Image Saliency Detection Based on Regional Label Fusion

Pengfei Guo, Yao Wang, Lan Wang and Haitao Zhang

October 3, 2022

Image saliency detection based on regional label fusion

Pengfei Guo*
Office of Research
Shantou Polytechnic
Shantou, China
* Corresponding author: lntuguopf@163.com

Yao Wang*
Department of Computing
Shantou Polytechnic
Shantou, China
* Corresponding author: 315910185@qq.com

Lan Wang*
Department of Computing
Shantou Polytechnic
Shantou, China
* Corresponding author: lntuwl@163.com

Haitao Zhang
Department of Computing
Shantou Polytechnic
Shantou, China
lntuzht@163.com

***Abstract*—In this paper, an image saliency detection method based on regional label fusion is proposed to solve the problems with fuzzy boundaries, unclear profile, and less interior density commonly existing in the researches of salient region detection. The image is segmented by super pixel segmentation algorithm, then the spectral clustering is carried out for the super pixel region to reduce the number of regions, thereby the label set with the boundary information could be obtained. Next, three salient features of the image have been fused under conditional random field model to generate the coarse saliency map. Afterwards, regional label fusion method is operated, which organically fuses the boundary information into the coarse saliency map by using the salient mean value calculated with the label information as the regional salient features, moreover, together with adaptive threshold segmentation algorithm to acquire reconstructed saliency map. At last, accurate salient region detection are achieved by calculating with a tag indicating vector defined and reconstructed coarse saliency map. Experimental results show that the salient regions obtained by this algorithm display clearer boundary contours and that the density of salient regions has been greatly improved compared with the other six significant detection methods prevailed in recent years. Particularly, it is confirmed that this method could give better detection for the images with high color similarity between salient and non-saliency regions. The multi-leveled salient detection run on the pixel level and at the regional level further enhance the accuracy of detection results. In short, regional label fusion method shows higher accuracy and stability in the field based on visual saliency such as image retrieval and image annotation.**

Keywords-saliency detection; super pixel; conditional random fields; spectral cluster; regional label

I. INTRODUCTION

The human visual system could find interest regions from different scenes as explained by the visual attention mechanism. Each image contains one or more saliency targets. The saliency detection [1] has been used to improve the efficiency and accuracy of the image processing by imitating the visual attention mechanism to obtain the important information in the image. Image saliency detection is widely used in the fields such as image annotation and retrieval [2], target recognition [3], image automatic cutting [4], and image compression, which became the focus in computer vision research.

Based on the visual attention mechanism, saliency detection has developed to form two models including bottom-up detection model and top-down detection model [5,29]. The bottom-up model has been applied to generate saliency map mainly by the salient features extracted from the difference between the saliency region and the non-saliency region calculated with underlying features such as color, texture, and edge. The top-down model is relatively more complex since it not only contains the processing procedure of bottom-up mode but also it is necessary to adjust the selection criteria according to the goal driven. Therefore, data-driven bottom-up models are widely applied contrasting to the top-down model. With the knowledge of visual attention mechanism, research on saliency has also grown up. Inspired by biology, Itti and Koch [6] proposed a model based on visual attention mechanism, which computed the underlying features of every scale in Gauss Pyramid structure and realized with cross-scale contrast to calculate saliency by employing central-peripheral operators. However, it is unfortunate that the saliency detection is not very accurate because more local information of the image has been chosen during the process of feature extraction, furthermore, this model was excessively dependent on the biological simulation process. Ma et al. [7] proposed to calculate saliency map by using local contrast difference, but which easily neglects the global features due to the requirement of high contrast while obtaining boundary information, thus leading to the salient regions not detailed. Zhai et al. [8] proposed an algorithm by employing the global contrast, which firstly classifies the pixels points with the characteristic histogram at the pixel level, and then uses the range formula to calculate the saliency. The global contrast method could

This work was supported by the project of high-level talent scientific research startup fund in Shantou Polytechnic, the science & technology research project of Liaoning Provincial Department of Education under grant LJ2017QL032.

highlight the salient region, but it is also possible to cause its edge fuzzy. With the ideas of multi-scale space, Achanta et al. [9] calculated the local contrast of the pixel blocks with various sizes using pixel block mean value as the eigenvalue. This algorithm has greatly improved the running speed due to simple consideration on the characteristics of color and brightness among pixels, but it tends to extract the features from higher contrast region since the calculation was merely completed in the multi-scale space. The context-aware saliency detection (CA) algorithm proposed by Goferman et al. [10] generated saliency maps by introducing context cognition theory and integrating local feature maps at different scales. It is worth mentioning that this algorithm uses the spatial position as a supplement besides the characteristics of the pixels itself. Liu et al. [11] firstly obtained the feature map calculated with three features including multi scale contrast, center peripheral histogram and color spatial distribution from three levels of local, regional, and global, respectively, and then used the conditional random field model to fuse those three feature maps to the final salient map. Cheng et al. [12] segmented images by Grab-cut method and calculated the global contrast based on spatial relations with the region as unit. This method calculated the saliency of each block at the regional level, which improved the speed of contrast calculation, saved the calculation time, also obtained a good saliency detection effect. However, the detection result was mostly affected by the segmentation effect, which increased error factors for the saliency detection. Yan et al. [13] obtained saliency maps by constructing tree structures on the segmentation image blocks with different size and adopting multi-scale analysis method, which reduced the interference of high texture area. Jiang et al. [14] firstly over- segmented the image by graph theory based on the thought of statistical learning, then carried out multi-feature description in the segmentation area, and used the random forest method to learn the feature mapping, finally operated a salient fusion on the multi-level.

