Depth Guided Selection of Adaptive Region of Interest for Grabcut-Based Image Segmentation

Garam Kim and Jae-Young Sim
School of Electrical and Computer Engineering, UNIST, Ulsan, Korea
E-mails: {gkim, jysim}@unist.ac.kr

***Abstract*—Grabcut is an efficient image segmentation technique which facilitates easy user interaction by locating a rectangular bounding box to include the foreground objects. However, when the foreground objects exhibit similar colors to that of the background, it often fail to work to accurately classify the pixels within the interior region of the bonding box. In this paper, we propose an adaptive region of interest selection algorithm for Grabcut-based image segmentation. We first obtain an initial segmentation result by performing the Grabcut on the depth image aligned to an input color image. Then we shrink and enlarge the depth segmentation mask by using the erosion and dilation operations. We regard the outside of the enlarged mask as background pixels and regard the interior of the shrunken mask as foreground pixels. The remaining pixels are classified into the foreground objects and the background by performing the Grabcut using the four-channel Gaussian mixture model of RGB colors and depth. Experimental results show that the proposed algorithm effectively suppress the false detection of objects and improves the segmentation performance compared with the existing algorithms by adaptively selecting the region of interest.**

I. Introduction

Image segmentation is one of the fundamental techniques of image processing and computer vision, which separates foreground objects from backgrounds in an input image. One of the conventional approaches of image segmentation employs user interaction to manually annotate some pixels as foreground objects and backgrounds and solves an optimization problem of labeling all pixels using graph-cut [1]. Grabcut [2] is an advanced version of this approach, which uses a rectangular bounding box to denote foreground objects by user interaction and employs the three dimensional Gaussian mixture model (GMM) to represent the distribution of color statistics. The Grabcut technique is relatively easy to use and yields good performance, and thus has been widely used as an efficient tool of image segmentation. Moreover, its performance has been improved by follow-up research works. For example, Chen et al. [3] improved the Grabcut algorithm by selecting optimal numbers of Gaussians in GMM modeling, and Kim et al. [4] used a saliency map of an image to resize the initial bounding box to improve the performance of the Grabcut method.

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(2013R1A1A2011920).

However, these methods use only the monochrome images or color images, and thus often fail to detect foreground objects accurately when the luminance or color characteristics of the foreground objects are similar to that of the backgrounds. To overcome this limitation, several research works have been performed by adopting other modalities of image data such as depth data, in addition to the colors. He et al. [5] linearly combined the energy terms of color and depth together into the energy minimization framework of the Grabcut algorithm. Vaiapury et al. [6] regarded the depth data as as an additional channel to the three color channels, and employed a four dimensional GMM based on the Grabcut algorithm. Such methods further improve the performance of the ordinary Grabcut. However, there is still a problem regarding the initialization of user-interacted bounding box. In particular, when the foreground objects exhibit complex shapes composed of convex and concave parts together, the rectangular shape of bounding box may not completely separate the foreground objects from the background.

In this paper, we propose an improved image segmentation algorithm based on Grabcut which adaptively select the region of foreground objects using depth data. We use a pair of aligned color and depth images. We locate rectangular bounding box on an input depth image and perform the ordinary Grabcut to get an initial segmentation result. Then we use the depth segmentation result as a candidate region of interest, which is then shrunken and enlarged by using the morphological operations. We partition an input image into the foreground object region, the background region, and the unknown region adaptively using the modified region of interest, which are then refined by applying the four-channel GMM based Grabcut. Experimental results show that the proposed algorithm yields a better performance compared with the existing Grabcut based segmentation algorithms.

The rest of this paper is organized as follows. Section II briefly describes the Grabcut algorithm. Section III introduces the proposed algorithm, and Section IV provides the experimental results. Finally, Section V concludes the paper.

