Ship detection and extraction using visual saliency and histogram of oriented gradient*

XU Fang (徐芳)[1,2] **and LIU Jing-hong** (刘晶红)[1]**

1. *Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China*
2. *The University of the Chinese Academy of Sciences, Beijing 100049, China*

(Received 13 August 2016)

A novel unsupervised ship detection and extraction method is proposed. A combination model based on visual saliency is constructed for searching the ship target regions and suppressing the false alarms. The salient target regions are extracted and marked through segmentation. Radon transform is applied to confirm the suspected ship targets with symmetry profiles. Then, a new descriptor, improved histogram of oriented gradient (HOG), is introduced to discriminate the real ships. The experimental results on real optical remote sensing images demonstrate that plenty of ships can be extracted and located successfully, and the number of ships can be accurately acquired. Furthermore, the proposed method is superior to the contrastive methods in terms of both accuracy rate and false alarm rate.

Document code: A **Article ID:** 1673-1905(2016)06-0473-5

DOI 10.1007/s11801-016-6179-y

Ship detection is an active research topic spanning several fields such as image processing, pattern recognition and computer vision[1-3]. Many approaches have been investigated to solve the problems of ship detection in optical remote sensing images. Some ones can be roughly considered as threshold segmentation and statistics methods. For instance, Corbane et al[4] proposed a method selecting ship candidates by morphology filtering, and removing false alarms by wavelet analysis and Radon transform. Xu et al[5] achieved multiscale contour extraction using level set. Yang et al[6] employed a linear function which combines pixel and region characteristics to select ship candidates after sea surface analysis. Proia et al[7] estimated Gaussian distribution of the sea background density function and applied Bayesian decision theory to discriminate small ships. These methods are sensitive to the complicated sea background including clouds, uneven illumination, sea clutters, and small islands. There are some detection methods using ship wakes[8,9]. But the variations of ship wake which relates to navigation speed are relatively large. The other ones are supervised classification-based methods. The great attention has been paid to the different kinds of features as well as various classifiers. Zhu et al[10] used the support vector machine (SVM) classifier based on shape and texture features to eliminate the false alarms. Kuma et al[11] introduced a classification algorithm using color and texture for ship detection. Tang et al[12] adopted compressed domain for fast ship target extraction combining with deep neural network (DNN) and extreme learning machine. Each of them has an improvement in either preprocessing or classification. They can achieve better performance. However, these methods need to make a large number of templates and rely on prior knowledge. Besides these, there are some detection methods which can resist interferences and detect objects with complex appearance effectively in remote sensing images with complicated background[13,14]. However, they have caused significant increase in computational complexity. It is not fit for real-time process. And some small targets may not be detected.

Therefore, although numerous approaches have been developed, it is still far from solving the problems of ship detection. In view of the above-mentioned facts, a novel unsupervised ship detection approach is proposed. Two important steps are emphasized: the first is suspected target region extraction, and the second is designing some efficient rules for discriminating targets. The second aspect is crucial for improving the detection accuracy and eliminating the false alarms. However, it is simplified or even not considered in some existing methods[6,7]. The former step used in this letter is visual saliency method which can quickly access to information associated with the current scene and tasks. We construct a combination saliency model in this step. The latter step is presented to characterize the gradient symmetry of

* This work has been supported by the National Natural Science Foundation of China (No.60902067), and the Key Programs for Science and Technology Development of Jilin Province of China (No.11ZDGG001).

** E-mail: liu1577@126.com

ship sides based on a novel histogram of oriented gradient (HOG).

Firstly, the visual saliency detection method is performed to extract the suspected ship target regions. Visual saliency has been widely used to highlight valuable targets while suppressing background in target detection. In this letter, a combination saliency model based on hypercomplex frequency domain transform (HFT) and phase quanternion Fourier transform (PQFT) is constructed for searching target region. The overall continuity of the same target region and the distinguishability between different target regions are enhanced simultaneously by this process.

Based on HFT model, we improve it in color, frequency domain transform and scale. The CIE *Lab* color system is used to describe the features in place of RGB color features. Then, the value of each pixel in an image is represented as follows:

$$q(x,y)=\left(L(x,y)-\overline{L}\right)u_1+$$
$$\left(a(x,y)-\overline{a}\right)u_2+\left(b(x,y)-\overline{b}\right)u_3, \quad (1)$$

where x and y denote the pixel positions in spatial domain, and u_1, u_2, u_3 are unit pure quaternions with $u_1\perp u_2$, $u_2\perp u_3$, $u_3\perp u_1$, $u_1u_2=u_3$, $u_1u_2u_3=-1$. L, a and b are color features, while $\overline{L}$, $\overline{a}$ and $\overline{b}$ are average values of L, a and b on the whole image.

