SUPERPIXEL SEGMENTATION FOR POLSAR IMAGES WITH LOCAL ITERATIVE CLUSTERING AND HETEROGENEOUS STATISTICAL MODEL

D. Xiang [a, *], W. Ni [a], H. Zhang [a], J. Wu [a], W. Yan [a], Y. Su [b]

[a] Northwest Institute of Nuclear Technology, 710024 Xi'an, China - xiangdeliang@gmail.com
[b] College of Electronic Science and Engineering, National University of Defense Technology, 410073 Changsha, China

Commission VI, WG VI/4

KEY WORDS: PolSAR, Image Segmentation, Heterogeneous model, Superpixel, Local clustering

ABSTRACT:

Superpixel segmentation has an advantage that can well preserve the target shape and details. In this research, an adaptive polarimetric SLIC (Pol-ASLIC) superpixel segmentation method is proposed. First, the spherically invariant random vector (SIRV) product model is adopted to estimate the normalized covariance matrix and texture for each pixel. A new edge detector is then utilized to extract PolSAR image edges for the initialization of central seeds. In the local iterative clustering, multiple cues including polarimetric, texture, and spatial information are considered to define the similarity measure. Moreover, a polarimetric homogeneity measurement is used to automatically determine the tradeoff factor, which can vary from homogeneous areas to heterogeneous areas. Finally, the SLIC superpixel segmentation scheme is applied to the airborne Experimental SAR and PiSAR L-band PolSAR data to demonstrate the effectiveness of this proposed segmentation approach. This proposed algorithm produces compact superpixels which can well adhere to image boundaries in both natural and urban areas. The detail information in heterogeneous areas can be well preserved.

1. INTRODUCTION

Object-based segmentation and classification are promising in remote sensing field, which significantly outperform the pixel-based image processing (Niu and Ban 2013; Ban and Jacob 2013). Therefore, object generation plays a key role in this kind of processing, where the images are segmented into many homogeneous regions. A superpixel is defined as a local region which preserves most of the object information and well adheres to the object boundaries (Xiang et al. 2013). To better preserve the polarimetric and statistical characteristics of the images and also overcome the influence of speckle noise in the meantime, superpixel generation and segmentation with regular size and shape seem promising for PolSAR data. Until now, numerous superpixel segmentation algorithms have been proposed for optical images, among them, the simple linear iterative clustering (SLIC) (Achanta et al. 2012) method is popular and shows good performance in superpixel generation. In contrast, there are very few superpixel generation and segmentation approaches proposed for SAR and PolSAR images. Xiang et al. (Xiang et al. 2013) developed a novel superpixel generation algorithm based on pixel intensity and location similarity, which modified the similarity measure of SLIC to make it applicative for SAR images. For PolSAR data, Liu et al. (Liu et al. 2013) incorporated the revised Wishart distance and edge map into the Normalized cuts algorithm to produce superpixels. On the basis of SLIC, Feng et al. (Feng, Cao, and Pi 2014), Song et al. (Song et al. 2015), and Qin et al. (Fachao, Jiming, and Fengkai 2015) utilized the symmetric revised Wishart distance, Bartlett distance and revised Wishart distance respectively as the similarity measures instead of the original one to generate superpixels. It can be seen that these methods are all designed based on the assumption of Wishart

distribution, which can well describe the backscatters of natural areas. However, for heterogeneous urban areas, this assumption is usually violated (Wenjin, Huadong, and Xinwu 2015; Soergel 2010), making the superpixels not well adhere to urban boundaries and preserve the polarimetric features. In the meantime, it is quite difficult to find a particular distribution to describe the backscatters of urban areas since they are extremely complex (Wenjin, Huadong, and Xinwu 2015). Therefore, superpixel generation and segmentation for PolSAR images, especially in heterogeneous urban areas, still remains unsolved. Another drawback of these methods lies on the non-adaptive selection of trade-off factor, which balances the polarimetric similarity and spatial proximity while simultaneously provides control over the shape and compactness of superpixels. This parameter is usually set manually to a constant value by trial and error, which might not be suitable in some areas.

Recent studies show that the higher scene heterogeneity leads to non-Gaussian clutter modelling (Vasile et al. 2010). The spherically invariant random vector (SIRV) proposed by Yao (Yao 1973), is a sort of non-Gaussian processes with random variance and has been already used in PolSAR data classification (Doulgeris, Anfinsen, and Eltoft 2008) and segmentation (Bombrun et al. 2011). Inspired by SLIC and the SIRV product model, this paper proposes an adaptive superpixel segmentation algorithm (Pol-ASLIC) for PolSAR images. The key points of this approach are the definition of pixel similarity measure and the adaptive setting of trade-off factor. The edge map is also essential to the final superpixel generation. The main contributions of this paper are listed as follows: 1) We utilize a new edge detector based on the SIRV model to detect PolSAR image edges. Traditional CFAR

---

* Corresponding author

detector is designed on the basis of Wishart distribution and cannot well detect the edges in complicated scenes. In contrast, our method can get a better edge map, especially in heterogeneous urban areas. 2) We define an effective similarity measure which contains multiple cues including polarimetric, texture, and spatial information for superpixel generation in PolSAR images. Therefore, more local information can be considered. 3) We propose an adaptive trade-off factor to control the shape and compactness of superpixels. Similar to (Fengkai, Jie, and Deren 2015), the equivalent number of looks (ENL) and new edge maps are combined to produce an effective polarimetric homogeneity measurement, which is then incorporated into the trade-off factor, making it flexible in homogeneous and heterogeneous regions. Thus, the number of parameters in traditional SLIC reduces to only one, i.e., the superpixel number. Furthermore, there are very few over- or under-segmentation in the final superpixel map compared to other methods.

