Efficient Interactive Segmentation using Modified Maximal Similarity Region Merging

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

Abstract

Interactive image segmentation has many applications in image processing, computer vision, computer graphics and medical image analysis. In medical applications, image segmentation is a fundamental process in most systems that support medical diagnosis, surgical planning and treatments. In many editing tasks, the aim is to separate a foreground object from its background. Therefore, we propose a fast and simple interactive image segmentation technique in this paper. The proposed method automatically merges the regions that are initially segmented by mean shift segmentation, and then effectively extracts the object contour by labeling all the non-marker regions as either background or object. Moreover, many experiments are tested and the results show that the proposed method is faster than the existing method. Therefore, the proposed method is effective and can quickly and accurately segment for both medical and natural scene images with ease.

Keywords: Interactive Image Segmentation, Initial Segmentation, RGB color histogram, Region Merging

1. Introduction

Interactive image segmentation techniques are of practical use for various applications including image analysis, image composition, key extraction, etc. Compared to fully automated image segmentation, interactive segmentation exploits user’s knowledge on the target object for tracing its boundary.

Segmentation in medical imaging is an intensively studied field and many algorithms have been developed during the past decades, including automatic, semi-automatic or interactive methods. In medical imaging, automatic segmentation is a challenging task and it is still an unsolved problem for many medical applications due to the wide variety of image modalities, scanning parameters and biological variability. Manual segmentation is time-consuming and frequently not applicable in clinical routine. Therefore, semi-automatic segmentation methods, i.e., methods which require user interaction, can be used in cases where automatic algorithms fail.

Medical image data such as CT/MRI images can be analyzed and visualized as two-dimensional and three-dimensional data using computers, minimizing the need of invasive surgery. It is now possible to store, process, and visualize vast amounts of medical information. Medical images provide tools for diagnosis, treatment, and education of pathologies and create opportunities that can benefit health care.

Due to the increasing proliferation of digital photos and photo management/editing tools such as Picasa and Adobe Photoshop CS5, image editing is now commonly used in desktop application. Interactive segmentation is a fundamental part of such image editing tools. Extracting the object of interest from the non-trivial background is a crucial step in many interactive multimedia applications. So, these applications motivate the development of effective method for interactive image segmentation.

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 plays an important role in the segmentation of medical and natural scene images.

In [1], a general technique for the recovery of image features is presented. In this paper, mean shift is used to cluster and segment images for color image segmentation. The advantage of mean shift clustering is that the user doesn't have to specify the number of clusters.

S. Steger and G. Sakas [10] presented a method for fast interactive segmentation of tumors from medical images. In this paper, the user needs to click in the center of the tumor and a belief propagation based iterative adaption process is initiated, thereby considering image gradients as well as local smoothness priors of the surface. The approach has successfully been applied to the segmentation of liver tumors in CT datasets. Further trials on oropharynx tumors, liver tumors and the prostate from MR images as well as lymph nodes and the bladder from CT volumes demonstrated the generality of the proposed method.

H. He et al. [3] proposed the combination of the Live-Wire method with the region growing algorithm based on fuzzy affinity. First, anisotropic diffusion filter is employed to process the images which smoothed the images while keeping the edge, and then the possible boundary is confined in applying the Live-Wire method to the over-segmentation found by the region growing algorithm. The speed and the reliability of the segmentation of the Live-Wire method are greatly improved by such combination. This method has been used for CT and MR image segmentation. The results confirmed that their method is practical and accurate in the medical image segmentation.

Guoheng and Chi-Man [4] presented an interactive image segmentation method based on initial segmentation and region merging that the users only need to roughly mark the regions of the object and background by using markers. First, the image is initially segmented by simple linear iterative clustering (SLIC) segmentation and the users only need to provide the number of super pixels and mark the object by using some markers. Second, a region feature based on color histogram and Contourlet transform is computed for each region. Third, merge the regions which are similar to marked object region and merge the regions similar to marked background region, until there are only two kinds of regions such as object and background.

