

# Natural Scene Interactive Segmentation Using Edge Detection and Maximal Similarity Region Merging

May Thu Win, Kay Thi Win  
University of Computer Studies, Mandalay  
maythu.ycc@gmail.com, kthiwin11@gmail.com

## Abstract

*A fundamental problem in image processing is image segmentation. The conventional image segmentation methods, to some extent, all suffer from the problem of inaccurate segmentation. A slightly easier and more approachable problem, interactive segmentation, has also received a lot of attentions over the years. In this paper, we propose a fast and simple interactive image segmentation technique. This segmentation process is conducted in two modules. First, the original image is detected by canny edge detection method. Second, the object of interest is segmented by using the region merging based on maximal similarity. In this work, color feature is used to measure the closeness between two regions and accordingly the label of the unmarked region is decided. The proposed method extracts the object from the complex background in the image. The effectiveness of the proposed method is validated by experimental results and compared with other method.*

**Keywords:** *Interactive Image Segmentation, Edge Detection, RGB color histogram, Region Merging*

## 1. Introduction

Image segmentation is an essential procedure in image analysis tasks such as image visualization, description, recognition, and object based compression. Many techniques have been proposed to achieve segmentation and can be classified into three groups: thresholding, edge-based, and region-based methods. Intensity thresholding technique which is the oldest segmentation method is still widely used in some applications since it is computationally inexpensive. An image in the thresholding method is assumed to be composed of several similar intensity objects on the background. However, the number of similar intensity classes is difficult to determine. On the other hand, edge-based methods are based on the assumption of that the abrupt change of intensity in an image corresponds to the boundary between objects. However, the edge detection procedure always

produced many false edges. Edge relaxation procedure is needed to obtain the true closed boundaries. Unfortunately, the edge procedure is complicated and very time consuming [13].

Interactive image segmentation involves minimal user interaction to incorporate user intention into the segmentation process and is an active research area in recent years because it can achieve satisfactory segmentation results that are unattainable by the state-of-the-art automatic image segmentation algorithms. This paper considers the same problem of how to interactively segment a foreground object out from its surrounding background. Our goal is to develop intuitive and intelligent image segmentation algorithms and tools that allow users to interactively guide the segmentation algorithm via a small amount of intuitive interactions until a satisfactory segmentation result that reflects both user intentions and photometric features is achieved [10].

The rest of this paper is organized as follows. Section II reviews the related works of the proposed method. Section III describes for the overview of the proposed method. Section IV discusses results and discussion. Section V compares the results with other method. Section VI concludes the paper.

## 2. Related Works

Interactive image segmentation is a popular field of research in computer vision. The aim of an interactive segmentation framework is to extract one or more foreground objects from the background. Applications for interactive image segmentation include photo editing and medical image analysis.

Most recent algorithms such as the graph cut based methods [11], [13], [2] consider both regional and boundary properties. With the graph cut based methods, an image is modeled as a graph where each node represents a pixel and two neighboring nodes are connected with a weighted edge defined as the distance between the pixel values. Moreover, the graph cut algorithm [11] models foreground and background pixel values according to histograms. Particularly, two

virtual nodes are added to the graph to denote the foreground and the background models. The max-flow/min-cut algorithm is employed to classify the nodes. To initialize the models, both foreground and background seeds are needed.

GrabCut [2] was introduced by Rother et al. in 2004. It is based on the works of Boykov et al. [11] by iteratively performing Graph Cuts. The user provides a rectangle around the desired object as an initial segmentation. This segmentation is used to approximate the appearance of the foreground and background respectively by Gaussian Mixture Models (GMMs). The negative log-likelihoods of the GMMs are assigned to each pixel and Graph Cut is used to update the segmentation. These steps are repeated iteratively.

Similarly, Scheuermann et al. [1] proposed an interactive segmentation scheme based on level-sets and Dempster's theory of evidence. Here, the user provides several strokes marking object and background regions. The histograms of these regions are included in a continuous energy function using Dempster's rule of combination. The defined energy function is minimized iteratively by solving the corresponding Euler-Lagrange equation [8].