2. SIRV PRODUCT MODEL FOR POLSAR DATA

2.1 Parameter Estimation

In the SIRV product model, the m-dimensional complex measurement $\mathbf{k}$ is defined as (Vasile et al. 2010)

$$\mathbf{k} = \sqrt{\tau}\mathbf{z} \qquad (1)$$

where $\mathbf{z}$ is an independent complex circular Gaussian vector with zero mean and normalized covariance matrix $\mathbf{M} = \mathrm{E}\{\mathbf{z}\mathbf{z}^\dagger\}$. The superscript † denotes the conjugate transpose operator. τ is a positive random variable and its PDF is not explicitly specified. For PolSAR data, the normalized covariance matrix $\mathbf{M}$ characterizes polarimetric diversity while the random variable τ can be considered as spatial texture, which represents the randomness of spatial variations and only affects the scattering power. The PDF of $\mathbf{k}$ is denoted as (Vasile et al. 2010)

$$p(\mathbf{k} \mid \tau, \mathbf{M}) = \frac{1}{(\pi)^{mN} \mid \mathbf{M} \mid^N} \prod_{i=1}^{N} \frac{1}{\tau_i^m} \exp\left(-\frac{\mathbf{k}_i^\dagger \mathbf{M}^{-1} \mathbf{k}_i}{\tau_i}\right) \qquad (2)$$

where N represents the number of independent samples used in the estimation. For PolSAR applications, the target vector $\mathbf{k}$ can be formed using the linear bases as

$$\mathbf{k} = \left[S_{\mathrm{HH}}, \sqrt{2}S_{\mathrm{HV}}, S_{\mathrm{VV}} \right]^{\mathrm{T}} \qquad (3)$$

here $S_{\mathrm{HH}}, S_{\mathrm{VV}}$ and S_{HV} are the elements of complex scattering matrix, T is the transpose operator. Therefore, in this paper, m equals to 3. For a given $\mathbf{M}$, the texture estimator $\hat{\tau}_i$ can be obtained by maximizing the log likelihood function of (2) like (Vasile et al. 2010)

$$\hat{\tau}_i = \frac{\mathbf{k}_i^\dagger \mathbf{M}^{-1} \mathbf{k}_i}{m}. \qquad (4)$$

Replacing τ_i in (2) with (4), we can obtain the maximum likelihood estimator of normalized covariance matrix as (Vasile et al. 2010)

$$\hat{\mathbf{C}} = f\left(\hat{\mathbf{C}}\right) = \frac{m}{N} \sum_{i=1}^{N} \frac{\mathbf{k}_i \mathbf{k}_i^\dagger}{\mathbf{k}_i^\dagger \hat{\mathbf{C}}^{-1} \mathbf{k}_i} \qquad (5)$$

which can be obtained by a recursive algorithm as

$$\hat{\mathbf{C}}_{i+1} = f(\hat{\mathbf{C}}_i). \qquad (6)$$

Note that the convergence of (6) can be assured with an arbitrary initialization of $\hat{\mathbf{C}}_i$ (Vasile et al. 2010). Therefore,

this algorithm can be initialized with the identity matrix $\hat{\mathbf{C}}_0 = \mathbf{I}_m$. The span for SIRV case is given by

$$P = \mathrm{E}\{\tau(\mathbf{z}^\dagger\mathbf{z})\} = \tau\mathrm{E}\{\mathbf{z}^\dagger\mathbf{z}\} = \tau\mathrm{Tr}\{\mid\mathbf{M}\mid\} = \tau \cdot m. \qquad (7)$$

The texture estimator in (4) can then directly be linked to the total scattering power (span) according to (7). Hence, the maximum likelihood estimator of span is defined as

$$\hat{P} = \mathbf{k}^\dagger \hat{\mathbf{C}}^{-1} \mathbf{k} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{k}_i^\dagger \hat{\mathbf{C}}_i^{-1} \mathbf{k}_i. \qquad (8)$$

It can be observed that the estimated covariance matrix is independent of the total scattering power and it only contains the polarimetric information. Finally, according to (1), the conventional covariance matrix $\mathbf{C}$ can be derived as

$$\mathbf{C} = \frac{\hat{P}}{m} \hat{\mathbf{C}}. \qquad (9)$$

In terms of the number of samples N used in the estimation, existing studies have shown that the span driven adaptive neighborhood (SDAN) (Vasile et al. 2008) can achieve a good tradeoff between preserving signal characteristics and collecting a large number of samples. According to the above estimation scheme, the conventional covariance matrix $\mathbf{C}$ of a PolSAR data set is decomposed into two parameters, i.e., the normalized covariance matrix $\hat{\mathbf{C}}$ which contains polarimetric information and the span $\hat{P}$ that contains scalar texture information. In the next sections, for simplicity, $\mathbf{M}$ and P are represented by the estimations $\hat{\mathbf{C}}$ and $\hat{P}$, respectively. These two parameters will be finally utilized for PolSAR edge detection and superpixel segmentation.

2.2 Distance Measure between Covariance Matrices

Measuring pixel similarity in PolSAR images is quite essential to our proposed framework. It will be used in the following edge map calculation and superpixel segmentation. This subsection introduces the SIRV distance between two normalized covariance matrices.

For normalized covariance matrices, since they do not respect the Wishart distribution, conventional Wishart distance cannot be directly applied. Similar to (Liu et al. 2013), the likelihood ratio test respects to the texture τ and the normalized covariance matrix $\hat{\mathbf{C}}$ is also adopted to derive the corresponding SIRV distance. The new hypothesis test is

$$\begin{cases} H_0 : \mathbf{\Sigma}_{\mathbf{M}i} = \mathbf{\Sigma}_{\mathbf{M}j} \\ H_1 : \mathbf{\Sigma}_{\mathbf{M}i} \neq \mathbf{\Sigma}_{\mathbf{M}j} \end{cases} \qquad (10)$$

where $\mathbf{\Sigma}_{\mathbf{M}i}$ and $\mathbf{\Sigma}_{\mathbf{M}j}$ are the center normalized covariance matrices of two regions Θ_i and Θ_j, respectively. Then the generalized likelihood ratio is

