is the common sharing atom dictionary, which includes two parts (i.e., the multiclass subdictionaries and common sharing subdictionary). In Fig. 4, the brown section is the common sharing subdictionary that can be expressed as D^0, and other different colors of subdictionary parts D'_nrepresent different oriented classes. Here, n is the class index. Then, $\bar{D} = [D'_n, D^0]$. $\bar{X}$is discriminative sparse code, which is produced by trained $\bar{D}$. Therefore, $\bar{X}$ also has two parts $\bar{X} = [X'_n, X^0]$.X'_n represents differential oriented sparse coding coefficients, and X^0 represents common sharing sparse coding coefficients. Here, we utilize $\bar{D}$ and $\bar{X}$ to achieve the implicit feature space similarity expression for inshore ship different orientation identification, where $\bar{X}$ is generated by the fast iterative shrinkage-thresholding algorithm (FISTA) [46]. λ_1, λ_2, and η are weight parameters to control the impacts of these three terms in (19). In (19), $f_{\bar{Y}}(\bar{D},\bar{X})$and $g(\bar{X})$ can leverage SSRM to generate a more discriminative dictionary and sparse coding coefficients, which are introduced as follows.

From (19), $f_{\bar{Y}}(\bar{D},\bar{X})$is used to ensure the minimized reconstruction error of observation samples according to common sharing subdictionary and differential oriented subdictionaries. Next, $\|\bar{X}\|_1$ is the sparsity constraint for generated discriminative sparse coding, and $\frac{\lambda_2}{2}g(\overline{X})$ is the common sharing sparse coding similarity constraint. Finally, $\|D^0\|_*$ is a low-rank constraint [39], [40] to ensure that the model sufficiently learned the common sharing atoms from different oriented inshore ship patches, and it can also prevent D^0 absorbing the discriminative atoms from differential oriented subdictionaries [47]. Then, in order to predict the inshore ship orientations of RP patches, $f_{\bar{Y}}(\bar{D},\bar{X})$ can be obtained as follows:

$$f_Y\left(\bar{D},\bar{X}\right) = \sum_{n=1}^{N} \left(\left\|Y_n-\bar{D}\bar{X}\right\|_2^2+\left\|Y_n-D'_nX'_n-D^0X^0\right\|_2^2+\sum_{j\neq n}\left\|D_jX_n^j\right\|_2^2\right). \quad (20)$$

In (20), it is the Fisher discrimination dictionary learning method [49] that maintains interclass distinguishability. N is the number of inshore ship orientation classes. First and second terms are reconstruction errors of observation sample Y_n by using differential orientations and common sharing atom subdictionaries. The third term represents an unrelated orientation atoms D_j with a small contribution to represent the observation sample Y_n. Related to constructed differential orientations and common sharing atom subdictionaries, here the constraint of sparse coding coefficients $g(\overline{X})$ can be obtained as follows:

$$g\left(\bar{X}\right)=\sum_{n=1}^{N}\left(\left\|\bar{X}_n-M_n\right\|_2^2-\left\|M_n-M\right\|_2^2\right)+\left\|X^0-M^0\right\|_2^2. \quad (21)$$

Here M_n represents the mean vector of sparse coding coefficients for identical orientated inshore ship samples. M is the sparse coding coefficients' mean vector of whole orientated inshore ship samples. M^0 is the mean sparse coding vector from common sharing subdictionary. Then, the first term in (21) is used to ensure the small difference of sparse coding coefficients

in each orientation intraclass, and the second term is used to make the big difference of sparse coding coefficients in each orientation interclasses. The last term in (21) can ensure that all oriented classes have more similar common sharing sparse coding coefficients for similar structure description.

Then, by the constraints presented in (20) and (21) of observation function (19), $\bar{D}$ and $\bar{X}$ are alternatively updated. The trained $\bar{D}$ can be used to generate $\bar{X}$ to predict the inshore ship orientation from RP patches. For $\bar{D}$ and $\bar{X}$ training process, when dictionary $\bar{D}$ is updating, then $\bar{X}$ is fixed. Let $\bar{D} = [D'_{\mathrm{n}}, D^0]$ solves D'_n and D^0 separately. Related to D'_n updating process, which must satisfy $Y_n \triangleq Y - D^0 X^0$. D'_n can employ the online dictionary learning [48] to update their atoms through the following equation:

$$D'_n = \arg\min_{D'_n} \left\{ -2\mathrm{trace}\left(ED^T\right) + \mathrm{trace}\left(FD^T D\right) \right\} \quad (22)$$

where $E = Y_n \Phi(X'^T_n)$ and $F = \Phi(X'_n X'^T_n)$. The operation $\Phi(\cdot)$ can be defined as follows:

$$\underbrace{\begin{bmatrix} A_{11} & \cdots & A_{1n} \\ A_{21} & \cdots & A_{2n} \\ \cdots & \cdots & \cdots \\ A_{n1} & \cdots & A_{nn} \end{bmatrix}}_{A} \mapsto A + \underbrace{\begin{bmatrix} A_{11} & \cdots & 0 \\ 0 & \cdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \cdots & A_{nn} \end{bmatrix}}_{\Phi(A)}. \quad (23)$$

Here, (23) is the given matrix A. Then, $\Phi(A)$ means double diagonal blocks of A. Next, D^0 updating follows the following equation:

$$D^0 = \arg\min_{D^0} \left\{ \mathrm{trace}\left(FD^{0T} D^0\right) - 2\mathrm{trace}\left(ED^{0T}\right) \right\} + \eta \left\| D^0 \right\|_*. \quad (24)$$

In (24), let $E = [Y - \frac{1}{2} D'_n \Phi(X'_n)] \cdot (X^0)^T$ and $F = X^0 (X^0)^T$. The alternating direction method of multipliers [50] is employed to update D^0. When the dictionary $\bar{D}$ is updated, the sparse coding $\bar{X}$ is updated by the FISTA [46] followed by

$$\nabla \bar{h}\left(\bar{X}\right) = \begin{bmatrix} \frac{\partial \bar{h}_{X^0}(X'_n)}{\partial X'_n} \\ \frac{\partial \bar{h}_{X'_n}(X^0)}{\partial X^0} \end{bmatrix}. \quad (25)$$

For (25), the gradient of sparse coding coefficients can be calculated by the following equations:.

$$\frac{\partial \bar{h}_{X^0}(X'_n)}{\partial X'_n} = \left(\Phi\left(\bar{D}^T \bar{D}\right) + 2\lambda_2 I\right) X'_n - \Phi\left(\bar{D}^T \bar{Y}\right) + \lambda_2 (M_n - 2M) \quad (26)$$

$$\frac{\partial \bar{h}_{X'_n}(X^0)}{\partial X^0} = 2D^{0T} D^0 X^0 - 2D^{0T}\left(Y - \frac{1}{2} D'_n \Phi(X'_n)\right) + \lambda_2 \left(X^0 - M^0\right). \quad (27)$$

For detailed derivations of (26) and (27), readers can be referred to [51]. Next, when $\bar{D}$ and $\bar{X}$ are alternatively updated, we set up the SSRM of common sharing atoms $\bar{D}$ and utilize (28) to generate discriminative sparse coding $\bar{X} = [X'_n, X^0]$.

