

Image Segmentation with Multi-feature Fusion in Compressed Domain based on Region-Based Graph

Hong-Chuan Luo

Postgraduate Admission Office of Graduate School, Southwest University, China
luohongchuan@swu.edu.cn

Hang-Kai Zhou

Department Civil Engineering, Zhejiang University of Technology, China
2112006052@zjut.edu.cn

Bo Sun

Department Civil Engineering, Zhejiang University of Technology, China
*Corresponding Author: sunbo2017@zjut.edu.cn

Wen-Sen Cao

Department Civil Engineering, Zhejiang University of Technology, China
2112006013@zjut.edu.cn

Abstract: Image segmentation plays a significant role in image processing and scientific research. In this paper, we develop a novel approach, which provides effective and robust performances for image segmentation based on the region-based (block-based) graph instead of pixel-based graph. The modified Discrete Cosine Transform (DCT) is applied to obtain the Square Block Structures (DCT-SBS) of the image in the compressed domain together with the coefficients, due to its low memory requirement and high processing efficiency on extracting the block feature. A novel weight computation approach focusing on multi-feature fusion from the location, texture and RGB-color information is employed to efficiently obtain weights between the DCT-SBS. The energy function is redesigned to meet the region-based requirement and can be easily transformed into the traditional Normalized cuts (Ncuts). The proposed image segmentation algorithm is applied to the salient region detection database and Corel1000 database. The performance results are compared with the state-of-the-art segmentation algorithms. Experimental results clearly show that our method outperforms other algorithms, and demonstrate good segmentation precision and high efficiency.

Keywords: Image segmentation, region-based graph, multi-feature fusion, compressed domain.

Received January 8, 2021; accepted February 10, 2022
<https://doi.org/10.34028/iajit/20/2/2>

1. Introduction

Image plays an important role in the process of receiving, processing and transmitting information with the rapid development of computer technology, artificial intelligence and thinking science [39]. Digital image processing has developed rapidly in the field of computer vision [1] such as object recognition [4], object tracking and image analysis [27, 34], aiming at more accurate and comprehensive techniques. As one of the most important components in image processing, image segmentation plays a significant role in scientific research and medical image processing, e.g., Computed Tomography (CT) [37] and Magnetic Resonance Imaging (MRI) [10, 38]. Moreover, image segmentation also has vital importance for our daily life such as weather forecasting [50], traffic monitoring and recognition [35], land exploration, behavior and face recognition [26], fingerprint match and bridge crack identification [30]. The image segmentation algorithm aims at partitioning an image into several non-overlapped regions based on color, intensity, texture or some other high level semantic features [25, 36].

Typically, image segmentation can be divided into the following five categories. The first one is the

threshold based segmentation [24, 60], which usually partitions images into two parts: background and foreground. The key factor of this type is to find the optimal threshold [52, 60]. The second one is the edge-based segmentation [31, 61]. This approach assumes that the intensity values connecting the foreground and background are distinct. The discontinuity is usually detected by the first or second order derivative method like gradient [43], Laplace [17], Sobel, Roberts and Prewitt [11] edge detectors. Those detective methods are easy to be implemented and can roughly detect the contour profile. The third category is the region-based segmentation [42], typical algorithms include region growing and region splitting-merging [22]. The fourth category is the watershed-based segmentation [9, 51, 54]. This approach views an image as a topological surface and the intensity value as height. The regional minimum values of the image are interpreted as the catchment basins and the maximum values between every two neighboring catchment basins are treated as a ridge line. The watershed-based approach aims at finding this ridge line (namely watershed) within the image. The fifth category is the energy-based segmentation [6, 57]. This approach constructs an objective (energy) function which reaches a minimum

value when images are segmented as expected results. Live wire [19], Active Contour [18], Level Sets [40, 56], Graph Cuts [5, 48] and Normalized cuts (Ncuts) [49] are all grouped into this category.

During the past decades, the graph-based (energy-based) approach [21, 49] has attracted significant attentions due to its good segmentation performance. This approach maps an image into a graph and solves the segmentation problem based on the relevant graph theory. Generally, a graph is constructed corresponding to an image, with a node representing a pixel and the edge reflecting the affinity between pixels. Then, the graph is divided by finding the minimum cuts to obtain the segmentation results. The Ncuts is one commonly used method [49]. This method takes image pixels as graph nodes. However, in most cases, the widespread application of digital imaging devices results in huge image size. Taking each pixel as a graph node can bring a huge increase of graph nodes on the pixel-level graph construction. Moreover, finding the minimum cuts for the segmentation in the whole graph turns out to be the NP-complete problem [49]. Besides, the normalized cuts usually computes weights by using intensity or texture; these low-level features generally result in poor segmentation results.

The rest of the paper is structured as follows. Section 2 reviews related works on the image segmentation based on the normalized cuts, discrete cosine transform and salient region detection. Section 3 presents the graph construction process of images used in this paper. Section 4 displays processes for the energy redesign of the traditional normalized cuts and the image segmentation through eigen-system. Experimental results are shown in section 5 and general conclusions are drawn in section 6.

2. Related Work

This paper focus on relevant literatures in the field of the normalized cuts criterion and saliency region detection in the graph-based approach for the image segmentation. Lots of attempts have been made to improve the performance of the normalized cuts criterion. Zhao [59] proposes a fast normalized cuts method, which applies the Simple Linear Iterative Clustering (SLIC) algorithm to extract super-pixel regions. The super-pixel regions are treated as the graph nodes to obtain segmentation results through the normalized cuts method. These processes improve the efficiency and largely reduce the computation cost as well. However, it is still of great significance to redesign the framework of the traditional normalized cuts by considering the properties of each region. This is because in region-based graph a node may be associated with an extremely large or small region of the image; both two types of nodes are of the same importance while the normalized cuts is tailored to them. Literature [47] resolve this problem by

modifying the framework (e.g., cost function) of the normalized cuts to adapt the new requirement. Since the over segmentation algorithm yields size-different regions, it is reasonable that the size of each region is considered in the energy function to get better results. Moreover, the properties of a region are different from a pixel. It is necessary to redesign the energy function to meet the new requirements. Literature [33] propose a clustering algorithm combined with the traditional normalized cuts to solve the above problems. The algorithm divides the high-resolution image into equal size of sub-images (image cells). The normalized cuts algorithm is then performed to segment the object of every image cell (graph node). Although those algorithms can reduce the computational time, the segmentation efficiency is inferior to the state-of-the-art algorithms. Besides, the segmentation precision still remains a great challenge.