Saliency detection based on frequency domain has been widely studied in recent years. Generally, what this kind of model does is firstly to transform the image from the spatial domain to the frequency domain, by further analyzing the frequency domain information, and finally to elaborate the relative relationship between the frequency domain information and the salient features in a certain way. The typical algorithm is the spectral residual method proposed by Hou et al [15]. The amplitude spectrum of the image obtained by Fourier transform was carried out the logarithm calculation, after that convoluting with the mean filter to get the residual information, which is convoluted once again with the image amplitude spectrum to generate the saliency information, after the saliency information was transformed to the space domain, the salient region was accordingly established. Guo et al. [16] defined the three-element broad approaches of the spatial domain as four-element and thus obtained the salient region by the Fourier phase spectrum. Zhang et al. [17] proposed to construct salient map in the space domain by using frequency domain space and thereby analyzing the amplitude spectrum and phase spectrum of image features extracted on multiple scales. This method could do favor to get clearer boundaries, but the lack of local features possibly results in not dense enough image salient region.

So far, considerable progress has been made in the research of saliency detection, there are still some problems being noticed. For example, it is often observed that an image with scattered background interference exists in a saliency region, which often causes the inhomogeneous inside the region. In addition, for an image with high foreground and background color similarity, the salient detection effect is reduced to a certain extent. In order to solve the problems that the boundary is not clear and that the interior area is not dense, to obtain more accurate and salient areas, a saliency detection method based on regional label fusion (abbreviated as RLF) is proposed in this paper.

RLF method organically combines a variety of salient features, conditional random field theory (CRF) [18], super pixel segmentation method [19], spectral clustering [20] and adaptive threshold segmentation method [21] by the label information, which thereby achieves more effective saliency detection.

II. Saliency Region Detection Based on Regional Label Fusion

Although the saliency region could be found and divided by the saliency region detection method based on multiple feature fusion, the overlapping of features emphasized by the superposition among features often leads to the problems of saliency regional boundary blurring, unclear outline details and insufficient internal density. In this paper, a saliency region detection method based on regional label fusion would be proposed and applied for avoiding above-mentioned problems. By this method, it could not only realize the fusion of a variety of salient features to get the coarse saliency map, but also acquire rich segmentation boundary information by image segmentation method [30]. Furthermore, the boundary information obtained by over-segmentation could be distributed by label assembly and then integrated into the coarse saliency map to accomplish the accurate segmentation of the saliency region.

RLF method is mainly summarized as follow. Firstly, carry out pretreatment and segment of the image by super pixel segmentation algorithm to get the over-segmented image. It should be noted that in order to reduce the number of regions and make the label of saliency region more unified, the Gauss kernel function is used to calculate the regional similarity according to the color and location information of the region. Afterwards, the spectral clustering of the super pixel region is performed with the regional similarity to obtain the label set of the image segmentation, which is saved as the boundary information of the image. Secondly calculate 3 salient features of the image, and multiple features obtained is fused under conditional random field model to get the coarse saliency map. At last, operate regional label fusion method. With the set of labels come from the segmentation image propagating the boundary information, and after which makes a comparison and fusion with the coarse saliency map, the coarse saliency map has been reconstructed. Then, the salient image is two valued by the adaptive threshold segmentation method, with the label indicator vector marking salient region as a unified label to refine the boundary contour of the saliency region, and the isolated points in the saliency region are further processed to ensure the denseness of the interior region, and thus a more accurate and saliency detection image is obtained. The flow chart of RLF for saliency region detection is shown in Figure 1.

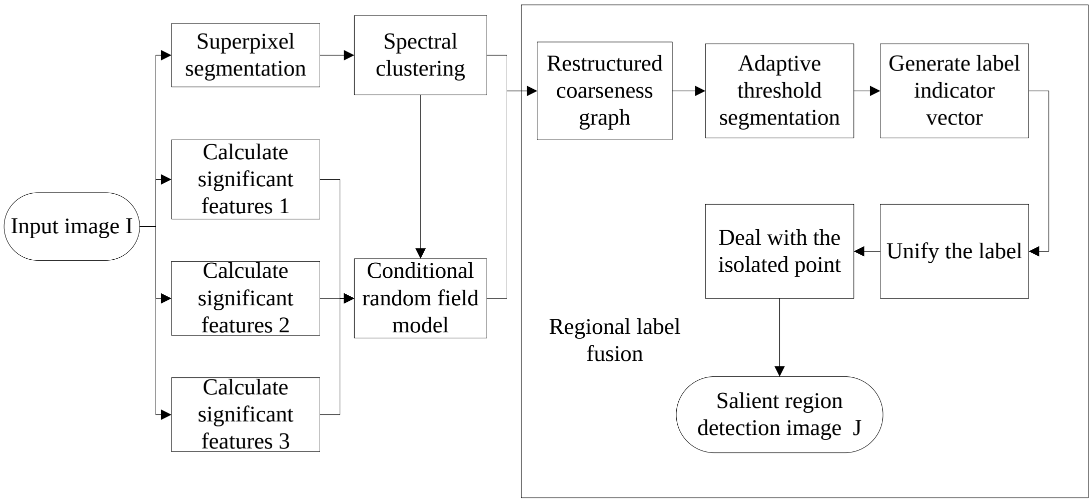


Figure 1. Regional label fusion-based saliency region detection flow chart

The algorithm step result schematic diagram is shown in Figure 2. It was known that simple linear iterative clustering algorithm (SLIC) could quickly and accurately segment the image boundary to achieve super pixel segmentation, so the input image (fig. 2 (a)) has been firstly carried out the SLIC algorithm to obtain better boundary degree of over segmentation image (fig. 2 (b)). Then, the spectral clustering algorithm was used to cluster the region label information to get a less regional segmented image (Fig. 2 (c)), which was also beneficial to reduce the amount of subsequent regional fusion. Next, three salient features were fused by conditional random fields to get the coarse degree saliency map (Fig. 2 (d)), but it could be seen that the contour of the image was not very clear and need to be carried on the fusion of regional labels. Accordingly, the coarse saliency map was regionally fused with the boundary information obtained from the segmenting image label (fig. 2 (e)), and because the saliency region marked as the same tag value, the non-saliency region (fig. 2 (f)) could be filtered out; and thus, a saliency area with clear boundary contour and interior density (Fig. 2 (g)) has been acquired with binarization of the segmented image by adopting adaptive threshold segmentation. Obviously, whether boundary contour clarity or interior density could be effectively generated by taking advantage of RLF method as shown in Figure 2. In addition, it was also confirmed that this method could give very good detection for the images with high color similarity between salient and non-saliency regions, and the results were closer to the ground-truth (GT) segmentation image.