Recently, the random walks algorithm and its extensions [6], [7], which also model an image as a graph, have been adopted for various image processing tasks. It has been demonstrated in [6] that the random walks algorithm can achieve better image segmentation performance than the graph cut algorithm. Similarly, the random walks algorithm requires the input of foreground and background seeds. However, the random walks algorithm lacks a global color distribution model; therefore it is very sensitive to the positions and quantities of foreground and background seeds. The random walks algorithm is essentially an approach that minimizes a Dirichlet energy with boundary conditions, where different boundary conditions (different input seeds) always result in different harmonic functions.

In this paper, the interactive region merging method based on the initial segmentation of Canny edge detection with a threshold is proposed. The main objective of the proposed method is to get the accurate result although the input image is the camouflage image. In the proposed scheme, the interactive information is introduced as markers, which are input by the users to roughly indicate the position and main features of the object and background. The markers can be the simple strokes (e.g. the green and blue lines). The proposed method will calculate the similarity of different regions and merge them based on the

proposed maximal similarity rule with the help of these markers. The object will then be extracted from the background when the merging process ends. And then, we analyze the effectiveness of the proposed method for comparing the different results between the proposed method and other method.

### 3. System Overview

In this paper, there are two major parts in the proposed system: the edge detection and the region merging for the interactive image segmentation.

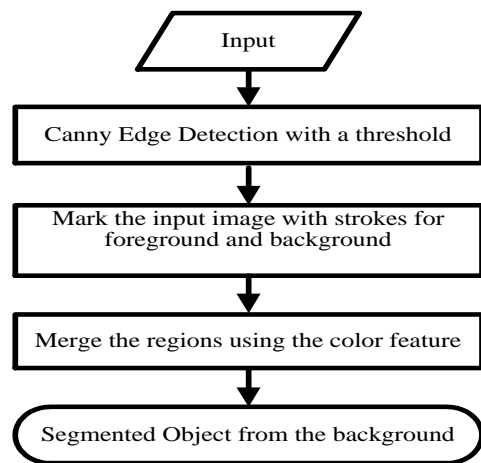


Figure 1. Flowchart of proposed system

There are many low level initial segmentation algorithms such as mean shift, watershed and level set. However, in our proposed system, firstly, the Canny edge detection with a threshold is used for detecting the edges of the input image. Therefore, there are many edge detection techniques in the state-of-the-art. Among them, we choose the Canny edge detection technique because it is the most powerful method and also preserves the edges of the input image well.

And then, the color histogram is used for representing the different regions. The object and background markers are marked by the user which object the user wants to segment. In this step, the user inputs are very important to get the accurate result for the interactive scheme. And then, the region merging based on the maximal similarity is used to extract the object of interest by using the input markers. Therefore, we get the object from the corresponding background by merging the different regions with the user input markers. Figure 1 shows the flowchart of the proposed system.

#### 3.1. Edge Detection

The Canny edge detector is an edge detection operator that uses a multi-stage algorithm to

detect a wide range of edges in images. It was developed by John F. Canny in 1986. Canny also produced a computational theory of edge detection explaining why the technique works. Canny's aim was to discover the optimal edge detection algorithm. In this situation, an "optimal" edge detector means:

1. Good Detection: There should be maximum probability of detecting real edge point and minimum of false edgepoints. This results in to the maximized signal to noise ratio.
2. Good Localization: There should be maximum possibility of detected edge closeness to the real edges.
3. Minimal Responses: One real image should result in to only one detected image.

### 3.1.1. Canny Edge Detection Algorithm

**Step 1:** Noise reduction by smoothing with Gaussian Filter.

It is well known that whatever we capture from a camera will contain some amount of noise, so there is a probability of mistaken noise as an edge, therefore noise is required to detect and removed. This is done by smoothing the image by using Gaussian Filter. Convolution of raw image  $I(i, j)$  is carried out with Gaussian filter  $G$ . Assuming the 2-D Gaussian function as:

$$G(i, j) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{i^2+j^2}{2\sigma^2}\right) \quad (1)$$

Where  $\sigma$  is indicates the standard variance of Gaussian function, and it determines the width of the Gaussian filter and outcome of smoothing.