$$Q_1 = \frac{L_{H_0}(\hat{\mathbf{\Sigma}}_{\mathbf{M}j} \mid \Theta_i)}{L_{H_1}(\hat{\mathbf{\Sigma}}_{\mathbf{M}i} \mid \Theta_i)} = \frac{\prod_{n=1}^{N_i} \frac{1}{(\pi)^m \tau_n^m \mid \hat{\mathbf{\Sigma}}_{\mathbf{M}j} \mid} \exp\left(-\frac{\mathbf{k}_n^\dagger \hat{\mathbf{\Sigma}}_{\mathbf{M}j}^{-1} \mathbf{k}_n}{\tau_n}\right)}{\prod_{n=1}^{N_i} \frac{1}{(\pi)^m \tau_n^m \mid \hat{\mathbf{\Sigma}}_{\mathbf{M}i} \mid} \exp\left(-\frac{\mathbf{k}_n^\dagger \hat{\mathbf{\Sigma}}_{\mathbf{M}i}^{-1} \mathbf{k}_n}{\tau_n}\right)}. \qquad (11)$$

$\hat{\mathbf{\Sigma}}_{\mathbf{M}i}$, $\hat{\mathbf{\Sigma}}_{\mathbf{M}j}$ are $(1/N_i)\sum_{i=1}^{N_i}\hat{\mathbf{C}}_i$ and $(1/N_j)\sum_{i=1}^{N_j}\hat{\mathbf{C}}_i$, respectively. Correspondingly, maximizing Q_1 in (11) is equivalent to minimizing the following SIRV distance like

$$D_{\mathrm{SIRV}}(\Theta_i, \Theta_j) = \ln\left(\frac{\mid \hat{\mathbf{\Sigma}}_{\mathbf{M}j} \mid}{\mid \hat{\mathbf{\Sigma}}_{\mathbf{M}i} \mid}\right) + \frac{m}{N_i} \sum_{n=1}^{N_i} \frac{\mathbf{k}_n^\dagger \hat{\mathbf{\Sigma}}_{\mathbf{M}j}^{-1} \mathbf{k}_n}{\mathbf{k}_n^\dagger \hat{\mathbf{\Sigma}}_{\mathbf{M}i}^{-1} \mathbf{k}_n}. \qquad (12)$$

It can be observed that this distance measure considers the polarimetric information from neighborhood pixels, which can reduce the effect of speckle noise to some extent. However, this distance measure is not symmetric, which makes it unsuitable for superpixel generation. Similar to the definition of symmetric revised Wishart distance in (Anfinsen, Jenssen, and Eltoft 2007), we define the symmetric SIRV distance as

$$SD_{SIRV}(\Theta_i, \Theta_j) = \frac{1}{2}\left(D_{SIRV}(\Theta_i, \Theta_j) + D_{SIRV}(\Theta_j, \Theta_i)\right)$$
$$= \frac{m}{2}\left(\frac{1}{N_i}\sum_{n=1}^{N_i}\frac{\mathbf{k}_n^\dagger \hat{\boldsymbol{\Sigma}}_{M_j}^{-1}\mathbf{k}_n}{\mathbf{k}_n^\dagger \hat{\boldsymbol{\Sigma}}_{M_i}^{-1}\mathbf{k}_n} + \frac{1}{N_j}\sum_{n=1}^{N_j}\frac{\mathbf{k}_n^\dagger \hat{\boldsymbol{\Sigma}}_{M_i}^{-1}\mathbf{k}_n}{\mathbf{k}_n^\dagger \hat{\boldsymbol{\Sigma}}_{M_j}^{-1}\mathbf{k}_n}\right) \quad (13)$$

where it can be seen that the first term of (12) is removed and the symmetric distance is dominated by the second term which takes into account the neighbor observed samples. This is beneficial for some distributed targets, such as the complex buildings in heterogeneous urban areas.

3. METHODOLOGY OF SUPERPIXEL GENERATION

3.1 SLIC Superpixel Segmentation

The basic idea of SLIC is the performance of local k-means clustering, and image pixels are iteratively assigned to neighborhood superpixels with close pixel gray value and spatial location (Achanta et al. 2012). The procedure includes three steps: 1) initialization of the cluster centers; 2) k-means clustering in a local region; and 3) post-processing.

Let N_p be the total pixel number, and K is the desired superpixel number. Initially, K cluster centers are sampled on a regular grid with uniform step size $S = \sqrt{N_p / K}$. The centers are then moved to the locations corresponding to the lowest gradient position in a 3×3 neighborhood, which avoids centering a superpixel on an edge and to reduce the chance of seeding a superpixel with a noisy pixel. Next, in the assignment step, each pixel is assigned to the nearest cluster center with the least distance whose search region overlaps its location, and then the superpixel centers will be updated. In terms of the distance measure D_{SLIC}, the pixel spectral information and the spatial location information are combined into a single distance measure as (Achanta et al. 2012)

$$D_{SLIC} = \sqrt{\left(\frac{d_p}{\lambda}\right)^2 + \left(\frac{d_s}{S}\right)^2} \quad (14)$$

where d_s is the Euclidean distance between two superpixels with center locations at (x_i, y_i) and (x_j, y_j). d_p represents the pixel spectral distance, which is the Euclidean distance in the three dimensional CIELAB color space. It should be noted that if simply defining D_{SLIC} to be five dimensional Euclidean distance, there will be inconsistencies in clustering behavior for different superpixel sizes. For example, for large superpixels, spatial distances outweigh color proximity, giving more relative importance to spatial proximity than color. This will produce compact superpixels those do not adhere well to image boundaries. In contrast, for smaller superpixels, the converse is true. Therefore, it is necessary to normalize color proximity and spatial proximity by their respective maximum distances within a cluster. The maximum spatial distance within a given cluster should be the sampling interval S. However, for color distances, since it can vary significantly in different clusters, it is not easy to determine the maximum color distance λ in (14). This problem can be avoided by modifying (14) as

$$D_{SLIC} = \sqrt{\left(d_p\right)^2 + \beta \cdot \left(\frac{d_s}{S}\right)^2} \quad (15)$$

where β is a weighting factor to be set manually that controls the shape and compactness of superpixels. Finally, after the clustering procedure, the pixels in the same cluster may be disjointed in space. Such pixels should be assigned the label of the nearest cluster center using a connected components algorithm.