(a) (b) (c) (d) (e) (f) (g)

Figure 2. The pipeline of schematic by RLF. ((a)input image；(b)SLIC；(c)spectral cluster；(d)roughness saliency image；(e)regional label fusion；(f)color saliency image；(g)saliency segmentation image)

A. The generation of regional label information

Super pixel segmentation is the first step preprocessing of the saliency region detection. Compared with other super pixel methods, SLIC algorithm is characterized by convenient use, fast running speed, small storage space, and super pixel segmented by SLIC shows better compactness and the fitting degree of the boundary. Therefore, SLIC has been used firstly to divide the image into several blocks. Each pixel in the super pixel block displays the characteristics of similar color and location. Super pixel segmentation could get a set of tags with boundary information. However, in order to obtain detailed boundary information, the phenomenon of super pixel over-segmentation is serious, which does not favor to the subsequent fusion of tags and the denseness of the images. Therefore, the spectral clustering algorithm is used to achieve regional clustering and at the same time to reduce the number of regional fusions. Literature [22] described different Laplacian matrix and their corresponding basic properties, introduced the currently common spectral clustering algorithms, and compared their advantages and disadvantages. A super pixel's spectral clustering method proposed in Literature [23] mapped super pixels into undirected graphs, constructed object sets, calculated similarity matrix in RBG space, and found out by spectral clustering the super pixel cluster corresponding to the basic unit of direct part sign (DPM) code region, and realized the location of the code area. Referenced above-mentioned researches, in LAB color space, with super pixel position information as auxiliary, with three-dimensional color features and Gauss kernel computing similarity, and further by using Laplacian matrix, spectral clustering has been achieved in our research.

The main algorithm steps of super pixel segmentation and spectral clustering are briefly described below.

- Firstly, the number of region segmentation is set as K, and the number of cluster centers is set as N, so the number of pixels in each region is N/K after an image is segmented. Each pixel could be written as $x_i = (l, a, b, x, y)$ in five dimensions, of which (l, a, b) is the value in color space CIELAB and (x, y) is the pixel point coordinate value.

- Initialize each seed point (cluster center), and calculates the color distance and space distance in the five-dimensional space to obtain the similarity between the pixel points and the seed points. Then, synchronizes the cluster center until convergence by k-means algorithm, after merging the small area the image is finally clustered into K regions to get the characteristic initial tag information for each super pixel region, which is called as $knlabels[i]\ (i = 1,2,3, \dots, K)$.

- Spectral clustering is the method based on graph theory, which could identify the sample space with arbitrary shape and converge to the global optimal solution. The undirected graph $\langle V, H \rangle$ could also be established. After super pixel segmentation, each super pixel could be seen as a point in graph V, and every two adjacent super pixels correspond to one edge in H. The Gauss kernel function is used to calculate the region similarity matrix $W \in R^{K \times K}$, in which the similarity of any two super pixel blocks is defined as:

$$W_{ij} = \begin{cases} exp\left(-d\left(R_i, R_j\right)\right) & H_{ij} = 1 \\ 0 & H_{ij} = 0 \end{cases} \tag{1}$$

$d\left(R_i, R_j\right) = (l_i - l_j)^2 + (a_i - a_j)^2 + (b_i - b_j)^2$ indicates the distance sum of squares of any two non-overlapping regions in LAB color space in the case of region adjacency. $H \in R^{K \times K}$ is the adjacency matrix that used to measure the proximity among the regions. Adj is the Gauss kernel parameter, generally, the value of adj is 20~40 (the experimental value is 27).

- Calculate the Graph Laplacians normalization matrix $L \in R^{K \times K}$ for the super pixel region based on region similarity.

$$L = D^{-\frac{1}{2}} * W * D^{\frac{1}{2}} \tag{2}$$

$W \in R^{K \times K}$ is a regional similarity matrix, $\mathrm{D} \in \mathrm{R}^{\mathrm{K \times K}}$ is diagonal matrix, and the diagonal value is the sum of the corresponding rows or columns in the W matrix, i.e., $D_{ii} = \sum_{j=1}^{K} W_{ij}$.

- Generate the smallest n eigenvalues and corresponding eigenvectors with the Laplacian matrix L. These n eigenvectors constitute the matrix $Y \in R^{K \times n}$, and each line of Y is recorded as one data point, which would be clustered to reduce the number of regions to $n(n < K)$, thus obtaining the super pixel spectral clustering segmentation image $C(x, y)$, and regional label information updated and merged is written as $clabels[i]\ (i = 1,2,3, \dots, K)$.

The super pixel segmentation could quickly acquire better boundary information that is further saved as the label. Spectral clustering not only preserves the edge superiority of the super pixel segmentation, but also reduces the number of over segmented regions, which provides more accurate boundary information for the subsequent fusion with saliency maps and speeds up the fusion process.

B. *Coarseness graph generation from multi feature fusion*

In order to extract the saliency region from the image, this research used an automatic top-down method recently proposed by Liu et al [11]. Three features related to image saliency have been extracted, including multi-scale contrast map, center surround histogram map and center weighted color spatial distribution map. And then the integral image technique proposed by F. Porikli et al. [25] was used to calculate the characteristic value. The conditional random field model was adopted to fuse three feature maps to obtain pixel-leveled saliency map.

The algorithm steps of multi-featured fusion under conditional random field are briefly described below.

- For the input image I, calculate the multiscale contrast map F_1, the center periphery histogram F_2, and the center weighted color space distribution map F_3, respectively.

