Image Salient Object Detection Based on K-means and Level Set Superpixel Segmentation

Jindiao Huang, Xing Jin, Jingjing Zhang
School of Automation, China University of Geosciences, Wuhan 430074, China
huangjindiao@cug.edu.cn

Abstract. In order to detect and segment the salient objects in digital images to solve complex machine visual problems, this paper proposes an image salient object detection algorithm based on K-means and level set superpixel segmentation. The algorithm segments a given target image into multiple superpixel regions with similar features to abstract unnecessary details in the images, and reduce the number of colors in all superpixels by Histogram acceleration to increase computational efficiency. The saliency map obtained by calculating the distance of all superpixel in the Lab color space, and optimizes the detection effect by background prior and multi-scale spatial fusion. A large number of experiments show that the saliency detection method proposed in this paper is superior to algorithms such as SR, AC, FT, MSS, LC, CA, SF, HC and RC in accuracy, recall rate, AUC value and average absolute error.

1. Introduction
Visual saliency is the ability of the human visual system to quickly search and locate the objects of interest and selectively ignore the areas of disinterest when facing natural scenes. It is also an important mechanism for human daily life to process visual information. An image salient object detection method which simulates human visual characteristics by algorithm, and allocate resources to more important areas in computer vision tasks. Saliency object detection has many applications in the field of computer vision, including object detection and recognition [1-2], image compression [3-4], image segmentation [5-6] and visual tracking [7], and other fields.

Visual saliency derives from the singularity, uniqueness, scarcity and unpredictability of vision, and has long been studied in many fields such as cognitive psychology [8], neurobiology [9] and computer vision [10]. The changes of color, lightness, angle, gradient, edge and boundary in visual scenes will attract the attention of human vision. Itti and Wolfe et al. put forward the theory of human visual attention mechanism in combination with previous experience: Human visual system usually goes through two main stages when processing visual information. Pre-attention stage, the visual system can simply and quickly detect different areas of the scene. Attention stage, the visual system perceive the fine areas in the scene and extract richer visual information.

2. Related Works
Visual saliency detection includes two main research directions: classical model, which establishes detection model through pure mathematical calculation method, mainly using color, lightness and position and other features; In the model based on deep learning, with CNN or FCN as the bottom layer, add some external components and lots of data for training to obtain a deep neural network framework that can automatically detect salient areas of the image. The detection accuracy of deep
learning model is higher than that of classical model, but it relies too much on data set, and the detection effect will become worse in special scenes.

AC, FT and MSS [11-13] are typical pixels-level comparison models. Their innovation is to transform images from RGB color space to Lab color space. Lab color space is very close to human visual perception and has an extremely wide gamut, making color contrast more accurate. The AC method obtained the saliency value of pixels by calculating the Euclidean distance between all pixels and the average value in the window of images h/2, H /4 and H /8. The MSS method obtains the saliency value of pixels by calculating the Euclidean distance between each pixel and the average value of its maximum symmetric space. The FT method directly calculates the Euclidean distance between each pixel and the average value of the whole image to obtain the saliency value. All the three algorithms only consider the color feature and cannot make full use of the internal information of the image.

CA and SF [14-15] are typical superpixel comparison model, its innovation is to segment the image into several local, compact and uniform super-pixels in the Lab color space by the SLIC [16] method. The two algorithms consider both color and spatial features to improve the calculation efficiency and detection accuracy, but SLIC cannot fully segment images and the segmentation efficiency is not high.

RC [17] is a typical regional comparison model, and its innovation is to use the graph-based segmentation method [18] to divide the image into several connected regional and calculate the color distance and spatial distance between each area to obtain the saliency value of the region. This method fully divides the image, speeds up the calculation speed and improves the detection accuracy.

The method proposed in this paper improves and innovates on a variety of classical methods to improve the detection accuracy. Its innovations include:

1. Level set combined with K-means clustering has higher accuracy and better connectivity in image segmentation, so as to improve calculation speed and overall detection efficiency.
2. Combined with background prior, the saliency of foreground area is improved. Compared with the method of using image edge as background, the background area obtained through level set is more accurate and the error rate is lower.
3. By fusing multi-scale space, the saliency of background area is inhibited, and the overall performance of detection is further improved.

In this paper, calculate the significance value by global comparison of all the segmented regions, compared with the local comparison method, the global comparison can better separate a wide range of targets from the surrounding environment, and uniformly show the significance area of the image.

3. Saliency Map Calculation
This section will introduce the specific process of level set superpixel segmentation and K-means cluster re-segmentation, as well as the calculation method of regional significance value and the details of regional color quantization. At the same time, analyse how to improve the detection results by background prior and multi-scale fusion.

