Optical Engineering

SPIEDigitalLibrary.org/oe

Automatic foreground extraction by background elimination based on multiscale segmentation

Ki Tae Park
Chan Woo Park
Young Shik Moon

Optical Engineering 50(6), 067004 (June 2011)

Automatic foreground extraction by background elimination based on multiscale segmentation

Ki Tae Park
Hanyang University
Ambient Intelligence Software Team
Institute of Hanyang BK21
17 Haengdang-dong, Seongdong-gu
Seoul 133-791 Republic of Korea

Chan Woo Park
Young Shik Moon
Hanyang University
Department of Computer Science and Engineering
1271 Sa-3-dong
Sangnok-gu, Ansan-si
Gyeonggi-do 426-791, Republic of Korea
E-mail: ysmoon@cse.hanyang.ac.kr

Abstract. We propose an approach to automatically extracting foreground regions. This is a novel method for segmenting salient objects from still images by background elimination. To extract foreground regions, a new method of background elimination based on multiscale segmentation is proposed to detect candidate object regions. To this end, we use a trimap consisting of foreground, background, and undefined regions and a region adjacency graph. A graph-cut technique is finally used to extract exact foreground regions from the candidates. Experimental results have shown that the proposed method yields a better foreground extraction than Kim's method under various environments containing multiple objects and clutter backgrounds in natural images. *© 2011 Society of Photo-Optical Instrumentation Engineers (SPIE).* [DOI: 10.1117/1.3582862]

Subject terms: foreground extraction; background elimination; multiscale segmentation; graph cuts.

Paper 100680RR received Aug. 23, 2010; revised manuscript received Mar. 31, 2011; accepted for publication Apr. 6, 2011; published online Jun. 9, 2011.

1 Introduction

Foreground segmentation is the task of identifying objects from backgrounds.[1–6] The automatic segmentation of foreground regions is an emerging research topic in various fields, such as content-based image retrieval, image compression, digital watermarking, and thumbnail cropping for small display screens in personal digital assistants (PDAs).[7] In order to effectively extract the foreground surrounded by backgrounds, it is necessary to mark background regions in still images. In recent years, there have been many research efforts on the background elimination in color images. Huang *et al*. proposed a method that segments color images into foreground and background regions using a minimum-description-length principle.[8] Lu and Guo proposed an object-based image retrieval method by eliminating backgrounds connected to image borders.[9] However, their methods are limited to images including smooth or monotonous textured backgrounds. Kim *et al*.[3] and Crevier[4] proposed a method for automatically extracting an object of interest from complex backgrounds. In order to extract correct object regions, they generated a salient probability map representing the probability of saliency for all pixels by manually marking the ground truth of object regions from test images. They showed that most object regions are close to the center of an image.

In this paper, we propose a method for extracting salient objects consisting of three steps. In the first step, three multiscale images by iteratively applying a bilateral filter[10,11] are generated and segmented by JSEG method.[12] In the second step, regions adjacent to the borders of segmented images (SIs) are eliminated as background regions and candidate foreground regions from multiscale images excluding the backgrounds are detected by using logical operations and a region adjacency graph (RAG).[13] In the final step, real foreground regions are extracted from the candidates by using a