3.2 Edge Detector

In CFAR edge detection algorithm, a set of filters with different orientations are applied on each pixel of a PolSAR image to calculate the edge map. The filter is displayed in Fig. 1 (a), which is controlled by four parameters, i.e., the length l_f, the width w_f, the spacing d_f between two rectangle regions, and the angular increment θ_f between two orientations. These filters estimate the average covariance matrix within the rectangle window on both sides of the center pixel and then calculate the Wishart distance as a measure of the probability of an edge pixel. The edge strength of each pixel is represented by the maximum distance from different sets of filters in this pixel. This method has been used in various PolSAR image applications (Lang et al. 2014; Liu et al. 2013), however, there are still two limitations:

i) Rectangle window functions are poor 2-D smoothing filters. Strong speckle in PolSAR data will diminish the average accuracy of covariance matrix since all the pixels are put equal weights;

ii) The Wishart distribution is not suitable for heterogeneous urban areas, resulting in incorrect covariance matrix estimation and the corresponding distance measure.

Filter banks have been proven to be effective for the edge detection since they can extract directional intensity variations. Inspired by this idea, we replace the rectangle-shaped filter with Gauss-shaped filter to overcome the first limitation of traditional CFAR edge detector, as shown in Fig. 1 (b). The horizontal Gauss-shaped window function is defined as:

$$W(x,y) = \frac{1}{\sqrt{2\pi}\sigma_x \sqrt{2\pi}\sigma_y}\exp\left(-\left(\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2}\right)\right) \quad (16)$$
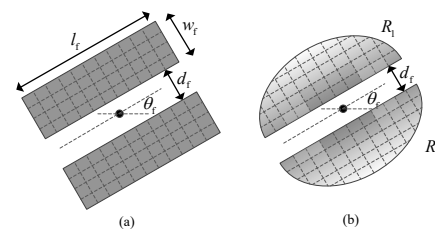

Fig. 1. Filter configuration. (a) Rectangle-shaped filter. (b). Gauss-shaped filter.

where σ_x and σ_y control the window length and width, respectively. W is the Gauss weight for each pixel, which will be used for the average of local covariance matrix of center pixel on both sides of the window. From (16) it can be observed that the pixels near the center pixel have larger weights than other pixels. This is in accordance with the fact that information contained at the pixels near the center pixel is more important than those at other pixels when deciding whether the center pixel is an edge pixel. At each orientation, the local average center covariance matrix for a PolSAR image can be computed as

$$\hat{\Sigma}_{\mathbf{M}} = \frac{\sum\limits_{(x,y)} W(x,y)\hat{\mathbf{C}}(x,y)}{\sum\limits_{(x,y)} W(x,y)}. \qquad (17)$$

After that, we utilize the SIRV distance given in (13) to calculate the similarity of center normalized covariance matrices on both sides of the central pixel.

3.3 Homogeneity Measurement

The new edge map introduced in previous subsection not only is beneficial to the initialization of cluster centers in SLIC, but also will be helpful to analyze the homogeneity in PolSAR images. Inspired by (Lang et al. 2014), in this paper, we utilize the ENL estimation method proposed in (Anfinsen, Doulgeris, and Eltoft 2009) to calculate the ENL value for each pixel and then further combine the proposed edge map to define a homogeneity measurement for PolSAR data. It is worth pointing out that this ENL estimator is less affected by texture and thus provides more accurate results than other estimators. Therefore, although the technique in (Anfinsen, Doulgeris, and Eltoft 2009) is based on the Wishart distribution, it can be used for ENL estimation in this study.

The random matrix $\mathbf{Z}$ is assumed to be positive semidefinite and satisfy complex Wishart distribution with L degrees of freedom and scale matrix $\Sigma = \mathrm{E}\{\mathbf{Z}\}/L$. Then the following moment of $\mathbf{Z}$ can be derived as (Anfinsen, Doulgeris, and Eltoft 2009):

$$\mathrm{E}\{\mathrm{Tr}(\mathbf{ZZ})\} = L^2\,\mathrm{Tr}(\Sigma\Sigma) + L\,\mathrm{Tr}(\Sigma)^2. \qquad (18)$$

Based on (18), the trace moment-based estimator of ENL is derived as

$$\mathrm{ENL} = \frac{\mathrm{Tr}(\Sigma)^2}{\langle\mathrm{Tr}(\mathbf{XX})\rangle - \mathrm{Tr}(\Sigma\Sigma)} \qquad (19)$$

where $\mathbf{X} = \mathbf{Z}/L$. The detailed derivation of this estimator can be found in (Anfinsen, Doulgeris, and Eltoft 2009).

Since the ENL and edge map are both related to target polarimetric information and more importantly, their value trends in homogeneous and heterogeneous regions are opposite, combining the ENL and edge map can significantly improve the probability of discriminating homogeneous and heterogeneous areas. According to (Fengkai, Jie, and Deren 2015), the homogeneity measurement can be represented as

$$\mathrm{HoM} = \frac{\mathrm{ENL}}{\mathrm{EDGE}} \qquad (20)$$

where EDGE denotes the proposed edge map.

3.4 Similarity Measure with Multiple Cues

This subsection gives the distance measure for PolSAR superpixel generation and segmentation, which considers the polarimetric, texture, and spatial information at the same time. The homogeneity measurement is incorporated into the distance measure, making the trade-off factor adaptive to balance the shape and compactness of the superpixels. We firstly introduce the polarimetric similarity cue and texture similarity cue, then give the complete distance measure which is incorporated with the homogeneity measurement.