3.1. Level set superpixel segmentation
Most of the current mainstream image saliency detection methods are based on the comparison between superpixels or regions, which can abstract the unnecessary details in the image, highlight the main object of the image, improve the detection effect, and reduce the calculation time compared with the pixel level. Fig.1 shows the image segmentation methods commonly used in saliency detection, in which Fig.1 (e) and Fig.1 (f) are the methods used and improved in this paper.

Fig.1 (b) is the effect diagram of SLIC segmentation, as you can see the image is divided into multiple pixel blocks with similar features. Although SLIC method can solve the feature jump deviation near the image target outline well, however, due to the artificial selection of the number of super pixels, the features of multiple superpixels obtained from the segmentation of large regions with the same features are basically similar, and will increase the time complexity of the detection algorithm.
Fig. 1 (c), (d) and (e) are effect diagram of graph-based segmentation, the image is divided into several main areas, and pixels with similar features are distributed to the same area. The segmentation effect is better than SLIC method, which can reduce the number of comparisons and improve detection efficiency when calculating the saliency value of the area. Where, $c$ is the number of regions, $m$ is the minimum number of pixels in a region, we can see that there is a strong correlation between the segmentation effect and the selection of parameters, different parameters have different segmentation effects and don’t have robustness.

![Fig.1 The image segmentation method used in saliency detection](image)

Aiming at the existing problems in the current image segmentation algorithm, this paper combines the level set and k-means clustering for image segmentation. Firstly, segment the target image by CV (Chan-Vese) level set geometric active contour model \[19\] to obtain the original superpixel block. The basic idea of the level set is to embed the lower one-dimensional curve into the higher one-dimensional level set function. During the curve evolution, the level set function is constantly updated to achieve the segmentation of the image contour, which has the advantage of maintaining the same topological structure during the evolution process. According to Fig. 1 (e), the superpixels based on level set segmentation has higher connectivity and accuracy than other methods, which greatly saves the computing resources of regional saliency detection in the later stage. There are two main features of horizontal set image segmentation: automatic segmentation, no need to set parameters, high universality, strong stability and robustness; to some extent, the background region can be directly obtained, and the detection results can be improved by combining the background prior. Secondly, the superpixels are further segmented by k-means clustering method to obtain more details of the image and improve the detection effect. As shown in Fig. 1 (f), by comparing the original figure, the segmented image by clustering is basically consistent with the detailed information of the original image.

### 3.2. Superpixels saliency calculation

When the image is segmented into $M$ superpixel regions $R_i (i=1,...,m)$, for any region $R_k$, its saliency value is determined by the sum of the color difference between the region and other regions. The calculation formula is shown in Equation (1).

$$ S(R_k) = \sum_{i=1 (i \neq k)}^{M} W(R_k)W(R_i)\text{Diff}_i(R_k, R_i) $$

(1)

$W(r_i)$ is the weight of the region $R_k$, which is related to the number of pixels, $\text{Diff}_i(R_k, R_i)$ is the color difference of the two regions, and the calculation formula is shown in Equation (2).

$$ \text{Diff}_i(R_k, R_i) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} F(c_{1,i})F(c_{2,j})\text{Dist}(c_{1,i}, c_{2,j}) $$

(2)

Where $F(c_{1,i})$ is the probability that the $i$th ($i=1,...,n_1$) color $c_{1,i}$ appears in the first region, and $F(c_{2,j})$ is similar to it. We can see, the longer the time and the lower the efficiency will be when the number of colors in the region is larger. Therefore, the key to speed up the calculation is to reduce the
total number of colors in the image, so this paper uses a histogram acceleration method to reduce the number of colors in the image.

Fig. 2 Compared with other classic saliency detection methods

3.3. Background prior
Combined with background prior, the saliency of foreground region can be improved. Most of the background areas in the natural image are continuous and transform gently, so in the level set segmentation, only the outline of the foreground area will be captured in most cases. For any foreground region $R_p$, the calculation formula of saliency value is shown in Equation 3.

$$ S(R_p) = \sum_{b=1}^{M_b} W(R_p) W(R_b) \text{Diff}_s(R_p, R_b) $$

(3)

Where, $R_b$ is the background region and $M_b$ is the number of background regions.

3.4. Multi-scale spatial fusion
The background area of the image will be blurred in multi-scale space, but the overall features of the foreground area will still be retained. Therefore, when the image is converted to multi-scale space, the background saliency in high-scale space will decrease while the foreground saliency will remain unchanged. In this section, combined with multi-scale fusion, the saliency of the image background area is reduced. The calculation formula is shown in Equation (4).

$$ \overline{S} = \frac{1}{M} \sum_{r \in R} S'^r $$

(4)

Where, $S'^r$ represents the saliency map obtained at the scale of $R$, and $M$ represents the number of scales in multi-scale space. The mean value of the saliency map at each scale is defined as the final saliency map. The scale space selected in this paper is $R = r, r / 2, r / 4$.

