Automatic Image Segmentation Using Saliency Detection and Superpixel Graph Cuts

Sandeul Kang, Hansang Lee, Jiwhan Kim, and Junmo Kim*

Dept. of Electrical Engineering, Korea Advanced Institute of Science and Technology,
291 Daehak-ro, Yuseong-gu, Daejeon 305-701, South Korea
junmo@ee.kaist.ac.kr

Abstract. Image segmentation, which divides an image into foreground and background, is an important task for several applications in vision area such as object detection and classification. In this paper, we introduce a novel algorithm for automatic image segmentation technique which does not require further learning processes to perform segmentation. To achieve this automatic image segmentation, we incorporate saliency map for an image as an initial cue for image segmentation. An enhanced saliency detection method for generating saliency map is proposed. With over-segmented superpixels for an image and the generated saliency map, we perform image segmentation using graph cuts. To adapt graph cut segmentation to superpixel graph and saliency map, we suggest edge costs for superpixel graph based on Gaussian mixture models (GMM). As a result, superpixel graph enhances computational efficiency for our image segmentation technique and saliency map provides helpful cue for foreground region. We evaluate the performance of our algorithm on MSRA database demonstrate experimental results.

Keywords: automatic image segmentation, saliency detection, graph cuts, superpixels.

1 Introduction

Object recognition and classification are still challenging problems in computer vision. One of the major issues in these problems is how we extract meaningful information from an image, an array of pixels. Foreground/background segmentation, which extracts meaningful regions, called foreground, from its surroundings, called background, is thus a crucial task for not only object recognition and classification problems, but also several applications in computer vision including image retrieval and annotation. Foreground/background segmentation has been widely investigated with various types of data clustering techniques. Most of them are semi-automated, which relies on users' guidance such as bounding box [15, 18], scribble [4], and set of points [12]. These interactive techniques, however, have limitations such as sensitivity to quantity and quality of manual cues, also called seeds. On the other hand, one common approach for automatic foreground/background segmentation is to identify foreground region which has highest correlation with trained foreground

* Corresponding author.

J.-H. Kim et al. (Eds.): *Robot Intelligence Technol. & Appl. 2012*, AISC 208, pp. 1023–1034.
DOI: 10.1007/978-3-642-37374-9_99

Fig. 1. Experimental results of our algorithm: Original test images (top row) and their segmented foreground images (bottom row)

detector or mask. [5, 17] Underlying concept of this approach is that a set of general images, which contain objects of single category, shares similar appearances of foreground regions. Since this approach usually requires a large training database and its learning step, it is not sufficient for some cases such as detecting unexpected object from an image or detecting object without training database.

In this work, we propose a fully automatic foreground segmentation algorithm which requires only a given single image. To achieve automatic segmentation, we incorporate saliency map as an initial cue for foreground segmentation. Saliency detection, which extracts visually salient regions from an image and creates saliency map, is widely researched as a useful tool for detecting visual information from general images. Several works has proposed various methods for saliency detection, including center-surround differences [8, 9], spatiotemporal cues [20], graph-based visual saliency [7], pixel-wise differences [1, 2], visual context [6], and psychological attention model [13]. In this research, we propose a novel saliency detection method which enhances a saliency map by combining global and local saliency features. With the proposed saliency map, we use graph cuts [4, 18], one of the most popular and powerful segmentation and optimization techniques, for foreground segmentation. To apply our saliency map into graph cut segmentation properly, we suggest modified terminal and neighboring edge costs based on statistical measures using Gaussian mixture models (GMM), which is in a similar way to them of grab cut approach [18]. Additionally, to reduce the computational complexity while still obtaining reliable boundaries, we utilize over-segmentation [14], which partitions an image into set of boundary-preserving segments called superpixels, prior to segmentation step, and construct the graph with these superpixels for graph cuts, not with whole image pixels. As a result, Fig. 1 shows some test images and their corresponding results segmented by the proposed algorithm. Our main contributions in this paper are summarized as follows:

- We propose a fully automatic foreground segmentation method, which does not require any user interactions.
- Neither training database nor learning process is required for the proposed segmentation algorithm.
- We propose a novel saliency detection method at full-resolution
- We propose superpixel graph cuts with modified edge costs based on GMM.