Mathematically, the smooth resultant image is given by

$$F(i, j) = G * I(i, j) \quad (2)$$

**Step 2:** Finding gradients

In this step we detect the edges where the change in grayscale intensity is maximal. Required areas are determined with the help of gradient of images. Sobel operator is used to determine the gradient at each pixel of smoothed image. Sobel operators in  $i$  and  $j$  directions are given as

$$D_i = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

$$D_j = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

These sobel masks are convolved with smoothed image and giving gradients in  $i$  and  $j$  directions.

$$G_i = D_i * F(i, j) \text{ and } G_j = D_j * F(i, j) \quad (3)$$

Therefore, edge strength or magnitude of gradient of a pixel is given by

$$G = \sqrt{G_i^2 + G_j^2} \quad (4)$$

The direction of gradient is given by

$$\theta = \arctan\left(\frac{G_i}{G_j}\right) \quad (5)$$

$G_i$  and  $G_j$  are the gradients in the  $i$  and  $j$  directions respectively.

**Step 3:** Non maximum suppressions:

Non maximum suppression is carried out to preserves all local maxima in the gradient image, and deleting everything else this results in thin edges. For a pixel  $M(i, j)$ :

- Firstly round the gradient direction  $\theta$  nearest  $45^\circ$ , then compare the gradient magnitude of the pixels in positive and negative gradient directions i.e. If gradient direction is east then compare with gradient of the pixels in east and west directions say  $E(i, j)$  and  $W(i, j)$  respectively.
- If the edge strength of pixel  $M(i, j)$  is largest than that of  $E(i, j)$  and  $W(i, j)$ , then preserve the value of gradient and mark  $M(i, j)$  as edge pixel, if not then suppress or remove.

**Step 4:** Hysteresis thresholding:

The output of non-maxima suppression still contains the local maxima created by noise. Instead choosing a single threshold, for avoiding the problem of streaking two thresholds  $t_{high}$  and  $t_{low}$  are used. For a pixel  $M(i, j)$  having gradient magnitude  $G$  following conditions exists to detect pixel as edge:

- If  $G < t_{low}$  discard the edge.
- If  $G > t_{high}$  keep the edge.
- If  $t_{low} < G < t_{high}$  and any of its neighbors in a  $3 \times 3$  region around it have gradient magnitudes greater than  $t_{high}$ , keep the edge.
- If none of pixel  $(x, y)$ 's neighbors have high gradient magnitudes but at least one falls between  $t_{low}$  and  $t_{high}$  search the  $5 \times 5$  region to see if any of these pixels have a magnitude greater than  $t_{high}$ . If so, keep the edge.
- Else, discard the edge [9].

### 3.2. Maximal Similarity Region Merging

In this section, we use the region merging based on the maximal similarity to get the accurate result. Therefore, after finishing the edge detection, the users need to specify the object and background

conceptually. Similar to the users can input interactive information by drawing markers, which could be lines, curves and strokes on the image. The regions that have pixels inside the object markers are called object marker regions, while the regions that have pixels inside the background markers are called background marker regions. The green markers to mark the object while using blue markers to represent the background are used.

And then, it is necessary to represent these regions using some descriptor and define a rule for merging. A region can be described in many aspects, such as the color, edge, texture, shape and size of the region. Among them the color histogram is an effective descriptor to represent the object color feature. The RGB color space is used to compute the color histogram in this work. Each color channel into 16 levels and then the histogram of each region is calculated in the feature space of  $16 \times 16 \times 16 = 4096$  bins are uniformly quantized. Denote by  $\text{Hist}_R$  the normalized histogram of a region R. Therefore, it is necessary to define a similarity measure  $\rho(R, Q)$  between two regions R and Q to accommodate the comparison between various regions. Bhattacharyya coefficient is used to measure the similarity between R and Q.

$$\rho(R, Q) = \sum_{\mu}^{4096} \sqrt{\text{Hist}_R^{\mu} - \text{Hist}_Q^{\mu}} \quad (6)$$

where  $\text{Hist}_R$  and  $\text{Hist}_Q$  are the normalized histograms of R and Q, respectively, and the superscript  $\mu$  represents the  $\mu$ th element of them. Bhattacharyya coefficient  $\rho$  is a divergence-type measure which has a straightforward geometric interpretation.

