Saliency Detection based on spectra Destiny Peaks Clustering

Ruochen Xia*, Guangwei Wangb, Xiang Chengc
School of Mathematics, Physics and Information Science, Zhejiang Ocean University, Zhoushan Zhejiang China
*Corresponding author e-mail:exrc9597@163.com, b gw_wang@outlook.com, c78468711@qq.com

Abstract. Saliency detection mapping the whole salient object by simulating the human visual system is one of the fundamental problems in computer vision. In this paper we proposed a novel method to map the spatial vector features of image points to spectra, and decompose an image into large scale perceptually homogeneous elements for efficient salient region detection, using spectra destiny peaks clustering. A multi-layer saliency mapping is built based the size of regions. The final saliency map is intergraded in a hierarchical model. This method can overcome the common problem in saliency detection that the detection accuracy could be adversely affected if salient foreground or background in an image containing small-scale high-contrast regions. The experimental results show that the proposed method outperforms are greatly improved greatly.

1. Introduction
Saliency detection in computer vision aims to find the most informative and interesting region in the scene by simulating the human visual system. Knowing the position of important region can pay more attention on and ignore the rest regions which can improve computational efficiency greatly. Saliency detection which gains much attention recently has been applied to numerous computer vision scenarios, such as objects-of-interest segmentation, image classification, image retargeting, image and video compression and so on.

Saliency detection algorithms can be broadly categorized as either bottom-up[1] or top-down methods according to the perspective of information processing mechanisms. bottom-up methods are stimuli-driven and get the visual saliency map by exploiting low-level cues such as color, textures, spatial distribution, frequency and so on. In contrast, top-down models are task-driven and require supervised learning with manually.

labeled ground truth. To better distinguish salient objects from background, high-level information and supervised methods are incorporated to improve the accuracy of saliency map.

In this paper, a new bottom-up saliency detection based on spectra destiny peaks clustering is proposed. Fist the image is over-segmented by superpixels technology, and graph is constructed with superpixels as nodes is mapped to spectra space. The image is segmented by spectra destiny peaks clustering. Second muti-layers are constructed and every layer saliency cues are abstracted. Last final saliency map are get by integrating the layer cues. The main contributions of this paper are 1) Image segments based on spectra destiny peaks clustering does not need predefined clustering parameter. 2) With our multi-level analysis and hierarchical inference, the model is able to deal with salient small-scale structure, so that salient objects are labeled more uniformly. 3) several refinement techniques are proposed improving the detection accuracy.
2. Related Work

Bottom-up saliency detection commonly interpreted in computer vision as detection and segmentation. The most standing out objects in the image using low-level visual information has long development history and is widely used in computer vision.

Local methods segmenting the image into several homogeneous regions, extracting saliency cues from regions and getting the final saliency map are the most commonly used in saliency detection at present? F. Liu et al\(^{[2]}\) get the saliency cues from multi-scales segmentation regions, and average the every scale saliency for the final result. Cheng et al. proposed RC method to consider region unique contrasting current region to the other regions as the most important factor, and get good detection results. Local methods are influent by the segmentations fulfilled the need are inexistence greatly. Lots of local methods based on superpixels over segmentation proposed such as graph theory \(^{[3]}\), Markov, probabilistic models \(^{[4]}\)and so on.

High-level priors are also commonly used based on common knowledge and experience, such as face prior \(^{[5]}\), center prior \(^{[6]}\). Recently saliency detection methods simulating human visual system getting the saliency map based on whole objects in the image \(^{[7]}\)have been developed.

3. Spectra Destiny Peaks Clustering

Image segmentation is the bridge of image processing to image analysis, is the key to saliency detection. Clustering as an efficient data analysis method has important applications in image segmentation such as K_Means, Mean_shift, k_mededoi and so on, but those methods have inevitable shortcomings such as: 1) color similarity and spatial similarity are considered only and the spatial continuity is ignored in those clustering methods, which will lead to discontinuous segmentations. 2) The initial clustering centers must be predefined and the local optimal solution can be obtained in most cases. 3) Multiple iterations for clustering results, increase the complexity of the algorithms.

In this paper, a new spectra destiny peaks clustering is proposed to improve the image segment.