As we discussed before, the conventional covariance matrix can be decomposed into two parts, i.e., the normalized covariance matrix $\hat{\mathbf{C}}$ which contains the polarimetric information and the span $\hat{P}$ that contains scalar texture information. Since the distance measure in (13) mainly considers the polarimetric information, we define a texture distance D_T based on the estimated span $\hat{P}$ like

$$D_\mathrm{T} = \frac{\left|\hat{P}(x_i,y_i) - \hat{P}(x_j,y_j)\right|}{\max(\hat{P})} \qquad (21)$$

where $\max(\hat{P})$ denotes the maximum value of $\hat{P}$ and $|\cdot|$ represents the absolute value operator. It should be noted that since the estimated span only contains the scalar texture information without polarimetric information, simple subtraction operation can be applied on $\hat{P}$ directly.

In (15), β is utilized as a weighting factor to balance the spectral similarity and spatial proximity. Similarly, in (Song et al. 2015; Fachao, Jiming, and Fengkai 2015; Feng, Cao, and Pi 2014), this parameter is also chosen to be a constant to balance the polarimetric and spatial similarity. This parameter is usually set manually by trial and error, which might cause over- or under- superpixel segmentation in some spatially complicated areas. In this paper, the parameter β is set adaptively according to the local spatial complexity of the scene, which can be defined as

$$\beta_\mathrm{adp} = \beta \cdot \frac{1}{2}\big(\mathrm{HoM}(x_i,y_i) + \mathrm{HoM}(x_j,y_j)\big). \qquad (22)$$

It can be seen that this adaptive parameter considers the homogeneity measurement of two compared pixels. Then the complete adaptive distance measure for superpixel generation is defined as

$$D_\mathrm{Pol\text{-}ASLIC} = \sqrt{\left(SD_\mathrm{SIRV}\right)^2 + \left(D_\mathrm{T}\right)^2 + \beta_\mathrm{adp}\cdot\left(\frac{d_\mathrm{s}}{S}\right)^2}. \qquad (23)$$

It is worth noting that since D_T is already normalized, there is no need to set another trade-off factor. β_adp can be used to balance the spatial proximity and other two similarity measures. Specifically, for homogeneous areas where β_adp is high, there is not too much edge information, the spatial proximity overweighs other two similarity measures, leading to compact superpixels. In contrast, for heterogeneous areas, β_adp is low and can suppress the spatial proximity, therefore, the superpixels are generated mainly on the basis of polarimetric and texture information, which can well adhere to image boundaries. In (22), since the homogeneity measurement is quite large in homogeneous areas and low in heterogeneous areas, β can be set to 1 in different cases. Thus, the number of parameters in original SILC algorithm is reduced from two to only one, i.e., the number of superpixels, making the proposed method easy to use.

The implementation procedure of Pol-ASLIC is similar to that of SLIC except the steps before local iterative clustering, which covers the following steps.
Input: original PolSAR image
1) Normalized covariance matrix and span estimation.
2) Edge map calculation based on the normalized covariance matrix.
3) ENL estimation using (19).
4) Homogeneity measurement calculation using (20).
5) Set the number of superpixels K and initialize the cluster centers.

6) Local iterative clustering with the adaptive distance calculated using (23).

7) Post-processing to eliminate the disjointed pixels.

Output: superpixel map.

4. RESULTS AND ANALYSIS

4.1 Dataset Description

Here we choose one PolSAR dataset to demonstrate the effectiveness of our proposed method with visual presentation and quantitative evaluation and comparison. This dataset, as shown in Fig. 2 (a), was acquired by ESAR L band system with study area located in Oberpfaffenhofen, Germany. The number of looks is four. There are a lot of man-made buildings in Fig. 2 (a), which is a heterogeneous urban area. For the visual presentation and quantitative evaluation, the manual segmentations are used as the ground truth. We select one area from the Pauli image and depict the ground truth segmentations, which are yellow lines in Fig. 2 (b).
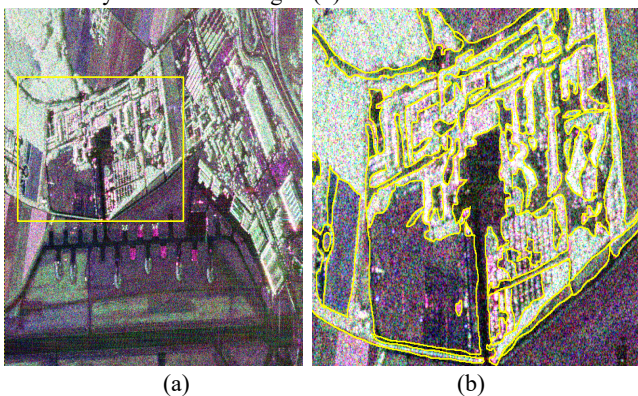

(a)   (b)

Fig. 2. ESAR L-band image from the Oberpfaffenhofen area in Germany. (a) Pauli RGB image. (b). The yellow lines are superimposed onto Pauli RGB image, depicting the ground truth segments.

In our experiment, two superpixel segmentation methods for PolSAR data, i.e., Qin's method in (Fachao, Jiming, and Fengkai 2015) and Liu's method in (Liu et al. 2013), are utilized for comparison. The former is a modified version of SLIC and the latter is designed based on Normalized cuts. In these two compared approaches, the edge maps produced by traditional CFAR edge detector are both incorporated into the superpixel segmentation.

