IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 9, 2018, accepted December 27, 2018, date of publication January 10, 2019, date of current version February 4, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2891941

Automatic Image Segmentation With Superpixels and Image-Level Labels

XINLIN XIE[1], GANG XIE[1,2], XINYING XU[1], LEI CUI[1], AND JINCHANG REN[1,3]

[1]College of Information and Computer, Taiyuan University of Technology, Taiyuan 030024, China
[2]School of Electronic and Information Engineering, Taiyuan University of Science and Technology, Taiyuan 030024, China
[3]Electronic and Electrical Engineering Department, University of Strathclyde, Glasgow G1 1XQ, U.K.

Corresponding author: Gang Xie (xiegang@tyut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61603267, and in part by the Shanxi Scholarship Council of China under Grant 2016-044.

ABSTRACT Automatically and ideally segmenting the semantic region of each object in an image will greatly improve the precision and efficiency of subsequent image processing. We propose an automatic image segmentation algorithm based on superpixels and image-level labels. The proposed algorithm consists of three stages. At the stage of superpixel segmentation, we adaptively generate the initial number of superpixels using the minimum spatial distance and the total number of pixels in the image. At the stage of superpixel merging, we define small superpixels and directly merge the most similar superpixel pairs without considering the adjacency, until the number of superpixels equals the number of groupings contained in image-level labels. Furthermore, we add a stage of reclassification of disconnected regions after superpixel merging to enhance the connectivity of segmented regions. On the widely used Microsoft Research Cambridge data set and Berkeley segmentation data set, we demonstrate that our algorithm can produce high-precision image segmentation results compared with the state-of-the-art algorithms.

INDEX TERMS Image segmentation, superpixels, image-level labels, disconnected regions.

I. INTRODUCTION

Automatically and ideally segmenting the semantic region of each object in an image will greatly improve the precision and efficiency of object recognition, image classification and image semantic segmentation [1]. Unfortunately, the groupings generated by the existing image segmentation algorithms are still difficult to be completely equivalent to the semantic regions [2]. This difficulty is mainly caused by the intra-class diversity and inter-class ambiguity [3]. Specifically, the object itself contains relatively high contrast regions or the contrast between object and background is small.

To obtain the perceptual grouping with high precision and high efficiency, using superpixel as basic processing unit for image segmentation has become an important manner recently [1]. The main reason is that it can significantly reduce the computational complexity of algorithms compared with using pixel. In addition, it can also provide powerful grouping cues to guide segmentation [4]. Furthermore, superpixel can preserve the original boundary information of the object, which guarantees the precision of superpixel-based image segmentation algorithms [5]. To date, a large number of superpixel-based image segmentation algorithms have been proposed. Most of them include the stages of superpixel segmentation and superpixel merging. At the first stage, popular superpixel segmentation algorithms mainly include: Simple Linear Iterative Clustering (SLIC) [6], Entropy Rate Superpixel Segmentation (ERS) [7], MeanShift (MS) [8], Watershed [9] and Efficient Graph-Based Image Segmentation (FH) [10]. However, none of these algorithms can adaptively generate the initial number of superpixels. Therefore, it is generally necessary to artificially set a large initial superpixel number to ensure the segmentation precision, which will bring a lot of redundancy for superpixel merging [1]. At the stage of superpixel merging, various graph cuts or classifiers models are widely used to measure the similarity of superpixels. However, how to automatically determine the termination condition of superpixel merging for each image is still a thorny problem [1].

According to whether supervised cues are needed, the superpixel-based image segmentation algorithms can be divided into unsupervised methods and supervised methods. Unsupervised methods are mainly based on the framework of superpixel segmentation and graph cuts. For example, Yin *et al.* [1] uses graph cuts to maximize fuzzy

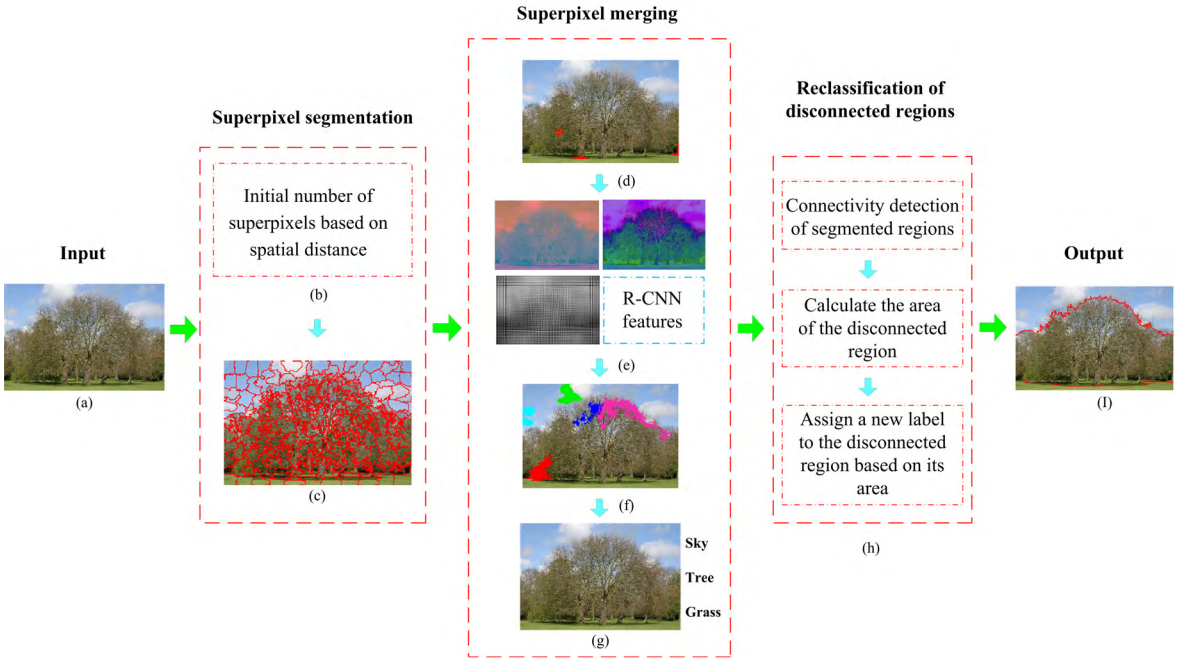


FIGURE 1. Overview of our proposed algorithm. (a) Input image. (b) Initial number of superpixels based on spatial distance. (c) LSC superpixel segmentation. (d) Merging small superpixels. (e) LAB, HSV, texture and R-CNN features map. (f) Merging similar superpixel pairs. (g) Termination condition based on image-level labels. (h) Reclassification of disconnected regions. (I) Output image.