Algorithm I: SSRM for Inshore Ship RP and OP.

1. **Input**: The input optical remote sensing harbor images *I* and proposed RP results *P*;
2. For S = 1, 2, … , n do
3. Selected the detection window scale *S*;
4. for i = 1, 2, … , n
5. for j = 1, 2, … , n do
6. get detection windows P_{ij} from input image *I*; Calculated the confident value for each P_i by SSRM-RP;
7. If confident value > threshold *T*
8. predict the orientation of P_{ij} by SSRM-OP
9. **Output** the patches P_{ij} and its' orientation θ
10. End

Finally, let generated $\bar{X}$ be substituted into (29) that can predict the inshore ship orientation from extracted high confidence value RP patches

$$\overline{X} = \arg\min_{\overline{X}} \frac{1}{2} \left\| \overline{Y} - \overline{DX} \right\|_2^2 + \frac{\lambda_2}{2} \left\| X^0 - M_0 \right\|_2^2 + \lambda_1 \left\| \overline{X} \right\|_1 \quad (28)$$

$$\mathrm{ORI}_{\mathrm{predict}} = \arg \min_{1 \le n \le N} \left(\omega \left\| \overline{Y} - D'_n X'_n - D^0 X^0 \right\| + (1 - \omega) \left\| X'_n - M_n \right\| \right). \quad (29)$$

In general, related to the OP process, the inshore ship RP and OP framework can be summarized as Algorithm I.

C. CSV for Refined Inshore Ship Contour Detection

In order to rapidly obtain the refined inshore ship contour detection results and eliminate false alarms, the CSV method is proposed followed by RP and OP guidance. Here, we consider using hull parallel lines and inshore ship symmetric texture distribution to achieve the inshore contour detection. These discriminative structures of the inshore ship had been proven in previous works [6]–[12]. Then, these parallel lines can be extracted by the line segmentation detection (LSD) method [52], which follows the OP guidance to rapidly find any two longest lines in each scale RP patch. Between these two lines, there is a great probability of inshore ship interior texture, which is composed of several symmetrical basic structure elements (i.e., lines and corners), as shown in Fig. 6.

Then, Hough line [53] and Harris corner [54] detections are used for symmetry element analysis, which can be expressed as

$$PG = \left\{ HT(\rho, \theta), R = |M| - k \cdot tr^2(M) \right\} \quad (30)$$

where PG represents the detected point group of lines and corners, $HT(\cdot)$ represents the Hough transform for line detection, and $R = |M| - k \cdot tr^2(M)$ is the Harris corner detection. Here, *M* is the eigenmatrix of the pixel's gradient between the parallel line area, and it can be expressed as $M = [I_x^2, I_x I_y; I_x I_y, I_y^2]$. I_x and I_y are the horizontal and vertical gradients, respectively, and k is the controlling parameter of corner detection performance. The detail steps of refined inshore ship contour detection is

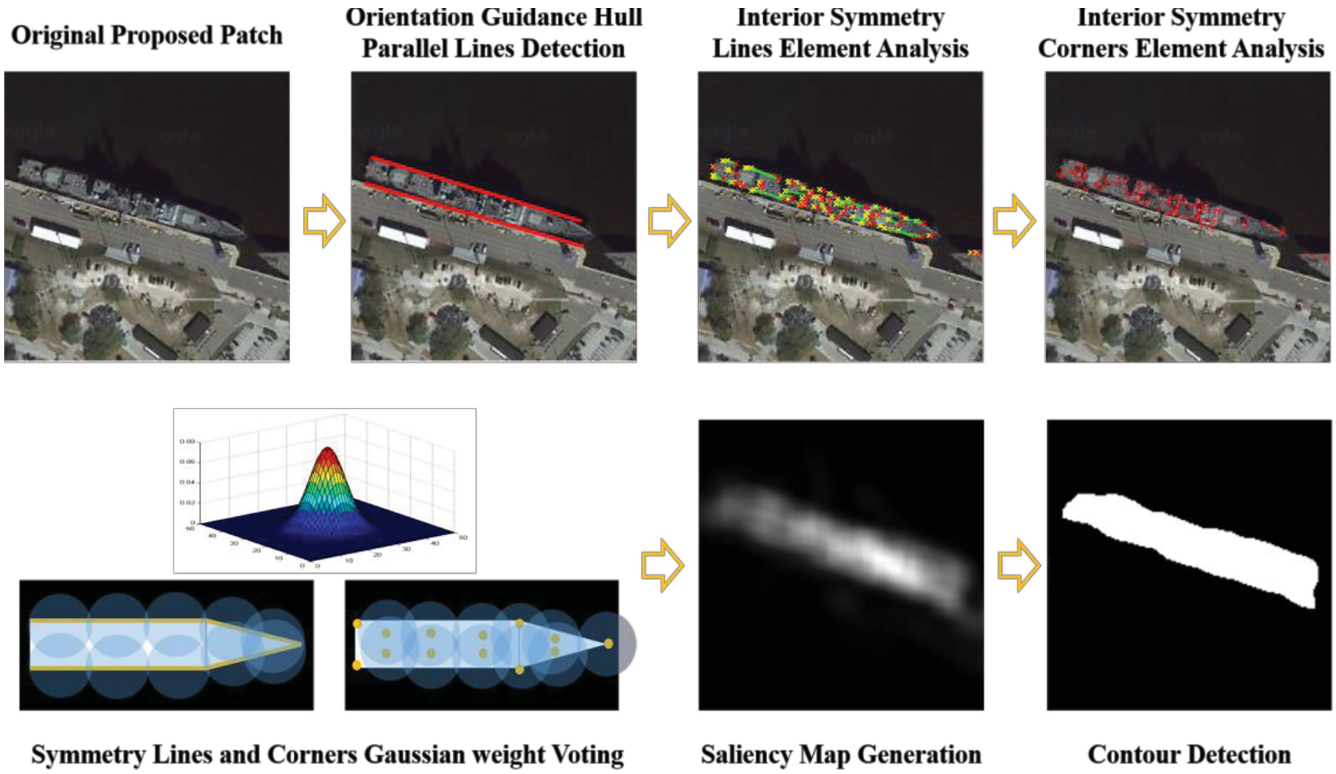


Fig. 6. CSV process for refined inshore ship detection.

Algorithm II: Comprehensive Structure Voting Method for Refined Inshore Ship Contour Detection.

1. **Input**: The proposed inshore ship patches P and predicted orientation θ;
2. For i = 1, 2, … , n do
3. LSD [52] lines detection for P_i;
4. If line degree is equal to θ
5. Mark the area A between the two longest parallel lines;
6. Obtain the PG points group of lines and corners detection for A_i.;
8. Generate the saliency map: $SM_j = PG{\cdot}G(\mu,\sigma)$
9. Winner take all and region growing analysis based on P_i and SM_i
10. End

shown in Algorithm II. Following the lines and corners analysis in the parallel lines areas, we can obtain the PG of the interior texture features. Then, the Gaussian weight is assigned to each point to generate the inshore ship saliency map from the proposed multiscale patches by aggregation of Gaussian weight overlap adding. Finally, based on each saliency map, the original patch region growing and saliency map winner-take-all algorithms are used to generate a refined inshore ship detection binary mask.

