SAR Image Segmentation Based on Hierarchical Visual Semantic and Adaptive Neighborhood Multinomial Latent Model

Fang Liu, *Senior Member, IEEE*, Yiping Duan, Lingling Li, Licheng Jiao, *Senior Member, IEEE*, Jie Wu, Shuyuan Yang, *Senior Member, IEEE*, Xiangrong Zhang, and Jialin Yuan

***Abstract*—A synthetic aperture radar (SAR) imaging system usually produces pairs of bright area and dark area when depicting the ground objects, such as a building or tree and its shadow. Many buildings (trees) are aggregated together to form urban areas (forests). It means that the pairs of bright and dark areas often exist in the aggregated scenes. Conventional unsupervised segmentation approaches usually segment the scenes (e.g., urban areas and forests) into different regions simply according to the gray values of the image. However, a more convincing way is to regard them as the consistent regions. In this paper, we aim at addressing this issue and propose a new SAR image segmentation approach via a hierarchical visual semantic and adaptive neighborhood multinomial latent model. In this approach, the hierarchical visual semantic of SAR images is proposed, which divides SAR images into aggregated, structural, and homogeneous regions. Based on the division, different segmentation methods are chosen for these regions with different characteristics. For the aggregated region, locality-constrained linear coding-based hierarchical clustering is used for segmentation. For the structural region, visual semantic rules are designed for line object location, and a geometric structure window-based multinomial latent model is proposed for segmentation. For the homogeneous region, a multinomial latent model with adaptive window selection is proposed for segmentation. Finally, these results are integrated together to obtain the final segmentation. Experiments on both synthetic and real SAR images indicate that the proposed method achieves promising performances in terms of the consistencies of the regions and the preservations of the edges and line objects.**

***Index Terms*—Adaptive neighborhood, hierarchical visual semantic, multinomial latent model, regional map, synthetic aperture radar (SAR) image segmentation.**

I. Introduction

With the rapid development of the spaceborne and airborne synthetic aperture radar (SAR) systems, massive quantities of SAR images are available for the increasing number of applications. Although many methods have got promising results, SAR image segmentation, as a crucial step for SAR image understanding and interpretation, is a challenging task for its heterogeneity of the urban areas and forests [1], [2]. The available SAR image segmentation methods can be roughly grouped into feature-based methods and model-based methods. The first category extracts features from SAR images, such as the gray level co-occurrence matrix (GLCM) [3], [4], Gabor [5], SAR-SIFT [6], and so on. These features are combined with the clustering methods [7]–[12] for segmentation. The low-level features extracted from the SAR images represent only pixel-level information and are easily influenced by the speckle noise. The second category often establishes the statistical models of SAR images and mainly includes level-set methods [13]–[15], random field methods and their variants [16]–[23], the multinomial latent model [24], and so on. Since the model-based methods incorporate the spatial context information into the segmentation, they are attracting the increasing attentions in SAR image segmentation.

The Markov random field (MRF) model [17] with Gibbs prior is a basic context model. However, this model may lead to oversegmentation, particularly in the urban areas and forests. In [18], a region-based MRF model was proposed. In the method, the traditional MRF model was built on the superpixel, and the edge penalty function was incorporated into the spatial context model for a better performance. Due to the usage of the region-based MRF model, the oversegmentation was largely reduced in the segmentation results. However, inconsistent semantic still can be found in the results, particularly in the urban areas and forests. In [19], using the Fisher distribution to model the local texture property, an improved MRF model was proposed for SAR image segmentation. Furthermore, a hierarchical MRF model [3] was proposed for SAR image segmentation.

Manuscript received September 28, 2014; revised June 4, 2015 and December 22, 2015; accepted February 21, 2016. Date of publication April 1, 2016; date of current version May 24, 2016. This work was supported in part by the National Basic Research Program (973 Program) of China under Grant 2013CB329402, by the National Natural Science Foundation of China under Grants 61573267, 61571342, and 61572383, by the Program for Cheung Kong Scholars and Innovative Research Team in University under Grant IRT_15R53, by the Fund for Foreign Scholars in University Research and Teaching Programs (111 Project) under Grant B07048, by the Major Research Plan of the National Natural Science Foundation of China under Grants 91438201 and 91438103, and by the Fundamental Research Funds for the Central Universities under Grants JB140317 and GK201603083.

F. Liu, Y. Duan, J. Wu, and J. Yuan are with the School of Computer Science and Technology, Xidian University, Xi'an 710071, China, and also with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, International Research Center for Intelligent Perception and Computation, Joint International Research Laboratory of Intelligent Perception and Computation, Xidian University, Xi'an 710071, China (e-mail: f63liu@163.com).

L. Li, L. Jiao, S. Yang, and X. Zhang are with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, International Research Center for Intelligent Perception and Computation, Joint International Research Laboratory of Intelligent Perception and Computation, Xidian University, Xi'an 710071, China.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2016.2539155

However, owing to the usage of the quad tree in the hierarchical model, blocking artifacts were always observed, and the details were often lost in the results. In [24], an unsupervised classification approach based on the multinomial latent model with amplitude and texture features (ATML-CEM) was proposed. In ATML-CEM, the amplitude and texture features of SAR images were assembled into a finite mixture model. A multinomial logistic function was used to describe the spatial context model. It is noted that the model is built on a fixed neighborhood. Practically, a large window will produce an accurate segmentation in the homogeneous regions but poor segmentation nearby the details (e.g., edges and line objects). A small window is more appropriate nearby the details. In summary, these methods based on statistical characteristics with low-level features have the following disadvantages: 1) the predefined spatial context model may not be suitable for regions with different characteristics; for example, the spatial contexts of the urban areas and edges are very different; 2) the semantic is not taken into account in these methods, which leads to the results without consistent semantic, particularly in the urban areas and forests; and 3) the line objects are not considered in most cases, which are confused with other regions. On the one hand, the content in SAR images is more related to the structural and shape characteristics of the land covers rather than the spectral responses of the surface materials [25]. On the other hand, SAR images are not only a random collection of the pixels but also a meaningful arrangement of the regions and objects. There is usually a large semantic gap between the low-level features and the high-level user semantic [26]. Therefore, segmentation methods without considering the structure and semantic of SAR images will produce the segmentation results with unsatisfied details and inconsistent semantic.

Both consistent regions and details' preservations are equally important for SAR image segmentation. It is well known that SAR images are essentially rich in structures that result from the surface roughness [25]. Moreover, owing to the side looking of the SAR imaging system, an object and its corresponding shadow are usually observed, for example, the building or tree and its shadow. Since SAR images are usually an observation of the Earth, an urban area (consisting of many buildings) and a forest (consisting of many trees) are commonly found in SAR images. It means that the pairs of the bright and dark areas are repeated in these aggregated scenes in the SAR image. Although these aggregated scenes are consistent in semantic, it is very difficult to obtain a region containing the whole forest or the whole urban area. Usually, the bright area is segmented into one class, while the dark shadow is segmented into another. Therefore, the main challenges of SAR image segmentation are the segmentation of the regions containing the whole aggregated scene and the preservations of the details simultaneously.

It is well known that the retina has the capability to interpret the images into meaningful regions. It indicates that our retina always groups those pixels into meaningful portions. Learning from the perception of the retina, some semantics are explored for SAR image segmentation. Many advanced methods have been done for this [27]–[29]. In [27], a primal sketch graph was proposed for natural images. By using the sketch information, the image is divided into sketchable and nonsketchable portions for the image reconstruction. Furthermore, the sketch map of the SAR image was proposed in [28]. In the article, a geometrical kernel function based on the sketch map was proposed for SAR speckle reduction, and a promising performance was obtained. The sketch map describes the edge and line features of SAR images. It is a sparse structure representation of SAR images.

In this paper, according to the SAR sketch map, a hierarchical visual semantic of SAR images is proposed, which divides SAR images into aggregated, structural, and homogeneous regions. Based on the division, different segmentation methods are chosen for different regions. For the aggregated region, locality-constrained linear coding (LLC) [30] based hierarchical clustering [8] is used for segmentation. For the structural region, in order to locate the line objects, the visual semantic rules are designed by analyzing the edge model and line model. Moreover, for the edges and line objects, a strong relationship exists along their directions rather than across them, so an orientation-based geometric structure window (GSW) is plugged into the multinomial latent model for segmentation. For the homogeneous region, in order to find an appropriate neighborhood to represent the central pixel, an adaptive window selection method is utilized. Then, a multinomial latent model with the adaptive window is used for segmentation. These results are integrated together to obtain the final segmentation. Compared with the available SAR image segmentation approaches, our proposed approach has the following characteristics: 1) a hierarchical visual semantic of SAR images is proposed, which divides SAR images into regions with different characteristics; 2) adaptive segmentation methods are proposed for different regions; and 3) visual semantic rules are designed to locate the line objects.

The rest of this paper is organized as follows. In Section II, the hierarchical visual semantic of SAR images is proposed. Section III describes the segmentation method based on the hierarchical visual semantic and adaptive neighborhood multinomial latent model (HVS-ANML). In Section IV, experimental results and analyses are presented. Section V concludes this paper and presents some perspectives of our future work.

II. Hierarchical Visual Semantic

In this section, according to the SAR sketch map, the hierarchical visual semantic of SAR images is proposed. On the hierarchical visual semantic, the first level is the original SAR image, the second level is the SAR sketch map, and the third level is the regional map.

Following Marr's insight [31], Guo *et al.* [27] proposed a primal sketch graph for natural images. However, the characteristics of SAR images are different from the natural images. Considering these different characteristics, Wu *et al.* [28] proposed a sketch map for SAR images, which is obtained mainly by the following steps.

1) By designing the edge and line templates at different scales and orientations, the responses obtained by the ratio of average, cross-correlation, and gradient operators are adaptively fused to calculate the edge-line intensity map of a SAR image.

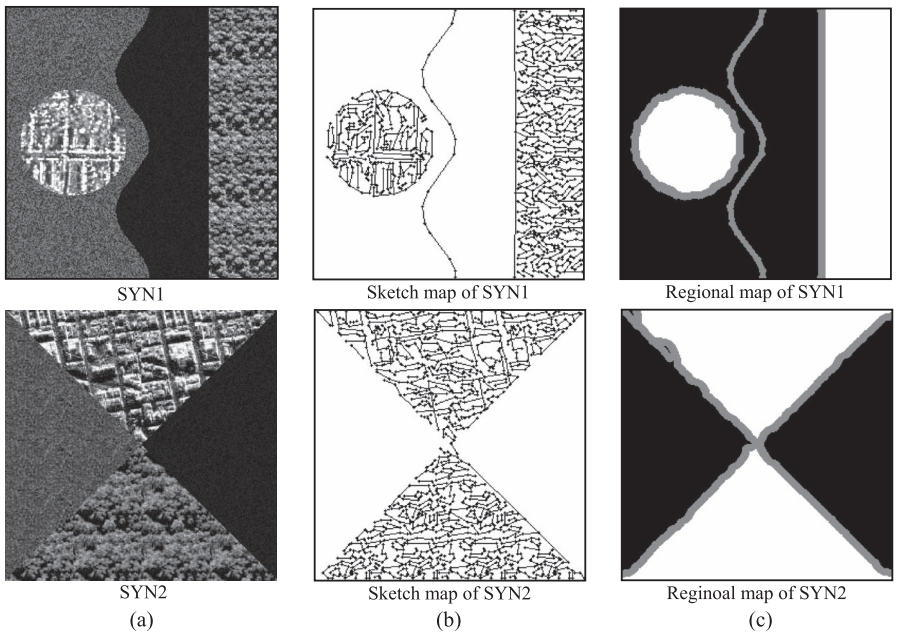

Fig. 1. (a) Original synthetic SAR images. (b) Sketch maps. (c) Regional maps, which include aggregated region (in white), structural region (in gray), and homogeneous region (in black).

