

# Saliency Cuts: Salient Region Extraction based on Local Adaptive Thresholding for Image Information Recognition of the Visually Impaired

Mukhriddin Mukhiddinov<sup>1</sup>, Rag-Gyo Jeong<sup>2</sup>, and Jinsoo Cho<sup>3</sup>

<sup>1</sup>Department of Hardware and Software of Control Systems in Telecommunications, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Uzbekistan

<sup>2</sup>Korea Railroad Research Institute, Uiwang, Gyeonggi-do 16105, Republic of Korea

<sup>3</sup>Department of Computer Engineering, Gachon University, Republic of Korea

**Abstract:** *In recent years, there has been an increased scope for assistive software and technologies, which help the visually impaired to perceive and recognize natural scene images. In this article, we propose a novel saliency cuts approach using local adaptive thresholding to obtain four regions from a given saliency map. The saliency cuts approach is an effective tool for salient object detection. First, we produce four regions for image segmentation using a saliency map as an input image and applying an automatic threshold operation. Second, the four regions are used to initialize an iterative version of the Grab Cut algorithm and to produce a robust and high-quality binary mask with a full resolution. Lastly, based on the binary mask and extracted salient object, outer boundaries and internal edges are detected by Canny edge detection method. Extensive experiments demonstrate that the proposed method correctly detects and extracts the main contents of the image sequences for delivering visually salient information to the visually impaired people compared to the results of existing salient object segmentation algorithms.*

**Keywords:** *Saliency region extraction, saliency map, saliency cuts, local adaptive thresholding, the visually impaired.*

*Received February 7, 2018; accepted January 6, 2020*  
<https://doi.org/10.34028/iajit/17/5/4>

## 1. Introduction

The task of accurately detecting the most visually significant foreground object in natural scene images salient object detection, has found a wide range of applications in computer vision and drawn a substantial amount of research interest [7]. The task is universally used as a first processing step in computer vision field including image and video compression, image retargeting, image retrieval, object-of-interest image segmentation, object recognition, content-aware image editing, sensation enhancement, image classification. Extracting salient objects is also crucial to generate tactile representation for the visually impaired to help them clearly understand important visual information contained in images. Salient object detection mainly consists of two steps -saliency map generation and saliency cuts. In the first step, a pixel- or region-level saliency map is generated, in which each pixel or region is assigned with a value proportional to its saliency. In [17], we presented a saliency detection method to produces a saliency map. Saliency cut is a binarization process to find a binary mask for salient objects using the saliency maps as its input.

We proposed novel automatic object segmentation method using local adaptive thresholding to estimate

three different threshold values for each pixel to segment the image into four regions using local neighboring pixel information utilizing integral images [20]. The threshold values are fed into Grab Cut algorithm [18] as seed to generate binary mask indicating the region of the salient object. For clearer tactile representation, outer contours and inner edges of the region is isolated in the final step.

The main contributions of the proposed method are summarized as:

- Extraction of salient regions from a saliency map using adaptive triple thresholding and the Grab Cut algorithm.
- Combination of the Grab Cut algorithm with the generated four-region seeds to refine the segmentation results.
- Processing of the extracted salient region suitable for the assistive technology system [11] for the visually impaired, allowing them to easily perceive and recognize natural scene images.

## 2. Literature Review

Saliency models based on bottom-up approach convert natural scene images into saliency maps, where each pixel, superpixel, or region is assigned with a saliency

value or probability [15]. These methods apply image segmentation techniques (e.g., graph-based [9], mean shift [8], or superpixels [1]) to the input image and segment homogeneous regions to extract their feature statistics to detect salient regions. Cheng *et al.* [7] demonstrate two saliency models: Histogram-Based Contrast (HC), which assigns a pixel-wise saliency value, and Region-Based Contrast (RC), which incorporates spatial relations at the cost of reduced computational efficiency. Zhou *et al.* [22] combine widely used contrast measurements, namely, center-surround, corner-surround, and global contrast to detect visual saliency.