2-partition entropy and segmentation smoothness on the basis of ERS [7]; Wang *et al.* [11] and Li *et al.* [4] both use bipartite graphs to spread grouping cues and aggregate multi-layer superpixels based on MS [8] and FH [10]. However, without the guidance of high-level cues, they are cannot achieve satisfactory segmentation results. In addition, supervised methods need to construct a classifier for training and learning. For example, Liu *et al.* [3] exploit large-margin framework and deep features generated by trained CNNs [12] for Conditional Random Fields (CRFs) learning respectively; Wang *et al.* [13] utilize a multinomial logistic regression classifier for parameter training. However, supervised methods have a strong dependence of pre-trained models, which require a large number of pixel-level labels for training.

As the most efficient form of weak supervision, image-level labels only need to give image-level cues of object and background in an image, which has attracted much attention and have been widely used for weakly-supervised image semantic segmentation [14]–[16]. The main advantage is that image-level labels can be cheaply and accurately available [17]. In addition, image-level labels can provide high-level cues for unsupervised methods, and can dramatically mitigate the time-consuming pixel annotation problem in supervised methods. In particular, the number of groupings contained in image-level labels provides an important guidance for the termination condition of superpixel merging. Unfortunately, the termination condition based on image-level labels may cause under-segmentation when multiple disjoint identical objects exit in an image or a single object has multiple disjoint regions.

Motivated by the above discussions, an automatic image segmentation algorithm using superpixels and image-level labels is proposed. The algorithm can not only generate the initial number of superpixels adaptively, but also automatically determine the termination condition of superpixel merging. The overview of the proposed algorithm is shown in Fig. 1. It includes three stages: superpixel segmentation based on spatial distance, superpixel merging based on image-level labels, and reclassification of disconnected regions. At the stage of superpixel segmentation, we first construct a spatial position matrix corresponding to the break-point pixels from color distance. Then, the initial number of superpixels is adaptively generated according to minimum spatial distance and the total number of pixels in the image. After that, the high-precision superpixel segmentation is achieved by Linear Spectral Clustering (LSC) [18]. At the stage of superpixel merging, in order to reduce the impact on final grouping number, we first define small superpixels and merge them rely on color and spatial distance. Second, feature extraction of each superpixel is performed based on the color, texture, and R-CNN features. Then, without considering the adjacency between superpixels, the most similar superpixel pairs in each iteration are merged. Finally, the superpixel merging converges until the number of superpixels is equal to the number of groupings contained in image-level labels. At the stage of reclassification of disconnected regions, we first merge smaller region into the region with the smallest spatial distance. Second, we add new label to larger disconnected regions to make it an independent region. Extensive experiments show that our algorithm can

obtain high-precision image segmentation results compared to the state-of-the-art algorithms on Microsoft Research Cambridge data set (MSRC-21) and Berkeley segmentation data set (BSDS500).

Our main innovations and contributions are as follows:

- We propose an automatic image segmentation algorithm using superpixels and image-level labels. It can adaptively generate the initial number of superpixels and automatically determine the termination condition of superpixel merging.
- We propose a method of generating the initial number of superpixels based on minimum spatial distance and the total number of pixels in the image.
- Without considering the adjacency, we automatically determine the termination condition of superpixel merging based on the number of groupings contained in image-level labels. In addition, we also define small superpixels and the number of merged superpixel pairs during the merging process.
- Extensive experimental results show that our algorithm can achieve high-precision image segmentation results on MSRC-21 and BSDS500 data set.

The organization of this paper is arranged as follows. In Section 2, we review superpixel segmentation and superpixel-based image segmentation respectively. Then, our proposed algorithm using superpixels and image-level labels is described in detail in Section 3. In Section 4, we carry out extensive experiments on MSRC-21 and BSDS500 data sets, and demonstrate the effectiveness of the proposed algorithm through visual and quantitative comparison. Finally, the conclusion is drawn in Section 5.

II. RELATED WORK

In this section, we will provide a detailed overview of superpixel segmentation and superpixel-based image segmentation. In the aspect of superpixel segmentation, we reviewed these algorithms from clustering-based methods and graph-based methods. In terms of superpixel-based image segmentation, we separately reviewed them from unsupervised methods and supervised methods.

A. SUPERPIXEL SEGMENTATION

The concept of superpixels was first proposed by Ren and Malik [19]. Afterwards, superpixel segmentation as an image preprocessing step is widely used in image segmentation, image classification, image parsing, object location and tracking [20]. At present, it can be divided into clustering-based methods and graph-based methods.

1) CLUSTERING-BASED METHODS

The clustering-based methods mainly rely on the feature of pixel for classification, which has the advantages of low computational complexity and high computational efficiency. The representative algorithms mainly include: Simple Non-Iterative Clustering (SNIC) [21], Density Based Spatial Clustering of Applications with Noise (DBSCAN) [22], Linear Spectral Clustering (LSC) [18], Simple Linear Iterative Clustering (SLIC) [6], Superpixels Extracted via Energy-Driven Sampling (SEEDS) [23], MeanShift [8], Watershed [9], whose corresponding complexities are: $O(N)$, $O(N)$, $O(N)$, $O(N)$, $O(N)$, $O(N^2)$, $O(N\log(N))$. N is the total number of pixels in the image. Among these clustering-based methods, SLIC [7] is widely used in superpixel-based image segmentation. However, it has limitations in terms of accuracy and boundary adherence, which will affect the final precision of image segmentation. For example, it is easy to produce some under-segmented superpixels. In addition, MeanShift [8] and Watershed [9] suffer from irregularly shaped superpixels, which make them difficult for the superpixel edges to coincide with the ground truth edges. Even worse, neither of them can control the initial number of superpixels. However, some high-precision algorithms proposed in recent years have not been applied to superpixel-based image segmentation widely. For example, SNIC [21] does not require an iterative process on the basis of improving segmentation precision. DBSAN [22] can find arbitrarily shaped clusters even for complex and irregularly shaped objects. LSC [18] uses kernel functions to map pixel values into a higher-dimensional feature space to improve the precision of the pixel clustering. It not only can produce compact and regular superpixels with high precision, but also has the ability to preserve the global properties of the image.

2) GRAPH-BASED METHODS

Graph-based methods see the whole image as an undirected graph. Compared with clustering-based methods, graph-based methods are less commonly used in superpixel-based image segmentation. The main reason is that, as a preprocessing step, the computational complexity of the graph-based methods is high especially when the initial number of superpixels is large. But graph-based methods generally adhere well to the boundaries of the object. These methods mainly include: Lazy Random Walks (LRW) [24], Entropy Rate Superpixel Segmentation (ERS) [7], Efficient Graph-Based Image Segmentation (FH) [10], Normalized cuts (N-cuts) [25]. Among them, LRW [24] is better for weak boundaries and the complicated texture regions, but the segmentation precision is not high. Its complexity is $O(I\cdot N^2)$, I is the number of iterations. The boundary recall of ERS [7] and FH [10] is higher, but the shape of superpixel they generated is irregular. Their complexity are $O(|V|\log(V))$ and $O(N\log(N))$, respectively. V is the vertex set. Moreover, N-cuts [25] can produce regular superpixels with a complexity of $O(N^{3/2})$, but the computational efficiency of the algorithm is unsatisfactory.