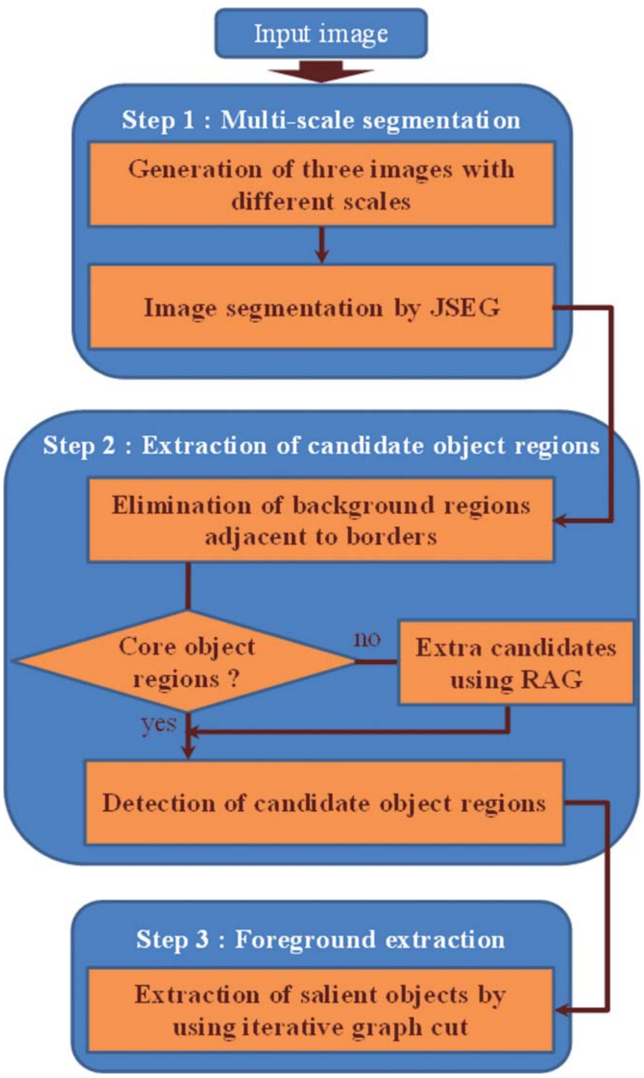


Fig. 1 Overall flowchart of the proposed method.

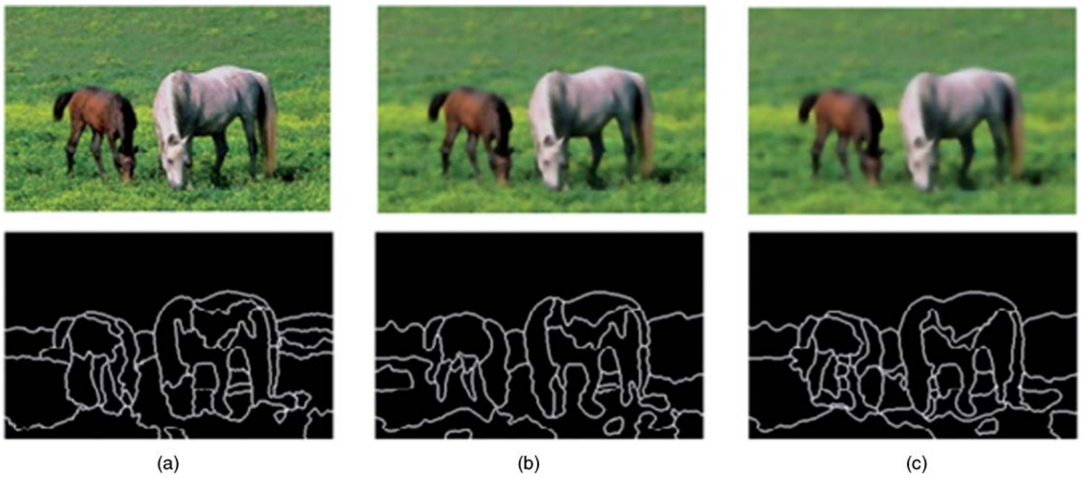


Fig. 2 Segmentation results of three multi-scale images: (a) I^1 image, (b) I^2 image, and (c) I^3 image.

graph-cut algorithm.[14] Figure 1 shows the overall flowchart of the proposed method.

2 Multiscale Image Segmentation by Bilateral Filter and JSEG Method

In order to extract foreground regions from an image, the JSEG method proposed by Deng and Manjunath[12] is first used for image segmentation. JSEG is performed on three multiscale images generated by iteratively applying a bilateral filter. The reason for using multiscale images for extracting foreground regions is that the JSEG method has a tendency of an oversegmentation caused by noises, colors, and textures. Because of the oversegmentation result, it is difficult to effectively extract foreground regions through eliminating background regions adjacent to the borders from the original image. Because a bilateral filter considering two Gaussian weights of color and spatial information between neighboring pixels smoothes a cluttered image while maintaining strong edges,[10,11] the oversegmentation can be effectively removed. Accordingly, by combining multiscale-based segmentation results of the three images, backgrounds are effectively eliminated. The bilateral filter is defined as

$$I^{n+1}(\mathbf{x}) = \int I^n(\mathbf{x}) \exp\left\{-\frac{(\mathbf{x}-\varepsilon)^2}{2\sigma_d^2}\right\} \exp\left\{-\frac{1}{2}[I^n(\mathbf{x}) - I^n(\varepsilon)]^T \times \sum\nolimits_r^{-1} [I^n(\mathbf{x}) - I^n(\varepsilon)]\right\} d\varepsilon, n \in \{1\ ,\ 2\}, \quad (1)$$

where I^n is an input image and I^{n+1} is the result image applying a bilateral filter. $I(\mathbf{x})$ is the color value at pixel $\mathbf{x}$. $\boldsymbol{\varepsilon}$ is a neighbor of pixel $\mathbf{x}$. The parameter $\boldsymbol{\sigma}_d$ is the standard deviation of spatial domain and Σ_r is the covariance matrix of color domain.