4. Results
In this paper, a large number of experiments were carried out in the database by comparing with several classical saliency detection algorithms: SR, AC, FT, MSS, LC, CA, SF, HC and RC. The saliency detection results of some images are shown in Fig. 2, in which KLS is the method we
proposed. It can be seen from the detection results that the detection effect of KLS method is better than other methods in terms of subjective feelings. We can see that the foreground area and background area of the saliency map obtained by KLS are effectively isolated.

Fig. 3 Evaluation index of saliency detection

In this paper, four quantitative indexes commonly used in saliency detection, PR curve, ROC curve, AUC value and Mean Absolute Error, are used to show more intuitively and concretively that our method is better than other classical algorithms. As can be seen from Fig. 3, the method proposed in this paper is superior to other methods in all indicators except that the MAE value is slightly lower than RC.

5. Conclusion
In the image salience detection method proposed in this paper, the superpixel region obtained is more accurate, has stronger connectivity, and can produce high quality saliency map. The experimental results show that the method proposed in this paper is more effective than other methods.

In future work, the level set segmentation algorithm will continue to be improved to obtain more accurate superpixel regions, and the prior knowledge such as face recognition and digital recognition will be added to obtain higher quality saliency map.

References

[1] Shen H, Li S, Zhu C, Chang H, Zhang J. Moving object detection in aerial video based on spatiotemporal saliency. Chinese Journal of Aeronautics, 2013,26(5): 1211-1217.
[2] Ren Z, Gao S, Chia L T, Tsang W H. Region-Based Saliency Detection and Its Application in Object Recognition. IEEE Transactions on Circuits and Systems for Video Technology, 2014,24(5): 769-779.
[3] Guo C, Zhang L. A Novel Multiresolution Spatiotemporal Saliency Detection Model and Its Applications in Image and Video Compression. IEEE Transactions on Image Processing, 2009,19(1): 185-198.
[4] Itti, L. Automatic Foveation for Video Compression Using a Neurobiological Model of Visual Attention. IEEE Transactions on Image Processing, 2004,13(10): 1304-1318.
[5] Johnson-Roberson M, Bohg J, Bjorkman M, Kragic D. Attention-based active 3D point cloud segmentation. IEEE/RSJ International Conference on Intelligent Robots Systems. IEEE, 2010.
[6] Qin C, Zhang G, Zhou Y. Integration of the saliency-based seed extraction and random walks for image segmentation. Neurocomputing, 2014,129: 378-1391.
[7] Borji A, Frintrop S, Sihite D N. Adaptive object tracking by learning background context. Computer Vision and Pattern Recognition Workshops (CVPR), 2012 IEEE Computer Society Conference on. IEEE, 2012.
[8] Wolfe J M, Horowitz T S. Opinion: What attributes guide the deployment of visual attention and how do they do it?. Nature Reviews Neuroscience, 2004,5(6): 495-501.
[9] Moore T, Zirnsak M. Neural Mechanisms of Selective Visual Attention. Annual Review of Psychology, 2017,68(1): 47-72.
[10] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis. IEEE TPAMI, 1998, 20(11): 1254-1259.

[11] Achanta R, Süsstrunk S. Saliency detection using maximum symmetric surround. in IEEE International Conference on Image Processing, 2010.

[12] Achanta R. Salient Region Detection and Segmentation. 2008.

[13] Achanta R. Frequency-tuned salient region detection. in IEEE Conference on Computer Vision Pattern Recognition. 2009.

[14] Stas G, Lihi Z M, Ayellet T. Context-aware saliency detection. IEEE Trans Pattern Anal Mach Intell, 2012, 34(10): 1915-1926.

[15] Perazzi F. Saliency Filters: Contrast Based Filtering for Salient Region Detection. in IEEE Conference on Computer Vision Pattern Recognition. 2012.

[16] Achanta R, Shaji A, Smith K. SLIC Superpixels Compared to State-of-the-Art Superpixel Methods. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(11): 2274-2282.

[17] Cheng M M. Global Contrast Based Salient Region Detection. in Computer Vision and Pattern Recognition. 2011.

[18] Felzenszwalb P F, Huttenlocher D P. Efficient Graph-Based Image Segmentation. International Journal of Computer Vision, 2004, 59(2): 167-181.

[19] Chan T F, Vese L. Active contours without edges. Image processing, IEEE transactions on, 2001, 10(2): 266-277.