

Hierarchical-Biased Random Walk for Urban Remote Sensing Image Segmentation

Xudong Zhao , *Student Member, IEEE*, Ran Tao , *Senior Member, IEEE*, Xuejing Kang,
and Wei Li , *Senior Member, IEEE*

Abstract—Random walk (RW) technique, with benefit of handling complicated boundaries, has recently drawn increasing attention in image segmentation. In this paper, RW is employed for urban remote sensing image segmentation. To deal with the complex spatial distribution with heterogeneous structures, a novel hierarchical-biased RW (HBRW) method is proposed. Firstly, edge regions extracted by fractional differential are combined with histograms to obtain plentiful features. Then, Dirichlet process mixture model is used to generate hierarchical global prior distribution and local seeds, which substitute manual scribbles. Moreover, the proposed model can adapt to different resolution segmentation tasks through adjusting the concentration parameter. Final segmentation output is obtained by biased RW. Experimental results on urban high-resolution remote sensing images demonstrate that the proposed algorithm achieves better performance than other state-of-the-art algorithms.

Index Terms—Dirichlet process mixture model (DPMM), fractional differential, hierarchical-biased random walk (HBRW), urban remote sensing image segmentation.

I. INTRODUCTION

SEGMENTATION is a typical and crucial step in automatic understanding and interpretations of urban remote sensing image. This task keeps challenging due to some heterogeneous appearance in high-resolution data, i.e., objects like buildings, streets, trees, and cars [11], [19], [26], [30]. During the past decades, many segmentation methods have been investigated. The classical segmentation methods include clustering [8], total variation [29], probabilistic [17], and thresholding [28] techniques. These standard methods simply use the amplitude value of pixels to produce the segmentation results without considering the characteristics of the remote sensing images. To improve the performance, graph-cut [9], normalized-cut [24], and random walk (RW) methods [13] were subsequently proposed

for segmentation. These graph-based methods attracts attention because of their solid mathematical foundation and capability of utilizing global and local spatial information during segmentation.

One of the typical graph-based segmentation methods, i.e., RW, has been widely used for different tasks in image segmentation [12], [13], [15], [25]. The RW model was first proposed by Grady and Funka-Lea [14] for medical image segmentation and then extended to natural images [13]. And then, many significant methods based on RW have been proposed [6], [12], [16], [25]. In the RW setting, the user marks some pixels in the image, then assumes that the random walker starts with each unlabeled pixel and calculates the probability that the random walker first arrives at the already tagged pixels. At each pixel, the tag with the maximum probability is selected as the final tag to obtain the final segmentation results.

To deal with the complex texture, the random walker with restarting (RWR) was introduced [18]. The RWR introduces restarting probability c , and the random walker would return to the starting node (pixel) with the probability of c or out to an adjacent pixel at the probability of $1 - c$. A lazy RW algorithm was proposed for superpixel segmentation [23], in which the random walker stays at the current node with probability $1 - c$ and transfers to adjacent nodes with probability c . These approaches are typically called graph-based since they use a regular grid as the image representation domain. One limiting property of these algorithms is that each segment has to be connected to a seed. In order to incorporate prior information into the RW approach and eliminate this constraint, Grady [12] proposed the use of an augmented graph where the additional nodes were connected to the original image's nodes. Similarly, Dong *et al.* [7] proposed a unifying sub-Markov RW method (subRW), which can be viewed as an extension of RWR where additional nodes were added to the original image graph providing various information. In [3], normalized graph-driven diffusion and RW schemes (normalized RW (NRW) and lazy NRW (LNRW)) are proposed for arbitrary graph image segmentation.

The common ground among all the aforementioned algorithms is that they consider only similarity of pixel values, which causes inadequacy of data features. Meanwhile, the output of RW algorithms relies largely on the user tags since they only use information contained in tags and the transition probability between adjacent pixels. Particularly, meticulous scribbles are needed for different segmentation tasks. Furthermore, once users ensure the precision of prior and scribbles, the number of segments is fixed and segmentation result is unadjustable.

Manuscript received September 27, 2018; revised December 28, 2018 and February 17, 2019; accepted March 9, 2019. This work was supported in part by the National Natural Science Foundation of China under Grants 61421001, 61701036 and U1833203. (Corresponding author: Ran Tao.)

X. Zhao, R. Tao, and W. Li are with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China, and also with the Beijing Key Laboratory of Fractional Signals and Systems, Beijing 100081, China (e-mail: zhaoxudong@bit.edu.cn; rantao@bit.edu.cn; liwei089@ieee.org).

X. Kang is with the School of Computer Science and Technology, Beijing University of Posts and Telecommunications, Beijing 100088, China (e-mail: kangxuejing@bupt.edu.cn).

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/JSTARS.2019.2905352

1939-1404 © 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.

See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

To solve these drawbacks, a novel hierarchical-biased RW (HBRW) model is proposed. First, the proposed method uses features including edge region extracted by fractional differential and histograms to incorporate more information, which mitigates the inadequacy of features. After feature extraction, learning method based on Dirichlet process mixture model (DPMM) is applied to adaptively get tags. Since DPMM is determined by the concentration parameter, which helps the users specify a level of resolution for segmentation rather than specifying a number of image segments. Then, the adaptive mixture model is established as *a priori* constraint for the RW algorithm operation. Finally, the probabilities are computed based on the random walker starting from each node reaching the prior nodes in the extended graph with multiple layers. On each layer, this is an RW process with global and class specific prior. Finally, segmentation results with different number of class for various tasks are obtained.

Compared with previous methods, the main contributions of the proposed HBRW are summarized as follows.

- 1) Except of using basic texture in spatial space, edge information with fractional differential operation is preserved, which makes image high-frequency more prominent while preserving smooth area.
- 2) Different from traditional RW methods that rely on manual tags, the proposed HBRW employs DPMM learning for hierarchical label prior, including global and local guides with more comprehensive information.

The outline of the paper is as follows. We first discuss related work about graph notation and RW in Section II. Then, the proposed HBRW model for image segmentation is discussed in Section III. In Section IV, the experimental results are presented. Finally, Section V ends this paper with some concluding remarks.

II. PRELIMINARIES

An image is considered as a weighted, undirected and connected graph $G = \{V, E, W\}$, where V is a finite set of vertices V with $|V| = N$, E is a set of edges, and W is a weighted adjacency matrix (e.g., [5] for all definitions in this section). The non-normalized graph Laplacian is defined as $L := D - W$, where D is the diagonal degree matrix with the n th element $d_n = \sum_{m \neq n} W_{mn}$ being the degree of vertex n . A weight $w_{ij} \in W$ of edge e_{ij} measures the likelihood that a random walker will cross this edge. As many other algorithms based on graph expressed, a weight w_{ij} is formulated as

$$w_{ij} = \exp\left(-\frac{\|I_i - I_j\|^2}{\sigma}\right) + \varepsilon \quad (1)$$

where I_i and I_j are the pixel intensities at two nodes v_i and v_j , σ is the controlling parameter and ε is a small constant as 10^{-6} . Standard RW on a graph can be described as a Markov process with transition probability matrix

$$Q_{ij} = \begin{cases} w_{ij}/d_i, & \text{if } i \sim j \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

which can also expressed as matrix operations $Q = D^{-1}W$. According to spectral graph theory [5], RW converges to a unique

stable probability distribution π , which satisfy the following balance equation $\pi_i = \frac{d_i}{\sum_{i=1}^N d_i}$.

One limiting property of the traditional RW is that each segment has to be connected to a seed. In order to overcome the limitation, biased RW (BRW) [1] was proposed [4]

$$r^{(t+1)}(i) = (1 - c) \sum_{j \sim i} r^{(t)}(j)w_{ij} + cz(i) \quad (3)$$

where $0 \leq c \leq 1$ is a controlling parameter, and z is called biased distribution. z is a probability distribution that $z(i) \geq 0$ and $\sum_i z(i) = 1$. Setting r^t as the distribution of the random walker at time t , the distribution at $t + 1$ is given by r^{t+1} . Equation (3) describes an RW process in which the RW hops on the graph G with the transition probability $1 - c$ and may hop to other vertices specified by prior distribution with probability c .

III. PROPOSED HBRW

The proposed HBRW is illustrated in Fig. 1, where we first extract edge regions using fractional differential, which is further combined with histograms to obtain plentiful features. Then, DPMM is used to get hierarchical prior distributions. At last, prior distribution and seeds for each class are used to compute final reaching probabilities. According to the task requirements, parameters are adjusted to provide output with particular number of segments.

A. Fractional Differential for Edge Detection

Various features play an important role in image segmentation pretreatment. However, traditional RW algorithms only consider the pixel values. For the lacking of other features, statistical information of pixel intensity and texture information are missed, which further results in low segmentation quality. To achieve reliable output, the proposed HBRW combines low-level features, such as pixel intensity and edge.

Given an image, observations (pixels) are extracted using the histogram clustering [22]. Specifically, multichannel images are converted into gray image to form the histogram vectors. Edge detection methods use the discontinuous distribution of the pixel intensity, and extract the external contour of the target object for the gray point mutation site. The traditional edge detection operators, such as gradient operator and Laplace operator, can only detect steep edges, but some edges formed by changing intensity slowly are ignored. Also, the antinoise performance of the traditional edge detection operator is poor. Thus, fractional differential operation is considered for edge detection [21]. Fractional differential is more sensitive and better in antinoise performance. The differential representation of fractional differential is

$$D^v f(i) \approx f(i) + (-v)f(i-1) + \dots + \frac{(-v)(-v+1)}{2} f(i-2) + \frac{\Gamma(-v+1)}{n! \Gamma(-v+n+1)} f(i-n) \quad (4)$$

where $D^v f(i)$ is the v order fractional differential of original signal $f(i)$, and $\Gamma(\alpha) = \int_0^\infty e^{-x} x^{\alpha-1} dx = (\alpha - 1)!$.