In this paper, three multiscale images I^1, I^2, and I^3 are used for extracting object regions. I^1 is the original image. I^2 and I^3 are the images resulting from one and two bilateral filtering steps, respectively. Segmentation results of multiscale images are shown in Fig. 2.

3 Background Elimination and Extraction of Candidate Object Regions

In order to extract foreground regions effectively, we first extract candidate object regions. To this end, three binary maps composed of candidate object regions and background regions are generated. Background regions in the binary maps are considered as regions adjacent to borders of each multiscale image segmented by JSEG method. If a binary map is used for extracting candidate object regions, then backgrounds that are composed of cluttered regions away from borders cannot be eliminated. Thus, by using three binary maps, a novel approach to detecting candidate foregrounds is proposed. To this end, we use two approaches of a logical operation and a RAG. In order to detect candidate foregrounds effectively, the logical operation on three binary maps is first performed for distinguishing candidate object regions and backgrounds, as defined in Eq. (2). Consequently, a trimap

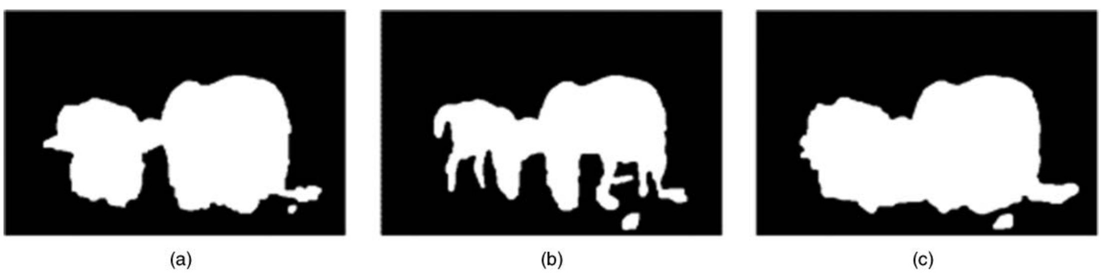


Fig. 3 Binary maps by distinguishing candidate object regions and backgrounds from three segmented multiscale images: (a) SI_1 map, (b) SI_2 map, and (c) SI_3 map.