B. SUPERPIXEL-BASED IMAGE SEGMENTATION

Image segmentation using superpixel as basic processing unit has been an important image segmentation manner recently. In this section, we review them from unsupervised methods [1], [4], [11], [26]–[29] and supervised methods [2], [3], [12], [13], [30]–[37].

1) UNSUPERVISED METHODS

Unsupervised methods mainly include two stages: the generation of superpixel and the merging of superpixel based on feature clustering. In the stage of superpixel merging, it is implemented in a hierarchical manner until the termination condition is satisfied. Among these unsupervised methods, graph cuts are most widely used in superpixel merging [1], [4], [11], [26], [27]. For example, Yin *et al.* [1] use fuzzy entropy on the basis of ERS [7] to find optimal threshold; Gao *et al.* [26] propose a Graph-Without-Cut structure for learning similarity graph and image segmentations simultaneously. Wang *et al.* [11] and Li *et al.* [4] both use the bipartite graphs to spread grouping cues and aggregate multi-layer superpixels on the basis of MS [8] and FH [10]; Rantalankila *et al.* [27] use SLIC [6] to generate superpixels, built a series of superpixel graphs for graph cut during the global search phase. In addition, Peng *et al.* [28] solve the problem of merging order and termination condition in the region merging process by constructing the sequential probability ratio test and the maximum likelihood criterion on the basis of Watershed [9]. Considering natural-image segmentation as a feature clustering problem for mixed data, Yang *et al.* [29] achieve image segmentation with lossy data compression. Although unsupervised methods are computationally efficient, they produce unsatisfactory segmentation results. The main reason is that they lack guidance of high-level cues. In particularly, they cannot automatically set the threshold of termination condition for each image.

2) SUPERVISED METHODS

Among the supervised methods, training and prediction through graph model and classifier are the most important manners [2], [3], [12], [13], [30]–[35]. These methods use superpixel as basic processing unit, which significantly reduces the computational complexity of graph model and classifier during the training. However, Most of them suffer from the problem of parameter optimization. Among supervised methods, the widely used superpixel segmentation algorithms are SLIC [6] and Watershed [9]. In addition, graph model based on conditional random fields is the most widely used [3], [12], [13], [30]–[32]. For example, Fulkerson *et al.* [30] and Gould *et al.* [31] use CRFs for image segmentation earlier; Liu *et al.* [3] and Lucchi *et al.* [32] on the basis of SLIC [6] exploit the large-margin framework and structured support vector machine to learn the parameters of CRFs, respectively; Liu *et al.* [12] use the deep features generated by trained CNNs for CRFs learning and achieve high image segmentation precision; Wang *et al.* [13] utilize a multinomial logistic regression classifier for parameter training of regional-level CRFs and propose a hybrid CRFs for image segmentation. In addition, Cheng *et al.* [33] propose a real-time image segmentation system using hierarchical feature selection and fusion strategy; Farag *et al.* [34] based on cascaded superpixels and image patch labeling propose a pancreas segmentation method. Furthermore, the tree model [2] and Markov random fields [35] are still always used in image segmentation. However, these methods have a strong dependence of pre-trained graph models or classifiers, which require time-consuming and labor-intensive pixel-level annotations.

As another supervised image segmentation manner based on superpixel, interactive image segmentation is also popular. For example, Jian and Jung [36] exploit semi-supervised kernel matrix learning to adaptively propagate interactive information in entire image based on MeanShift [8]. Ning *et al.* [37] utilize markers to guide the process of merging on the basis of MeanShift [8], and propose a merging scheme based on maximal-similarity. Although the interactive image segmentation algorithm can improve segmentation precision by using markers that are labeled in advance, annotating a large number of region-level markers for large-scale data sets is also a labor-intensive task.

III. PROPOSED METHOD

Based on superpixels and image-level labels, we propose an automatic image segmentation algorithm. It solves the problem of the generation of initial superpixel number in superpixel segmentation and the determination of termination condition in superpixel merging. As is described in Fig. 1, our algorithm consists of three stages: superpixel segmentation based on spatial distance, superpixel merging based on image-level labels, and reclassification of disconnected regions. The detailed description of each stage is shown below.

A. SUPERPIXEL SEGMENTATION BASED ON SPATIAL DISTANCE

Generating superpixels with high boundary adherence is the prerequisite of superpixel-based image segmentation algorithms. Therefore, under high boundary adherence, proper initialization of the number of superpixels is critical for the computational efficiency of subsequent superpixel merging. Therefore, we propose a method of generating the initial superpixel number based on minimum spatial distance and the total number of pixels in the image.

In the field of superpixel segmentation, the initial number of superpixels K is determined by the total number of pixels in the image N and the number of pixels in each initial superpixel n:

$$K = N/n, \tag{1}$$

$$n = S' \times S', \tag{2}$$

where S' is the grid step. Thus, we can indirectly determine the initial number of superpixels for each image by finding the optimal grid step. In an image, the optimal grid step is approximately equal to the length of the circumscribed rectangle corresponding to the smallest salient object, which can be obtained from the spatial position information of the pixels in the image.

First, by calculating the color distance between adjacent pixels of each row or column, the spatial position matrix B

corresponding to the breakpoint pixels is obtained by:

$$\mathrm{B}(i,j) = \begin{cases} 1| \sum_{V \in L,A,B} |V(i,j) - V(i,j+1)| > T_1, & m \geq n \\ 1| \sum_{V \in L,A,B} |V(i,j) - V(i+1,j)| > T_1, & m < n, \end{cases} \tag{3}$$

where 1 refer to the spatial position that is a breakpoint. $V(i,j)$ is the value of LAB color component. $|\cdot|$ is an absolute value symbol. T_1 is the threshold. m and n are the rows and columns of the image respectively.

Secondly, from the spatial position matrix B, the set of breakpoint spatial positions $P(x)$ and $P(y)$ for each row and column can be calculated as follows:

$$\begin{aligned} P(x) &= \{j | B(i,j) == 1\}, \quad m \geq n \\ P(y) &= \{i | B(i,j) == 1\}, \quad m < n. \end{aligned} \tag{4}$$

After removing the adjacent breakpoint spatial positions, the minimum spatial distance D' for any row and column is obtained by:

$$D' = \begin{cases} \min(|P(x) - P(x+1)| > T_2 \& \, |P(x) - P(x+1)| \\ \quad < \sqrt{N/N_K}), \quad m \geq n \\ \min(|P(y) - P(y+1)| > T_2 \& \, |P(y) - P(y+1)| \\ \quad < \sqrt{N/N_K})), \quad m < n, \end{cases} \tag{5}$$

where T_2 is a threshold, N_K is the minimum initial number of superpixels.