IV. Experiments and Analysis

To demonstrate the effectiveness of the proposed small sample set inshore ship detection framework. Here, several relevant datasets and evaluation indexes are involved, and the optimal parameter discussion and comparisons are described in detail as follows.

A. Dataset and Evaluation Indexes

To demonstrate the method proposed in this article, Google Earth service, HRSC 2016 [31], and DOTA [32] are employed. Here, Google Earth service data were manually downloaded from remote sensing browsing platform by using geographic coordinate information (i.e., Mombasa, Kenya, Baltimore, and the California Santiago naval base). Then, we randomly select 233 VHR optical remote sensing harbor scene testing images from Google Earth service, HRSC 2016 [31], and DOTA [32], which include the 1021 inshore ships. Next, we train all the comparison methods based on a built small sample set. Here, the small sample set contains 350 identical orientation training samples of inshore ships, and the total number of relevant false alarms are 2100. Then, we also employ indexes of *accuracy*, *precision*, *recall*, and *Overall Accuracy* to evaluate comparison methods' performance. They can be expressed as follows:

$$accuracy = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{P}+\mathrm{N}} \tag{31}$$

$$precision = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \tag{32}$$

$$recall = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{33}$$

$$overall\ accuracy = \frac{\mathrm{PL}}{M}. \tag{34}$$

TABLE I
DIFFERENT SIZES OF STRUCTURED SR MODEL DISCRIMINATIVE DESCRIPTION WITHOUT ERROR MATRIX PERFORMANCE

Dictionary Size	600	960	1320	1680
Iteration Times	8	10	35	52
Learning Rate	0.3	0.3	0.3	0.3
Learning Decline Rate	1×10^{-3}	1×10^{-3}	1×10^{-3}	1×10^{-3}
Batch Size	300	300	300	300
Regularization weight υ	1×10^{-6}	1×10^{-6}	1×10^{-6}	1×10^{-6}
Dictionary Initialization Iterations	10	10	10	10
Sparsity Constraint Weight γ	1×10^{-6}	1×10^{-6}	1×10^{-6}	1×10^{-6}
Training Time	510.3s	1025.1s	2749.2s	4397.6s
Accuracy	43%	49%	86%	83%

Here, TP is the true positive. FP is the false positive. FN is the false negative. P is the total number of positive samples in the testing dataset, and N is the total number of negative samples in the testing dataset, PL is the number of correctly predicted, and M is the total number of samples in testing dataset. In addition, to evaluate refined inshore ship contour detection, if IoU [56] is more than 0.7 comparing with ground truth, the actual detection window should be considered. Otherwise, it should be the false alarm. Finally, the SR-based methods were performed on Windows 10 environment by using MATLAB 2016b. The computer configuration is an Intel Corel i7-4500U CPU with 2.00 GHz and 7.71 GB RAM. Then, deep learning based methods were performed on an Ubuntu 16.04 operation system, which includes Python 4.0, CUDA 8.0, cudnn 7, and Pytorch 1.0. Then, the experiments were performed on an NVIDIA TITAN XP GPU workstation.

B. Optimization Parameter Discussion

1) SSRM for Inshore Ship RP: In the proposed SSRM for inshore ship RP, the error matrix is a very important part, which can ensure intraclass completeness description and facilitate to build lower dimensional structured SR dictionary. Thus, first, we set different sizes of structured SR models without error matrix part to discuss their dimension impact, as shown in Table I. Here, when the structured SR dictionary sizes are 600 (e.g., 300 inshore ship atoms and 300 background atoms) and 960 (e.g., 480 inshore ship atoms and 480 background atoms), they could not provide better inshore ship RP performance because of their insufficient dictionary dimensions. Then, when the dictionary dimensions increase to 1320 (e.g., 660 inshore ship atoms and 660 background atoms) and 1680 (e.g., 840 inshore ship atoms and 840 background atoms), we can see that they can obtain good performances. Otherwise, in Table I, the results show that 1680 dimensions require the considerably more time and training samples for its model convergence than 1320 dimensions', and performances of 1680 and 1320 also could not support practical detection systems. To further improve the inshore ship RP performance and achieve intraclass completeness description, error matrix is embedded into 1320 and 1680 dimensional structured SR models. Table II shows RP performances of structured SR model with error matrix embedded or without error matrix embedded.

TABLE II
PERFORMANCE OF THE PROPOSED STRUCTURED SR MODEL

Size of Structured SR model	Precision	Recall	Score
576×1320	76.2%	73.5%	0.90
576×1680	72.1%	72.7%	0.90
576×(1320+576)	82.5%	91.3%	0.90
576×(1680+576)	79.3%	85.4%	0.90

Here, we employ selected 233 VHR harbor area images that contain 1021 inshore ships to test these structured SR models in Table II, and inshore ship RP confidence value S is set more than 0.9. For inshore ship RP confidence value calculation, the structured SR model without error matrix does not consider error matrix sparse coding weights in formulas (17) and (18). Results presented in Table II show that the error matrix embedded into structured SR could improve the inshore ship RP performance, because error matrix sparse code coefficients significantly reflect inshore ship and background intraclass variance difference, and they can also ensure the intraclass completeness description. When some observation samples are not included in small training sample set, the proposed SSRM can utilize the intraclass variance difference to discriminate whether they belong to inshore ships or backgrounds by formulas (17) and (18), as shown in Fig. 7, because the inshore ship has a smaller intraclass difference than background's, which make inshore ship regions more discriminable.

In Fig. 7, these 12 sparse coding distributions are from $576 \times (1320 + 576)$ dimensional SSRM. For sparse coding, blue color represents sparse coding coefficients that are located in the inshore ship part. Red color represents sparse coding coefficients that are located in the background part. Green color represents sparse coding coefficients that are located in the error matrix part. In Fig. 7, six inshore ship observation samples are shown, and their sparse coding is shown in Fig. 7(a) and (b). We can see that the most of large sparse coding coefficients are located in inshore ship part, and a few of small coefficients are located in background and error matrix parts. For six background observation samples shown in Fig. 7(c) and (d), most of large sparse coding coefficients are located in the background part, and many coefficients appear in the error matrix part. Following the phenomena mentioned before, it has been proved that the priori information of presetting Q in (8) can certainly leverage the interclass discriminative description of inshore ships and background. These results also show that background usually has a larger intraclass variance than inshore ships, and we can utilize this character with a small sample set to build a