II. Grabcut Algorithm

Let $\mathbf{I} = \{i_1, i_2, ..., i_N\}$ be the pixel array in an input image, where i_n represents the RGB color value at the n-th pixel. The segmentation problem is to determine an opacity array $\underline{\alpha} = \{\alpha_1, \alpha_2, ..., \alpha_N\}$, where α_n means the label assigned to

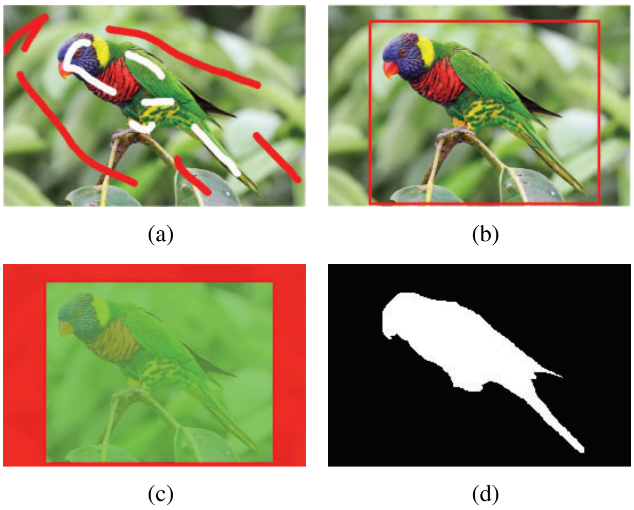

Fig. 1: User interaction for image segmentation. (a) Graph-cut: the white and red strokes denote the pixels in the foreground object and the background, respectively. (b) Grabcut: the red rectangle is an initial bounding box including the foreground object. (c) The pixels used to construct GMMs for the background and the foreground object region. (d) The segmentation result of Grabcut.

the n-th pixel. We set the opacity value as 0 for the background pixels and 1 for the foreground object pixels, respectively.

Grabcut is an user interaction based image segmentation technique which employs the GMM and performs the graph-cut based optimization iteratively. As shown in Fig. 1(a), the users draw strokes on the foreground object regions and the background region respectively, in the ordinary graph-cut based segmentation method. On the other hand, a rectangular bounding box is simply located to include the foreground objects in the Grabcut technique, as shown in Fig. 1(b), and therefore, the user interaction burden is much alleviated.

Using the user interaction, the pixels located outside of the bounding box are assigned the label 0, as shown in red in Fig. 1(c). Also, these pixels are used to construct the color distribution of the background based on GMM. The pixels inside of the box, as shown in green in Fig. 1(c), are used to construct the GMM for the foreground object region. Each GMM is composed of K components. For each green pixel, we assign the labels of the foreground objects and the background by solving the optimization problem iteratively. We design an energy function, which is composed of the regional cost term and the boundary cost term. The regional cost measures the proximity of each GMM between the foreground object region and the background. The boundary cost gives the penalty to the pixels which have different labels to their neighboring pixels.

III. Proposed Algorithm

While the Grabcut provides a relatively easy user interaction scheme and yields good performance of segmentation. it often fails to work on an input image where the foreground object exhibit similar colors to that of the background. When an initial bounding box is located on an input image, the outside region of the box is definitely belong to the background. However, the inside region of the box is uncertain region where each pixel should be assigned a label between 0 and 1. It means that the initially obtained GMM for the foreground object region also reflect the background pixels as well, which may degrade the performance of accurate classification of the associated pixels.

We call the user-interacted initial region of foreground objects as region-of-interest (ROI). In this work, we propose an adaptive ROI selection algorithm to alleviate the ambiguity of color similarity between the foreground object and the background. We adopt the depth image synchronized to an input color image, and use the depth data of pixels to first select the pixels of the foreground object region. Then we perform the Grabcut using the four-dimensional GMM modeling the RGB color and depth data together and obtain the final segmentation results.

We use the Kinect sensor to obtain the depth image synchronized to an input color image. A pair of color and depth images are shown in Fig. 2(a) and (b), respectively. We put an ROI as depicted in the red rectangle on the color image in Fig. 2(a), and apply the ordinary Grabcut. The segmentation result is shown in Fig. 2(c), where we see that a large amount of background areas are classified to the foreground objects. We also apply the Grabcut to the depth image in Fig. 2(b) using the same ROI, and obtain the segmentation result as shown in Fig. 2(d). Whereas the color image cannot be clearly classified due to the color similarity between the inside and the outside of the ROI, the depth image yields an accurate segmentation result.