Another issue worth attention is that images are usually stored in compressed formats due to the large storing space caused by the high resolution. Most images in the internet-based applications are stored in the compressed domain of Joint Photographic Experts Group (JPEG) [8, 29]. This compressed format significantly reduces the storage space and increases the downloading speed. Features for the compressed images can be directly extracted from their compressed format by using the Discrete Cosine Transform (DCT). Recently, there are numerous approaches focus on DCT block processed images for information extraction [15, 16, 20, 41, 58]. The advantage of DCT is explained by its de-correlation property, feature preservation and reduction in complexity [59]. It is evident that the adoption of DCT-SBS, in combination with other techniques, results in good segmentation performance [53].

The saliency region detection usually refers to various feature attributes such as intensity, color, edge and shape [62]. Literature [14] proposes the Histogram-based Contrast method (HC) and Region-based Contrast method (RC). Both two methods evaluate the saliency value of an image region by its contrast with respect to the entire image. The HC method calculates the saliency value of each pixel within the image to determine whether this pixel belongs to foreground or background. Instead of using full color space, the HC method reduces the number of color channels to speed up, due to the expensive computational time even for medium sized images. The RC method improves the computation efficiency of the saliency value for each region. A mask is generated by binarizing the saliency map based on the saliency value and then used to initialize the Grabcut method [45] to finally obtain the segmentation results. However, the RC method is highly sensitive to the contrast in the visual signal. Images of complex background or low contrast between the background and foreground lead to poor segmentation results.

In this paper we employ the modified DCT to obtain the Square Block Structures (DCT-SBS) together with the coefficients, due to its low memory requirement and high processing efficiency on extracting the block feature. Then, a novel weight computation approach is proposed, combining various information from the location, texture and Red, Green, Blue Color (RGB-color). The perceptual Euclidean distance with the DCT-SBS coefficients is used to implement the extraction of the color information. The original RGB-color space is transformed into the YCrCb color space to obtain the texture information. A novel weight computation formula is developed to integrate the texture information together with the color information and location information within the block. Finally, the novel energy function is considered when optimizing the regional clustering to obtain the segmentation results.

3. Graph Construction

The general form of an undirected graph is in the following structure $G=\{V, E\}$ where V denotes the vertices and E denotes the edges. The set of edges E is a collection of local correlations between each pair of vertices. This section describes the formulation process of the DCT-SBS and the efficient calculation process of the weight for edges.

3.1. Vertices and Edges

As discussed in section 2, the traditional normalized cuts method adopts the smallest elements in the image (pixels) as vertices [49]. In the graph construction step, a full collected graph is corresponding to an image, finding the optimal cut in this graph is equivalent to the exhaustive search. We apply DCT-SBS instead of pixels to represent the graph vertices in the experiments of his paper. This is because a DCT-SBS contains several pixels, the nodes of the full collected graph can be reduced by using DCT-SBS instead. An example of graph construction is shown in Figure 1.

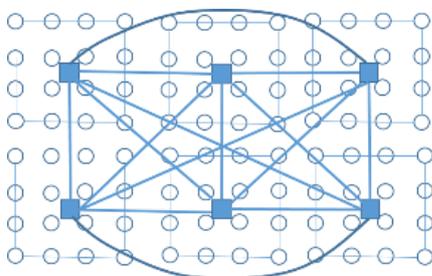


Figure 1. Full collected graph construction using DCT-SBS.

3.2. Weight

The RGB color space is adopted due to its efficiency advantage and no conversion is required. We use the perceptual Euclidean distance with DCT-SBS coefficients in the following form:

$$d_v^c(i, j) = \sum_v W_v \sqrt{(v_i - v_j)^2}, v \in \{R, G, B\} \quad (1)$$

Where $d_v^c(i, j)$ is the color dissimilarity vector between block i and j , v_i and v_j represent the constructed coefficient feature vectors in R; G; B channel as shown in Figure 2. The weight W_v for different color channel satisfies the condition $W_R + W_G + W_B = 1$.

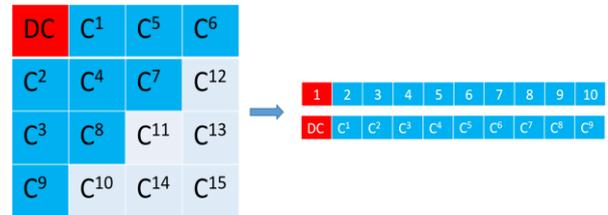


Figure 2. The process of DCT-coefficients: (Left) Blue colored AC-coefficients are chosen to construct the feature vector of each 4×4 block; (Right) Feature vectors, the first line is the index.

The zigzag storage principle of DCT coefficient is shown in Figure 3. We select 9 Alternating Current (AC) coefficients out of all 15 AC-coefficients in each block instead of using all the AC-coefficients. The Direct Current (DC) coefficient and the selected AC-coefficients are utilized to construct the feature vector of each block. This feature vector is then used to compute the weight of color information:

$$d^c(i, j) = \sum_n d_v^c(i, j)_n \times e^{\frac{1}{n}} \quad (2)$$

Where n denotes the serial number of coefficients stored in the feature vector which is shown in Figure 2. and e denotes the irrational number.

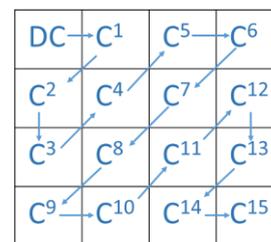


Figure 3. DCT coefficients and zigzag scanning in one 4×4 block.

Now the color dissimilarity can be transformed to color similarity between blocks using the following equation:

$$S^c(i, j) = 1 - d_{norm}^c(i, j) \quad (3)$$

Where $d_{norm}^c(i, j)$ is the normalized color information and $S^c(i, j)$ denotes the color similarity between blocks i and j .

The DCT coefficients in one block are comprised of the DC coefficient and AC coefficients. In each block, the DC coefficient is the average energy over all the 4×4 pixels while the 15 AC coefficients represent the frequency properties of the block. The AC coefficients are ordered by the zigzag scanning: most energy is included in the first several coefficients (in the left-

upper corner) and the coefficients in the right-bottom corner are almost close to zero. Thus, we define Low Frequency (LF), Medium Frequency (MF) and High Frequency (HF) zones for the left-upper corner and simply neglected the coefficients in the right-bottom corner, as show in Figure 4.

DC	LF	MF	HF
LF	MF	HF	
MF	HF		
HF			

Figure 4. Using AC-coefficients to construct texture feature.