3.1. Spectra Clustering

The input image with n nodes \(P=(P_1,\cdots,P_n)\) will be segmented into m areas using spectra clustering, the segmentation process is:

**Step1:** Construct a graph \(G=(V,E)\) with nodes as \(V\), and \(E\) is a set of undirected edges which contains the connectivity between two nodes, the value is:

\[
E_{i,j} = \omega_{ij} = \begin{cases} 
    e^{-\|x_i-x_j\|^2/\sigma^2} & \text{i,j is adjacent} \\
    0 & \text{else} 
\end{cases}
\]

where \(x_i\) and \(x_j\) are the value vector of two nodes, and \(\sigma\) is a constant that controls the strength of the weight. The affinity matrix of \(G\) is \(W=[\omega_{ij}]_{n\times n}\).

**Step2:** Compute the Laplace matrix of \(G\) as \(L=D^{-1/2}WD^{1/2}\), where \(D=diag(d_{11},\cdots,d_{nn})\), \(d_{ii} = \sum\omega_{ij}\).

**Step3:** Get the eigenvalues \((\lambda_1,\cdots, \lambda_n)\) of the matrix \(L\), and \(\lambda_1 > \cdots > \lambda_n\). The corresponding eigenvectors are \(X=[X_1,\cdots, X_n]^{T}\).

**Step4:** Construct the matrix \(R=[X_1,\cdots, X_n]^{T}\) extracting the first m dimensions from the eigenvectors matrix \(X\). Every row of matrix \(R\) is considered as the node value vector, and all nodes are clustered into m clusters by K_Means.

Spectral clustering belongs to the category of the graph theory, and inevitably exist the algorithm complexity increases greatly as the increasing number of input nodes. Clustering results are affected significantly by predefine value \(m\) which is difficult to accurate predetermined in a complex scene.
3.2. Spectra destiny peaks clustering

In this method, the clustering results are getted without predefining the value m. The algorithm has high precision, low computational complexity characteristics.

SuperPixel graph: Superpixel algorithms group pixels into perceptually meaningful atomic regions can effectively reduce the image resolution and reduce the complexity of subsequent image processing tasks. The superpixel algorithm has been widely application in computer vision recently. Simple linear iterative clustering (SLIC) method has the outperforms of high rapid segmentation, high boundary recall, computing complexity is not affected by superpixels quantity, will be worked well in the application.

Based on the superpixels, we represent the image as a graph $G = (V, E)$, where the vertices $V$ are represented by the superpixels, and $E$ is a set of undirected edges which contains the connectivity between two superpixels. The corresponding vector matrix $X$ is computed by the method mentioned in the former chapter.

density peaks clustering: This Innovative and efficient clustering algorithm is proposed by Alex Rodriguez and Alessandro Laio[9], considers that some point has higher local density than surrounding points and far away from points with higher local density than this point will be considered as clustering center. The method need find the point with highest local density as the reference point:

$$\rho_i = \max_{i \in P} \rho_i$$

(2)

Where $\rho_i = \sum_j e^{-d_{ij}}$, $d_{ij} = \|X_i - X_j\|$. $X_i$ is vector of point $i$ in spectra space. The distance to the reference point $n$ called reference distance, that is:

$$\xi_i = d_{in}$$

(3)

The effective distance $\delta$ is the key factor. Whether this point belongs to the clustering center set. The value of $\delta$ is the shortest distance to higher reference distance.

$$\delta_i = \begin{cases} 
\xi_i & \text{if } \xi_i = \max(\xi) \\
\min(d_{ij}) & \text{else}
\end{cases}$$

(4)

The sets of $\delta$ are sorted by descend, if the clustering number $m$ is determined, the center sets are

$$C = \{P_{n_1}, P_{n_2}, \ldots, P_{n_m}| \delta_1 > \cdots > \delta_{m-1}\}$$

(5)

The clustering is finished by the principle every point is assigned to the nearest center point. The clustering result is $R = \{R_1, \ldots, R_m\}$.

Clustering parameter: According to input image feature automatically determine the clustering parameters is a key factor guaranteeing the algorithm robustness. Principal Component Analysis (PCA) is an effective tool for reduction of data dimension. Main Idea of the algorithm is finding the principal axis of datas and constructing a new coordinate system based on the principal axis, re-projecting the input datas to this coordinate system forming new data sets which can be represented by part of dimensions with least error at same dimensions, further more margin of err can be calculated by the power of each dimension.

Matrix $X$ is the projection of matrix $L$ whose eigenvalues are $\lambda = \{\lambda_1, \cdots, \lambda_n\}$ sorted by descend calculating from affinity matrix $W$ is raw data vector set of input image. The eigenvalue $\lambda_i$ can quantitatively describe how important the $ith$ dimension of the vector in $X$ is. The experiments show that 90% energy can get better clustering effect. Percent of

Energy extraction ways is showed as follows:

$$\chi = \arg\min_j \{\sum_{i=1}^j \lambda_i \sum_{i=1}^n \lambda_i > 0.9\}$$

(6)
Where $\lambda_i = (\lambda_i / \text{max}(\lambda_i))^2$, and in the spectra density peaks clustering $\mathcal{Z}$ can be used as the dimension length in PCA methods and clustering number. In order to avoid the segment results the effected by special images, the value of $\mathcal{Z}$ is set to $\mathcal{Z} \in [5,32]$.