- Use Linearly fuse the three features to get the coarse degree saliency map by the CRF model. The conditional distribution function of CRF is

$$p(A|I) = \frac{1}{z} exp\left(-E(A|I)\right) \tag{3}$$

The E represents the energy function and the Z represents the distribution function. In this paper, the energy function E is defined as

$$E(A|I) = \sum_x \sum_{h=1}^{3} j_h * F_h + \sum_{x,x',h} b_h C(l_x, l_{x'}, I) \tag{4}$$

In which F_h represents the h feature map of the image; φ_h and β_h are weighted parameters, respectively; and $C(l_x, l_{x'}, I)$ are pair functions, of which $l_x, l_{x'}$ is a pair of pair two value tags corresponding to pixel point x;

$l_x = 1$ and 0 indicate that the pixel points belong to the saliency region and background area, respectively. The coarse saliency graph obtained from conditional random field model is recorded as $G(x,y)$.

C. Saliency region segmentation based on label fusion

As above described, the saliency map from three salient features through the fusion of the conditional random field model could be thought to generate from the pixel level prominence, thus which might cause blurred edge and unclear contour. While the image segmentation method at the regional level might promote boundary contour clear, there still existed the over-segmentation phenomenon and lower computed accuracy. Therefore, fusion method of regional label was used to further optimize the as-obtained results in this research. The boundary information stemmed from the super pixel and spectral clustering fused with the coarse saliency map, then the adaptive threshold segmentation was used for the majorization of the two-value segmentation process of the saliency graph, by which finally exporting the two-value graph with clear boundary. This method retains the advantage of multi-feature fusion, on the other side, it further fuses with the boundary information brought by the image over segmentation, thereby realizing the fusion of the salient information of the pixel level and the region level. More importantly, the fusion of the two parts greatly improves the anti-interference ability of pixel-leveled salient detection and the accuracy of the region level saliency detection. The concrete steps of the algorithm are as follows.

- Input the segmentation image $C(x,y)$, the coarse saliency map $G(x,y)$ and the area label information $clabel(x,y) = i$ obtained from the super pixel clustering, respectively.

- Based on label information, compute the salient mean value of each super pixel, which represents the salient features of each region. The formula for calculating the salient mean value U is

$$U[i] = \frac{\sum_{clabel(x,y)=i} G(x,y)}{m_i} \quad i = 1,2,3,\dots,n \qquad (5)$$

  The m_i represents the total number of pixels with a tag value of i; $\sum_{clabel(x,y)=i} G(x,y)$ represents the sum of the salient values of the pixels with the pixel point $p(x,y)$ and the label i in the coarse saliency map.

- Use the salient mean value of each super pixel to reconstruct the gross visibility map at the regional level. For each area $R_i(i \in [1,n])$, the salient value of the pixel $p(x,y)$ is replaced by the salient mean value of the region, that is

$$G'(x,y) = U[i] \qquad p(x,y) \in R_i \qquad (6)$$

  After reconstructing, the coarse degree saliency map has shown more obvious regionalization characteristics and less salient value, which accelerates the adaptive threshold segmentation process as well.

- Realize binarization of the coarse degree saliency map by adopting adaptive threshold segmentation algorithm.

  Traditionally, the image is two valued by comparing the salient value and the threshold value T set in advance indicated as the fixed threshold segmentation method. In this way, the processing speed is fast but it is often accompanied by the lower stability. Particularly, it needs to set thresholds repeatedly in order to find the most suitable segmentation for the different images, and it is also difficult to show excellent image segmentation results for the image with high foreground and background similarity. Therefore, adaptive threshold segmentation was presented to deal with these problems. Different from the gray image used in the paper [21], the reconstructed coarse value saliency map was adaptively divided into the two values in this research. The saliency value was firstly multiplied by 255 to make its value between 0 and 255, and then it is needed to find a suitable threshold to make the two parts of the foreground and the background satisfy the minimum inter class variance of the intra class variance, which finally got a two valued coarse salience graph $g(x,y)$.

- Generate the label indication vector.

  This paper defines a label indicator vector LI indicating the belonging of the tags to the salient or the background areas when the label information is fused with the rough graph. The generating formula of the indicator vector LI is as follows.

$$\begin{cases} LI(i) = 1 & \left|\frac{S(i)-\varphi}{\varphi}\right| \le \rho \\ LI(i) = 0 & \left|\frac{S(i)-\varphi}{\varphi}\right| > \rho \end{cases} \qquad (7)$$

  Among them, $S(i)$ represents the number of pixels with a salient value of 1 for the area labelled i in the coarse salient map $g(x,y)$. $S(i)$ is defined as:

$$S(i) = \sum_{clabels(x,y)=i}^{x,y} g(x,y) \qquad (8)$$

  The $clabel(x,y) = i$ is the label information; ρ is the threshold parameter; $\varphi = max|S(i)|$ represents the maximum among the pixels number with the salient value as 1 corresponding to different labels. $LI(i)$ is the indicator vector of the tag i, and 1 and 0 mean that the label is assigned to the salient region and background area, respectively.

- Unify the label value and divide the front background area. Get the final label information with the tag indicator vector to, that is

$$\begin{cases} label(x,y) = co & LI(i) = 1 \\ & clabel(x,y) = i \\ label(x,y) = clabel(x,y) & LI(i) = 0 \\ & clabel(x,y) = i \end{cases} \qquad (9)$$

- co is the tag value corresponding to the number of pixels as φ; $LI(i)$ represents the indicator vector of the tag, and $clabels(x,y)$ corresponds to the label information of pixel point $p(x,y)$. The salient area is marked as the same label value co, meaning that the salient area $cc(x,y)$ domain could be recognized by the tag value co . The formula is

$$\begin{cases} cc(x,y) = 0 & label(x,y) \neq co \\ cc(x,y) = c(x,y) & label(x,y) = co \end{cases} \quad (10)$$

Among them, the $label(x,y)$ is the tag value and the $c(x,y)$ represents the segmentation image of super pixel clustering.

- Treat the "solitary point" in the salient area.

