IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium | 978-1-7281-6374-1/20/$31.00 ©2020 IEEE | DOI: 10.1109/IGARSS39084.2020.9323433

ADAPTIVE SUPERPIXEL SEGMENTATION WITH FISHER VECTORS FOR SHIP DETECTION IN SAR IMAGES

Xueqian Wang[1], *Gang Li*[1*], *and Antonio Plaza*[2]

[1] Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
[2] Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, 10003 Cáceres, Spain
[*]Corresponding author: Gang Li. Email: gangli@tsinghua.edu.cn.

ABSTRACT

In this paper, we propose an improved superpixel segmentation algorithm for ship target detection in synthetic aperture radar (SAR) images, called adaptive Fisher vector-based simple linear iterative clustering (AFVSLIC). Compared with existing algorithms, three new features produced by Fisher vectors, i.e., zero-order, first-order and second-order features, are exploited by the proposed AFVSLIC algorithm to enhance segmentation performance. Besides, AFVSLIC adaptively adjusts the weights of the features to maintain the segmentation performance in different signal-to-clutter ratio (SCR) scenarios. Experimental results demonstrate that the proposed AFVSLIC algorithm outperforms existing, commonly used algorithms for superpixel segmentation and (accordingly) improves the performance of ship target detection.

Index Terms—Adaptive superpixel segmentation, fisher vectors, ship detection, synthetic aperture radar (SAR)

1. INTRODUCTION

Superpixel segmentation aims at producing locally coherent regions in a processed image. Recently, simple linear iterative clustering (SLIC) and its variants [1], [5], [7] have been proposed for superpixel segmentation, including the problem of ship detection in synthetic aperture radar (SAR) images. Only simple intensity and spatial features of pixels are exploited in superpixel segmentation algorithms based on SLIC [1], [5], [7], which may lead to inaccurate segmentation especially in low signal-to-clutter ratio (SCR) cases and degrade the subsequent detection performance [2]. Moreover, the weights of the features need to be manually adjusted in [1], [5], [7] to achieve superpixel segmentation in different SCR cases. Ref. [3] has proposed an algorithm to achieve adaptive superpixel segmentation of urban/land SAR images. In [3], the features of pixels and their corresponding weights are selected based on the image gradient, which is helpful to extract edge information in urban/land SAR images. However, strong and heterogeneous sea clutter background in marine SAR images makes it difficult to characterize the boundaries of ship targets by using the image gradient.

The Fisher vector (FV) of an image pixel is computed according to the gradient of the log-likelihood function with respect to the global Gaussian mixture model (GMM) parameters [6]. FV contains multi-order information (i.e., zero-order, first-order and second-order information) of a pixel. In [1], [2], it has been demonstrated that the multi-order features in FV can effectively discriminate ship targets and the sea clutter, even in low SCR cases. Note that FV is only utilized at the step of decision-making in existing superpixel-based detectors in [1], [2]. To the best of our knowledge, multi-order features in FVs have not been exploited for superpixel segmentation of SAR images thus far.

In this paper, we introduce an adaptive FV-based simple linear iterative clustering (AFVSLIC) algorithm for superpixel segmentation of SAR images. AFVSLIC considers not only the intensity and spatial features, but also the multi-order features produced by FVs, where the weights of the features are adaptively adjusted in the iterative process. Mainly, there are three steps in the iteration of AFVSLIC: 1) locally clustering pixels to generate supeprixels based on the pixel distance with aggregated features; 2) updating each superpixel center by averaging the features of its pixels; 3) adjusting the weights of the features to obtain more compact superpixels. Experiments with real SAR data show that the proposed AFVSLIC algorithm achieves better segmentation performance than the commonly used algorithms, improving the performance of ship target detection.

2. DISTANCE MEASURE

Superpixel segmentation is regarding as a local clustering process [4]. A key problem for clustering is how to define the distance between two pixels in the image. In this section, we consider five distance measures based on different aspects, such as intensity, spatial location, zero-order, first-order and second-order characteristics produced by FVs, respectively.