Another popular approach is saliency object detection methods based on deep Convolutional Neural Networks (CNNs). Lee *et al.* [12] introduce the Encoded Low-Level Distance Map (ELD-map), which directly encodes the feature distance between each pair of superpixels in an image. The encoded feature distance map has a strong discriminative power to evaluate similarities between different parts of an image with precise boundaries among superpixels. Liu and Han [14] demonstrate an end-to-end saliency detection model to detect salient objects with a new hierarchical refinement model, which can refine saliency maps hierarchically and progressively to recover image details by integrating local context information without using over-segmentation methods.

In recent, approaches using saliency cuts have been widely explored, as these methods focus on providing a binary mask of salient objects with the aid of a saliency map. Saliency cuts automatically segment a salient object from the background. With a saliency map input and using the iterative Grab Cut algorithm [18], we can extract a precise image mask [7, 10, 13, 18]. Fu *et al.* [10] modify the graph cut method by exploring the effects of labels for graph-based segmentation. Aytekin *et al.* [3] propose a link between quantum mechanics and spectral graph clustering, referred to as Quantum Cuts, which forms a graph among superpixels extracted from an image, then optimizes a criterion related to the image boundary, local contrast, and area information. Mehrani and Veksler [16] exploit the standard features often used in vision based on factors such as color and texture. Properly normalized, these simple features yield performances superior to those of the methods based on hand-crafted features specifically designed for saliency detection. Li *et al.* [13] propose a saliency cuts approach using the Otsu thresholding technique and the Grab Cut algorithm. They improve the Otsu algorithm to calculate three level thresholds, and the saliency map is split into four regions using the three thresholds.

We also surveyed relevant literature on global and local adaptive thresholding. AlSaeed *et al.* [2] proposes an unsupervised method of automatic thresholding, which is improved version of the most widely used

Otsu's global thresholding techniques. In contrast, local adaptive thresholding used for binarization can account for variations in illumination. Sauvola and Pietikinen [19] demonstrate adaptive document image thresholding, in which a page is considered a collection of subcomponents such as text, background, and picture. Bradley and Roth [5] propose real-time adaptive thresholding using an integral image of the input image. Another extension of Bradley's algorithm, a new local adaptive thresholding technique, is proposed by Benny and Soumya [4]. They propose a hand-written character recognition system using a local thresholding method for binarization and a dynamic self-organizing feature map technique for classifying feature vectors extracted from characters.

Visual information generation based on tactile graphics is one of the most important works since it can provide the visually impaired with graphic information. Jungil *et al.* [11] developed an education assistive technology system based on a graphic haptic electronic board. This system enables authoring, automatic image/graphic conversion, and real-time distribution of education materials for low-vision and blind students. Chen and Takagi [6] suggest a method for automatically translating hand-drawn maps into tactile maps using a pattern recognition technique for extracting and classifying objects in hand-drawn maps. Takagi and Chen [21] introduce a method for extracting character strings from scene images using edge detection, a morphology operator, and a fuzzy inference technique. The above works help the visually impaired to recognize important visual information from images and even enable them to walk more independently on the street.

### 3. Proposed Method

#### 3.1. Overview

Figure 1 illustrates the overall steps of the proposed saliency cuts method.

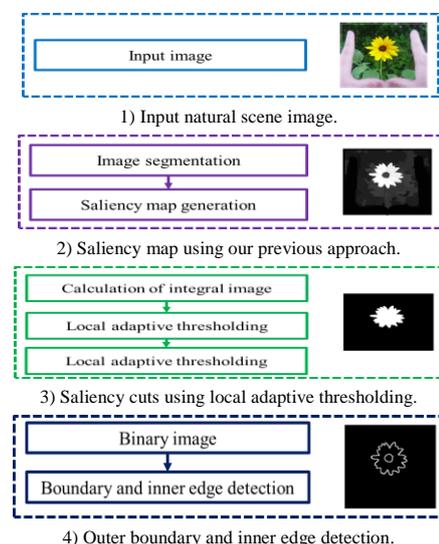


Figure 1. The proposed method.