  A small number of non-salient labels in salient regions are processed by 8 neighborhood method. It is defined that the parameter np represents the number of label value in the current test area identical with that in the 8 neighborhoods, and they would be merged if $np > \alpha$ according to the threshold α ($\alpha = 5$ in the experiment). In order to prevent the reduction of boundary clarity caused by the excessive merger, it is necessary to merge not only the small area which is negligible due to less label number in the salient area, but also the salient points in the non-saliency region, in this way the dense and saliency regional map J could be finally achieved.

III. EXPERIMENTAL RESULTS AND DISCUSSION

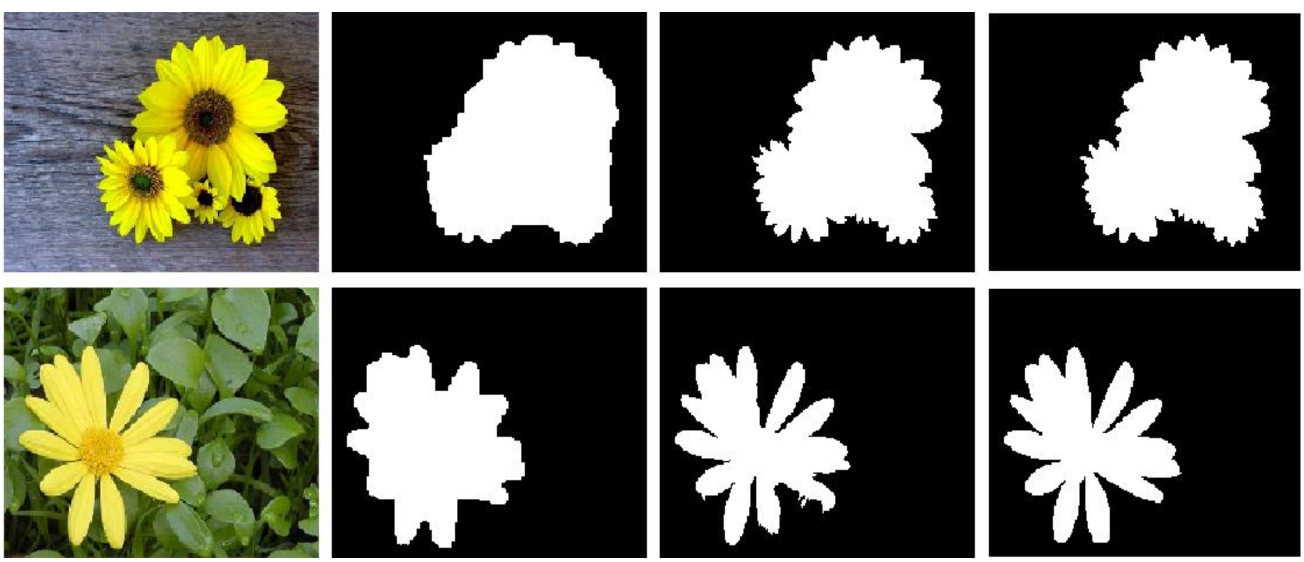
In this paper, the MSRA10k dataset provided by the Media Computing Laboratory of Nankai University was used as the test data set. The dataset contains 10000 images and their corresponding GT segmentation. The characteristic of our algorithm was verified and analyzed by using this dataset. The three evaluation indexes including Precision, Recall, and F - measure were used to compare the algorithm RLF with other algorithms. The F-measure metric formula is

$$F_\alpha = \frac{(1+\alpha^2)*Precision*Recall}{\alpha^2*Precision+Recall} \quad (11)$$

α is the weight parameter of precision, and generally, the value of α^2 is 0.3. This weight parameter reflects the significance of precision in the formula.

A. *Comparative test of gross detection and precision detection*

In this algorithm, the detection of salient regions includes roughness detection and label fusion boundary information. In order to verify the validity of label fusion, the experiment would be operated based on the combined algorithm with literature [11] and adaptive threshold segmentation (recorded as CSD). This method firstly fused three salient features based on conditional random fields, and then performed the adaptive threshold segmentation of the coarse saliency map to get the coarseness graph. The gross and precision contrast graphs are shown in Figure 3. Figure 3 (a) is the original input graph; figure 3 (b) is the saliency region partition graph stemmed from the CSD algorithm; figure 3 (c) is the saliency region partition graph after running the RLF algorithm, and figure 3 (d) is the ground-truth (GT) partition graph of the input graph. The flower picture with more detailed boundaries has been chosen as the experimental sample with the aim to fully illustrate the difference in the aspects of the boundary contour and dense interior of the salient region between CSD and RLF algorithm. Compared with the GT image, it could be seen that the prominent area divided by RLF showed more precise and detailed boundary contour than that treated by the CSD algorithm, and the petal pattern was clearer. It might be the reason that with the utility of the precise boundary information got from the super pixel and spectral clustering algorithm, the RLF algorithm effectively divided the background area from the adjacent salient region, resulting in the segmentation image closer to GT.

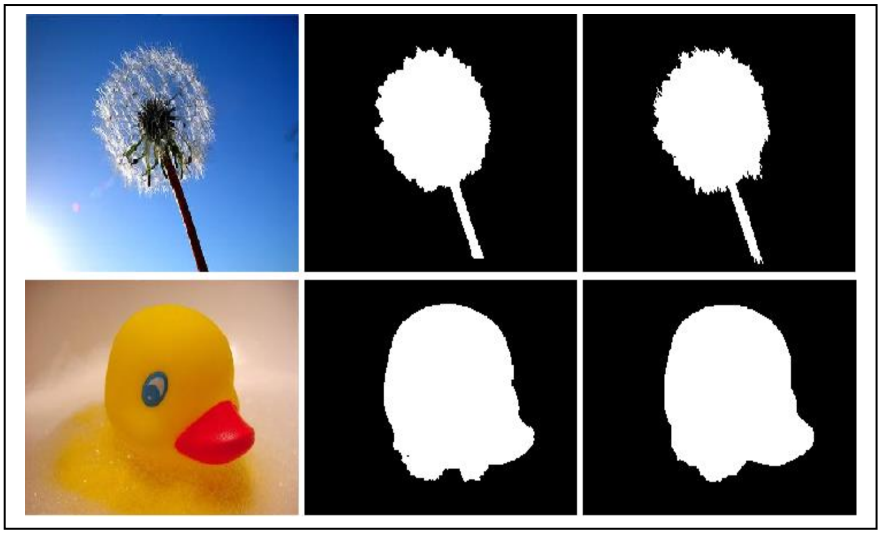


(a) (b) (c) (d)