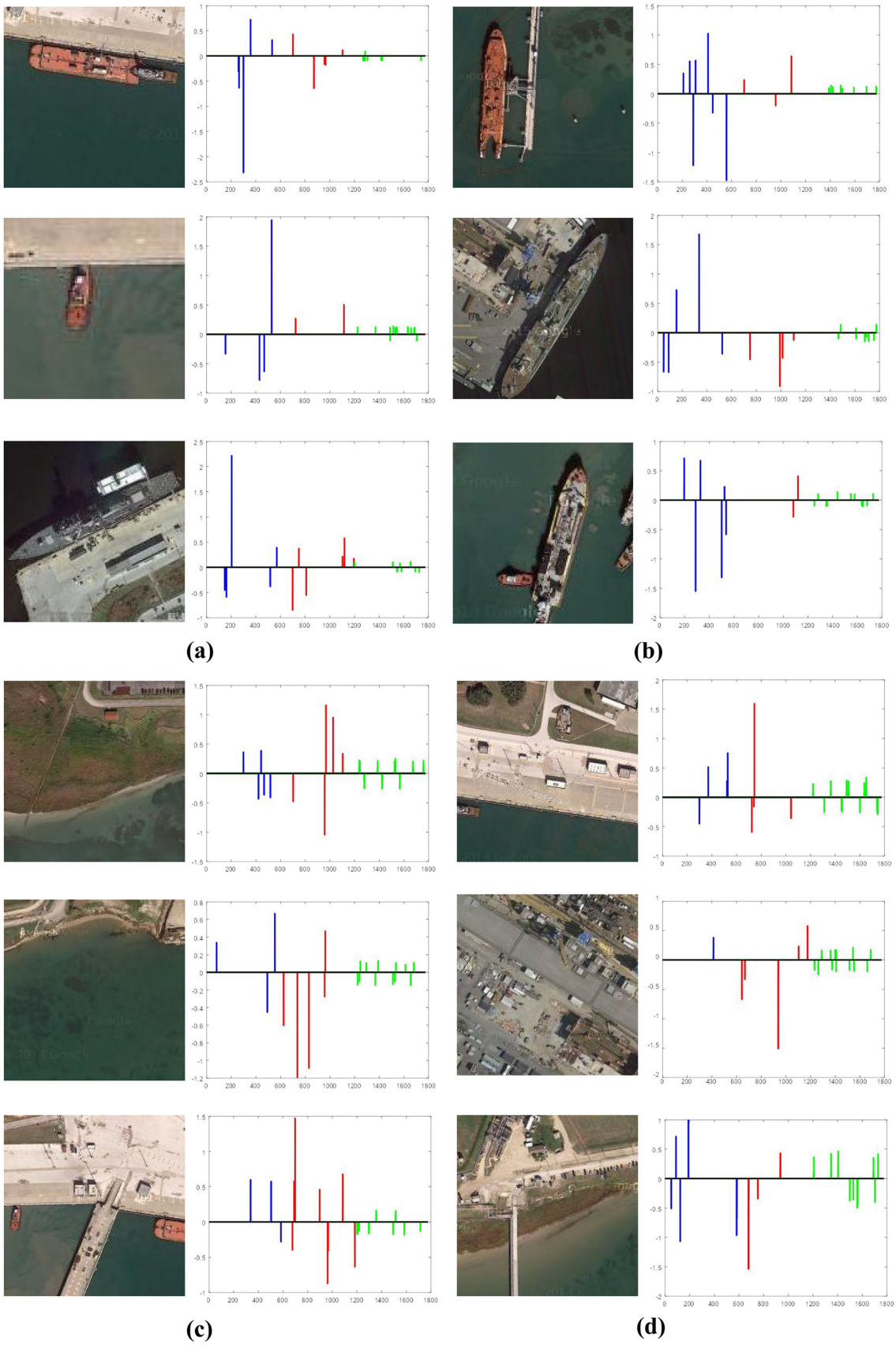


Fig. 7. Proposed SSRM sparse coding coefficients character analysis. (a) and (b) Inshore ship sparse coding coefficients distribution. (c) and (d) Background sparse coding coefficients distribution.

TABLE III
COMMONLY SHARED ATOM MODEL PARAMETERS

Parameter Setting	λ_1	λ_2	η
Parameter 1	/	0.01	0.02
Parameter 2	/	/	0.02
Parameter 3	/	0.01	/
Parameter 4	0.01	0.01	/
Parameter 5	0.01	/	0.02
Parameter 6	0.01	/	/
Parameter 7	0.01	0.01	0.02

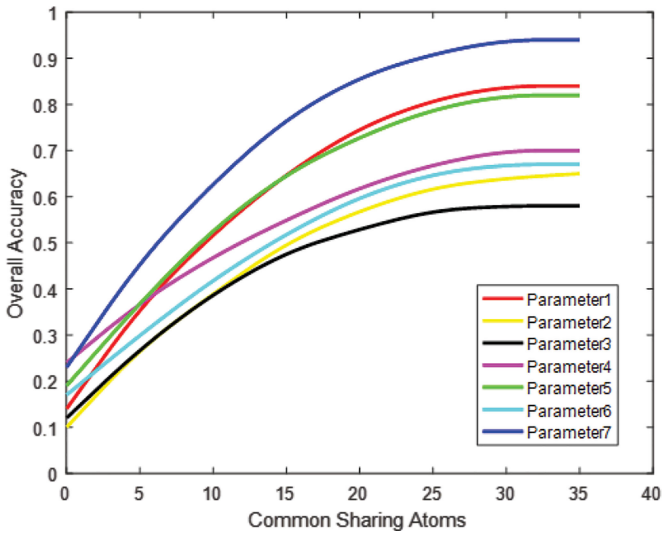


Fig. 8. Inshore ship OP performances with different common sharing atoms and impact factors setting.

powerful low-dimensional dictionary with completeness intr-aclass description. Finally, the proposed SSRM can rapidly extract inshore ship regions based on thresholding of confidence score.

2) SSRM for Inshore Ship OP: After the inshore ship RP stage, the SSRM for the inshore ship OP from each RP patch should be discussed in detail. As introduced in Section III-B, the impact factors of common sharing atoms and their low-rank constraint are mainly effects of OP performance according to (19). First, these multiorientation inshore ships, which are generated from small sample set, are used to test the proposed common sharing atom structured model. For sample set partition, we set 40 atoms for each oriented inshore ship to create the common sharing atom structured model and use 210 training samples for training. Next, we use the rest of 140 testing samples of each oriented inshore ship sample set to demonstrate the inshore ship OP performance. Before testing common sharing atom SSRM, some impact weights in (19) have to be set. Table III shows the seven different parameter settings that are used to demonstrate the effectiveness of each regulation term constraint in (19). Related to parameter settings in Table III, when the total number of atoms in common sharing subdictionary is changed, their performances are shown in Fig. 8.

TABLE IV
CSV ANALYSIS OF THE PROPOSED INSHORE SHIP DETECTION FRAMEWORK

Stages	Precision	Recall
RP + Oriented Guidance + CSV (Parallel lines)	84.4%	91.3%
RP + Oriented Guidance + CSV (Parallel lines & Internal lines)	86.5%	91.0%
RP + Oriented Guidance + CSV (Parallel lines & Internal Corner)	85.3%	90.7%
RP + Oriented Guidance + CSV (Parallel lines & Internal Corner & Internal Lines)	87.2%	90.2%

From Fig. 8, we can see that parameter 7 in Table III is proved to be more accurate parameter setting of inshore ship OP. Therefore, the three terms in (19) are important to affect the OP accuracy. Then, from Fig. 8, when more common sharing atoms are absorbed in the low-rank subdictionary, it can provide better OP performance, which means they have to use more than 28 common sharing atoms combined with low-rank constraint to make SSRM sufficiently learn the commonly shared part in feature description space and figure out the different inshore ship orientations.