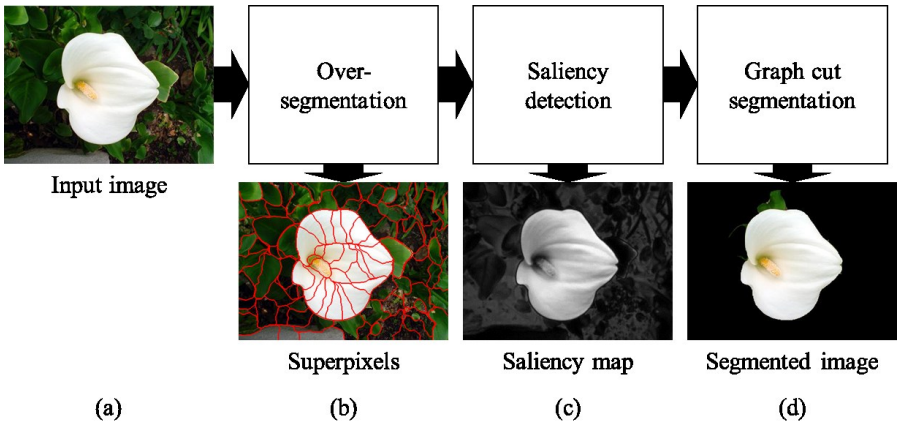

Fig. 2. Schematic overview of our segmentation algorithm: (a) An input image, (b) an over-segmentation step and its output superpixels, (c) a saliency detection step and its output saliency map, (d) a graph cut segmentation step and its output segmented foreground image

The outline of the paper is as follows: In section 2, we introduce the structure of our segmentation algorithm and its implementation. In section 3, we show experimental results of our segmentation algorithm on test images and evaluations of their performances. In section 4, we discuss our work and conclude the paper.

2 Algorithm

Fig. 2 shows the main structure of our segmentation method. First, we apply over-segmentation [14] on the given image to obtain superpixels. With the given image and obtained superpixels, we perform saliency detection to create the saliency map. In the saliency detection step, we find salient regions which are visually outstanding in two levels: local and global. With the saliency map created in this saliency detection, we create a tri-map for initial foreground and background GMMs using thresholding. In graph cut segmentation step, superpixel graph with modified edge costs is constructed and graph cuts [4] are performed on this graph to segment foreground and background from the given image. In the following subsections, we describe details of each step.

2.1 Over-segmentation

Over-segmentation is an algorithm which clusters pixels of an image into groups of pixels called superpixels which have similar properties such as color, brightness, and texture. Fig. 2 (b) shows an example of over-segmented image. As shown in the figure, over-segmented superpixels preserve boundaries inside an image and each superpixel is visually consistent. Furthermore, this over-segmentation step simplifies the computational complexity of later steps including graph cuts, by decreasing the number of graph nodes from the total number of pixels in an image to the number of superpixels, which can be determined by the user in the over-segmentation step. As

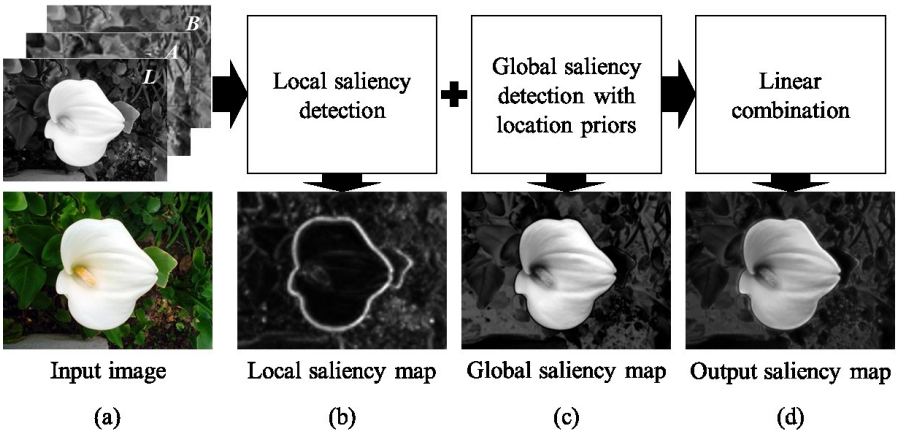

Fig. 3. Outline of saliency detection method: (a) An input image and its CIE Lab color components, (b) local saliency detection step and its output saliency map, (c) global saliency detection step with location prior maps and its output saliency map (d) output saliency map

shown in Fig. 2, both our saliency detection step and graph cut segmentation step use over-segmented superpixels for the given image. We use an over-segmentation method proposed by Mori et al. [14] which is based on the normalized cuts [16]. In our experiments, we choose the number of superpixels as $N_{SP} = 100$.