Figure 3. Comparison diagram of roughness and precision saliency detection ((a) input image; (b) CSD; (c) our method RLF; (d) GT))

B. *Salient detection experiment of special image*

The special image in this experiment refers to the image with higher similarity whether the color or the texture between the background area and the saliency region. A saliency map of a special image is shown in Figure 4. Figure 4 (a) is the original input graph; Figure 4 (b) is the saliency detection graph by algorithm RLF, and Figure 4 (c) is GT partition graph. Compared with GT graph, the RLF algorithm could effectively divide the dandelion from the background area (blue sky), and divide the head of the yellow duck out of the whole yellow duck in the water, although the background area of the two pictures were very close to the color of the saliency region. It also could be observed that the segmented dandelion boundary contour was very clear and detailed, and the head segmentation of the small duck was very complete, which ensured the internal density of the two images, so the segmentation image was very close to that from GT treatment. The experimental results showed that the proposed RLF algorithm divide the dandelion from the background area (blue sky), and divide the head of the yellow duck out of the whole yellow duck in the water, although the background area of the two pictures were very close to the color of the saliency region. It also could be observed that the segmented dandelion boundary contour was very clear and detailed, and the head segmentation of the small duck was very complete, which ensured the internal density of the two images, so the segmentation image effectively segment the salient region to acquire a salient detection map with detailed and clear boundaries though the difference. In this research, a five-dimensional feature (including three-dimensional color features and two-dimensional position features) was chosen as the boundary information in the region segmentation, and then three salient features were fused under the conditional random field model, which effectively reduced the noise effect of the background area. Additionally, even for the special images with strong background interference, the RLF method could carry out the multi-levelled salient detection, i.e., on the pixel level and at the regional level, thus generating the salient region and the salient contour with clear outline.



(a) input image (b) our method RLF (c) GT

Figure 4. Saliency detection of special images

C. *Comparison experiment of regional segmentation method*

The

As above-mentioned, the SLIC algorithm was used during the process of regional segmentation to get the over segmentation area, and spectral clustering algorithm was used to speed up the fusion of label in the subsequent experiment and to improve the internal density due to the decrease of the regional number. So, in the experiments aiming to verify the effectiveness of the spectral clustering procedure in the RLF algorithm, SLIC segmentation and SLIC+ spectral clustering was carried out to get the boundary information, respectively, under the same condition to construct the coarse saliency map and use the regional label fusion method proposed in this research. As shown in figure 5, figure 5 (a) was the result constituted with 200 blocks of segmented images by only SLIC algorithm; figure 5 (b) was the result constructed with 200 blocks by SLIC segmentation clustering to 90 blocks with spectral clustering algorithm; and figure 5 (c) was the image formed by1000 pieces of super pixel segmentation by only using SLIC algorithm; Figure 5 (d) was the result stemmed from clustering 1000 super pixels by SLIC segmentation to 510 block regions by spectral clustering algorithm. Obviously, the segmentation with clear outline could be obtained by SLIC algorithm only, but it could also be seen that the super-pixel over-segmentation was more serious and thus the graph displayed cell convex, and the internal density was not enough in the salient region. In contrast, RLF method, despites reducing the number of regions, SLIC super pixel segmentation together with the use of spectral clustering algorithm could not only obtain reliable boundary information in the case of less super pixel blocks, but also enhance the density within the salient region, which finally generate more ideal salient regional segmentation results.

(a)only SLIC (b)SLIC+spectral clustering (c)only SLIC (d)SLIC+spectral clustering

Figure 5. Comparison diagram of saliency region segmentation

D. Comparative Experiment with Other Algorithms

In order to verify the effectiveness and accuracy of the regional segmentation and boundary superiority of the segmentation result, the experiments used to evaluate RLF have also been carried on the other 6 new salient regional detection methods developed and prevailed in recent years, FT [26], SEG [27], RC [28], CA, CB [29], and CSD, respectively. The results were shown in Figure 6. Figure 6 (a) was the original input image, figure 6 (b) ~ (g) were the segmentation results from the other six algorithm methods, respectively, and figure 6 (h) came from RLF algorithm proposed in this paper; and figure 6 (i) was the GT segmentation image of the original image. As can be seen from Figure 6, compared with other methods, the results obtained by the algorithm RLF could not only have better density within the region, but also display a clearer and more accurate boundary contour than the other algorithms, and the salient regional cut results were closer to that come from the GT segmentation image.

Figure 7 shows the results run on RLF and the other 6 algorithms evaluated with the three indexes. It could be seen that RLF is superior to FT, SEG and CA in terms of overall performance, and all indexes are higher than those of the three algorithms. Compared with RC, RLF is basically flat on Precision and F-measure, while on Recall, RLF is 11% higher than RC. Compared with CB, on Precision, CB is only 1% higher than RLF, while on Recall and F-measure, RLF is 19.5% and 2% higher than CB, respectively. Compared with CSD, Recall is superior to RLF, while RLF is superior to CSD on Precision and F-measure. Overall, the RLF algorithm has obvious advantages compared with other algorithms, which could result from that the RLF algorithm not only combines three salient features in the salient region segmentation, but also combines the more accurate boundary information obtained by the super pixel and spectral clustering algorithm. Besides, this algorithm could find the salient area accurately and effectively obstruct the interference in the non-salient region, and thereby the resulting segmentation image shows a clearer boundary and the image segmentation result is closer to the ground-truth segmentation image.