We first generate a saliency map from a natural scene image using our previous work [17]. Then we describe how an integral image is calculated a given saliency map image which is computationally effective pre-processing for the next step. We then obtain locally adaptive three-level threshold values based on comparison. Three threshold levels classify each pixel into one of four categories-certain background, probable background, probable foreground, and certain foreground, which are used as for types of seeds in the next step. The seeds are assigned as iterative variants in the Grab Cut method which to produce a full-resolution binary mask. Finally, we detect the boundary and the inner edges of salient region to refine visually salient information for the visually impaired. We describe each step in detail in the following subsections.

### 3.2. Saliency Map Generation

In [17], we presented a method to generate saliency map. The step involves, first, contrast enhancement followed by segmentation with additional improvement using spatial prior and color space smoothing. The output of this step is a gray scale image in which the value of each pixel corresponds to the strength of saliency. The reader is referred to [17] for further details.

### 3.3. Saliency Cuts Using Local Adaptive Thresholding

#### 3.3.1. Local Adaptive and Global Thresholding

We briefly compare the local thresholding, which we employ, and the global thresholding before describing the method in detail in the following subsection.

Global techniques assign one threshold to the entire image, while local adaptive thresholding techniques assign a varying threshold value to each pixel/region determined by neighboring pixels. Global binarization approaches are more adequate for images with uniform contrast distribution of background and foreground, such as single-object or bi-modal histogram images. However, in complex natural images that contain considerable background noise or variations in contrast and illumination, many pixels cannot be easily classified as foreground or background. In such cases, binarization with global thresholding is not a suitable option.

Local thresholding techniques apply a unique threshold value to a single pixel or a certain region. The local threshold value can be calculated using various information contained in a given image. While local adaptive thresholding approaches generally achieve better results, they often rely on individual parameters and require a substantially higher amount of computation than global thresholding. In this paper, we reduce the computational cost for adaptive

thresholding by using integral image, which is explained in the next sub-section.

#### 3.3.2. Integral Images

An integral image (also known as a summed-area table) is a fast and effective means for computing the sum of pixel intensity values in a given image-or a rectangular subset of a grid of the given image. Mathematically, it can be expressed as

$$I(x, y) = \sum_{\substack{x' \leq x \\ y' \leq y}} i(x', y') \quad (1)$$

Where  $I$  is the integral value at point  $(x, y)$  and  $i$  is the intensity at point  $(x, y)$  in a grayscale image (saliency maps). Figure 2 shows how an integral image is produced. Figure 3 shows the an example where  $I(3, 3)$  is the sum of the intensity values. Integral image  $I(x, y)$  can be quickly calculated using,

$$I(x, y) = i(x, y) + I(x - 1, y) + I(x, y - 1) - I(x - 1, y - 1) \quad (2)$$

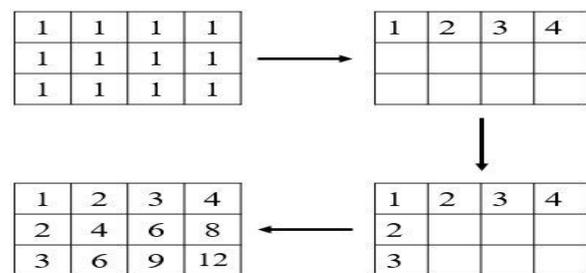


Figure 2. Steps for calculating integral image.

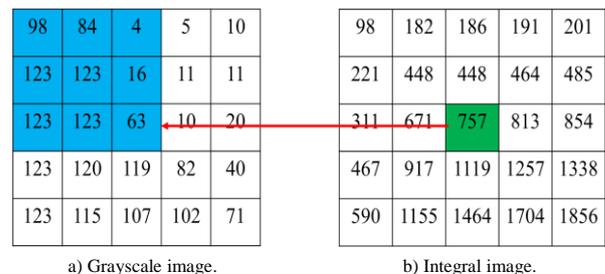


Figure 3. Example of integral image calculation.