The discrete cosine transform is used instead of discrete Fourier transform. The spectrum scale-space is used for handling amplitude spectra at different scales, defined as:

$$Q[u,v]=DCT(q(x,y)), \quad (2)$$

$$A(u,v)=\left|Q[u,v]\right|, \quad (3)$$

$$G(u,v;k)=\frac{1}{\sqrt{2\pi}2^{k-2}}\mathrm{e}^{-(u^2+v^2)/(2^{2k-3})}, \quad (4)$$

$$\Lambda(u,v;k)=(G(.,.,;k)A)(u,v), \quad (5)$$

where u and v denote the pixel positions in frequency domain; $DCT(\cdot)$ denotes the discrete cosine transform. A represents the amplitude spectra. $G(\cdot)$ denotes Gaussian kernels. A collection of derived signals $\Lambda(u, v; k)$ are defined by the convolution of the amplitude spectra A with the series of Gaussian kernels. k is the scale parameter, and it is set to be k=1, 2, 3.

The better one ($S'(x, y)$) from a series of saliency maps S_k is obtained based on entropy criterion. Then, we smooth it by Gaussian filtering. The saliency map $S_1(x,y)$ based on HFT is expressed as:

$$S_1(x,y)=g\otimes\left(S'(x,y)\right)^2, \quad (6)$$

where g is a 2-D Gaussian filter, and $\otimes$ denotes convolution.

The similar procedure is applied in improving the PQFT method. The value of each pixel in an image is represented as a quaternion as follows:

$$q(x,y)=L(x,y)u_1+a(x,y)u_2+b(x,y)u_3. \quad (7)$$

The discrete cosine transform is used instead of discrete Fourier transform. The logarithm value of amplitude spectra can be calculated using Eq.(8). The saliency map $S_2(x, y)$ based on PQFT is obtained.

$$AL=\log(A)=\log(\left|Q[u,v]\right|). \quad (8)$$

The saliency map $S_1(x, y)$ and saliency map $S_2(x, y)$ are fused based on the following formula. The fusion result combines the advantages of two models. Before fusing, the saliency maps from two models are scaled to [0, 1]. The final map $S(x,y)$ is calculated as:

$$S(x,y)=z_1\times S_1(x,y)+z_2\times S_2(x,y), \quad (9)$$

where z_1 and z_2 are weights, and they are set to be z_1=0.3 and z_2=1−z_1, empirically. The results of our saliency detection method are shown in Fig.1. As we can see, the proposed method generates sharper and more uniform highlighted salient regions compared with other methods. Most of thin clouds, mist and sea clutters are eliminated. Different sizes of ships are highlighted fast and accurately.

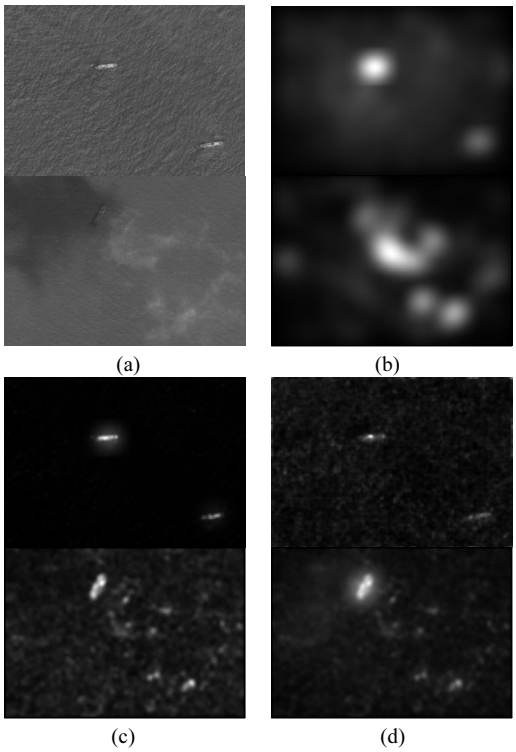

Fig.1 Visual comparison of saliency maps with different methods: (a) Original image; (b) HFT; (c) PQFT; (d) The proposed