4.2 Comparison of the Superpixel Segmentation Results

In this subsection, we will compare the superpixel segmentation results of three methods by visual assessment and quantitative evaluation. Fig. 3 gives the superpixel segmentation results of ESAR dataset using Liu's, Qin's, and our proposed methods, respectively. The numbers of superpixel are all set to 2200. To compare the results in detail, two areas marked with yellow rectangles A and B are selected from Fig. 3 (c)-(f) and are shown in Fig. 4. Area A mainly includes the buildings while Area B covers natural targets, as well as some man-made targets. Fig. 3 (a), (c), and (e) present the final superpixel maps of three methods, where the red lines superimposed onto the Pauli images depict the superpixel boundaries. Fig. 3 (b), (d), and (f) give the corresponding representation maps, in which the coherency matrix of each pixel is replaced by the average coherency matrix of the superpixel this pixel belongs to. From Fig. 3 (a) and (b), we can see that the edges of the superpixels

are very smooth, and the shapes of the superpixels are quite regular. In natural areas, the results are acceptable. However, in urban areas, these superpixels cannot adhere well to image edges, the points and lines in the image cannot be preserved and most of the urban information is lost. There are two reasons for this result. On one hand, the edge map calculated by traditional CFAR edge detector does not work well in heterogeneous urban areas. On the other hand, unlike SLIC, Normalized cuts algorithm does not consider the pixel local information. Compared with Fig. 3 (a), the result in Fig. 3 (c) seems much better, where most of the edges and points are preserved. The superpixels can well adhere to image boundaries and capture the local information.
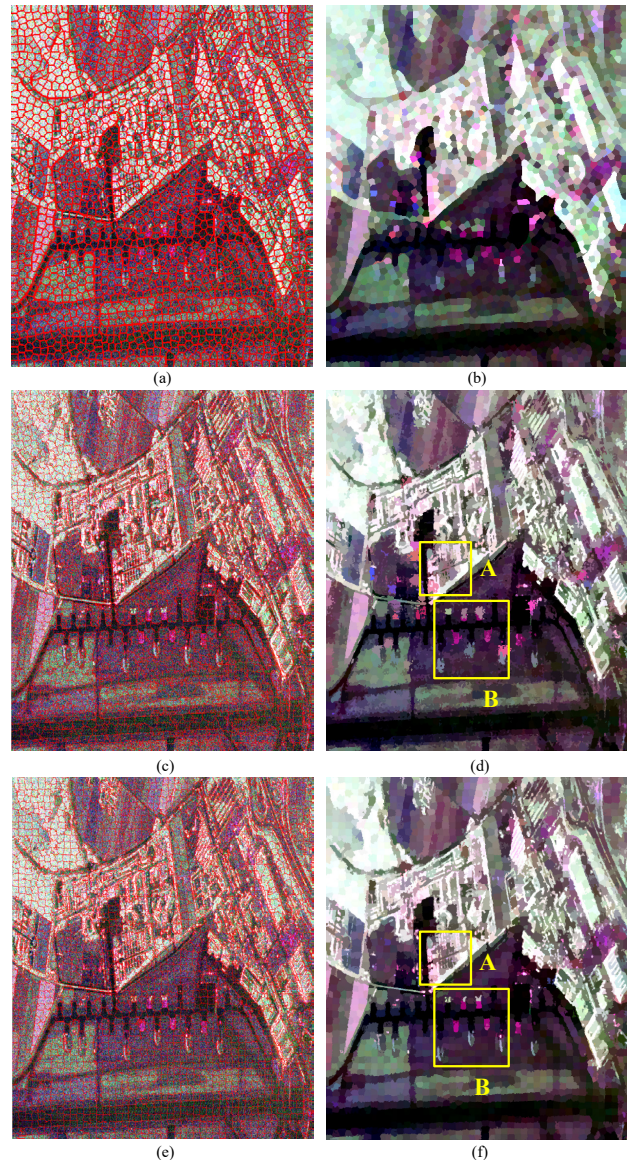

(a)   (b)

(c)   (d)

(e)   (f)

Fig. 3. Superpixel generation results of Liu's, Qin's, and our proposed approaches with $K = 2200$ for ESAR image. The first column denotes the final superpixel maps of different methods. The red lines superimposed onto the Pauli images depict the superpixel boundaries. The second column gives the representation maps, where the coherency matrix of each pixel is replaced by the average value of the superpixel this pixel belongs to.

However, the shape of the superpixel is very irregular and the edges are not smooth, even in the homogeneous natural areas,

as shown in Fig. 4 (c) and (g). In this method, to well preserve the edges and points, the trade-off factor which balances the polarimetric similarity and spatial proximity is set to 1.0.

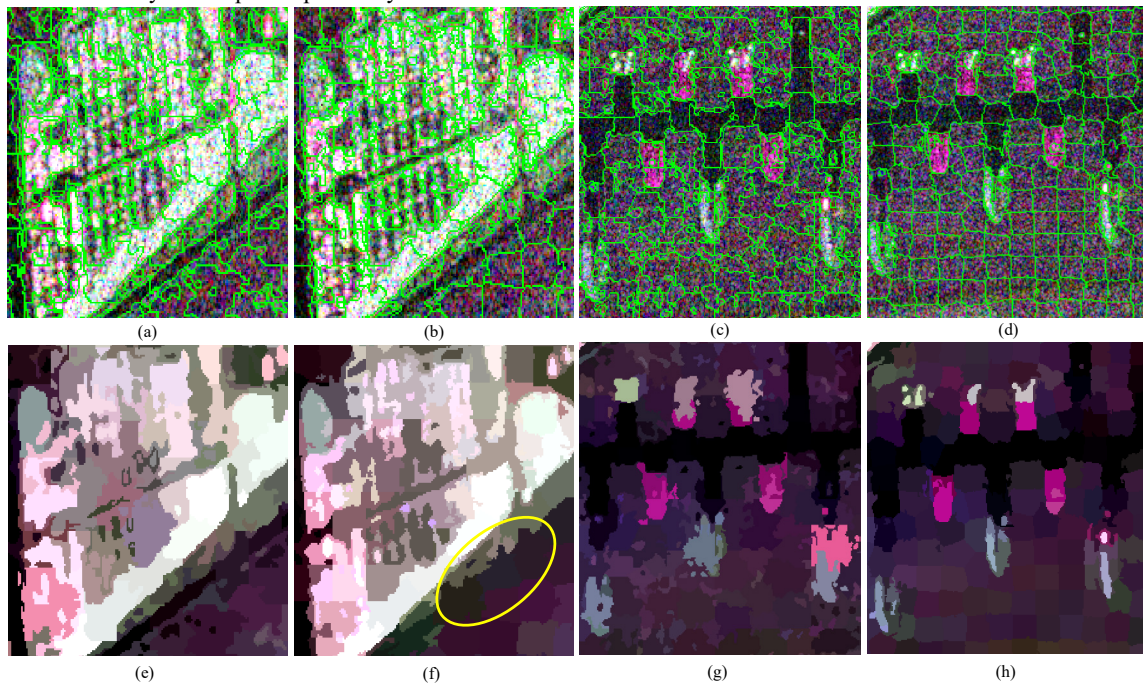Therefore, the polarimetric similarity overweighs spatial proximity, leading to irregular superpixels.