2.2 Saliency Detection

Saliency detection is to extract salient regions from images, which are visually outstanding and perceptually attractive. In photographic environment, foreground is represented as salient region since people usually focus cameras on the foreground. Under this assumption, we use saliency of the given image as an initial cue for foreground/background segmentation. For saliency detection, we construct three principles which are modified from principles suggested in Goferman et al. [6] based on psychological and biological evidence [11]:

P1. (*Local saliency*) Salient region has different feature properties, such as color, intensity, and texture, from its surroundings.

P2. (*Global saliency*) Salient region appears uniquely among the image, i.e. it does not occur concurrently.

P3. (*Location prior*) Salient region is usually located at the center of the image, or it includes the center of the image.

Based on these three principles, we propose a novel saliency detection method which is outlined in Fig. 3. As shown in the figure, our saliency detection method consists of three major steps: First, we extract locally salient region from an image according to the principle P1. Next, we find globally salient region from an image with the principle P2. In global saliency detection step, we apply the location prior map, which is determined in the principle P3, into global saliency map generation. We then create

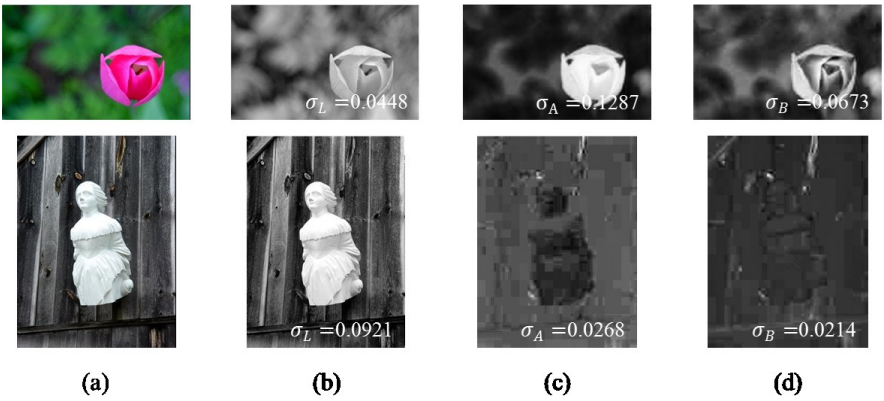

Fig. 4. Color channel selection for global saliency map: For each row, (a) input images, (b), (c), (d) visualized color channel components for L, a, b channels, respectively. Standard deviation values for each channel are written at the bottom of each image.

the saliency map by computing linear combination of local and global saliency maps we obtained. In our saliency detection method, we use CIELAB color space for the given image, which is known to represent color components proper for human visual perception.

According to the principle P1, people tend to give more attention to region which is locally distinctive from its neighbors. This local level approach has been accepted to several early saliency detection methods [7, 8, 9]. In this research, we find the locally salient region from the given image using center-surround differences which is originally proposed by Itti et al. [8, 9]. First, we make several scales for the input image in *L*, *a*, and *b* color channels. Then, we compute differences between a "center" fine scale $c \in \{1,2,3,4\}$ and a "surround" coarser scale $s = c + \delta$, $\delta \in \{3,4\}$ for each color channel as follows:

$$I(c,s) = \left| I(c) - I(s) \right| \tag{1}$$

where $I \in \{L,a,b\}$ is color channel for each scale and the across-scale difference between two images is obtained by expanding both images to the size of original scale image with neighbor-padding interpolation. We compute the local saliency map by combining center-surround differences for each scale pair and for each color channel:

$$S_L = \sum_{I \in \{L,a,b\}} \sum_{c=1}^{4} \sum_{s=c+3}^{4} \overline{N}(I(c,s)) \tag{2}$$

where $\overline{N}(I(c,s))$ is a normalized center-surround difference for each scale pair (c,s). Fig. 3 (b) shows an example of local saliency map extracted by our local saliency detection method.