3) CSV for Inshore Ship Contour Detection: Following inshore ship RP and their oriented information guidance, the effectiveness of the refined inshore ship detection CSV method is discussed. Here, the CSV method can avoid the false alarms and provide refined inshore ship detection results rapidly. In Table IV, structures of hull parallel lines, internal line, and corner distributions in the CSV method are tested on selected 233 harbor images. Here, from Table IV, first line shows that hull parallel lines are effective structures for refined inshore ship detection, because compared with the RP stage, the *precision* rate is improved from 82.5% to 84.4%, and the *recall* is kept as 91.3%. Then, second and third lines show the texture distributions of interior lines and corners also having false-alarm removing ability because of inshore ships having rich textural and symmetrical structural features. Finally, in fourth line, the proposed CSV method provides the *precision* and *recall* indexes of 87.2% and 90.2%, respectively. Fig. 9 shows some inshore ship detection examples by the proposed CSV method. In Fig. 9(a), there are four 1024×1024 harbor images with a 0.5-m spatial resolution. In Fig. 9(b) shows manually labeled GTs of Fig. 9(a). Fig. 9(c) and (d) shows inshore ship Gaussian voting saliency maps and their corresponding binary masks generated by the proposed CSV method. From Fig. 9(e), we can see that the proposed CSV method can provide more refined inshore ship detection results based on inshore ship RP and OP guidance.

C. Results and Comparison

After optimal parameter discussion, we select several state-of-the-art methods to compare with our proposed inshore ship detection method. First, these SR-based methods of SR-Hough

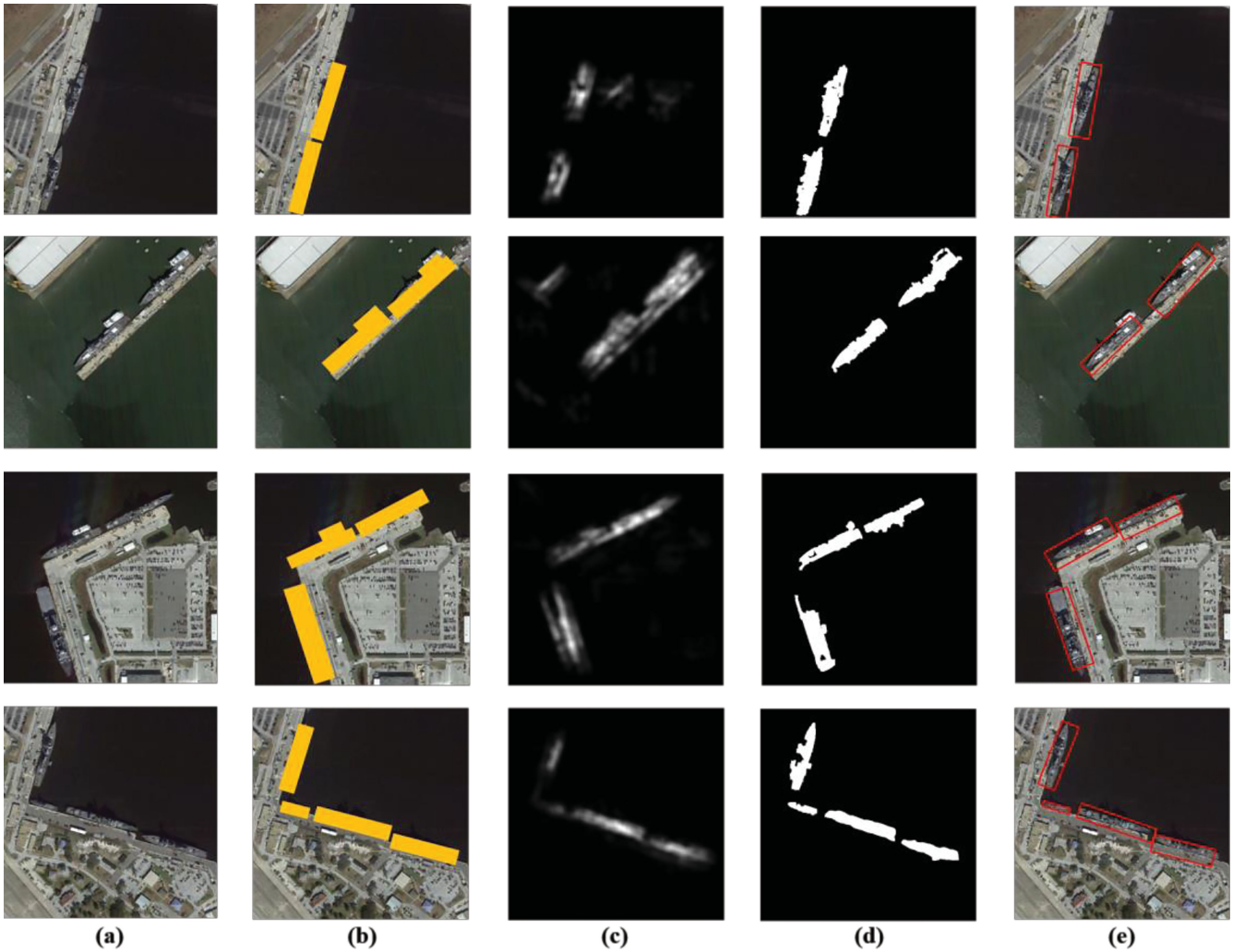


Fig. 9. Results of the proposed structured SR refined inshore ship detection. (a) Original harbor images. (b) GTs of inshore ships. (c) Saliency maps generated by the CSV method. (d) Region growing binary mask from saliency maps. (e) Refined inshore ship detection results of the proposed method.

[27], multilayer sparse coding (MLSC) [28], and SR-superpixels [29] are considered as comparison methods, because they use the SR principle to construct the structured SR dictionary and achieve detection, and these comparisons can better demonstrate the improvement and innovation of our work. Then, we also choose the PST methods, including weight pose voting (WPV) [17] and ship rotated bounding box space (SRBBS) [16], as comparison methods to prove that the proposed method has a better refined inshore ship locating ability. Finally, several famous deep learning networks are also involved because they perform well in object detection tasks. These methods are based on single-shot multibox detector (SSD) [55], FCN [21], and fast-RCNN [56].

Aiming at exploring the small sample set learning ability, we gradually reduce the number of training samples to train and test all of comparing methods. During the testing processing, all methods are tested on selected 233 harbor images from Google Earth service, HRSC 2016 [31], and DOTA [32]. Related to 1021 inshore ships of 233 harbor images, they can be roughly divided into two categories, which are the docking and moving ships. Performances of comparison methods are shown in Fig. 10. In Fig. 10(a), when training samples are expanded by selected 350 identical oriented inshore ships, here, most of the comparing methods have acceptable performances. However, related to deep-learning-based methods, these samples are not enough to train the robust CNN architectures. Therefore, deep-learning-based methods produce many false alarms and have the lower *recall* ratios, as shown in Table V. Next, following reducing training samples shown in Fig. 10(b)–(d), all comparing methods show severely decreasing *precision* and *recall* ratios. Only the method proposed in this article produces 72.3% *precision* and 83.7% *recall* with 170 identical oriented inshore ship training samples presented in Table V.

