

Received September 1, 2017, accepted November 16, 2017, date of publication November 24, 2017, date of current version February 14, 2018.

Digital Object Identifier 10.1109/ACCESS.2017.2777444

Visual Saliency Modeling for River Detection in High-Resolution SAR Imagery

FEI GAO¹, FEI MA¹, JUN WANG¹, JINPING SUN¹, ERFU YANG², AND HUIYU ZHOU³

¹School of Electronic and Information Engineering, Beihang University, Beijing 100191, China

²Space Mechatronic Systems Technology Laboratory, Department of Design, Manufacture and Engineering Management, Strathclyde Space Institute, University of Strathclyde, Glasgow G1 1XJ, U.K.

³School of Electronic, Electrical Engineering and Computer Science, Queen's University Belfast, Belfast BT3 9DT, U.K.

Corresponding author: Fei Ma (mafeimf@buaa.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61771027, Grant 61071139, Grant 61471019, Grant 61501011, and Grant 61171122, in part by the Aeronautical Science Foundation of China under Grant 20142051022, in part by the Pre-Research Project under Grant 9140A07040515HK01009. The work of E. Yang was supported by the RSE-NNSFC Joint Project (2017–2019) with the China University of Petroleum (Huadong) under Grant 6161101383. The work of H. Zhou was supported in part by the U.K. EPSRC under Grant EP/N508664/1, Grant EP/R007187/1, and Grant EP/N011074/1, and in part by the Royal Society-Newton Advanced Fellowship under Grant NA160342.

ABSTRACT Accurate detection of rivers plays a significant role in water conservancy construction and ecological protection, where airborne synthetic aperture radar (SAR) data have already become one of the main sources. However, extracting river information from radar data efficiently and accurately still remains an open problem. The existing methods for detecting rivers are typically based on rivers' edges, which are easily mixed with those of artificial buildings or farmland. In addition, pixel-based image processing approaches cannot meet the requirement of real-time processing. Inspired by the feature integration and target recognition capabilities of biological vision systems, in this paper, we present a hierarchical method for automated detection of river networks in the high-resolution SAR data using biologically visual saliency modeling. For effective saliency detection, the original image is first over-segmented into a set of primitive superpixels. A visual feature set is designed to extract a regional feature histogram, which is then quantized based on the optimal parameters learned from the labeled SAR images. Afterward, three saliency measurements based on the specificity of the rivers in the SAR images are proposed to generate a single layer saliency map, i.e., local region contrast, boundary connectivity, and edge density. Finally, by exploiting belief propagation, we propose a multi-layer saliency fusion approach to derive a high-quality saliency map. Extensive experimental results on three airborne SAR image data sets with the ground truth demonstrate that the proposed saliency model consistently outperforms the existing saliency target detection models.

INDEX TERMS Synthetic aperture radar (SAR), remote sensing, rivers, object detection, biological system modeling.

I. INTRODUCTION

River networks have great significance to water resource monitoring, flood assessment, wetlands' dynamics monitoring, and ship navigation. Compared with visible and infrared light sensor systems, SAR can achieve all-day and all-weather processing for geographical objects and therefore becomes one of the popular means to extract river information. Especially, the latest technology advance in SAR imaging sensors, such as Terra SAR-X and COSMO-SkyMed, has improved the imaging resolution greatly, providing more

accurate spatial information and more detailed features for the detection of rivers [1]–[3].

At present, one class of the river detection approaches is to employ thresholding, blob tracking, Bayesian network [50] or Wavelet edge detection [51] to extract shorelines and then conduct target outlining with the help of a Snake algorithm. For example, Yamada et al. [4] combined the morphological methods with a traditional thresholding method to detect the spatial information of rivers. Sun et al. [5] used the edge extraction and a baseline tracking approach to detect

river networks. Thresholding and dynamic contour modeling were applied for water area extraction in the ERS-2 SAR images by Hallikainen *et al.* [6]. Since texture and other high-level features were ignored, these methods were reluctant to deal with the interference of complex background.

Other methods mainly use semi-supervised or supervised classification algorithms to detect rivers. For example, Klemenjak *et al.* [7] proposed to exploit adaptive features and Support Vector Machine (SVM) to detect small rivers. Tian *et al.* [8] combined the corner features with SVM for river detection. However, these classification algorithms require a large amount of labeled samples for learning and training, which limits their applications in practice due to the significant time consumption in manually labeling samples.

As a computationally efficient signal processing system, biological visual systems have significant advantages in target search, feature extraction and integration applications. Taking the human visual systems (HVS) as an example, the retina and visual cortex located in a human brain can handle about 10^8 - 10^9 bit data per second. In recent years, the researchers have attempted to employ the biological vision theories for the SAR image interpretation to improve efficiency and precision. Inspired by the achievements in cognitive psychology, neurobiology, biology and computer science, some biological models have been recently developed and applied in SAR image interpretation successfully, e.g. [52].

Focus of Attention (FOA) is an important regulatory mechanism in the biological vision systems, which refers to the ability that the mammals selectively spend more computing resources on the regions of interest (ROI) or targets in the scenes. Several models have been proposed to simulate the FOA mechanism, e.g., fixation prediction models (FP). This class of models mainly utilize the feature integration theories and guided the search models to predict the regions that people look at (free viewing of natural scenes usually takes 3–5 seconds). They aim to obtain a saliency map where the pixels with higher saliency are more likely to be the fixations. Examples include Itti's model [11] proposed in 1997 and later Spectral Residual (SR) [12], Graph-Based Visual Saliency (GBVS) [13], Saliency Using Natural Statistics (SUN) [14], and Attention-based on Information Maximization (AIM) [15]. Inspired by these developments, Zhang *et al.* [9] utilized FOA to improve the detection efficiency on the remote sensing images. Similarly, in order to obtain more representative patches for scene classification, Zhang *et al.* [10] designed an unsupervised feature selection strategy based on the FP models, which achieved better classification performance than random selection schemes. Although these fixation prediction models can effectively improve the detection efficiency, they have larger prediction errors, e.g., up to 30 pixels [53].

For more accurate region detection, the pluralities of Saliency Object Detection models (SOD) [53] have been proposed since 2007. Such models are theoretically based on the Gestalt perception theory, which states that the

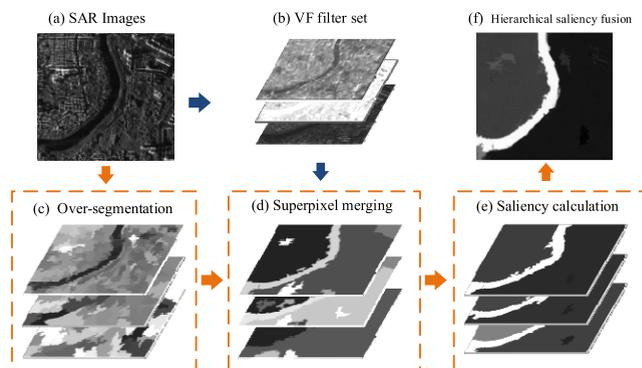


FIGURE 1. An overview of the proposed method. (a) A SAR image. (b) Six different features are extracted from the input SAR image. (c) Several image layers are hierarchically constructed from the input image by over-segmenting the image. (d) Some superpixels from each of the layers are merged. (e) The single-layer saliency maps are calculated using the three saliency cues. (f) The final saliency map is obtained by integrating all the single-layer saliency maps.

primate's visual systems divide the scenes into many small sub-regional units before conducting object recognition, searching or scene understanding [54]. Therefore, the SOD models generally consist of two steps. In the first step, images are divided into irregularly shaped sub-regions (superpixels) using Simple Linear Iterative Clustering (SLIC) [16], Mean-Shift [17], or watershed [18] methods. In the second step, a full-resolution saliency map is obtained using conditional random field [19], random forests [20], random walk [21], cellular automata [22], absorbing Markov chain [23], sparse reconstruction [24], or Bayesian networks [25].

The SOD models lead to new ideas for river detection in SAR images. However, currently SOD models are mainly used to process optical images, which are quite different from SAR images. Firstly, SAR images accompany strong speckle noise due to the coherence between radar echo signals. Secondly, SAR images are in grayscale and carry less information. Moreover, the diversified terrain structures and artificial buildings in SAR images greatly increase the difficulty of detecting targets. Hence, the application of SOD models in the SAR images faces intractable problems.

In this paper, we propose a new bottom-up SOD method to detect rivers in high-resolution SAR images (an overview of the proposed method is shown in Fig. 1). SAR images are firstly over-segmented to superpixels. Then the superpixels at the same layers are merged via feature-based distance. Afterwards, single-layer saliency maps are derived so that we can generate the final full-resolution saliency map. The novel aspects of our proposed algorithm consist of the following aspects.

- 1) A Simple-Complex Cell (SCC) filter set, inspired by observations in the primary visual cortex, is developed to extract features from each superpixel.
- 2) We propose a quantized parameter learning method for merging superpixels.
- 3) Three saliency cues are designed to construct signal-layer saliency maps considering the specialty of the rivers in the SAR images.