From the principle P2, since the salient region usually occurs uniquely among entire image, we extract global saliency by measuring outlying degrees for superpixels of an image. First, we pick the superpixel sp_j and calculate the

differences between the image for each color channel $I \in \{L,a,b\}$ and $m(sp_j)$, the mean value of sp_j. We then compute the average of differences among all the superpixels, and add the difference between the image I and the mean value of entire pixels $m(I)$:

$$I_G = \frac{1}{N_{SP}}\left(\sum_{j=1}^{N_{SP}}\left(I - m\left(sp_j\right)\right)^2\right) + \left(I - m\left(I\right)\right)^2 \tag{3}$$

Before combining computed global saliency maps for each color channel to produce the final global saliency map, we select color channels to be combined among three color channels, *L*, *a*, and *b*. Fig. 4 shows some motivating examples for this channel selection process. If the given image has enough color information including both intensity and contrast, as shown at the top row in Fig. 4, only *a* and *b* channels provide a sufficient result without *L* channel, since *L* channel represents luminance, which usually interrupts color-based salient region. On the contrary, if the given image doesn't have enough color information as shown at the bottom row in Fig. 4, only *L* channel provide a sufficient result without *a* and *b* channels. Thus, we select channels for each image adaptively, based on comparing standard deviation values for each color channel which scale is normalized. For standard deviation values σ_L, σ_a, and σ_b for *L*, *a*, and *b* channels, respectively, the set of selected color channels $D \subseteq \{L,a,b\}$ is determined by following conditions:

$$D = \begin{cases} \{L\} & \text{where } \sigma_L \gg \sigma_a, \sigma_b \\ \{a,b\} & \text{where } \sigma_a, \sigma_b \gg \sigma_L \\ \{L,a,b\} & \text{otherwise} \end{cases} \tag{4}$$

where $\gg$ means "twice greater than." With the selected color channels, we then compute global saliency map by combining global saliency maps for selected channels:

$$S_G = \sum_{I \in D} I_G \tag{5}$$

According to the principle P3, people tend to give more attention at near the center of the image [11]. We thus create a location prior map and refine our global saliency map by applying this map to each color channel image. The location prior map is defined by a Gaussian distribution based on the distance between the pixel and the center of the image C:

$$P = \exp\left(-dist\left(p,C\right)/2\sigma_N{}^2\right) \tag{6}$$

where $dist(p,C)$ is the Euclidean distance between the pixel p and the center of the image C, and $\sigma_N{}^2$ is a constant. In our experiment, we choose σ_N as 2. We multiply this prior map with the global saliency computed for each color channel before combining them. Fig. 3 (c) shows an example of global saliency map extracted by our

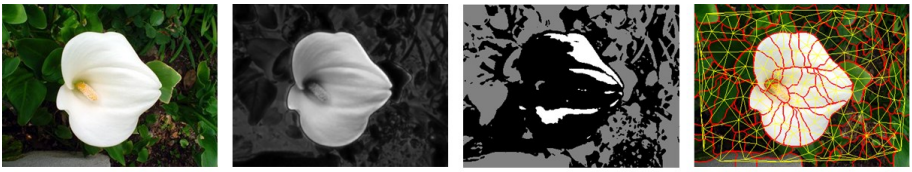

Fig. 5. Tri-map generation: (a) An input image, (b) our saliency map, (c) tri-map generated by thresholding, (d) Superpixel graph with yellow edges

global saliency detection method. With local saliency map and refined global saliency map, we finally obtain our saliency map by computing linear combination of them:

$$S = \alpha S_L + \beta S_G \tag{7}$$

where α, β are constants. In our experiments, we set α, β as 0.25 and 0.75, respectively. Fig. 3 (d) shows our final saliency map obtained from our saliency detection method.

2.3 Graph Cut Segmentation

We then perform foreground/background segmentation using graph cuts with the saliency map for the given image obtained in the previous subsection as initial cues. First, to apply the saliency map as prior information for graph cuts, we transform the saliency map into tri-map, which consists of tri-nary regions: foreground, background, and unknown regions. We define two threshold values Th_f and Th_b for foreground and background regions, respectively, and determine the tri-map by thresholding with these two thresholds as follow:

$$M = \begin{cases} \text{foreground} & \text{if } S > Th_f \\ \text{background} & \text{if } S < Th_b \\ \text{unknown} & \text{otherwise} \end{cases} \tag{8}$$

Fig. 5 shows the saliency map and its corresponding tri-map. In Fig. 5 (c), initial foreground, background, and unknown regions are colored white, gray, and black, respectively. In our experiments, we set Th_f and Th_b as 5% and 60%, respectively.