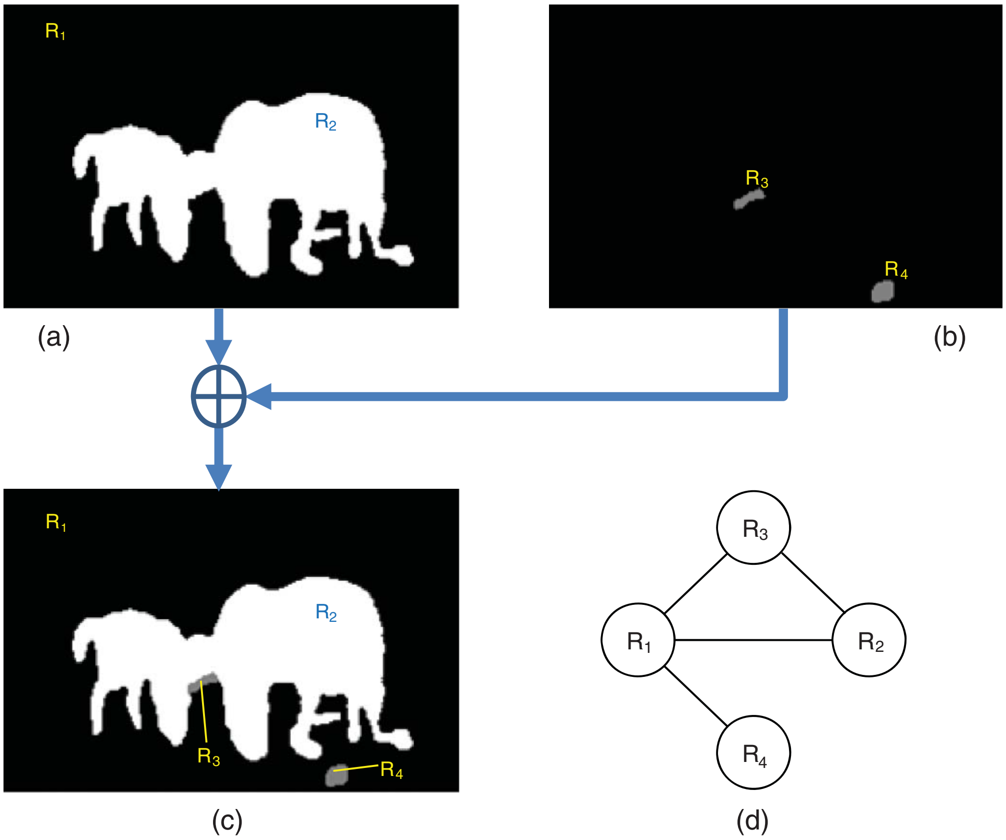


Fig. 4 Example of a RAG construction: (a) A labeled binary map composed of core candidate region and background region, (b) undefined regions with region labeling, (c) a trimap combined (a) and (b), and (d) a RAG corresponding to (c).

composed of three parts, such as candidate object regions, backgrounds, and undefined regions, is obtained by

$$\mathrm{Trimap}(x, y) = \begin{cases} \text{foreground} \quad \text{if } \mathrm{SI}_1(x, y) \\ \quad = 1\&\mathrm{SI}_2(x, y) = 1\&\mathrm{SI}_3(x, y) = 1, \\ \text{background} \quad \text{else if } (\mathrm{SI}_1(x, y) = 0\& \\ \quad \mathrm{SI}_2(x, y) = 0\&\mathrm{SI}_3(x, y) = 0) \\ |(\mathrm{SI}_1(x, y) \otimes \mathrm{SI}_2(x, y) \otimes \mathrm{SI}_3(x, y) = 1), \\ \text{undefined} \quad \text{otherwise.} \end{cases} \tag{2}$$

In Eq. (2), SI_1, SI_2, and SI_3 denote binary maps composed of candidate object regions and background regions. A pixel of binary map SI_k is set to zero if the pixel belongs to background regions adjacent to the borders of the corresponding segmented image and is equal to 1 otherwise. Figure 3 shows three binary maps obtained from three multiscale images. Operators of &, |, and ⊗ represent "AND," "OR," and "XOR" operation, respectively.

By combining three binary maps, a trimap is generated. If all pixels at same coordinate of three binary maps belong to candidate object regions in each binary map, then Trimap(*x*,*y*) is considered as core candidate region. Trimap(*x*,*y*) is regarded as background if at least two pixels of three binary maps belong to background regions in corresponding binary maps. Otherwise, it is considered as undefined region. However, as shown in Fig. 3, because core candidate regions cannot be effectively extracted due to over- and undersegmentation caused by the JSEG method, it is difficult to extract exact foregrounds by using only the core candidate regions in a trimap. Therefore, in order to detect final candidate foregrounds effectively, we perform an approach to determining whether undefined regions are included to extra candidates or backgrounds. Undefined regions are distinguished into three kinds as shown in

$$\mathrm{Undef\_}R_k = \{(x, y) : \mathrm{SI}_k(x, y) = 0 \text{ and } \mathrm{SI}_l(x, y) = 1 \text{ if } l \neq k\}, \; k \text{ and } l \in \{1, 2, 3\}. \tag{3}$$

In order to distinguish extra candidates from the undefined regions, we use a technique of a RAG. For constructing a RAG, region labeling based on eight-connectivity is first performed. In RAG, core candidate regions, backgrounds, and undefined regions are represented as nodes and each pair of adjacent regions is connected by an edge. In this paper, three RAGs are generated for distinguishing extra candidates. RAG_1 of the core candidate regions, and Undef_R_1 is first constructed. Then, by using Undef_R_2 and Undef_R_3, RAG_2 and RAG_3 are constructed by the same method for RAG_1, respectively. Figure 4 shows an example of RAG_1 construction. Figure 4(a) shows a binary map composed of two labeled regions, such as background (R_1) and core candidate regions (R_2) obtained by Eq. (2). Figure 4(b) represents two undefined regions labeled as R_3 and R_4, respectively. Figure 4(c) is a trimap generated by combining Figs. 4(a) and 4(b). Figure 4(d) represents a RAG corresponding to Fig. 4(c).