2.1. Distance Measures of Intensity and Spatiality

The intensity distance $d_{\mathrm{int}}(i,j)$ and the spatial distance $d_{\mathrm{spa}}(i,j)$ between the pixel i and the pixel j are respectively calculated as:

$$d_{\mathrm{int}}(i,j) = |\delta_i - \delta_j|, \tag{1}$$

$$d_{\mathrm{spa}}(i,j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}, \tag{2}$$

where (x_i, y_i) and δ_i represent the spatial coordinates and the intensity value of the pixel i, respectively, i, $j \in \{1,2,\ldots,P\}$, and P is the number of pixels in the whole SAR image.

2.2. Distance Measures of FVs

The basic idea of FV lies in: 1) establishing a GMM to approximate the global SAR image, and 2) calculating the gradients of the log-likelihood function with respect to the model parameters in GMM as the coding vectors of pixels. It is well known that GMM can model continuous distributions with arbitrary precision [10]. The probability density function (PDF) of GMM for the SAR image is expressed as:

$$f(\delta) = \sum_{k=1}^{K} \omega_k \hat{f}_k(\delta), \tag{3}$$

where $\hat{f}_k(\delta) = 1/\sqrt{2\pi\sigma_k^2} \exp\left[-\frac{1}{2\sigma_k^2}(\delta - \mu_k)^2\right]$, ω_k, μ_k, σ_k represent the weight, mean value and standard deviation of the k-th Gaussian component, respectively, k =1,2,…,K, $\sum_{k=1}^{K} \omega_k = 1$ and K is the number of Gaussian components in GMM. The GMM parameter set $\boldsymbol{\Theta} \triangleq \{\omega_k, \mu_k, \sigma_k, k = 1,2,\ldots, K\}$ can be estimated by using the expectation-maximization (EM) algorithm [11].

The FV $\boldsymbol{\alpha}_i$ of the i-th pixel is the normalized gradient vector of the log-likelihood function with respect to the parameter set $\boldsymbol{\Theta}$ [6]:

$$\boldsymbol{\alpha}_i = [\underbrace{\alpha_{i,\omega_1}, \ldots, \alpha_{i,\omega_K}}_{\left(\boldsymbol{\alpha}_i^{(0)}\right)^T}, \underbrace{\alpha_{i,\mu_1}, \ldots, \alpha_{i,\mu_K}}_{\left(\boldsymbol{\alpha}_i^{(1)}\right)^T}, \underbrace{\alpha_{i,\sigma_1}, \ldots, \alpha_{i,\sigma_K}}_{\left(\boldsymbol{\alpha}_i^{(2)}\right)^T}]^T, \tag{4}$$

where

$$\alpha_{i,\omega_k} = \left(\phi_{i,k} - \beta_k\right)/\sqrt{\beta_k}, \tag{5}$$

$$\alpha_{i,\mu_k} = \phi_{i,k}(\delta_i - \mu_k)/\left(\sigma_k\sqrt{\beta_k}\right), \tag{6}$$

$$\alpha_{i,\sigma_k} = \left(\phi_{i,k}/\sqrt{2\beta_k}\right)[(\delta_i - \mu_k)^2/\sigma_k^2 - 1], \tag{7}$$

i =1,2,…, P, k =1,2,…, K, $(.)^T$ represents the transpose operation, $\phi_{i,k} = \omega_k \hat{f}_k(\delta_i)/\sum_{k=1}^{K} \omega_k \hat{f}_k(\delta_i)$, and $\beta_k = \exp(\omega_k)/\sum_{k=1}^{K} \exp(\omega_k)$. We refer readers to [6] for the detailed derivations of (5)-(7). In (4), we use $\{\boldsymbol{\alpha}_i^{(\eta)}, \eta = 0,1,2\}$ to denote the sub-vectors in FV corresponding to the weights, mean values and standard deviations in GMM, respectively, for i=1,2,…,P. Then, the signed square-rooting operation is performed on $\{\boldsymbol{\alpha}_i^{(\eta)}, \eta = 0,1,2\}$, $\forall i$, like in [6] to avoid FV close to null.

Note that $\{\boldsymbol{\alpha}_i^{(\eta)}, \eta = 0,1,2\}$ in (4) represent the zero-order, the first-order and the second-order information (see (5)-(7)) of the i-th pixel, respectively. In [1], [2], the multi-order features in FV have shown the good discrimination ability between ship targets and the clutter background in SAR images, while they have not been exploited for the problem of superpixel segmentation.