In this paper, the interactive region merging method is proposed to get the satisfying result for medical and natural scene images. The proposed method will extract the object from the corresponding 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 the existing method.

3. System Overview

There are two major parts in the proposed system: the initial segmentation and the modified maximal similarity region merging for the interactive image segmentation.

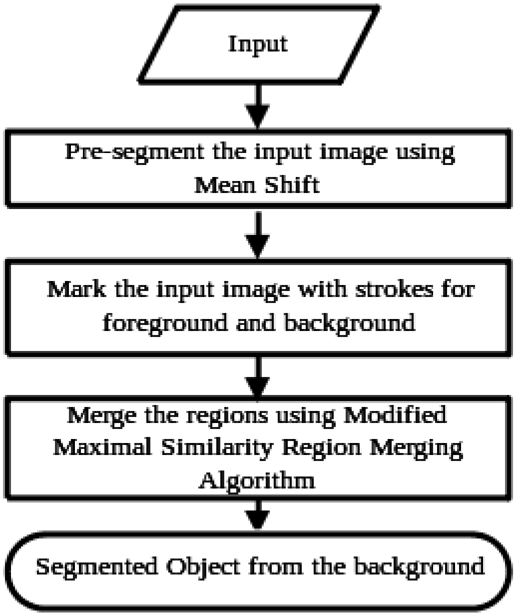

Figure 1. Flowchart of the Proposed System

There are many low level initial segmentation algorithms such as mean shift, watershed and level set. Any existing low level segmentation methods, such as super-pixel [9], mean shift [6, 12], watershed [8] and level set [5], can be used for this step. In this system, we choose to use the mean shift method for initial segmentation because it has less over segmentation and can well preserve the object boundaries. Particularly, we use the mean shift segmentation software, the EDISON System [2], to obtain the initial segmentation map. And then, the color histogram is used for representing the different regions. The object and background markers are marked by the user which object to be segmented. In this step, the user inputs are very important to get the accurate result for the interactive scheme. And then, the modified maximal similarity region merging algorithm 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. Initial Segmentation using Mean Shift

An initial segmentation is required to partition the image into homogeneous regions for merging. The mean-shift segmentation for initial segmentation has been chosen. Mean shift is a non-parametric feature-space analysis technique for locating the maxima of a density function, a so-called mode seeking algorithm. Application domains include cluster analysis in computer vision and image processing.

Mean-shift algorithm is a powerful clustering procedure that estimates the gradient of a probability density function using a generalized kernel approach. It has been successfully used for image segmentation in [1]. Being given a set of n points, $x_1, \ldots, x_n$, in the d-dimensional Euclidean space, the kernel density estimate is defined as:

$$\hat{f}(x) = \frac{i}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \quad (1)$$

Where h is the window radius (bandwidth parameter) of the used kernel K(x).

The estimate of the density gradient is defined as the gradient of the kernel density estimate:

$$\hat{\nabla} \hat{f}(x) = \frac{i}{nh^d} \sum_{i=1}^{n} \nabla K\left(\frac{x - x_i}{h}\right) \quad (2)$$

The kernel K(x) is a function of $\|x\|^2$: $K = c_{k,d}\, k\left(\|x\|^2\right)$. K(x) is called the profile of K(x) and ck, d is a normalization constant, which makes K(x) integrate to one. This class of kernels are called radially symmetric kernels.

The density estimator can be rewritten as:

$$\hat{f}_{h,K(x)} = \frac{c_{k,d}}{n h^{d}} \sum_{i=1}^{n} k\left(\left\|\frac{x-x_i}{h}\right\|^2\right) \quad (3)$$

Two commonly used kernels are multivariate Gaussian kernel:

$$K_{G(x)} = 2\pi^{\frac{-d}{2e}-\frac{1}{2}\|x\|^2} \quad (4)$$

And the Epanechnikov kernel:

$$K_{E(x)} = \begin{cases} \frac{1}{2} c_d^{-1}(d+2)\left(1-\|x\|^2\right), 0 \le \|x\| \le 1 \\ 0, \|x\| > 1 \end{cases} \quad (5)$$