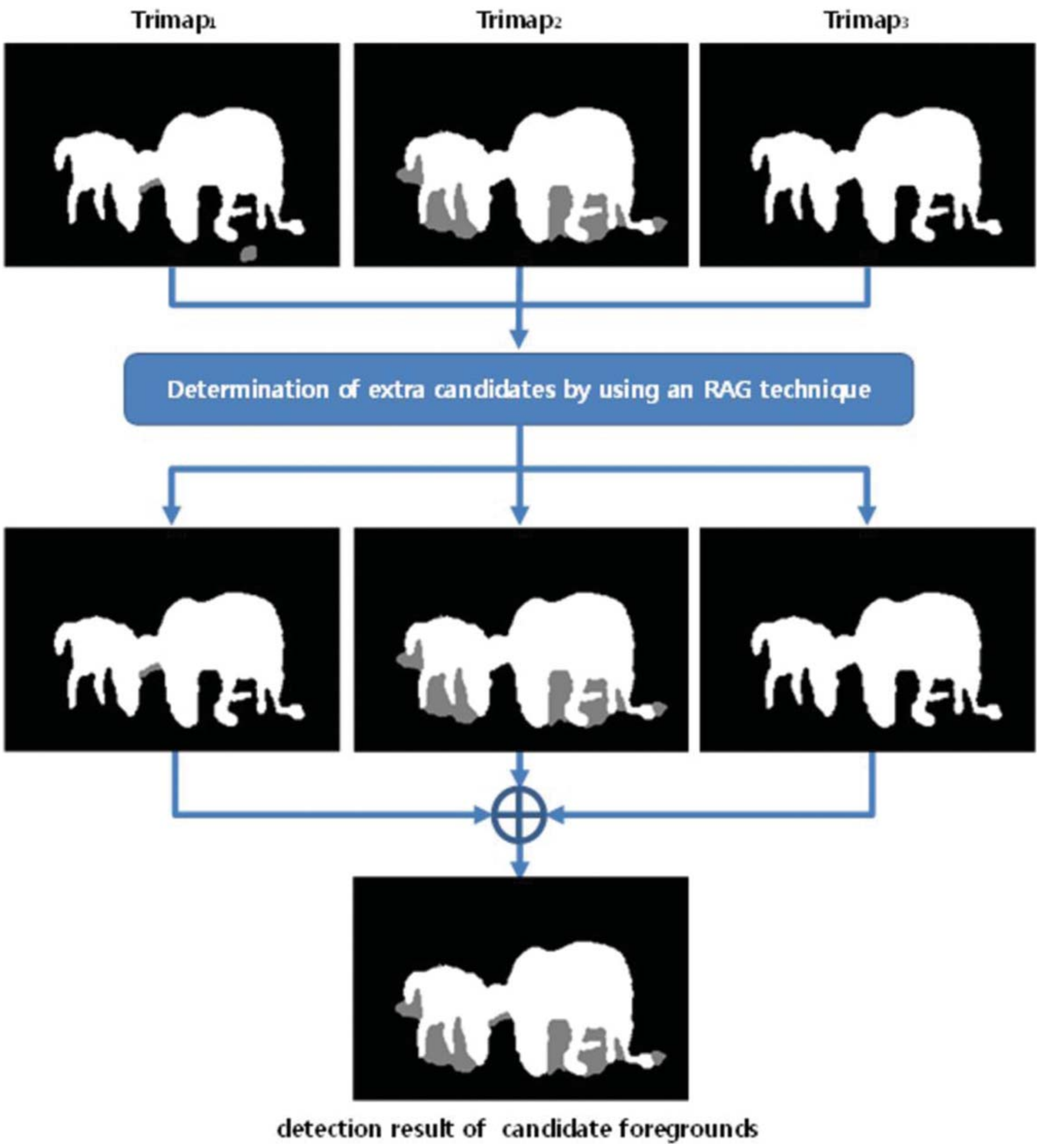
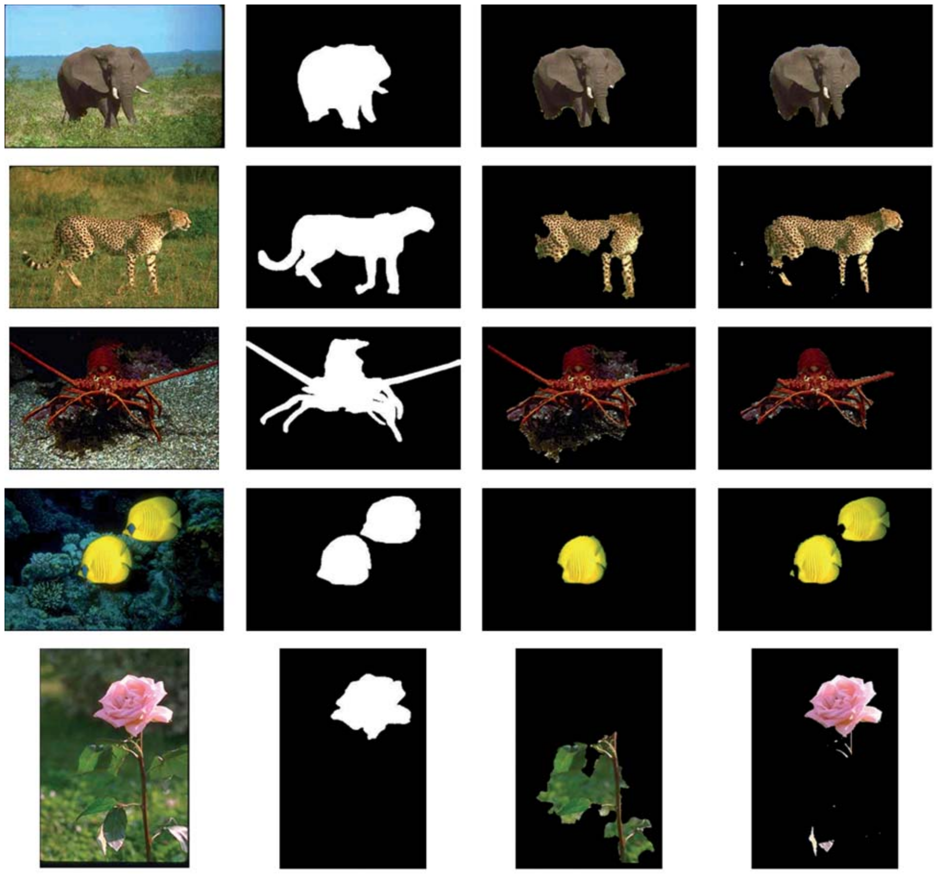