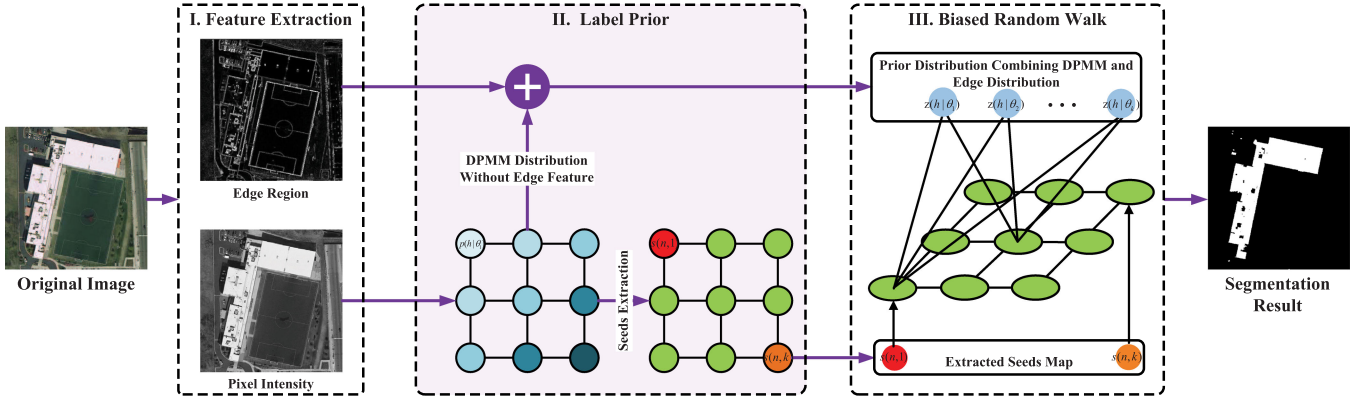


Fig. 1. Flowchart of the proposed HBRW model for urban remote sensing image segmentation. In Part II, blue nodes $p(h|\theta_k)$ are DPMM distribution without edge feature. Green nodes are unseeded nodes while other color nodes are extracted seeds $s(n, k)$. Then, in Part III, blue nodes $z(h|\theta_k)$ are prior distribution of BRW. Ellipse nodes denote the original nodes and circle nodes are auxiliary nodes. Only part of the transition edges are shown for simplification.

For digital image processing, the shortest distance from the gray change of the two-dimensional (2-D) image signal is between two pixels, so the duration of the 2-D digital image in two axes can only be measured in pixels. The partial fractional differential of pixel $I(i, j)$ is defined as

$$\begin{aligned} D_i^v(I) &\approx I(i, j) + (-v)I(i-1, j) + \frac{(-v)(-v+1)}{2}I(i-2, j) \\ D_j^v(I) &\approx I(i, j) + (-v)I(i, j-1) + \frac{(-v)(-v+1)}{2}I(i, j-2). \end{aligned} \quad (5)$$

The fractional differential can improve high-frequency signals while nonlinearly preserving low-frequency signals [21]. Fractional differential is used to make the image edge more prominent while preserving the image texture information of smooth area. By subtracting the corresponding pixel value of the original image, we can get the edge information, which has been changed by fractional differential operation, i.e., $ER = D^v(I) - I$. As step I in Fig. 1, both pixel intensity feature and edge feature are extracted for segmentation.

B. DPMM Learning for Label Prior

A core issue of segmentation models is how many segments should be inferred from an image. Standard RW algorithms need great manual work in scribbles marking to decide the number of segments. To avoid this, DPMM learning is employed to replace the number of segments under human decision with adaptively generated prior distribution (more details about DPMM learning can be seen in [10] and [22]). The Dirichlet process (DP) is defined from the aspect of set partitioning. Suppose that H is a distribution on the measure space Θ and α is a positive value. For arbitrary finite measure division A_1, \dots, A_r , the distribution G_0 and vector $(H(A_1), \dots, H(A_r))$ are arbitrary. Then, the DP is defined by the base distribution H and the concentration parameter α , which can be expressed as $G_0 \sim DP(\alpha, H)$ when

$$(G_0(A_1), \dots, G_0(A_r)) \sim DP(\alpha H(A_1), \dots, \alpha H(A_r)). \quad (6)$$

Let $h_i \in \mathbb{R}$ denote the i th observation, where i is the number of pixels in I . We then model the distribution where h_i is generated. Draw a parameter $\theta_i \sim G_0$ and then $h_i \sim F(\cdot|\theta_i)$, where F is the probability model associated with a mixture component. DPMM is based on the DP priors with parameter θ_i . Suppose that the prior distribution G_0 is uncertain and drawn from a DP $G_0 \sim DP(\alpha H)$, the hierarchical model of DPMM is represented as

$$\begin{aligned} h_i &\sim F(\cdot|\theta_i) \\ \theta_i &\sim G_0 \\ G_0 &\sim DP(\alpha H). \end{aligned} \quad (7)$$

In Section III-A, observations are identified with sites $i = 1, \dots, n$. A histogram window is placed around each site and draw h_i from the intensity values I_i of all the pixels in the window. The size of window is described as data value number Nbin for the whole image. Each histogram bin is expressed as $h_i = h_{i1}, \dots, h_{iNbin}$. Assuming that a label k has an intensity distribution P_k for each node, where the elements indicated by $p(h_i|\theta_k)$ means the probability belonging to P_k at the i th observation. The distribution learnt by DPMM is as follows:

$$p(h_i|\theta_k) = G_0(\theta_k) \prod_{i|S_i=k} F(h_i|\theta_k) \quad (8)$$

where $G_0(\theta_k)$ is the base measure. $F(h_i|\theta_k)$ is the likelihood of k th class.

In the proposed HBRW method, $p(h|\theta_k)$ is the label prior for all nodes in V , which can be used for both seeded and unseeded nodes. Unlike traditional RW algorithms, which need user tags at particular pixels (seeds), the proposed method draw partial nodes from the hierarchical label prior $p(h|\theta_k)$ as tags. Since these nodes are uniformly drawn from prior distribution, the number of classes is decided by the seeds. Naturally, a lower percentage of seeds is corresponding to less effects of prior distribution. Hierarchical prior distribution and its corresponding seeds as global and local guide are designed for BRW.

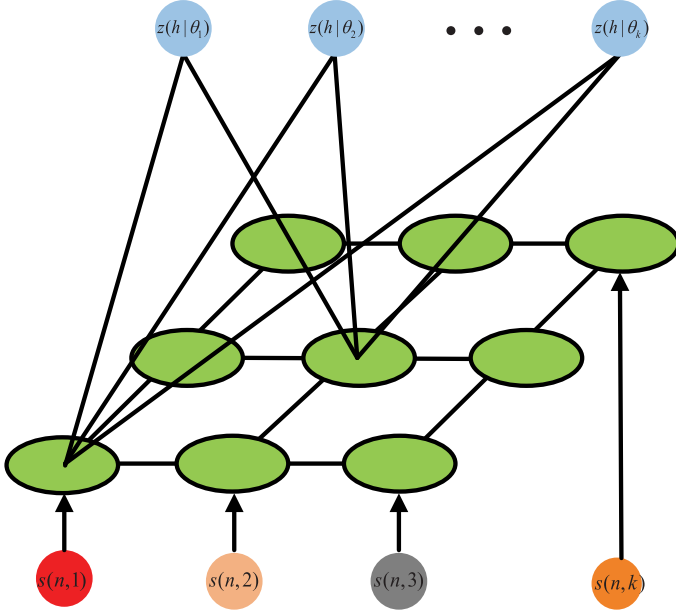


Fig. 2. Nodes graph with hierarchical prior nodes of the proposed HBRW. Ellipse nodes denote the original nodes and circle nodes are auxiliary nodes. Green nodes are unseeded nodes, blue nodes $z(h|\theta_k)$ are prior distribution combining DPMM and edge distribution, other color nodes are seeds $s(n, k)$. Only part of the transition edges are shown for simplification.

C. Segmentation via BRW

Three parts of prior information are obtained including global probability distribution $p(h|\theta_k)$, edge distribution $ER(\theta_k)$, and seeds map S . The prior information is added to construct a graph with prior \bar{G} as shown in Fig. 2. For the global prior distribution, each node is connected with all nodes in V . To combine two kinds of global prior information, the weights $w_{iz(z(h|\theta_k))}$ is defined as

$$w_{iz(z(h|\theta_k))} = (1 - c)\lambda z(h|\theta_k) \quad (9)$$

where the prior distribution $Z = z(h|\theta_k)$ is decided by DPMM posterior distribution $P = p(h|\theta_k)$ and edge distribution $ER(\theta_k)$. Using Bayes theorem, given prior edge region distribution and the likelihood function $p(h|\theta_k)$, the posterior distribution of h is derived as

$$z(h|\theta_k) \propto ER(\theta_k) * p(h|\theta_k) \quad (10)$$

the posterior distribution $z(h|\theta_k)$ of h is then used as the global prior distribution of BRW.