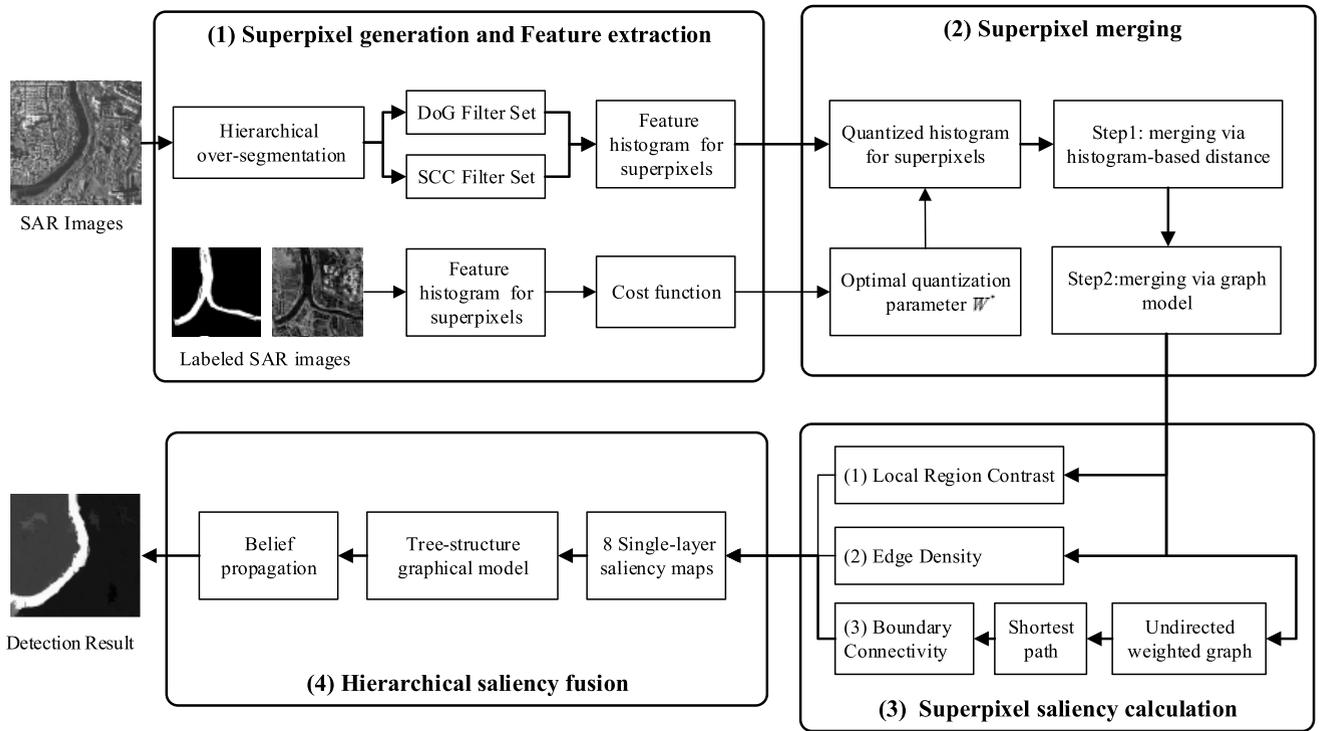


FIGURE 2. The flowchart of the proposed method.

4) We introduce a new saliency fusion algorithm via tree-structured graphical modeling and belief propagation to derive optimal full-resolutions saliency maps.

The rest of this paper is organized as follows. Section II presents the framework of the proposed method. In Section III, we introduce the principle and constitution of the visual feature (VF) filter set. In Section IV, we describe the process of hierarchical over-segmentation and superpixel merging based on the quantized histograms. Section V shows the strategy of calculating and integrating single-layer saliency maps. Our experiments and the results are presented in Section VI. Section VII concludes this paper with discussion.

II. FRAMEWORK OF THE PROPOSED METHOD

The proposed river detection method is illustrated in Fig. 2. This method consists of four major steps. In the superpixel generation and feature extraction step, SAR images are hierarchically over-segmented to superpixels. Then, a VF filter set, composing of a DoG filter set and a SCC filter set, is applied to extracting feature histograms from the superpixels.

In order to improve the computational efficiency of the saliency extraction, feature histograms are quantized in order to reduce their dimensions in the superpixel merging section. The corresponding optimal quantization parameters are learned from the labeled SAR samples via minimizing a cost function that we design here. Then a merging mechanism is introduced to reduce the number of the superpixels.

This merging strategy is performed at two steps: merging via histogram-based modeling and merging via graph modeling.

In the superpixel saliency extraction phase, three saliency cues based on the specificity of the rivers in the SAR images are proposed to produce a single layer saliency map, i.e., Local Region Contrast (LRC) [46], Boundary Connectivity (BC) [45], and Edge Density (ED). Among them, BC is a measure to quantify how confident a superpixel is connected to the image boundaries.

The last step of the process is saliency fusion, where we consider the hierarchical saliency maps as a tree-structured graphical model and we design an energy function to obtain an optimal full-resolution saliency map exploiting a belief propagation algorithm.

III. FEATURE EXTRACTION USING VF FILTER SET

We here design a VF filter set to extract features based on the research outcome on the receptive fields in the primary visual cortex. This filter set contains the following two different filter sets: DoG filter set (two filters) and SCC filter Set (three filters).

A. DoG FILTER SET

The DoG filter set originates from the “receptive fields” [26], which are specific regions receiving a light stimulation in the retina’s internal ganglion cells. The receptive fields can result in the reduction of the number of signals. Their functions can be approximated as the difference of two Gaussian

TABLE 1. The parameters of three SCC filters.

SCC filters	Filter size S	Effective width σ	Wave-Length λ	Spatial pooling grid $N_s \times N_s$	Overlaps ΔS
SCC 1	5×5	2	2.5	6×6	3
	7×7	2.8	3.5		
SCC 2	9×9	3.7	4.6	8×8	4
	11×11	4.6	5.7		
SCC 3	13×13	5.5	6.8	10×10	5
	15×15	6.4	8		

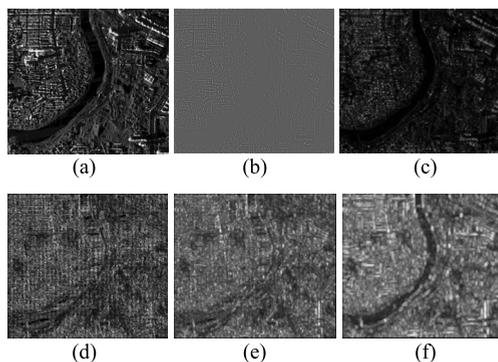


FIGURE 3. An airborne SAR image and the filtering results of the VF filter set. (a) The input SAR image. (b)–(c) The results of the two DoG filters. (d)–(f) The results of the three SCC filters.

functions [27], which is defined as

$$f(x; \mu, \sigma_1, \sigma_2) = \frac{1}{\sigma_1 \sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma_1^2}\right) - \frac{1}{\sigma_2 \sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma_2^2}\right) \quad (1)$$

where μ, σ_1, σ_2 respectively represent the mean and variances of the two Gaussian functions; x represents a pixel. As the imaging height of the airborne SAR images changes over time, the extracted features are scale dependent. Inspired by the approach reported in [32], we select two DoG filters with different parameters, where $\sigma_1 = 0.5, 1$ and $\sigma_2 = 2\sigma_1$. Fig. 3 (a) illustrates an airborne SAR image with a river in the urban area and the corresponding results of the two DoG filters are shown in Fig.3 (b)–(c). As can be seen, the edges of the SAR image have been effectively enhanced.

B. SCC FILTER SET

According to Hubel and Wiesel’s theory [32], the cells within the primary visual cortex are divided into simple cells [32]–[35] and complex cells [36]–[38]. The simple cells have relatively smaller receptive fields [39], [40]. Its direction and frequency selectivity for stimulations can be modeled by a two-dimensional Gabor function [29], [30]. On the other

hand, the complex cells not only can sense edges or rod-like features but also have a good tolerance for the change of the feature’s size and location. Inspired by the hierarchical maximization model (HMAX) [31], we here propose three SCC filters to extract increasingly invariant features, which closely correspond to the simple and complex cells of the human visual cortex.

Firstly, Gabor filters are used to approximate the simple cells:

$$\begin{cases} g_{\lambda, \theta, \sigma, \varphi}(x, y) = \exp\left(-\frac{u^2 + \gamma v^2}{2\sigma^2}\right) \times \cos\left(\frac{2\pi}{\lambda}u + \varphi\right) \\ u = x \cos \theta + y \sin \theta \quad v = x \sin \theta - y \cos \theta \end{cases} \quad (2)$$

where $g_{\lambda, \theta, \sigma, \varphi}(x, y)$ denotes the filtering result of a pixel locating at (x, y) ; γ is a constant, called the spatial aspect ratio, which determines the ellipticity of the receptive field. The phase offset parameter φ determines the symmetry of $g_{\lambda, \theta, \sigma, \varphi}(x, y)$ with respect to the origin. The parameters λ, θ, σ respectively denote the wavelength, preferred orientation and the size of the filter. In order to ensure the sensitivity of the features with various scales and orientations, the parameters of the Gabor filters (i.e., λ, θ, σ and filter size s) should be reasonably defined. Similar to [31], we choose 24 (6 sizes \times 4 directions) Gabor filters to compose pyramid-shaped filter banks, which contain six filter sizes: $5 \times 5, 7 \times 7, \dots, 15 \times 15$; six wavelengths: 2.5, 3.5, 4.6, 5.7, 6.8; six effective width: 2, 2.8, 3.7, 4.6, 5.5, 6.4 and four directions: $0^\circ, 45^\circ, 90^\circ, -45^\circ$. In summary, these simple cells will be divided into 3 bands and each band is corresponding to a SCC filter, as shown in Table 1.

As shown in Fig. 4, the results of a Gabor filter in the S th SCC filter are expressed as $G_{l, \theta}^S$, where $l = [1, 2]$ represents the filter size and θ is its orientation. For example, $G_{1, \theta}^1$ indicates that the four filtered images from the four Gabor filters with a size of 5×5 belong to SCC filter 1.