Let $S' = \min\{D'\}$, the initial number of superpixels K of each image is determined indirectly:

$$K = N/(S' \times S'). \tag{6}$$

After obtaining the initial number of superpixels of each image, we use the Linear Spectral Clustering superpixel segmentation (LSC) [18] to generate superpixels. The generated superpixels can be represented as $S = \{S(k), k = 1, \ldots, K\}$, where $S(k)$ refers to the k-th superpixel and K refers to the initial number of superpixels.

B. SUPERPIXEL MERGING BASED ON IMAGE-LEVEL LABELS

In this section, we propose an automatic superpixel merging scheme based on image-level labels. Its distinctions are: 1) adding a small superpixels processing module; 2) merging superpixels without considering the adjacency; 3) using the number of groupings contained in image-level labels as the termination condition of superpixel merging.

In the iterative process, there are some small superpixels that are difficult to merge due to the larger contrast with adjacent superpixels. They affect the number of final groupings and interfere with the judgment of termination condition. Therefore, we introduce small superpixel detection and merging module during the iteration process. The small superpixel is defined as:

$$J(S(k)) = \begin{cases} 1, & N(S(k)) \leq N/(a \cdot K_r) \&\& N(S(k)) \leq t \cdot \sqrt{N} \\ 0, & \text{otherwise}, \end{cases} \tag{7}$$

where $N(S(k))$ is the number of pixels in the superpixel $S(k)$. N is the total number of pixels in the image. K_r is the number of superpixels during the iteration. a is a constant, we set $a = 10$. t is a proportional parameter. If $J(S(k)) = 1$, the neighborhood set $NS = \{S(p), p = 1, 2, \ldots m\}$ of superpixel $S(k)$ is calculated. After that, the distance D of the superpixel $S(k)$ to the neighborhood set NS is calculated by:

$$D(S(k), S(p)) = \sqrt{d_{LAB}(S(k), S(p))^2 + d_{XY}(S(k), S(p))^2}, \tag{8}$$

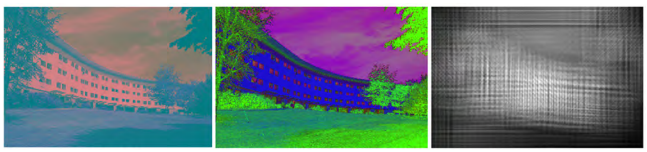
$$d_{LAB}(S(k), S(p)) = \sqrt{LAB((S(k)) - (S(p)))^2}, \tag{9}$$

$$d_{XY}(S(k), S(p)) = \sqrt{(x(S(k)) - x(S(p)))^2 + (y(S(k)) - y(S(p)))^2}, \tag{10}$$

$$S(L_k) = \{S(p) | \min(D(S(k), S(p))\}, \tag{11}$$

where $d_{LAB}(S(k), S(p))$ and $d_{XY}(S(k), S(p))$ are the color and spatial distance, respectively. $S(L_k)$ is the superpixel corresponding to the minimum distance. Finally, the superpixel $S(k)$ is merged into $S(L_k)$.

The accurate representation of superpixel feature is important for superpixel merging. Therefore, we adopt color feature representation based on LAB and HSV color space. On the other hand, texture feature based on Gabor filter bank [38] is also adopted. The reason is that color and texture are still the most basic features that characterize each superpixel. Specifically, we choose the average of each color component $\{L(k), A(k), B(k)\}$ and $\{H(k), S(k), V(k)\}$ to represent the color feature of each superpixel. The LAB and HSV color feature map is shown in Fig. 2(a) and Fig. 2(b). In texture feature extraction based on Gabor filter bank [38], we extract the average of 5 scales and 8 orientations in the frequency domain to represent the texture feature of image. Each superpixel texture feature $Gb(k)$ is represented by the average of the texture features within the superpixel. The texture feature map based on the Gabor filter bank is shown in Fig. 2(c).



(a) (b) (c)

FIGURE 2. LAB, HSV and Texture feature map.

In addition, in order to more accurately describe the appearance of superpixel, we also extract the 4096-dimensional R-CNN features of each superpixel. Instead of using bounding box, we use the resized inscribed rectangle

of superpixel as input to the AlexNet [39] pre-trained on ImageNet.

After normalizing the color, texture and R-CNN features of each superpixel, the similarity between superpixels can be defined as:

$$D(i,j) = \frac{w_1}{3} \cdot \sum |LAB(i) - LAB(i)| + \frac{w_2}{3} \cdot \sum |HSV(i) - HSV(j)| + w_3 \cdot |Gb(i) - Gb(j)| + \frac{w_4}{4096} \cdot \sum_{u=1}^{4096} |r(i) - r(j)|, \quad (12)$$

where w_1, w_2, w_3 and w_4 are the weights of LAB, HSV, texture and R-CNN features.

In order to solve the problem that there are multiple disjoint identical objects in an image or a single object has multiple disjoint region, we do not consider the adjacency between superpixels in the stage of superpixel merging. Therefore, we will directly merge the most similar superpixel pairs. The correspondence between the number of merged superpixel pairs $N(r)$ with highest similarity and current number of superpixels K_r is defined as:

$$N(r) = \begin{cases} \frac{K/b - 1}{K - T} \cdot K_r + \frac{K - K \cdot T/b}{K - T}, & K_r > c \\ 1, & K_r \leq c, \end{cases} \quad (13)$$

where K is the initial number of superpixels. K_r is the number of superpixels during the iteration. T and b are constant, c is a threshold. From Eq.(13), $N(r)$ is defined as a piecewise function. Finally, the $N(r)$ superpixel pairs with the highest similarity are merged until the termination condition is satisfied.

As a form of weak supervision, the number of groupings contained in an image provides the final superpixel number in superpixel merging. For example, if we annotate the image-level labels of {sky, tree, grass} for an image, then the final grouping number is three. Therefore, as long as the number of superpixels in the merging process is equal to the number of groupings contained in image-level labels, the termination condition converges.

C. RECLASSIFICATION OF DISCONNECTED REGIONS

Since the adjacency is not considered in superpixel merging based on image-level labels, the outputted label will contain some disconnected regions. Although these disconnected regions do not significantly affect the boundaries of the final segmentation, they can affect subsequent image processing tasks with the segmented regions as the basic processing unit.

Therefore, we will detect the connectivity of each segmented region, and reclassify the disconnected regions according to their area. First, for the smaller disconnected region, we will assign it a label corresponding to the region with the smallest spatial distance. Secondly, for the larger disconnected regions, we add new label to make it an independent region. The main steps of the whole segmentation algorithm are shown in Algorithm 1.