Then, the transition probability \bar{Q} on $V \cup S \cup Z$ is formulated as

$$\bar{q}(i, j) = \begin{cases} c, & \text{if } i \in V, j \in S \\ (1 - c)\frac{\lambda z(h_i|\theta_k)}{d_i + \lambda g_i}, & \text{if } i \in V, j \in Z \\ (1 - c)\frac{w_{ij}}{d_i + \lambda g_i}, & \text{if } j \sim i \in V \\ p(h_i|\theta_k), & \text{if } i = j \in V \cup S \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

where $g_i = \sum_{k=1}^K z(h_i|\theta_k)$.

Given the transition probability \bar{Q} on a graph with prior \bar{G} , the reaching probability $\bar{r}(h_i|\theta_k)$ that a random walker from a node $v_i \in V$ reaching seeds $s(n, k)$ with label k or prior node $z(h|\theta_k)$ is formulated as

$$\bar{r}(h_i|\theta_k) = \sum_{j \sim i \in V} (1 - c)\frac{w_{ij}\bar{r}(h_j|\theta_k)}{d_i + \lambda g_i} + \dots \\ + (1 - c)\frac{\lambda z(h_i|\theta_k)}{d_i + \lambda g_i} + cp(h_i|\theta_k)s(n, k). \quad (12)$$

Rewriting $[r(h_i|\theta_k)]_{N \times 1}$ as a vector $\bar{\mathbf{r}}^k$, the vector formulation can be shown as

$$\bar{\mathbf{r}}^k = (\mathbf{I} - \mathbf{D}_c)\bar{\mathbf{W}}\bar{\mathbf{r}}^k + (\mathbf{I} - \mathbf{D}_c)\bar{\mathbf{z}}^k + \mathbf{D}_c\bar{\mathbf{p}}^k\mathbf{S}^k \\ = (\mathbf{I} - (\mathbf{I} - \mathbf{D}_c)\bar{\mathbf{W}})^{-1}((\mathbf{I} - \mathbf{D}_c)\bar{\mathbf{z}}^k + \mathbf{D}_c\bar{\mathbf{p}}^k\mathbf{S}^k) \\ = \bar{\mathbf{O}}^{-1}((\mathbf{I} - \mathbf{D}_c)\bar{\mathbf{z}}^k + \mathbf{D}_c\bar{\mathbf{p}}^k\mathbf{S}^k) \quad (13)$$

where $\bar{\mathbf{O}} = (\mathbf{I} - (\mathbf{I} - \mathbf{D}_c)\bar{\mathbf{W}})$, the transition probability matrix $\bar{\mathbf{W}} = [\bar{w}_{ij}]_{N \times N}$ is defined as $\bar{w}_{ij} = \frac{w_{ij}}{d_i + \lambda g_i}$, the prior distribution $\bar{\mathbf{z}}^k = [\bar{z}(h_i|\theta_k)]_{N \times 1}$ is defined as $\bar{z}(h_i|\theta_k) = \frac{\lambda z(h_i|\theta_k)}{d_i + \lambda g_i}$ while $\bar{\mathbf{p}}^k\mathbf{S}^k = [p(h_i|\theta_k) * s(n, k)]_{N \times 1}$ indicates the extracted seeds distribution.

Then, the segmentation target can be obtained as

$$\bar{R}_i = \arg \max_{l_k} \bar{r}(h_i|\theta_k) \quad (14)$$

where \bar{R}_i is the result of single resolution segmentation under the influence of fixed prior distribution. \bar{R}_i does not only consider the similarity between adjacent positions as the classical RW algorithm, but also contain the prior distribution information determined by the image features including pixel intensity and edge region. Although the HBRW algorithm is based on RWs, it can also be interpreted as a general energy optimization problem (the optimization explanation of the proposed HBRW is given in the Appendix).

IV. EXPERIMENTATION RESULTS AND ANALYSIS

In this section, experimental results are presented to illustrate the effectiveness of the HBRW model for high-resolution remote sensing images segmentation. The compared methods include traditional RW [13], RWR [18], state-of-the-art subRW [7], NRW, and LNRW [2] as well as other probabilistic methods including probabilistic graph matching (PGM) [17], DPMM [22], and variational method (TV) [29]. The codes of these algorithms are offered by the authors and the suggested parameters in their experimentation are used as well in our experimentations. The proposed method is described in Algorithm 1. These compared methods with their corresponding parameters are described as follows.

- 1) RW: The standard RW algorithm only consider spatial information and at each pixel, the tag with the maximum probability is selected as the final tag to obtain the final segmentation results.

Algorithm 1: HBRW model.

Require: original image img , concentration parameter α , weight of seeds c , weight of prior λ , order of fractional differential v , number of iterations $Niter$.

Ensure: posteriors distribution $posteriors$, labeled image $labelimg$, number of classes K .

- 1: **Step1: Initialize**
- 2: Get histogram feature $h_i = (h_{i1}, \dots, h_{iNbin})$.
- 3: Find edge region $ER = D^v(I) - I$ using v order fractional differential.
- 4: **Step2: Hierarchical Prior**
- 5: **for** $iteration < Niter$ **do**
- 6: Generate base measure G_0 with concentration parameter α .
- 7: **for** site $i = 1, \dots, n$ **do**
- 8: Assign probability for each class by (8).
- 9: **end for**
- 10: Update the class parameters.
- 11: $iteration = iteration + 1$
- 12: **end for**
- 13: **Step3: Biased Random Walk**
- 14: Generate seeds $s(n, k)$.
- 15: Build graph \bar{G} with weight matrix W and compute transition matrix \bar{Q} .
- 16: Compute optimal random walk path by (12) and obtain segmentation result \bar{R}_i .

- 2) *RWR*: The RW with restart algorithm not only consider spatial transition probability, but also consider the probability for random walker to restart. The restart probability is set as 0.0004.
- 3) *subRW*: Based on the former RW algorithms, the subRW algorithm take prior distribution, effect of seeds and spatial constraint into consideration. The restart probability is set as 0.0004. The parameter for unitary is set as $2e - 10$.
- 4) *NRW and LNRW*: Graph-driven NRW and LNRW are applied to image segmentation by developing diffusion processes defined on arbitrary graphs. The restart probability is set as 0.0001. Region adjacency graph are used as suggested.
- 5) *DPMM*: Nonparametric Bayesian method (DPMM) is applied to image segmentation problems for freedom of class number parameter. The concentration parameters α and number of histogram bins $Nbin$ are optimized the same as our method.
- 6) *PGM*: PGM is an unsupervised and semiautomatic image segmentation approach where the segmentation is formulated as an inference problem. λ and K are optimized for specific image.
- 7) *TV*: Variational method (TV) is applied to image segmentation by researching on total variation minimization. β is set to 0.5 as suggested.

TABLE I
PRIOR KNOWLEDGE OF REAL URBAN REMOTE SENSING IMAGES
FOR FOREGROUND OR BACKGROUND SEGMENTATION

Urban Image	Size	City
Chicago1	500*500	Chicago
Chicago2	600*400	Chicago
Chicago3	500*400	Chicago
Chicago4	500*500	Chicago
Chicago5	1392*1346	Chicago
Chicago6	1146*1318	Chicago
Vienna1	1500*1500	Vienna
Vienna2	1200*1560	Vienna

TABLE II
PRIOR KNOWLEDGE OF REAL URBAN REMOTE SENSING
IMAGES WITH MULTIPLE TARGETS

Urban Image	Size	City
Shanghai1	1000*1000	Shanghai
Shanghai2	1000*1000	Shanghai
Beijing1	1000*1000	Beijing
Beijing2	1000*1000	Beijing

A. Experimental Data

In our experiments, eight real high-resolution urban remote sensing images are utilized to demonstrate the effectiveness of the proposed method in foreground or background segmentation. As shown in Fig. 9, four of these images are simple scenes in Chicago including few buildings while the others are complicated scenes in Chicago and Vienna [20]. The detailed information about these images are depicted in Table I. As listed in Table II, another four real high-resolution urban remote sensing images in Beijing and Shanghai are utilized to demonstrate the effectiveness of our method in multiple targets segmentation.

B. Parameter Analysis

The segmentation performance is closely related to the designed architecture of our HBRW. In order to validate the proposed HBRW, we compare the segmentation results using different parameters in this paper. The fractional differential order v is discussed with comparison to other edge detection operators. Then, the number of classes K in segmentation output is influenced by two resolution parameters α and $Nbin$. At last, the influence of weight parameters λ and c is analyzed.