Fig. 5 Detection of candidate foregrounds by using a RAG method.

(a) (b)

Fig. 6 Foreground regions extracted by using the graph cut technique from candidate foreground: (a) Candidate foreground and (b) extracted foreground regions.



(a) Original image (b) Ground truth (c) Results of Kim's method (d) Results of proposed method

Fig. 7 Experimental comparison of salient object extraction.

After constructing these RAGs, some regions of undefined regions that are connected with core candidate regions from each RAG are additionally considered as extra candidates, whereas other regions not connected with core candidate regions are regarded as backgrounds. Accordingly, candidate foregrounds composed of both core candidate regions and extra candidates are used for exact foreground extraction. Figure 5 shows a flowchart of distinguishing extra candidates from undefined regions. As shown in Fig. 5, candidate foregrounds are finally extracted. In Fig. 5, Trimap_k ($k = 1,2,3$) is a trimap corresponding to RAG_k. The $\oplus$ implies the combination of three images.

4 Extraction of Salient Object Regions by Iterative Graph Cut

GrabCut[13] is an interactive tool based on an iterative graph cut for foreground extraction in still images. To segment a foreground object using the GrabCut, users should draw a rough bounding box, including foreground regions in an image for establishing an initial region. However, in this paper, candidate foreground regions that are automatically detected by background elimination are used as initial region, instead of a bounding box established by users.

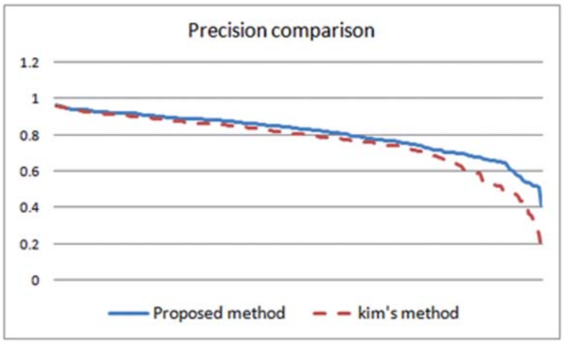
Let I be an image as an array $\mathbf{z} = (z_1, \ldots, z_n, \ldots, z_N)$ of pixels indexed by a single index n, where z_n is in RGB space. A segmentation result is expressed by an array of "opacity" values $\alpha = (\alpha_1, \ldots, \alpha_n, \ldots, \alpha_N)$, where α_n is 0 for background and 1 for foreground at each pixel. For the purpose of segmentation, GrabCut constructs two separate Gaussian mixture models (GMMs) to express color distributions for the background and foreground. Each GMM, one for foreground and one for background, is set to be a full covariance Gaussian mixture with K components. GrabCut defines an energy function E such that its minimum should correspond to a good segmentation in the sense that it is guided by the observed foreground and background GMMs and that the opacity is "coherent." This is achieved by "Gibbs" energy in

$$\mathbf{E}(\underline{\alpha}, \mathbf{k}, \underline{\theta}, \mathbf{z}) = U(\underline{\alpha}, \mathbf{k}, \underline{\theta}, \mathbf{z}) + V(\underline{\alpha}, \mathbf{z}), \tag{4}$$

