Multi-Layers Saliency Detection Based on Spectral Density Peaks Clustering

Xiang Cheng, Guangwei Wang and Ruochen Xia

ABSTRACT

Saliency detection obtains the whole salient object by simulating human visual system, but the detection accuracy is seriously influenced by the small-scale and high-contrast regions in foreground or background, especially when dealing with objects with complex construe. This problem is common in natural images and forms a fundamental challenge for prior methods. To solve the above problem, we propose the multi-layers approach based on spectral density peaks clustering, in which the information from different saliency layers are fused to extract saliency object in complex scene, and we propose a criterion of salient region merging in order to light whole salient object. The experimental results show that the proposed method improves saliency detection on many images that cannot be handled well traditionally.

KEYWORDS
Salient Object Detection, Multi-layer, Spectral Clustering, Super-Pixel, Image Segmentation, Markov Mmodel.

INTRODUCTION

Human vision system can distinguish the image salient region quickly and accurately, then focuses attention on the region and ignores the other regions. Such, through saliency detection, more computing resources can be applied to the salient region, and the calculating efficiency can be improved apparently[1]. In recent years, saliency detection has been used widely at many fields of computer vision, such as object-of-interest extraction[2,3], image segmentation[4], image video compression [5,6], image retrieval[7], and so on. Saliency detection has been an important part of computer vision.

Xiang Cheng, Guangwei Wang, Ruochen Xia
School of Mathematics, Physics and Information Science, Zhejiang Ocean University, Zhoushan 316022, China
According to different calculation models, saliency detection methods can be classified into the top-down task driving and the down-top data driving. In the top-down task driving detection, the occurring probability of specified object in image is regarded as the saliency degree of image. In order to improve the detection robustness, a large number of annotation figures with different occasions, different scales and different environments should be used to accomplish the feature extraction by using supervised learning method[8,9].

To solve the influence of image segmentation on saliency detection, the cognitive psychology theory is introduced, which is that there is more than one scale during human visual perception and there are complex interactions between different scales. According cognitive psychology theory, a hierarchical saliency detection based on spectral density peaks clustering model is proposed, which can effectively decrease the influence of incomplete segmentation on saliency detection. The advantages of the new method is shown in the following aspects: (1) clustering in the model does not need predefine clustering parameters or complicated iterative computations; (2) the layer model can solve the influence of incomplete segmentation on detection through fusing the saliency information of different layers; (3) the application of more basic saliency information can improve the detection accuracy.

**SPECTRAL DESTINY PEAKS CLUSTERING**

As an efficient data analysis method, clustering had been extensively applied in image segmentation, such as K_Means, Mean_shift, k_medoids and so on[11]. However, these methods have the following problems:(1) only color similarity and spatial similarity are considered. The spatial continuity is ignored in those clustering methods, which will lead to discontinuous segmentations;(2) the initial clustering centers must be predefined and in most cases, only local optimal solution can be obtained;(3) for clustering results, multiple iterations increase the complexity of the algorithms. To improve the robustness of image segmentation, a new spectral destiny peaks clustering is proposed in this paper.

**SPECTRAL CLUSTERING ALGORITHM**

Spectral clustering relies on the eigenstructure of affinity matrix to accomplish the image clustering segmentation. The main steps of spectral segmentation on the input image with n nodes $P = \{p_1, \cdots, p_n\}$ are shown as follows:

Step 1: construct a graph $G(V,E)$, in which V is the image node, and E is an undirected edge of any two nodes; the affinity matrix $A \in \mathbb{R}^{n \times n}$ defined as follow:
In which \( D(i, j) \) is the color Euler distance of two nodes, and \( \sigma \) is a constant to control the weight coefficients.

Step 2: calculate the Laplace matrix of \( A \) as 
\[
L = D^{1/2} \cdot \Lambda \cdot D^{1/2}
\]
where \( D = \text{diag}(d_1, \ldots, d_m) \) and \( \Lambda = \sum_j A_{ij} \).

Step 3: solve the eigenvalues \( \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n \) of matrix \( L \) and the corresponding normalized eigenvectors \( X = [X_1, \ldots, X_n]^{\text{norm}} \), and then extract the first \( m \) dimensions from the \( X \) to form the new map matrix \( R = [X_1, \ldots, X_n]^{\text{norm}} \).

Step 4: consider every row of matrix \( R \) as a node value vector, and all nodes are clustered with K_Means.

Spectral clustering method is effective, but it inevitably has some problems, such as the algorithm complexity greatly increases with the increasing of input nodes; large area similar region is easily clustered into multiple blocks[12]. In Fig.1, the first row is the input images; the second row is the clustering results. It shows that clustering results are greatly influenced by clustering parameters, but it is difficult to get accurate parameters in a complex scene.

![Figure 1. Spectral clustering segmentation.](image)

SPECTRAL DESTINY PEAKS CLUSTERING

The spectral destiny peaks clustering method does not need predefine clustering parameters and iterative computation. The algorithm has high precision, low computational complexity and good robustness.
SUPER-PIXELS GRAPH

Super-pixels algorithm splits the image into pieces with more consistent inner regions, which can effectively reduce the image resolution and the complexity of subsequent image processing. The super-pixel algorithm has been widely applied in computer vision recently.