IV. CONCLUSION

In this paper, an image saliency detection method based on regional label fusion is proposed for solving the problems of unclear boundary contours and insufficiently dense internal density in salient regions. Under the framework of conditional random fields, the center periphery histogram, multi-scale contrast map and center weighted color space distribution map are combined to get the coarse saliency maps. In order to get more accurate boundary information, the SLIC algorithm is used to get the super pixel map, and then the spectral clustering algorithm is further used to get the corresponding boundary information. By using a regional label fusion method, which combines the saliency map with the set of labels with boundary information and refines the boundary of the coarse saliency map, the salient region with clear boundary contour could be obtained. The experiment verifies the effectiveness of the algorithm from multiple perspectives. In brief, the RLF algorithm could be effectively used to get the salient regions with clear boundaries, and ensure the density inside the salient regions. Moreover, the RLF algorithm is also advantageous to treating the images with similar colors in the salient region and the non-salient region. It could be envisioned that the RLF algorithm might be applied to the field of mass image processing, such as image retrieval and image annotation after further optimizing.

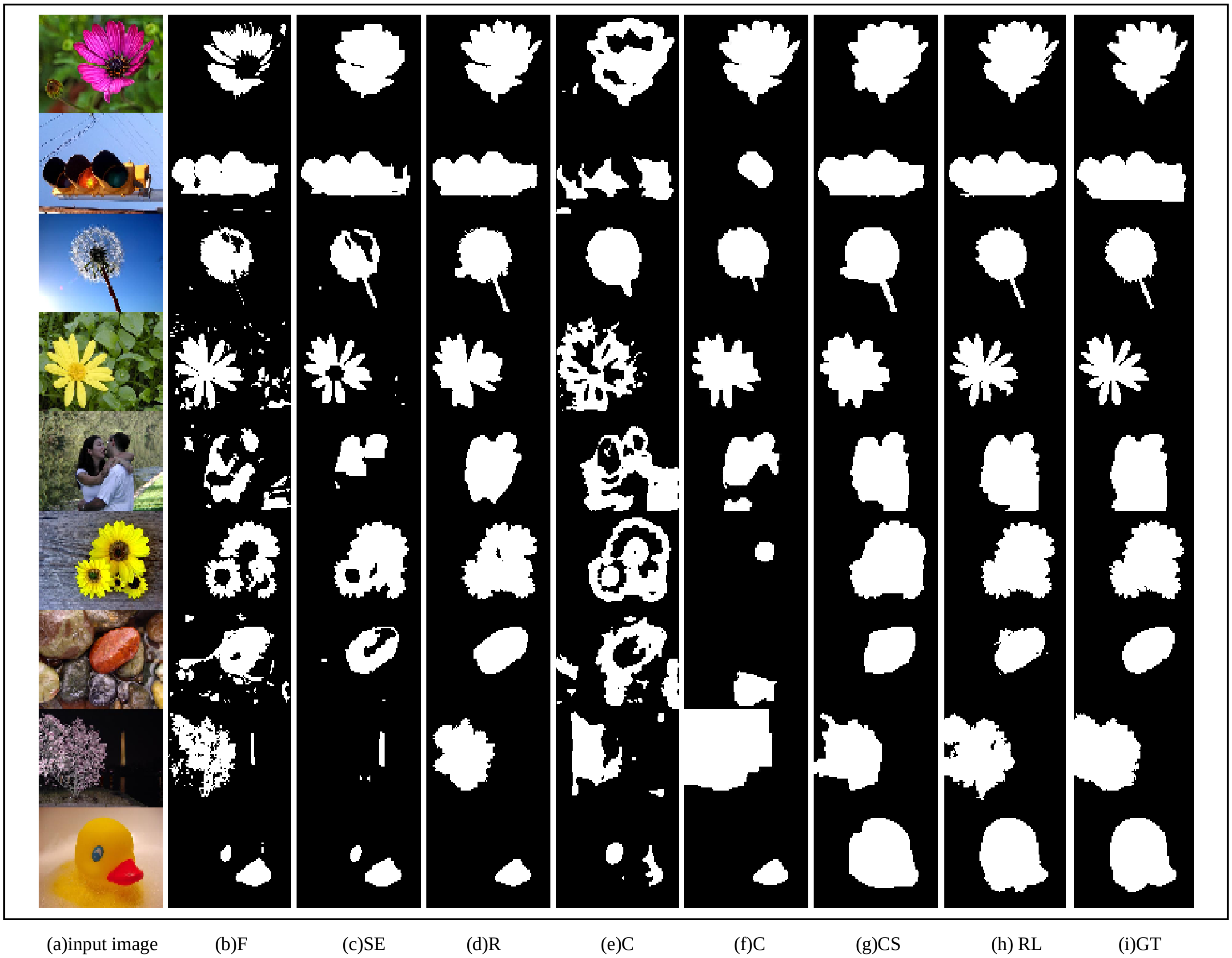


Figure 6. Comparison diagram of region segmentation methods((a)input image; (b)saliency map using FT; (c)saliency map using SEG; (d)saliency map using RC; (e)saliency map using CA; (f)saliency map using CB; (g)saliency map using CSD; (h)saliency map using our RLF; (i)ground-truth)

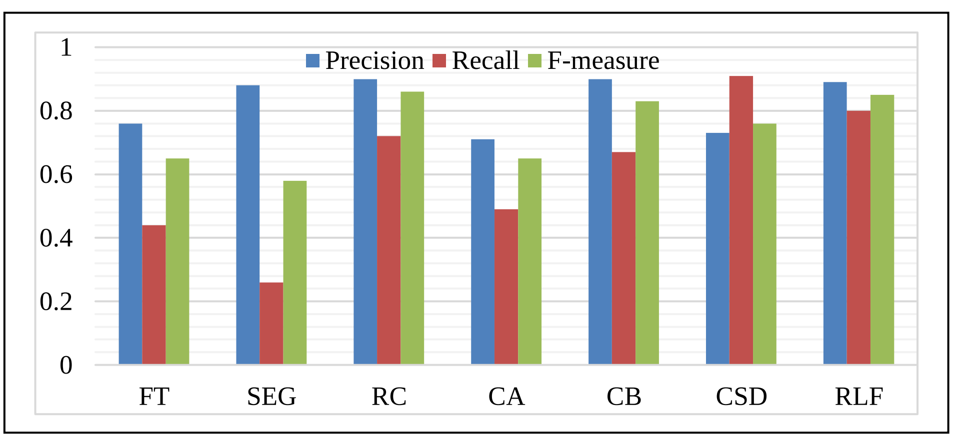


Figure 7. Comparison diagram of evaluation index with different methods

REFERENCES

[1] Triesman A M, Gelade G. A feature-integration theory of attention[J]. Cognitive Psychology, 1980, 12(1):97-136.