#### 3.3.3. Local Adaptive Triple Thresholding

We derive three thresholding values based on the local mean for each pixel calculated over a window  $R_s$ . Using the integral image generated in the previous step, this step takes constant time for each  $R_s$  pixel independently from the size of the image. Using this technique, we can achieve a binarization speed close to those of global binarization methods. Bradley and Roth [5] used a window size that is 1/8th of the image width to calculate the average value. We chose our window size  $S$  as half of the image width, as it is the typical size of the saliency object size in natural scene images. Local adaptive triple thresholding process is performed as follows.

1. Sum of pixel values over a rectangle  $R$  within a

moving window with a size  $S \times S$  is calculated. In Figure 4, sum of pixel values in shadow area  $R$  is calculated with

$$R_s = I(x_2, y_2) - I(x_2, y_1 - 1) - I(x_1 - 1, y_2) + I(x_1 - 1, y_1 - 1) \quad (3)$$

Where  $x_1, x_2 = x \pm S/2$  and  $y_1, y_2 = y \pm S/2$ .

$(x_1, y_1)$ 98	84	$(x_2, y_1)$ 4	5	10
123	$(x, y)$ 123	16	11	11
123	123	$(x_2, y_2)$ 63	40	20
123	120	119	82	40
123	115	107	102	71

Figure 4. Sum over rectangle  $R$  in grayscale image.

2. The number of pixels in  $R$  is given with

$$N = (x_2 - x_1) \times (y_2 - y_1) \quad (4)$$

3. The local arithmetic mean  $m(x, y)$  at  $(x, y)$  is the average of the pixels within the window of size  $S \times S$ , and can be calculated as

$$m(x, y) = R_s / N \quad (5)$$

Benny and Soumya [4] calculated the optimal threshold value based on comparison: if the value of the current pixel is 15 percent less than the average, it is set to black; otherwise, it is set to white for document images. We improved this threshold value by assigning it to the local arithmetic mean value using (5) and (6). Because saliency map of natural scene images mainly consists of black (certain background) with some variety of grayscale values (unknown), we experimentally found that the local arithmetic mean value as a first-level threshold value yields good results for a variety of images.

4. We choose an initial threshold values as the average of the pixel values within the window.

$$t_l = m(x, y) \quad (6)$$

We also improved the Bradley algorithm [5] to obtain three-level thresholds, including  $t_l$ ,  $t_m$  and  $t_h$  ( $0 < t_l < t_m < t_h < L$ ) by decomposing the histogram of a saliency map into four regions.  $t_m$  and  $t_h$  will be defined below.

5. We calculate the additional threshold values  $t_m$  and  $t_h$  as follows:

$$t_h = t_l + ((L - t_l) / 2) \quad (7)$$

Where  $L$  is the maximum pixel value in the image, and

$$t_m = t_l + ((t_h - t_l) / 2) \quad (8)$$

These three threshold values categorizes each pixel segmenting the image into four regions,  $T_{cb}$  (Certain Background),  $T_{pb}$  (Probable Background),  $T_{pf}$  (Probable Foreground) and  $T_{cf}$  (Certain Foreground). Certain background regions are retained while other regions are to be refined during Grab Cut optimization.

### 3.3.4. Auto-Generated Seeds for Grab Cut

With adaptive three-level threshold values, we have already acquired seeds of four types for the masking requirement of the Grab Cut method [18]. For an image pixel value greater than  $t_m$  in (8), the largest connected region is considered the initial candidate region of the most dominant salient object. This candidate region is marked as probable foreground and certain foreground, while other regions are marked as probable background and certain background.