Algorithm 1 Automatic Image Segmentation With Superpixels and Image-Level Labels

Input: Data sets, image-level labels, parameters.
Output: Pixel labels.
Step 1: Generate the initial number of superpixels K by Eq.(3)∼Eq.(6).
Step 2: LSC superpixel segmentation, $S = \{S(k), \mathrm{k} = 1, 2, ..., K\}$.
Step 3: **While** $K_r >$ |image-level labels|
1. For each superpixel $S(k)$, if $J(S(k)) = 1$, $S(L_k) \leftarrow S(L_k) \cup S(k)$.
2. Update K_r.
3. **If** $K_r > |image - levellabels|$
4. Extract the features of each superpixel: $LAB(k)$, $HSV(k)$, $Gb(k)$, $r(k)$.
5. Calculate the similarity $D(i,j)$ by Eq.(12).
6. Calculate the number of superpixel pairs $N(r)$ to be merged by Eq.(13).
7. Merging the most similar superpixel pairs without considering the adjacency, $S(i_v) \leftarrow S(i_v) \cup S(j_v)$, $v = 1, 2, \ldots, N(r)$.
8. **End**
9. Update K_r.
End.

Step 4: Reclassification of disconnected regions.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In order to quantitatively and qualitatively demonstrate the validity of the proposed algorithm (AIS-SIL), we first give some descriptions of data sets and main parameters. Then, the initial number of superpixels of each image is given on Microsoft Research Cambridge data set (MSRC-21) [40] and Berkeley segmentation data set (BSDS500) [41], and the effectiveness of the proposed method of generating initial superpixel number is further verified on different superpixel segmentation algorithms. In addition, we compared with the state-of-the-art methods under standard metrics, and validated the performance of the proposed algorithm under the guidance of image-level labels.

A. DATA SETS AND PARAMETER SETTING

We chose the widely used MSRC-21 [40] and BSDS500 [41] to evaluate the performance of the proposed algorithm. Among them, MSRC-21 is a popular multiclass data set with 591 images, which contains 21 different categories. The BSDS500 is composed of 500 natural scene images, and each image has multiple manually labeled ground truth segmentations.

In the parameter setting, we will give the value or range of the key parameters that appear in the proposed algorithm. Among them, the ratio r_c of color similarity and space proximity in LSC [18] is set to 0.075. Moreover, thresholds T_1 and T_2 have an important effect on the calculation of the minimum spatial distance, and their ranges are

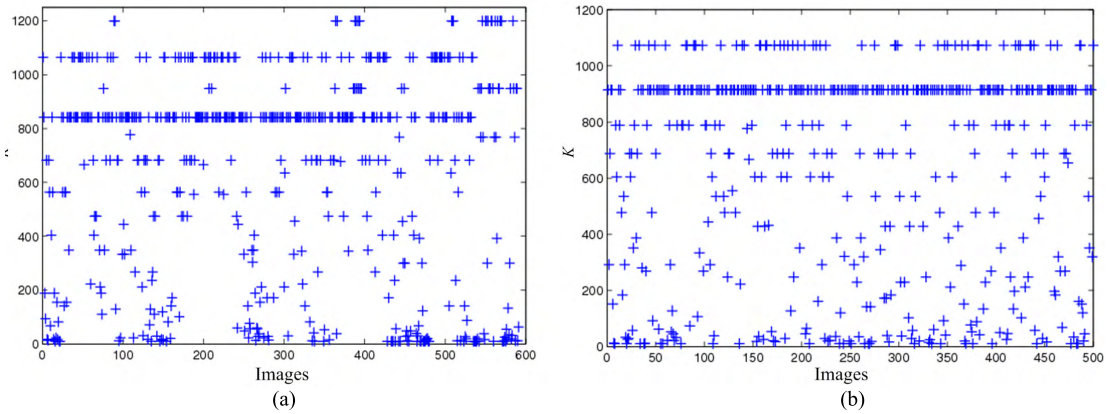


FIGURE 3. Distributions of the initial number of superpixels K.

$T_1 \in [0.5, 1.0]$, $T_2 \in [5, 15]$. $n_K = 10$. Proportional parameter in the definition of small superpixel $t \in [1, 2]$. The weight ranges for LAB, HSV, texture and R-CNN features are: $w_1, w_2, w_3, w_4 \in [0.1, 1.0]$. Furthermore, $T = 20$, $b = 25$, $c = 20$.

B. VERIFICATION OF GENERATING THE INITIAL NUMBER OF SUPERPIXELS

In this section, we use two-part experiments to verify the effectiveness of generating the initial number of superpixels based on minimum spatial distance and the total number of pixels in the image. In the first part, we give the initial number of superpixels K of each image on MSRC-21 and BSDS500 data sets to visually display the distribution of the initial number of superpixels. In the second part, under the same initial number of superpixels, the performance comparison and analysis of the representative superpixel segmentation algorithms are used to further verify the validity of the proposed method.

First, from minimum spatial distance and the total number of pixels in the image, we can obtain the initial number of superpixels K of each image On MSRC-21 and BSDS500 data sets. The distributions of the initial number of superpixels K for each image are shown in Fig. 3. The T_1 and T_2 of MSRC-21 are 0.6 and 8, and the T_1 and T_2 of BSDS500 are 0.5 and 12.

From Fig. 3, we can see that each image can generate the corresponding initial number of superpixels K according to its own characteristic on MSRC-21 and BSDS500 data sets. Furthermore, the initial superpixel number of many images falls into the lower interval, which is lower than the number of artificial settings in most superpixel-based image segmentation algorithms.

Second, under condition of the same initial number of superpixels, it is necessary to verify whether the representative superpixel segmentation algorithms can produce high-quality superpixel segmentation results. The representative superpixel segmentation algorithms mainly include: LSC [18], SNIC [21], DBSCAN [22], SLIC [6], ERS [7]. The complexity of these algorithms are: $O(N)$, $O(N)$, $O(N)$, $O(|V|\log(V))$. V is the vertex set. To evaluate the performance of these algorithms, three standard superpixel segmentation metrics were selected: Boundary Recall (BR) [18], Under-segmentation Error (UE) [6] and Achievable Segmentation Accuracy (ASA) [18]. BR is used to measure the degree of coincidence between the superpixel segmentation boundary and the ground truth boundary. UE evaluates the quality of segmentation boundary by punishing superpixel for overlapping multiple objects. The smaller the UE of the algorithm is, the higher the accuracy is. ASA is defined as the upper bound of the object segmentation accuracy that can be achieved.

Under condition of the same initial number of superpixels, the performance comparisons of the representative superpixel segmentation algorithms are shown in Table 1 and Table 2, respectively. Bold values show the best performance.

TABLE 1. Performance comparison of different algorithms on MSRC-21. BR (Boundary Recall), UE (Under-segmentation Error) and ASA (Achievable Segmentation Accuracy).

	BR	UE	ASA
ERS [7]	**0.8576**	0.0718	0.9723
SLIC [6]	0.8059	0.0798	0.9679
DBSCAN [22]	0.8263	0.1086	0.9707
SNIC [21]	0.7985	0.0841	0.9740
LSC [18]	0.8541	**0.0677**	**0.9729**

According to the comparisons shown in Table 1 and Table 2, LSC achieved the best segmentation results on both MSRC-21 and BSDS500 data sets, which represents the best segmentation performance and quality in superpixel segmentation. Therefore, under the method of generating the initial number of superpixels, it is appropriate and feasible to achieve superpixels using LSC. Meanwhile, ERS obtained close segmentation performance with LSC. Furthermore, SNIC and DBSAN also achieved competitive

TABLE 2. Performance comparison of different algorithms on BSDS500. BR (Boundary Recall), UE (Under-segmentation Error) and ASA (Achievable Segmentation Accuracy).