Previous studies have shown that the AC coefficients can be used to represent the texture information for image blocks in YCrCb color space [13]. Hence, we adopt the YCrCb color space when extracting the texture feature for each 4×4 block. The Cr and Cb components of the YCrCb color space represent the color information and their AC coefficients provides little information for the texture. Thus, we use the AC coefficients from the Y component only to extract the texture feature. For the AC coefficients in one 4×4 DCT block, the LF components capture most part of the detailed information while the HF components include little information. According to [13, 23], the coefficients in each LF, MF, HF parts are summed as one value to obtain three corresponding elements to represent the texture feature T for each DCT block:

$$T = \{t^{LF}, t^{MF}, t^{HF}\} \quad (4)$$

Where t^{LF} , t^{MF} , t^{HF} is the sums of all the coefficients in the LF, MF and HF parts, respectively.

After the texture feature of each DCT block is obtained, the Euclidean distance is adopted to calculate the difference between two texture feature vectors from different blocks. The Euclidean distance has widely been used to calculate the dissimilarity between pixels or pixel sets in image processing. Since the texture feature is expressed by low/medium/high frequency, we use the Euclidean distance to calculate the difference of low frequency, medium frequency and high frequency respectively. Thus, the texture dissimilarity $d'(i, j)$ between two blocks i and j can be calculated as follows:

$$d'(i, j) = \sum_F \alpha^F \sqrt{(t_i^F - t_j^F)^2}, F \in \{LF, MF, HF\} \quad (5)$$

Where t_i^F and t_j^F denote the components of T_i and T_j . F is the calculation coefficient where $\alpha^{LF} + \alpha^{MF} + \alpha^{HF} = 1$. Similar to color information, the texture information is obtained by the following equation:

$$S'(i, j) = 1 - d'_{norm}(i, j) \quad (6)$$

Where $d'_{norm}(i, j)$ is the normalized texture information vector and $S'(i, j)$ denotes the texture similarity between blocks i and j .

The weight on each edge should reflect the likelihood that two blocks belong to one object. So in this paper the location information $S'(i, j)$ is further considered by using the following equation:

$$S^l(i, j) = \begin{cases} e^{-\frac{\|X(i) - X(j)\|^2}{\delta_x}}, & \text{if } \|X(i) - X(j)\| < T \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

Where $X(i)$ is the spatial location of block i and $X(j)$ is the spatial location of block j together with the texture information $S'(i, j)$ and color information, the semantic affinity information is calculated by the following equation:

$$w(i, j) = S^l(i, j) \times (\delta_t \times S^t(i, j) + \delta_c \times S^c(i, j)) \quad (8)$$

Where $w(i, j)$ denotes the semantic affinity between block i and j . δ_c and δ_t represent the weight of texture and color, respectively. As we can see, if the location of two blocks is large, the weight between this pair of blocks are small. Note that the weight $w(i, j)$ is equal to 0 for any pair of blocks that are more than T blocks apart.

4. Improved Normalized Cuts

4.1. Traditional Normalized Cuts

Given a graph $G=(V, E)$, the cut is defined as the total weight of the edges which has been removed between two disjoint sets A and B

$$cut(A, B) = \sum_{i \in A, j \in B} w(i, j) \quad (9)$$

Where A and B are sub-graphs with constraint of $A \cup B = V$, $A \neq \emptyset$, $B \neq \emptyset$, and $A \cap B = \emptyset$.

Graph partition cuts out edges with low values of weight, since low weight indicates low similarity between the paired pixels [49]. Hence, the optimal partition of a graph is the one that minimizes this cut value. Literature [55] proposes a clustering method based on the minimum cut criterion. This method partitions a graph into several sub-graphs such that the maximum cut across the sub-graphs is minimized [49]. However, the minimum cut criterion is a global optimal criterion for image segmentation; it is implemented while the image segmentation problem is viewed by seeking all the possible solutions. With the minimum cut values obtained, the segmentation result is formed by removing the edges indicted in Equation (9). Besides, the minimum cut criterion favors cutting small sets of isolated nodes in the graph. Figure 5 illustrates one such case. Since minimum cuts cannot give satisfactory results in image segmentation,

Literature [49] propose the normalized cuts algorithm to improve the minimum cuts. The normalized cuts algorithm alleviates the isolated pixels problem by computing the cut cost as a fraction of the total edge connections to all the nodes in the graph, i.e.

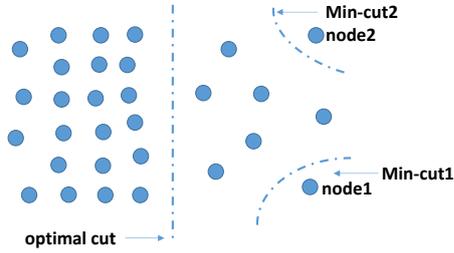


Figure 5. An example of minimum cut which gives a bad result.

$$Ncut(A, B) = \frac{cut(A, B)}{assoc(A, V)} + \frac{cut(A, B)}{assoc(B, V)} \quad (10)$$

Where $assoc(X, V) = \sum_{i \in X, j \in V} w(i, j)$ is the total connection from nodes in X to all nodes in the graph, $X=A$ or B . Traditional Ncuts problem can be solved by formulating into a generalized eigenvalue problem in the field of spectral graph theory [49].

4.2. Our Formulation: DCT-SBS based Spectral Clustering

To improve the performance of the normalized cuts criterion for the region-based graph, we try to recall that the intention of the normalized cuts results in high intra-similarities and low inter-similarities. Following these rules, we start from the energy function of the normalized cuts and define a theorem as follows:

- **Theorem:** Given a graph G with the node set $V=\{1,2,\dots,n\}$ and weight matrix W , let $H=[h_{ii}]$ denotes the diagonal matrix. We define

$$h(A) = \sum_{i,j \in A, i \neq j} \frac{1}{e^{(DC_i - DC_j)^2}} \quad (11)$$

Where A is the subset of the graph nodes, DC_i and DC_j denote the means of block i and j , respectively.

Following the definition of h , we define the energy function as follows:

$$\arg \min_{A \subset N} \frac{cut(A, \bar{A})}{h(A) \times h(\bar{A})} \quad (12)$$

Note that A and \bar{A} are two subsets of the graph node set V . Based on the following proof, the optimal solution of the energy function is corresponding to the second smallest eigenvector of the generalized eigenvalue problem.

$$(D - W)x = \lambda Hx \quad (13)$$

With x as the eigenvector and λ as eigenvalue.