1) *Fractional Differential Order v* : First, the influence of the fractional differential order and its comparison with integer-order differential are analyzed. As shown in Fig. 3, fractional differential of image enhances the high-frequency edge information, which makes the edge region of original image enhanced. At the same time, the pixel intensities are kept to a certain extent by fractional differential, which makes the relative change of the gray level in the image smoothing region get nonlinear enhancement. The application of proper-order fractional differential can

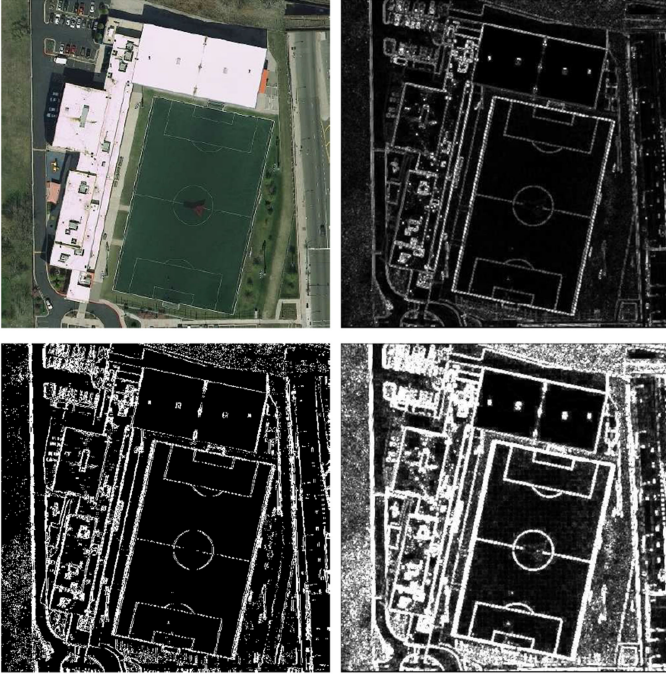


Fig. 3. Edge region of HBRW using different orders of fractional differential. Upper: original image (left) and edge region detected by 0.5 order fractional differential (right). Bottom: edge region detected by average gradient (left) and 1.5 order fractional differential (right).

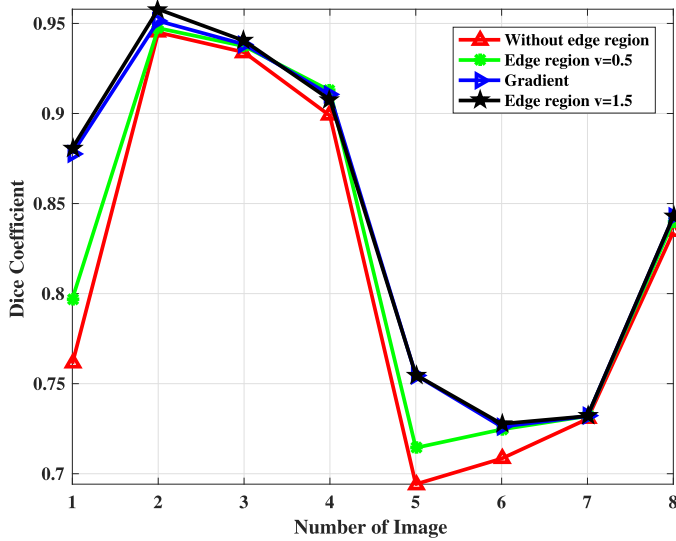


Fig. 4. Influence of with/without edge features on segmentation performance (Higher Dice coefficient means better segmentation [27]).

not only extract edge changing sharply, but also keep the continuity of edges. A larger order $v \in [0, 1]$ means approaching integer different while a small order means approaching the original signal. For a fractional order $v \in [1, 2]$, the performance of edge detection is between gradient operator and Laplacian operator. By considering the experimental analysis of edge region extracted by fractional differential, the segmentation performance of with and without edge regions in Fig. 4 illustrate that the HBRW with

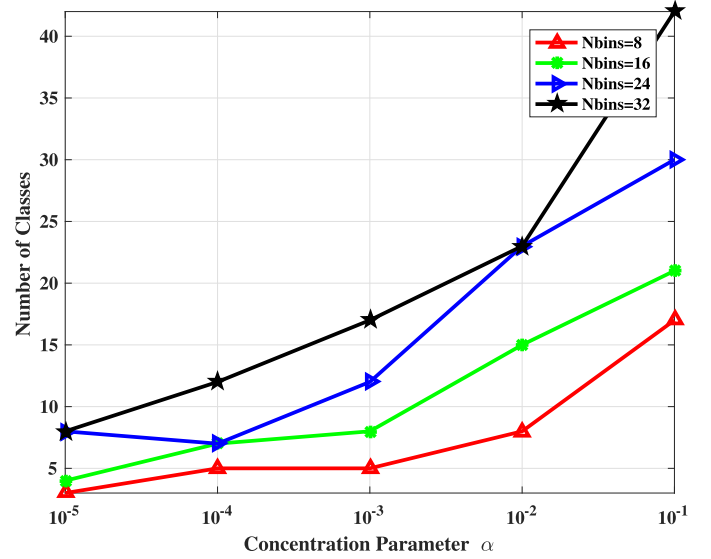


Fig. 5. Class number with different resolutions of HBRW for different images. The number K of classes (vertical) are drawn against dispersion coefficient α (horizontal), with each curve indicates numbers of pixels in each histogram.

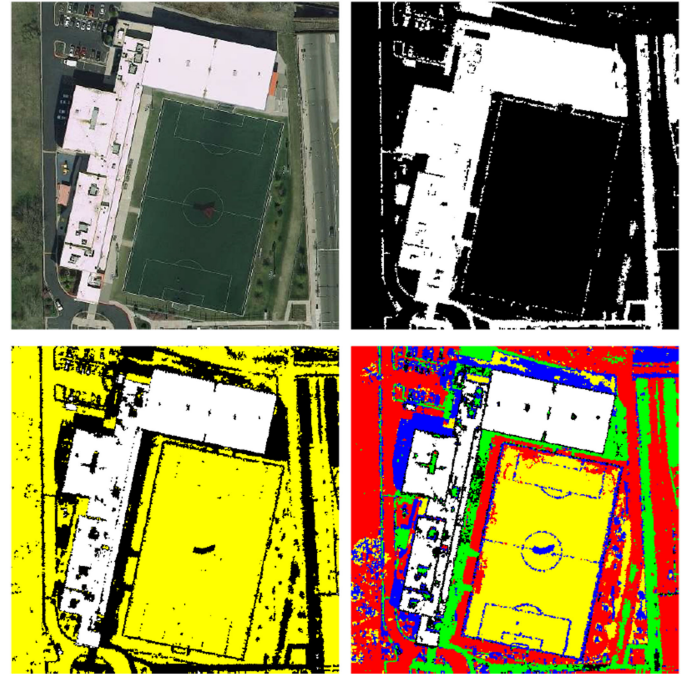


Fig. 6. Segmentation results with different concentration parameters. Upper row: Input image (left) and segmentation result for $\alpha = 1e-8$ (right). Bottom row: Segmentation results for $\alpha = 1e-6$ (left) and $\alpha = 1e-4$ (right).

edge features achieves better performance than that without edge features.

2) *Resolution Parameters α and Nbin*: As mentioned before, the segmentation is influenced by the concentration parameter α and the number of bins concluded in each histogram as well. In detail, a larger concentration parameter α lead to more classes. Here, two examples of segmentations illustrate this relationship between α and number of classes K in Fig. 5.