Next, the multi-order features produced by FVs are introduced to measure the distance of two pixels in the image. Based on (4)-(7), the distances between FVs of pixel i and pixel j are defined as follows:

Algorithm 1: AFVSLIC

***Input*:** The SAR image with P pixels, the superpixel size S, the amplification factor λ, the number of components in GMM K, and the maximum iteration count *Iter*.

***Initialization*:** The expected number of superpixels is $L = P/S^2$. Estimate the GMM parameters via the EM algorithm and calculate the FV for each pixel in the SAR image by (4)-(7). Set the center coordinate $c_{(l)} = [x_{(l)}, y_{(l)}]$ of the l-th superpixel with the regular grid interval S, for l = 1,2,…,L. Initialize distance $\mathbf{D} = [D_1, D_2, \ldots, D_P]$, $D_p = +\infty$, p = 1,2,…,P, label $\boldsymbol{\Omega} = \mathbf{0}, \boldsymbol{\Omega} \in \mathbb{R}^P$, and the weights h_r = 1/R, for r = 1,2,...,R.

***Repeat*:**

Step 1: For each center coordinate $c_{(l)}$, l = 1,2,…,L,

For each pixel (q = 1,2,…,$4S^2$) in the area 2S × 2S centered by $c_{(l)}$,

1) Calculate the distance d_{all} between the center $c_{(l)}$ and the pixel q via (9).

2) If $d_{\text{all}} < D_q$, then $D_q = d_{\text{all}}$, $\boldsymbol{\Omega}_q = l$.

End

End

Step 2: Update each superpixel center $\{c_{(l)}, \forall l\}$ by averaging the intensities, spatial locations and FVs of its pixels.

Step 3: Adjust the weights $\{h_r, \forall r\}$ of the features via (11).

Until the count of iterations is larger than *Iter*.

***Postprocessing*:** If there is a disjoint superpixel denoted by C in the segmentation result $\boldsymbol{\Omega}$, 1) find its neighboring supeprixels and calculate their averaged intensity values $\{\bar{\delta}_1, \bar{\delta}_2, \ldots, \bar{\delta}_{N_C}\}$, where N_C is the number of neighboring supeprixels; 2) assign the label of the n^*-th supepixel to C, where $n^* = \arg\min_{n \in \{1,2,\ldots,N_C\}} |\bar{\delta}_n - \bar{\delta}_C|$, $\bar{\delta}_C$ is the averaged intensity value of the disjoint superpixel C.

***Output*:** The superpixel segmentation result $\boldsymbol{\Omega}$.

$$d_{\text{FV},\eta} = \left\| \sum_{t=1}^{9} G(t) \left(\boldsymbol{\alpha}_{i_t}^{(\eta)} - \boldsymbol{\alpha}_{j_t}^{(\eta)} \right) \right\|_2, \tag{8}$$

where $\eta \in \{0,1,2\}$, $G(.)$ is the standard Gaussian kernel, $t \in \{1,2,\ldots,9\}$, i_t represents the t-th element in the 3×3 patch centered by pixel i, and i, $j \in \{1,2,\ldots,P\}$. Here, the Gaussian kernel $G(.)$ is used to suppress the speckle noise in SAR images [5].

3. PROPOSED AFVSLIC ALGORITHM

The proposed AFVSLIC algorithm aggregates not only the distances of the (commonly used) intensity and spatial features, but also the distances of the multi-order features in FVs. The weights of features are adaptively selected in the iteration of AFVSLIC to enhance the compactness of superpixels.