[2] Zhang J, Hu W W, Chen Z H, et al. Multi-model fused framework for image annotation[J]. Journal of Computer-Aided Design&Computer Graphics, 2014, 26(3):472-478.

[3] Makovski T, Jiang Y V. Feature binding in attentive tracking of distinct objects[J]. Visual Cognition, 2009, 17(1−2):180−194.

[4] Kim W, Kim C. A novel image importance model for content-aware image resizing[J]. IEEE International Conference on Image Processing. 2011, 263(4):2469−2472.

[5] Koch C, Ullman S. Shifts in selective visual attention:towards the underlying neural circuitry[J]. Hum Neurobiol, 1985, 4(4):219-227.

[6] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11):1254-1259.

[7] Ma Y F, Zhang H J. Contrast-based image attention analysis by using fuzzy growing[C]. Proceedings of the 1lth ACM International Conference on Multimedia. Berkeley, CA, USA：ACM, 2003:374-381.

[8] Zhai Y, Shah M. Visual attention detection in video sequences using spatiotemporal cues[C]. Proceedings of the 1 4th annual ACM international conference on Multimedia. ACM, 2006:815-824.

[9] Achanta R, Estrada F'Wils P, et a1. Salient region detection and segmentation[M]. Computer Vision Systems. Springer Berlin Heidelberg, 2008:66-75.

[10] Goferman S, Zelnik-Manor L, Tal A. Context-aware saliency detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(10): 1915−1926.

[11] Liu T, Yuan Z, Sun J, et al. Learning to detect a salient object[J]. IEEE Transactions on Software Engineering, 2011, 33(2):353−367.

[12] Cheng MM, ZhangG , Mitra NJ, et a1. Global contrast based salient region detection[C]. 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2011:409-416.

[13] Yan Q, Xu L, Shi J P, et al. Hierarchical saliency detection[C]. In: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, Oregon, USA: IEEE, 2013:1155−1162.

[14] Jiang H Z, Wang J D, Yuan Z J, et al. Salient object detection: a discriminative regional feature integration approach[C]. In: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, Oregon, USA: IEEE, 2013:2083−2090.

[15] Hou X D, Zhang L. Saliency detection: a spectral residual approach[C]. In: Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, Minnesota State, USA: IEEE, 2007:1−8.

[16] Guo C L, Ma Q, Zhang L M. Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform[C]. In: Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, Alaska, USA: IEEE, 2008:1−8.

[17] Zhang Q R, Gu G C, Liu H B, et al. Salient region detection using multi-scale analysis in the frequency domain[J]. Journal of Harbin Engineering University, 2010, 31(3):361-365.

[18] Qian S, Chen Z H, Lin M Q, et al. Saliency detection based on conditional random field and image segmentation[J]. Acta Automatica Sinica, 2015, 41(4): 711–724.

[19] Achanta R, Shaji A, Smith K, et al. SLIC super pixels compared to state-of-the-art super pixel methods[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(11): 2274−2282.

[20] Sareh Shirazi, Mehrtash T, Conrad Sanderson, et al. Lovell. Clustering on grassmann manifolds via kernel embedding with application to action analysis[J]. IEEE International Conference on Image Processing, 2012, 19:781-784.

[21] Otsu N. A threshold selection method from gray-level histograms[J]. Automatiea, 1975, 11(285-296):23-27.

[22] Ulrike von Luxburg. A Tutorial on Spectral Clustering[J]. Statistics and Computing, 2007, 17(17):395-416.

[23] Wang J, Wang P, Wang G. Stippled Direct Part Mark Location Based on Self-adaptive Super-pixels Segmentation[J]. Acta Automatia Sinica, 2015, 41(5):991-1003.

[24] Porikli F, Tuzel O, Meer P. Covariance tracking using model update based on lie algebra[C]. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York: Institute of Electrical and Electronics Engineers Computer Society, 2006:728-735.

[25] Achanta R, Hemami S, Estrada F,et al. Frequency-tuned salient region detection[C]. In: Proceedings of the 2009 IEEE International Conference on Computer Vision and Pattern Recognition. Miami Beach, Florida, USA: IEEE, 2009:1597−1604.

[26] Rahtu E, Kannala J, Salo Met al. Segmenting salient objects from images and videos[C]. In: Proceedings of the 11th European conference on Computer Vision — Part V. Berlin, Heidelberg: Springer-Verlag, 2010:366−379.

[27] Cheng M M, Zhang G X, Mitra N J, et al. Global contrast based salient region detection[C]. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition. Colorado Springs, Colorado, USA:2011 IEEE, 2011:409−416.

[28] Jiang H Z, Wang J D, Yuan Z J, et al. Automatic salient object segmentation based on context and shape prior[C]. In: Proceedings of the 2011 British Machine Vision Conference. Dundee, Scotland, UK: BMVA Press, 2011:1−12.

[29] Zhang Y Y, Zhang S, Zhang P, et al. Saliency detection via background and foreground null space learning[J]. Signal Processing: Image Communication, 2019, 70: 271-281.

[30] Khan A I, Wani M A.Patch-based segmentation of latent fingerprint images using convolutional neural network[J]. Applied Artificial Intelligence,2019, 33(1): 87-100.