We construct graph \( G(V, E) \) with nodes \( V \) and edge \( E \) which represent super-pixels and the histogram correction of adjacent super-pixels, respectively. The histogram correction has greater accuracy than color average value in reflecting the similarity between adjacent regions. The edge is calculated as:

\[
A_{ij} = \begin{cases} 
  e^{-D(h_i, h_j)/\sigma} & \text{if } i, j \text{ is adjacent} \\
  0 & \text{else}
\end{cases}
\]  

(2)

\[
D(h_i, h_j) = \sum_{N=3}^{5} \lambda_N \sum_{k=0}^{255} \frac{2(h_{ik} - h_{jk})^2}{h_{ik} + h_{jk}}
\]  

(3)

Here \( \lambda_N \) is the weight of every color channel in histogram. The experiments show that results will be better when \( \lambda = \{0.5, 0.3, 0.2\} \). \( h_i \) is the relative histogram of super-pixel \( i \) in one color channel.

DENSITY PEAKS CLUSTERING

The innovative and efficient clustering algorithm is proposed by Alex Rodriguez and Alessandro Laio[13], which considers that the points with higher local density than surrounding points are more likely to be clustering centers. The method needs find the point with highest local density as the reference point to finish the clustering segmentation, the peak point is calculated as:

\[
\rho_{\eta} = \max_{i \in P} \rho_i
\]  

(4)

Where \( \rho_i = \sum_j e^{-d_{ij}} \), \( d_{ij} = \|X_i - \bar{X}_j\| \), \( \bar{X}_i \) is value of point \( i \) in vector space. Point \( \eta \) is considered as the center of maximum similar region in image. The distance of other points to the reference point \( \eta \) is called reference distance, which is defined as:
\[ \xi_i = d_{ij} \]  \hspace{1cm} (5) 

The effective distance \( \delta \), which is applied to determine whether this point is the clustering center, is the shortest distance between the current point and point set with maximum reference distance, and it is calculated as:

\[ \delta_i = \begin{cases} 
\xi_i & \xi_i = \max(\xi) \\
\min(d_{ij}) & \text{else} 
\end{cases} \]  \hspace{1cm} (6) 

The sets of \( \delta \) are sorted by descend, if the clustering number is \( m \), the clustering center is \( C \)

\[ C = \{P_{\delta_1}, P_{\delta_2}, \ldots, P_{\delta_m} | \delta_1 > \ldots > \delta_{m-1} \} \]  \hspace{1cm} (7) 

The clustering is finished according the nearest distance principle. The clustering result is \( R = \{R_1, \ldots, R_m\} \).

**SALIENCY DETECTION**

The saliency map is obtained through 3 steps, firstly the saliency map on the SLIC layer (layer I) is calculated, then the saliency map on spectral density peaks segmentation layer (layer II) is finished, finally two saliency maps are integrated to get the final results.

**LAYER I SALIENCY CUES**

We use the Markov model[14] based on saliency cues of border prior to get the saliency map. Hence, when dealing with over-segmentation layer by using regional contrast algorithm, we can effectively avoid detection results with only image borders or small region with high contrast.

Give the states of a set of points \( S = \{S_1, S_2, \ldots, S_n\} \), a Markov chain can be determined by the transition probability matrix \( P \), in which, \( P_{ij} \) is the transforming proba-
bility from state $S_i$ to state $S_j$. In Markov model, the point with invariable state ($P_{ii} = 1$) is considered as the absorbing node and there is at least one absorbing node, while the other points are considered as the transient nodes. Image edge super-pixels are considered as absorbing nodes because of photo habits. If there are $t$ transient nodes and $r$ absorbing nodes, the transition matrix $P$ of input image can be shown as:

$$P = \begin{bmatrix} Q & R \\ 0 & I \end{bmatrix}$$

(9)

Where $Q \in [0,1]^{t \times t}$ is the transition probability matrix to express the transition probability among transient nodes, $R \in [0,1]^{r \times t}$ is the absorbing probabilities matrix to express the transition probability between transition nodes and absorbing nodes, $0$ is a zero matrix with $r$ columns and $t$ rows and $I$ is the unit matrix. The transition times of a transient node to some absorbing nodes can be obtained based on transition probability matrix, then the value of the transient node relative to the absorbing node is calculated, and the transition times can be calculated as follows:

$$y = NC$$

(10)

Where $N=(I-Q)^{-1}$, $C$ is a $t$ dimensional all ones vector.