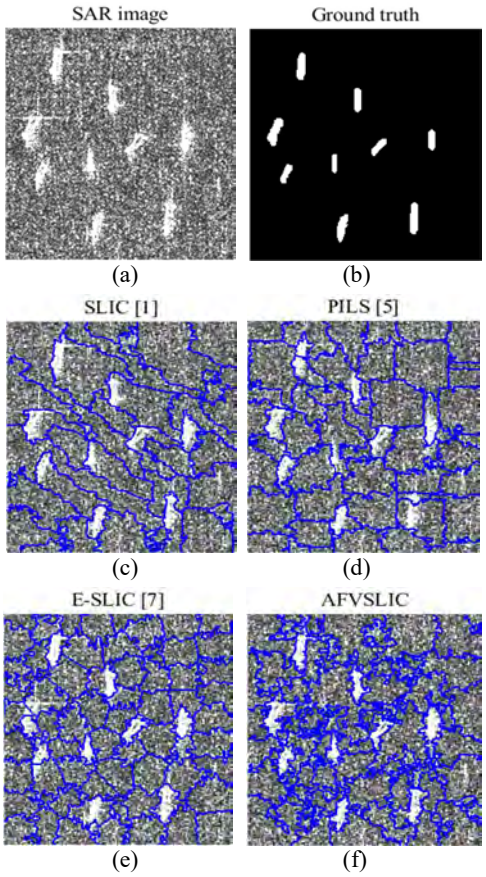


Fig. 1 Results of superpixel segmentation. (a) The original SAR image (SCR = -1dB), (b) the ground truth, (c) SLIC [1], (d) PILS [5], (e) E-SLIC [7], (f) proposed AFVSLIC algorithm.

The iterative process of the proposed AFVSLIC algorithm mainly contains three steps:

1) **Step 1**: Set the centers of superpixels and the weights of features, then solve the pixel labels. The search area for each superpixel center is its neighboring $2S\times 2S$ region, where S is the predefined size of the superpixel. The distance between the l-th superpixel center $c_{(l)}$ and the q-th pixel in its $2S\times 2S$ search area is defined by aggregating the five feature distances introduced in Section 2:

$$d_{\text{all}} = \sqrt{\sum_{r=1}^{R} h_r^{\lambda} d_r^2(c_{(l)}, q)}, \quad (9)$$

where $R = 5$ is the number of features, $r = 1,2,\ldots,R$ represent 'int', 'spa', 'FV, 0', 'FV, 1' and 'FV, 2', respectively (see (1), (2) and (8)), h_r is the weight parameter of the r-th feature, $\sum_{r=1}^{R} h_r = 1$, $\lambda > 1$ represents the amplification factor of the weights [8], $q = 1, 2,\ldots,4S^2$, $l = 1, 2,\ldots, L$, and $L = P/S^2$ is the expected number of superpixels in the SAR image. Based on the distance measure in (9), each pixel is assigned to its closest superpixel center. Note that the values in $\{d_r^2(c_{(l)}, q), \forall r\}$ are pre-normalized before (9) by using their corresponding maximum values in the $2S\times 2S$ search area of the l-th superpixel center to guarantee that they have the same dynamic range [0, 1], $\forall l$.

2) **Step 2**: Obtain the superpixel centers with the fixed pixel labels and the weights of features. In detail, each superpixel center is updated by averaging the intensity values, spatial coordinates and FVs of the pixels corresponding to it.

3) **Step 3**: Update the weights of features with the fixed pixel labels and superpixel centers. First, the sum of the within-superpixel variances (SSV) [10] is defined as:

$$\text{SSV} = \sum_{l=1}^{L}\sum_{p=1}^{P}\sum_{r=1}^{R} \text{sign}(p \in \mathbf{s}_l) h_r^{\lambda} d_r^2(c_{(l)}, p), \quad (10)$$

where $\mathbf{s}_l$ denotes the l-th superpixel with the center $c_{(l)}$, for $l = 1,2,\ldots,L$, sign(.) is 1 if its input is true and 0 otherwise. Then, the optimal weights $\{h_r, \forall r\}$ are obtained by minimizing SSV in (10), i.e., the weights $\{h_r, \forall r\}$ are adjusted to make each superpixel more compact. By setting the first derivative of SSV with respect to $\{h_r, \forall r\}$ into zero, we have:

$$\frac{\partial \text{SSV}}{\partial h_r} = 0 \Rightarrow h_r = 1/\sum_{\acute{r}=1}^{R} [\text{sSSV}_r/\text{sSSV}_{\acute{r}}]^{1/(\lambda-1)}, \forall r, \quad (11)$$

where $\text{sSSV}_r = \sum_{l=1}^{L}\sum_{p=1}^{P} \text{sign}(p \in \mathbf{s}_l) d_r^2(c_{(l)}, p)$.