Object and background markers provide some key features of object and background, respectively. The region merging method also starts from the initial marker regions and all the non-marker regions will be gradually labeled as either object region or background region. Let Q be an adjacent region of R and denote by  $S_Q = \{S_i^Q\} \ i = 1, 2, q$  the set of Q's adjacent regions. The similarity between Q and all its adjacent regions, i.e.  $\rho(Q, S_i^Q), \ i = 1, 2, \dots, q$ , are calculated. Obviously R is a member of  $S_Q$ . If the similarity between R and Q is the maximal one among all the similarities  $\rho(Q, S_i^Q)$ , R and Q will be merged. The following merging rule is defined: Merge R and Q

$$\rho(R, Q) = \max_{i=1,2,\dots,q} \rho(Q, S_i^Q) \quad (7)$$

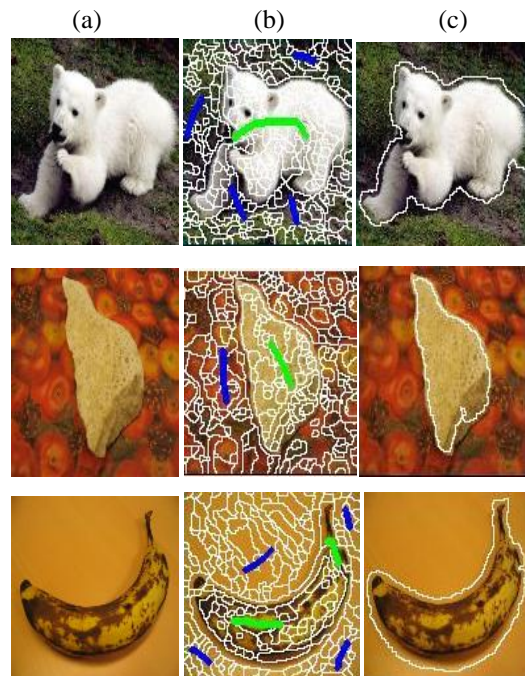
The merging rule (2) is very simple but it establishes the basis of the proposed region merging process. One important advantage of (7) is that it

avoids the presetting of similarity threshold for merging control [5].

The whole maximal similarity based region merging process can be divided into two stages, which are repeatedly executed until no new merging occurs. In the first stage, the non-marker regions are merged with background marker regions as many as possible. After the region merging of this stage, some non-marker background regions will be merged with the corresponding markers. Therefore, to complete the task of target object extraction, the second stage will focus on the non-marker regions to merge each other under the guidance of the maximal similarity rule. After merging all the background regions, it is equivalent to extracting the desired object.

## Results and Discussion

The objective of the proposed system is to achieve robust and accurate segmentation in different natural scene images. The Berkeley segmentation dataset [15] is used for the experiment of the proposed method. In our interactive image segmentation, the input images are the natural scene images. And then, the edge detection technique is used for detecting the different edge regions of the input image. Therefore, the canny edge detection technique with a suitable threshold is used for the initial segmentation.






**Figure 2.** (a) Original input image (b) The source images with user input strokes, where blue strokes are used to mark the background regions, and green strokes are used to mark the foreground regions (c) Result obtained by the proposed method

In our experiment, the threshold value is 0.01 with the canny edge detection. The initial edge detection is very valuable for segmenting the object from its background in the interactive scheme. And then, the input markers are marked by the user upon the result of the edge detection. The region merging based on the maximal similarity is used for segmenting the object of interest from the corresponding background. In figure 2, the satisfying result is achieved by providing the user inputs.