Then, an adaptive coarse segmentation method based on OTSU is used to extract the candidate regions:

$$S(x,y)=\begin{cases}1, & S(x,y)\geq T\\ 0, & S(x,y)<T\end{cases}. \tag{10}$$

$S(x,y)$ is binarized by setting any pixel larger than T to 1, otherwise the rest of the pixels to 0. According to the binary map, we define the regions covered by the bounding rectangle of each connected area as candidate regions. Many candidate chips can be obtained in this step. But it is still possible to detect some heavy clouds, islands or coast. Candidate chips may contain real and false targets. So, the discrimination is necessary. The next work is applied to solve this problem. In order to ensure that the target has a good integrity, we extend the range of coordinate of each candidate chip to 10 pixels.

Different from the segmentation method in Ref.[15], an effective method based on GrabCut is proposed to segment the target chips accurately in this letter. This process can be done automatically. This method is based on the texture and color information to extract the target from the candidate chips precisely. In order to overcome the shortcomings of manual operation in previous algorithm, these candidate chips are disposed directly instead of segmenting the whole image. The ranges of (4, col-4) and (4, row-4) in horizontal and vertical directions of the chip are the foreground. The rest is the background. In the candidate target chip, only a single target and a small portion of the surrounding area are contained. Besides, the targets are relatively small. In general, the number of iterations is set as 2. If the sea background is complex or the target chip is large, it can be set as 3.

Given a target chip after fine segmentation, it should be symmetric by rotating and aligning the ship axis to the vertical direction. Radon transform is used to perform this task. The illustration is shown in Fig.2 in detail.

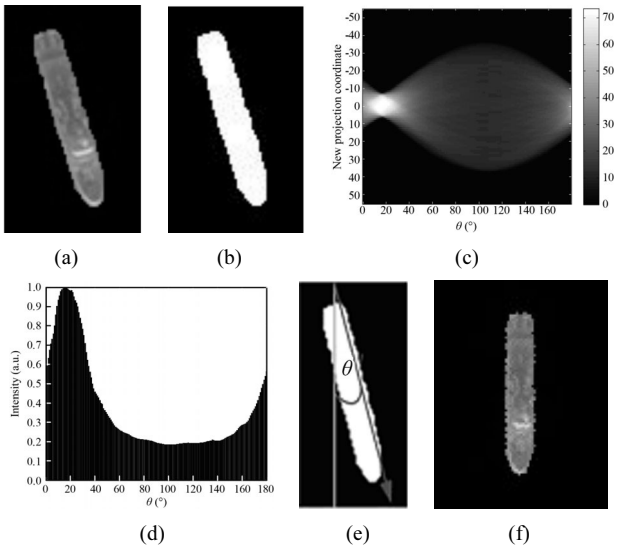

Fig.2 The process of Radon transformation: (a) Gray image; (b) Binary image; (c) Radon transform (*θ* presents the rotation angle ranging from 0 ° to 180 °.); (d) The angle with the highest Radon correlation; (e) Rotating; (f) Regular chip

An effective descriptor identifying the real ship is critical for the final discrimination. In general, a robust ship descriptor should meet the requirement that it is applicable to ships with different sizes firstly. Inspired by the facts that a ship always has a long bar-type shape, and the gradients of the two ship sides are symmetrical, a novel descriptor based on HOG is applied in this letter. The traditional HOG descriptor identifies an object by the gradients from its multi-parts. But it is sensitive to the orientation small targets. To overcome this, we align the target axis to the vertical direction by Radon transform. And the rotation invariance is fulfilled. The gradient orientations are divided into eight specific bins, h1—h8 as shown in Fig.3(a). The angle in each bin is 45°. To identify the target more accurately, we divide the body of ship along the axis direction into three blocks, B1, B2 and B3. They are one whole and two halves as shown in Fig.3(b). Theoretically, magnitudes in bins h1 and h5 acquire higher statistical quantities relative to others in all blocks. This can be explained by that the ship target has a pair of parallel long sides, which leads to strong and symmetric responses on the gradient. So, the descriptor presented has a good robustness in identifying ships. Moreover, the proposed descriptor is very effective for different sizes of ships.