2) Using the nonmax suppression and double-threshold-based connection methods, the curves contained in the SAR image are extracted.
3) Each extracted curve is represented as a sketch line via an approximation method.
4) According to the properties of SAR images, a pair of paradoxical hypotheses is built to evaluate the significance of each sketch line, and only the significant sketch lines are preserved to constitute the sketch map of the SAR image.

In the SAR sketch map [see Figs. 1(b) and 2(b)], the black lines are the sketch lines. Each sketch line consists of several sketch line segments, which are connected end to end. The end points of each sketch line segment are marked as the solid dots. Moreover, the sketch lines in the aggregated scenes are aggregated, while the sketch lines indicating the edges and line objects are not aggregated. It means that the sketch lines in different regions usually have different characteristics. According to these different characteristics, the aggregated degree is used to classify the sketch lines into aggregated and nonaggregated sketch lines. Then, the region extractor is implemented on the aggregated sketch lines to obtain the aggregated region [32] (the details are introduced in the Appendix). The GSW is operated on the nonaggregated sketch lines to obtain the structural region, where GSW is an oriented rectangular window. Specifically, two sides of GSW are parallel to the sketch line, and the other two sides are vertical to the sketch line. The remaining part of the sketch map is the homogeneous region. The aggregated, structural, and homogeneous regions constitute the regional map. According to the obtained regional map, a SAR image is reasonably divided into aggregated, structural, and homogeneous regions. It is a kind of sparse representation [33] of SAR images. Specifically, the aggregated scenes are usually contained in the aggregated portion. The details are always contained in the structural portion. The regions with a slight change of gray value are mostly contained in the homogeneous portion. The sketch maps and regional maps of two synthetic SAR images and seven real SAR images are shown in Figs. 1 and 2, respectively. In the sketch maps [shown in Figs. 1(b) and 2(b)], the black lines are the sketch lines. We can see that the sketch lines in the real SAR images are more complicated than that in the synthetic SAR images. The regional maps [shown in Figs. 1(c) and 2(c)] include three portions. The white, gray, and black portions are the aggregated, structural, and homogeneous regions, respectively.

From the above, we can see that the sketch map is a primal semantic space which represents the structure of the SAR image. The primitive of the sketch map is a sketch line. The regional map is a higher semantic space, whose primitive is a region. Moreover, the semantic is based on the framework of Marr's vision. Therefore, we name it as hierarchical visual semantic, which is used to guide the SAR image segmentation. Specifically, the regional map is used to divide SAR images into aggregated, structural, and homogeneous regions. It can improve the consistencies of the aggregated land covers in the segmentation process. The sketch map is used to design GSW and visual semantic rules. It can improve the location accuracy of the edges and line objects in the segmentation process. Since different levels of semantics are used to deal with regions with different characteristics, the consistencies of the regions and the details' preservations will be achieved simultaneously in the segmentation process. The hierarchical visual semantic and its corresponding semantic representation of SAR images are shown in Fig. 3. Fig. 3(a) shows the hierarchical visual semantic. The first level is the original SAR image, the second level is the sketch map, and the third level is the regional map. In Fig. 3(b), the regional map is mapped into the SAR image, which is divided into aggregated, structural, and homogenous regions.

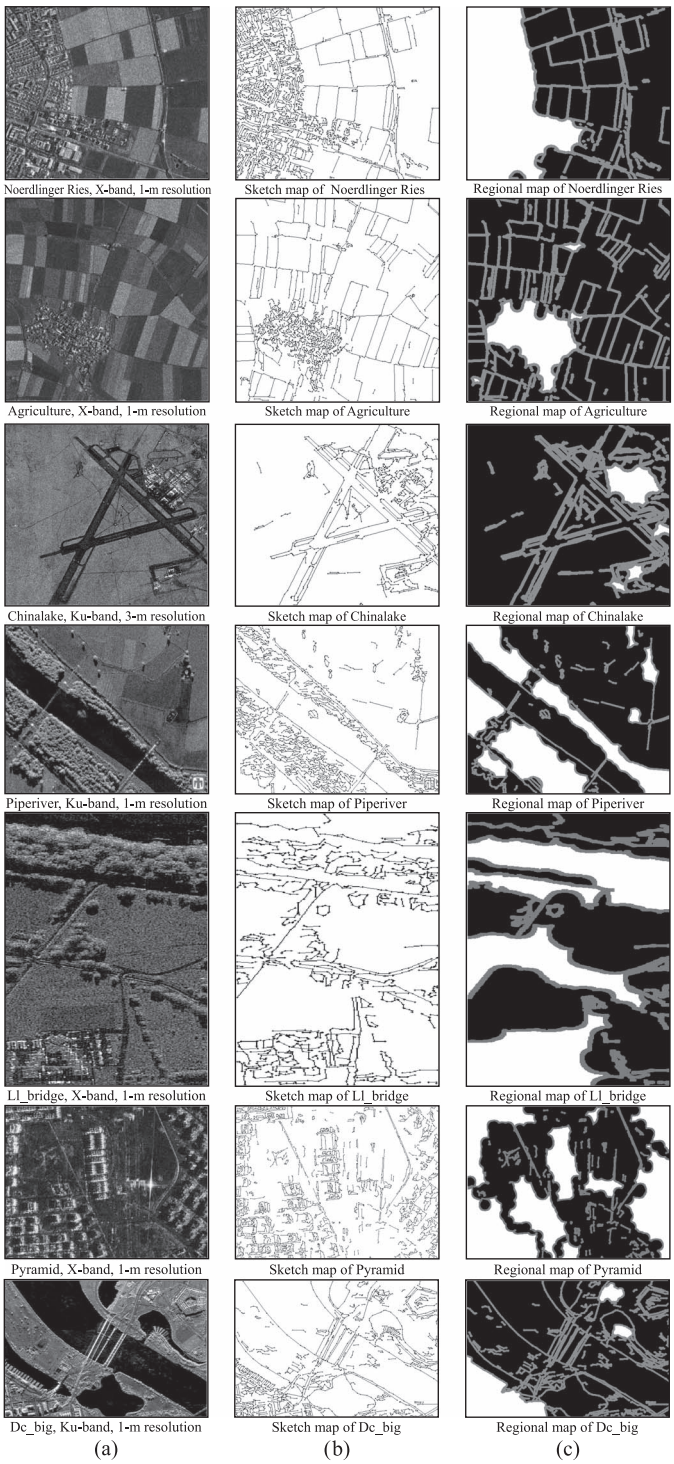

Fig. 2. (a) Original real SAR images. (b) Sketch maps. (c) Regional maps, which include aggregated region (in white), structural region (in gray), and homogeneous region (in black).

The aggregated region mainly includes urban areas, forests, and so on. The structural region mainly includes edges, line objects, and so on. The homogenous region mainly includes water, farmland, and so on.

III. Segmentation Based on HVS-ANML

In this section, considering the properties of different regions, the LLC-based hierarchical clustering and adaptive neighborhood multinomial latent model are used for adaptive segmentation.

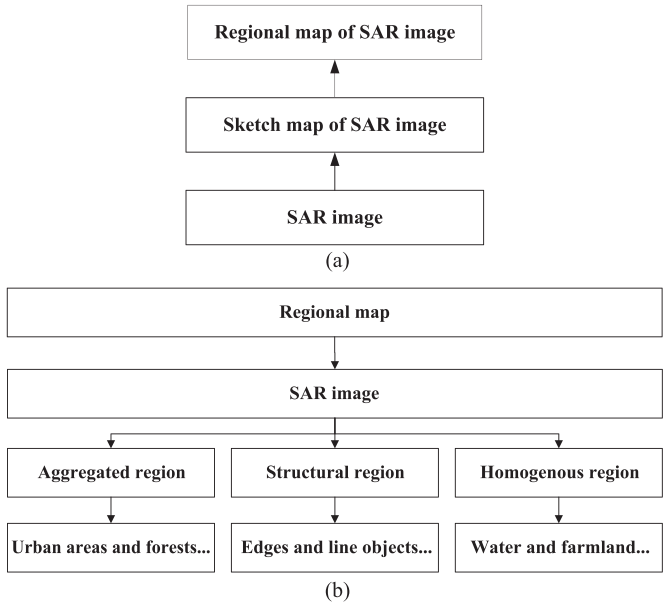

Fig. 3. (a) Hierarchical visual semantic. (b) Semantic representation of different regions in SAR images.

A. Segmentation of the Aggregated Region

The aggregated region is generated by the morphological close operator (see the Appendix), where the edges may not be accurate. Since the level-set method can represent the contours of the complex topology and is able to handle topological change, the edge-based level-set method [34] is used to refine the edges of the aggregated region. In the evolving process of the level-set method, the ratio operator [35] is used to compute the gradient of the original SAR image.

The edges of the aggregated region are located according to the aforementioned scheme. However, it is not known whether the several regions included in the aggregated portion belong to the same class. A well-chosen representation is helpful for the segmentation. Motivated by bag-of-words approaches [36], we use GLCM as the feature descriptors and the LLC method to obtain the final representation. Then, hierarchical clustering with this final presentation is used for unsupervised segmentation. In addition, from Fig. 2(c), we can see that the areas of some aggregated regions are very small. This leads to imbalanced classification [37]. Therefore, we consider the region whose area is less than 1% of the biggest region as a separate class.

B. Segmentation of the Structural Region and Homogeneous Region

In this part, for a better location of the line objects, the visual semantic rules are designed to extract the sketch lines of line objects. Moreover, an adaptive neighborhood-based multinomial latent model is constructed for the segmentation of the structural region and homogeneous region.

1) Visual Semantic Rules: In order to locate the line objects, we need to distinguish the edges and line objects in the

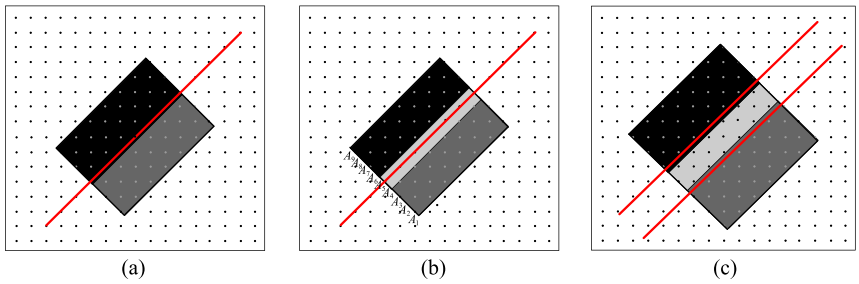

Fig. 4. Edge model and line model. (a) Edge model. (b) Line object is represented by the single sketch line. (c) Line object is represented by two sketch lines.

structural region. It is well known that line objects are modeled by two close and parallel edges. It means that the line model consists of three regions, while the edge model consists of two regions [38], [39]. In addition, from the sketch lines in the structural region, we find that the narrower line object is represented by a single sketch line and the wider line object is represented by two close and parallel sketch lines. These attributes can be described as follows: 1) the close and parallel sketch lines are the sketch lines of line objects, and 2) for the single sketch line, if there are three regions in its GSW, then this single sketch line is the sketch line of line objects. It means that the number of the amplitude change in the GSW is two. Moreover, orientations are very important for the edges and line objects, so their models are analyzed in GSW. Fig. 4 shows the edge model and line model.