To evaluate the effectiveness of the proposed method, we manually labeled the desired objects in the test images and took them as ground truth. Then we computed the true positive rate (TPR) and false positive rate (FPR) for these segmentation results.

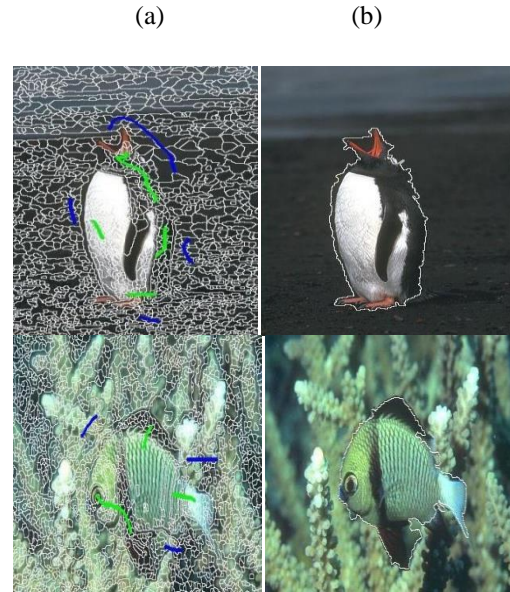
**Table 1. The TPR and FPR values on the test images**

Images	TPR (%)	FPR (%)
Bear1 	82.17	1.86
Stone2 	95.83	0.94
Banana1 	92.58	1.15

The TPR is defined as the ratio of the number of correctly classified object pixels to the number of total object pixels in the ground truth, and the FPR is defined as the ratio of the number of background pixels but classified as object pixels to the number of background pixels in the ground truth. Obviously, the higher the TPR is and the lower the FPR is, the better the method is. Table 1 lists the TPR and FPR results on the test images.

## 5. Comparison of Results

For evaluating the advantages of the proposed method, we use the maximal similarity based region merging based on the initial segmentation of mean shift [5] is used. The mean shift segmentation is good but it can't provide boundaries well if the input image is overlapped with its foreground and background colors. Therefore, we use the canny edge detection with a suitable threshold instead of mean shift because we want to get the appropriate initial boundary regions of the input image.



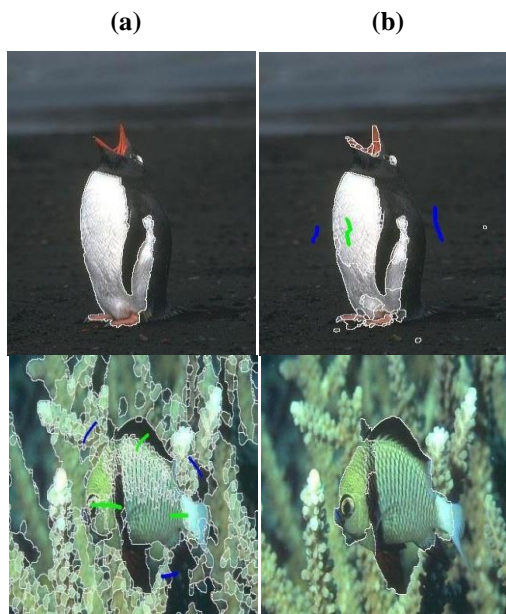
**Figure 3.(a)The source images with user input strokes, where blue strokes are used to mark the background regions, and green strokes are used to mark the foreground regions (b) Result obtained by the proposed method**

Figure 3 shows the result with our proposed method. The objective of using the canny edge detection in our proposed method is that to get the satisfying and accurate result for different natural scene images. Some natural scene input images are overlapped with its foreground color and background colors. Therefore, the satisfying result is achieved with our proposed method because the initial segmentation result is very important for high level process such as region merging. In our proposed method, the boundaries of the input image can provide well to get the accurate result from the corresponding background although the foreground and background colors of the input image are overlapped with each other.