We perform a pooling operation for the filtered images that simulate the complex cells. The pooling operation is a local maximum operation. A grid of size $N_s \times N_s$ is firstly used to scan the filtered images with the overlapping of ΔS . Only the strongest pixel in the grid is maintained. In the next stage,

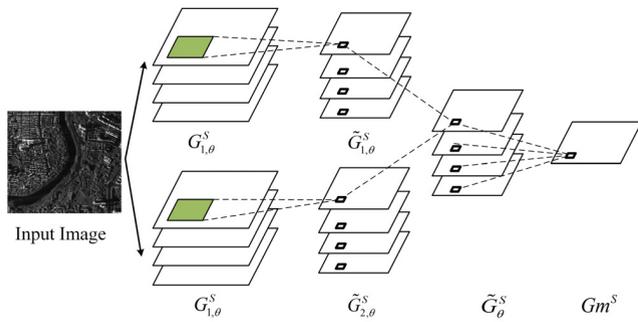


FIGURE 4. Architecture of S th SCC Filter. The input image is firstly filtered by 8 Gabor filters (4 orientations \times 2 scales). Then a pooling operator is implemented to obtain 8 features maps $\tilde{G}_{l,\theta}^S$, followed by a maximization operation over different scales and orientations to sustain the most intense stimulation map Gm^S .

we take a maximization operator over the two sub-sampled images with the same orientation in the same SCC filter. For example, two sub-sampled maps $\tilde{G}_{1,45^\circ}^1$ and $\tilde{G}_{2,45^\circ}^1$ in the SCC filter 1 will be compressed into a new map \tilde{G}_θ^S after the pooling operation is performed. Finally, we obtain a feature map Gm^S retaining the maximum values over the pixels at the same position of 4 maps \tilde{G}_θ^S in the S th SCC filter. The whole pooling operation can be described as

$$Gm^S = \max_{l,\theta} \tilde{G}_{l,\theta}^S \quad (3)$$

To some extent, the pooling operation is consistent with the “winner-take-all (WTA)” paradigm in the visual cortex. In other words, the interaction between nerve cells ensures that intense stimulation can be kept within a certain area while the others fade away.

The SCC filters can eliminate redundant information and ensure computational stability by maximizations. As shown in Fig. 3 (d)–(f), the strong features in different scales can be reserved by the SCC filters with certain parameters. Together with the brightness features, six feature maps can be produced after the VF filtering.

IV. OVER-SEGMENT AND MERGE USING QUANTIZED HISTOGRAM

A. SUPERPIXEL GENERATION

Efficient graph-based image segmentation (EGBIS) is a graph-based segmentation method proposed in [42], which can preserve significant areas in an image and its computational complexity is $2N \log N$, meeting the efficiency requirement. We adopt the EGBIS to achieve a hierarchical over-segmentation. Specifically, the input images are segmented to superpixels with various sizes by changing the standard deviation of the Gaussian smoothing function used in the EGBIS. The i th layer of segmentation results is denoted as $L_i \{i = 1, \dots, N_{seg}\}$ and its j th superpixel is $R_j^i \in L_i \{j = 1, \dots, M_i\}$, where N_{seg} is the iteration and experimentally set $N_{seg} = 8$. M_i is the number of superpixels in L_i .

The reference [42] indicates that the standard deviation of the Gaussian function σ be set to 0.8. In this case, the Gaussian smoothing function will not produce any visible change to an image but helps remove the artifacts. Based on this suggestion, we experimentally set the eight variances of the Gaussian functions in our method, changing from 0.3 to 1 with an increment step of 0.1 to obtain good detection results.

B. QUANTIZED HISTOGRAM FOR SUPERPIXELS

The Scale-Invariant Feature Transform (SIFT) algorithm [28] has proven that the histogram-based descriptors have sufficient robustness against scale change and rotation. Thus, for irregularly shaped superpixels, histograms are promising descriptors. Traditional saliency target detection models often use a 1×256 dimensional gray histogram. In this way, the VF space contains 256^6 features, which are computationally expensive even for medium sized images. To reduce the number of features, we first quantize the k th feature channel to have w_k values. This can reduce the dimension of the histogram from 1×256^6 to $1 \times [w_1, w_2, \dots, w_8]$, which compose the quantization parameter vector $W = [w_1, \dots, w_6]$.

We design a cost function based on the labeled SAR images to obtain the optimal quantization parameter vector W^* . The cost function is defined as:

$$W^* = \arg \min_W \left\{ \exp \left(- \sum_{i=1}^{N_g} \left\| H^{B,i} - H^{O,i} \right\|_2 \right) + 0.1 \|W\|_1 \right\} \quad (4)$$

where N_g is the number of the labeled SAR images as the learning samples. The quantized histograms of the background regions and the river regions in the i th images are described as $H^{B,i}$ and $H^{O,i}$ respectively.

The cost function contains two parts: the discriminability constraint (the first term) and the computational complexity constraint (the second term) in (4). The first one reveals that a well quantized histograms vector is to ensure enough discriminability between the background and the rivers, where $\|\cdot\|_2$ denotes the L_2 norm. The second term is to ensure that the histogram dimension is small enough to maintain the efficiency of the detection.

For the simplicity, it is assumed that the 6 feature channels are independent of each other. Therefore, the above cost function can be transformed to:

$$w_k^* = \arg \min_{w_k} \left\{ \exp \left(- \sum_{i=1}^{N_g} \left\| H_k^{B,i} - H_k^{O,i} \right\|_2 \right) + 0.1 \|w_k\|_1 \right\} \quad (5)$$

where w_k^* is the optimal quantization parameter of the k th feature channel. $H_k^{B,i}$ and $H_k^{O,i}$ are the quantized histograms of the k th feature channel from the background and the river areas, respectively. The solution w_k^* can be derived by setting the derivative of the above function to be zero. Fig. 5 shows the change of the cost function shown in (13) as w_k increases

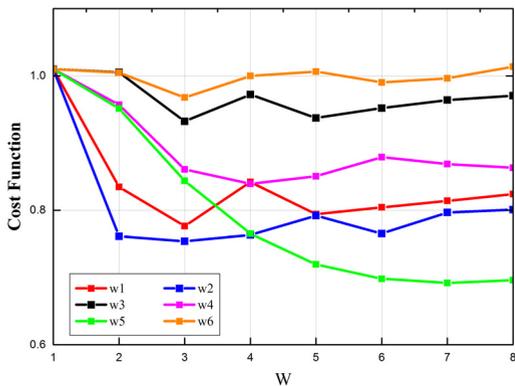


FIGURE 5. The cost function curve of 6 feature channels with different w_k , which reveals the optimal quantization parameter vector W . For example, the cost of w_1 is minimal when $w_1 = 3$. Hence, the optimal quantization of the first feature channel is $w_1^* = 3$.

from 1 to 8, whereby we can get the optimal quantization parameter vector $W = [3 \ 2 \ 3 \ 4 \ 7 \ 3]$.

C. SUPERPIXEL MERGING

In this section, we design a superpixel merging framework to conduct accurate river detection using histogram-based regional distance. The whole merging process can be achieved in the following two steps:

1) *Step 1:* For each region R_j^i , since the EGBIS algorithm does not strictly limit the size of superpixels, their scale value may vary greatly, measured by the number of pixels. A coarse superpixel may lead to less detailed histograms, and hence affect the detection accuracy. Therefore, we first merge these coarse superpixels with their nearest regions. Specifically, we sort all the superpixels according to their sizes in an ascending order. During each loop, each superpixel is merged to its most similar neighboring regions using a histogram-based matrix. For the computational simplicity, the distance between the regions $s(R_j^i, R_k^i)$ is defined as the Euclidean distance of the corresponding quantized histograms: $s(R_j^i, R_k^i) = \|H_j^i - H_k^i\|_2$, where H_j^i and H_k^i are the histograms of superpixels R_j^i and R_k^i at the L_i layer. Note that $s(R_j^i, R_k^i)$ is denoted as $s_{j,k}^i$ in the following section. After the merging, the sizes and histograms of the superpixels are updated. The loop will stop till the number of the remaining areas is less than the pre-defined number \tilde{N} . Comprehensive experiments show that generally better detection results can be obtained when \tilde{N} fall in the range of 30 and 80.

2) *Step 2:* Updating the quantized histogram and calculating the histogram-based regional distance are computationally expensive, and thus a new merging method is introduced in the second stage.

We define a graph $G^i = (V^i, E^i)$ of \tilde{N} nodes at the L_i layer, where the nodes V^i correspond to the superpixels and the undirected edges E^i are weighted by a distance matrix $S^i = [s_{j,k}^i]_{\tilde{N} \times \tilde{N}}$. In order to exploit the spatial relationship,

each node is only connected to those neighboring nodes. Next, the binarization operation is implemented on the graph. A small edge value $s_{j,k}^i$ means that the node pair (R_j^i, R_k^i) is likely to be the same target, where $s_{j,k}^i$ is reset to 1, otherwise $s_{j,k}^i$ is 0. With the constraints on the edges, it is clear that the constructed graph is a sparsely connected. That suggests that most the elements of the distance matrix S^i are zero. Therefore, all the linked nodes in the graph may be merged to a region whilst the corresponding quantized histogram is re-calculated.

Fig. 6 shows the merging results of $L_1 - L_4$ layers, where each region is illustrated by a different color. The number of the remaining areas at each layer generally drop to 10 after the second step, much simplifying the subsequent saliency computation. In addition, this merging process effectively helps separate the rivers and the background in most of the layers. Although the river areas are mixed with the background in some layers, such as L_1 , these areas still can be picked up in the following saliency detection phase.

V. SUPERPIXEL SALIENCY CALCULATION AND INTEGRATION

A. SINGLE-LAYER SALIENCY MAP

Existing saliency detection models are able to define the regions' uniqueness, but most of them can be directly used in the detection of rivers in SAR images. For instance, [43] strongly relies on the center-prior with the assumption that the salient targets are commonly located in the center of the optical images, whereas the positions of rivers in SAR images are stochastic in reality. Reference [19] shows the commonly employed spatial distribution prior is related to the rarity of the superpixel colors. Considering that SAR images have no color channels, the color-based saliency measures cannot be used in the detection of rivers from SAR images. In this paper, we define three measures to compute region saliency values based on the characteristics of rivers in SAR images.