It should be noted that one sketch line consists of several sketch line segments. In other words, one sketch line consists of several orientations. This makes the orientation of the sketch line hard to be determined. Considering this, the operators are implemented on the sketch line segments to constitute the sketch lines of the line objects. We assume that l_s is the sketch line segment, where $s \in \{1, 2, \ldots, S\}$ and S is the number of the sketch line segment. The Euclidean distance of l_i and l_j is D_{ij}. O_s is the orientation of the sketch line segment l_s. The distance threshold between two sketch line segments is T_1. In GSW, A_i is the average amplitude of the ith column. The average amplitude difference is defined as AD_i, where $\text{AD}_i = abs(A_i - A_{i+1})$. The label vector is written as $\mathbf{z}_s$, whose elements are $z_{\text{si}} \in \{0, 1\}$. If $\text{AD}_i > T_2$, $z_{\text{si}} = 1$; otherwise, $z_{\text{si}} = 0$, where T_2 is a threshold to measure the amplitude change. $z_{\text{si}} = 1$ represents that there is an amplitude change between the ith column and the $(i + 1)$th column. lsl is the set of the sketch lines of line objects. The rule is defined as follows.

Rule 1:
If $D_{\text{ij}} < T_1$ and $abs(O_i - O_j) < 10^\circ$, then $l_i, l_j \in lsl$;
If $\text{sum}(\mathbf{z}_s) = 2$, then $l_s \in lsl$.

Here, $\text{sum}(\cdot)$ is the sum of the elements in vector $\mathbf{z}_s$. T_1 is the distance threshold of two sketch line segments. The Euclidean distance D_{ij} is written as

$$D_{\text{ij}} = \sqrt{(lx_i - lx_j)^2 + (ly_i - ly_j)^2} \tag{1}$$

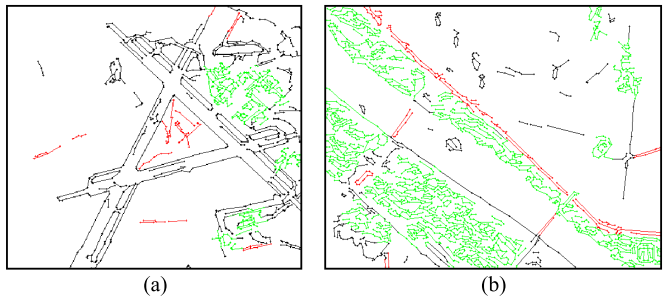

Fig. 5. Sketch lines of line objects (in red) and sketch lines (in green) in the aggregated region. (a) Sketch lines of line objects (in red) in Chinalake. (b) Sketch lines of line objects (in red) in Piperiver.

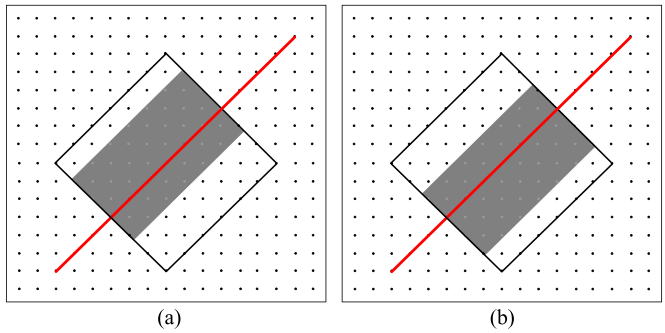

Fig. 6. SR (in gray) in GSW. (a) Most pixels in SR are located on the left of the sketch line. (b) Most pixels in SR are located on the right of the sketch line.

where (lx_i, ly_i) and (lx_j, ly_j) are the midpoint coordinates of l_i and l_j.

According to Rule 1, the sketch lines of line objects in the structural region are extracted. The sketch lines of line objects (in red) in Chinalake and Piperiver are shown in Fig. 5. We can see that the sketch lines of line objects are all in the structural region.

In addition, GSW is operated on the nonaggregated sketch lines to obtain the structural region. The size of GSW is set empirically, so the structural region needs to be refined. In GSW, the pixel with the larger edge-line feature [28] is the truth pixel of the structural region. Otherwise, it belongs to the homogeneous region. The refined structural region in GSW is shown in Fig. 6. We define that the pixel amplitude is y_n, $n \in \{1, 2, \ldots, N\}$ and N is the number of the pixels. G_n is the edge-line intensity [28] of y_n. SR is the refined structural region. HR is the homogeneous region. The rule is defined as follows.

TABLE I
ALGORITHM DESCRIPTION OF HVS-ANML

Algorithm SAR image segmentation based on hierarchical visual semantic and adaptive neighborhood multinomial latent model

Input: SAR image,

1. Extracting the sketch map of the input SAR image.
2. Extracting the regional map of the input SAR image.
3. The input SAR image is divided into aggregated, structural and homogeneous regions by the regional map.
4. For the aggregated region, the level set method is firstly used to refine the edges. Then LLC based hierarchical clustering is used to segment the aggregated region into c_1 classes.
5. For the structural and homogeneous regions:
   (1)The sketch lines of line objects are extracted in the structural region by using Rule 1,
   (2)Refined structural region SR and homogeneous region HR are obtained by using Rule 2,
   (3)For SR and HR, initialize the two regions into c_2 classes,
   (4)While the number of the label changes in SR and HR >0,do
      Estimate the parameters using the equations (23)-(29) in [24];
      If $y_n \in SR$, calculate the posterior probability by the equation (7) with $w_o(n)$;
      If $y_n \in HR$, calculate the posterior probability by the equation (7) with $w(n)_{max}$;
      Classifying the pixels according to the posterior probability;
   End,
   (5)For the close and parallel sketch lines of the line objects, the regions between them are labeled as the line objects; for the single sketch line of line objects, the elongated region covers this sketch line is labeled as the line object. All line objects are labeled as one separate class.
6. Perform the labeling from 1 to $c_1 + c_2 + 1$ on the whole image to obtain the final segmentation result.

Output: the labels of the SAR image.

Rule 2:
If $G_n > T$, then $y_n \in$ SR; if $G_n \leq T$, then $y_n \in$ HR, where T is obtained by Otsu's method.

2) Segmentation: Considering the properties of different regions, a multinomial latent model with GSW and an adaptive window selection are proposed for the segmentation of SR and HR, respectively.

Prior Probabilities: To estimate the prior probabilities on the obtained segmentation map, different neighborhoods are considered for SR and HR. For the edges and line objects in SR, a strong relationship exists along their directions rather than across them. Therefore, orientation-based GSW is used to capture the context in SR. Moreover, in order to preserve the details, the size of the window is usually small. For the homogeneous region, an adaptive window selection method is used to find the suitable window for each pixel. It makes the spatial relationship for each central pixel to be extracted only from the related pixels in the window. In order to find the suitable window, the standard deviation σ is used as the homogeneity measure of the window [40]. The variation of the standard deviation $\Delta\sigma$ is used to control the evolution of σ when the window size is increased. The process is initiated with a small window of 3×3 centered at the current pixel. At each step, the window size is increased by two rows and two columns (e.g., $5 \times 5, 7 \times 7, 9 \times 9, \ldots$) until $\Delta\sigma < T_3$. It is actually a process of window increasing. We assume that $w(n)_{\max}$ is the window centered at y_n after the window increasing. $Y = \{y_n | n \in N\}$ is the observed field, and $X = \{x_n | n \in N\}$ is the label field. The prior model is expressed as

$$p(x_n) = \frac{\exp\left(\eta\left(1 + \sum_{m \in W(n)} \delta_{x_m = k}\right)\right)}{\sum_{i=1}^{K} \exp\left(\eta\left(1 + \sum_{m \in W(n)} \delta_{x_m = i}\right)\right)} \tag{2}$$

where K is the number of classes. $\delta_{x_m=k} = 1$ if $x_m = k$, and $\delta_{x_m=k} = 0$ if $x_m \neq k$. η is the model parameter. $W(n)$ is the window defined around y_n. We define $W(n)$ as

$$W(n) = \begin{cases} w_o(n), & \text{if } y_n \in \text{SR} \\ w(n)_{\max}, & \text{if } y_n \in \text{HR} \end{cases} \tag{3}$$

where $w_o(n)$ is the GSW and $w(n)_{\max}$ is the square window after the window increasing.

Likelihood Probabilities: The likelihood model is regarded as the feature model. However, the single feature often cannot completely describe the SAR image. Therefore, it is necessary to combine different features to improve the performance of the

SAR image segmentation. In [24], the amplitude and texture features of SAR images are assembled into a finite mixture model. The model is written as

$$p(y_n|x_n;\theta_k) = p_A(y_n|x_n;\theta_k)p_T(\mathbf{y}_{\partial n}|\mathbf{x}_{\partial n};\theta_k) \quad (4)$$

where $\mathbf{y}_{\partial n}$ denotes the surrounding pixels of y_n and $\mathbf{x}_{\partial n}$ denotes the surrounding labels of x_n. $p_A(\cdot)$ is the amplitude density, and $p_T(\cdot)$ is the texture density. θ_k denotes the model parameters.

The amplitude density is described as the Nakagami distribution

$$p_A(y_n|x_n;r_k,u_k) = \frac{2}{\Gamma(u_k)}\left(\frac{u_k}{r_k}\right)^{u_k} y_n^{2u_k-1} e^{\left(-u_k\frac{y_n^2}{r_k}\right)} \quad (5)$$

and an autoregressive (AR) model [24] is used to describe the texture feature. The texture density is written as a t distribution such that

$$p_T(\mathbf{y}_{\partial n}|\mathbf{x}_{\partial n};\boldsymbol{\alpha}_k,\beta_k,\varphi_k) = \frac{\Gamma\left((1+\beta_k)/2\right)}{\Gamma(\beta_k/2)(\pi\beta_k\varphi_k)^{1/2}} \times \left[1+\frac{\left(y_n-\mathbf{y}_{\partial n}^T\boldsymbol{\alpha}_k\right)^2}{\beta_k\varphi_k}\right]^{-\frac{\beta_k+1}{2}} \quad (6)$$

where $\boldsymbol{\alpha}_k$ is the AR coefficient and r_k, u_k, β_k, and φ_k are the distribution parameters associated with a specified class.