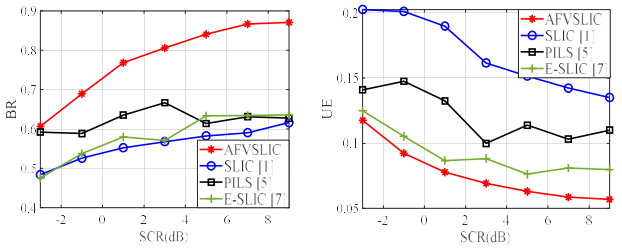
When the number of iterations in AFVSLIC exceeds a predefined maximum iteration count *Iter*, postprocessing is performed to eliminate the disjoint superpixels as [1], [5], [7]. The implementation of the proposed AFVSLIC algorithm is summarized in Algorithm 1.

4. EXPERIMENTAL RESULTS

In this section, we compare the performance of the SLIC [1], pixel intensity and the location similarity (PILS) [5], exponential-SLIC (E-SLIC) [7] and the proposed AFVSLIC algorithm by using SAR images collected by the Gaofen-3 satellite. For these segmentation algorithms, the predefined superpixel size is $S = 26$ and the maximum iteration count is *Iter* = 10, which are selected according to [1] and [4]. The other hyper-parameters in SLIC, PILS and E-SLIC are well-tuned in the case of SCR = 15dB. The amplification factor λ in AFVSLIC is 7. The number of Gaussian components in GMM is $K = 7$, which is sufficient to model the SAR images [2]. To evaluate the robustness of above superpixel segmentation algorithms in different SCR cases, semi-controlled SAR images are generated according to the guideline in [9]. Let $\mathbf{U}$ and $\mathbf{V}$ denote the original SAR image containing ship targets and the pure sea surface SAR image, respectively. Given the value of SCR in dB, the semi-controlled SAR image $\mathbf{U}'$ is obtained by $\mathbf{U}' = \mathbf{U} + \rho\mathbf{V}$, where $\rho = \sqrt{10^{-\text{SCR}/10}\frac{\sum_{i\in\mathbf{U}}|\delta_i|^2}{\sum_{j\in\mathbf{V}}|\delta_j|^2}}$ [9].

4.1. Segmentation Performance

Fig. 1 shows the segmentation performances of SLIC [1], PILS [5], E-SLIC [7] and the proposed AFVSLIC algorithm in the case of SCR = -1dB. From Fig. 1, it can be observed that AFVSLIC provides more distinct boundaries to distinguish ship targets and sea clutter. This is because 1) the newly introduced FV features in AFVSLIC enhance the discrimination between ship targets and the sea clutter, and 2) the weights of the features are adaptively adjusted to improve



(a) (b)

Fig. 2 Segmentation performance: (a) BR, (b) UE.

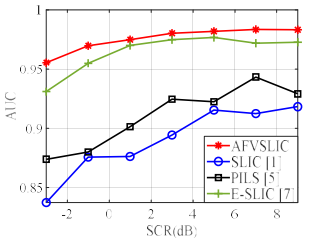


Fig. 3 Detection performance: AUC of SP-CFAR [7] versus SCR.

the compactness of superpixels.

In Fig. 2, we quantitatively evaluate the segmentation performance by two commonly used metrics, i.e., boundary recall (BR) and undersegmentation error (UE) [4]. BR and UE are defined as

$$\mathrm{BR} = \frac{\sum_{b\in\mathbf{B}(\boldsymbol{\theta})} \mathrm{sign}\left[\left(\min_{a\in\mathbf{B}(\mathbf{s})} \sqrt{(x_a-x_b)^2+(y_a-y_b)^2}\right)\le\epsilon\right]}{|\mathbf{B}^{(\boldsymbol{\theta})}|}, \quad (12)$$

$$\mathrm{UE} = \frac{\sum_{z=1}^{Z}\sum_{l=1}^{L}|\mathbf{s}_l|\times\mathrm{sign}(|\mathbf{s}_l\cap\boldsymbol{\theta}_z|>\xi|\mathbf{s}_l|)}{\sum_{z=1}^{Z}|\boldsymbol{\theta}_z|} - 1, \quad (13)$$