1) LOCAL REGION CONTRAST (LRC)

Based on the observation that the human vision system is sensitive to contrast changes in the visual signals, we propose a Local Region Contrast method based on the regional quantized histograms to define the regional saliency values. Specifically, the saliency of a superpixel R_j^i at L_i layer is defined as the weighted sum of its histogram-based distance to n adjacent regions,

$$C_j^i = \sum_{k=1}^n w(R_k^i) \|H_j^i - H_k^i\|_2 \quad (6)$$

where H_j^i, H_k^i are the quantized histograms of superpixels R_j^i and R_k^i , respectively. The weight $w(R_k^i)$ is set to the area ratio of (R_k^i) and (R_j^i) . It means that the adjacent regions (R_k^i) with more pixels contribute more than those regions only containing a few pixels.

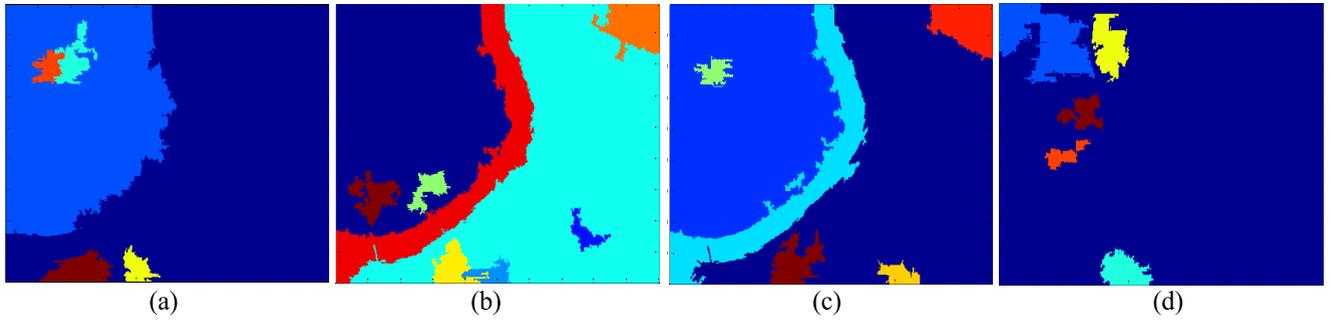


FIGURE 6. Superpixel merging results after steps 1 and 2 at four layers. (a) merging results at L_1 layer. (b) merging results at L_2 layer. (c) merging results at L_3 layer. (d) merging results at L_4 layer.

2) BOUNDARY CONNECTIVITY (BC)

We observe that rivers in SAR images generally link with borders. Inspired by the saliency optimization theory reported in [45], we propose a measure to quantify how confident a superpixel R_j^i is connected to specific image boundaries, e.g. rivers, called boundary connectivity (BC). BC is defined as

$$BC(R_j^i) = \sum_{k=1}^{\tilde{M}_i} w_{Spa}(R_j^i, R_k^i) \delta(R_k^i \in Bnd) \quad (7)$$

$BC(R_j^i)$ is computed as the summation of its spatial weights to all the boundary superpixels, where Bnd is the set of image boundary superpixels and \tilde{M}_i denotes the number of the superpixels at the L_i layer. $\delta(\cdot)$ is 1 for the superpixels on the image boundary and 0 otherwise. $w_{Spa}(R_j^i, R_k^i)$ is the spatial weight related to the centers of superpixels R_j^i and R_k^i , which is defined as follows:

$$w_{Spa} = \exp(-Spa(R_j^i, R_k^i)) \quad (8)$$

The geodesic distance between any two superpixels $Spa(R_j^i, R_k^i)$ is defined as the accumulated edge weights along their shortest path of an undirected weighted graph. The graph is constructed by connecting all the adjacent superpixels (R_j^i, R_k^i) and assigning their weight as the Euclidean distance between their quantized histograms. In this paper, we use the Dijkstra algorithm [44] to search for the shortest path between two nodes:

$$Spa(R_j^i, R_k^i) = \sum_{t=1}^{n-1} \min_{R_1^i=R_j^i, R_n^i=R_k^i} s(p_t, p_{t+1}) \quad (9)$$

where p_{t+1} is an adjacent region of p_t and n is the number of the nodes on the shortest path except R_j^i and R_k^i . According to (9), the geodesic distance is related to both spatial layout characteristics and VF features of the superpixels. Smaller geodesic distance means a higher similarity between two regions, leading to a larger spatial weight $w_{Spa}(R_j^i, R_k^i)$ and boundary connectivity $BC(R_k^i)$. In this case, R_j^i is more likely to be part of the river area.

3) EDGE DENSITY (ED)

Compared with the farmland, mountains and urban areas in the SAR imagery, river areas commonly remain unchanged in terms of their intensities. In other words, the density of edges is high in the areas around the rivers. The edge map is obtained by the SCC filters that we have introduced above. The ED of superpixel R_j^i at the L_i layer is computed as the density of edges:

$$ED(R_j^i) = (Gm_j^{1,i} + Gm_j^{2,i} + Gm_j^{3,i}) / w(R_j^i) \quad (10)$$

where $Gm_j^{1,i}$, $Gm_j^{2,i}$, $Gm_j^{3,i}$ respectively denote the sum of the filtered results by three SCC filters for superpixels R_j^i . $w(R_j^i)$ is used to count the number of the pixels in the region $w(R_j^i)$. A higher ED cue indicates a stronger possibility that the region is part of the river areas.

Fig.7 (a)–(c) shows that these three cue maps at the L_2 layer. These cues are complementary. The single-layer saliency map of the L_2 layer is defined as the product of the three cue maps.

$$f(R_j^i) = C_j^i \cdot BC(R_j^i) \cdot ED(R_j^i) \quad (11)$$

After having obtained initial saliency maps from different layers, as illustrated in Fig.7 (d), we propose an integration scheme to fuse them for the multi-scale saliency detection in the next section.

B. HIERARCHICAL SALIENCY FUSION

Cue maps reveal different saliency possibilities at different layers and none of the single layers can accurately reflect satisfactory spatial details. Also, it is hard to determine which layer leads to the best representation. Multi-layer fusion by naively averaging all the saliency maps cannot achieve the desired results [46], considering various terrain structures and complex background in SAR images. In this paper, we propose hierarchical saliency fusion rules based on a new energy function for achieving the best detection results.

For superpixel R_j^i , its final saliency $f(R_j^i)$ can be mapped to all the inner pixels. $f_{k,l}^i$ represents the saliency of the pixel with the coordinates (k, l) at the L_i layer. We design an energy

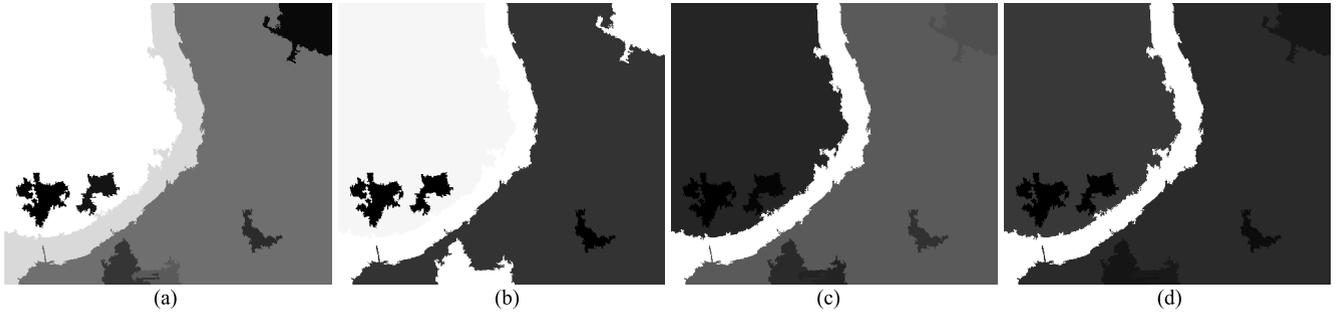


FIGURE 7. Three saliency cue maps and a single-layer saliency map of the L_2 layer. (a) LRC map. (b) BC map. (c) ED map. (d) single-layer saliency map.

function to respectively assign the river region value to 1 and the background region value to 0. The optimal solution $\hat{f}_{k,l}^i$ is then obtained by minimizing the following energy function:

$$E(f) = \arg \min \left\{ \sum_{i=1}^{N_{seg}} \sum_{k,l} E_D(\hat{f}_{k,l}^i) + \sum_{n=1}^{N_{seg}} \sum_{k,l} E_S(\hat{f}_{k,l}^i) \right\} \quad (12)$$

where f is the set consisting of all $\hat{f}_{k,l}^i$. The function contains two terms. The first term is to ensure that the predicted saliency $\hat{f}_{k,l}^i$ matches the initial saliency value $f_{k,l}^i$ calculated using (11). Hence, the data term $E_D(\hat{f}_{k,l}^i)$ is defined as the Euclidean distance between $f_{k,l}^i$ and $\hat{f}_{k,l}^i$:

$$E_D(\hat{f}_{k,l}^i) = \left\| \hat{f}_{k,l}^i - f_{k,l}^i \right\|_2^2 \quad (13)$$

The second term $E_S(\hat{f}_{k,l}^i)$ enforces the consistency between the corresponding pixels at different layers, e.g. $(\hat{f}_{k,l}^i)$ and $\hat{f}_{k,l}^n$,

$$E_S(\hat{f}_{k,l}^i) = \left\| \hat{f}_{k,l}^i - \hat{f}_{k,l}^n \right\|_2^2 \quad (14)$$