Therefore, we use an arbitrary shape of ROI instead of rectangular bounding box, and set an initial ROI to be adapted to the shape of foreground objects using the depth information. Specifically, we first take the mask R_C of the labels of 1 in the segmentation result of depth image, which is used as a shape-adaptive ROI. Fig. 2(e) shows this candidate ROI aligned to the input color image. Then we perform the morphology operations of erosion and dilation to R_C to obtain the shrunken mask R_E and the enlarged mask R_D, respectively,

$$R_E = R_C \ominus S_p, \quad (1)$$
$$R_D = R_C \oplus S_q, \quad (2)$$

where $\ominus$ and $\oplus$ denote the erosion and dilation operator, respectively, and S_p is the $p \times p$ structuring element. We set $p = 10$ empirically. Note that $R_E \subset R_D$.

The shrunken ROI adaptively aligned to the shape of the foreground objects is highly probable to include only the foreground pixels. The outside of the enlarged ROI is highly probable to include only the background pixels. Therefore, using the resulting regions of R_E and R_D, we determine an initial trimap, which partitions an input image into the foreground object region T_F, background region T_B, and

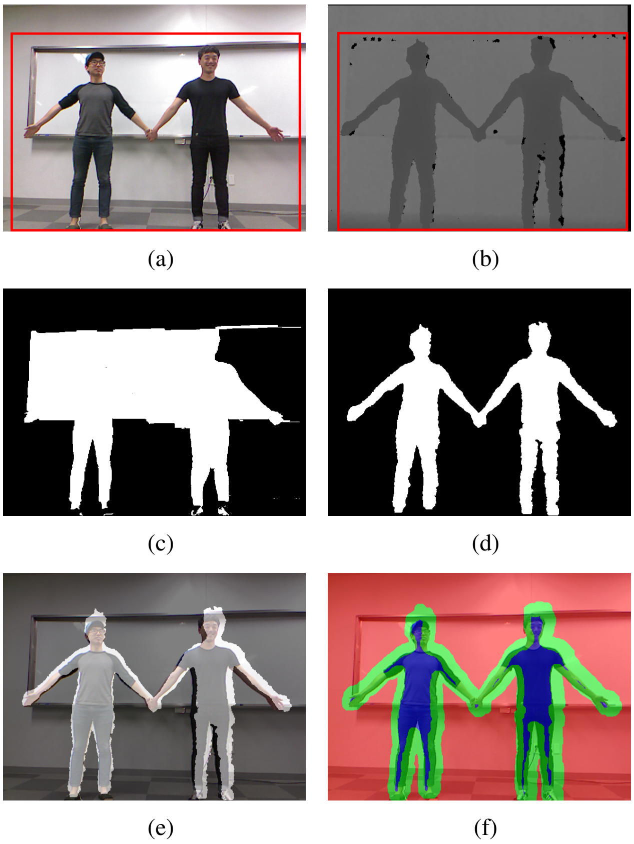

Fig. 2: Adaptive ROI selection. (a) Color image and (b) depth image, where the user-interacted bounding box is shown in red. (c) and (d) are the segmentation results by using the Grabcut on the color image and the depth image, respectively. (e) The ROI of the depth image segmentation result aligned to the color image. (f) The initial trimap generated by shrinking and enlarging the ROI, where the foreground, background, and unknown regions are shown in blue, red, and green, respectively.

unknown region T_U, respectively.

$$T_F = R_E, \tag{3}$$
$$T_B = \overline{R_D}, \tag{4}$$
$$T_U = R_D - R_E, \tag{5}$$

where $\overline{R}$ is the set of the pixels which do not belong to R. Fig. 2(f) shows the resulting trimap, where T_F, T_B, and T_U are depicted in blue, red, and green colors, respectively. Using the initial trimap, we apply the Grabcut using the four-channel GMM considering the RGB color and depth data together [6].

IV. Experimental Results

We evaluate the performance of the proposed algorithm using 90 test images capturing various human poses and indoor objects and structures using the Microsoft Kinect for Windows v2. We used the aligned color and depth images. Fig. 3 shows 12 images of the test dataset.

TABLE I: Comparison of quantitative performances.