Posterior Probabilities: The aim of the segmentation here is to estimate the labels X given a set of observations Y by maximizing the posterior probability. According to Bayesian's rule, the posterior probability is defined as

$$p(x_n|y_n,\theta_k) = p(y_n|x_n;\theta_k)\frac{\exp\left(\eta\left(1+\sum_{m\in W(n)}\delta_{x_m=k}\right)\right)}{\sum_{i=1}^{K}\exp\left(\eta\left(1+\sum_{m\in W(n)}\delta_{x_m=i}\right)\right)}. \quad (7)$$

MAP estimation is used to obtain the image labels. $\theta_k = \{r_k, u_k, \boldsymbol{\alpha}_k, \beta_k, \varphi_k\}$ are estimated by the expectation–maximization (EM) algorithm in [24]. Their analytic expressions are given in [24, eqs. (23)–(27)]. η is estimated by Newton's method. Its corresponding expression is given in [24, eq. (29)].

3) Line Object Location: The sketch lines of line objects (see Fig. 5) and the aforementioned segmentation result work together to locate the line objects. Based on the segmentation map, for two close and parallel sketch lines of line objects, the regions between them are labeled as the line objects; for the single sketch line of line objects, the elongated region covering this sketch line is labeled as the line object.

The main steps of our algorithm are shown in Table I.

The final segmentation result consists of three parts. The first is the segmentation result of the aggregated region. The number of the class in the aggregate region is adaptively determined by hierarchical clustering, which is an unsupervised clustering method. The second is the result of the structural and homogeneous regions. The number of the class in structural and homogeneous regions is set empirically according to the experimental results and the corresponding optical images. The third is the line objects, which are labeled as one class. Based on the hierarchical visual semantic (shown in Fig. 3), we can see that the three parts are uncrossed. It means that the classes in the aggregated region do not belong to the other two parts and vice versa. Therefore, we combine the results of the three parts together directly. We assume that the numbers of the class in the aggregated region, structural and homogeneous regions, and line objects are c_1, c_2, and 1, respectively. The final result consists of $c = c_1 + c_2 + 1$ classes. We perform the labeling from 1 to c on the whole image to obtain the final segmentation result.

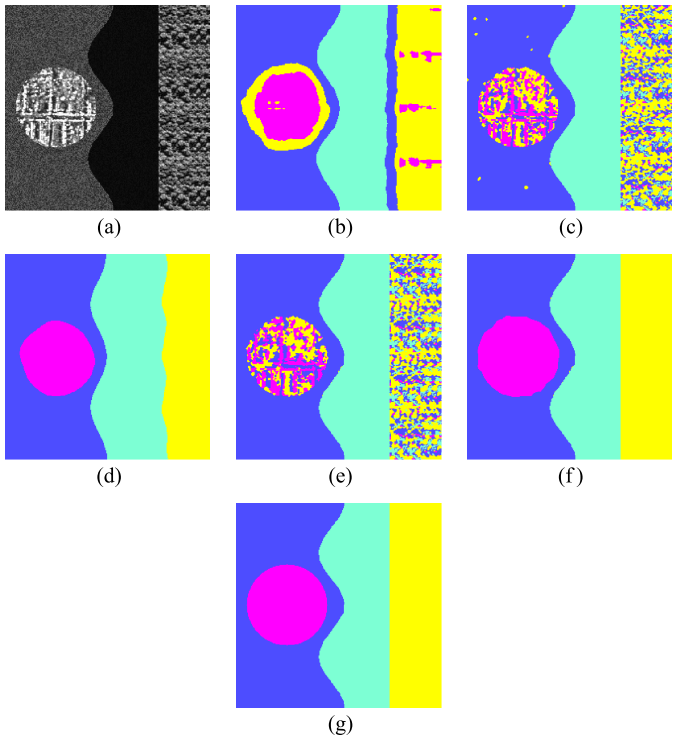

Fig. 7. Segmentation results of SYN1. (a) SYN1. (b) GLCM-FCM. (c) MRF. (d) ATML-CEM. (e) SM-ANML. (f) HVS-ANML. (g) Ground truth.

IV. Experiments and Analyses

In this section, we perform some experiments on both synthetic and real SAR images using our proposed method and some related segmentation approaches, including GLCM-FCM (fuzzy C-means) [7], MRF [17], ATML-CEM [24], and sketch map-based adaptive neighborhood multinomial latent model (SM-ANML). GLCM-FCM is a feature-based method. MRF, ATML-CEM, and SM-ANML are model-based methods. By comparing with the first three of them, it shows that the division of the SAR images is reasonable. The comparison with SM-ANML is used to prove that the hierarchical visual semantic is very important for the segmentation of the aggregated land covers.

These compared approaches and their corresponding parameters are described as follows. 1) GLCM-FCM: This approach is based on the GLCM features, and FCM is used for unsupervised segmentation. The GLCM features are extracted in a 13×13 window. 2) MRF: It is a statistics-based approach integrating the feature model and the spatial context model. MAP is used to obtain the labels of the image. 3) ATML-CEM: In

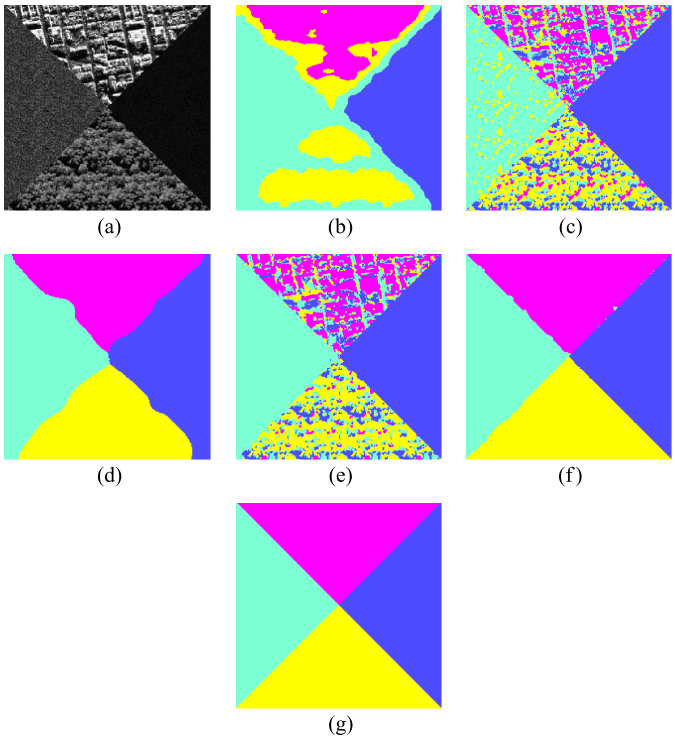

Fig. 8. Segmentation results of SYN2. (a) SNY2. (b) GLCM-FCM. (c) MRF. (d) ATML-CEM. (e) SM-ANML. (f) HVS-ANML. (g) Ground truth.

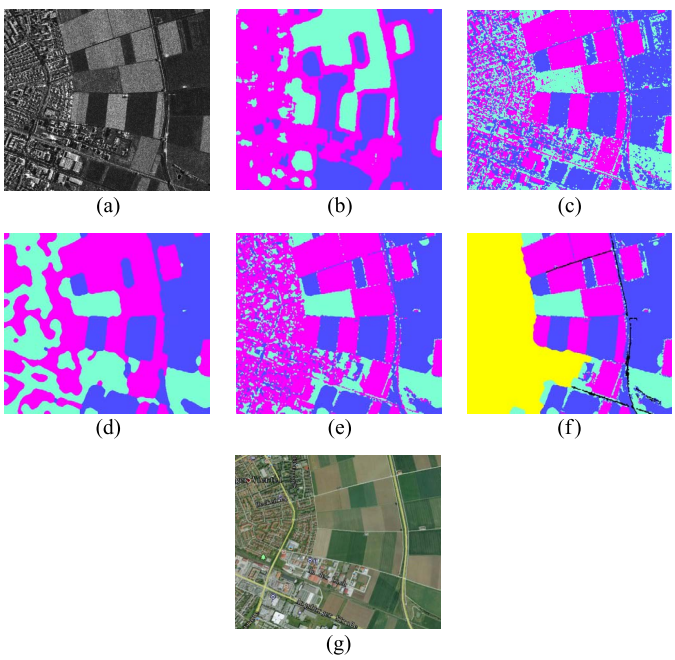

Fig. 9. Segmentation results of Noerdlinger Ries. (a) Noerdlinger Ries. (b) GLCM-FCM. (c) MRF. (d) ATML-CEM. (e) SM-ANML. (f) HVS-ANML. (g) Optical image.

the approach, the amplitude and texture features are assembled into a finite mixture model, and a multinomial logistic function is used to describe the spatial context model. The labels are obtained by the classification EM (CEM) [24] algorithm. The size of the window for the texture feature is selected to be 3×3. The window sizes for the spatial context models are selected to be 7×7 for SYN1, SYN2, Noerdlinger Ries, Piperiver,

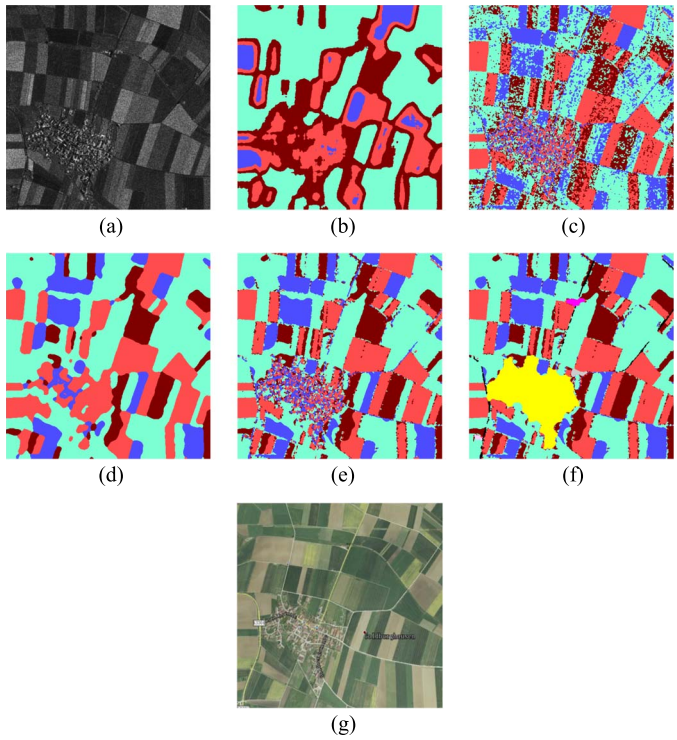

Fig. 10. Segmentation results of Agriculture. (a) Agriculture. (b) GLCM-FCM. (c) MRF. (d) ATML-CEM. (e) SM-ANML. (f) HVS-ANML. (g) Optical image.