The density gradient estimator of $f_{h,K(x)}$ is obtained as:

$$\hat{\nabla} f_{h,K(x)} = \nabla \hat{f}_{h,K(x)} = \frac{2c_{k,d}}{n h^{d+2}} \sum_{i=1}^{n} (x-x_i)\, \acute{k}\left(\left\|\right.\right.$$

, (6)

We denote: $g(x) = -\acute{k}(x)$. Using g(x) for profile, the kernel G(x) is defined as:

$$G(x) = c_{k,g} . g . \|x\|^2 \quad ,$$

Where $c_{k,g}$ is a positive constant (the normalization coefficient). The kernel K(x) is called the shadow of G(x). The estimate of the density gradient becomes:

$$\hat{\nabla} f_{h,K(x)} = \frac{2c_{k,d}}{n h^{d+2}} \sum_{i=1}^{n} g\left\|\frac{x-x_i}{h}\right\|^2 \frac{\sum_{i=1}^{n} g\left(\left\|\frac{x-}{h}\right.\right.}{\sum_{i=1}^{n} g\left(\left\|\frac{x-}{h}\right.\right.}$$

The density estimator computed with the kernel G(x) can be written as:

$$\hat{f}_{h,K(x)} = \frac{c_{k,d}}{n h^{d}} \sum_{i=1}^{n} g\left(\left\|\frac{x-x_i}{h}\right\|\right)$$

The mean shift vector (or sample mean shift) is defined as the difference between the weighted mean using kernel G(x) and x, as the center of the kernel:

$$m(x) = \frac{\sum_{i=1}^{n} g\left(\left\|\frac{x-x_i}{h}\right\|\right) x_i}{\sum_{i=1}^{n} g\left(\left\|\frac{x-x_i}{h}\right\|\right)} - x$$

The relation captured in (4.10) is intuitive, the local mean is shifted toward the region in which the majority of the points reside. Since the mean shift vector is aligned with the local gradient estimate it can define a path leading to a stationary point of the estimated density. The modes of the density are such stationary points. The mean shift procedure, obtained by successive

- computation of the mean shift vector $m(x)$
- translation of the kernel (window) by G(x)

is guaranteed to converge at a nearby point where the estimate (8) has zero gradient.

The mean shift procedure based image segmentation is a straightforward extension of the discontinuity preserving smoothing algorithm. Each pixel is associated with a significant mode of the joint domain density located in its neighborhood, after nearby modes were pruned as in the generic feature space analysis technique [1].

3.2. Modified Maximal Similarity Region Merging Algorithm

The modified maximal similarity region merging process can be divided into two stages, which are repeatedly executed until no new merging occurs. Our strategy is to merge object regions as much as possible while keeping object regions from being merged. Once we merge all the object regions, it is equivalent to extract the desired object. Some two-step strategies have been used for image pyramid construction. The proposed strategy aims for image segmentation and it is guided by the marker's input by users.

The objective of the proposed method is not only to get the accurate result but also to be faster than existing method. Therefore, there are three markers: object, background and non-marker in the proposed region merging based on the user input markers. The user supports the object and background markers for the interactive image segmentation. Therefore, it can’t be possible the non-marker regions are firstly merged because our segmentation method mainly depends on the object and background markers. On the one hand, we need to consider the object and background markers that are provided by the user.

In maximal similarity region merging (MSRM) [7], the background marker and non-marker regions are merged in the first stage. Therefore, the running time of MSRM is very expensive. To overcome this, the object markers are mainly considered in the first stage of the proposed method. Therefore, the object and non-marker regions are merged in the first stage to get the desired object as much as quickly.