With the obtained tri-map, we extract initial cue information for foreground and background regions using Gaussian mixture models in a similar way to grab cut approach [18]. From these extracted Gaussian components, we define sets of mean values for each Gaussian as follow:

$$G_F = \{\mu_{F1}, \mu_{F2}, ..., \mu_{FK}\} \tag{9}$$

$$G_B = \{\mu_{B1}, \mu_{B2}, ..., \mu_{BK}\} \tag{10}$$

where G_F, G_B are sets of mean values for foreground and background regions, respectively, K is the number of Gaussian components. In our experiment, we

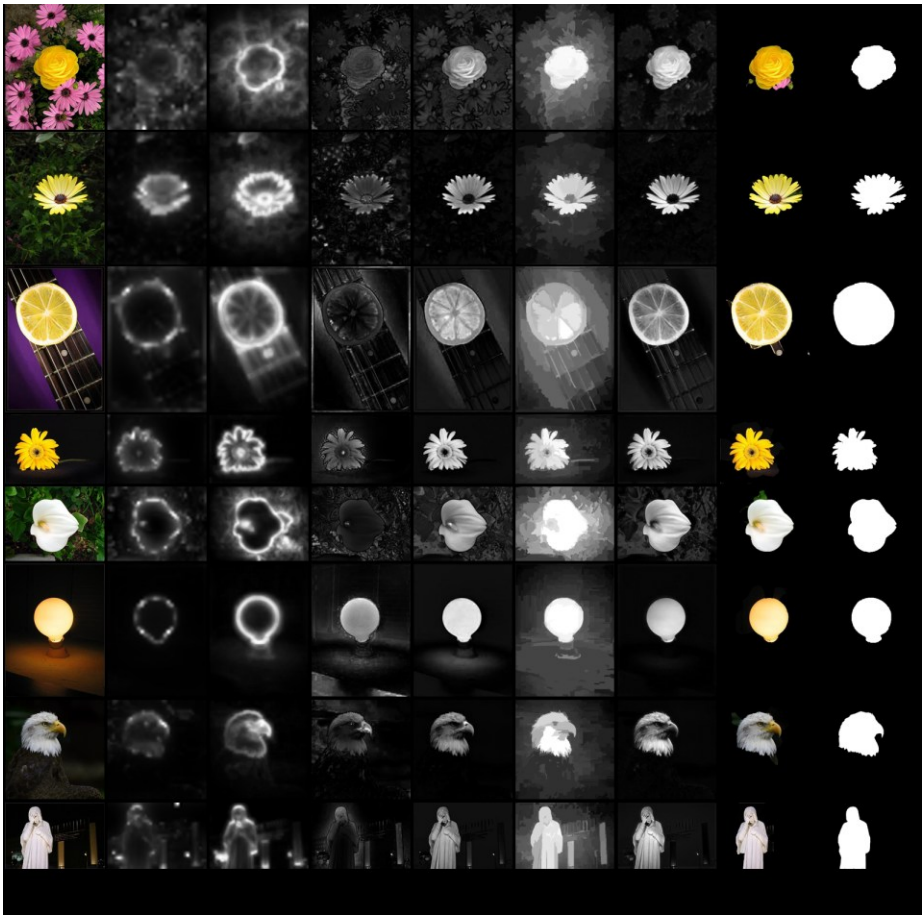

Fig. 6. Experimental results and their comparisons: (a) Input images, (b)~(g) saliency maps generated by (b) HR [7] , (c) CA [6], (d) AC [21], (e) IG [1], (f) RC [22], and (g) ours, (h) our segmentation results, (i) ground truth masks

choose the number of Gaussians as 5. Once initial GMMs are created, then we construct graph with over-segmented superpixels for graph cut segmentation. We use Delaunay triangulation to connect each superpixel to adjacent superpixels and for each superpixel sp_j, we compute mean color μ_j and use it as pixel color in pixel-wise graph cuts [4]. Fig. 5 (d) shows superpixel graph with edges connecting adjacent superpixels. For N-links which connect between adjacent superpixels, we compute modified edge costs defined as follows:

$$E_N(sp_j, sp_k) = \exp\left(-\left(\mu_j - \mu_k\right)^2 \Big/ 2\sigma_T^{\,2}\right) \tag{11}$$