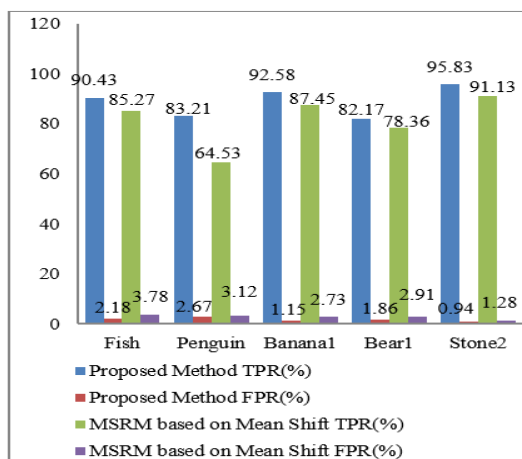
Whereas, figure 4 shows the result of the maximal similarity region merging based on mean shift. As we can see in this figure, the satisfying result is not achieved because the final segmentation result mainly depends on the initial segmentation result.

The initial segmentation regions are less than the regions of our edge detection. Therefore, the segmentation result of our method is better than the segmentation result of the maximal similarity region merging based on mean shift. Moreover, figure 5 shows the comparison of TPR and FPR between the proposed method and MSRM based on Mean Shift. Therefore, we can easily see in this bar chart, the proposed method is better than existing method with TPR and FPR values.





**Figure 4.(a)The source images with user input strokes, where blue strokes are used to mark the background regions, and green strokes are used to mark the foreground regions (b) Result obtained by MSRM based on mean shift**



**Figure 5.Comparison of TPR and FPR between proposed method and MSRM based mean shift**

## 6. Conclusion

In this paper, an interactive image segmentation method based on the edge detection is proposed. Since there are a lot of techniques like mean-shift, watershed, and level set for initial segmentation. In this case, the canny edge detection is used instead of initial segmentation algorithm. The user input strokes are very important for the foreground and background to get a desired object accurately. Color factor in feature is a critical human visual perception. In this paper, the

color feature is mainly used in the region merging process. Moreover, for the performance evaluation, the proposed method is compared with other method such as the maximal similarity region merging based on mean shift. So, it is very convenient for segmenting the object from the complex background in interactive scheme.

## References

- [1] B. Scheuermann and B. Rosenhahn, "Interactive image segmentation using level sets and dempster-shafer theory of evidence," SCIA, pages 656–665, 2011.
- [2] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: Interactive foreground extraction using iterated graph cuts," in ACM SIGGRAPH, 2004.
- [3] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002.
- [4] Haimin Ma, "Research on Interactive Segmentation Algorithm Based on Search Path Optimization," International Conference on Intelligent Human-Machine Systems and Cybernetics, 2009.
- [5] J. Ning, L. Zhang, D. Zhang, and C. Wu, "Interactive image segmentation by maximal similarity based region merging", Pattern Recognition, 2009.
- [6] L. Grady, "Random walks for image segmentation," IEEE Trans. PAMI, vol. 28, no. 11, pp. 1768–1783, Nov. 2006.
- [7] L. Grady, "Multi label random walker image segmentation using priormodels," in CVPR, 2005, pp. 763–770.
- [8] Oliver Jakob Arndt, Bjorn, and Bodo Rosenhahn, "RegionCut- Interactive Multi-Label Segmentation Utilizing Cellular Automaton," IEEE, 2012.
- [9] Rashmi, Mukesh, A.K.Jaiswal, and Rohini, "Performance Analysis of Adaptive Canny Edge Detector Using Bilateral Filter," IJIRCC, 2013.
- [10] W. Yang, J. Cai, J. Zheng, and J. Luo, "User-friendly interactive image segmentation through unified combinatorial user inputs," IEEE Transactions on Image Processing, 2010.
- [11] Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images," in ICCV, 2001.
- [12] Yu-Len Huang, Xun-Yao Lin, "Texture-Based Region Merging for Watershed Segmentation in Sonography."
- [13] Y. Li, J. Sun, C.K. Tang, and H.Y. Shum, "Lazy snapping," in ACM Siggraph, 2004, pp. 303–308.
- [14] <http://www.eecs.berkeley.edu/Research/Projects/CS/vison/grouping/segbench/>