From Fig. 10 and Table V, the results demonstrated that the small training sample set can severely affect the inshore ship detection performance. First, related to SR-based comparison methods, the sizes of the training sample set and SR dictionary are important factors that affect detection performance. However, in our small sample set inshore ship detection experiments, there are insufficient samples to train 3000-dimension inshore ship subdictionary for setup SR-Hough [27]. In addition, its 100-dimension background subdictionary presented in [27] also

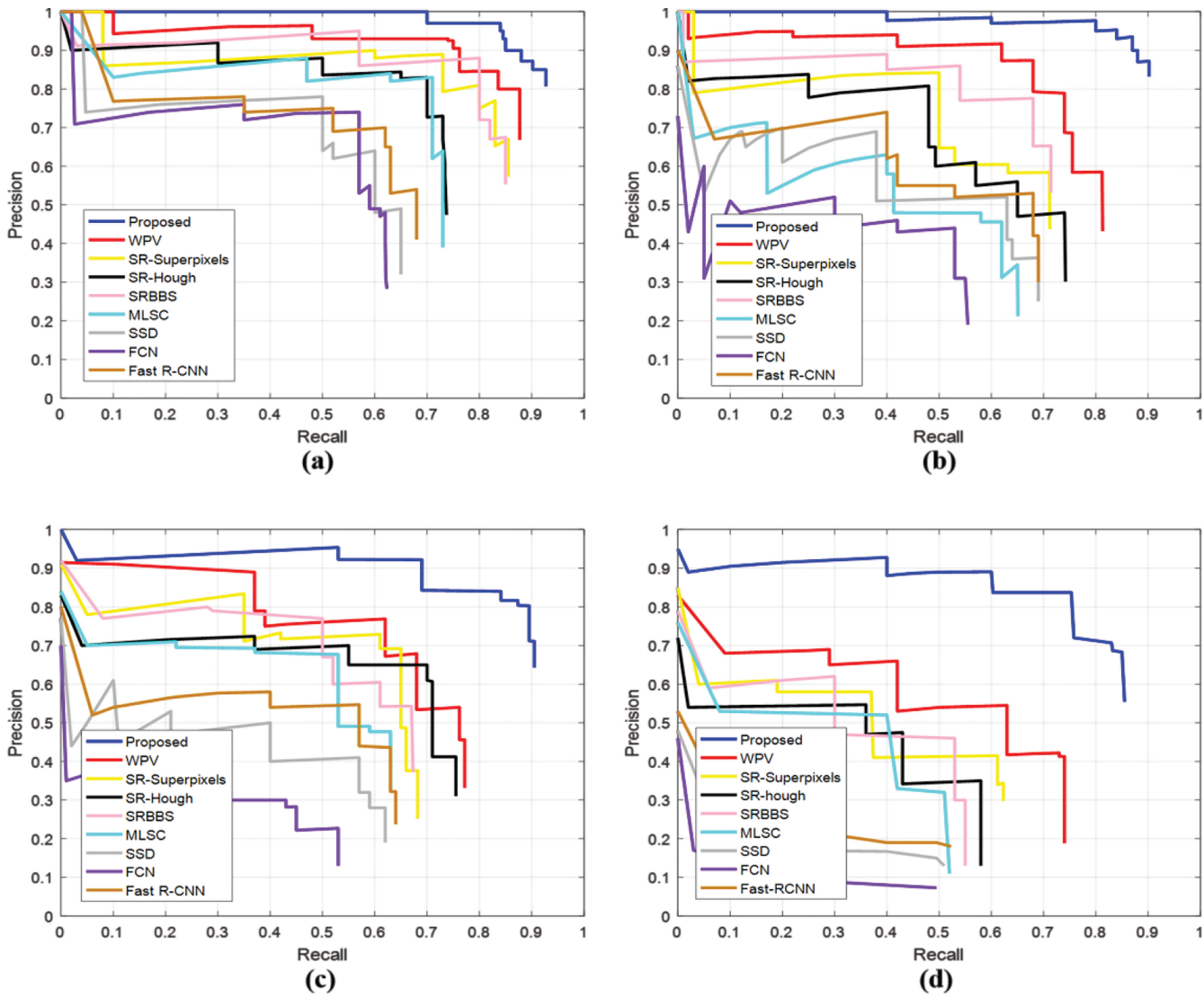


Fig. 10. Performances of comparison methods with different sizes of small sample set. (a) Performances of 350 training samples. (b) Performances of 290 training samples. (c) Performances of 230 training samples. (d) Performances of 170 training samples.

TABLE V
PERFORMANCE OF INSHORE SHIP DETECTION WITH COMPARISON METHODS

Method	WPV	SR-Superpixels	SR-Hough	SRBBS	MLSC	SSD	FCN	fast R-CNN	Proposed
350 Inshore Ship Training Samples									
Precision	84.6%	87.0%	73.0%	72.1%	83.0%	64.0%	74.1%	70.0%	**87.2%**
Recall	83.6%	83.0%	73.0%	82.2%	71.0%	60.0%	57.1%	62.0%	**90.2%**
290 Inshore Ship Training Samples									
Precision	78.9%	60.5%	57.0%	65.2%	45.6%	52.0%	44.0%	53.0%	**87.2%**
Recall	74.0%	63.2%	60.0%	68.0%	62.0%	63.0%	53.0%	68.0%	**90.3%**
230 Inshore Ship Training Samples									
Precision	53.5%	48.7%	60.7%	60.5%	49.1%	32.0%	28.2%	43.6%	**71.2%**
Recall	76.2%	66.0%	71.0%	61.0%	59.0%	59.0%	45.0%	63.0%	**89.5%**
170 Inshore Ship Training Samples									
Precision	52.0%	41.0%	35.0%	46.0%	32.0%	15.0%	9.0%	19.0%	**72.3%**
Recall	63.0%	61.2%	58.0%	53.0%	51.0%	49.5%	27.0%	49.5%	**83.7%**

Fig. 11. Proposed inshore ship detection framework results from large view scale optical remote sensing images.

cannot satisfy the intraclass completeness description of uncertain interferences from variance harbor background. These problems also exist in most SR-based object detection methods, such as SR-superpixel [29] and MLSC [28], because it is also difficult to train the background subdictionary with the infinity dimensional to adapt uncertain interferences by using these methods. Otherwise, in general SR-based detection methods, the lower dimensional dictionary or a few training samples has the certain limitation for inshore ship detection model training. Therefore, these state-of-the-art SR-based methods could not provide a good detection performance under small sample set. Next, related to PST-based methods, WPV [17] and SRBBS [16] have a certain ability for small sample set learning. However, these methods are seriously affected by docking ships, because the docking ships hardly achieve a completed contour edge extraction for robust parameter space description. Finally, the one-stage detector SSD [55] and two-stage Fast R-CNN [56] methods heavily rely on large quantities of data annotations, and therefore, they produce a poor detection performance by a small sample set. Then, the FCN [21] method considers the task partition of fore and aft distinguish parts to detect inshore ships, but these characters would also lead the extra false alarms from complex harbor background. By experimental analysis, these selected state-of-the-art methods could not provide good detection performance under inshore ship sample access limitation situation. Therefore, the proposed SSRM combined with the CSV method not only can achieve the small sample set learning for inshore ship detection, but also have better performance than state-of-the-art methods. Fig. 11 shows several large view scale inshore ship detection results that are produced by the proposed structured SR method.