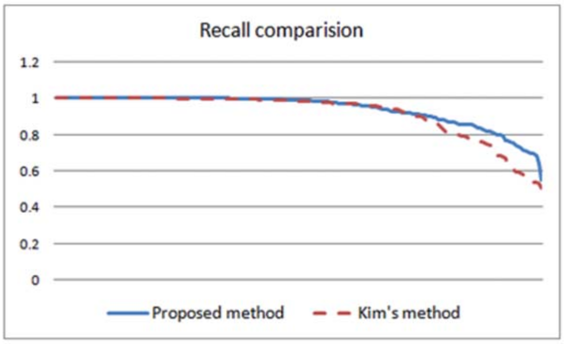
where U evaluates the fit of the opacity distribution α to data $\mathbf{z}$ under the color GMM models. V is called the smoothness term. In this paper, each of five components is used for expressing color distributions for background and foreground, respectively. Figure 6 shows the extraction result of foreground regions by using the GrabCut method.

5 Experimental Results

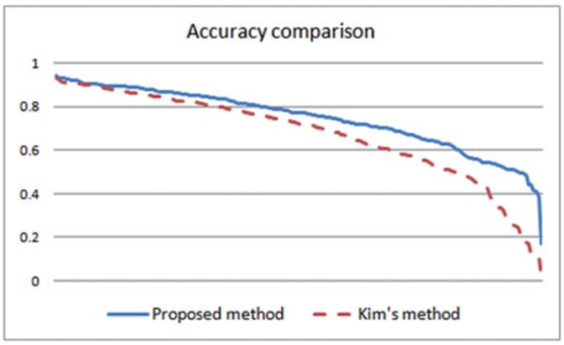
In order to evaluate the performance of the proposed method, a test set of 200 images containing multiple objects and clutter backgrounds is tested and the performance is compared to a conventional method. For applying the bilateral filter, two parameters $\boldsymbol{\sigma}_d$ and Σ_r were set to 3 and $\lfloor 30^2\ 0\ 0;0\ 30^2\ 0;0\ 0\ 30^2 \rfloor$, respectively. Figure 7 shows the experimental



(a) Precision comparison between the proposed method and Kim's method



(b) Recall comparison between the proposed method and Kim's method



(c) Accuracy comparison between the proposed method and Kim's method

Fig. 8 Performance evaluation of the proposed method and Kim's method.

comparison of foreground regions extracted by the method of Kim *et al*.[3] and the proposed method. As shown in Fig. 7, because Kim's method extracts the connected regions located close to the center of the image as a foreground region, it is impossible to extract object regions that are not located at the center of image and multiple objects, which are separately located. On the other hand, by extracting the foregrounds close to real object regions through background elimination and the estimation of foreground regions from the candidate object regions using graph-cut technique, the proposed method shows better performance of foreground extraction, compared to the method of Kim *et al*.[3] Foreground regions extracted by the proposed method are compared to those of the ground truth.

In this paper, the precision, recall, and accuracy are used as the performance measures, which are defined as

$$\text{precision} = \frac{N(S_M \cap S_A)}{N(S_A)} \times 100,$$

$$\text{recall} = \frac{N(S_M \cap S_A)}{N(S_M)} \times 100, \tag{5}$$

$$\text{accuracy} = \frac{\max\{N(S_M) - [N(S_U) + N(S_O)]\ ,\ 0\}}{N(S_M)} \times 100,$$

where $N(S_M)$ and $N(S_A)$ are the number of pixels in the ground truth and in the extracted foreground regions, respectively. $N(S_M \cap S_A)$ denotes the number of identical pixels between the ground truth and the extracted foreground regions. The precision is a measure of how accurately real foreground regions are extracted compared to the extracted regions. The recall is a measure of how well the extracted regions represent the ground truth, and the accuracy is a measure of the percentage of the foreground regions excluding errors, with respect to the ground truth.[3] The performance comparisons of these measures have been shown in Fig. 8. As shown in Fig. 8, the precision, recall, and accuracy of the proposed method have been improved, on an average, by 5.2, 2.2, and 11.9%, respectively, compared to the method of Kim *et al*.[3]

6 Conclusion

In this paper, we have proposed a novel approach to automatic extraction of foreground regions. In order to extract exact foreground regions, the detection of candidate object regions through multiscale segmentation-based background elimination using trimap and RAG techniques have been proposed and the graph-cut technique was used for extracting foreground regions. By using the proposed method, more exact foreground regions have been extracted, compared to the method of Kim *et al*.,[3] regardless of the location of objects, whereas existing method did not extract some foreground regions not at the center of an image.