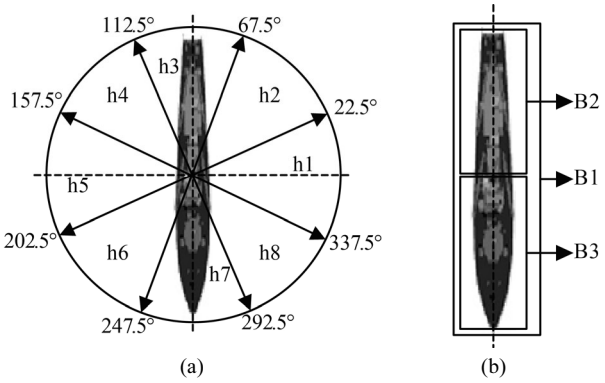

Fig.3 Illustration for bins and blocks: (a) Intervals of eight bins; (b) Three blocks

We summarize three rules for HOG in characterizing ship targets as follows: 1) Magnitudes in h1 and h5 should be bigger than those of others; 2) Magnitudes in h1 and h5 should be matchable; 3) The three blocks should satisfy the two rules above simultaneously.

However, the image is often disturbed by clouds, clutters, islands and illumination variances in fact. So, ships in real images might not strictly comply with these rules. Regarding the degradation, the relaxation parameters α_1, α_2 and γ are introduced to relax the constraints. Let H={hi, i=1, 2, 3, … , 8}, H_f={h1, h5}, H_p={h2, h3, h4, h6, h7, h8}, where $\overline{H_f}$ is the average of H_f, and $\overline{H_p}$ is the average of H_p. The following conditions should be satisfied to decide the suspected target is a real ship: $\frac{\overline{H_p}}{\min(H_f)}<\alpha_1, \frac{\max(H_p)}{\max(H_f)}<\alpha_2, \frac{\min(H_f)}{\max(H_f)}>\gamma$. Empirically,

the parameters are set as follows: α_1=0.6, α_2=0.7, γ=0.65.

To give an objective performance evaluation of the algorithms, we test them on 80 panchromatic satellite images with size of 300×210 from Google Earth under different weather conditions. These images are evenly divided into two groups: one group with simple sea surface and the other with complicated sea surface.

In the first step, the proposed saliency detection method is evaluated according to the values of *Recall*, *Precision* and *F-Measure*. The PQFT and HFT methods are used to be compared with the proposed saliency detection method. Given a ground-truth saliency map G and the detected saliency map S for an image, we have the following formulas. The corresponding curves are shown in Fig.4. As we can see, the proposed method has a better effect in generating sharper and more uniform highlighted salient regions compared with other methods. It is also more effective for eliminating background interferences and detecting ship targets with different sizes.

$$Recall = \frac{\sum_x g_x \times s_x}{\sum_x g_x}, \tag{11}$$

$$Precision = \frac{\sum_x g_x \times s_x}{\sum_x s_x}, \tag{12}$$

$$F\text{-}Measure = \frac{2Precision \times Recall}{Precision + Recall}. \tag{13}$$

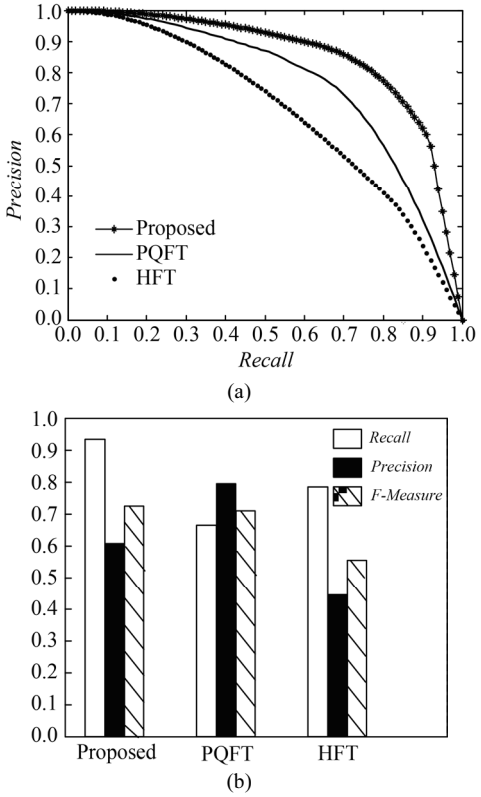

Fig.4 Performance comparison of different saliency models: (a) RP curves; (b) *F-Measure*

In the second step, we evaluate the total detection performance of the proposed method with discrimination using accuracy rate (*Cr*) and false alarm rate (*Far*). They are defined as follows.

$$Cr = \frac{N_{tt}}{N_t}, \tag{14}$$

$$Far = \frac{N_{fa}}{N_{tt} + N_{fa}}, \tag{15}$$

where N_t is the total number of real ships, N_{tt} is the number of real detected ships, and N_{fa} is the number of false alarms. The detection results are shown in Tab.1. Two unsupervised methods proposed in Ref.[4] and Ref.[6] are used for comparison with the proposed method. In addition, the proposed method without the step 2 and that with the step 2 are also compared to test the performance of the HOG descriptor.