V. Conclusion

In this article, we presented a novel inshore ship detection method for VHR optical remote sensing images by using the

proposed SSRM combined with the CSV method. First, related to small sample set, we constructed the SSRM with data argument. Here, SSRM contains three parts, which are target, background, and error matrix subdictionaries, and it can achieve rapid and accurate inshore ship RP from complex harbor background. Second, following built small sample set, the common sharing atom SSRM was proposed to predict inshore ship orientation based on RP results by low-rank constraint. Third, based on inshore ship RP and OP guidance, the CSV method used parallel lines, corner, and line distribution characteristics to highlight inshore ship regions. Then, the winner-take-all strategy and region growing algorithm were employed to generate refined inshore ship contour detection results. Finally, the proposed inshore ship detection framework was tested on selected datasets of the Google Earth service, HRSC 2016 [31], and DOTA [32]. Then, comparison experiments demonstrated that the proposed method can provide a better inshore ship detection performance based on a small sample set. It can be applied to a practical detection system, regardless of some inshore ship samples cannot be obtained, and harbor background includes various unknown interference situations.

The future work of this article will concentrate on the small sample set semisupervised way for inshore ship detection. Then, some problems, for example, the small inshore ship has unclear visual features and uncertain direction information, which can lead to small inshore ship miss detection, also need to be addressed in the future work. In addition, the inshore ship intensive parking situation is a hard example to occur leakage detection. Therefore, we will continue to address these problems in our future work.

REFERENCES

[1] H. Greidanus and N. Kourti, "Findings of the DECLIMS project—Detection and classification of marine traffic from space" in *Proc. SEASAR 2006: Adv. SAR Oceanogr. ENVISAT ERS Missions*, Frascati, Italy, Apr. 2006, pp. 1–6.

[2] R. Muller *et al.*, "Optical satellite services for EMSA (Opsserve)- near real-time detection of vessels and activites with optical satellite imagery," in *Proc. ESA Living Planet Symp.*, Edinburgh, U.K., Sep. 2013, p. 309.

[3] G. Máttyus, "Near real-time automatic marine vessel detection on optical satellite images" in *Proc. Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, Hannover, Germany, vol. XL-1/W1, Sep. 2013, pp. 233–237.

[4] Periodic Report Summary - DOLPHIN (Development of Pre-Operational Services for Highly Innovative Maritime Surveillance Capabilities), Publication Office CORDIS, Rep. 159422, Apr. 1, 2015.

[5] G. Lemoine, J. Chesworth, G. Schwartz-Juste, N. Kourti, and I. Shepherd, "Near real time vessel detection using spaceborne SAR imagery in support of fisheries monitoring and control operations," *in Proc . IEEE Int. Geosci. Remote Sens. Symp.*, Anchorage, AK, USA, Sep. 2004, pp. 4825–4828.

[6] G. Liu, Y. Zhang, X. Zheng, X. Sun, K. Fu, and H. Wang, "A new method on inshore ship detection in high-resolution satellite images using shape and context information," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 3, pp. 617–621, Sep. 2014.

[7] Z. Shi, X. Yu, Z. Jiang, and B. Li, "Ship detection in high-resolution optical imagery based on anomaly detector and local shape feature," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4511–4523, Aug. 2014.

[8] S. Qi, J. Ma, J. Lin, Y. Li, and J. Tian, "Unsupervised ship detection based on saliency and S-HOG descriptor from optical satellite images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 7, pp. 1451–1455, Jul. 2015.

[9] S. Haigang and S. Zhina, "A novel ship detection method for large-scale optical satellite images based on visual LBP feature and visual attention model," in *Proc. Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, Prague, Czech Republic, Jul. 2016, vol. XLI-B3, pp. 917–921.

[10] S. Li, Z. Zhou, B. Wang, and F. Wu, "A novel inshore ship detection via ship head classification and body boundary determination," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1920–1924, Dec. 2016.

[11] F. Xu, J. Liu, C. Dong, and X. Wang, "Ship detection in optical remote sensing images based on wavelet transform and multi-level false alarm identification," *Remote Sens*., vol. 9, no. 10, Sep. 2017, Art. no. 985.

[12] B. Jin, Y. Cong, W. Zhou, and G. Wang, "A new method for detection of ship docked in harbor in high resolution remote sensing image," in *Proc. IEEE Int. Conf. Prog. Informat. Comput.*, Shanghai, China, May 2014, pp. 341–344.

[13] F. Bi, J. Chen, Y. Zhuang, M. Bian, and Q. Zhang, "A decision mixture model-based method for inshore ship detection using high-resolution remote sensing images," *Sensors*, vol. 17, no. 7, p. 1470, Jun. 2017.

[14] W. G. Luo, W. Wang, F. Lang, and G. Gui, "Ship detection of remote sensing image on FRHT and multi-points curvature based polygon approximation," *Res. J. Appl. Sci., Eng.Technol.*, vol. 4, no. 15, pp. 2590–2599, Aug. 2012.

[15] J. Xu, X. Sun, D. Zhang, and K. Fu, "Automatic detection of inshore ships in high-resolution remote sensing images using robust invariant generalized hough transform," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 12, pp. 2070–2074, Dec. 2014.

[16] Z. Liu, H. Wang, L. Weng, and Y. Yang, "Ship rotated bounding box space for ship extraction from high-resolution optical satellite images with complex backgrounds," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 8, pp. 1074–1078, Aug. 2016.

[17] H. He, Y. Lin, F. Chen, H. Tai, and Z. Yin, "Inshore ship detection in remote sensing images via weighted pose voting," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3091–3107, Jun. 2017.

[18] X. Yang *et al.*, "Automatic ship detection in remote sensing images from google earth of complex scenes based on multiscale rotation dense feature pyramid networks," *Remote Sens*., vol. 10, no. 1, Jan. 2018, Art. no. 132.

[19] R. Zhang, J. Yao, K. Zhang, C. Feng, and J. Zhang, "S-CNN ship detection from high-resolution remote sensing images," in *Proc. Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, Prague, Czech Republic, Jun. 2016, vol. XLI-B7, pp. 423–430.

[20] Z. Zou and Z. Shi, "Ship detection in spaceborne optical image with SVD networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 5832–5845, Oct. 2016.

[21] H. Lin, Z. Shi, and Z. Zou, "Fully convolutional network with task partitioning for inshore ship detection in optical remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1665–1669, Oct. 2017.

[22] Z. Liu, J. Hu, L. Weng, and Y. Yang, "Rotated region based CNN for ship detection," in *Proc. IEEE Int. Conf. Image Process.*, Beijing, China, Sep. 2017, pp. 900–904.

[23] Q. Li, M. Lichao and L. Qingjie *et al.* HSF-Net: Multiscale deep feature embedding for ship detection in optical remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 12, pp. 7147–7161 Dec. 2018.

[24] H. Zhou, Y. Zhuang, L. Chen, and H. Shi, "Ship detection in optical satellite images based on sparse representation," in *Signal and Information Processing, Networking and Computers*, S. Sun, N. Chen, and T. Tian, Eds. Singapore: Springer, 2017, pp. 164–171.

[25] X. Wang and C. Chen, "Ship detection for complex background SAR images based on a multiscale variance weighted image entropy method," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 2, pp. 184–187, Feb. 2017.