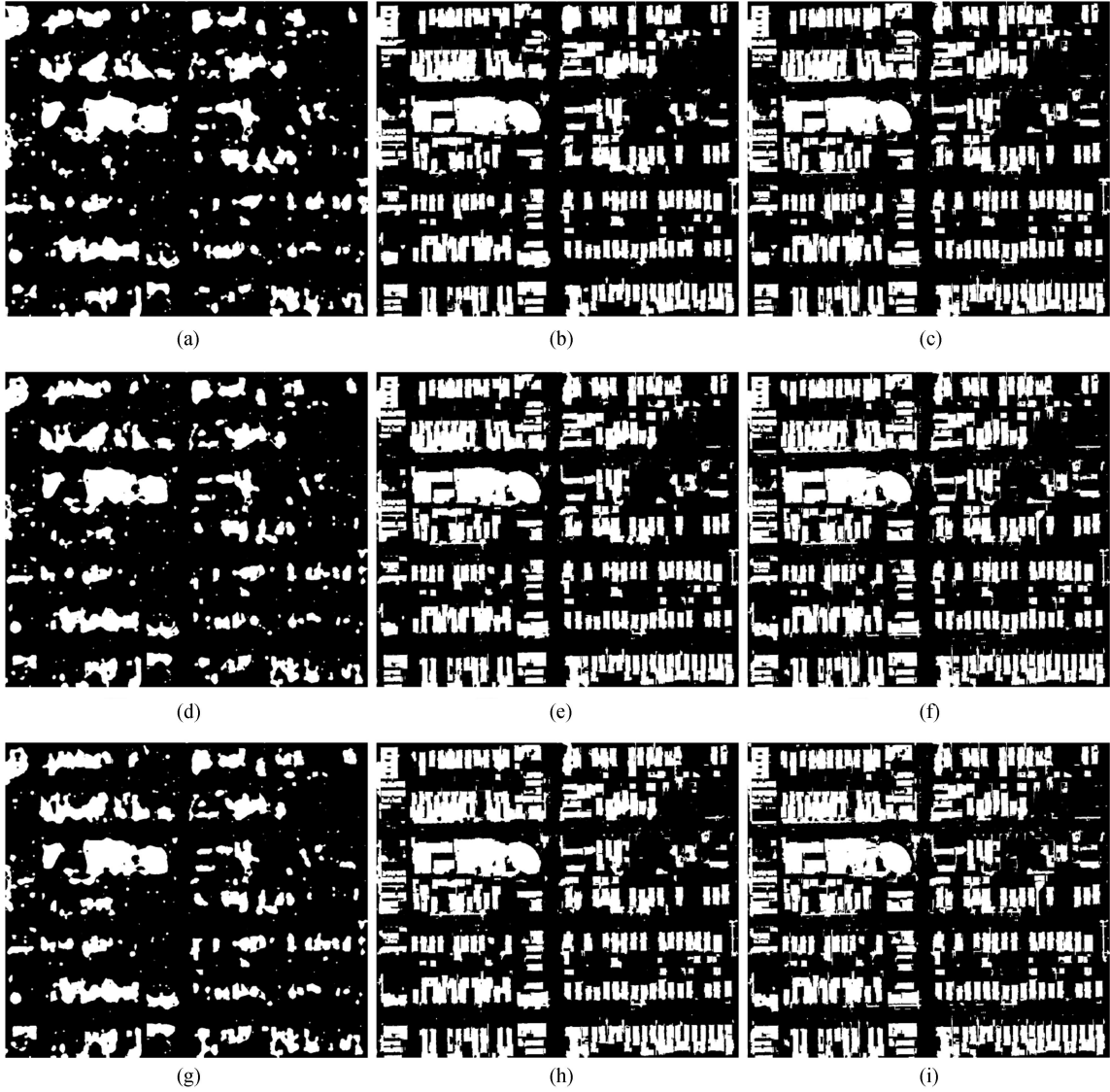


Fig. 7. Segmentation results with different weight parameters c and λ with other parameters $v = 1.5, \alpha = 1e - 5, N_{bins} = 8$ using Chicago6 as test image. (a) $c = 0.1, \lambda = 0.1$. (b) $c = 0.1, \lambda = 0.5$. (c) $c = 0.1, \lambda = 0.9$. (d) $c = 0.5, \lambda = 0.1$. (e) $c = 0.5, \lambda = 0.5$. (f) $c = 0.5, \lambda = 0.9$. (g) $c = 0.9, \lambda = 0.1$. (h) $c = 0.9, \lambda = 0.5$. (i) $c = 0.9, \lambda = 0.9$.

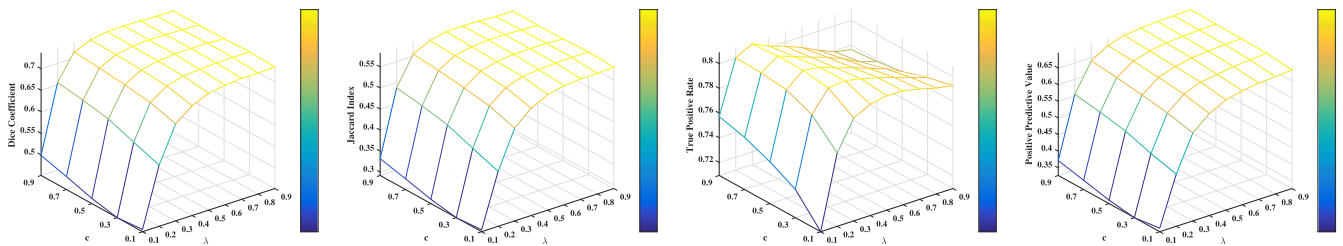


Fig. 8. Segmentation performance with different weight parameters c and λ with other parameters $v = 1.5, \alpha = 1e - 5, N_{bins} = 8$ using Chicago6 as test image.

The observations are as follows: 1) A larger value of α leads to creating more classes; 2) a tremendous value of α brings too many classes, which means part of the actual result is treated as a new class; and 3) when the value is smaller than a threshold, the number of classes will be 1, this means no new

classes are created from the initial distribution. Thus, a proper concentration parameter chosen for segmentation is rather significant. With same concentration parameter, the number of pixels concluded in each histogram also influence the class number. It is obviously depicted in Fig. 5 that more pixels concerned

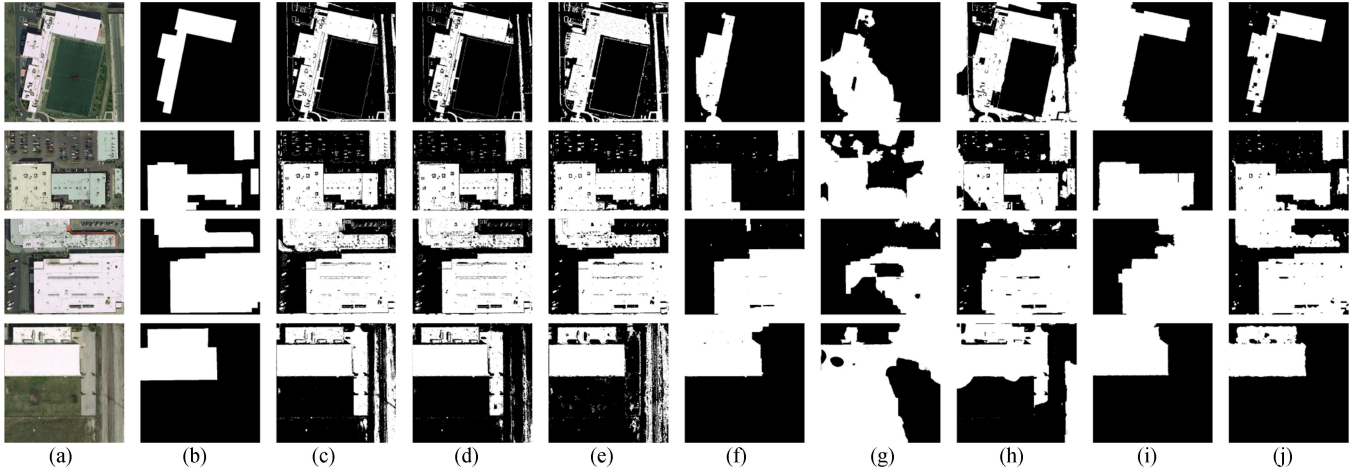


Fig. 9. Comparison results of the proposed HBRW with other algorithms for foreground/background problem of simple scenes. (a) Input urban remote sensing images. (b) Ground truth (GT), (c)–(i) are segmentation results of TV [29], PGM [17], DPMM [22], RW [13], RWR [18], subRW [7], LNRW [3]. (j) Segmentation results of the proposed HBRW method.

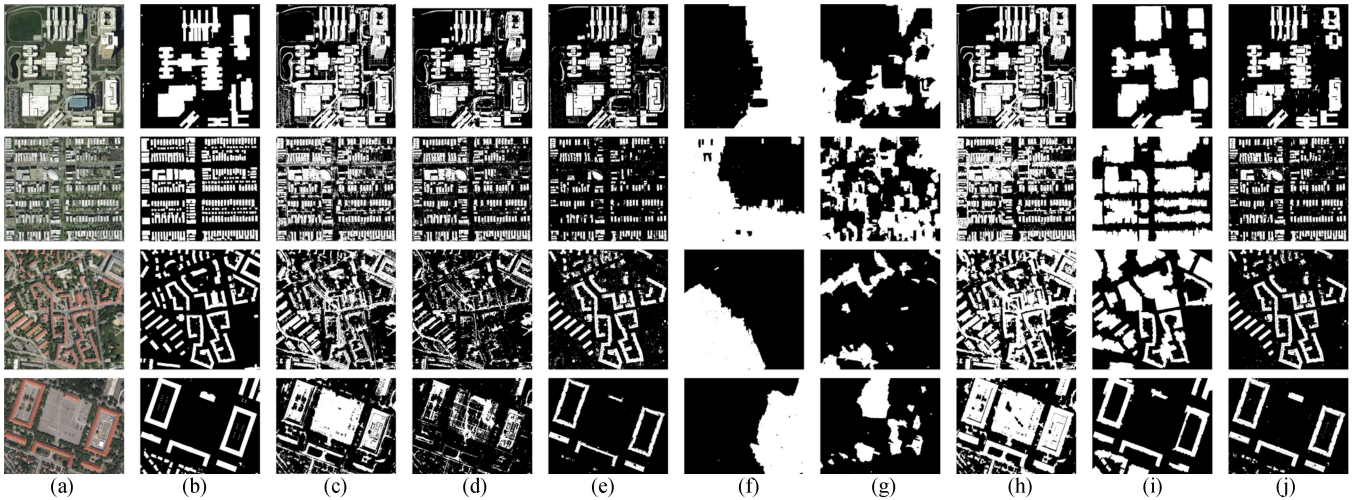


Fig. 10. Comparison results of the proposed HBRW with other algorithms for foreground/background problem of complicated scenes. (a) Input urban remote sensing images. (b) GT, (c)–(i) are segmentation results of TV [29], PGM [17], DPMM [22], RW [13], RWR [18], subRW [7], LNRW [3]. (j) Segmentation results of the proposed HBRW method.

in each histogram, more classes created in the segmentation output.

The example in Fig. 6 shows that the classes are accurately segmented under different levels of resolution (α varying from $1e-8$ to $1e-4$). When $\alpha = 1e-8$, the prior of RW is too weak so that image segments are joined erroneously. It can be seen from the images that the proposed algorithm can distinguish more details as concentration parameter growing.

3) *Weight Parameter λ and c* : The weight parameters also affect the performance of proposed HBRW. The parameter λ means the weight of prior distribution comparing with the transition probability on particular nodes. A larger λ leads to greater influence of prior distribution, which makes the segmentation containing more details as well as noise from prior distribution.