Tab.1 Detection results of different methods

Image group	Methods	N_t	N_{tt}	N_{fa}	*Cr* (%)	*Far* (%)
Simple sea surface	Ref.[4]	93	75	14	80.645	15.730
	Ref.[6]	93	79	10	84.946	11.236
	The proposed method without step 2	93	89	7	95.699	7.292
	The proposed method with step 2	93	89	2	95.699	2.198
Complex sea surface	Ref.[4]	116	85	36	73.276	29.752
	Ref.[6]	116	93	30	80.172	24.390
	The proposed method without step 2	116	106	24	90.517	18.461
	The proposed method with step 2	116	103	11	88.793	9.649

The results in Tab.1 indicate that the proposed method with the step 2 performs better for different sea groups due to the saliency detection and the novel HOG descriptor. The method in Ref.[4] is sensitive to the noises when target candidates are extracted and its detection results are relatively poor. The method in Ref.[6] can overcome the cluttered background by utilizing length-width ratio and compactness. But it is less successful when detecting small target and clouds cover. Compared with the proposed method without the step 2, we note that the *Far* after the discrimination has been reduced largely. The *Cr* decreases a little because the ship target may be deformed on the complicated sea surface. Some detection results output by our method are shown in Fig.5 respectively. Some heavy clouds, islands and coast are eliminated and different sizes of ship targets can be detected successfully.

Furthermore, the relaxation parameters α_1, α_2 and γ are crucial in ship discrimination. We briefly test their effects by varying one parameter while fixing others as empirical values. The relationship curves are shown in Fig.6. Based on these results, they are set as follows after testing plenty of images: α_1=0.6, α_2=0.7, γ=0.65.