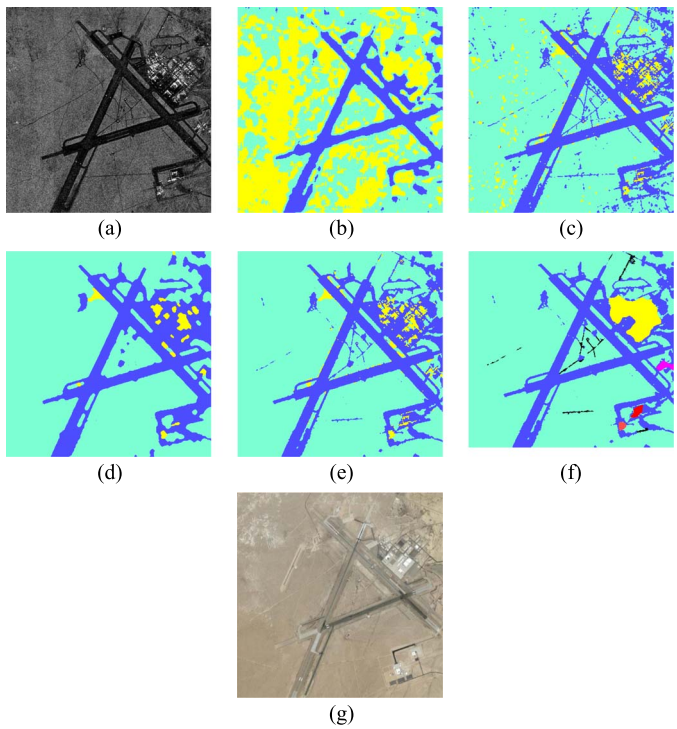

Fig. 11. Segmentation results of Chinalake. (a) Chinalake. (b) GLCM-FCM. (c) MRF. (d) ATML-CEM. (e) SM-ANML. (f) HVS-ANML. (g) Optical image.

Ll_bridge, Pyramid, and Dc_big. The window sizes for the spatial context models are selected to be 5×5 for Agriculture and Chinalake. 4) SM-ANML: It is one different version of our approach. The sketch map is used to guide the SAR image segmentation. This method is named as SM-ANML. The size of GSW is 3×3.

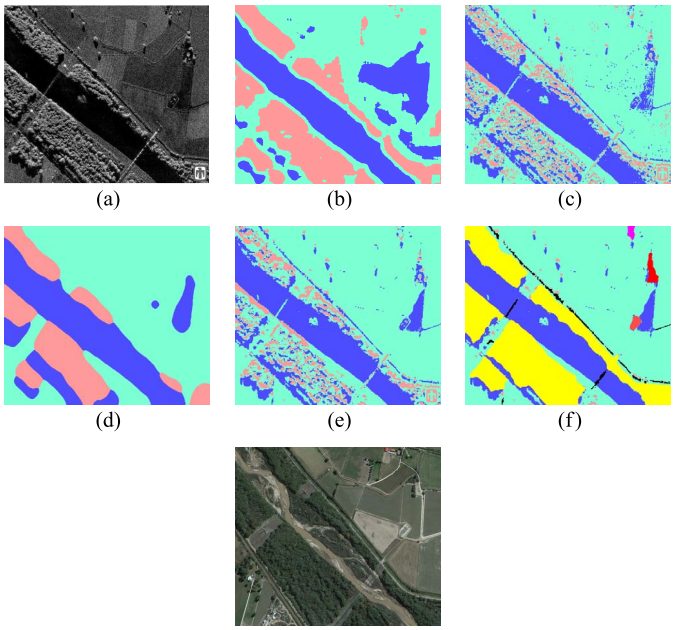

Fig. 12. Segmentation results of Piperiver. (a) Piperiver. (b) GLCM-FCM. (c) MRF. (d) ATML-CEM. (e) SM-ANML. (f) HVS-ANML. (g) Optical image.

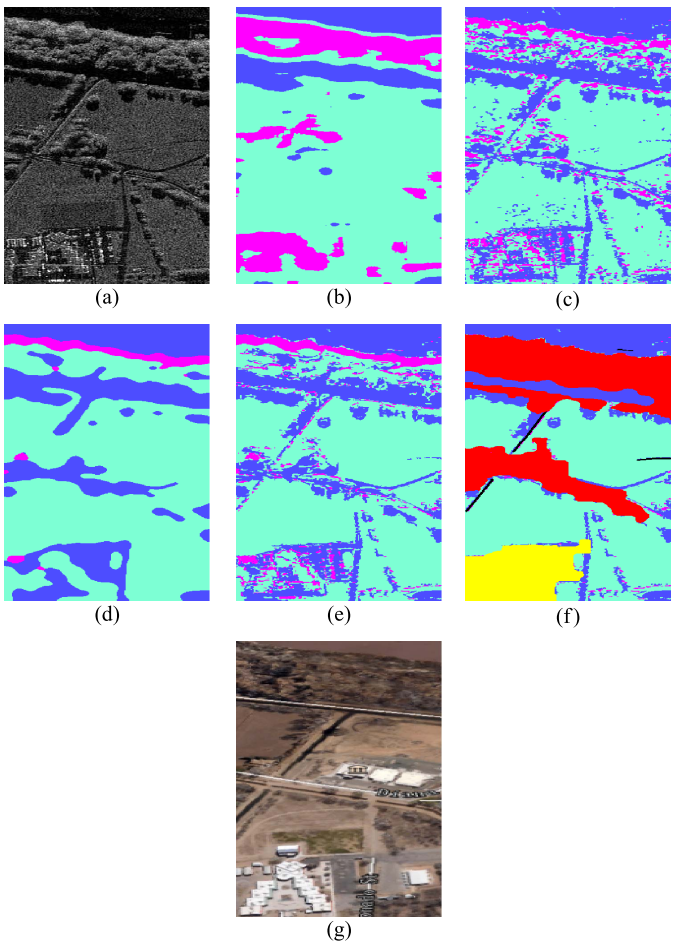

Fig. 13. Segmentation results of Ll_bridge. (a) Ll_bridge. (b) GLCM-FCM. (c) MRF. (d) ATML-CEM. (e) SM-ANML. (f) HVS-ANML. (g) Optical image.

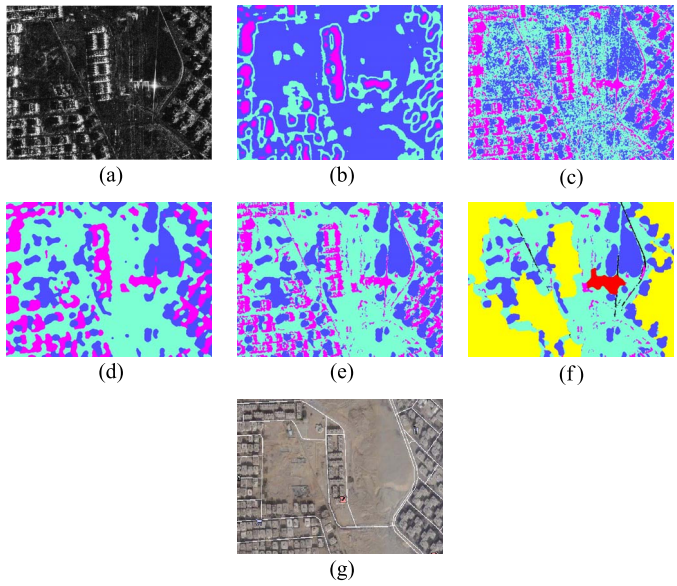

Fig. 14. Segmentation results of Pyramid. (a) Pyramid. (b) GLCM-FCM. (c) MRF. (d) ATML-CEM. (e) SM-ANML. (f) HVS-ANML. (g) Optical image.

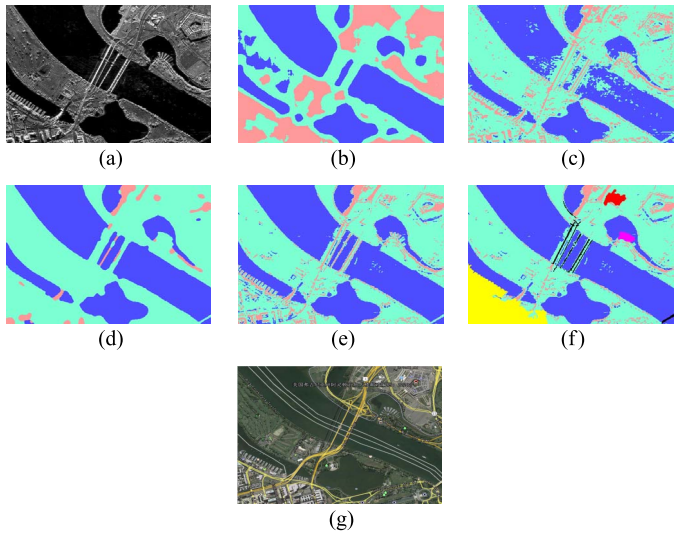

Fig. 15. Segmentation results of Dc_big. (a) Dc_big. (b) GLCM-FCM. (c) MRF. (d) ATML-CEM. (e) SM-ANML. (f) HVS-ANML. (g) Optical image.

The parameters of HVS-ANML are set as follows. The size of GSW is 3×3. T_1 and T_2 are used in the process of the line object location, and they are computed only on the real SAR images. Due to the different resolution and image content, T_1 is estimated for each image and set as 10, 11, 5 16, 10, 8, and 6 for Noerdlinger Ries, Agriculture, Chinalake, Piperiver, Ll_bridge, Pyramid, and Dc_big, respectively. T_2 is used to measure the amplitude change in GSW and set as 34 for all real SAR images. T_3 is used to measure the homogeneity of the window and set as 3. The detailed analyses of these parameters are given in the subsequent section.

A. *Test SAR Images*

We test these approaches on two synthetic SAR images and seven real SAR images. The synthetic SAR image SYN1 [shown in Fig. 7(a)] is constituted by four parts, including

TABLE II
PRIOR KNOWLEDGE OF THE SAR IMAGES

Images	Size	Resolution	Band	Polarization	Modalities	Location	Sensor
SYN1	256×256	-	-	-	-	-	-
SYN2	256×256	-	-		-	-	-
Noerdlinger Ries	500×440	1m	X	HH	High resolution spotlight mode	Swabian Jura	TerraSAR
Agriculture	512×512	1m	X	HH	High resolution spotlight mode	Swabian Jura	TerraSAR
Chinalake	440×420	3m	Ku	-	Stripmap	California	Airborne SAR
Piperiver	600×432	1m	Ku	-	Stripmap	New Mexico	Airborne SAR
Ll_bridge	210×520	1m	X	-	Stripmap	New Mexico	Airborne SAR
Pyramid	700×525	1m	X	HH	High resolution spotlight mode	Giza	TerraSAR
Dc_big	550×370	1m	Ku	-	Stripmap	Washington D.C	Airborne SAR

TABLE III
ACCURACY (IN PERCENT) AND KAPPA OF HVS-ANML, SM-ANML, ATML-CEM, MRF, AND GLCM-FCM FOR SYN1

Methods	Water	Land	Urban	Forest	Average	Kappa
HVS-ANML	**100**	99.13	**99.58**	**99.98**	**99.67**	**0.9951**
SM-ANML	**100**	**99.78**	38.61	45.29	70.81	0.6963
ATML-CEM	98.89	96.58	84.04	87.99	91.87	0.9102
MRF	**100**	98.94	36.09	45.82	70.21	0.6899
GLCM-FCM	94.96	91.17	70.13	81.34	84.40	0.8222

TABLE IV
ACCURACY (IN PERCENT) AND KAPPA OF HVS-ANML, SM-ANML, ATML-CEM, MRF, AND GLCM-FCM FOR SYN2

Methods	Water	Land	Urban	Forest	Average	Kappa
HVS-ANML	**100**	99.87	**99.04**	**99.46**	**99.59**	**0.9946**
SM-ANML	**100**	99.69	52.49	53.81	76.50	0.6867
ATML-CEM	99.99	**100**	92.45	91.8	96.06	0.9474
MRF	**100**	89.34	44.94	48.29	70.64	0.6086
GLCM-FCM	95.15	94.89	65.56	58.53	78.53	0.7135

water, land, urban area, and forest, which are from the real SAR images. The contents of the four parts in SYN2 [shown in Fig. 8(a)] are the same as that in SYN1, while the scale of the texture in SYN2 is larger than that of SYN1 and the shapes of edges in SYN2 are also different from that of SYN1. The seven real SAR images used are Noerdlinger Ries [shown in Fig. 9(a)], Agriculture [shown in Fig. 10(a)], Chinalake [shown in Fig. 11(a)], Piperiver [shown in Fig. 12(a)], Ll_bridge [shown in Fig. 13(a)], Pyramid [shown in Fig. 14(a)], and Dc_big [shown in Fig. 15(a)]. The line objects are only analyzed in the real SAR images. We also download the corresponding optical images [shown in Figs. 9(g)–15(g)] from Google Earth for comparison. The prior information of the used SAR images is shown in Table II. "–" represents that the information is not provided. More information about these images can be found from the website http://www.sandia.gov/radar/imagery/index.html.