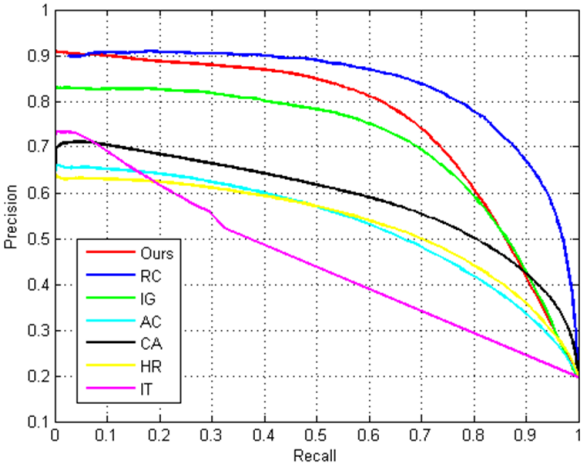

Fig. 7. Precision-recall curves for our saliency detection method with 6 state-of-the-art methods (RC [22], IG [1], AC [21], CA [6], HR [7], IT [9])

where σ_T is a constant. In our experiment, we choose σ_T as 1. For T-links which connect between superpixels and terminals representing foreground and background, we also compute modified edge costs defined as follows:

$$E_T(sp_j, F) = \min_{\mu_B \in G_B} \left(\mu_j - \mu_B\right)^2 \tag{12}$$

$$E_T(sp_j, B) = \min_{\mu_F \in G_F} \left(\mu_j - \mu_F\right)^2 \tag{13}$$

where *F*, *B* are terminals for foreground and background regions, respectively. With the constructed superpixel graph, we perform graph cut segmentation using min-cut max-flow algorithm [4] to obtain the final segmented foreground region. Fig. 2 (d) shows an example of segmentation results.

3 Experimental Results

We experiment our algorithm on the MSRA salient object database, which includes 1000 images selected by Achanta et al. [1] as test images with ground truth binary masks. On this dataset, we compare our saliency detection results with 6 state-of-the-art saliency detection methods including Cheng et al. (RC) [22], Achanta et al. (IG) [1], Achanta et al. (AC) [21], Goferman et al. (CA) [6], Harel et al. (HR) [7], and Itti et al. (IT) [9]. We also measure the performance of our segmentation results by comparing with the ground truth masks of test images. Our parameter settings on the experiment are presented in the previous sections.

We test our algorithm on an Intel® Core™ i5 with 2.67 GHz CPU with 8GB memory. The algorithm is implemented on the MATLAB platform. Except the over-segmentation step, our saliency map generation takes average 0.67 seconds per image. To compare saliency detection results, we test our method on every 1000 images provided in the given MSRA dataset. Fig. 6 shows selected test images and their corresponding saliency maps resulted from several saliency detection methods mentioned above and ours. With these results, we obtain the precision-recall rate curve by changing threshold values from 0 to 255 where the entire saliency maps are represented as grayscale images. We compare these thresholded binary masks of salient with ground truth masks and compute average precision and recall values over 1000 images. Fig. 7 shows precision-recall curves for our method with others. As shown in the figure, our saliency maps present high precision and recall as their own. We also measure precision and recall rates to evaluate the performance of our segmentation algorithm. With 1000 segmented masks on MSRA database, we compute overlapping ratios between our segmented masks and ground truth masks to calculate mean precision and recall rates. As a result, we obtain average values of precision and recall as 73.81% and 73.57%, respectively.

4 Conclusion

In this paper, we proposed a fully automatic foreground segmentation algorithm based on saliency detection and superpixel graph cuts. We introduced a novel saliency detection method which combines local and global saliency features to produce an accurate saliency map. Compared to state-of-the-art saliency detection algorithms, our saliency maps show fine performances as saliency detection itself. We further used our saliency map as an initial cue for the graph cut segmentation. In graph cuts, we utilized over-segmentation to construct the superpixel graph and computed modified edge costs on this superpixel graph. To evaluate our segmentation method quantitatively, we tested on 1000 images of MSRA database and compared with the ground truth and segmented regions resulted from state-of-the-art saliency detection algorithms. As a result, our saliency model shows proper performances for graph cut segmentation compared to other saliency detection algorithms. Moreover, our segmentation method achieves accurate segmentation performances without user inputs and learning processes, which makes our segmentation method a useful pre-processing tool for object detection and classification.