- **Proof:** Given $A \subset V$, let a binary vector $f=(f_1, f_2, \dots, f_n) \in R^n$ has the following entries:

$$f_i = \begin{cases} \sqrt{\frac{h(\bar{A})}{h(A)}}, & \text{if } i \in A \\ -\sqrt{\frac{h(\bar{A})}{h(A)}}, & \text{if } i \in \bar{A} \end{cases} \quad (14)$$

The proof starts from the laplacian matrix L of graph G , where $L=D-W$. following literature [47, 49], we have

$$\begin{aligned} f'Lf &= f'Df - f'Wf = \sum_{i=1}^n d_i \cdot f_i^2 - \sum_{i,j=1}^n w_{ij} \cdot f_i f_j \\ &= \frac{1}{2} [\sum_{i=1}^n d_i \cdot f_i^2 - 2 \sum_{i,j=1}^n w_{ij} \cdot f_i f_j + \sum_{j=1}^n d_j \cdot f_j^2] \\ &= \frac{1}{2} \sum_{i,j=1}^n w_{ij} (f_i - f_j)^2 \\ &= \frac{1}{2} [\sum_{i \in A, j \notin A} w_{ij} \cdot (\sqrt{\frac{h(\bar{A})}{h(A)}} + \sqrt{\frac{h(\bar{A})}{h(A)}})^2 \\ &\quad + \sum_{i \in \bar{A}, j \notin \bar{A}} w_{ij} \cdot (-\sqrt{\frac{h(\bar{A})}{h(A)}} - \sqrt{\frac{h(\bar{A})}{h(A)}})^2] \\ &= \frac{1}{2} [cut(A, \bar{A}) \times (\frac{h(\bar{A})}{h(A)} + \frac{h(A)}{h(\bar{A})} + 2) \\ &\quad + cut(\bar{A}, A) \times (\frac{h(A)}{h(\bar{A})} + \frac{h(\bar{A})}{h(A)} + 2)] \\ &= \frac{cut(A, \bar{A}) \cdot [h(A) + h(\bar{A})]^2}{h(A) \times h(\bar{A})} \\ &= tr^2(H) \cdot \frac{cut(A, \bar{A})}{h(A) \times h(\bar{A})} \end{aligned} \quad (15)$$

Where $tr(H)$ is the trace of the matrix H . Moreover, we have

$$\begin{aligned} (Hf)' \cdot \bar{1} &= \sum_{i=1}^n h_{ii} f_i = \sum_{i \in A} h_{ii} \sqrt{\frac{h(\bar{A})}{h(A)}} - \sum_{i \in \bar{A}} h_{ii} \sqrt{\frac{h(\bar{A})}{h(A)}} \\ &= h(A) \sqrt{\frac{h(\bar{A})}{h(A)}} - h(\bar{A}) \sqrt{\frac{h(\bar{A})}{h(A)}} = 0 \end{aligned} \quad (16)$$

and

$$f'Hf = \sum_{i=1}^n h_{ii} f_i^2 = \sum_{i \in A} h_{ii} \frac{h(\bar{A})}{h(A)} + \sum_{i \in \bar{A}} h_{ii} \frac{h(A)}{h(\bar{A})} = tr(H) \quad (17)$$

Therefore, solving the energy function of Equation (17) is equivalent to find

$$\arg \min_f f'Lf \text{ s.t. } Hf \perp \bar{1} \text{ and } f'Hf = tr(H) \quad (18)$$

Note that finding the optimal solution for the above optimization problem is NP-hard, since the elements in the vector f are only allowed to take two distinct values. The optimization solution is obtained by relaxing f_i to take real value as follows [47, 49]:

$$\arg \min_{f \in R^n} f'Lf \text{ s.t. } Hf \perp \bar{1} \text{ and } f'Hf = tr(H) \quad (19)$$

Suppose $g = H^{\frac{1}{2}} f$, we have the relaxed problem

$$\arg \min_{g \in R^n, g \perp H^{\frac{1}{2}} \bar{1}, \|g\|^2 = tr(H)} g^T H^{-\frac{1}{2}} L H^{-\frac{1}{2}} g \quad (20)$$

Where $H^{-\frac{1}{2}} L H^{-\frac{1}{2}}$ is a symmetric matrix. According to

literature [47], we have the properties of the following:

- *Property 1*: The matrix $H^{-\frac{1}{2}}LH^{-\frac{1}{2}}$ is positive semidefinite.
- *Property 2*: The smallest eigenvalue of $H^{-\frac{1}{2}}LH^{-\frac{1}{2}}$ is zero and its corresponding eigenvector is $H^{-\frac{1}{2}}\mathbf{1}$.

Following Properties 1 and 2, the solution of the optimization problem in Equation (20) is the second smallest eigenvector of the eigenvalue problem:

$$H^{-\frac{1}{2}}LH^{-\frac{1}{2}}g = \lambda g \quad (21)$$

Substituting $H^{-\frac{1}{2}}f$ with g , we have $H^{-\frac{1}{2}}Lg = \lambda H^{-\frac{1}{2}}g$ and

$$Lg = \lambda Hg \quad (22)$$

Therefore, the optimum solution is the second smallest eigenvector of Equation (22).

The generalized process constructed above in sections 3 and 4 combines the multi-feature fusion method and the DCT-SBS based spectral clustering method to finish the image segmentation task. The flow chart of the proposed algorithm is shown in Figure 6.

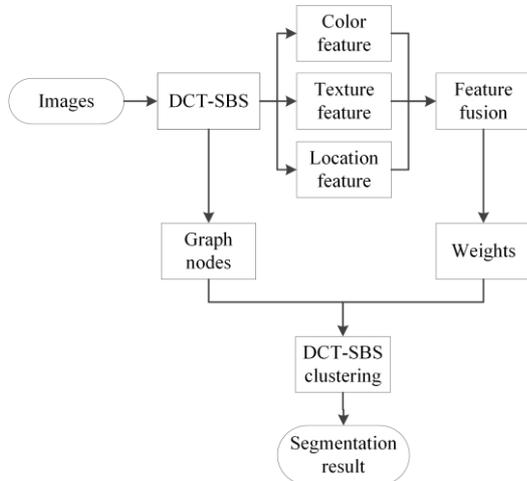


Figure 6. Flow chart: image segmentation with multi-feature in compressed domain and spectral clustering.

5. Experiments

5.1. Segmentation Benchmark

The difficulty of evaluating the segmentation results falls on the small number of the canonical test datasets available to provide ground truth segmentations. This is partly because the manual delineation of segments in complex images can be laborious. Furthermore, people tend to incorporate their subjective considerations which are beyond the scope of the data-driven segmentation Algorithms (1). In our experiments we adopt the benchmark presented by [14]. This dataset precisely marks the pixels in salient object regions and makes each image containing a clear and unambiguous

object. We evaluate the segmentation results by comparing their consistency with the ground truth segmentations. With no loss of generality, we use the average precision, recall, and F-measure [3] to do the comparison with the followings:

1. Traditional normalized cuts.
2. Mean shift.
3. Region-based contrast method, which is the most challenge algorithm evaluating the segmentation results by automatic estimation of salient regions without any prior assumption or knowledge of the corresponding scenes.