However, minimizing (12) needs to take pixels of all the layers into consideration, possibly resulting in unnecessary calculations. To simplify this process, we consider all the pixels in different layers as a tree-structured graphical model. Every node in the model represents a pixel in its corresponding layer and its saliency is only related to its parent node and the child node. The parent and child nodes are respectively the corresponding pixels in the higher and lower level layer. Inspired by the research of [46], [47], we design a new ‘‘Roll’’ belief propagation to achieve saliency fusion in our work, which consists of Top-down and Bottom-up steps. In the Top-down step, the energy function is simplified as follows:

$$E_{T-B}(\hat{f}^i) = \arg \min \left\{ \sum_{k,l} \left\| \hat{f}_{k,l}^i - f_{k,l}^i \right\|_2^2 + \sum_{k,l} \left\| \hat{f}_{k,l}^i - \hat{f}_{k,l}^{i+1} \right\|_2^2 \right\} \quad (15)$$

According to the equation, the optimal saliency of a pixel in i th layer $\hat{f}_{k,l}^i$ depends on its initial saliency value f^i (the first term) and the optimal saliency of its parent node \hat{f}^{i+1} obtained in the previous loop (the second term). Supposing the $L_{N_{seg}}$ layer is the highest level, the propagation starts from the $L_{N_{seg}-1}$ layer. The initial saliency map f^i of each layer is updated to the optimal result \hat{f}^i via minimizing (15) in every loop. Also, the updated saliency values \hat{f}^i will join the next loop as the parent node.

The top-down step progressively passes over the information from the top layer to the bottom layer in the tree model and stops when optimal nodes in the bottom layer are found. Then, a down-top procedure is performed. Its corresponding energy function is defined as:

$$E_{B-T}(\hat{f}^i) = \arg \min \left\{ \sum_{k,l} \left\| \hat{f}_{k,l}^i - f_{k,l}^i \right\|_2^2 + \sum_{k,l} \left\| \hat{f}_{k,l}^i - \hat{f}_{k,l}^{i-1} \right\|_2^2 \right\} \quad (16)$$

The difference between Eqs. (15) and (16) is that the optimal saliency $\hat{f}_{k,l}^i$ lies in the optimal saliency of its child node \hat{f}^{i-1} (the second term) instead of the parent node. The saliency of the child node \hat{f}^{i-1} has been obtained in the previous loop via minimizing (16). This saliency propagation starts from the L_2 layer. After all the nodes are updated in the bottom-up propagation procedure, the final saliency map is obtained at the highest layer, i.e., $\hat{f}^{N_{seg}}$.

Fig. 8(a) shows the fusion result for the SAR image shown in Fig. 3(a) using the above propagation method; (b) is the result after we have averaged all the single layer saliency maps; (c) is the ground-truth result. Our method can obtain more accurate river borders than the averaging method.

VI. EXPERIMENTAL RESULTS

In this section, we first describe the data sets and evaluation measures used for the experiments. The empirical evaluation and analysis of the proposed model against the state-of-the-art methods on high-resolution SAR images are then discussed.

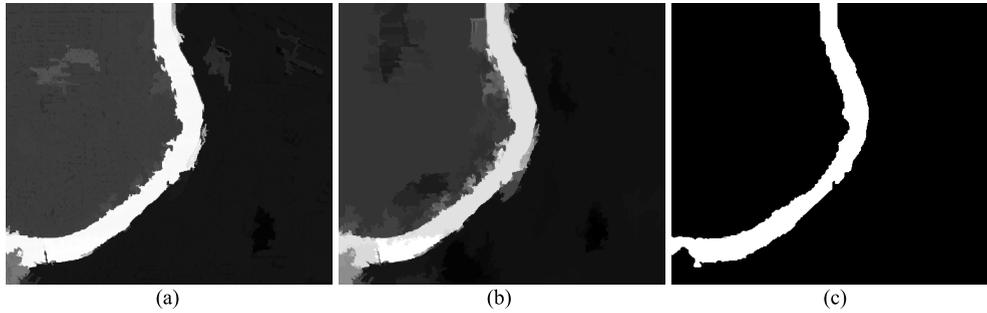


FIGURE 8. Integration results of multi-layer saliency maps. (a) our final saliency map. (b) results by averaging. (c) Ground-truth.

A. DESCRIPTION OF THE DATA SETS

We evaluate the proposed method on three SAR image datasets collected in March 2016. The first one contains the imaging results of Fangchenggang, Guangxi Province, China. As an important port in Southwest China, ship navigation needs precise river network information. The second one is an image dataset of a city in China's Shandong Province, which has complicated water networks. The image sizes for the two datasets is not uniform, ranging between 500×500 and 1000×1000 , and the resolution is 0.5 meter. In order to verify the effectiveness of our method for the low-resolution SAR images, Data Set3 is introduced in the experiments. It is the imaging results of Maoming, a coastal city of Guangdong Province, China. Its resolution is 3 meter, and each image is of larger with more than 1,000,000 pixels.

B. EVALUATION MEASURES

In the experiments, we use the receiver operating characteristic curve (ROC) and F-measure to evaluate the performance of the proposed method [48].

The ROC curve is the plot of true positive rate (TPR) versus false positive rate (FPR) by varying the threshold T_f . Specifically, a saliency map firstly is converted to a binary image using a threshold T_f changing in the range of 0 and 255. Based on each threshold, a pair of FPR and TPR scores can be computed, which are finally combined to form a false-true positive rate curve to describe the system performance at different situations:

$$FPR = \frac{|M \cap \bar{G}|}{|\bar{G}|}, \quad TPR = \frac{|M \cap G|}{|G|} \quad (17)$$

where M is the binary image, G is the ground-truth binary mask of the salient objects and \bar{G} denotes the opposite of G . For a binary mask, we use $|\cdot|$ to represent the number of non-zero entries in the mask. FPR is the ratio between the number of negative events wrongly categorized as positive (false positives) and the total number of the actual negative events, while TPR is the proportion of the salient pixels correctly identified in all the pixels of the positive regions. The area under the ROC curve (AUC) is proportional to the performance of the algorithm. A perfect model will score an AUC of 1.

We also introduce F-measure to comprehensively evaluate the overall performance of a saliency map. F-measure is

computed by the weighted harmonic of precision and recall as follows.

$$F_\beta = \frac{(1 + \beta^2) \cdot \text{precision} \times \text{recall}}{\beta^2 \cdot \text{precision} + \text{recall}} \quad (18)$$

The precision value corresponds to the ratio of the salient pixels correctly assigned to all the pixels of the extracted regions, while the recall value is defined as the percentage of the detected salient pixels in relation to the ground-truth number. A larger β^2 means the more important role of the precision value in the F-measure, and β^2 is set to 0.3 as shown in [24], [43], and [45].

C. PERFORMANCE EVALUATION

1) DATA SET1

Fig. 9 illustrates some exemplar images of the aerial scenes included in Data Set1. The river networks are characterized by different widths and shapes. Moreover, various topographic and geomorphic conditions exist, e.g., many farmlands appear in (a), complex buildings and secondary roads in (b), a highway and mountains in (c), which have to be discriminated from the river networks.

Fig. 10(a) shows the three SAR images corresponding to the optical images shown in Fig. 9, respectively. It can be seen that the shapes of the rivers in the three maps are quite different. The first scene (row 1 in Fig. 10(a)) contains a wide river, whereas the second and third scenes (rows 2 and 3 in Fig. 10(a)) have O- and Y-shaped rivers. Moreover, the complicated environment around these rivers, such as block-like farmlands in the first scene and shadows caused by mountains in the second scene, makes it difficult to detect the rivers. Fig. 10(b)–(c) show the final saliency maps and the binary results. Due to the application of multi-scale segmentation and feature processing, the proposed saliency maps tend to capture and delineate the whole river region accurately. In the binary results of our method (Fig. 10(c)), the saliency values of most the background regions are close to zero, which proves that the three measures (LRC, BC and ED) defined for the single-layer saliency estimation in the proposed method well suppressed the above-mentioned interference.



FIGURE 9. Three optical images of imaging regions in Data Set1.