On the contrary, a small λ leads to greater influence of transition probability, which means more weights on spatial constriction for smoothing the segmentation. As shown in Fig. 7, a balance between details and smoothness is decided by λ . At the same time, another significant weight parameter c controls the weight of seeds information from the prior distribution. The proposed algorithm uses the seeds weight parameter c to control the contribution of seeds to the global optimum. A larger c means greater influence of prior fixed seeds, while smaller c means more influence caused by transition probability and prior distribution. Fig. 7 shows the segmentation results with different c and λ , which control the weight of seeds and prior distribution. To select optimal weight parameters, an extensive parameter search is conducted, as shown in Fig. 8.

TABLE III
QUANTITATIVE MEASURES FOR FOREGROUND OR BACKGROUND SEGMENTATION WITH SIMPLE SCENES

Image	Measures	Model								HBRW		
		TV	PGM	DPMM	RW	RWR	subRW	NRW	LNRW	HBRW	Parameter	Value
Chicago1	DICE	0.7091	0.7368	0.6704	0.6275	0.4178	0.5024	0.6092	0.6075	0.8808	α	$1e-7$
	JAC	0.5493	0.5833	0.5042	0.4571	0.2648	0.3354	0.4380	0.4363	0.7871	Nbin	8
	TPR	0.5974	0.6495	0.5276	0.5637	0.5478	0.9181	0.4438	0.4392	0.9228	λ	0.2
	PPV	0.8721	0.8512	0.9194	0.7075	0.3389	0.3458	0.9708	0.9846	0.8425	c	0.2
											v	1.5
Chicago2	DICE	0.9088	0.9194	0.9288	0.7400	0.6600	0.8455	0.8322	0.8312	0.9579	α	$1e-7$
	JAC	0.8329	0.8509	0.8567	0.5873	0.4925	0.7324	0.7126	0.7112	0.9193	Nbin	8
	TPR	0.8816	0.9157	0.8884	0.6012	0.8528	0.9045	0.8609	0.8588	0.9398	λ	0.2
	PPV	0.9379	0.9231	0.9600	0.9621	0.5383	0.7938	0.8054	0.8054	0.9768	c	0.2
											v	1.5
Chicago3	DICE	0.9031	0.9159	0.9255	0.7919	0.6064	0.7821	0.7662	0.7748	0.9404	α	$1e-7$
	JAC	0.8233	0.8448	0.8613	0.6555	0.4352	0.6421	0.6211	0.6324	0.8875	Nbin	8
	TPR	0.8773	0.9168	0.9339	0.7002	0.4884	0.7135	0.7851	0.7874	0.9360	λ	0.2
	PPV	0.9304	0.9150	0.9172	0.9113	0.7996	0.8652	0.7483	0.7626	0.9449	c	0.2
											v	1.5
Chicago4	DICE	0.6338	0.7149	0.7039	0.8647	0.3871	0.6936	0.8685	0.8721	0.9127	α	$1e-7$
	JAC	0.4639	0.5563	0.5431	0.7617	0.2400	0.5310	0.7676	0.7732	0.8394	Nbin	8
	TPR	0.5054	0.6275	0.6077	0.9154	0.7068	0.9096	0.8498	0.8390	0.9380	λ	0.01
	PPV	0.8498	0.8306	0.8363	0.8193	0.2665	0.5605	0.8881	0.9079	0.8887	c	0.01
											v	0.5

TABLE IV
QUANTITATIVE MEASURES FOR FOREGROUND OR BACKGROUND SEGMENTATION WITH COMPLICATED SCENES

Image	Measures	Model								HBRW		
		TV	PGM	DPMM	RW	RWR	subRW	NRW	LNRW	HBRW	Parameter	Value
Chicago5	DICE	0.6373	0.6947	0.7074	0.3647	0.1492	0.6486	0.7370	0.7386	0.7548	α	$1e-5$
	JAC	0.4677	0.5323	0.5473	0.2230	0.0806	0.4799	0.5836	0.5856	0.6062	Nbin	16
	TPR	0.5382	0.6555	0.7024	0.3093	0.1438	0.5526	0.7054	0.7233	0.7979	λ	0.5
	PPV	0.7812	0.7390	0.7125	0.4442	0.1150	0.7848	0.7717	0.7547	0.7162	c	0.4
											v	1.5
Chicago6	DICE	0.6732	0.7103	0.6346	0.4550	0.4062	0.6684	0.6590	0.6651	0.7277	α	$1e-5$
	JAC	0.5074	0.5508	0.4648	0.2945	0.2548	0.5020	0.4914	0.4982	0.5719	Nbin	16
	TPR	0.6268	0.7672	0.9580	0.3841	0.3660	0.6097	0.5376	0.5535	0.8750	λ	0.7
	PPV	0.7270	0.6613	0.4744	0.5582	0.4562	0.7397	0.8513	0.8329	0.6228	c	0.4
											v	1.5
Vienna1	DICE	0.5465	0.4086	0.7022	0.2823	0.0905	0.6266	0.6000	0.6165	0.7304	α	$1e-22$
	JAC	0.3760	0.2568	0.5411	0.1643	0.0474	0.4562	0.4285	0.4456	0.5752	Nbin	8
	TPR	0.4494	0.4156	0.7494	0.2450	0.1287	0.4713	0.4661	0.4879	0.8619	λ	0.6
	PPV	0.6973	0.4018	0.6606	0.3329	0.0698	0.9346	0.8417	0.8373	0.6336	c	0.8
											v	1.5
Vienna2	DICE	0.2740	0.1278	0.7383	0.2808	0.1081	0.4810	0.8181	0.8141	0.8428	α	$1e-22$
	JAC	0.1587	0.0682	0.5852	0.1633	0.0571	0.3166	0.6922	0.6865	0.7284	Nbin	8
	TPR	0.2035	0.1300	0.9962	0.2264	0.1128	0.3342	0.7487	0.7752	0.9494	λ	0.6
	PPV	0.4193	0.1256	0.5865	0.3695	0.1038	0.8574	0.9017	0.8571	0.7578	c	0.8
											v	1.5

Specific weight parameters are used for better segmentation performance.

C. Segmentation Performance

In this section, the segmentation performance of the proposed method is compared with other RW algorithms. First, in the foreground or background segmentation, we test the effectiveness of the proposed method on real urban high-resolution remote sensing images as shown in Figs. 9(a) and 10(a). The corresponding GT segmentation results are illustrated in Figs. 9(b) and 10(b). Figs. 9 and 10(c)–(i) provide the segmentation results of TV, PGM, DPMM, RW, RWR, subRW,

and LNRW. Figs. 9(f) and 10(f) show the segmentation results by the proposed HBRW. Comparing with other algorithms, the proposed HBRW performed better in identifying details qualitatively. It segmented out not only the main part of the target, but also edge region and twigs details. This is due to the edge detected by fractional differential.

Due to the availability of the GT segmentation, several metrics are employed for evaluation [27]. For two segmentations S_g and S_t , the confusion matrix consists of the four common cardinalities that reflect the overlap between the two segmentations, namely the true positives TP , the false positives FP , the false negatives FN , and the true negatives TN . These cardinalities provide for each pair of subsets $i \in S_g$ and $j \in S_t$; the sum of

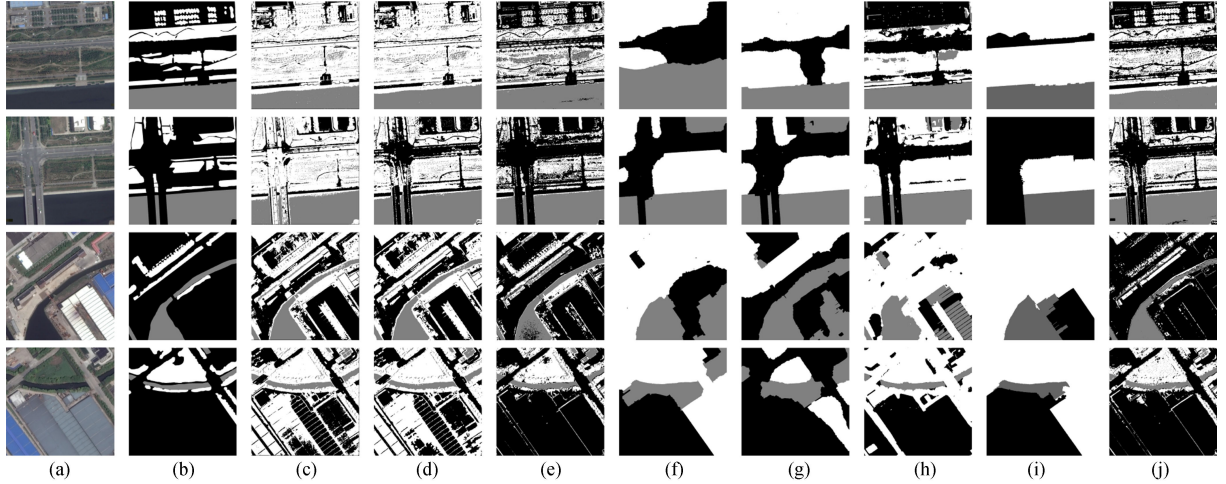


Fig. 11. Comparison results of the proposed HBRW with other algorithms for multiple targets problem. (a) Input urban remote sensing images. (b) GT, (c)–(i) are segmentation results of TV [29], PGM [17], DPMM [22], RW [13], RWR [18], subRW [7], LNRW [3]. (j) Segmentation results of the proposed HBRW method.

