Non-interactive object extraction based on saliency detection

Ting Pa, Xuan Wang, Shouxun Liu
Information Engineering School Communication University of China Beijing, China

Abstract. In traditional content-based image retrieval (CBIR), it is very important to segment the regions of interest for retrieval. In order to realize the non-interactive target region segmentation, this paper proposes the object segmentation based on significance region detection. The Grab Cut segmentation algorithm is initialized through the detection of the significance region, and the interactive operation is removed. The segmentation precision of the algorithm proposed in this paper is 89.51%.

1. Introduction
With the rapid development of internet technology, a large number of image information rely on the Internet for rapid dissemination, sharing and even management. Due to the wide distribution, fast growth and large number of image information, it is more and more important to accurately and quickly retrieve the target image from the image library. Nowadays, the main research direction of image retrieval is content-based image retrieval technology (CBIR) [1], This technique automatically analyzes the image content by the machine, uses the visual features of the image (color, texture, shape, etc.) to establish the index, and retrieves the image according to the similarity comparison of feature vectors. The content of retrieval is usually part of the region of human interest. The accurate segmentation of the object of interest will greatly affect the accuracy of the retrieval image.

This paper mainly studies the automatic target segmentation technology applied to image retrieval. Before image retrieval, significant region detection is used to find the retrieval target regions of interest, and automatic image segmentation and retrieval target extraction are realized. In this way, the interference of background pixels on retrieval results can be reduced and the precision of retrieval can be improved.

Visual significance technology is proposed based on the visual attention mechanism. By detecting more significant targets and removing the interference of background and noise, the purpose of highlighting the target is achieved [2]. Literature [3] divides them into three categories: 1) significance analysis algorithm based on underlying visual features. The representative algorithm is the Itti algorithm proposed in [4]. 2) Pure mathematical calculation methods that are not based on any principle of biological vision, which mainly include the full resolution algorithm (AC algorithm) proposed in [5] and the residual spectrum algorithm (SR algorithm) based on spatial frequency domain analysis in [6]. 3) Merge the first two, the representative algorithm is the algorithm based on graph theory (GBVS) [7]. This algorithm is similar to Itti algorithm to simulate the principle of vision in the process of feature extraction, but markov chain is introduced in the process of significant graph generation to obtain the significant value by pure mathematical calculation. Combining visual significance detection technology in image segmentation can not only improve the visual mechanism of segmentation results, but also quickly determine the segmentation target and improve the segmentation efficiency.

There are many common image segmentation methods, including threshold segmentation, fuzzy method, edge detection and feature space clustering [1]. According to the different characteristics of the segmentation method, it can be roughly divided into image segmentation based on threshold, image
segmentation based on edge, image segmentation based on region, image segmentation based on model, image segmentation based on graph theory and image segmentation based on artificial intelligence. Although there are many methods for image segmentation, it is not easy to segment images correctly. At present, there is no effective segmentation algorithm for all images. In recent years, researchers have done a lot of research on image segmentation, Graph Cut [8] is an efficient algorithm, and it is an interactive segmentation algorithm. The user marks some foreground and background pixels as seed pixels, and then compares the gray value of unlabelled pixels with that of labelled pixels to calculate the minimum energy value of image segmentation. Graph Cut algorithm has been studied and applied in many aspects, such as image segmentation [9] and clothes [10], facial or posture restoration [11], etc. Rother and other scholars extended Graph Cut method in 2004, and this new algorithm is Grab Cut [12]. This algorithm adopts the method of "incomplete labelling" to simplify the required user interaction, the user can simply place a rectangle loosely around an object to extract the object. Compared with Graph Cut algorithm, Grab Cut mainly has the following three differences: First, the target and background model of Graph Cut is the grayscale histogram, and Grab Cut replaces the mixed Gaussian model GMM of RGB three-channel; Second, Graph Cut minimizes the energy (segmentation) at one time, while Grab Cut replaces an interactive iterative process of continuous segmentation estimation and model parameter learning. Third, Graph Cut requires the user to specify some seed points for the target and background, but Grab Cut only needs to provide a set of pixels for the background area. In other words, Grab Cut allows incomplete annotations.