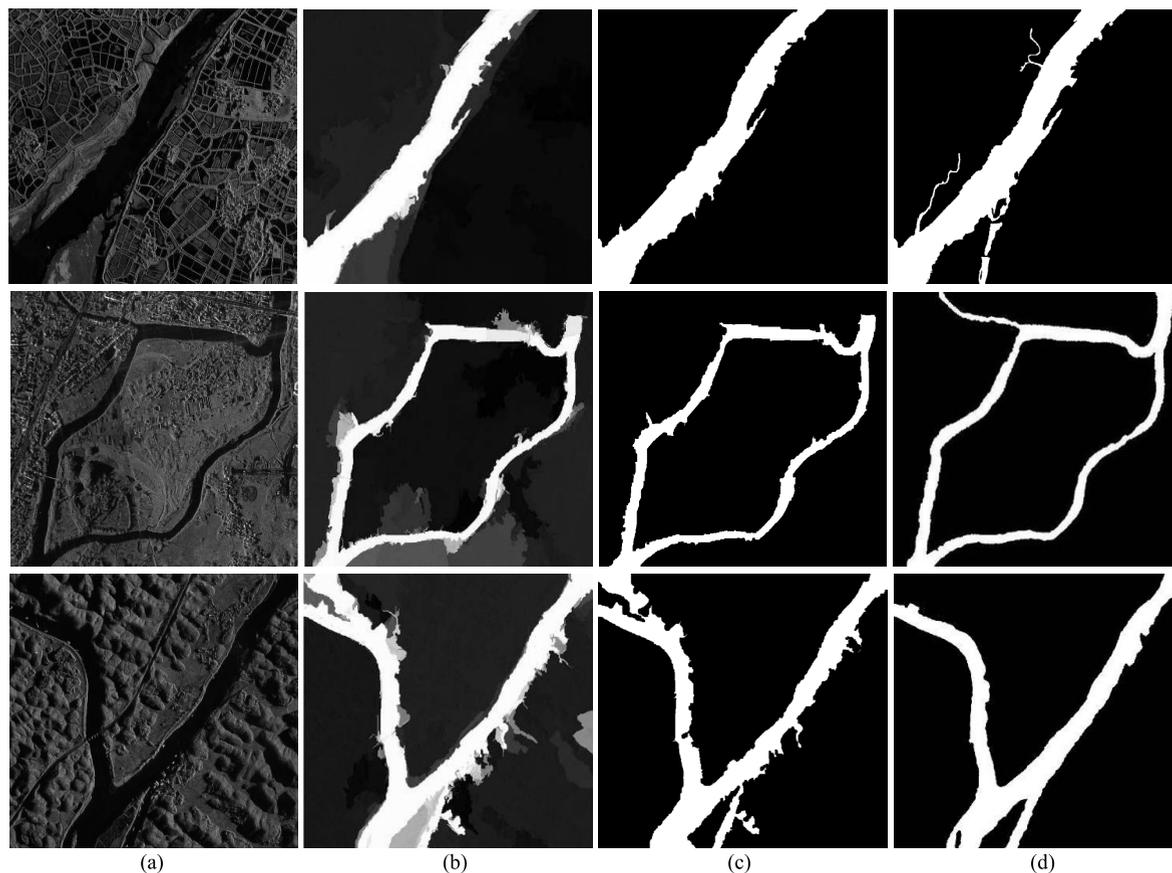


FIGURE 10. Some detection results of the proposed method for Data set1. (a) original SAR images. (b) final saliency maps. (c) binary processing results for the final saliency maps using a proper threshold in the range of 0 and 255. (d) Ground-truth.

2) DATA SET2

Some saliency maps of our proposed method for Data Set2 are shown in Fig.11. Compared with Data set1, the SAR images in Data set2 have worse azimuth ambiguity and more speckle noise, leading to a low contrast between the objects and the background. We observe that, because of the use of the VF filter set and the quantized histograms, the proposed method can suppress the heterogeneous background regions and highlight the whole rivers regions effectively. Moreover, the proposed hierarchical framework helps address the issue of object scaling. For the small-scale rivers (such as

rows 1 and 2 in Fig. 11) and large-scale rivers (such as row 3 in Fig. 11), the proposed model achieves effective detection.

3) DATA SET3

Fig.12 shows the results of the proposed method for two images in the Data Set3, where Fig. 12(a) refers to an image of 3300×1200 and Fig. 12(b) is 1250×665 . In the low-resolution SAR images, the boundaries of the rivers are blurred. There are also many trees and artificial buildings around the rivers, making the detection very difficult.

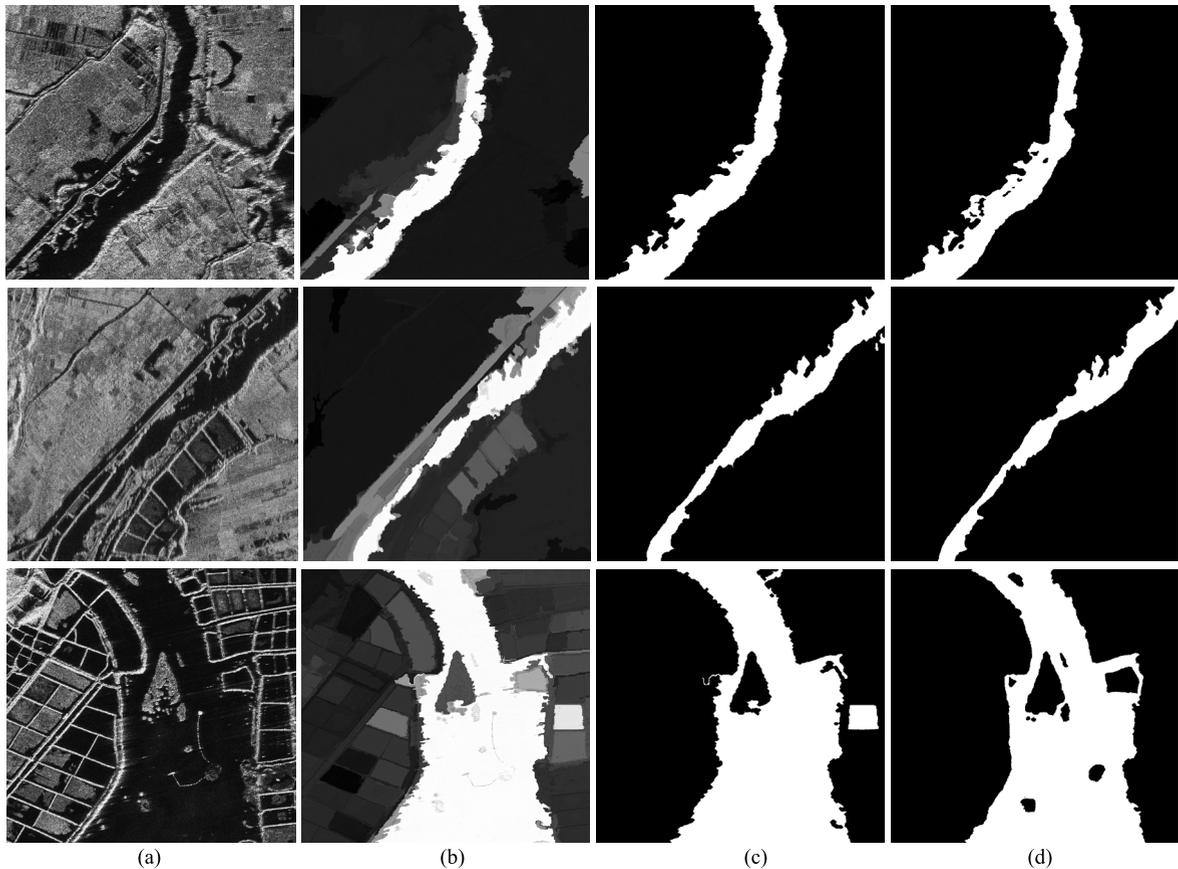


FIGURE 11. Some detection results of the proposed method for Data Set2. (a) Three representative input images. (b) Final saliency maps. (c) Binary processing results for the final saliency maps using a proper threshold in the range of 0 and 255. (d) Ground-truth.



FIGURE 12. Some detection results of the proposed method for Data Set3. (a) Image 1 and binary processing result. (b) Image 2 and binary processing result.

Our method is capable of accurately detecting the areas of the rivers in low-resolution SAR images.

4) COMPARISONS WITH THE STATE-OF-THE-ART METHODS

In order to qualify the saliency detection performance of our saliency model, we use the average ROC and the

average F-measure curves of the three data sets, shown in Figs. 13 and 14. The corresponding curves of the existing multiple saliency object detection methods are also given in the comparison, i.e., Saliency Detection via Dense and Sparse Reconstruction (DSR) [24], Saliency Optimization from Robust Background Detection (RBD) [45],

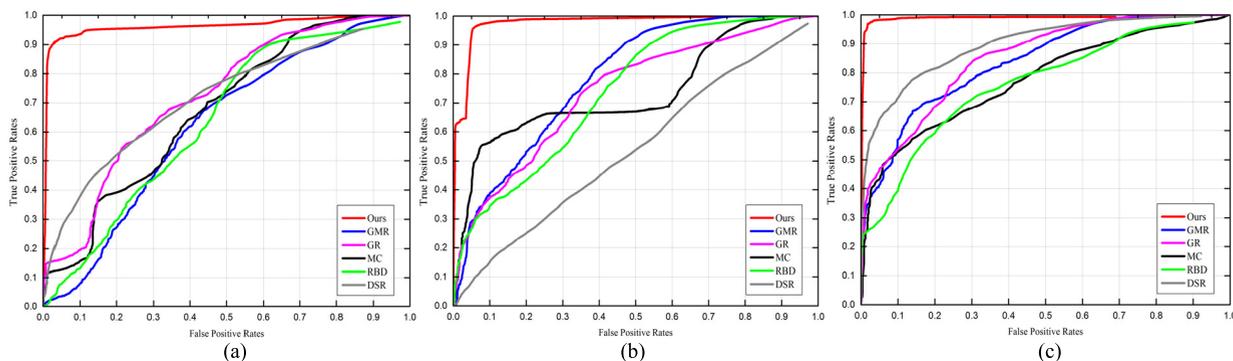


FIGURE 13. Comparison of the average ROC curve with 5 state-of-the-art algorithms, GMR, GR, MC, RBD and DSR on three data sets. (a) Data Set1. (b) Data Set2. (c) Data Set3.

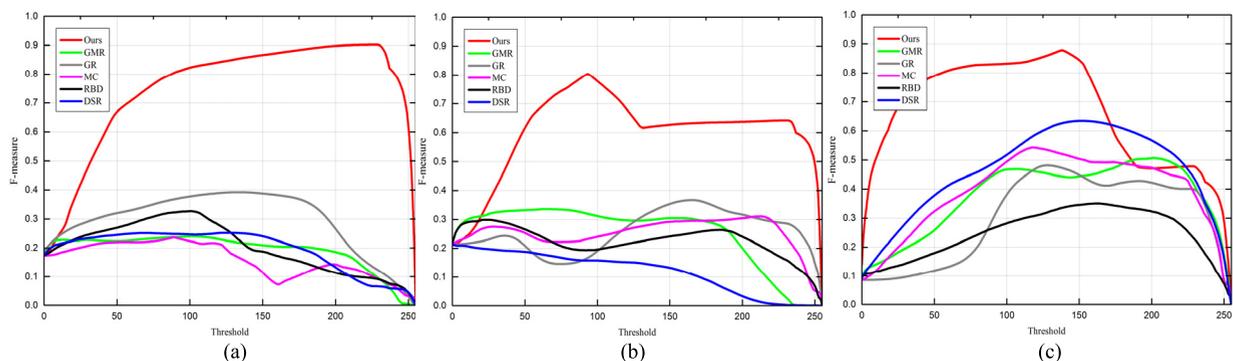


FIGURE 14. Comparison of the average F-measure curve with GMR, GR, MC, RBD and DSR on three data sets. (a) Data Set1. (b) Data Set2. (c) Data Set3.