[26] Y. Liu, M. Zhang, P. Xu, and Z. Guo, "SAR ship detection using sea-land segmentation-based convolutional neural network," in *Proc. Int. Workshop Remote Sens. Intell. Process.*, Shanghai, China, May 2017, pp. 1–4.

[27] N. Yokoya and A. Iwasaki, "Object detection based on sparse representation and Hough voting for optical remote sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 5, pp. 2053–2062, May 2015.

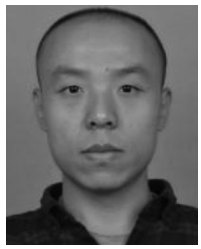
[28] H. Li, Z. Li, Z. Chen, and D. Yang, "Multi-layer sparse coding model-based ship detection for optical remote-sensing images," *Int. J. Remote Sens.*, vol. 38, no. 22, pp. 6281–6297, Jul. 2017.

[29] Z. Chen *et al.*, "Vehicle detection in high-resolution aerial images via sparse representation and superpixels," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 103–116, Jan. 2016.

[30] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, Jun. 2016.

[31] Y. Yang, L. Weng, L. Yuan, and Z. Liu, "A high resolution optical satellite image dataset for ship recognition and some new baselines," in *Proc. 6th Int. Conf. Pattern Recognit. Appl. Methods*, Jan. 2017, pp. 324–331.

[32] G.-S. Xia *et al.*, "DOTA: A large-scale dataset for object detection in aerial images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3974–3983.
[33] S. Agarwal and D. Roth, "Learning a sparse representation for object detection," in *Proc. 7th Eur. Conf. Comput. Vision-Part IV*, 2002, pp. 113–130.
[34] Z. Z. Li *et al.*, "Sparse representation for infrared dim target detection via a discriminative over-complete dictionary learned online," *Sensors*, vol. 14, no. 6, pp. 9451–9470, Jun. 2014.
[35] W. Shi, C. Chen, F. Jiang, D. Zhao, and W. Shen, "Group-based sparse representation for low lighting image enhancement," in *Proc. IEEE Int. Conf. Image Process.*, Phoenix, AZ, USA, Sep. 2016, pp. 4082–4086.
[36] P. Gifani, H. Behnam, F. Haddadi, Z. A. Sani, and M. Shojaeifard, "Temporal super resolution enhancement of echocardiographic images based on sparse representation," *IEEE Trans. Ultrason., Ferroelect., Freq. Control*, vol. 63, no. 1, pp. 6–19, Jan. 2016.
[37] J. Mairal, F. Bach, and J. Ponce, "Task-driven dictionary learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 791–804, Apr. 2012.
[38] Z. Jiang, Z. Lin, and L. S. Davis, "Label consistent K-SVD: Learning a discriminative dictionary for recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2651–2664, Nov. 2013.
[39] T. H. Vu and V. Monga, "Fast low-rank shared dictionary learning for image classification," *IEEE Trans. Image Process.*, vol. 26, no. 11, pp. 5160–5175, Nov. 2017.
[40] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
[41] J. Yang, K. Yu, and T. Huang, "Supervised translation-invariant sparse coding," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recognit.*, San Francisco, CA, USA, Jun. 2010, pp. 3517–3524.
[42] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.
[43] N. Morioka and S. I. Satoh, "Generalized Lasso based approximation of sparse coding for visual recognition," in *Proc. 24th Int. Conf. Neural Inf. Process. Syst.*, Granada, Spain, Dec. 2011, pp. 181–189.
[44] G. Golub, P. Hansen, and D. O'Leary, "Tikhonov regularization and total least squares," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 1, pp. 185–194, 1999.
[45] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proc. COMPSTAT'2010*, Y. Lechevallier and G. Saporta, *Eds*. Heidelberg, Germany: Physica-Verlag HD, 2010, pp. 177–186.
[46] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.
[47] B. Recht, M. Fazel, and P. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, no. 3, pp. 471–501, 2010.
[48] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, pp. 19–60, 2010.
[49] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Fisher discrimination dictionary learning for sparse representation," in *Proc. Int. Conf. Comput. Vision*, Barcelona, Spain, Nov. 2011, pp. 543–550.
[50] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.
[51] S. Kong and D. Wang, "A dictionary learning approach for classification: Separating the particularity and the commonality," in *Computer Vision—ECCV 2012*, A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, Eds. Berlin, Germany: Springer, 2012, pp. 186–199.
[52] R. G. V. Gioi, J. Jakubowicz, J. Morel, and G. Randall, "LSD: A fast line segment detector with a false detection control," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 4, pp. 722–732, Apr. 2010.
[53] J. J. Chen, "The improved approach of straight line segments detection based on Hough transform" *Microcomput. Inf.*, vol. 26, no. 21, pp. 211–213, 2010.
[54] B. F. Chen and Z. X. Cai, "Harris corner detection based on theory of scale-space," *J. Central South Univ. Technol.*, vol. 36, pp. 751–754, Oct. 2005.
[55] W. Liu *et al.*, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vision*, 2016, pp. 21–37.
[56] X. Wang, A. Shrivastava, and H. Mulam, "A-fast-RCNN: Hard positive generation via adversary for object detection" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Apr. 2017, pp. 2606–2615.



Yin Zhuang (Member, IEEE) was born in Henan, China, in 1990. He received the B.S. degree from the University of Sussex, Brighton, U.K., in 2013, and the Ph.D. degree from the Beijing Institute of Technology, Beijing, China, in 2018.

From 2018 to 2020, he was a Postdoctoral Researcher with the School of Electronic Engineering and Computer Sciences, Peking University, Beijing, China. His research interests include remote sensing target detection and recognition.

Lianlin Li (Senior Member, IEEE) was born in Shan'xi, China, in 1980. He received the Ph.D. degree in electrical engineering from the Institute of Electronics, Chinese Academy of Sciences, Beijing, China, in 2006.

In July 2012, he joined the School of Electronic Engineering and Computer Sciences, Peking University, Beijing, China, as a Hundred Talented Program Professor. He has authored more than 80 peer-reviewed journal papers in *Nature Communications*, *Advanced Science*, and a series of the IEEE.

Dr. Li serves as a Guest Editor of the special issue on radar imaging and detection for concealed target in *Journal of Radar* (Chinese). He serves as a regular Reviewer for the multiple journals, such as the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, *Inverse Problem*, and *Chinese Physics Letters*. He was the recipient of the Excellent Doctoral Dissertation Award from the Chinese Academy of Sciences in 2008, the National Excellent Doctoral Dissertation Nomination Award in 2009, and the URSI Young Scientist Award in 2011 and 2014. He served as a Technical Program Committee Member and a Session Chair for URSI Atlantic Radio Science Conference in 2015. Since 2014, he has been a Senior Member of the Institute of Electronics, China.

He Chen was born in Shenyang, China in 1970. She received the Ph.D. degree in electronic engineering from the Harbin Institute of Technology, Harbin, China, in 1998.

She is currently a Professor and the Dean of the School of Information, Beijing Institute of Technology, Beijing, China. Her current research interests include system-on-chip design, remote sensing data intelligent processing, and VLSI architectures for real-time image and signal processing.