	BR	UE	ASA
ERS [7]	0.8908	0.2148	0.9547
SLIC [6]	0.8632	0.2432	0.9398
DBSCAN [22]	0.8481	0.2395	0.9468
SNIC [21]	0.8575	0.2291	0.9493
LSC [18]	**0.9129**	**0.2135**	**0.9560**

segmentation results on three standard metrics. However, the widely used SLIC still has much room for improvement under the three metrics compared to other superpixel segmentation algorithms. In particularly, all of these superpixel segmentation algorithms can achieve good segmentation performance, which proves that the proposed method of generating the initial number of superpixels is effective.

C. COMPARISON WITH OTHER METHODS

In order to comprehensively evaluate the effectiveness of the proposed algorithm, we compare it with other image segmentation methods on two widely-used data sets of MSRC-21 and BSDS500. Furthermore, in order to prove that the proposed algorithm can produce high-precision image segmentation results under the guidance of image-level labels, we not only give the segmentation examples on each data set, but also give a quantitative comparison under the standard metrics. The standard metrics are:

(1) Segmentation Covering (SC) [41]: It is a measure of the coincidence of proposed segmentations and ground truth segmentations. It assigns greater weight to proposed segmentations with high coincidences, and assigns smaller weight to proposed segmentations with lower coincidence. It is defined as:

$$SC(S, S_g) = \frac{1}{N} \sum_{s_i \in S} |s_i| \max_{s_j \in S_g} \frac{|s_i \cap s_j|}{|s_i \cup s_j|}, \tag{14}$$

where N is the total number of pixels in the image. S and S_g are proposed segmentations and ground truth segmentations, respectively.

(2) Variation of Information (VI) [41]: It is defined as the relative entropy between proposed segmentations and ground truth segmentations. The smaller the Variation of Information is, the greater the similarity between proposed segmentations and ground truth segmentations is. It is defined as:

$$VI(S, S_g) = H(S|S_g) + H(S_g|S), \tag{15}$$

where $H(S|S_g)$ and $H(S_g|S)$ are conditional image entropies.

(3) Probabilistic Rand Index (PRI) [41]: It is a measure of likelihood of a pair of pixels being grouped consistently between two segmentations. The higher the Probabilistic Rand Index is, the closer the proposed segmentations and ground truth segmentations are. The Probabilistic Rand Index is defined as:

$$PRI(S, S_g) = \frac{1}{T} \sum_{i<j} [c_{ij} p_{ij} + (1 - c_{ij})(1 - p_{ij})], \tag{16}$$

where c_{ij} and p_{ij} are the event that pixels i and j have the same label and its probability.

On MSRC-21 and BSDS500 data sets, all experimental results are calculated at optimal dataset scale (ODS) and optimal image scale (OIS) [41].

1) COMPARISON ON MSRC-21

In this section, we first give some segmentation examples on MSRC-21 to visually demonstrate the performance of the proposed algorithm. Since MSRC-21 is a multiclass segmentation data set, we also give the pixel-wise accuracy for each category, average per-category scores and global pixel-wise accuracy. After that, we compare and analyze with other image segmentation methods under three standard metrics to evaluate the performance of each algorithm. All the experimental results are performed on the entire MSRC-21 data set.

In the proposed image segmentation using superpixels and image-level labels, some segmentation examples are shown in Fig. 4.

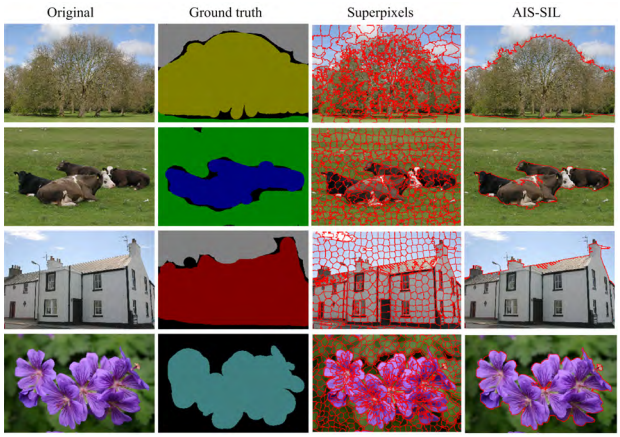


FIGURE 4. Some segmentation examples on MSRC-21.

From Fig. 4, we can find that LSC superpixel segmentation based on the method of generating the initial number of superpixels can obtain more regular superpixels, and there are no obvious under-segmentation superpixels. In addition, it can be seen from these segmentation examples that the proposed algorithm can obtain better segmentation results.

In order to more intuitively compare the performance of each algorithm on MSRC-21, the pixel-wise accuracy for each category, average per-category scores and the global pixel-wise accuracy on the entire data set of comparison algorithms are shown in Table 3.

To further illustrate the performance of the proposed algorithm, under the three standard metrics, the performance comparison of the proposed algorithm and other comparison algorithms is shown in Table 4.

TABLE 3. Segmentation results on the MSRC-21 data set (%). Bold entries are used to indicate best performance.

	building	grass	tree	cow	sheep	sky	aeroplane	water	face	car	bicycle	flower	sign	bird	book	chair	road	cat	dog	body	boat	**Average**	**Global**
Shotton et.al. [42]	49	88	79	**97**	**97**	78	82	54	**87**	74	72	74	36	24	93	51	78	75	35	66	18	67	72
Gonfaus et al. [43]	60	78	77	91	68	88	87	76	73	77	93	**97**	73	**57**	95	81	76	81	46	56	46	75	77
Lucchi et al. [32]	59	90	**92**	82	83	94	**91**	80	85	**88**	**96**	89	73	48	**96**	62	81	**87**	33	44	30	76	82
AIS-SIL	**74**	**95**	89	87	83	**96**	73	**93**	86	75	88	96	**82**	56	93	**89**	**93**	59	**72**	**78**	**56**	**82**	**88**

TABLE 4. Performance comparison on MSRC-21. SC(segmentation covering), VI (variation of information), PRI (probabilistic rand index), ODS (optimal dataset scale) and OIS (optimal image scale).

	SC		VI		PRI	
	ODS	OIS	ODS	OIS	ODS	OIS
gPb [41]	0.65	0.75	1.28	0.99	0.78	0.85
GWC [26]	0.68	0.76	1.24	0.98	0.78	0.85
HMT [2]	0.67	0.77	1.23	0.93	0.79	0.86
AIS-SIL	**0.74**	**0.79**	**1.04**	**0.75**	**0.83**	**0.88**

From Table 3, on average per-category scores and the global pixel-wise accuracy on the entire data set, the proposed algorithm achieves the best segmentation performance. Moreover, although there are some fluctuations in the pixel-wise accuracy for each category, the proposed algorithm still achieves better results. From Table 4, we can see that the proposed algorithm can also obtain high-precision segmentation performance. We attribute it to the guidance of termination condition based on image-level labels.