Saliency Detection via Graph-based Manifold Ranking (GMR) [43], Graph-regularized Saliency Detection with Convex-hull-based Center Prior (GR) [49], Saliency Detection via Absorbing Markov Chain (MC) [23].

In Fig.13, the proposed method achieves a competitive performance compared to the other five methods. The areas under the ROC curves of our method are both close to 1 and the TPR values are larger than 0.9, while keeping a low FPR. The other 5 methods achieved less satisfactory results and their areas under the ROC curves are below 0.8.

Likewise, we make the threshold vary from 0 to 255 to obtain the average F-measure curves of all the six saliency detection methods on the three data sets. From Fig.14, it can be seen that the proposed method has a higher F-measure value at most of the thresholds, indicating its better detection capability than the others.

Fig. 15 shows the average maximum F-measure and the AUC values of the six saliency models on each data set. Specifically, Fig. 15(a) shows the average maximum F-measure of the six algorithms when the threshold varies from 0 to 255, which represents the best detection results a detection algorithm can achieve. Our three average maximum F-measures are larger than 0.8, higher than those of the other five methods. Fig. 15(b) shows the average AUC values of the six methods on the three data sets. The larger the AUC values,

TABLE 2. Computational complexity of the proposed method.

Step	Computational Complexity
1.Features Extraction	$1,600 N$
2.Superpixel Generation	$2 N \log N$
3.Saliency Calculation	$(W \cdot N_{seg} / S_{superpixel}) N^2$

the better the detection results. Thanks to the appropriate saliency calculation method, the average AUC of our method is larger than 0.9 on all the images, far better than the other five methods. This is because the priors (contrast, center and smoothness) used in the saliency detection of the other methods capture less difference between the rivers and the background.

D. RUNNING TIME

To evaluate the computational efficiency of our method, we present the computational complexity of each step of the method in Table 2. We assume that one addition and one multiplication are the same in terms of computational complexity and the input SAR image has N pixels. As can be seen from Table 2, the third step ‘‘Saliency Calculation’’ requires the largest amount of computational efforts, where W represents the optimal quantization parameter vector, N_{seg} is the segment times, and $s_{superpixel}$ is the average number of

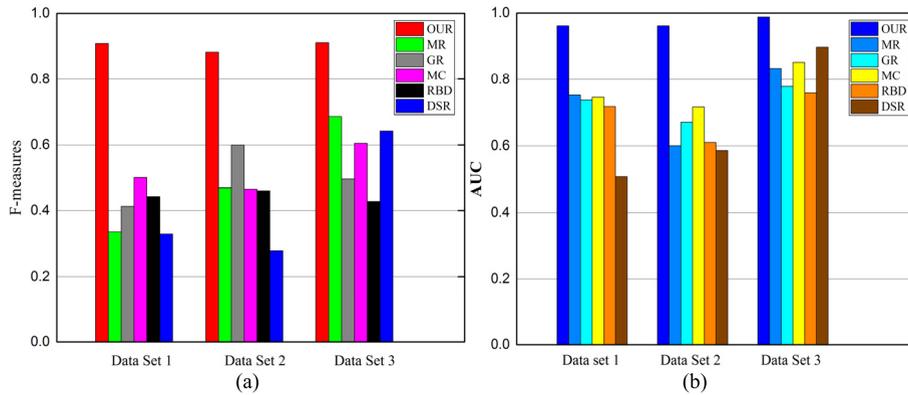


FIGURE 15. Comparison of the average maximal F-measure and average AUC with 5 state-of-the-art algorithms, GMR, GR, MC, RBD and DSR on three data sets. (a) Average Maximum F-measure. (b) Average AUC.

TABLE 3. Comparison of average run time (seconds per image) with 5 state-of-the-art algorithms the longest running time is highlighted in red.

Method	Ours	GMR	GR	MC	RBD	DSR
Time(s) of Data Set1	4.83	3.44	3.41	0.66	0.92	0.50
Time(s) of Data Set2	6.39	4.71	4.86	0.83	1.38	0.73
Time(s) of Data Set3	13.75	16.08	16.26	2.89	16.92	3.49

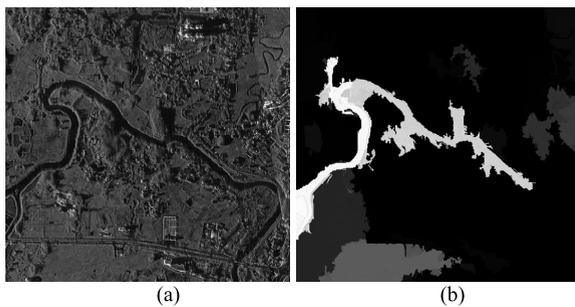


FIGURE 16. An example of the failure cases. (a) Input image. (b) Our result.

pixels in a superpixel. In our experiment, we set $N_{seg} = 8$, $s_{superpixel} = 10000$, $W = [3\ 2\ 3\ 4\ 7\ 3\ 3\ 2]$ and the total calculation complexity of the proposed method approximately is $0.5N^2 + 2N \log N + 1600N$.

Moreover, we compare the average computational time of our method on the three data sets with the other five methods and show the results in Table 3. The simulation software is MATLAB 2013b and the main configuration of the computer includes 4GB RAM and Intel Core i7-4790 CPU. For data set 1 and 2, the computational time of our method is nearly equal to that of GMR and GR, but far more than that of MC, RBD and DSR. For the large scale images of data set 3, the average computational time of GMR, GR and RBD increases rapidly, more than that of our method. The comparison shows that our method has better efficiency in handling large-size images.

E. FAILURE CASES

It is worth noting that our method still has some limitations. It is difficult to reach the accuracy of a pixel level by the

proposed method because it takes superpixels as the processing units. Moreover, the number of the scale spaces is limited in the multi-scale segmentation and some features are not positively contributing to the classification. Hence our results for some images are unsatisfactory. Fig. 16 shows an example of the failing cases, where the width of the river is about 20 pixels. Our method does not accurately distinguish the whole river region from the shores and part of the river is even missed.

VII. CONCLUSION

In this study, a new river detection algorithm based on biological visual saliency models has been presented for airborne high-resolution SAR data. Initial visual features are first extracted using a VF filtering set simulating the signal processing mechanism of the receptive field in the retina and visual cortex. Then, inspired by the bottom-up transfer theory used in the biological visual system, the hierarchical segmentation, regional histogram quantization, region merging, single-layer saliency measurement and multilayer saliency fusion were systematically undertaken to obtain a full-resolution saliency map.

The experiments showed: 1) saliency modeling is a useful and efficient way to detect rivers from a SAR image; 2) hierarchical merging and integration based on the quantized histograms ensure identifying the targets in various scale spaces effectively; 3) by using three saliency cues i.e., LRC, BC, ED, which were designed according to the prior information of rivers, the proposed approach well handled the heterogeneous background and effectively outlined the target regions. In addition, both subjective and objective evaluations on the three datasets demonstrate that the proposed approach

achieved consistently better saliency detection performance than the other state-of-the-art saliency models.

ACKNOWLEDGMENT

The SAR images used in the experiments are the courtesy of Beijing Institute of Radio Measurement. The authors would like to thank their support in this work.