Grab Cut algorithm has an ideal segmentation effect for images in the image library to be segmented in this paper, and the specific effect will be shown in the third part of this paper. However, Grab Cut still has a certain amount of user interaction and cannot automatically select and segment the target region. In order to solve this problem, this paper proposes an automatic Grab Cut segmentation algorithm based on GBVS significance model on the basis of sufficient experimental research on image features of the subject.

The rest of this article is organized as follows: The second part introduces the technical background, which include the Grab Cut and GBVS significance model. The third part introduces the background of the subject and the specific features of the images in the design image library. Then this part introduces the methods proposed in this paper. The fourth part will show the results of the algorithm experiment in this paper. The fifth part summarizes this paper and looks forward to the future work.

2. Technical Backgrounde

2.1. GBVS
The leading models of visual saliency may be organized into the three stages:
(s1) extraction: extract feature vectors at locations over the image plane
(s2) activation: form an "activation map" (or maps) using the feature vectors
(s3) normalization/combination: normalize the activation map (or maps, followed by a combination of the maps into a single map).

GBVS is a bottom-up model of visual significance, consisting of two stages: the first stage is to extract the feature map and create the activation map, construct a fully connected graph of a node for each pixel, and make use of the difference between pixels in the feature map and its distance weighted directed edge. The weight is normalized to 1 as the transition probability of the markov chain, and then the activation map is calculated according to the equilibrium distribution of the resulting chain. The second stage uses the calculated activation map and distance to construct another diagram. Again, the graph is used as a markov chain, and the normalized mapping is derived from an equilibrium distribution.

2.2. Grab Cut Algorithm
The idea of Grab Cut algorithm is to map the entire image to the s-t network graph based on the graph cutting theory, where the source point s represents the foreground endpoint, the confluence point t represents the background endpoint, and the edge set E contains two parts: The weight of the t-links of the junction point of the source point and all other nodes is determined by the "regional energy item",
which reflects the similarity between the pixel point and the foreground and background; The weight of edge n-links between adjacent nodes in the figure is determined by the "boundary smoothing energy item", reflecting the color difference between adjacent pixels.

After the image is constructed into a network graph, the foreground and background samples are set by user interaction. Then searches for points in the image and calculates the distance to the user’s specified foreground and background, as well as the distance between them in pixels. Taking the linear combination of these two distance as the segmentation energy weight of the edge in the graph, the image segmentation problem is transformed into problem of finding the minimum energy of the edge in the graph. The energy function model of s-t network is constructed, and the minimum cut of graph is solved by the minimum cut algorithm of maximum flow in graph theory, that is, the optimal solution of corresponding energy function.

The segmentation effect of this algorithm for the research object is shown in figure 1 a-d.

![Figure 1](image)

**Figure 1.** Different images and their segmentation results by Grab Cut

3. **Proposed Algorithm**

3.1. **Project Background**

This project is applied to develop a retrieval system for advertising images of bus stops. In order to improve the accuracy of image retrieval segmentation of early target recognition research. There is no general algorithm for image segmentation, so this paper proposes Grab Cut automatic segmentation algorithm based on GBVS significance model.

The main features of the image studied in this paper are as follows:

1) The image background is complex, and part of the image target area is blocked. the background of the target area is mainly urban environment, including buildings, green vegetation, etc.;
2) The target image is greatly affected by light shadow and shooting Angle;
3) Image color information is rich;
4) The target region has obvious edge characteristics;
Figure 2. Figure a-d shows several representative images in the image library to be processed: (a) with few background interference terms and no occlusion of the target, it is an ideal state; (b) the existence of natural background trees and building background railings; (c) large shadow area caused by light; (d) The shooting angle causes the geometry of the target to deform.