respectively, where $l = 1,2,\ldots,L$, $\mathbf{B}^{(\mathbf{s})}$ represents the set of all the boundary pixels in the segmentation result, $\boldsymbol{\theta}_z$ is the z-th ground truth of segment, for $z = 1,2,\ldots,Z$, $\mathbf{B}^{(\boldsymbol{\theta})}$ is the ground truth set of boundary pixels, ϵ is the bound parameter, and ξ is the threshold parameter. In this paper, we set ϵ =2 and ξ =0.05. A larger value of BR means that the generated superpixels have better consistency with the ground truth of boundaries. A smaller value of UE implies that the generated superpixels show less overlapping of targets and clutter. From Fig. 2, we can conclude that AFVSLIC provides larger BR and smaller UE, i.e., better segmentation performance is achieved by AFVSLIC.

4.2. Detection Performance

Superpixel-based constant false alarm rate (SP-CFAR) [7] is a commonly used superpixel-based detector. Here, SLIC [1], PILS [5], E-SLIC [7] and AFVSLIC are considered for the superpixel segmentation step of SP-CFAR, respectively. The detection performance is evaluated by the area under the receiver operating characteristic curve (AUC) [9]. In Fig. 3, we show the AUC of the SP-CFAR detector. We can see that the proposed AFVSLIC algorithm leads to better detection performance, due to its more accurate superpixel segmentation.

5. CONCLUSION

In this paper, we introduce a new superpixel segmentation algorithm called AFVSLIC. Compared with existing algorithms, the multi-order characteristics contained in FVs are utilized in AFVSLIC to measure the distance between pixels in SAR images. In addition, AFVSLIC selects adaptively the weights of the features. Our experiments demonstrate that AFVSLIC provides better segmentation performance, leading to better performance in ship target detection than existing state-of-the-art algorithms.

6. ACKNOWLEDGEMENT

This work was supported by National Natural Science Foundation of China under Grants 61790551 and 61925106, and in part by the National Key R&D Program of China 2018YFC0825800. This work was also funded by FEDER and Junta de Extremadura (GR18060).

7. REFERENCES

[1] H. Lin, H. Chen, K. Jin, L. Zeng and J. Yang, "Ship Detection With Superpixel-Level Fisher Vector in High-Resolution SAR Images," *IEEE Geosci. Remote Sens. Lett.*, 2019.

[2] X. Wang, G. Li, X.-P. Zhang and Y. He, "Ship Detection in SAR Images via Local Contrast of Fisher Vectors" Submitted to *IEEE Trans. Geosci. Remote Sens.*, Major Revision, Nov. 2019.

[3] D. Xiang, T. Tang, S. Quan, D. Guan and Y. Su, "Adaptive Superpixel Generation for SAR Images With Linear Feature Clustering and Edge Constraint," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3873-3889, June 2019.

[4] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua and S. Süsstrunk, "SLIC Superpixels Compared to State-of-the-Art Superpixel Methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274-2282, Nov. 2012.

[5] D. Xiang, T. Tang, L. Zhao and Y. Su, "Superpixel Generating Algorithm Based on Pixel Intensity and Location Similarity for SAR Image Classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 6, pp. 1414-1418, Nov. 2013.

[6] J. Sánchez, F. Perronnin, T. Mensink and J. Verbeek, "Image classification with the fisher vector: theory and practice," *Int. J. Comput. Vis.*, vol. 105, no. 3, pp. 222-245, 2013.

[7] W. Yu, Y. Wang, H. Liu and J. He, "Superpixel-Based CFAR Target Detection for High-Resolution SAR Images," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 5, pp. 730-734, May 2016.

[8] J. Z. Huang, M. K. Ng, H. Rong and Z. Li, "Automated variable weighting in k-means type clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 657-668, May 2005.

[9] G. Xiong, F. Wang, L. Zhu, J. Li and W. Yu, "SAR Target Detection in Complex Scene Based on 2-D Singularity Power Spectrum Analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9993-10003, Dec. 2019.

[10] T. Celik and T. Tjahjadi, "Automatic image equalization and contrast enhancement using Gaussian mixture modeling," *IEEE Trans. on Image Process.*, vol. 21, no. 1, pp. 145–156, Jan. 2012.

[11] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Process. Mag.*, vol. 13, no. 6, pp. 47-60, Nov. 1996.