Acknowledgments

This work was supported by a National Research Foundation of Korea Grant funded by the Korean Government (Grant No. 2009–0077434).

References

1. K. T. Park and Y. S. Moon, "Automatic extraction of salient objects using feature maps," in *Proc. of IEEE Conf. on Acoustics, Speech and Signal Processing*, pp. 617–620 (2007).
2. K. T. Park, J. H. Lee, and Y. S. Moon, "Unsupervised foreground segmentation using background elimination and graph cut techniques," *Electron. Lett.* **45**(20), 1025–1027 (2009).
3. S. Kim, S. Park, and M. Kim, "Central object extraction for object-based image retrieval," *Lect. Notes Comput. Sci.* **2723**, 38–47 (2004).
4. D. Crevier, "Extracting salient objects from operators-framed images," in *Proc. of 4th Canadian Conf. on Computer and Robot Vision*, pp. 36–43 (2007).
5. M. Kimura and M. Yamauchi, "A method for extracting regions of interest based on attractiveness," *IEEE Trans. Consumer Electron.* **52**(2), 312–316 (2006).
6. G. Ciocca, C. Casso, F. Gasparini, and R. Schettini, "Self-adaptive image cropping for small displays," *IEEE Trans. Consumer Electron.* **53**(4), 1622–1627 (2007).
7. B. Ko and J. Nam, "Object-of-interest image segmentation based on human attention and semantic region clustering," *J. Opt. Soc. Am. A* **23**(10), 2462–2470 (2006).
8. Q. Huang, B. Dom, D. Steels, J. Ashely, and W. Niblack, "Foreground/background segmentation of color images by integration of multiple cues," in *Proc. of IEEE Conf. on Image Processing*, pp. 246–249 (1995).
9. Y. Lu and H. Guo, "Background removal in image indexing and retrieval," in *Proc. Int. Conf. on Image Analysis and Processing*, pp. 933–938 (1999).
10. C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. of IEEE Conf. on Computer Vision*, pp. 839–846 (1998).

11. S. Paris, P. Kornprobst, J. Tumblin, and F. Durand, "Bilateral filtering: theory and applications," *Found. Trends Comput. Graphics Vis.* **4**(1), 1–73(2009).
12. Y. Deng and B. S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video," *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(8), 800–810 (2001).
13. S. Feng, D Xu, and X Yang, "Attention-driven salient edge(s) and regions(s) extraction with application to CBIR," *Signal Process.* **90**(1), 1–15 (2010).
14. C. Rother, V. Kolmogorov, and A. Blake, "GrabCut-Interactive foreground extraction using iterated graph cut," *ACM Trans. Graphics* **23**(3), 309–314 (2004).

Ki Tae Park received the BS, MS, and PhD in computer science from Hanyang University, Republic of Korea, in 2000, 2002, and 2007, respectively. He had previously been worked in Samsung Electronics Co., Ltd., Suwon, Republic of Korea. He is currently working as research professor at Hanyang University. His major research interests are in the areas of digital signal processing, content-based image retrieval, object segmentation, multimedia processing, and video indexing.

Chan Woo Park received his MS in computer science and engineering from Hanyang University, Republic of Korea, in 2007. He is currently working toward his PhD at the Computer Vision and Pattern Recognition Laboratory, Department of Computer Science and Engineering, Hanyang University, Republic of Korea. His major interests include face processing, object segmentation, and target tracking.

Young Shik Moon received his BS and MS in electronics engineering from Seoul National University and Korea Advanced Institute of Science and Technology, Republic of Korea, in 1980 and 1982, respectively. From 1982 to 1985, he had been a researcher at the Electronics and Telecommunications Research Institute, Republic of Korea. He received his PhD in computer engineering from University of California at Irvine, in 1990. Since 1992, he has been a professor in Department of Computer Science and Engineering at Hanyang University, Republic of Korea. His research interests include content-based image retrieval, object segmentation, target tracking, and data association.