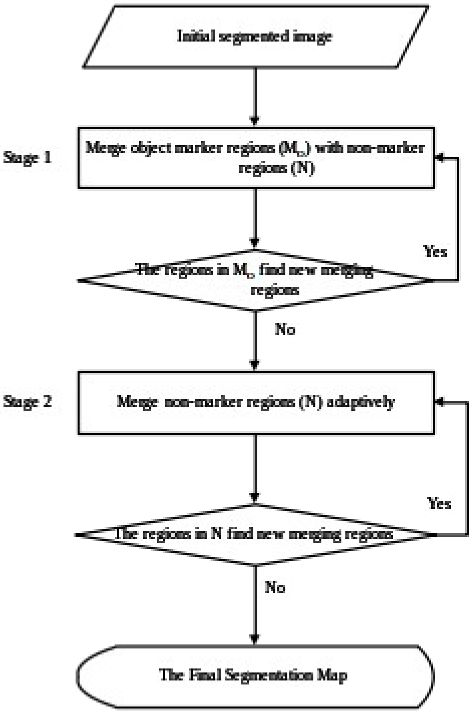

Figure 2. Flowchart of Modified MSRM

Figure 2 shows the flowchart of the modified maximal similarity region merging algorithm. 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 object marker regions as many as possible. Therefore, we try to merge object marker regions with their adjacent regions in the first stage. For each object region $O \in M_O$, the set of its adjacent regions $\acute{S}_O = \{A_i\}_{i=1,2,\ldots,r}$ was formed. each A_i and $A_i \notin M_O$, its set of adjacent regions $S_{A_i} = \{\acute{S}_j^{A_i}\}_{j=1,2,\ldots\ldots,k}$ was formed.

It is obvious that $O \in \acute{S}_{A_i}$. The similarity between A_i and each element in S_{A_i}, i.e. $\rho(A_i, S_{A_i})$, is calculated. If O and A_i satisfy the rule,

$$\rho\left(A_i,O\right)=max_{j=1,2,\ldots\ldots..k}\,\rho\left(A_i,S_j^{A_i}\right)$$
(11)

then O and A_i are merged into one region and the new region will have the same label as region O:

$O = O \cup A_i$

Otherwise, O and A_i will not merge.

The above procedure is iteratively implemented. The sets M_O and N will be updated in each iteration. Specifically, M_O expands and N shrinks. The iteration stops when the entire marker object regions M_O will not find new merging regions. After the region merging of this stage, some non-marker background regions will be merged with the corresponding markers.

After the first stage, for each non-marker (background or object) region P ∈ N, we form the set of its adjacent regions $\acute{S}_P=\left\{H_i\right\}_{i=1,2,\ldots..p}$. Then for each H_i that $H_i \notin M_B$ and $H_i \notin M_O$, we form its set of adjacent regions $\acute{S}_{H_i}=\left\{S_j^{H_i}\right\}_{i=1,2,\ldots..k}$. There is P $\in \acute{S}_{H_i}$. The similarity between Hi and each element in $\acute{S}_{H_i}$, i.e. $\rho\left(H_i,S_j^{H_i}\right)$, is calculated. If P and H_i satisfy the rule (2), i.e.

$$\rho\left(P,H_i\right)=\max_{j=1,2,\ldots\ldots,k}\rho\left(H_i,S_j^{H_i}\right)$$

then P and H_i are merged into one region

$P = P \cup H_i$

Otherwise, P and H_i will not merge.

To complete the task of the 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.

4. Results and Discussion

The objective of the proposed system is to achieve fast, robust and accurate segmentation for medical images and different natural scene images. We tested 100 medical images for interactive image segmentation. In our system, the brain, liver and lung images are used to segment the tumors of these images from the corresponding background for the medical image analysis.