In view of the above features, the traditional segmentation method based on threshold, edge and geometric features has a poor effect. GBVS significance model can simulate the visual significance of human eyes. Instead of manual interaction, the initialization step of Grab Cut algorithm is completed. Experiments show that the Grab Cut algorithm can improve the efficiency and ensure the segmentation accuracy.

3.2. Proposed Algorithm

Based on the above analysis, a non-interactive segmentation algorithm based on GBVS significance model is proposed to achieve the non-interactive object segmentation of this topic image. Because the target and background of the image to be processed are obviously different, and the advertising image has the characteristics of rectangular border, the significance model algorithm and appropriate threshold setting can be used to obtain better significance image effect. Therefore, the position of the rectangular box is automatically determined and the optimal segmentation is obtained by Grab Cut algorithm. The algorithm steps are as follows:

1) Input color image ‘a’ and use GBVS significance algorithm to calculate the significant value of color image pixel. On this basis, the target image ‘a1’ with the significant value in the top 45% was obtained, as shown in figure 3-a;

The GBVS algorithm is used to calculate the original image significance region as follow steps [13]:

i. An algorithm similar to Itti is used to extract the feature graph of input image I, and the feature graph is obtained by linear filtering and some basic nonlinear filtering M: [n]2 → R;

ii. The weight of edges between two nodes in the figure is defined as:

\[ w_1 ((i,j)||| (p,q)) \triangleq d((i,j)||| (p,q)) \cdot F(i-p,j-q) \cdot d((i,j)||| (p,q)) \triangleq \left| \log \frac{M(i,j)}{M(p,q)} \right| \]

Represents the difference between two pixel points;

\[ F(a, b) \triangleq \exp \left( -\frac{a^2 + b^2}{2\sigma^2} \right) \]

Type of \( \sigma \) is the free parameters in the algorithm;

iii. Define markov chains. The edge weights between two nodes are normalized between [0, 1]. This weight is defined as the transition probability from one node to another, nodes are defined as states and the equilibrium state of a markov chain represents the time spent at each point. The differences between
a node and its neighbors are greater, the residence time will be longer. Residence time can reflect the degree of attention of visual system, that is, the value of the equilibrium state of markov chain can be regarded as the significance value of each node.

2) After extracting edges by binarization of image ‘a1’, all connected regions were marked with rectangular boxes and the maximum connected regions were found, as shown in figure 3-b;

3) The rectangle box of the maximum connected region is marked as the background region TB outside the box and the unknown region TU inside the box, which is divided into TB, TU and TF, and GMM model is initialized accordingly.

4) Establish Gibbs energy function in Grab Cut algorithm. Gibbs energy function E includes data item U and smooth item V, then the energy function of the segmentation result converges by iteration (the energy function E doesn't change much anymore). So as to achieve optimal segmentation, an ideal a3 image with retrieval is obtained, as shown in figure 3-c;

5) After the above segmentation, the experiment shows that some images are over-segmented. In order to ensure the integrity of the image, the segmentation results are further improved: The extracted image is extracted from the edge of rectangle, the mask was used to extract the full segmentation result. To some extent, the influence of over-segmentation is avoided, as shown in figure 3-d;

4. Experimental Results And Evaluation
Based on the experiment of image library in this project, it can be seen that the proposed algorithm achieves non-interactive target segmentation and ensures the segmentation accuracy of Grab Cut algorithm. It also can effectively overcome the influence of complex background, light shadow, shooting Angle and other problems in the image on the segmentation accuracy of the target. The experimental comparison between automatic Grab Cut segmentation algorithm and manual initialization of Grab Cut segmentation in this paper is shown in figure 4.
Figure 4. Grab Cut interactive segmentation results and the non-interactive segmentation results that presented in this paper.