2) COMPARISON ON BSDS500

On BSDS500 data set, the proposed algorithm and other image segmentation algorithms are compared and analyzed under three standard metrics. All the experimental results are performed on the entire BSDS500 data set, which consists of test set, train set and validation set.

In this section, we still first give some segmentation examples of the proposed algorithm for visual verification. Some segmentation examples on the BSDS500 data set are shown in Fig. 5.

As is shown in Fig. 5, the groupings generated by the proposed algorithm can achieve a better segmentation performance. At the same time, the superpixels generated by LSC can preserve the boundary information of salient object as much as possible. However, from the last column of Fig. 5, some small objects are forced to merge. Moreover, few different objects with similar features are merged together. Therefore, we hope to introduce more semantic information in the process of image segmentation to improve its accuracy.

In addition, under the three standard metrics, the performance of the proposed algorithm and other comparison methods are shown in Table 5.

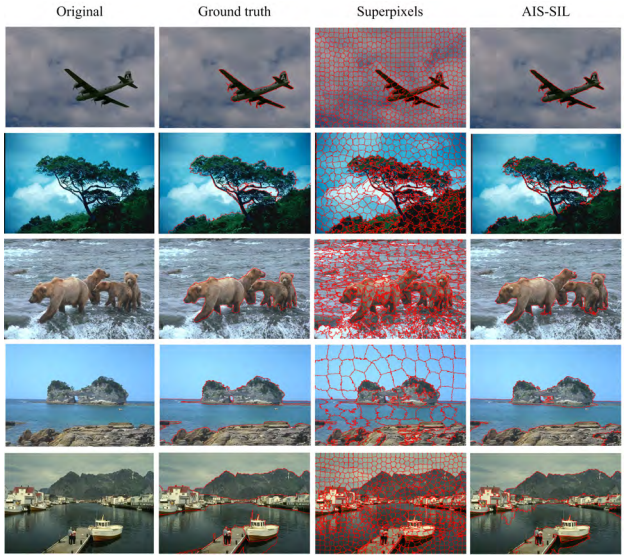


FIGURE 5. Some segmentation examples on BSDS500.

From Table 5, the proposed algorithm can also achieve better performance under three standard metrics. The performances of the proposed algorithm are improved on Segmentation Covering and Variation of Information metrics. One of the reasons lies in the high-precision superpixel segmentation boundary, making the coincidence rate of the proposed segmentations and ground truth segmentations high.

In summary, the experimental results on the MSRC-21 and BSDS500 data sets proved that the proposed method of generating the initial number of superpixels is effective, and proved that the proposed algorithm based on superpixels and image-level labels can produce high-precision image segmentation results.

TABLE 5. Performance comparison on BSDS500. SC(segmentation covering), VI (variation of information), PRI (probabilistic rand index), ODS (optimal dataset scale) and OIS (optimal image scale).

	SC		VI		PRI	
	ODS	OIS	ODS	OIS	ODS	OIS
MS [8]	0.54	0.58	1.85	1.64	0.79	0.81
FH [10]	0.52	0.57	2.21	1.87	0.80	0.82
gPb [41]	0.59	0.65	1.69	1.48	0.83	0.86
GWC [26]	0.61	0.66	1.62	1.41	0.83	0.87
HMT [2]	0.63	0.68	1.53	1.38	**0.84**	0.87
AIS-SIL	**0.70**	**0.76**	**1.36**	**1.23**	0.81	**0.87**

V. CONCLUSION

In this paper, we have proposed an automatic image segmentation algorithm based on superpixels and image-level labels. The proposed algorithm adaptively generates the initial superpixel number for superpixel segmentation, and automatically determines the termination condition of superpixel merging. In the process of superpixel merging, we dynamically define the small superpixels and the number of merged superpixel pairs. In addition, in order to avoid the under-segmentation caused by the disjoint regions in an image, we adopt a merge manner that does not consider adjacency. Furthermore, we added a module of reclassification of disconnected regions after superpixel merging to enhance the connectivity of segmented regions. Extensive validations on MSRC-21 and BSDS500 data sets show that our algorithm achieves high-precision image segmentation results compared to the state-of-the-art algorithms.

REFERENCES

[1] S. Yin, Y. Qian, and M. Gong, "Unsupervised hierarchical image segmentation through fuzzy entropy maximization," *Pattern Recognit.*, vol. 68, pp. 245–259, Aug. 2017.

[2] T. Liu, M. Seyedhosseini, and T. Tasdizen, "Image segmentation using hierarchical merge tree," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4596–4607, Oct. 2016.

[3] F. Liu, G. Lin, R. Qiao, and C. Shen, "Structured learning of tree potentials in CRF for image segmentation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 6, pp. 2631–2637, Jun. 2018.

[4] Z. Li, X. M. Wu, and S. F. Chang, "Segmentation using superpixels: A bipartite graph partitioning approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Providence, RI, USA, Jun. 2012, pp. 789–796.

[5] A. Schick, M. Bäuml, and R. Stiefelhagen, "Improving foreground segmentations with probabilistic superpixel Markov random fields," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Providence, RI, USA, Jun. 2012, pp. 27–31.

[6] 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.

[7] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Colorado Springs, CO, USA, Jun. 2011, pp. 2097–2104.

[8] 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.

[9] 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.

[10] 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.

[11] X. Wang, Y. Tang, S. Masnou, and L. Chen, "A global/local affinity graph for image segmentation," *IEEE Trans. Image Process.*, vol. 24, no. 4, pp. 1399–1411, Apr. 2015.

[12] F. Liu, G. Lin, and C. Shen, "CRF learning with CNN features for image segmentation," *Pattern Recognit.*, vol. 48, no. 10, pp. 2983–2992, Oct. 2015.

[13] F. Wang, Y. Wu, M. Li, P. Zhang, and Q. Zhang, "Adaptive hybrid conditional random field model for SAR image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 537–550, Jan. 2017.

[14] Y. Wei, H. Xiao, H. Shi, Z. Jie, J. Feng, and T. S. Huang, "Revisiting dilated convolution: A simple approach for weakly-and semi-supervised semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, May 2018, pp. 7268–7277.

[15] Z. Huang, X. Wang, J. Wang, W. Liu, and J. Wang, "Weakly-supervised semantic segmentation network with deep seeded region growing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7014–7023.

[16] F. S. Saleh, M. S. Aliakbarian, M. Salzmann, L. Petersson, J. M. Alvarez, and S. Gould, "Incorporating network built-in priors in weakly-supervised semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 6, pp. 1382–1396, Jan. 2018.

[17] X. Qi, Z. Liu, J. Shi, H. Zhao, and J. Jia, "Augmented feedback in semantic segmentation under image level supervision," in *Proc. IEEE Conf. Eur Conf. Comput. Vis.*, Sep. 2016, pp. 90–105.