Acknowledgements. This research was supported by the MKE (The Ministry of Knowledge Economy), Korea, under the Human Resources Development Program for Convergence Robot Specialists support program supervised by the NIPA. (National IT Industry Promotion Agency) (NIPA-2012-H1502-12-1002).

References

1. Achanta, R., Hemami, S., Estrada, F., Süsstrunk, S.: Frequency-tuned salient region detection. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1597–1604 (2009)
2. Achanta, R., Süsstrunk, S.: Saliency detection using maximum symmetric surround. In: Proceedings of the 2010 17th IEEE International Conference on Image Processing, pp. 2653–2656 (2010)
3. Arbelaez, P., Maire, M., Fowlkes, C., Malik, J.: From contours to regions: An empirical evaluation. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 2294–2301 (2009)
4. Boykov, Y.Y., Jolly, M.-P.: Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. In: Proceedings of the 2001 8th IEEE International Conference on Computer Vision, vol. 1, pp. 105–112 (2001)
5. Brox, T., Bourdev, L., Maji, S., Malik, J.: Object segmentation by alignment of poselet activations to image contours. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, pp. 2225–2232 (2011)
6. Goferman, S., Zelnik-Manor, L., Tal, A.: Context-aware saliency detection. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition, pp. 2376–2383 (2010)
7. Harel, J., Koch, C., Perona, P.: Graph-based visual saliency. In: Proceedings of the Conference on Advances in Neural Information Processing Systems, vol. 19, pp. 545–552 (2006)
8. Itti, L., Braun, J., Lee, D.K., Koch, C.: Attentional modulation of human pattern discrimination psychophysics reproduced by a quantitative model. In: Proceedings of the 1998 Conference on Advances in Neural Information Processing Systems, vol. 2, pp. 789–795 (1998)
9. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(11), 1254–1259 (1998)
10. Jung, C., Kim, C.: A unified spectral-domain approach for saliency detection and its application to automatic object segmentation. IEEE Transactions on Image Processing 21(3), 1272–1283 (2012)
11. Judd, T., Ehinger, K., Durand, F., Torralba, A.: Learning to predict where humans look. In: Proceedings of the 2009 IEEE 12th International Conference on Computer Vision, pp. 2106–2113 (2009)
12. Kim, T.H., Lee, K.M., Lee, S.U.: Nonparametric higher-order learning for interactive segmentation. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition, pp. 3201–3208 (2010)
13. Klein, D.A., Frintrop, S.: Center-surround divergence of feature statistics for salient object detection. In: Proceedings of the 2011 IEEE International Conference on Computer Vision, pp. 2214–2219 (2011)
14. Mori, G.: Guiding model search using segmentation. In: Proceedings of the 2005 10th IEEE International Conference on Computer Vision, vol. 2, pp. 1417–1423 (2005)
15. Pham, V.-Q., Takahashi, K., Naemura, T.: Foreground-background segmentation using iterated distribution matching. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, pp. 2113–2120 (2011)
16. Ren, X., Malik, J.: Learning a classification model for segmentation. In: Proceedings of the 2003 9th IEEE International Conference on Computer Vision, pp. 10–17 (2003)

17. Rosenfeld, A., Weinshall, D.: Extracting foreground masks towards object recognition. In: Proceedings of the 2011 IEEE International Conference on Computer Vision, pp. 1371–1378 (2011)
18. Rother, C., Kolmogorov, V., Blake, A.: GrabCut: interactive foreground extraction using iterated graph cuts. ACM Transactions on Graphics 23(3), 309–314 (2004)
19. Sinop, A.K., Grady, L.: A seeded image segmentation framework unifying graph cuts and random walker which yields a new algorithm. In: Proceedings of the 2007 IEEE 11th International Conference on Computer Vision, pp. 1–8 (2007)
20. Zhai, Y., Shah, M.: Visual attention detection in video sequences using spatiotemporal cues. In: Proceedings of the 14th Annual ACM International Conference on Multimedia, pp. 815–824 (2006)
21. Achanta, R., Estrada, F., Wils, P., Süsstrunk, S.: Salient region detection and segmentation. In: Proceedings of the 6th International Conference on Computer Vision Systems, pp. 66–75 (2008)
22. Cheng, M.-M., Zhang, G.-X., Mitra, N.J., Huang, X., Hu, S.-M.: Global contrast based salient region detection. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, pp. 409–416 (2011)