In pattern recognition and information retrieval, True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN) are 4 types of each pixel expressed in Table 1. Accuracy for a segmentation task is the number of true positives (number of pixels correctly labeled as belonging to the object class) divided by the total number of elements labeled as belonging to the positive class:

$$Precision = TP / (TP + FP) \quad (23)$$

Where TP and TN are the numbers of the true positives and true negatives, respectively. And the recall is defined as the number of true positives divided by the total number of elements that actually belong to the positive class [7]:

$$Recal = TP / (TP + FN) \quad (24)$$

Where FN indicates the number of the false negatives. Then, the F-measure is defined as

$$F_{\beta} = \frac{(1 + \beta^2)R \times P}{\beta^2 \times P + R} \quad (25)$$

Where P and R denote the precision and recall as described in Equations (23), and (24), respectively.

Table 1. Precision-recall contingency.

		Observation
Prediction		TP – Correct result
		FP – Unexpected positive result
		TN – Correct absence of result
		FN – Missing negative result

In this paper, we employ databases Salient Region Detection [14] and Corel1000 [32]. The segmentation algorithm proposed in this paper is compared with the mean shift algorithm [12] and traditional normalized cuts algorithm [49]. The intuitively segmentation results and quantitative evaluations are shown and discussed in the following.

5.2. Evaluation

Several segmentation results are chosen to illustrate that the algorithm proposed in this paper outperforms other methods. Figure 7 shows a sample result obtained by applying our algorithm to images from the

Salient Region Detection database compared to some other algorithms. The fourth column of the image matrix in Figure 7 is the most challenging part. While it has the most complex background and the shape of the object in this image is anomalous, our algorithm still detects the right object compared with the benchmark.

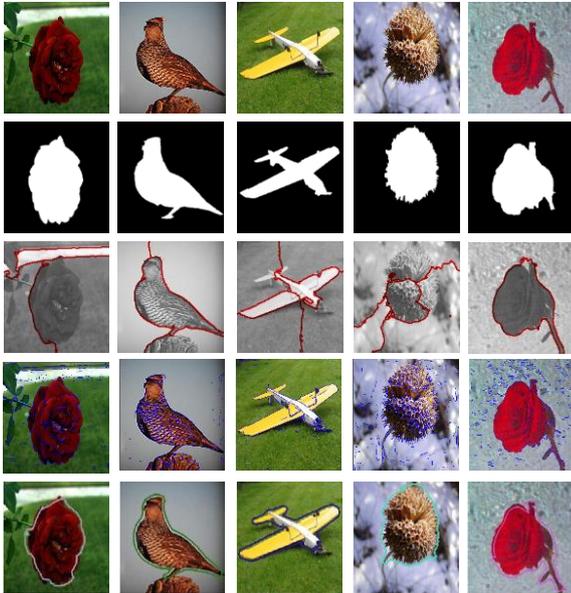


Figure 7. A sample of the results obtained by applying our algorithm to images from the salient region detection database compared to some other algorithms. From top to bottom: Original images, benchmark, Normalized Cuts, mean shift, and our method.

Since our method is based on the DCT-SBS, the final segmentation contour is made up of boundary blocks, which means our segmentation results are based on the block-level instead of pixel-level. Therefore, large size of block will lead to poor segmentation results and even cannot get the correct segmentation results as shown in Figure 8. It is clear that the block size with 4×4 can get the best segmentation results. As the block size increases, some images cannot get the correct segmentation results. This is partly because the block with larger size may contain the background and foreground which have the similar color information. Thus, the feature vector in this block is difficult to get, leading to the wrong results. Figure 8 also shows that the results with block sizes 2×2 and 4×4 is much more similar. However, the efficiency is obvious poor with respect to those block sizes, e.g., block size with 2×2 requires twice the time to finish the discrete cosine transform on each block.

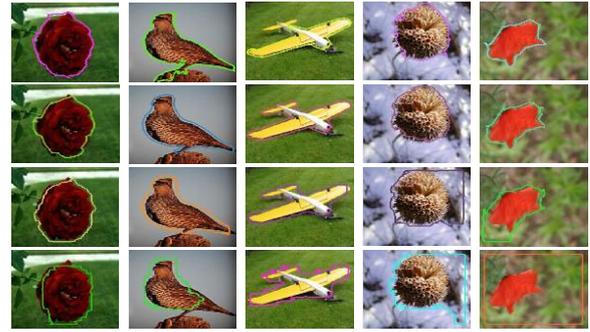


Figure 8. Different size of block result in different segmentation results, from top to bottom, block size is 2×2 , 4×4 , 6×6 , 8×8 respectively.

Figure 9 shows the average precision, recall and F-measure for different segmentation methods:

1. Gaussian blob (Gau) [44].
2. Frequency-Tuned (FT) [2].
3. Segmenting salient objects (SEG) [46].
4. GrabCut [45].
5. Context and shape prior (CB) [28].
6. Histogram based Contrast method (HC) [14]
7. Ours (method proposed in this paper).

For the parameters in our method, parameters δ_c , δ_t , β , the values of each weight in color information and texture information are shown in Table 2. β is set to be 0.3 according to [2, 14] to weigh precision more than recall. Other parameters are the default values with respect to the traditional normalized cuts. The other segmentation algorithms are running with their optimal parameters in the experiments. The test results (Figure 9) clearly shows that our method achieves the highest average recall score (0.921) and F-measure score (0.9003). The most challenging algorithm is the HC method: it achieves the best score in precision and the F-measure score (0.878) is close to our method. However, the recall score for the HC algorithm is inferior to that of our method.

Table 2. Part parameters setting in the experiments.

parameter	W_R	W_G	W_B	δ_c	δ_t
value	0.26	0.70	0.04	0.716	0.284
parameter	α^{LF}	α^{MF}	α^{HF}	β	
value	0.5	0.33	0.17	0.3	

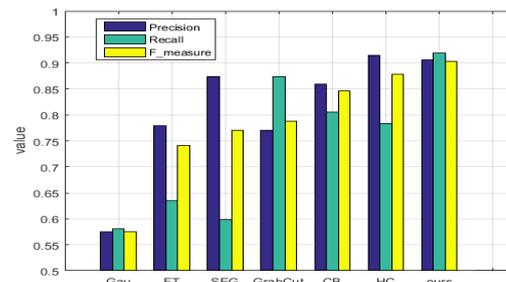


Figure 9. Comparison average precision, recall and F measure for different segmentation methods: Gaussian blob [44], frequency-tuned [2], Segmenting salient objects [46], CB [28], HC [14] and ours.