We obtain the contour of the salient object using the binary mask produced by Grab Cut method. These outer boundaries can be translated into tactile graphics to provide tactile information about a natural scene image to the visually impaired. For clearer representation for the visually impaired, we also demonstrate a technique for detecting the inner edges in the following as a post-processing step.

### 3.4. Boundary and Inner Edge Detection

In the first step of post-processing, we recover a salient object in full color space using the binary mask.

$$S_o = B_m(x, y) * I_i(x, y) \quad (9)$$

Where  $S_o$  is the salient image in full color space,  $B_m(x, y)$  is the binary mask, and  $I_i(x, y)$  is the input image. The recovered salient region is further processed using bilateral filtering with sigma value of 150 and Canny edge detection with filter size 9 to obtain visual representation suitable for tactile graphics for assistive technology systems. Figure 5 shows an example of the masking method using the proposed binary mask.

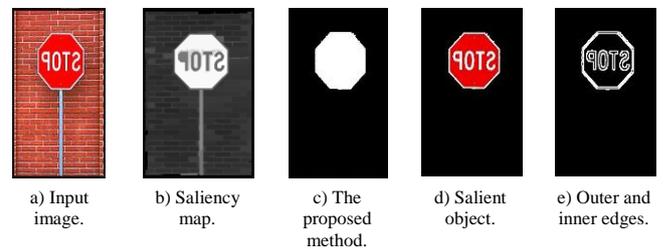


Figure 5. Example of using the masking method.

## 4. Experimental Results

For evaluation of the performance of the proposed saliency cuts method, we implemented the full processing pipeline in C++ and conducted experiments on a desktop Personal Computer (PC) with 2.60G Hz Intel processor and 4GB of Random-Access Memory (RAM). We used MSRA 10k dataset [7], which

includes 10,000 natural scene images and manually labeled ground truth images, exact and fully salient object(s) for the given image. We performed subjective qualitative evaluation and quantitative evaluation between the proposed method and others.

### 4.1. Qualitative Evaluation

We performed visual comparisons between the proposed method and two other state-of-the-art saliency cuts methods: saliency cuts method using Otsu Automatic Thresholding (AT-Cuts) [13] and saliency cuts method using Fixed Thresholding (FT-Cuts) [7].

In Figure 6, the first row displays typical input images containing people and various objects. Ground truth images are shown in the second row. The third row shows the saliency maps of the first row. The fourth and fifth rows are saliency cuts using the AT-Cuts and FT-Cuts, respectively. The last row shows the results of the proposed method. As shown in Figure 6 (6), the results of the proposed method for saliency map binarization are close to ground truth images. The proposed method extracts salient objects even when background and foreground regions have very similar information. In comparison, FT-Cuts can extract only a single salient object from a given image and fail to segment salient objects when the pixels in an image have low saliency values. AT-Cuts has a similar drawback of misclassifying a salient object as a background region when the saliency values of a given image are low. However, it can detect multiple objects if the saliency values of pixels are sufficiently high. In comparison, the proposed method can extract salient objects without misclassifying them as background regions, and is able to detect multiple salient objects.

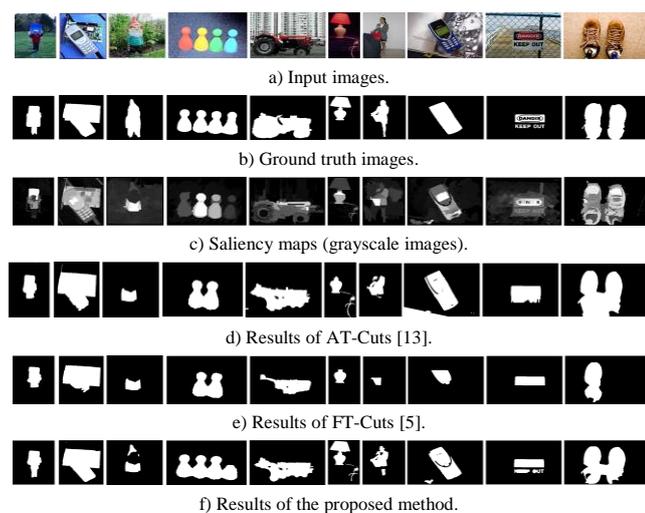


Figure 6. Visual comparisons between different methods.