B. Segmentation Results of the Synthetic SAR Images

We produce two synthetic SAR images SYN1 and SYN2 to test the performance of our method. Due to the availability of the ground truth, two quantitative indexes, including accuracy and Kappa coefficient, are used for evaluation. The higher accuracy means a better performance of the method. A higher Kappa coefficient implies a better consistency in the segmentation. The visual results of SYN1 are shown in Fig. 7, and the numerical indexes are calculated in Table III. The visual results of SYN2 are shown in Fig. 8, and the corresponding numerical indexes are shown in Table IV.

From the segmentation results of GLCM-FCM [shown in Figs. 7(b) and 8(b)], we can see that the edges are not well preserved. The main reason is that the GLCM is extracted in a predefined and larger window. For the results of MRF [shown in Figs. 7(c) and 8(c)], the edges are well preserved, but some oversegmentation occurs, particularly in the urban areas and forests. It is because the predefined spatial context model is not suitable for the aggregated scenes. In the ATML-CEM method [shown in Figs. 7(d) and 8(d)], both the amplitude and texture densities are considered in the likelihood model. The smooth constraints are included in the spatial context model. Therefore, the regions in ATML-CEM are more homogeneous than those in GLCM-FCM and MRF, while the edges are not accurate; this is because the window size of the spatial context model is set empirically. In the SM-ANML method [shown in Figs. 7(e) and 8(e)], the SAR image is divided into structural and nonstructural regions. The urban areas and forests are regarded as the structural portions to be preserved. This leads to the extremely poor consistency in the urban areas and forests. From the results of HVS-ANML [shown in Figs. 7(f) and 8(f)], it is noted that various classes such as water, lands, urban areas, and forests are identified clearly. Particularly in the urban areas and forests, consistent regions are obtained rather than the noisy results. Meanwhile, the edges are preserved clearly. This is because the SAR image is divided into regions with different characteristics by the hierarchical visual semantic. For each region, the adaptive method is chosen for the segmentation. The comparisons between Figs. 7(e) [8(e)] and Figs. 7(f) [8(f)] indicate that the hierarchical visual semantic is crucial to the consistencies of the urban areas and forests. Figs. 7(g) and 8(g) are the ground truths of the two synthetic SAR images. Among all the visual results in Figs. 7 and 8, Figs. 7(f) and 8(f) are the closest to the ground truths. From the above, we can see that no one single method performs perfectly in SAR image segmentation. Each method has its own benefits. Therefore, different characteristics should be considered in SAR image segmentation, and corresponding

segmentation methods should be adopted.

The numerical results of SYN1 and SYN2 are listed in Tables III and IV, respectively. In Table III, the accuracy of the water is high in most methods. HVS-ANML obtains the highest accuracies in the urban areas and forests, while the accuracy of the land in HVS-ANML is slightly lower than that of SM-ANML. In Table IV, the accuracies of the water, urban area, and forest are the highest with HVS-ANML, while the accuracy of the land is slightly lower than that of ATML-CEM. However, from the average accuracy, we can see that HVS-ANML is obviously better than SM-ANML, ATML-CEM, MRF, and GLCM-FCM. The Kappa coefficients are given in the last columns of Table III and IV. The highest Kappa coefficient is obtained by HVS-ANML. From the above, it demonstrates that a better performance in terms of the consistencies of the regions and the preservations of the edges is obtained by HVS-ANML.

C. Segmentation Results of the Real SAR Images

In this section, seven real SAR images are used for a further analysis. The compared methods are the same as those used on the synthetic SAR images. The visual results are shown in Figs. 9–15. Since the ground truths of the real SAR images are hard to obtain, the corresponding optical images of the same scene from Google Earth are used for comparison.

From the visual results in Figs. 9–15, it is noted that the segmentation results of HVS-ANML are different from that of the compared methods. The number of the class in HVS-ANML is larger than that of the other methods. This is because the segmentation result of HVS-ANML consists of three parts, which are the aggregated region, the structural and homogeneous regions, and the line objects. In Fig. 9(f), the aggregated region is labeled as one class (in yellow). The structural and homogeneous regions are segmented into three classes (in blue, red, and light green). The line objects are labeled as one class (in black). The final segmentation result includes five classes. This segmentation map is closer to the optical image [shown in Fig. 9(g)]. Therefore, it is more reasonable. The results of the compared methods include three classes. In Fig. 10(f), the four portions in the aggregated region are imbalanced, which are labeled as four different classes. The structural and homogeneous regions are segmented into four classes. The line objects are labeled as one class. The final segmentation map includes nine classes. The results of the compared methods include four classes. In Fig. 11(f), the final segmentation result of HVS-ANML includes seven classes. The compared methods include three classes. In Fig. 12(f), the final result of HVS-ANML includes eight classes. The compared methods include three classes. In Fig. 13(f), the final result of HVS-ANML includes six classes. The compared methods include three classes. In Fig. 14(f), the final result of HVS-ANML includes six classes. The compared methods include three classes. In Fig. 15(f), the final result of HVS-ANML includes seven classes. The compared methods include three classes.

The results will be analyzed further on. In the results of GLCM-FCM [shown in Figs. 9(b)–15(b)], edges and line objects are not well preserved. Obviously, the bridges are lost in Figs. 12(b) and 15(b). This is because the window size of GLCM is larger than the width of the bridges. The edges and line objects are not well preserved in ATML-CEM either. An oversmooth spatial context model causes the edge generalization and the loss of the line objects. Particularly in Figs. 9(d) and 12(d), almost no details exist in the segmentation results. The edges are well kept in the results of Figs. 9(c)–15(c) by the MRF method. From the results of SM-ANML [shown in Figs. 9(e)–15(e)], we can see that the edges are also well preserved, while the consistencies in the urban areas and forests are not satisfied. It is because the different semantics of the sketch lines in different regions are not analyzed. In our results [shown in Figs. 9(f)–14(f)], the water, lands, urban areas, forests, and line objects are segmented clearly. The consistencies of the urban areas and forests are much better than that of the other methods. The reason is that the aggregated regions are obtained by the hierarchical visual semantic. With the adaptive neighborhood, the consistencies of the homogeneous regions are well kept, and the edges are accurately located simultaneously. The visual semantic rules are designed by analyzing the structure of line objects, so the line objects (in black) are located only in our method. The optical images are found [shown in Figs. 9(g)–14(g)] from Google Earth for comparison. By comparing the results with the optical images, it is obvious that the results of VSH-ANML are more correct. In Fig. 15(f), three aggregated regions are segmented clearly. It is noteworthy that the ports and the urban area are labeled into one aggregated region (in yellow). That is because the aggregated degree of the urban area and the port are similar and the urban area is close to the port. In order to solve this problem, the shape and orientation of the sketch lines can be used to distinguish them. For example, the port has the linear-type structure. The structure is extended only in one direction. The urban area has the sphere-type structure. The structure is extended in every direction. We will explore these characteristics in our future work.

D. Parameter Analysis and Discussion

Parameter Analysis of T_1 and T_2: Parameters T_1 and T_2 are used to judge whether the sketch line belongs to the sketch lines of line objects. There are two types of sketch lines to represent the line objects. One is two close and parallel sketch lines. T_1 is the distance threshold of the two sketch line segments. The other is the single sketch line. There are two amplitude changes in the GSW of the single sketch line. T_2 is used to measure the amplitude change.

T_1 is the distance threshold of the two sketch line segments. Due to the different resolution and image content, T_1 is estimated for each image according to the experimental results. We take Chinalake as an example to explain the effect of the parameter T_1. From Fig. 16, we can see that a small value of T_1 will lose some sketch line of line objects. A large value of T_1 will introduce some false sketch lines of line objects. According to the original SAR image Chinalake, the selection of T_1 as 5 keeps the true sketch lines of line objects and avoids the false sketch lines of the line object. Therefore, T_1 is selected as 5 for Chinalake. With the same strategy, T_1 is chosen as 10, 11,

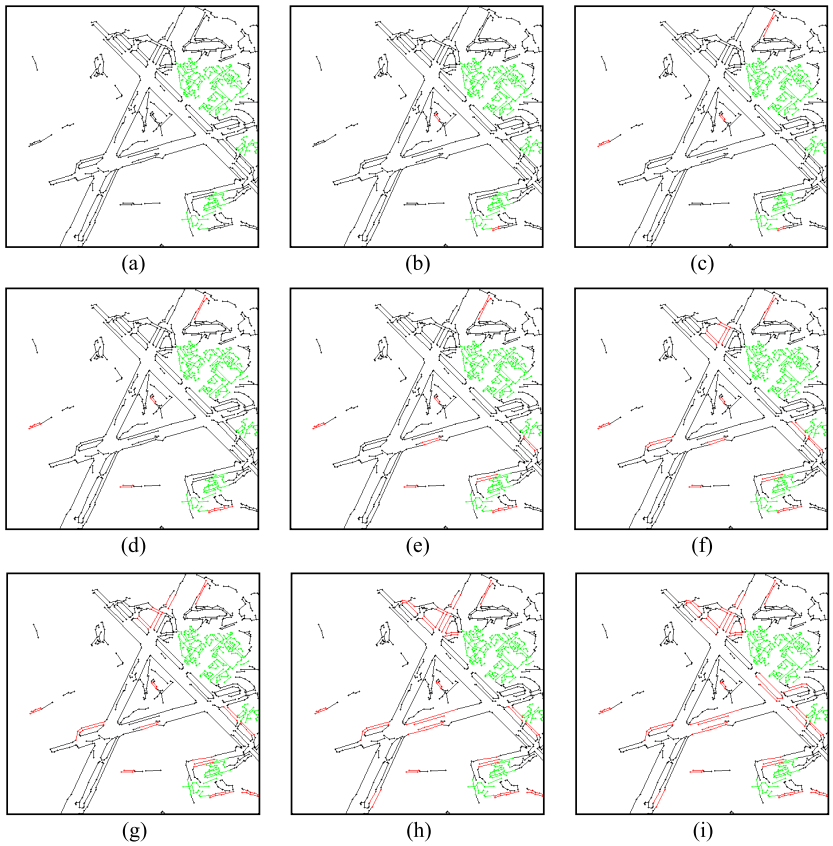

Fig. 16. Different values of T_1 on Chinalake. Sketch lines of line objects (in red) and sketch lines (in green) in aggregated region. (a) $T_1 = 2$. (b) $T_1 = 3$. (c) $T_1 = 4$. (d) $T_1 = 5$. (e) $T_1 = 6$. (f) $T_1 = 7$. (g) $T_1 = 8$. (h) $T_1 = 9$. (i) $T_1 = 10$.