In addition to this evaluation, we also evaluate our method on the Corel1000 database [32] and compared with the mean shift algorithm and traditional normalized cuts. Sample results of applying our method to images from Salient Region Detection database and Corel1000 database are shown in Figures 10, and 11. Experimental results clearly show that our method outperforms other algorithms.

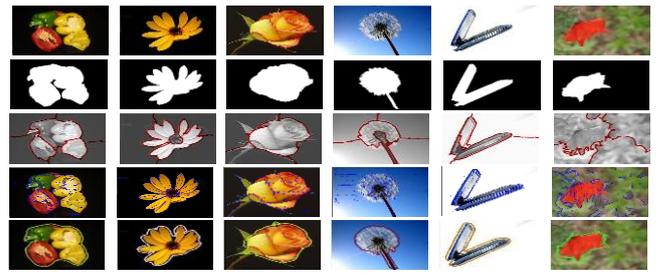


Figure 10. A sample of the results obtained by applying our algorithm to images from the Salient Region Detection database compared to some other algorithms. From top to bottom: Original images, benchmark, normalized cuts, mean shift, and our method.

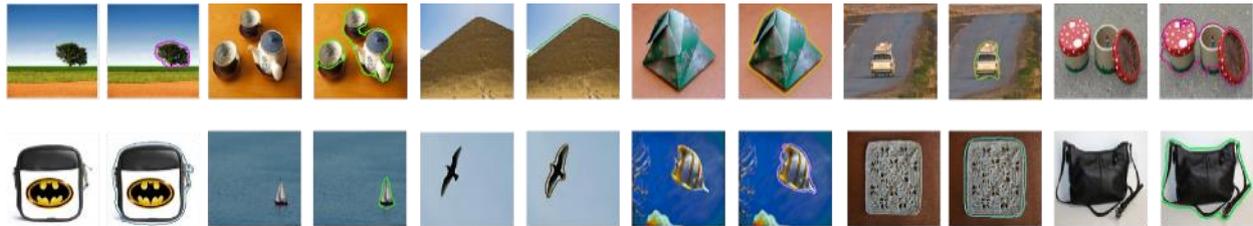


Figure 11. A sample of the results obtained by applying our algorithm to images from the Corel1000 database.

All the segmentation tasks in our experiments are implemented on a desktop equipped with a 2.7 GHz Pentium (R) Dual-Core CPU and 4GB RAM. Most programs are implemented in the MATLAB software, except some time-consuming operations in the Visual Studio 2010 software interfaced with MATLAB through mex-files. Some traditional program frameworks such as the normalized cuts and discrete cosine transform weight are available online with MATLAB version. The segmentation efficiency of our method has been largely improved compared with the traditional normalized cuts. When solving a standard eigenvalue problem, a 400×300 image in the traditional normalized cuts has 120000 graph nodes. In our method, when taking the discrete cosine transform block size 4×4 , the number of graph nodes is reduced to 7500 graph nodes. The large decrease on the graph nodes number dramatically improves the computation efficiency.

6. Conclusions

This paper proposes a novel graph-based (block-based) approach to image segmentation based on the traditional normalized cuts algorithm. The traditional normalized cuts algorithm is recalled by treating the image segmentation task as a graph partitioning problem. The discrete cosine transform is utilized to obtain the features (color, texture and location information) of each image, considering the huge image size of the high-resolution digital image. A novel computational method is developed and applied to compute weight between graph nodes efficiently. The proposed segmentation algorithm is applied to the Salient Region Detection database and Corel1000 database. Experimental results clearly show that our method outperforms other algorithms.

Another contribution of this paper is the transformable framework which is derived from the traditional normalized cuts. Through Section 5, it turns out to be easy to get suboptimal results by solving a generalized eigenvalue system. The key goal of our framework is to maximize the association within the regions, and meanwhile, minimize the disassociation between the regions. The experimental results demonstrate that the good segmentation precision and high efficiency of the novel approach proposed in this paper. Future work falls on the consideration of the semantic information into the link of the edges.

Acknowledgement

The authors wish to thank all the members of the Digital Media and Communication Research Group in Southwest University, for helpful discussions and suggestions.