REFERENCES

- [1] D. C. Mason, I. J. Davenport, J. C. Neal, G. J.-P. Schumann, and P. D. Bates, "Near real-time flood detection in urban and rural areas using high-resolution synthetic aperture radar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 8, pp. 3041–3052, Aug. 2012.
- [2] F. Covello *et al.*, "COSMO-SkyMed an existing opportunity for observing the Earth," *J. Geodyn.*, vol. 49, nos. 3–4, pp. 171–180, Apr. 2010.
- [3] N. Pierdicca, L. Pulvirenti, M. Chini, L. Guerriero, and L. Candela, "Observing floods from space: Experience gained from COSMO-SkyMed observations," *Acta Astron.*, vol. 84, pp. 122–133, Mar./Apr. 2013.
- [4] Y. Yamada, T. Sakurai-Amano, and M. Takagi, "Detection of flood damaged areas in the entire Chao Phraya River Basin from JERS-1/SAR images with a help of spatial information," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, vol. 5, Jun. 2002, pp. 2877–2879.
- [5] J. Sun and S. Mao, "River detection algorithm in SAR images based on edge extraction and ridge tracing techniques," *Int. J. Remote Sens.*, vol. 32, no. 12, pp. 3485–3494, 2011.
- [6] P. Ahtonen and M. Hallikainen, "Automatic detection of water bodies from spaceborne SAR images," in *Proc. IGARSS*, vol. 5, 2005, pp. 3845–3848.
- [7] S. Klemenjak, B. Waske, S. Valero, and J. Chanussot, "Automatic detection of rivers in high-resolution SAR data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 5, pp. 1364–1372, Oct. 2012.
- [8] Z. Tian, C. Wu, D. Chen, X. Yu, and L. Wang, "A novel method of river detection for high resolution remote sensing image based on corner feature and SVM," in *Proc. Int. Symp. Neural Netw.*, 2012, pp. 266–273.
- [9] L. Zhang, H. Li, P. Wang, and X. Yu, "Detection of regions of interest in a high-spatial-resolution remote sensing image based on an adaptive spatial subsampling visual attention model," *GISci. Remote Sens.*, vol. 50, no. 1, pp. 112–132, 2013.
- [10] F. Zhang, B. Du, and L. Zhang, "Saliency-guided unsupervised feature learning for scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2175–2184, Apr. 2015.
- [11] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.
- [12] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
- [13] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proc. Adv. Neural Inf. Process. Syst.*, 2006, pp. 545–552.
- [14] L. Zhang, M. H. Tong, T. K. Marks, H. Shan, and G. W. Cottrell, "SUN: A Bayesian framework for saliency using natural statistics," *J. Vis.*, vol. 8, no. 7, p. 32, Dec. 2008.
- [15] N. D. B. Bruce and J. K. Tsotsos, "Saliency based on information maximization," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 18, Jun. 2006, pp. 155–162.
- [16] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.
- [17] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.
- [18] L. Vincent and P. Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 6, pp. 583–598, Jun. 1991.
- [19] T. Liu *et al.*, "Learning to detect a salient object," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 2, pp. 353–367, Feb. 2011.
- [20] J. Wang, H. Jiang, Z. Yuan, M.-M. Cheng, X. Hu, and N. Zheng, "Salient object detection: A discriminative regional feature integration approach," *Int. J. Comput. Vis.*, vol. 123, no. 2, pp. 251–268, 2017.
- [21] V. Gopalakrishnan, Y. Hu, and D. Rajan, "Random walks on graphs for salient object detection in images," *IEEE Trans. Image Process.*, vol. 19, no. 12, pp. 3232–3242, Dec. 2010.
- [22] Y. Qin, H. Lu, Y. Xu, and H. Wang, "Saliency detection via cellular automata," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 110–119.
- [23] B. Jiang, L. Zhang, H. Lu, C. Yang, and M.-H. Yang, "Saliency detection via absorbing Markov chain," in *Proc. IEEE Int. Conf. Comput. Vis.*, Jun. 2013, pp. 1665–1672.
- [24] X. Li, H. Lu, L. Zhang, R. Xiang, and M.-H. Yang, "Saliency detection via dense and sparse reconstruction," in *Proc. IEEE Int. Conf. Comput. Vis.*, Jun. 2013, pp. 2976–2983.
- [25] Y. Xie, H. Lu, and M.-H. Yang, "Bayesian saliency via low and mid level cues," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1689–1698, May 2013.
- [26] D. H. Hubel and T. N. Wiesel, "Ferrier lecture—Functional architecture of macaque monkey visual cortex," *Proc. Roy. Soc. B, Biol. Sci.*, vol. 198, no. 1130, pp. 1–59, 1977.
- [27] R. W. Rodieck, "Quantitative analysis of cat retinal ganglion cell response to visual stimuli," *Vis. Res.*, vol. 5, no. 12, pp. 583–601, Dec. 1965.
- [28] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
- [29] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 2, no. 7, pp. 1160–1169, 1985.
- [30] J. P. Jones and L. A. Palmer, "An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex," *J. Neurophysiol.*, vol. 58, no. 6, pp. 1233–1258, Dec. 1987.
- [31] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, "Robust object recognition with cortex-like mechanisms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 3, pp. 411–426, Mar. 2007.
- [32] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex," *J. Physiol.*, vol. 195, no. 1, pp. 215–243, 1968.
- [33] E. Peterhans and R. von der Heydt, "The role of end-stopped receptive fields in contour perception," in *Proc. 15th Göttingen Neurobiol. Conf.*, 2012, p. 29.
- [34] A. Dobbins, S. W. Zucker, and M. S. Cynader, "Endstopped neurons in the visual cortex as a substrate for calculating curvature," *Nature*, vol. 329, no. 6138, pp. 438–441, 1987.
- [35] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat," *J. Physiol.*, vol. 28, no. 2, pp. 229–289, 1965.
- [36] R. Baumann, R. von der Zwan, and E. Peterhans, "Figure-ground segregation at contours: A neural mechanism in the visual cortex of the alert monkey," *Eur. J. Neurosci.*, vol. 9, no. 6, pp. 1290–1303, 1997.
- [37] B. Heider, V. Meskenaitė, and E. Peterhans, "Anatomy and physiology of a neural mechanism defining depth order and contrast polarity at illusory contours," *Eur. J. Neurosci.*, vol. 12, no. 11, pp. 4117–4130, 2000.
- [38] E. Peterhans and F. Heitger, "Simulation of neuronal responses defining depth order and contrast polarity at illusory contours in monkey area V2," *J. Comput. Neurosci.*, vol. 10, no. 2, pp. 195–211, 2001.
- [39] R. von der Heydt, E. Peterhans, and M. R. Dursteler, "Periodic-pattern-selective cells in monkey visual cortex," *J. Neurosci.*, vol. 12, no. 4, pp. 1416–1434, 1992.
- [40] C. Grigorescu, N. Petkov, and M. A. Westenberg, "Contour detection based on nonclassical receptive field inhibition," *IEEE Trans. Image Process.*, vol. 12, no. 7, pp. 729–739, Jul. 2003.
- [41] H. Jiang, J. Wang, Z. Yuan, T. Liu, N. Zheng, and S. Li, "Automatic salient object segmentation based on context and shape prior," in *Proc. BMVC*, vol. 6, 2011, pp. 110.1–110.12.
- [42] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.*, vol. 59, no. 2, pp. 167–181, Sep. 2004.
- [43] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 3166–3173.
- [44] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numer. Math.*, vol. 1, no. 1, pp. 269–271, Dec. 1959.
- [45] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2814–2821.
- [46] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1155–1162.
- [47] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

- [48] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1597–1604.
- [49] C. Yang, L. Zhang, and H. Lu, "Graph-regularized saliency detection with convex-hull-based center prior," *IEEE Signal Process. Lett.*, vol. 20, no. 7, pp. 637–640, Jul. 2013.
- [50] A. D'Addabbo, A. Refice, G. Pasquariello, F. P. Lovergine, D. Capolongo, and S. Manfreda, "A Bayesian network for flood detection combining SAR imagery and ancillary data," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3612–3625, Jun. 2014.
- [51] A. Niedermeier, E. Romaneessen, and S. Lehner, "Detection of coastlines in SAR images using wavelet methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 5, pp. 2270–2281, Sep. 2000.
- [52] Y. Yu, B. Wang, and L. Zhang, "Hebbian-based neural networks for bottom-up visual attention and its applications to ship detection in SAR images," *Neurocomputing*, vol. 74, no. 11, pp. 2008–2017, 2011.
- [53] A. Borji, M.-M. Cheng, H. Jiang, and J. Li, "Salient object detection: A survey," *Eprint ArXiv*, vol. 16, no. 7, p. 3118, 2014.
- [54] J. Wagemans et al., "A century of Gestalt psychology in visual perception: II. Conceptual and theoretical foundations," *Psychol. Bull.*, vol. 138, no. 6, pp. 1217–1218, 2012.



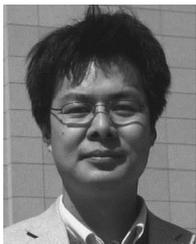
FEI GAO received the B.S. and M.S. degrees from the Xi'an Petroleum Institute, Xi'an, China, in 1996 and 1999, respectively, and the Ph.D. degree from the Beijing University of Aeronautics and Astronautics (BUAA), Beijing, China, in 2005.

He is currently an Associate Professor with the School of Electronic and Information Engineering, BUAA. He is interested in radar signal processing, moving target detection, and image processing.



FEI MA received the M.S. degree in electronic and information engineering from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 2016, where he is currently pursuing the Ph.D. degree.

His research interests include radar signal processing, image processing, machine learning, and target detection.



JUN WANG received the B.S. degree from the North Western Polytechnical University, Xi'an, China, in 1995, and the M.S. and Ph.D. degrees from the Beijing University of Aeronautics and Astronautics (BUAA), Beijing, China, in 1998 and 2001, respectively.

He is currently a Professor with the School of Electronic and Information Engineering, BUAA. He is interested in signal processing, DSP/FPGA real-time architecture, target recognition, and tracking. His research has resulted in 38 papers in journals, books, and conference proceedings.



JINPING SUN received the M.Sc. and Ph.D. degrees from the Beijing University of Aeronautics and Astronautics (BUAA), Beijing, China, in 1998 and 2001, respectively.

He is currently a Professor with the School of Electronic and Information Engineering, BUAA. His research interests include high-resolution radar signal processing, image understanding, and robust beam forming.



ERFU YANG received the Ph.D. degree in robotics in the interdisciplinary area of robotics and autonomous systems from the School of Computer Science and Electronic Engineering, University of Essex, Colchester, U.K. He is currently a Lecturer with the Department of Design, Manufacture and Engineering Management (DMEM), University of Strathclyde, Glasgow, U.K.

His main research interests include robotics, autonomous systems, mechatronics, manufacturing automation, computer vision, image/signal processing, nonlinear control, process modeling and simulation, condition monitoring, fault diagnosis, multi-objective optimizations, and applications of machine learning and artificial intelligence, including multi-agent reinforcement learning, fuzzy logic, neural networks, bio-inspired algorithms, and cognitive computation. He has over 70 publications in these areas, including over 30 journal papers and ten book chapters.



HUIYU ZHOU received the B.Eng. degree in radio technology from the Huazhong University of Science and Technology, China, in 1990, the M.Sc. degree in biomedical engineering from the University of Dundee, Dundee, U.K., and the Ph.D. degree in computer vision from Heriot-Watt University, Edinburgh, U.K.

He is currently with the School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Belfast, U.K. He will be with the University of Leicester, U.K., from 2018. He has authored or co-authored widely in international journals and conferences. His research interests include computer vision, human motion analysis, and intelligent systems.

...