Saliency Detection

The saliency map is get through 3 steps, first the hierarchical structure is built according to the size of segmented regions, second saliency cues are computed for each layer. They are finally fused into one single map.

3.3. Hierarchical structure

All salient object should be marked in ideal saliency detection, but in natural images small area but high contrast patches will be detected instead of whole object, that is inevitable insaliency detection with the single scale model. Aiming to solve this notorious and universal problem, we propose a hierarchical model according to the theory of psychology the selection process in human attention system operates from more than one levels, and the interaction between levels is more complex than a feed-forward scheme. In the process of structuring one level the feature size of block less than the layer size thresh will be merged to the most similar adjacent block, as showed:

\[
S_{\text{scale}}(\mathcal{I}_i) = \arg \min_S (S \geq \tau_i)
\]  

(7)

Where $\tau_i$ is the layer size thresh, $S$ is the feature size is calculated balancing the accuracy and efficiency as showed:

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

(8)

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

(9)

Where $W_i, H_i$ are the width and the height of axis-aligned bounding box of this block, $C_i$ is the compactness. Themulti-layers segmented result is showed in figure 1.

![Fig.1 multi-layers segment results](image)

(a) Image (b)Layer I (c)Layer II (d)Layer III

3.4. Single Layer Saliency Cues

Global uniqueness (GU). Visual uniqueness in terms of high contrast to other image regions is believed to be the most important indicator of low-level visual saliency. The current block global uniqueness is defined as:

\[
GU(i) = \sum_{j \neq i} \exp \left(-\frac{D(R_i,R_j)}{\sigma^2} \right) \cdot \omega_j \cdot \| \tilde{c}_i - \tilde{c}_j \|
\]  

(10)

$D(R_i,R_j)$ Calculating as Eq.(11) is different to the traditional measurements getting the spatial distance based on centroid of two blocks $C_i$ and $C_j$, which can avoid the influence of irregular shape, as showed in Fig.2. $\tilde{c}_i$ and $\tilde{c}_j$ are the color mean of those blocks, $\omega_j$ is area percentage of block j, and we use $\sigma^2 = 0.4$ as in to allow distant regions to also contribute to the global uniqueness.
Fig. 2: the influence of irregular shape. The traditional spatial distance $D(R_i, R_{ii}) > D(R_{ii}, R_{i})$ doesn’t reflect the reality.

Spatial Center Degree (SCD). The psychologists researches show that when we take a picture, salient object will be placed near the center surrounding around the background. The spatial center degree is calculated as:

$$SC(i) = \exp\left(-\sum_{x,y\in R_i}(x-0.5)^2 + (y-0.5)^2 \times \lambda\right)$$

(11)

Where $x,y$ is the center of suppixel in the block $i$, $|R_i|$ is the superpixels quantity in block $i$, $\lambda$ is the weighting factor.

Single layer saliency integrated GU and SCD and normalized as showed:

$$\hat{S}(i) = \text{Norm}(GU(i) \times SC(i))$$

(12)

3.5. Saliency Cues Integration

The single layer cues produce the current saliency map at one scale. As also discussed in, combining individual saliency maps using weights may not be a good choice, since better individual saliency maps may become worse after they are combined with others. Through the analysis of the size of each salient object in some salient dataset with ground truth annotations, the integration method balancing the precision and robustness is due to their small size threshold, all mergers will produce within the specified size in layer I. Mergers satisfying One of the two conditions will be happened in layer II, and Mergers satisfying $2\text{th}$ will be happened, two conditions showed as:

$$\begin{cases} 
|c_i - c_{j}| \leq c_{i} \text{ condition I} \\
\min(S, S_{j}) \geq S' \text{ condition II}
\end{cases}$$

(13)

The saliency detection results using this method showed as Fig. 3.
4. Experiments

We exhaustively compare our results on the most widely used public available dataset with pixel accurate ground truth annotations, and compare with 10 alternate methods IT, SR, FT, LC, HC, RC, GB, SF[10], SIA, BS[4]. Results of the alternative methods are obtained by the original authors or the authors’ publicly available software running results. For objective evaluation, we use two evaluation standards of the precision recall analysis and mean absolute errors. The main experiment parameters are set as: the superpixels num is 250, the value of $\tau_1, \tau_2, \tau_3$ are 10, 20, 50.