16, 10, 8, and 6 for Noerdlinger Ries, Agriculture, Piperiver, Ll_bridge, Pyramid, and Dc_big, respectively.

T_2 is used to measure the amplitude change in GSW. A large value of T_2 will lose some sketch lines of line objects; on the contrary, a small value of T_2 will misclassify some edges as line objects. Therefore, a too big or too small value of parameter T_2 will lead to the unsatisfactory results. In order to explain how to set parameter T_2, we use different T_2 to extract the sketch lines of line objects in Chinalake (shown in Fig. 17). From the results in Fig. 17, we find that some sketch lines indicating the true line objects are lost by using a too large threshold. According to the original SAR image Chinalake, the selection of T_2 as 34 keeps the true sketch lines of line objects. Therefore, T_2 is selected as 34. All experiments are conducted under this selected value.

According to the threshold T_1, the close and parallel sketch lines indicating the line objects are obtained. By using the threshold T_2, the single sketch line indicating the line objects is obtained. The two parts are integrated together to obtain the sketch lines of line objects.

Parameter Analysis of T_3: T_3 is used to measure the homogeneity of the window. In order to make a reasonable selection of T_3, SYN1 and SYN2 are used in the experiments. The corresponding curves that represent the variation of the average accuracy and Kappa coefficient to different T_3 are given in Fig. 18. As we have observed, T_3 has a significant effect on the segmentation results at the beginning. The peaks of these curves are observed at $T_3 = 3$; after that, the average accuracy and the Kappa coefficient change slowly. Therefore, T_3 is set to be 3 in our approach.

V. Conclusion

In this paper, we have proposed a SAR image segmentation method based on HVS-ANML. Using the hierarchical visual semantic, the SAR image is divided into aggregated, structural, and homogeneous regions. Considering the different characteristics, the multinomial latent model with the adaptive neighborhood is proposed for segmentation. Moreover, the visual semantic rules are designed for a better segmentation of the line objects. The contribution of this paper lies in three aspects. First, the hierarchical visual semantic of SAR images is proposed by analyzing the semantic of the sketch map. Second, the SAR image is divided into regions with different characteristics, and the methods considering these characteristics are proposed for segmentation. Third, the visual semantic rules are derived for the line object location. Experiments on the synthetic and real SAR images show that the proposed method obtains better consistencies of the regions and preserves more details.

However, some parameters are set by the interactive operation in our method. More adaptive parameter selection methods will be explored by considering the prior information, such as the statistical and structural characteristics. In addition, the semantic is important for SAR image segmentation. More semantics based on computer vision will be explored in our future work.

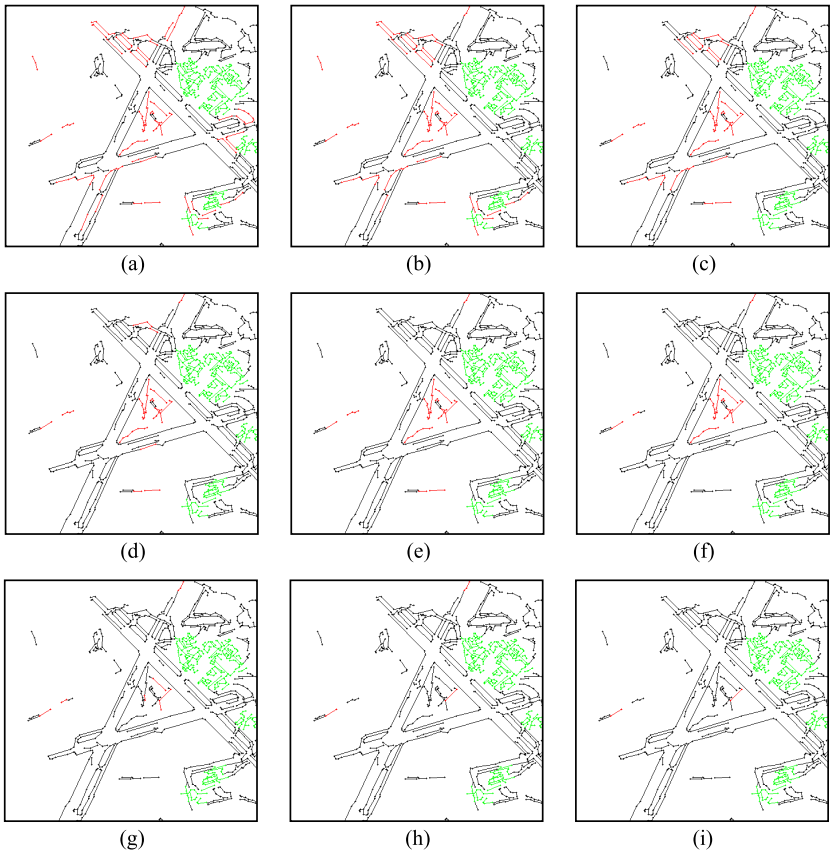

Fig. 17. Different values of T_2 on Chinalake. Sketch lines of line objects (in red) and sketch lines (in green) in aggregated region. (a) $T_2 = 18$. (b) $T_2 = 22$. (c) $T_2 = 26$. (d) $T_2 = 30$. (e) $T_2 = 34$. (f) $T_2 = 38$. (g) $T_2 = 42$. (h) $T_2 = 46$. (i) $T_2 = 50$.

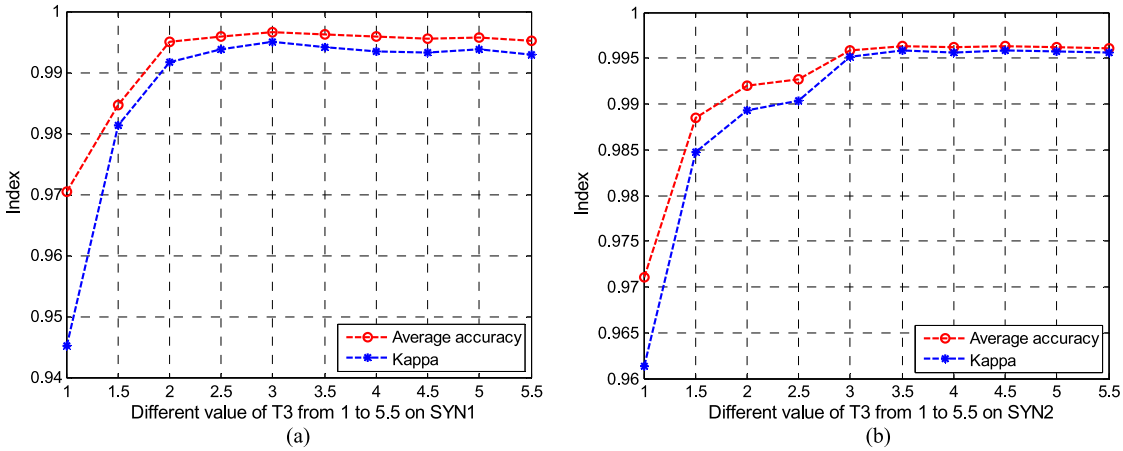

Fig. 18. Experimental indexes with different parameter values of T_3. (a) Experimental indexes with different values of T_3 on SYN1. (b) Experimental indexes with different values of T_3 on SYN2.

APPENDIX
AGGREGATED REGION EXTRACTION

The aggregation degree of the sketch line segment l_s is defined as the average distance between l_s and its v-nearest neighbors. The aggregation degree is given as

$$\text{aggregation}(l_s) = \frac{1}{v}\sum_{i=1}^{v} \text{sort}(D_{\text{si}}),\ i = 1, 2, \ldots, S \quad (8)$$

where $\text{aggregation}(l_s)$ is the aggregated degree of l_s and $\text{sort}(\cdot)$ denotes a function that arranges the vector elements in a non-decreasing order. v is a parameter. The distance between the sketch line segments D_{si} is defined in (7).

The aggregation degree is used to classify the sketch line segments by a threshold T_4, where T_4 is selected according to the histogram of the aggregated degree. The sketch line segments are classified into the aggregated sketch line segments when its aggregated degree is less than or equal to T_4. Otherwise, the sketch line segments are classified into the nonaggregated sketch line segments. The aggregated sketch line segments constitute the aggregated sketch lines A, and the nonaggregated sketch line segments constitute the nonaggregated sketch lines

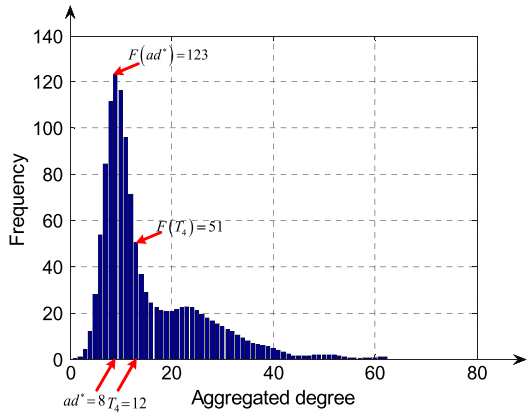

Fig. 19. Histogram of the aggregation degree.

Fig. 20. (a) Aggregated sketch lines (in blue) and nonaggregated sketch lines (in red). (b) Aggregated region (in gray).

NA. The morphological close operator is implemented on the aggregated sketch lines A to obtain the corresponding aggregated region.

Taking the sketch map of Agriculture as an example, the histogram of the aggregation degree is shown when $v = 7$ in Fig. 19. The horizontal axis represents the aggregated degree, which is represented by ad. The vertical axis represents the number of the sketch line segments, which is represented by $F(ad)$. The highest peak of the histogram is expressed as $F(ad^*)_{\max}$, and its corresponding value of horizontal axis is ad^*. In order to guarantee the aggregated sketch line segments to be preserved, T_4 is larger than the aggregated degree ad^* and usually set as the aggregated degree whose frequency is near the half of the peak $F(ad^*)_{\max}$. Some fine adjustments may be needed for a specified SAR image. In this example, T_4 is selected as 12. According to the threshold $T_4 = 12$, the sketch lines are classified into aggregated sketch lines and nonaggregated sketch lines. In Fig. 20(a), the blue sketch lines are the aggregated sketch lines, and the red sketch lines are the nonaggregated sketch lines. Then, the morphological close operator is implemented on the aggregated sketch lines to obtain the corresponding aggregated region. In Fig. 20(b), the gray region is the aggregated region.

The main steps of the aggregated region extraction are given as follows:

1) Extracting the sketch map of the input SAR image;
2) Computing the aggregation degree of every sketch line segment by using (8);
3) Classifying the sketch lines into A and NA by T_4;
4) The morphological close operator is implemented on A to generate the aggregated region.