- **Funding Statement:** this work was supported by the Key Project on Postgraduate Education Innovation of Chongqing (No. yjg212012), the Postgraduate Education Management Project of Chinese Society of Academic Degrees and Graduate Education (No. 2021-NLZX-YB53), the National Natural Science Foundation of China (No. 51908503) and the Humanities and Social Sciences Project of Basic Scientific Research Business Expenses of Central Universities (No. SWU2009406).
- **Conflicts of Interest:** the authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] Andrew A., "Multiple View Geometry in Computer Vision," *Kybernetes*, vol. 30, no. 9/10, pp. 1333-1341, 2001.
- [2] Achanta R., Hemami S., Estrada F., Susstrunk S., "Frequency-Tuned Salient Region Detection," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Miami, pp. 1597-1604 2009.
- [3] Alpert S., Galun M., Brandt A., and Basri R., "Image Segmentation by Probabilistic Bottom-Up Aggregation and Cue Integration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 2, pp. 315-327, 2012.
- [4] Alom M., Hasan M., Yakopcic C., Taha T., Asari and V., "Improved Inception-Residual Convolutional Neural Network for Object Recognition," *Neural Computing and Applications*, vol. 32, pp. 279-293, 2020.
- [5] Boykov Y. and Funka-Lea G., "Graph Cuts and Efficient N-D Image Segmentation," *International Journal of Computer Vision*, vol. 70, no. 2, pp. 109-131, 2006.
- [6] Birchfield S., Natarajan B., and Tomasi C., "Correspondence as Energy-Based Segmentation," *Image and Vision Computing*, vol. 25, no. 8, pp. 1329-1340, 2007.
- [7] Burdescu D., Stanescu L., Brezovan M., and Spahiu C., "Efficient Volumetric Segmentation Method," in *Proceedings of Federated Conference on Computer Science and Information Systems*, Warsaw, pp. 659- 668, 2014.
- [8] Belalia A., Belloulata K., and Kpalma K., "Region-Based Image Retrieval using Shape-Adaptive DCT," *International Journal of Multimedia Information Retrieval*, vol. 4, pp. 1-16, 2015.
- [9] Bargoti S. and Underwood J., "Image Segmentation for Fruit Detection and Yield Estimation in Apple Orchards," *Journal of Field Robotics*, vol. 34, no. 6, pp.1039-1060, 2017.
- [10] Bai X., Zhang Y., Liu H., and Chen Z., "Similarity Measure-Based Possibilistic FCM With Label Information for Brain Mri Segmentation," *IEEE Transactions on Cybernetics*, vol. 49, no. 7, pp. 2618-2630, 2019.
- [11] Cherri A. and Karim M., "Optical Symbolic Substitution: Edge Detection using Prewitt, Sobel, and Roberts Operators," *Applied Optics*, vol. 28, no. 21, pp. 4644-4648, 1989.
- [12] Comaniciu D. and Meer P., "Mean Shift: A Robust Approach toward Feature Space Analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603-619, 2002.
- [13] Choong M., Kow W., Chin Y., Angeline L., and Teo K., "Image Segmentation Via Normalised Cuts and Clustering Algorithm," in *Proceedings of IEEE International Conference on Control System, Computing and Engineering*, Penang, pp. 430-435, 2012.
- [14] Cheng M., Mitra N., Huang X., Torr P., and Hu S., "Global Contrast Based Salient Region Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 569-582, 2015.
- [15] Chen T., Chuang K., Wu W., and Lu Y., "Compressed Medical Image Quality Determination Using The Kolmogorov- Smirnov Test," *Current Medical Imaging Reviews*, vol. 12, no. 999, pp. 1-1, 2017.
- [16] Dabbaghchian S., Ghaemmaghami M., and Aghagolzadeh A., "Feature Extraction Using Discrete Cosine Transform and Discrimination Power Analysis with A Face Recognition Technology," *Pattern Recognition*, vol. 43, no. 4, pp. 1431-1440, 2010.
- [17] Ding K., Xiao L., and Weng G., "Active Contours Driven by Region-Scalable Fitting and Optimized Laplacian of Gaussian Energy for Image Segmentation," *Signal Processing*, vol. 134, pp. 224-233, 2017.
- [18] Ding K., Xiao L., and Weng G., "Active Contours Driven by Local Pre-Fitting Energy for Fast Image Segmentation," *Pattern Recognition Letters*, vol. 104, pp. 29-36, 2018.
- [19] Falco A., Udupa J., and Flvio K., "An Ultra-Fast User-Steered Image Segmentation Paradigm: Live-Wire-On-The-Fly," *IEEE Transactions on Medical Imaging*, vol. 19, no. 1, pp. 55-62, 2000.
- [20] Feng G. and Jiang J., "JPEG Compressed Image Retrieval via Statistical Features," *Pattern Recognition*, vol. 36, no. 4, pp. 977-985, 2003.
- [21] Felzenszwalb P. and Huttenlocher D., "Efficient Graph-Based Image Segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167-181, 2004.
- [22] Fan J., Zeng G., Body M., and Hacid M., "Seeded Region Growing: an Extensive and Comparative Study," *Pattern Recognition Letters*, vol. 26, no. 8, pp. 1139-1156, 2017.
- [23] Gijssenij A., Geversa T., and Lucassen M., "A Perceptual Comparison of Distance Measures for Color Constancy Algorithms," in *Proceedings of European Conference on Computer Vision*, Marseille, pp. 12-18, 2008.
- [24] Gallivanone F., Interlenghi M., Canervari C., and Castiglioni I., "A Fully Automatic, Threshold-Based Segmentation Method for the Estimation of the Metabolic Tumor Volume from PET Images: Validation on 3D Printed Anthropomorphic Oncological Lesions," *Journal*

- of Instrumentation, vol. 11, no. 01, pp. 01022-01022, 2016.
- [25] Garcia-Lamont F., Cervantes J., López A., and Rodriguez L., "Segmentation of Images by Color Features: A Survey," *Neurocomputing*, vol. 292, pp. 1-27, 2018.
- [26] Guo B., Zhang Y., Zhang D., and Wang Z., "Special Issue on Device-Free Sensing for Human Behavior Recognition," *Personal and Ubiquitous Computing*, vol. 23, no. 1, pp. 1-2, 2019.
- [27] Haut J., Paoletti M., Plaza J., and Plaza A., "Cloud Implementation of the K-Means Algorithm for Hyperspectral Image Analysis," *The Journal of Supercomputing*, vol. 73, no. 1, pp. 514-529, 2017.
- [28] Jiang H., Wang J., Yuan Z., Liu T., Zheng N., and Li S., "Automatic Salient Object Segmentation Based on Context and Shape Prior," in *Proceedings of the British Machine Vision Conference*, pp. 1-15, 2011.
- [29] Javed M., Nagabhushan P., and Chaudhuri B., "A Review on Document Image Analysis Techniques Directly in the Compressed Domain," *Artificial Intelligence Review*, vol. 50, no. 4, pp. 539-568, 2017.
- [30] Kim H., Lee J., Ahn E., Cho S., Shin M., and Sim S., "Concrete Crack Identification using A UAV Incorporating Hybrid Image Processing," *Sensors*, vol. 17, no. 9, pp. 2052, 2017.
- [31] Khadidos A., Sanchez V., and Li C., "Weighted Level Set Evolution Based on Local Edge Features for Medical Image Segmentation," *IEEE Transactions on Image Processing*, vol. 26, no. 4, pp. 1979-1991, 2017.
- [32] Li J. and Wang J., "Automatic Linguistic Indexing of Pictures by A Statistical Modeling Approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1075-1088, 2003.
- [33] Li C., Lin L., Zuo W., Wang W., and Tang J., "An Approach to Streaming Video Segmentation with Sub-Optimal Low-Rank De-Composition," *IEEE Transactions on Image Processing*, vol. 25, no. 5, pp. 1947-1960, 2016.
- [34] Litjens G., Kooi T., Bejnordi B., Setio A., Ciompi F., Ghafoorian M., Laak J., Ginneken B., and Sánchez C., "A Survey on Deep Learning in Medical Image Analysis," *Medical Image Analysis*, vol. 42, no. 9, pp. 60-88, 2017.
- [35] Li J. and Wang Z., "Real-Time Traffic Sign Recognition Based on Efficient Cnns in the Wild," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 3, pp. 975-984, 2019.
- [36] Liu X., Dong Z., and Yang Y., "Recent Progress in Semantic Image Segmentation," *Artificial Intelligence Review*, vol. 52, pp. 1089-1106, 2019.
- [37] Maire E. and Withers P., "Quantitative X-Ray Tomography," *International Materials Reviews*, vol. 59, no. 1, pp. 1-43, 2017.
- [38] Maier O., Menze B., Gablentz J., Häni L., Heinrich M., and et al, "ISLES 2015-A Public Evaluation Benchmark for Ischemic Stroke Lesion Segmentation from Multispectral MRI," *Medical Image Analysis*, vol. 35, pp. 250-269, 2016.
- [39] Mary M., Padma L., and John M., "Modified Image Segmentation Method Based on Region Growing and Region Merging," *The International Arab Journal of Information Technology*, vol. 13, no. 64, pp. 899-907, 2016.
- [40] Ngo T., Lu Z., and Carneiro G., "Combining Deep Learning and Level Set for the Automated Segmentation of the Left Ventricle of the Heart from Cardiac Cine Magnetic Resonance," *Medical Image Analysis*, vol. 35, pp. 159-171, 2017.
- [41] Nie G., Fu Y., Zheng Y., and Huang H., "Image Restoration from Patch-Based Compressed Sensing Measurement," *arXiv preprint arXiv:1706.00597*, 2017.
- [42] Niu S., Chen Q., Sisternes L., Ji Z., Zhou Z., and Rubin D., "Robust Noise Region-Based Active Contour Model Via Local Similarity Factor for Image Segmentation," *Pattern Recognition*, vol. 61, pp. 104-119, 2017.
- [43] Parameswaran V. and Shukla A., "Processing and Characterization of A Model Functionally Gradient Material," *Journal of Materials Science*, vol. 35, no. 1, pp. 21-29, 2000.
- [44] Prabhakar R., "Evaluation of The Gaussian Blob Model for Coarse-Graining Hydrodynamic Interactions in Isolated Polymer Molecules," in *Proceedings of The XV International Congress on Rheology: The Society of Rheology 80th Annual Meeting; Albert Co, and Gary L. Leal; AIP Conference Proceedings 1027*, Monterey, pp. 309-311, 2008.
- [45] Rother C., Kolmogorov V., and Blake A., "'Grabcut': Interactive Foreground Extraction Using Iterated Graph Cuts," *Acm Trans Graph*, vol. 23, no. 3, pp. 309-314, 2004.
- [46] Rahtu E., Kannala J., Salo M., and Heikkilä J., "Segmenting Salient Objects from Images and Videos," in *Proceedings of European Conference on Computer Vision*, pp. 366-379, 2010.
- [47] Rezvanifar A. and Khosravifard M., "Including The Size of Regions in Image Segmentation by Region-Based Graph," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 635-44, 2014.
- [48] Rashno A., Nazari B., Koozekanani D., Drayna P., Sadri S., Rabbani H., and Parhi K., "Fully-Automated Segmentation of Fluid Regions in