	OGC [2]	OGD	[5]	[6]	Proposed
Precision	0.71	0.74	0.82	0.83	0.90
Recall	0.90	0.83	0.87	0.93	0.91
F-score	0.79	0.78	0.85	0.88	0.91

In Fig. 3, we compare the segmentation results of the proposed algorithm with that of the existing techniques: ordinary Grabcut on color image (OGC) [2], ordinary Grabcut on depth image (OGD), advanced Grabcut using linear combination of color and depth costs [5], and the advanced Grabcut using four-dimensional GMM [6]. We see that OGC [2] results in false detection for most test images, for example, the background pixels located on the concave regions of the foreground objects are detected as foreground pixels in ”TWO PEOPLE1”, ”CHAIR1”, ”PERSON1”, ”CHAIR2”, and ”BASKET” images. These falsely detected pixels are removed in the segmentation results of OGD by using the depth discontinuity of foreground objects compared to the backgrounds. However, the OGD also cause the false detection in ”TWO PEOPLE2”, ”TWO PEOPLE3”, and ”TWO PEOPLE4” images due to the ambiguity of depth values between the feet and the floors. The two advanced algorithms which use the color and depth data together achieve better performances than OGC in general, but also have large errors in ”FOUR PEOPLE”, ”TWO PEOPLE2”, ”TWO PEOPLE3”, and ”TWO PEOPLE4” images. On the contrary, the proposed algorithm captures the foreground objects more accurately and reduces the false detection more faithfully compared with the existing methods, especially on ”FOUR PEOPLE”, ”TWO PEOPLE2”, and ”TWO PEOPLE4” images.

In Table 1, we also compare the quantitative performance of the segmentation algorithms in terms of the precision, recall, and F-score measured on the 90 test images. We see that the proposed algorithm yields the best performance in terms of the F-score compared with the existing methods.

V. Conclusions

We proposed an adaptive region of interest selection algorithm for Grabcut-based image segmentation. We employed the depth image aligned to an input color image to get an initial segmentation result. Then we used the depth segmentation mask as a candidate ROI for the foreground objects. The ROI is shrunken and enlarged by using the erosion and dilation operators, which results in a trimap adapted to the shape of the foreground objects. We used the four-channel GMM to model the distribution of RGB colors and depth and apply iterative optimization of Grabcut algorithm to obtain the final segmentation results. Experimental results demonstrated that the proposed algorithm improved the performance of segmentation compared with the existing Grabcut-based algorithms.

References

[1] Y. Boykov, and V. Kolmogorov, “An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision,” *IEEE. Trans-*


(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 3: The segmentation results of the proposed algorithm compared with that of the four algorithms: ordinary Grabcut on color image (OGC) [2], ordinary Grabcut on depth image (OGD), advanced Grabcut using linear combination of color and depth costs [5], and the advanced Grabcut using four-dimensional GMM [6]. From top to bottom, ”TREE”, ”TWO PEOPLE1”, ”CHAIR1”, ”FOUR PEOPLE”, ”PERSON1”, ”TWO PEOPLE2”, ”CHAIR2”, ”CHAIR3”, ”TWO PEOPLE3”, ”TWO PEOPLE4”, ”BASKET”, and ”PERSON2”, respectively. (a) Input color images and (b) the aligned depth images. The segmentation results of the (c) OGC [2], (d) OGD, (e) [5], (f) [6], and (g) the proposed algorithm, respectively. (h) The ground truth results.

actions on pattern analysis and machine intelligence, 26.9, 1124-1137, 2004.

[2] C. Rother, V. Kolmogorov, and A. Blake, “Grabcut: Interactive foreground extraction using iterated graph cuts,” *ACM. Transactions on Graphics,* vol. 23, no. 3, 2004.

[3] D. Chen, B. Chen, G. Mamic, C. Fookes, and S. Sridharan, “Improved grabcut segmentation via gmm optimisation,” *IEEE. Digital Image Computing Techniques and Applications,* pp.39-45, 2008.

[4] K. S. Kim, Y. J. Yoon, M. C. Kang, J. Y. Sun, and S. J. Ko, “An improved GrabCut using a saliency map,” *IEEE. Global Conference on Consumer Electronics,* pp.317-318, 2014.

[5] H. He, D. McKinnon, M. Warren, and B. Upcroft, “Graphcut-based interactive segmentation using colour and depth cues,” , 2010.

[6] K. Vaiapury, A. Aksay, and E. Izquierdo, “Grabcutd: improved grabcut using depth information,” *AMC. Workshop on Sur. Med. Vir. Clo.* pp.57-62 , 2010.