TABLE V
QUANTITATIVE MEASURES FOR MULTIPLE TARGETS SEGMENTATION

Image	Average Measures	Model									HBRW	
		TV	PGM	DPMM	RW	RWR	subRW	NRW	LNRW	HBRW	Parameter	Value
Shanghai1	DICE	0.6669	0.4846	0.6493	0.4986	0.5286	0.6963	0.6938	0.7604	0.7527	α	$1e-5$
	JAC	0.5143	0.6373	0.5167	0.3530	0.3881	0.5832	0.6037	0.6636	0.6080	Nbin	16
	TPR	0.5827	0.3715	0.5615	0.6572	0.5789	0.7073	0.6279	0.6714	0.8117	λ	0.6
	PPV	0.8375	0.6966	0.7791	0.4154	0.5042	0.6950	0.8154	0.9560	0.7084	c	0.5
Shanghai2	DICE	0.6757	0.5457	0.7670	0.6934	0.7379	0.7678	0.5872	0.6102	0.7533	α	$1e-5$
	JAC	0.5535	0.6755	0.6506	0.5576	0.6052	0.6553	0.4627	0.4891	0.6272	Nbin	16
	TPR	0.5977	0.4418	0.7519	0.8358	0.8724	0.9497	0.6560	0.6684	0.8616	λ	0.6
	PPV	0.8986	0.7133	0.7834	0.5957	0.6418	0.6710	0.5322	0.5654	0.6854	c	0.5
Beijing1	DICE	0.6069	0.5335	0.6244	0.3303	0.4871	0.3621	0.3226	0.3838	0.6667	α	$1e-6$
	JAC	0.4908	0.4185	0.5618	0.1998	0.3219	0.2263	0.1939	0.2475	0.5478	Nbin	16
	TPR	0.5189	0.4315	0.6259	0.2185	0.9180	0.8550	0.1967	0.2863	0.6833	λ	0.7
	PPV	0.9233	0.9409	0.8147	0.7801	0.3317	0.2482	0.9226	0.8346	0.6552	c	0.5
Beijing2	DICE	0.5739	0.3263	0.7035	0.4234	0.4306	0.6686	0.4910	0.5073	0.7277	α	$1e-6$
	JAC	0.4095	0.4106	0.5449	0.2709	0.2747	0.5194	0.3254	0.3403	0.5726	Nbin	16
	TPR	0.4302	0.2094	0.8098	0.8571	0.6694	0.8891	0.3612	0.3841	0.8507	λ	0.7
	PPV	0.9090	0.6364	0.6131	0.2841	0.3590	0.5726	0.7995	0.7989	0.6358	c	0.5
										v	1.5	

agreement m_{ij} between them is as follows:

$$m_{ij} = \sum_{r=1}^{|X|} f_g^i(x_r) f_t^j(x_r) \quad (15)$$

where $TP = m_{11}$, $FP = m_{10}$, $FN = m_{01}$, and $TN = m_{00}$. Then, we have the calculation of overlap based metrics as the Dice coefficient and the Jaccard index. The Dice coefficient (DICE) is also called the overlap index. In addition to the direct comparison between automatic and GT segmentations, it is common to use the DICE to measure reproducibility. The DICE is shown as

$$\text{DICE} = \frac{2|S_g^1 \cap S_t^1|}{|S_g^1| + |S_t^1|} = \frac{2TP}{2TP + FP + FN}. \quad (16)$$

Similarly, the Jaccard index (JAC) is defined as the intersection between the two segmentations divided by their union

$$\text{JAC} = \frac{|S_g^1 \cap S_t^1|}{|S_g^1 \cup S_t^1|} = \frac{TP}{TP + FP + FN}. \quad (17)$$

A higher score indicates better segmentation performance. These metrics are effective for objective evaluation. Some specific measures for sensitivity and precision evaluation are needed. True positive rate (TPR) measures the portion of positive pixels in the GT that are also identified as positive by the segmentation being evaluated, which is defined as

$$\text{TPR} = \frac{TP}{TP + FN}. \quad (18)$$

TABLE VI
EXECUTION TIME (IN SECONDS) IN THE THREE EXPERIMENTAL DATA SETS

Dataset	Table.I	Table.II	Table.I
Image Size	500*500	1000*1000	1500*1500
TV	0.37	1.53	2.39
PGM	24.88	640.39	850.22
DPMM	42.13	205.65	367.56
RW	1.44	7.92	12.09
RWR	1.66	8.32	19.54
subRW	1.72	10.25	17.861
NRW	21.96	78.37	333.25
LNRW	27.15	78.37	371.88
HBRW	44.75	217.84	394.74

Another related measure is the precision, also called the positive predictive value (PPV), which is defined as

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \quad (19)$$

The quantitative comparison shown in Tables III and IV depicts that the improvements over other RW algorithms are evident. To obtain proper number of classes under similar scenes, fixed resolution parameter α and Nbins are used. To improve segmentation performance, optimal weight and fractional-order parameters of the proposed model are founded by an extensive parameter search. Specific parameters are listed in Tables III and IV for clarity.

Another important segmentation problem is the multiple target segmentation, we choose the real urban remote sensing images including two or more objects as shown in Fig. 11(a) with their detail information listed in Table II. The corresponding GT maps are shown in Fig. 11(b). Fig. 11(c)–(i) provides the segmentation results of TV, PGM, DPMM, RW, RWR, subRW, and LNRW. Fig. 11(j) illustrates the segmentation results by the proposed method, which performs better in identifying details for multilabel problem qualitatively. In quantitative comparisons, the measures in Table V also confirm the significant improvement of the proposed algorithm.

The computational complexity of the aforementioned RW methods are summarized in Table VI. All the average run time in Table VI are measured in seconds with our MATLAB implementation. The configuration of the used PC is Intel Core i7-6700 CPU @3.40 GHz with 16 GB RAM. The proposed approach runs slower than other RW algorithms due to additionally building label prior and optimization process of DPMM. Longer run time is a shortcoming of the proposed HBRW algorithm.

V. CONCLUSION

In this paper, a novel HBRW model was proposed for segmentation of urban high-resolution remote sensing images. This model can be explained as a traditional random walker that walks on the graph by adding hierarchical prior distributions. While the region information was obtained from histograms and edge information by fractional differential, the prior distribution was

computed with proper resolution parameters. The experiment results demonstrated that the proposed HBRW performed well on and was proved applicability and practicability in complex images segmentation. Some other quantitative analysis confirmed its effectiveness, especially when compared with other state-of-the-art algorithms.

APPENDIX

OPTIMIZATION EXPLANATION OF HBRW

We describe the optimization explanation of the proposed HBRW with more details for completeness. Take a fixed pair of nodes i and j as an example, a random walker is considered start from node i and finally reached node j . And let R be the set of all paths from nodes i to nodes j in the expanded graph with prior \bar{G} . Given L , one can describe the squared variations of the signal with respect to G using the graph Laplacian regularizer

$$h^T L h = \frac{1}{2} \sum_{(i,j) \in e} (h_i - h_j)^2 W_{i,j}. \quad (20)$$

Assuming the whole path r energy is $E(r)$ and the total energy is additive that for a path r , the following objective function is considered for class k

$$\begin{aligned} E(r(\theta_k)) &= \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N w_{ij} (r(h_i|\theta_k) - r(h_j|\theta_k))^2 + \dots \\ &+ \frac{1}{2} \sum_{i=1}^N \frac{(d_i + \lambda g_i)c}{1-c} (r(h_i|\theta_k) - z(h_i|\theta_k))^2 + \dots \\ &+ \frac{1}{2} \sum_{i=1}^N \lambda z(h_i|\theta_k) (r(h_i|\theta_k) - 1)^2 + \dots \\ &+ \frac{1}{2} \sum_{m=1, m \neq k}^K \sum_{i=1}^N \lambda z(h_i|\theta_m) r(h_i|\theta_k)^2. \end{aligned} \quad (21)$$

Correspondingly, with graph Laplacian regularizer in (20), the vector formulation of above equation is

$$\begin{aligned} \mathbf{E}(\mathbf{r}^k) &= \frac{1}{2} [\mathbf{r}^k]^T (\mathbf{D} - \mathbf{W}) \mathbf{r}^k + \dots \\ &+ \frac{1}{2} (\mathbf{r}^k - \mathbf{z}^k)^T \frac{(\mathbf{D} + \lambda \mathbf{D}_g) \mathbf{D}_c}{\mathbf{I} - \mathbf{D}_c} (\mathbf{r}^k - \mathbf{z}^k) + \dots \\ &+ \frac{\lambda}{2} (\mathbf{r}^k - \mathbf{I})^T \mathbf{D}_z^k (\mathbf{r}^k - \mathbf{I}) + \frac{\lambda}{2} \sum_{m=1, m \neq k}^K [\mathbf{r}^k]^T \mathbf{D}_z^m \mathbf{r}^k. \end{aligned} \quad (22)$$

In order to get the minimum energy of paths in every class, the partial derivative of \mathbf{r}^k is taken

$$\begin{aligned} \frac{\partial \mathbf{E}(\mathbf{r}^k)}{\partial \mathbf{r}^k} &= (\mathbf{D} - \mathbf{W}) \mathbf{r}^k + \frac{(\mathbf{D} + \lambda \mathbf{D}_g) \mathbf{D}_c}{\mathbf{I} - \mathbf{D}_c} (\mathbf{r}^k - \mathbf{z}^k) + \dots \\ &+ \lambda \left(\mathbf{D}_z^k \mathbf{r}^k + \sum_{m=1, m \neq k}^K \mathbf{D}_z^m \mathbf{r}^k - \mathbf{z}^k \right). \end{aligned} \quad (23)$$

Setting $\frac{\partial E(\mathbf{r}^k)}{\partial \mathbf{r}^k} = 0$, the optimal solution is obtained for every class, which is the same with the result of HBRW in (13). From the representation of energy, we can see four parts of energy constitute it. The first component represent the energy of unbiased RW as a smoothing term. Minimizing this part means keeping nodes gather to the prior nodes. The second part can be considered as the energy of random walker reaching seeds, which is decided by prior distribution. Minimizing this part means keeping nodes gather to the prior nodes. The third part and last part respectively indicate reaching probabilities to nodes in the prior distribution of current class labels and labels for other classes.