- Exudative Age-Related Macular Degeneration Subjects: Kernel Graph Cut in Neutrosophic Domain,” *PLoS ONE*, vol. 12, no. 10, pp. e0186949, 2017.
- [49] Shi J. and Malik J., “Normalized Cuts and Image Segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888-905, 2000.
- [50] Son L. and Thong P., “Some Novel Hybrid Forecast Methods Based on Picture Fuzzy Clustering for Weather Nowcasting from Satellite Image Sequences,” *Applied Intelligence*, vol. 46, no. 1, pp. 1-15, 2017.
- [51] Subudhi A., Jena S., and Sabut S., “Delineation of the Ischemic Stroke Lesion Based on Watershed and Relative Fuzzy Connectedness in Brain MRI,” *Medical and Biological Engineering and Computing*, vol. 56, no. 5, pp. 795-807, 2018.
- [52] Satapathy S., Raja N., Rajinikanth V., Ashour A., and Dey N., “Multi-Level Image Thresholding Using Otsu And Chaotic Bat Algorithm,” *Neural Computing and Applications*, vol. 29, pp. 1285-1307, 2018.
- [53] Singh C. and Bala A., “A DCT-Based Local and Non-Local Fuzzy C-Means Algorithm for Segmentation of Brain Magnetic Resonance Images,” *Applied Soft Computing*, vol. 68, pp. 447-457, 2018.
- [54] Vincent L. and Soille P., “Watersheds in Digital Spaces: an Efficient Algorithm Based on Immersion Simulations,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 6, pp. 583-598, 1991.
- [55] Wu Z. and Leahy R., “An Optimal Graph Theoretic Approach to Data Clustering: Theory and its Application to Image Segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1101-1113, 1993.
- [56] Wang B., Gao X., Tao D., and Li X., “A Nonlinear Adaptive Level Set for Image Segmentation,” *IEEE Transactions on Cybernetics*, vol. 44, no. 3, pp. 418-428, 2017.
- [57] Xu Z., Huang T., Wang H., and Wang C., “Variant of the Region-Scalable Fitting Energy for Image Segmentation,” *Journal of the Optical Society of America A*, vol. 32, no. 3, pp. 463, 2015.
- [58] Yang C., Liu S., and Yang M., “Hallucinating Compressed Face Images,” *International Journal of Computer Vision*, vol. 126, no. 9, pp. 1-18, 2017.
- [59] Zhao C., “Image Segmentation Based on Fast Normalized Cut,” *Open Cybernetics and Systemics Journal*, vol. 9, no. 1, pp. 28-31, 2015.
- [60] Zhang C., Xie Y., Liu D., and Wang L., “Fast Threshold Image Segmentation Based On 2D Fuzzy Fisher and Random Local Optimized QPSO,” *IEEE Transactions on Image Processing*, vol. 26, no. 3, pp. 1355-1362, 2017.
- [61] Zhao Y., Xu J. Q., Li H. W., Zhang P., “Edge Information Diffusion based Reconstruction (EIDR) for Cone Beam Computed Laminography,” *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4663-4675, 2018.
- [62] Zhang Q., Lin J., Tao Y., Shi Y., “Salient Object Detection Via Color and Texture Cues,” *Neuro Computing*, vol. 243, pp. 35-48, 2017.



Hong-Chuan Luo is a researcher at the Postgraduate Admission Office of Graduate School in Southwest University, Chongqing, China. He obtained his Master degree in Computer Science and Technology at the College of Computer and Information Science in Southwest University. His research area includes computer vision, and machine learning for various applications.



Bo Sun is an associate professor at the Department of Civil Engineering in Zhejiang University of Technology, Hangzhou, China. He obtained his Ph.D. degree in Bridge and Tunnel Engineering at the Department of Civil Engineering in Tongji University, Shanghai, China. His research interests focus on the application of computer vision and image technique in engineering problems and practices.



Hang-Kai Zhou is a postgraduate student at the Department of Civil Engineering in Zhejiang University of Technology, Hangzhou, China. His research area focuses on information processing by different computer techniques.



Wen-Sen Cao is a postgraduate student at the Department of Civil Engineering in Zhejiang University of Technology, Hangzhou, China. His research areas include machine learning, data processing and reliability analysis.