**LAYER II SALIENCY CUES**

Global uniqueness (GU) means that the region with high contrast in image has high saliency, which is the low level information applied in saliency detection and is suitable for layer II saliency detection especially under appropriate segmentation. The global uniqueness is defined as:

$$GU(i) = \sum_{j} \exp\left(-\frac{D(R_i,R_j)}{\sigma^2}\right) \times \omega_j \times D(h_i,h_j)$$

(11)

Here $D(h_i,h_j)$ is the histogram correlation of two regions. $\omega_j$ is area percentage of block j, and $\sigma^2 = 0.4$ as in Ref.[10] to allow distant regions to also contribute to the
global uniqueness, \( D(R_i, R_j) \) is the distance between two regions and it can be calculated as follows:

\[
D(R_i, R_j) = \left( \sum_{k \in R_i} \|\text{Sup}(k) - C_j\| + \sum_{k \in R_j} \|\text{Sup}(k) - C_i\| \right) / 2
\]

(12)

Different from the traditional methods to get the spatial distance between the center of two blocks \( C_i \) and \( C_j \), the method can effectively avoid the influence of irregular region, as shown in Fig.3. The traditional spatial distance \( D >> R_i, R_m >> D(R_u, R_m) \) does not reflect the reality.

![Figure 3. The influence of irregular shape.](image)

Spatial Center Degree (SCD) means that we usually place salient object near the center and set background around when we take pictures. The spatial center degree is calculated as:

\[
SC(i) = \exp\left( -\sum_{x,y \in R} \frac{\max((x-0.5)^2, (y-0.5)^2)}{|R|} \times \lambda \right)
\]

(13)

Where \((x,y)\) is the super-pixel center of region \( i \), \(|R_i|\) is the super-pixel quantity of region \( i \), \( \lambda \) is the weighting factor.

\[
S(i) = \text{Norm}(GU(i) \times SC(i))
\]

(14)
INTEGRITY DETECTION OF SALIENCY OBJECTS

The whole salient object should be marked in ideal saliency detection. But in natural images, small area with high contrast patches will be detected instead of whole object, and it is inevitable in saliency detection. Aiming to solve this notorious and universal problem, we propose a method to merge the patches based on their feature size, which is shown as:

\[
S_i = \arg \min_{W, H} (W_i, H_i) \times C_i
\]

\[
C_i = \frac{|R_i|}{W_i \times H_i}
\]

Where \( W_i \) and \( H_i \) are the width and the height of axis-aligned bounding box in local block, respectively. \( C_i \) is the block compactness.

The main principles to remove patches are shown as follows:

1. Only the image with the most saliency and the feature size smaller than predefined value \( S_i \) should be processed.
2. Only the objects with the maximum saliency in adjacent regions and the saliency are higher than two times of average should be merged.
3. Consider the new region after merging with the highest saliency.

CONCLUSIONS

We propose a bottom-up method to detect salient regions based on spectral destiny peaks clustering without predefining the clustering parameters or iterating. In the new method, we segment the input image twice by SLIC and spectral destiny peaks. We extract and integrate the salience cues under the layer information and size information. The experiments show that results from the new method are better than the previous best results (compared with 9 alternate methods).

REFERENCES

1. L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. IEEE T. PAMI, 20(11): 1254–1259, 1998.
2. J. Han, K. Ngan, M. Li, and H. Zhang. Unsupervised extraction of visual attention objects in color images. IEEE TCSV, 16(1): 141–145, 409, 414, 416, 2006.
3. C. Guo and L. Zhang. A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression. IEEE TIP, 2010.
4.  Z. Liu, R. Shi, L. Shen, Y. Xue, K. N. Ngan, and Z. Zhang. Unsupervised salient object segmentation based on kernel density estimation and two-phase graph cut. IEEE Transactions on Multimedia, 14(4): 1275–1289, 201.

5.  Y.-S. Wang, C.-L. Tai, O. Sorkine, and T.-Y. Lee. Optimized scale-and-stretch for image resizing. ACM Trans. Graph., 27(5), 2008. 1.

6.  L. Itti. Automatic foveation for video compression using a neurobiological model of visual attention. IEEE Trans. Image Process., 13(10): 1304–1318, 2004.

7.  G. Sharma, F. Jurie, and C. Schmid. Discriminative Spatial Saliency for Image Classification. In CVPR, 2012.

8.  J. Yang, M. Yang. Top-down visual saliency via joint crf and dictionary learning. In CVPR, 2012. 1.

9.  J. Xu, M. Jiang, S. Wang, M. S. Kankanhalli, and Q. Zhao. Predicting human gaze beyond pixels. J. Vision, 2014.

10.  R. Achanta, S. S. Hemami, F. J. Estrada, and S. Süsstrunk. Frequency-tuned salient region detection. In CVPR, pages 1597–1604, 2009.

11.  L. Kaufman, P. J. Rousseeuw. Finding Groups in Data: An Introduction to Cluster Analysis. vol. 344 (Wiley-Interscience, New York, 2009).

12.  P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 33(5):898–916, May 2011.

13.  Alex Rodriguez, Alessandro Laio. Clustering by fast search and find of density peaks. Science 344, 1492 (2014).

14.  Jingang Sun, Huchuan Lu, Xiuping Liu. Saliency region detection based on Markov absorption probabilities. in CVPR 2015, pp. 1639-1649.