REFERENCES

- [1] Y. Azar, A. Z. Broder, A. R. Karlin, N. Linial, and S. Phillips, "Biased random walks," in *Proc. 24th Annu. ACM Symp. Theory Comput.*, 1992, pp. 1–9.
- [2] C. G. Bampis, P. Maragos, and A. C. Bovik, "Graph-driven diffusion and random walk schemes for image segmentation," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 35–50, Jan. 2017.
- [3] C. G. Bampis, P. Maragos, and A. C. Bovik, "Graph-driven diffusion and random walk schemes for image segmentation," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 35–50, Jan. 2017.
- [4] X. Cai, H. Wang, H. Huang, and C. Ding, "Simultaneous image classification and annotation via biased random walk on tri-relational graph," in *Proc. 12th Eur. Conf. Comput. Vis.*, 2012, pp. 823–836.
- [5] F. R. K. Chung, *Spectral Graph Theory*, 92 of CBMS Regional Conference Series. Providence, RI, USA: Am. Math. Soc., 1997.
- [6] C. Couprie, L. Grady, L. Najman, and H. Talbot, "Power watershed: A unifying graph-based optimization framework," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 7, pp. 1384–1399, Jul. 2011.
- [7] X. Dong, J. Shen, L. Shao, and L. Van Gool, "Sub-Markov random walk for image segmentation," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 516–527, Feb. 2016.
- [8] J. Fan, M. Han, and J. Wang, "Single point iterative weighted fuzzy c-means clustering algorithm for remote sensing image segmentation," *Pattern Recognit.*, vol. 42, no. 11, pp. 2527–2540, 2009.
- [9] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.*, vol. 59, no. 2, pp. 167–181, 2004.
- [10] T. S. Ferguson, "A bayesian analysis of some nonparametric problems," *Ann. Statist.*, vol. 1, no. 2, pp. 209–230, 1973.
- [11] R. Gaetano, G. Scarpa, and G. Poggi, "Hierarchical texture-based segmentation of multiresolution remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2129–2141, Jul. 2009.
- [12] L. Grady, "Multilabel random walker image segmentation using prior models," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2005, vol. 1, pp. 763–770.
- [13] L. Grady, "Random walks for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1768–1783, Nov. 2006.
- [14] L. Grady and G. Funka-Lea, "Multi-label image segmentation for medical applications based on graph-theoretic electrical potentials," in *Proc. ECCV Workshops CVAMIA MMBIA*, 2004, vol. 3117, pp. 230–245.
- [15] L. Grady and A. K. Sinop, "Fast approximate random walker segmentation using eigenvector precomputation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2008, pp. 1–8.
- [16] L. J. Grady and J. R. Polimeni, *Discrete Calculus: Applied Analysis on Graphs for Computational Science*. New York, NY, USA: Springer, 2010.
- [17] A. Heimowitz and Y. Keller, "Image segmentation via probabilistic graph matching," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4743–4752, Oct. 2016.
- [18] T. H. Kim, K. M. Lee, and S. Uk Lee, "Generative image segmentation using random walks with restart," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 264–275.
- [19] D. Li, G. Zhang, Z. Wu, and L. Yi, "An edge embedded marker-based watershed algorithm for high spatial resolution remote sensing image segmentation," *IEEE Trans. Image Process.*, vol. 19, no. 10, pp. 2781–2787, Oct. 2010.
- [20] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Can semantic labeling methods generalize to any city? The inria aerial image labeling benchmark," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2017, pp. 3226–3229.
- [21] B. Mathieu, P. Melchior, A. Oustaloup, and C. Ceyral, "Fractional differentiation for edge detection," *Signal Process.*, vol. 83, no. 11, pp. 2421–2432, 2003.
- [22] P. Orbanz and J. M. Buhmann, "Nonparametric bayesian image segmentation," *Int. J. Comput. Vision*, vol. 77, nos. 1–3, pp. 25–45, 2008.
- [23] J. Shen, Y. Du, W. Wang, and X. Li, "Lazy random walks for superpixel segmentation," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1451–1462, Apr. 2014.
- [24] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.
- [25] A. K. Sinop and L. Grady, "A seeded image segmentation framework unifying graph cuts and random walker which yields a new algorithm," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, 2007, pp. 1–8.
- [26] W. Song, M. Li, P. Zhang, Y. Wu, L. Jia, and L. An, "Unsupervised polar image classification and segmentation using dirichlet process mixture model and Markov random fields with similarity measure," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 8, pp. 3556–3568, Aug. 2017.
- [27] A. A. Taha and A. Hanbury, "Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool," *BMC Med. Imag.*, vol. 15, no. 1, 2015, Art. no. 29.
- [28] O. J. Tobias and R. Seara, "Image segmentation by histogram thresholding using fuzzy sets," *IEEE Trans. Image Process.*, vol. 11, no. 12, pp. 1457–1465, Dec. 2002.
- [29] M. Unger, T. Pock, W. Trobin, D. Cremers, and H. Bischof, "Tvseg - interactive total variation based image segmentation," in *Proc. Brit. Mach. Vis. Conf.*, 2008, pp. 40.1–40.10.
- [30] M. Zhang, W. Li, and Q. Du, "Diverse region-based cnn for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2623–2634, Jun. 2018.



Xudong Zhao (S'19) received the B.S. degree in science and technology from the Department of Electronic Information, Beijing Institute of Technology (BIT), Beijing, China, in 2016. He is currently working toward the Ph.D. degree in information and communication engineering from BIT, under the supervision of Dr. R.Tao.

His research interests include fractional Fourier transform, graph theory, and remote sensing image process.



Ran Tao (M'00–SM'04) received the B.S. degree from Electronic Engineering Institute, PLA, Hefei, China, in 1985, and the M.S. and Ph.D. degrees from the Harbin Institute of Technology, Harbin, China, in 1990 and 1993, respectively.

He was a Senior Visiting Scholar with the University of Michigan, Ann Arbor, MI, USA, and the University of Delaware, Newark, DE, USA, in 2001 and 2016, respectively. He is currently a Professor with the School of Information and Electronics, Beijing Institute of Technology, Beijing, China. His current research interests include fractional Fourier transform and its applications,

theory, and technology for radar and communication systems. He has authored/coauthored three books and more than 100 peer-reviewed journal articles.

Dr. Tao is a Fellow of the Institute of Engineering and Technology and the Chinese Institute of Electronics. He was a recipient of the National Science Foundation of China for Distinguished Young Scholars in 2006, and a Distinguished Professor of Changjiang Scholars Program in 2009. He has been a Chief-Professor of the Creative Research Groups, National Natural Science Foundation of China, since 2014, and he was a Chief-Professor of the Program for Changjiang Scholars and Innovative Research Team in University during 2010–2012. He is currently the Vice-Chair of the IEEE China Council. He is also the Vice-Chair of the International Union of Radio Science (URSI) China Council and a member of Wireless Communication and Signal Processing Commission, URSI. He was a recipient of the First Prize of Science and Technology Progress in 2006 and 2007, and the First Prize of Natural Science in 2013, both awarded by the Ministry of Education.



Xuejing Kang received the B.S. and M.S. degrees from the Tianjin University of Technology, Xiqing, China, in 2008 and 2012, respectively, and the Ph.D. degree from the Beijing Institute of Technology, Beijing, China.

She is currently a Lecturer with the School of Computer Science and Technology, Beijing University of Posts and Telecommunications, Beijing, China. Her research interests include fractional Fourier transform, image processing, and computer vision.



Wei Li (S'11–M'13–SM'16) received the B.E. degree in telecommunications engineering from Xidian University, Xi'an, China, in 2007, the M.S. degree in information science and technology from Sun Yat-Sen University, Guangzhou, China, in 2009, and the Ph.D. degree in electrical and computer engineering from Mississippi State University, Starkville, MS, USA, in 2012.

Subsequently, he spent one year as a Postdoctoral Researcher with the University of California, Davis, CA, USA. He is currently a Professor with the School of Information and Electronics, Beijing Institute Technology, Beijing, China. His research interests include hyperspectral image analysis, pattern recognition, and data compression.

Dr. Li is an active Reviewer for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE GEOSCIENCE REMOTE SENSING LETTERS, and the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING. He is currently serving as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS. He has served as Guest Editor for special issue of *Journal of Real-Time Image Processing, Remote Sensing*, and IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING. He received the 2015 Best Reviewer Award from IEEE Geoscience and Remote Sensing Society for his service for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.