Fig. 4. Comparison of detailed superpixel generation results in area A and B. The first row denotes the final superpixel maps. The green lines superimposed onto the Pauli images depict the superpixel boundaries. The second row gives the corresponding representation maps. (a) and (b) are the results of area A in Fig. 3 (c) and (e), respectively. (c) and (d) are the results of area B in Fig. 3 (c) and (e), respectively.

The results in Fig. 3 (e) and (f) indicate that our proposed algorithm can generate promising superpixels for PolSAR images. The target points and edges can be preserved very well. Moreover, the compactness of the superpixels is adaptive. In homogeneous areas, the edges of superpixels are very smooth and the superpixel shapes are quite regular, which can be seen in Fig. 4 (d) and (h). This is because in such areas, the homogeneity measurement is high, making the spatial proximity overweigh other two similarities. Therefore, the superpixels are compact and regular. In contrast, within heterogeneous areas, the homogeneity measurement is low. To preserve the detailed information, spatial proximity is not as important as polarimetric and texture similarities any more. Therefore, the superpixels have irregular shapes and can well preserve the image edges and points. From Fig. 4, we can also see that in heterogeneous areas, our method can achieve better results than Qin's method, where the building edges are clearer and the man-made targets are better extracted. This is because our new edge detector based on the SIRV product model can detect more accurate edges. In addition, the proposed distance measure considers more local information for superpixel generation, such as the span information. Compared with other two methods, it can be concluded that the proposed method provides smoother approximations in homogeneous areas, and also can keep better details in heterogeneous areas.

4.3 Quantitative Evaluation

In this paper, to perform a quantitative comparison of different methods, we adopt two commonly used evaluation metrics: i.e., boundary recall (BR) and achievable segmentation accuracy (ASA). BR is defined as the fraction of ground truth boundaries correctly recovered by the superpixel edges. If a true boundary

pixel falls within 2 pixels from at least one superpixel edge, it can be regarded to be recovered correctly. Therefore, a high BR indicates that the superpixels can well adhere to image edges and very few true boundaries are missed. ASA is defined as the highest achievable accuracy of object segmentation when regarding the superpixels as units. By labeling each superpixel with the ground truth segments of the largest overlapping area, ASA can be obtained as the fraction of labeled pixels those are not leaked from the ground truth boundaries. Thus a high ASA means that the superpixels comply well with objects in the PolSAR image. These two indicators can evaluate the final superpixel maps.
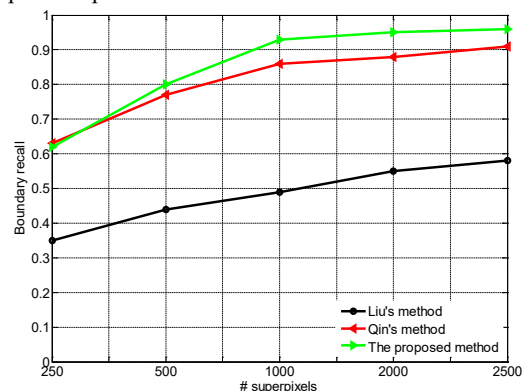

Fig. 5. Boundary recall (BR) of three methods with different superpixel numbers.

Fig. 5 and Fig. 6 depict the BR and ASA of three methods with different numbers of superpixels, respectively. This number is set from 250 to 2500 with different step sizes. According to these two figures, Liu's method performs the worst in terms of

boundary adherence and achievable segmentation accuracy. In addition, another drawback is its extremely low time efficiency. Qin's method and our proposed method have similar BR values when the superpixel number does not exceed 500. However, if we increase this value, our method has a better boundary adherence than Qin's method. In Fig. 6, these two methods have similar results but our method still performs slightly better than Qin's approach.
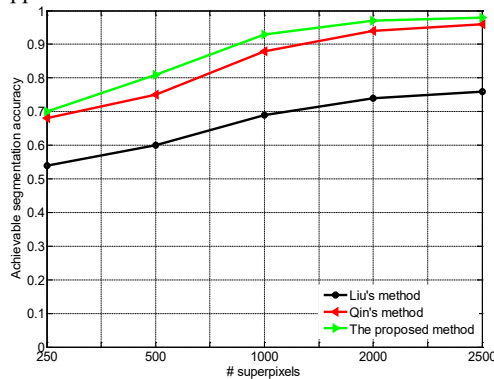

Fig. 6. Achievable segmentation accuracy (ASA) of three methods with different superpixel numbers.

It is worth mentioning that the number of superpixels in these methods is set to 2200. Although increasing this value can achieve better results for all the three methods, it is not recommendatory to produce too many superpixels for PolSAR data. The main advantage of superpixel generation is that superpixels can cause substantial speed-up of subsequent processing since the number of superpixels of an image is significantly lower than the number of pixels. Therefore, the goal of our approach is to get better segments for PolSAR data with a limited number of superpixels.