The segmentation performance of the algorithm in this paper is quantitatively calculated by using "segmentation accuracy". Segmentation precision is defined as the percentage of the accurate segmentation area in the GT (ground truth) image. The formula is as follows:

$$SA = \left(1 - \frac{|R_s - T_s|}{R_s}\right) \times 100\%$$

Where, $R_s$ represents the ideal segmentation area manually sketched, $T_s$ represents the real area of the image segmented by the algorithm in this paper, and $|R_s-T_s|$ represents the area wrongly segmented. According to the statistics of the segmentation results of 30 images in the image library, the average segmentation accuracy of the algorithm in this paper is 89.51%. Table 1. This table shows the segmentation accuracy of some randomly selected images.

| Image number | Ideal segmentation area (Number of pixels) | Real area of the image segmented | SA  |
|--------------|------------------------------------------|---------------------------------|-----|
| 1            | 810660                                    | 709632                          | 87.5% |
| 2            | 796950                                    | 703472                          | 88.3% |
| 3            | 739508                                    | 655944                          | 88.7% |
| 4            | 598800                                    | 466334                          | 77.9% |
| 5            | 808886                                    | 768690                          | 95.0% |
| 6            | 206057                                    | 202420                          | 98.2% |
| 7            | 804960                                    | 698544                          | 86.8% |

5. Conclusion
The algorithm proposed in this paper solves the problem of zero interaction and segmentation accuracy, but there are still the following problems: 1) The specific parameters in the algorithm are proposed for the image features of this project, which are not universally applicable; 2) The image color information is rich, the background is complex, and the calculation of significant value takes a long time, so real-time segmentation cannot be achieved. 3) Grab Cut has some drawbacks. If the difference between foreground and background pixels is not obvious, the segmentation effect is not good. The computation is heavy and the speed is slow. In the future research work, we will focus on solving the above three problems. It can be considered to process images into hyper pixel images and build network graphs with hyper pixels as nodes, so as to reduce the computation amount of GBVS significance algorithm and enhance real-time performance.

References
[1] G. Quellec, M. Lamard, G. Cazuguel. Wavelet optimization for content-based image retrieval in medical databases [J]. Medical Image Analysis, 2010 (4): 227 - 241.
[2] Zhang L, Zhang Y, Wei W, et al. An associative saliency segmentation method for infrared targets
[3] R. Achanta, S. Hemami, F. Estrada, & S. Süsstrunk. Frequency-tuned salient region detection. IEEE International Conference on Computer Vision and Pattern Recognition, 2009, pp. 1597 - 1604.

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

[5] R. Achanta, F. Estrada, P. Wils, & S. Süsstrunk. Salient region detection and segmentation. International Conference on Computer Vision Systems, 2008, pp. 66 - 75.

[6] X. Hou & L. Zhang. Saliency Detection: A spectral residual approach. IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1 - 8.

[7] J. Harel, C. Koch, & P. Perona. Graph-based visual saliency. Advances in Neural Information Processing Systems, 19: 545 - 552, 2007.

[8] Boykov Y Y, Jolly M P. Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in ND Images. Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on. IEEE, 2001, 1: 105 - 112.

[9] V. Gulshan, V. Lempitsky and A. Zisserman. Humanising GrabCut: Learning to segment humans using the Kinect. IEEE International Conference on Computer Vision Workshops, 1127 – 1133, 2011.

[10] M. Wang, L. Shen, and Y. Yuan. Automatic foreground extraction of clothing images based on grabcut in massive images. IEEE International Conference on Information Science and Technology, 238–242, 2012.

[11] A. Hernandez, M. Reyes, S. Escalera and P. Radeva. Spatio-temporal grabcut human segmentation for face and pose recovery. IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, 33–40, 2010.

[12] Rother C, Kolmogorov V, Blake A. Grabcut: Interactive Foreground Extraction Using Iterated Graph Cuts. ACM Transactions on Graphics (TOG). ACM, 2004, 23 (3): 309 - 314.

[13] Harel J, Koch C, Perona P. Graph-Based Visual Saliency. [J]. Proc of Neural Information Processing Systems, 2006 19: 545 - 55.