The results of boundary detection and the comparisons between the proposed method and other methods are shown in Figure 7. The first row displays given input images containing people, birds, and

objects. The second and third rows show the saliency maps and ground truths generated from the input images, respectively. Results obtained using FT-Cuts and AT-Cuts are illustrated in the fourth and fifth rows, respectively. The results of the proposed method are shown in the sixth row, and the last row displays the detected boundaries of salient objects. As shown in Figure 7, the boundaries of salient objects are well detected using binary images and the Canny edge operator, providing valuable information to the visually impaired.

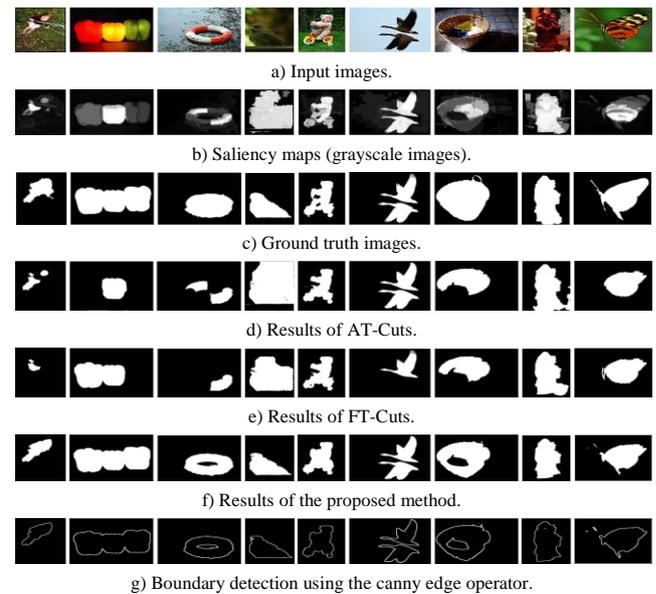


Figure 7. Experimental results of outer contour generation.

In some situations, however, the visually impaired might not be able to fully identify an object solely by touching the outer boundary. Thus, we also produced inner edges of salient objects as shown in Figure 8.

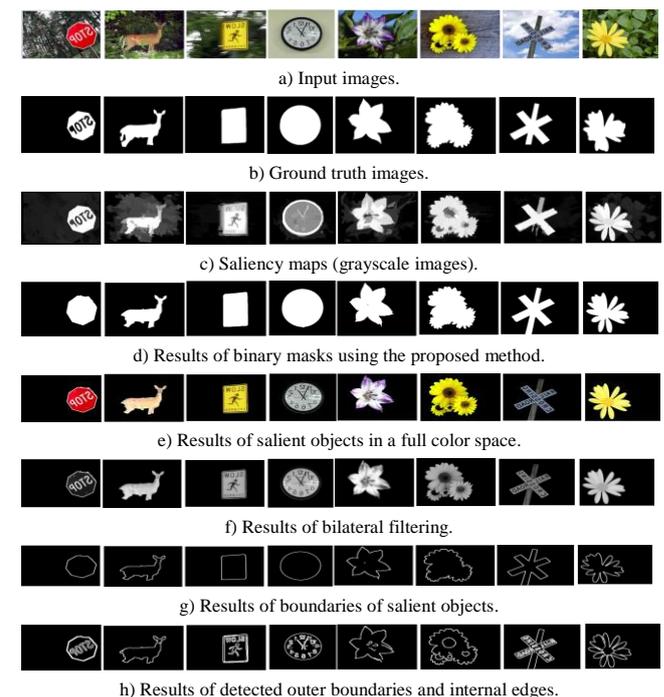


Figure 8. Experimental results of salient object extraction.