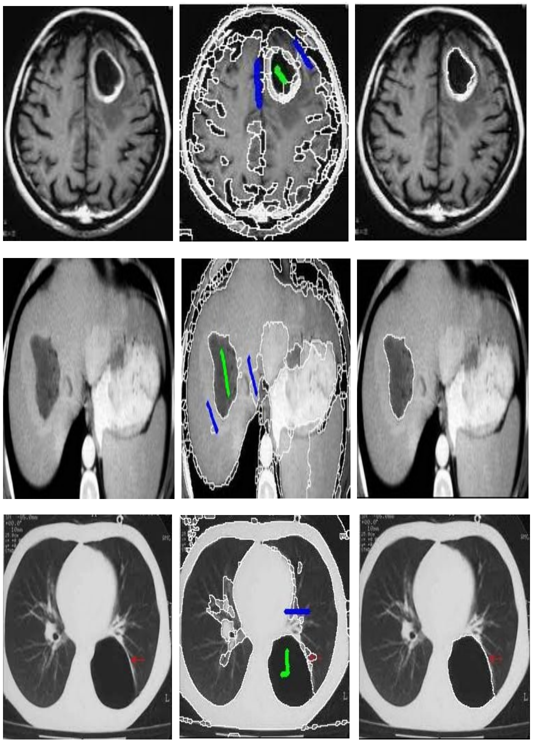

Figure 3. (a) Original Input Image (b) The Source Images with User Input Stokes (c) Result obtained by the Proposed Method of Medical Images

The Berkeley segmentation dataset [13] and Grab cut [10] dataset are used for the experiments of the proposed method. We tested all 100 images, which are natural images. The initial segmentation is very beneficial for segmentation the object from its background in the interaction scheme. Figure 3 shows the result of the proposed method for medical images. Similarly, Figure 4 also shows the result of the proposed method for natural scene images.

Figure 4. (a) Original Input Image (b) The Source Images with User Input Stokes (c) Result obtained by the Proposed Method of Natural Scene Images

5. Comparison of Results

For evaluating the advantages of the proposed method, we used the existing MSRM to compare the results of both the proposed method and the existing method. Therefore, Figure 5 and Figure 6 show the comparison of running time between the proposed method and the existing method for both medical and natural scene images. In these figures, we can easily see that the proposed method is preferable than the existing method via the running time.

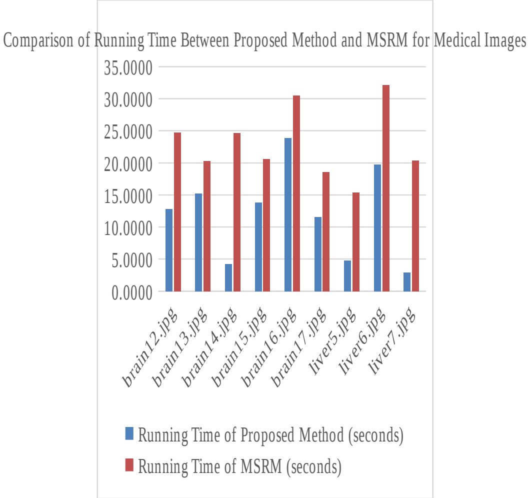

Figure 5. Comparison of Running Time for Medical Images

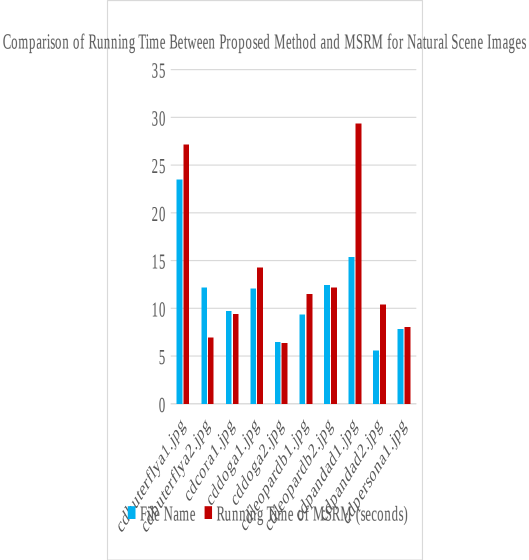

Figure 6. Comparison of Running Time for Natural Scene Images

Table 1. The TPR and FPR Values of Medical Images

Name of Images	**TPR by Proposed**	**FPR by Propo**	**TPR by Existing**	**FPR by Existi**