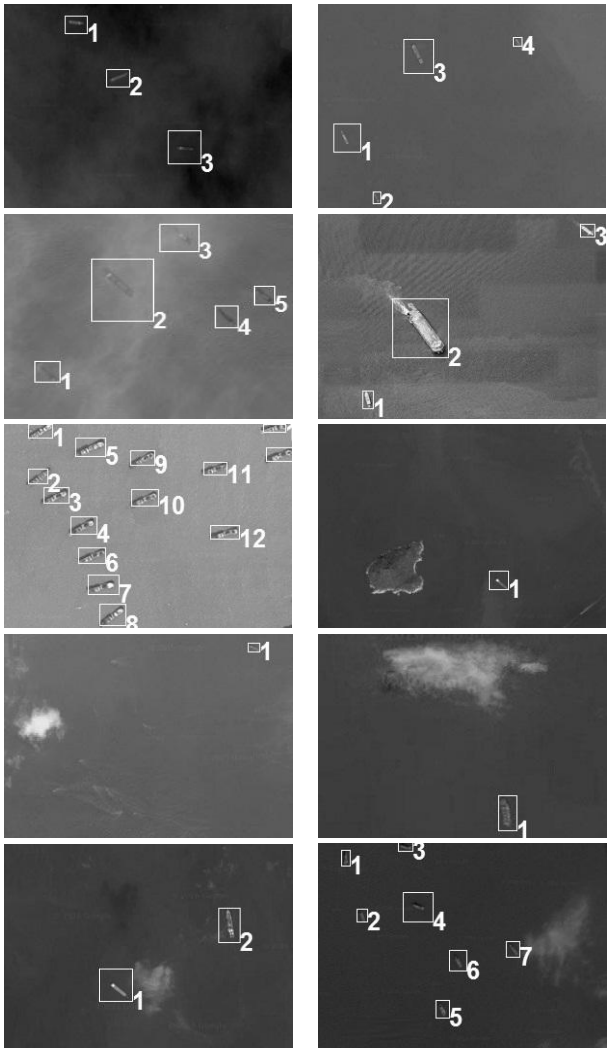

Fig.5 Some detection results of the proposed method

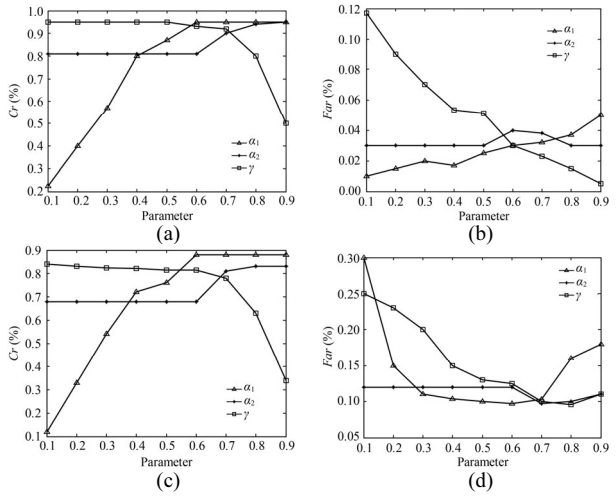

Fig.6 *Cr* and *Far* curves for different relaxation parameters: (a) and (b) simple sea surface; (c) and (d) complicated sea surface

With regard to the time consumption, our method can basically achieve the near-real-time. The proposed saliency detection method consumes a total of 1.693 s to process an image in the first step. In the second step, taking an example of ship target chip with size of 55*92, the average running time of coarse segmentation, fine segmentation, Radon transform and HOG is 0.097 s, 0.025 s, 0.217 s and 0.166 s, respectively.

In this letter, a novel unsupervised method is proposed to detect and extract ships from optical remote sensing images. In order to generate sharper and more uniform highlighted salient regions, a saliency model is constructed which combines the merits of HFT and PQFT. Then, some regular target chips are extracted after fine segmentation and Radon transform. To characterize ship targets and discriminate them more accurately, a novel descriptor based on HOG is introduced. Due to the contribution, the false alarms in the candidate chips are greatly suppressed and most real ship targets are well preserved. The experiments on real satellite images confirm that the proposed method not only outperforms the state-of-the-art methods, but also is robust for scenes with clouds, islands, sea clutters, and different sizes of ships.

References

[1] Z W Shi, X R Yu, Z G Jiang and B Li, IEEE Transactions on Geosciences and Remote Sensing **52**, 4511 (2014).

[2] Li Zheng-zhou, Chen Jing, Shen Mei-rong, Hou Qian, Ding Hao and Jin Gang, Journal of Optoelectronics·Laser **25**, 588 (2014). (in Chinese)

[3] J Wang, Y G Zhang, Z K Pan, W Z Zhang and G D Wang, Optoelectronics Letters **11**, 0307 (2015).

[4] C Corbane, L Najman, E Pecoul, L Demagistri and M Petit, International Journal of Remote Sensing **31**, 5837 (2010).

[5] Q Z Xu, B Li, Z F He and C Ma, IEEE Geoscience and Remote Sensing Letters **8**, 854 (2011).

[6] G Yang, B Li, S F Ji, F Gao and Q Z Xu, IEEE Geoscience and Remote Sensing Letters **11**, 641 (2014).

[7] N Proia and V Pagé, IEEE Geoscience and Remote Sensing Letters **7**, 226 (2010).

[8] S Shiotani and N Fujitomi, Journal of Maritime Sciences Technology **14**, 443 (2009).

[9] S Jeong, S W Ban, S Choi, D H Lee and M Lee, IEEE Journal of Oceanic Engineering **37**, 456 (2012).

[10] C R Zhu, H Zhou, R S Wang and J Guo, IEEE Transactions on Geosciences and Remote Sensing **48**, 3446 (2010).

[11] S S Kumar and M U Selvi, Journal of Computer Applications in Engineering Sciences **1**, 59 (2011).

[12] J X Tang, C W Deng, G B Huang and B J Zhao, IEEE Transactions on Geosciences and Remote Sensing **53**, 1174 (2015).

[13] N Yokoya and A Iwasaki, Journal of Selected Topics in Applied Observations and Remote Sensing **8**, 2053 (2015).

[14] X Wang, S Q Shen, C Ning, F C Huang and H M Gao, Applied Optics **55**, 1381 (2016).

[15] S X Qi, J Ma, J Lin, Y S Li and J W Tian, IEEE Geoscience and Remote Sensing Letters **12**, 1451 (2015).