5. CONCLUSION

This paper proposes an adaptive superpixel segmemtaton method for PolSAR images. The whole framework is designed on the basis of local iterative clustering and a heterogeneous statistical model, which has three main contributions, i.e., edge map calculation, homogeneity measurement evaluation, and adaptive distance measure definition. The Gauss-shaped filter and SIRV model have been utilized to improve the traditional CFAR edge detector for PolSAR images. The edges of urban areas can be effectively extracted and the locations are also accurate. Multiple clues are combined in the distance measure for superpixel segmentation, making it more effective to cluster pixels since more local information is considered. Moreover, this distance measure can balance different similarities adaptively, according to the homogeneity measurement. The performance of our proposed method is demonstrated on one PolSAR dataset from the ESAR platform, with visual and quantitative evaluation. The superpixel generation and segmentation results show that our proposed method has a better performance and can balance over segmentation and under segmentation more effectively than other methods.

REFERENCES

Achanta, Radhakrishna, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Susstrunk. 2012. "SLIC superpixels compared to state-of-the-art superpixel methods." Review of. *IEEE Trans. Pattern Anal. Mach. Intell.* 34 (11):2274-82.

Anfinsen, Stian Normann, Anthony P Doulgeris, and Torbjørn Eltoft. 2009. "Estimation of the equivalent number of looks in polarimetric synthetic aperture radar imagery." Review of. *IEEE Trans. Geosci. Remote Sens.* 47 (11):3795-809.

Anfinsen, Stian Normann, Robert Jenssen, and Torbjørn Eltoft. 2007. "Spectral clustering of polarimetric SAR data with Wishart-derived distance measures." In *Proc. POLinSAR*, 1-9.

Ban, Yifang, and Anoj Jacob. 2013. "Object-based fusion of multitemporal multiangle ENVISAT ASAR and HJ-1B multispectral data for urban land-cover mapping." Review of. *IEEE Transactions on Geoscience and Remote Sensing* 51 (4):1998-2006.

Bombrun, Lionel, Gabriel Vasile, Michel Gay, and Felix Totir. 2011. "Hierarchical segmentation of polarimetric SAR images using heterogeneous clutter models." Review of. *IEEE Trans. Geosci. Remote Sens.* 49 (2):726-37.

Doulgeris, A. P., S. N. Anfinsen, and T. Eltoft. 2008. "Classification with a non-Gaussian model for PolSAR data." Review of. *IEEE Trans. Geosci. Remote Sens.* 46 (10):2999-3009. doi: 10.1109/TGRS.2008.923025.

Fachao, Qin, Guo Jiming, and Lang Fengkai. 2015. "Superpixel segmentation for Polarimetric SAR imagery using local iterative clustering." Review of. *IEEE Geosci. Remote Sens. Lett.* 12 (1):13-7. doi: 10.1109/LGRS.2014.2322960.

Feng, Jilan, Zongjie Cao, and Yiming Pi. 2014. "Polarimetric contextual classification of PolSAR images using sparse representation and superpixels." Review of. *Remote Sens.* 6 (8):7158-81.

Fengkai, Lang, Yang Jie, and Li Deren. 2015. "Adaptive-window Polarimetric SAR image speckle filtering based on a homogeneity measurement." Review of. *IEEE Trans. Geosci. Remote Sens.* 53 (10):5435-46. doi: 10.1109/TGRS.2015.2422737.

Lang, Fengkai, Jie Yang, Deren Li, Lingli Zhao, and Lei Shi. 2014. "Polarimetric SAR image segmentation using statistical region merging." Review of. *IEEE Geosci. Remote Sens. Lett.* 11 (2):509-13.

Liu, Bin, Hao Hu, Huanyu Wang, Kaizhi Wang, Xingzhao Liu, and Wenxian Yu. 2013. "Superpixel-based classification with an adaptive number of classes for polarimetric SAR images." Review of. *IEEE Trans. Geosci. Remote Sens.* 51 (2):907-24.

Niu, Xin, and Yifang Ban. 2013. "Multi-temporal RADARSAT-2 polarimetric SAR data for urban land-cover classification using an object-based support vector machine and a rule-based approach." Review of. *International Journal of Remote Sensing* 34 (1):1-26.

Soergel, U. 2010. *Radar remote sensing of urban areas*. Berlin, Germany: Springer-Verlag.

Song, H, W Yang, Y Bai, and X Xu. 2015. "Unsupervised classification of polarimetric SAR imagery using large-scale spectral clustering with spatial constraints." Review of. *Int. J. Remote Sens.* 36 (11):2816-30.

Vasile, G., J. P. Ovarlez, F. Pascal, C. Tison, L. Bombrun, M. Gay, and E. Trouve. 2008. "Normalized coherency matrix estimation under the SIRV model. alpine

glacier Polsar data analysis." In *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, I-74-I-7. Boston, MA.

Vasile, G., J. Ovarlez, F. Pascal, and C. Tison. 2010. "Coherency matrix estimation of heterogeneous clutter in high-resolution Polarimetric SAR images." Review of. *IEEE Trans. Geosci. Remote Sens.* 48 (4):1809-26. doi: 10.1109/TGRS.2009.2035496.

Wenjin, Wu, Guo Huadong, and Li Xinwu. 2015. "Urban area SAR image man-made target extraction based on the product model and the time-frequency analysis." Review of. *IEEE J. Selected Topics in Applied Earth Observations and Remote Sensing (JSTARS)* 8 (3):943-52. doi: 10.1109/JSTARS.2014.2371064.

Xiang, Deliang, Tao Tang, Lingjun Zhao, and Yi Su. 2013. "Superpixel generating algorithm based on pixel intensity and location similarity for SAR image classification." Review of. *IEEE Geosci. Remote Sens. Lett.* 10 (6):1414-8.

Yao, K. 1973. "A representation theorem and its applications to spherically-invariant random processes." Review of. *IEEE Trans. Inf. Theory* 19 (5):600-8. doi: 10.1109/TIT.1973.1055076.