4.1. Precision and recall

Precision measures the percentage of salient pixels correctly assigned, while recall measures the percentage of salient pixel detected. There is no necessary correlation between precision and recall but some certain restrictive correlations are existing, so precision/recall can effectively reflect the performance of the algorithm.

Saliency maps are binarized at each possible threshold within range [0, 255] in fixed thresh, each resulting in a precision and recall value. The resulting precision recall curves is shown in Fig. 3(a), and binarized at value of 2 times of mean saliency map value. In many application high precision and high recall are both required, so we thus estimate the F-Measure as:

$$F_\beta = \frac{(1 + \beta^2) \cdot \text{precision} \cdot \text{recall}}{\beta^2 \cdot \text{precision} + \text{recall}}$$

(14)

Where $\beta^2$ is set to 0.3. The experiment results of adaptive Thresholding is showed in Fig.4.

4.2. Mean absolute error

Saliency maps must be binarized when precision/recall is used to measure algorithm performance. Lots of image information will be lost in binarization, and the detection effects are significantly influenced by threshold value. In paper, results show the limitation of precision/recall metrics. For a more balanced comparison, the mean absolute error (MAE) evaluating the mean err between a continuous saliency map $S$ and the binary ground truth $G$ for all image pixels $I_x$, is proposed as a new saliency detection metric, defined as:

$$MAE = \frac{1}{|I|} \sum |S(I_x) - G(I_x)|$$

(15)

Where $|I|$ is the number of image pixels. The experiments show our method has lowest MAE value, result is showed in Fig.5.
Fig. 6 Mean absolute error

Fig. 7 Saliency detection results of different methods

(a) Image (b) GT (c) LC (d) FT (e) HC (f) RC (g) SIA (h) SF (i) BS (i) Ours

Table 1. Average run time for images in the benchmark (most images have resolution 300 × 400).

| method   | RC  | SF  | BF  | SIA | OURS |
|----------|-----|-----|-----|-----|------|
| Time(s)  | 0.204 | 0.221 | 5.1 | 0.151 | 0.302 |
| code     | C++ | C++ | matlab | C++ | matlab |

4.3. Run Time

We compare the performance of our method in terms of run time with methods with most competitive accuracy (SF, SIA), or based on the similar region segmentation method (BF, RC). Computed results are presented in Table 1 based on a machine with Intel Dual Core i7 5500U CPU, and 4GB RAM. The algorithm proposed in this paper has lower complexity algorithm, can satisfy the actual needs.

5. Conclusion

We propose a bottom-up method to detect salient regions based on spectra destiny peaks clustering without predefining the clustering parameters and iterating for accurate result. The multi-layers built
based on the size of segmentation can resolve the fundamental problem that small-scale structures would adversely affect salient detection. The experiments show that our proposed method achieves high performance and broadens the feasibility to apply saliency detection to more applications handling different natural images.

Reference

[1] Yao Qin, Huchuan Lu, Yiqun Xu and He Wang (2015) Saliency Detection via Cellular Automata, CVPR.
[2] F. Liu and M. Gleicher (2006). Region enhanced scale-invariant saliency detection. ICME, pp. 1477–1480.
[3] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang (2013). Saliency detection via graph-based manifold ranking. CVPR.
[4] Y. Xie, H. Lu and M.-H. Yang (2013). Bayesian saliency via low and mid level cues. IEEE TIP, vol. 22, no. 5.
[5] S. Goferman, L. Zelnik-Manor and A. Tal (2010). Context-aware saliency detection. IEEE CVPR, pages 2376–2383.
[6] G. Sharma, F. Jurie and C. Schmid (2012). Discriminative spatial saliency for image classification. IEEE CVPR, pages 3506–3513.
[7] P. Isola, J. Xiao, A. Torralba, and A. Oliva (2011). What makes an image memorable? IEEE CVPR, pp 145–152.
[8] B. M’t Hart, H. C. Schmidt, C. Roth and W. Einh¨auser (2013). Fixations on objects in natural scenes: dissociating importance from salience. Frontiers in psychology, vol. 4.
[9] Alex Rodriguez and Alessandro Laio (2014). Clustering by fast search and find of density peaks. Science, Vol. 344, Issue 6191, pp. 1492-1496.
[10] F. Perazzi, P. Kr¨ahenb¨uhl, Y. Pritch and A. Hornung. Saliency filters (2012). Contrast based filtering for salient region detection. IEEE CVPR, pages 733–740.