[18] Z. Li and J. Chen, "Superpixel segmentation using linear spectral clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1356–1363.

[19] X. Ren and J. Malik, "Learning a classification model for segmentation," in *Proc. IEEE Int. Conf. Comput Vis.*, Nice, France, Oct. 2003, pp. 10–17.

[20] M. Wang, X. Liu, Y. Gao, X. Ma, and N. Q. Soomro, "Superpixel segmentation: A benchmark," *Signal Process. Image Commun.*, vol. 56, pp. 28–39, Aug. 2017.

[21] R. Achanta and S. Süsstrunk, "Superpixels and polygons using simple non-iterative clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 4895–4904.

[22] J. Shen, X. Hao, Z. Liang, Y. Liu, W. Wang, and L. Shao, "Real-time superpixel segmentation by DBSCAN clustering algorithm," *IEEE Trans. Image Process.*, vol. 25, no. 12, pp. 5933–5942, Dec. 2016.

[23] M. Van den Bergh, X. Boix, G. Roig, B. Capitani, and L. Van Gool, "SEEDS: Superpixels extracted via energy-driven sampling," in *Proc. IEEE Conf. Eur. Conf. Comput. Vis.*, Oct. 2012, pp. 13–26.

[24] 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.

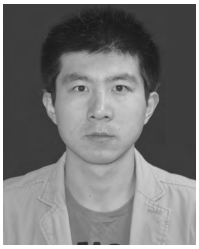
[25] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.

[26] L. Gao, J. Song, F. Nie, F. Zou, N. Sebe, and H. Shen, "Graph-without-cut: An ideal graph learning for image segmentation," in *Proc. AAAI*, 2016, pp. 1188–1194.

[27] P. Rantalankila, J. Kannala, and E. Rahtu, "Generating object segmentation proposals using global and local search," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2417–2424.

[28] B. Peng, L. Zhang, and D. Zhang, "Automatic image segmentation by dynamic region merging," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3592–3605, Dec. 2011.

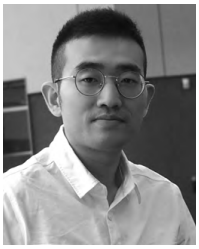
[29] A. Y. Yang, J. Wright, Y. Ma, and S. Sastry, "Unsupervised segmentation of natural images via lossy data compression," *Comput. Vis. Image Understand.*, vol. 110, no. 2, pp. 212–225, May 2008.
[30] B. Fulkerson, A. Vedaldi, and S. Soatto, "Class segmentation and object localization with superpixel neighborhoods," in *Proc. IEEE Int. Conf. Comput. Vis.*, Kyoto, Japan, Sep./Oct. 2009, pp. 670–677.
[31] S. Gould, J. Rodgers, D. Cohen, G. Elidan, and D. Koller, "Multi-class segmentation with relative location prior," *Int. J. Comput. Vis.*, vol. 80, no. 3, pp. 300–316, 2008.
[32] A. Lucchi, Y. Li, K. Smith, and P. Fua, "Structured image segmentation using kernelized features," in *Proc. IEEE Conf. Eur. Conf. Comput. Vis.*, Sep. 2012, pp. 400–413.
[33] M.-M. Cheng *et al.*, "HFS: Hierarchical feature selection for efficient image segmentation," in *Proc. IEEE Conf. Eur. Conf. Comput. Vis.*, Sep. 2016, pp. 867–882.
[34] A. Farag, L. Lu, H. R. Roth, J. Liu, E. Turkbey, and R. M. Summers, "A bottom-up approach for pancreas segmentation using cascaded superpixels and (deep) image patch labeling," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 386–399, Jan. 2017.
[35] Y. Duan, F. Liu, L. Jiao, P. Zhao, and L. Zhang, "SAR image segmentation based on convolutional-wavelet neural network and Markov random field," *Pattern Recognit.*, vol. 64, pp. 255–267, Apr. 2017.
[36] M. Jian and C. Jung, "Interactive image segmentation using adaptive constraint propagation," *IEEE Trans. Image Process.*, vol. 25, no. 3, pp. 1301–1311, Mar. 2016.
[37] J. Ning, L. Zhang, D. Zhang, and C. Wu, "Interactive image segmentation by maximal similarity based region merging," *Pattern Recognit.*, vol. 43, no. 2, pp. 445–456, Feb. 2010.
[38] D. A. Clausi and H. Deng, "Design-based texture feature fusion using Gabor filters and co-occurrence probabilities," *IEEE Trans. Image Process.*, vol. 14, no. 7, pp. 925–936, Jul. 2005.
[39] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[40] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "*TextonBoost*: Joint appearance, shape and context modeling for multi-class object recognition and segmentation," in *Proc. IEEE Conf. Eur. Conf. Comput Vis.*, May 2006, pp. 1–15.
[41] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 898–916, May 2011.
[42] J. Shotton, M. Johnson, and R. Cipolla, "Semantic texton forests for image categorization and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Anchorage, AK, USA, Jun. 2008, pp. 1–8.
[43] J. M. Gonfaus, X. Boix, J. van de Weijer, A. D. Bagdanov, J. Serrat, and J. Gonzàlez, "Harmony potentials for joint classification and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, San Francisco, CA, USA, Jun. 2010, pp. 3280–3287.



XINLIN XIE received the B.S. degree from the Polytechnic Institute, Taiyuan University of Technology, Shanxi, China, in 2012, where he is currently pursuing the Ph.D. degree. His current research interests include image semantic segmentation and granular computing.

GANG XIE received the B.S. degree in control theory and the Ph.D. degree in circuits and systems from the Taiyuan University of Technology, China, in 1994 and 2006, respectively, where he has been a Professor, since 2008. His main research interests include intelligent information processing, computer vision, and big data. He holds five invention patents. He has attained six provincial science and technology awards and has authored over 100 papers.

XINYING XU received the B.S. and Ph.D. degrees from the Taiyuan University of Technology, China, in 2002 and 2009, respectively, where he is currently an Associate Professor. His research interests include granular computing and computer vision.



LEI CUI received the B.S. degree from the College of Electrical and Power Engineering, Taiyuan University of Technology, Shanxi, China, in 2010, where he is currently pursuing the Ph.D. degree. His research topics include security and privacy issues in the IoT, computer vision, and Big Data.

JINCHANG REN received the B.E. degree in computer software, the M.Eng. degree in image processing, and the D.Eng. degree in computer vision from Northwestern Polytechnical University, Xi'an, China, and the Ph.D. degree in electronic imaging and media communication from Bradford University, U.K. He is currently with the University of Strathclyde, Glasgow, U.K., and also a Joint Distinguished Professor with the Taiyuan University of Technology. His research interests include visual computing and multimedia signal processing, especially on semantic content extraction for video analysis and understanding and, more recently, hyperspectral imaging.