REFERENCES

[1] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*, 3rd ed. Norwood, MA, USA: Artech House, 2004.
[2] L. C. Jiao, X. Tang, B. Hou, and S. Wang, "SAR images retrieval based on semantic classification and region-based similarity measure for earth observation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 8, pp. 3876–3891, Aug. 2015.
[3] A. Voisin, V. A. Krylov, G. Moser, S. B. Serpico, and J. Zerubia, "Classification of very high resolution SAR images of urban areas using copulas and texture in a hierarchical Markov random field model," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 1, pp. 96–100, Jan. 2013.
[4] L. K. Soh and C. Tsatsoulis, "Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 2, pp. 780–795, Mar. 1999.
[5] H. Yu, X. R. Zhang, S. Wang, and B. Hou, "Context-based hierarchical unequal merging for SAR image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 995–1009, Feb. 2013.
[6] F. Dellinger, J. Delon, Y. Gousseau, J. Michel, and F. Tupin, "SAR-SIFT: A SIFT-like algorithm for SAR images," in *Proc. IEEE Int. Geosci. Sens. Symp.*, Munich, Germany, Jul. 2012, pp. 3478–3481.
[7] J. Feng, L. C. Jiao, X. R. Zhang, M. G. Gong, and T. Sun, "Robust non-local fuzzy c-means algorithm with edge reservation for SAR image segmentation," *Signal Process.*, vol. 93, no. 2, pp. 487–499, Feb. 2013.
[8] S. C. Johnson, "Hierarchical clustering schemes," *Psychometrika*, vol. 32, no. 3, pp. 241–254, Sep. 1967.
[9] X. R. Zhang, L. C. Jiao, F. Liu, L. F. Bo, and M. G. Gong, "Spectral clustering ensemble applied to SAR image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2126–2136, Jul. 2008.
[10] L. C. Jiao, J. Feng, F. Liu, T. Sun, and X. R. Zhang, "Semisupervised affinity propagation based on normalized trivariable mutual information for hyperspectral band selection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2760–2773, Jun. 2015.
[11] L. C. Jiao, F. H. Shang, F. Wang, and Y. Y. Liu, "Fast semi-supervised clustering with enhanced spectral embedding," *Pattern Recognit.*, vol. 45, no. 12, pp. 4358–4369, Dec. 2012.
[12] L. C. Jiao *et al.*, "Natural and remote sensing image segmentation using memetic computing," *IEEE Comput. Intell. Mag.*, vol. 5, no. 2, pp. 78–91, May 2010.
[13] O. Germain and P. Refregier, "Edge location in SAR images: Performance of the likelihood ratio filter and accuracy improvement with an active contour approach," *IEEE Trans. Image Process.*, vol. 10, no. 1, pp. 72–78, Jan. 2001.
[14] R. C. P. Marques, F. N. Medeiros, and J. S. Nobre, "SAR image segmentation based on level set approach and G model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 2046–2057, Oct. 2012.
[15] J. L. Feng, Z. J. Cao, and Y. M. Pi, "Multiphase SAR image segmentation with G-statistical-model-based active contours," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 4190–4199, Jul. 2013.
[16] S. Z. Li, *Markov Random Field Modeling in Image Analysis*. London, U.K.: Springer, 2001.
[17] H. Derin and H. Elliott, "Modeling and segmentation of noisy and textured images using Gibbs random fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 9, no. 1, pp. 39–55, Jan. 1987.
[18] Q. Y. Yu and D. A. Clausi, "IRGS: Image segmentation using edge penalties and region growing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 12, pp. 2126–2139, Dec. 2008.
[19] C. Tison, J. M. Nicolas, F. Tupin, and H. Maitre, "A new statistical model for Markovian classification of urban areas in high-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vo1. 42, no. 10, pp. 2046–2057, Oct. 2004.
[20] J. Li, A. Najmi, and R. M. Gray, "Image classification by a two-dimensional hidden Markov model," *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 517–533, Feb. 2000.
[21] D. E. Melas and S. P. Wilson, "Double Markov random fields and Bayesian image segmentation," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 357–365, Feb. 2002.
[22] F. Galland *et al.*, "Unsupervised synthetic aperture radar image segmentation using Fisher distribution," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2966–2972, Aug. 2009.
[23] P. Zhang *et al.*, "Unsupervised multi-class segmentation of SAR images using fuzzy triplet Markov fields model," *Pattern Recognit.*, vol. 45, no. 11, pp. 4018–4033, Nov. 2012.

[24] K. Kayabol and J. Zerubia, "Unsupervised amplitude and texture classification of SAR images with multinomial latent model," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 561–572, Feb. 2013.
[25] O. Aytekin, M. Koc, and I. Ulusoy, "Local primitive pattern for the classification of SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2431–2441, Apr. 2013.
[26] N. Rasiwasia, P. J. Moreno, and N. Vasconcelos, "Bridge the gap: Query by semantic example," *IEEE Trans. Multimedia*, vol. 9, no. 5, pp. 923–938, Aug. 2007.
[27] C. E. Guo, S. C. Zhu, and Y. N. Wu, "Primal sketch: Integrating texture and structure," *Comput. Vis. Image Understand.*, vol. 106, no. 1, pp. 5–19, Apr. 2007.
[28] J. Wu *et al.*, "Local maximal homogeneous region search for SAR speckle reduction with sketch-based geometrical kernel function," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 1–14, Sep. 2014.
[29] C. Shi *et al.*, "Learning interpolation via regional map for pan-sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3417–3431, Jun. 2015.
[30] J. J. Wang *et al.*, "Locality-constrained linear coding for image classification," in *Proc. IEEE CVPR*, 2010, pp. 3360–3367.
[31] D. Marr, *Vision*. San Francisco, CA, USA: Freeman, 1982.
[32] F. Liu *et al.*, "Polarimetric hierarchical semantic model and scattering mechanism based PolSAR image classification," *Pattern Recognit.*, Mar. 2016, in press.
[33] F. Liu *et al.*, "Nonconvex compressed sensing by nature-inspired optimization algorithms," *IEEE Trans. Cybern.*, vol. 45, no. 5, pp. 1028–1039, May 2015.
[34] C. M. Li, C. Y. Xu, C. F. Gui, and M. D. Fox, "Distance regularized level set evolution and its application to image segmentation," *IEEE Trans. Image Process.*, vol. 19, no. 12, pp. 3243–3254, Dec. 2010.
[35] C. J. Oliver, D. Blacknell, and R. G. White, "Optimum edge detection in SAR," in *Proc. Inst. Elect. Eng.—Radar, Sonar Navig.*, vol. 143, no. 1, pp. 31–40, Feb. 1996.
[36] Y. Z. Huang, Z. F. Wu, L. Wang, and T. N. Tan, "Feature coding in image classification: A comprehensive study," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 3, pp. 493–506, Mar. 2014.
[37] G. M. Weiss and F. Provost, "Learning when training data are costly: The effect of class distribution on tree induction," *J. Artif. Intell. Res.*, vol. 19, pp. 315–354, 2003.
[38] O. Hellwich and H. Mayer, "Extracting line features from synthetic aperture radar (SAR) scenes using a Markov random field model," in *Proc. Int. Conf. Image Process.*, Lausanne, Switzerland, Sep. 1996, pp. 883–886.
[39] F. Tupin, H. Maitri, J. F. Mangin, J. M. Nicolas, and E. Pechersky, "Detection of linear features in SAR images: Applications to road network extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 2, pp. 434–453, Mar. 1998.
[40] J. M. Nicolas, F. Tupin, and H. Maitre, "Smoothing speckled SAR images by using maximum homogeneous region filters: An improved approach," in *Proc. IEEE Int. Geosci. Sens. Symp.*, Sydney, NSW, Australian, Jul. 2001, pp. 1503–1505.

Fang Liu (SM'07) received the B.S. degree in computer science and technology from the Xi'an Jiaotong University, Xi'an, China, in 1984 and the M.S. degree in computer science and technology from the Xidian University, Xi'an, in 1995.

She is currently a Professor with Xidian University. She has authored or coauthored five books and over 80 papers in journals and conferences. Her current research interests include image perception and pattern recognition, machine learning, evolutionary computation, and data mining. She won the second prize of the National Natural Science Award in 2013.

Yiping Duan received the B.S. degree from the School of Computer Science and Technology, Henan Normal University, Xinxiang, China, in 2010. She is currently working toward the Ph.D. degree in the School of Computer Science and Technology, Xidian University, Xi'an, China.

Her current research interests include semantic mining, machine learning, and synthetic aperture radar image processing.

Lingling Li received the B.S. degree from the School of Electronic Engineering, Xidian University, Xi'an, China, in 2011, where she is currently working toward the Ph.D. degree.

Her current research interests include community detection in networks and multiobjective optimization.

Licheng Jiao (SM'89) received the B.S. degree from Shanghai Jiaotong University, Shanghai, China, in 1982 and the M.S. and Ph.D. degrees from Xi'an Jiaotong University, Xi'an, China, in 1984 and 1990, respectively.

He is currently a Distinguished Professor with the School of Electronic Engineering, Xidian University, Xi'an. He has led approximately 40 important scientific research projects and has published over 10 monographs and 100 papers in international journals and conferences. He is the author of the three books *Theory of Neural Network Systems* (Xidian University Press, 1990), *Theory and Application on Nonlinear Transformation Functions* (Xidian University Press, 1992), and *Applications and Implementations of Neural Networks* (Xidian University Press, 1996). He has authored or coauthored over 150 scientific papers. His current research interests include signal and image processing, natural computation, and intelligent information processing.

Prof. Jiao is a member of the IEEE Xi'an Section Executive Committee and the Chairman of the Awards and Recognition Committee. He is an Executive Committee Member of the Chinese Association of Artificial Intelligence. He was the recipient of the second prize of the National Natural Science Award in 2013.

Jie Wu received the B.S. degree from Xidian University, Xi'an, China, in 2008, and the Ph.D. degree in computer science and technology from Xidian University and the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, Xi'an, in 2015.

He is currently a Lecturer with the School of Computer Science, Shaanxi Normal University, Xi'an. His current research interest is synthetic-aperture-radar image processing, including despeckling, segmentation, and change detection.

Shuyuan Yang (SM'15) received the B.S. degree in electrical engineering and the M.S. and Ph.D. degrees in circuit and system from Xidian University, Xi'an, China, in 2000, 2003, and 2005, respectively.

She is currently a Professor with the School of Electronic Engineering, Xidian University. Her research interests include compress sensing, machine learning, and intelligent information processing.

Xiangrong Zhang received the B.S. and M.S. degrees from the School of Computer Science, Xidian University, Xi'an, China, in 1999 and 2003, respectively, and the Ph.D. degree from the School of Electronic Engineering, Xidian University, in 2006.

She is currently a Professor with the Key Laboratory ofIntelligentPerception and ImageUnderstanding of the Ministry of Education, Xidian University. Her research interests include pattern recognition, machine learning, and image analysis and understanding.

Jialin Yuan received the B.S. and M.S. degrees in computer science and technology from Xidian University, Xi'an, China, in 2010 and 2013, respectively.

She is currently with Kiwi-image Technologies Company, Xidian University, working as a Video Algorithm Engineer for high-end TV sets. Her main research interests include synthetic-aperture-radar image processing, 3-D visualization, and motion estimation motion compensation.