	Method (%)	sed Metho d (%)	Method (%)	ng Metho d (%)
brain1	71.4495	0.0086	71.4495	0.0086
brain2	76.0000	0.0275	76.0000	0.0275
brain3	83.7743	0.0482	83.7743	0.0482
liver29	92.9032	0.1007	92.9032	0.1007
lung10	94.5338	0.2268	94.5338	0.2268

Table 2. The TPR and FPR Values of Natural Scene Images

ID of Images	TPR by Proposed Method (%)	FPR by Proposed Method (%)	TPR by Existing Method (%)	FPR by Existing Method (%)
21077	92.0111	0.9714	75.5008	0.6424
208001	98.0208	0.2660	97.9451	0.1447
209070	94.3748	1.5614	94.2118	0.3107
376043	97.8263	1.1049	95.3860	0.3376
304074	92.6229	1.0209	91.6378	0.2170

Table 1 and Table 2 show the comparison of accuracy between the proposed method and the existing method for all input images in terms of True Positive Rate (TPR) and False Positive Rate (FPR). 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.

6. Conclusion

This proposed scheme is a region merging for interactive image segmentation method. The image is initially segmented by mean shift segmentation. Since the object regions will have high similarity to the marked object regions and so do the background regions, a maximal similarity based region merging mechanism was proposed to extract the object. According to our experiments, the running time of the proposed method is faster than the running time of the existing method for both medical and natural scene images. On the one hand, the running time of the proposed algorithm is much less than the existing algorithm, while the segmentation quality remains stable.

References

[1] D. Comaniciu, P. Meer, "Mean shift: a robust approach toward feature space Analysis", IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (5), pp. 603–619, 2002.

[2] Georgescu, Bogdan, and C. M. Christoudias, "Edge detection and image segmentation (EDISON) system", Robust Image Understanding Laboratory, Rutgers University, http://www.rutgers.edu/riul/research/ code/EDISON/index.html , 2003.

[3] H. He, J. Tian, Y. Lin, and K. Lu, "A New Interactive Segmentation Scheme Based on Fuzzy Affinity and Live-Wire", Springer-Verlag Berlin Heidelberg, pp. 436-443, 2005.

[4] G. Huang and C-M Pun, "Interactive Segmentation based on Initial Segmentation and Region Merging", in IEEE, 10th International Conference Computer Graphics, Imaging and Visualization, 2013.

[5] N. Li, M. Liu and Y. Li, "Image Segmentation Algorithm Using Watershed Transform and Level Set Method", In Acoustics, Speech and Signal Processing, ICASSP, IEEE International Conference on, vol. 1, pp. I-613, 2007.

[6] Q. Luo, T. M. Khoshgoftaar, "Efficient image segmentation by mean shift clustering and MDL-guided region merging", in IEEE Proceedings of the International Conference on Tools with Artificial Intelligence, pp. 337–343, November, 2004.

[7] J. Ning et al., “Interactive image segmentation by maximal similarity based region merging”, Pattern Recognition 43, no. 2, pp. 445-456, 2010.

[8] B. Peng, L. Zhang, and D. Zhang, “Automatic Image Segmentation by Dynamic Region Merging”, Image Processing, IEEE Transactions on 20, no. 12: 3592-3605, 2011.

[9] X. Ren, J. Malik, “Learning a classification model for segmentation”, In Computer Vision, Proceedings, Ninth IEEE International Conference on, vol.1, pp. 10-17, 2003.

[10] Georgescu, Bogdan, and C.M. Christoudias, “Edge detection and image segmentation (EDISON) system”, Robust Image Understanding Laboratory, Rutgers University,http://www.rutgers.edu/riul/research/code/EDISON/index.html , 2003.

[11] S. Steger and G. Sakas, “FIST: Fast Interactive Segmentation of Tumors”, Springer-Verlag Berlin Heidelberg, pp. 125–132, 2012.

[12] C. Rother, V. Kolmogorov, and A. Blake, “Grabcut: Interactive foreground extraction using iterated graph cuts,” in ACM SIGGRAPH , 2004.

[13]http://www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench/.