In Figure 8, the results of salient object segmentation from natural scene images and the results of detecting outer boundaries and internal edges illustrated. Figure 8 (4) shows the results of the proposed method for obtaining full resolution binary masks, which are almost identical to ground truth images (Figure 8 (2)) in most cases. Outer boundaries detected using Canny edge operator are shown in Figure 8 (7), and both outer and internal edges detected are shown in Figure 8 (8).

## 4.2. Quantitative Evaluation

We performed quantitative analysis by averaging precision and recall rates along with F-measures using MSRA 10k dataset. Precision and recall rates can be obtained as follows:

$$precision = \frac{TDO \cap GT}{ADO} \quad (10)$$

$$recall = \frac{TDO \cap GT}{GT} \quad (11)$$

Where  $TDO$  denotes truly detected salient objects,  $ADO$  all detected objects which is true positives plus false positives as shown in Figure 8 (4), and  $GT$  manually labeled ground truth as shown in Figure 8 (2). In other words,  $TDO$  is true positives and can be obtained by comparing all detected regions with manually labelled ground truth and select only regions that are matched with GT regions. As shown in Table 1, the proposed method accomplished the highest precision rate of 0.94. FT-Cuts [7] and AT-Cuts [13] yielded precision rates of 0.93 and 0.90, respectively. In addition, we computed the F-measure value, which balanced the measurements between the mean of precision and recall rates. A higher F-measure means a higher performance. F-measure is defined as follows:

$$F_{\beta} = \frac{(1 + \beta^2) precision \times recall}{\beta^2 \times precision + recall} \quad (12)$$

Where  $\beta^2 = 0.3$  to weigh precision more than recall as in many other research, including FT-Cuts and AT-Cuts. The proposed method achieved the highest F-measure value of 0.90, while the other two methods showed 0.82 and 0.88, respectively. The proposed method produced saliency objects with the highest precision rate, while AT-Cuts showed the highest recall rate.

Table 1. Quantitative analysis of the proposed method and other two saliency cuts methods: average precision, recall rate, and F-measure using MSRA 10k dataset.

	FT-Cuts [5]	AT-Cuts [13]	Proposed method
<b>Precision</b>	0.93	0.90	0.94
<b>Recall</b>	0.77	0.82	0.80
<b>F-measure</b>	0.82	0.88	0.90

Although all three methods produced similar results, the proposed method is proved to be the most robust in

complex images containing foreground and background with similar appearances, as demonstrated in Figures 6 and 7.

## 4.3. Performance Analysis

We compared performance analysis by comparing running times of FT-Cuts, AT-Cuts, and the proposed method are shown in Table 2 for the images in the MSRA 10k dataset (typical resolution of 400×300) using a PC with 2.60G Hz Intel processor and 4GB of RAM. Note that, running times for the each method were obtained by adding the time of saliency map generation and the time of saliency cuts. The FT-Cuts method is the fastest (1.24 seconds per image in average) since the fixed global threshold, followed by the proposed method (1.53 seconds in average) since the local adaptive thresholds per a certain region require more calculation time than fixed global threshold and saliency map generation method [17] uses global contrast enhancement technique but these give the most robust results, AT-Cuts method (1.86 seconds in average).

Table 2. Performance analysis comparison of average processing times per image of the proposed method and other two saliency cuts methods.

	FT-Cuts [5]	AT-Cuts [13]	Proposed method
<b>Average processing times per image (s)</b>	1.24	1.86	1.53

## 4.4. Tactile Graphic Generation

We generated tactile graphics of salient objects using the proposed method and the assistive technology system software [11] in order to actually help the visually impaired better perceive and recognize natural scene images, as shown in Figure 9. Outer and internal edges of salient object as well as tactile graphics based on dot representation for the visually impaired are illustrated in Figure 9 (2) and Figure 9 (3) respectively.

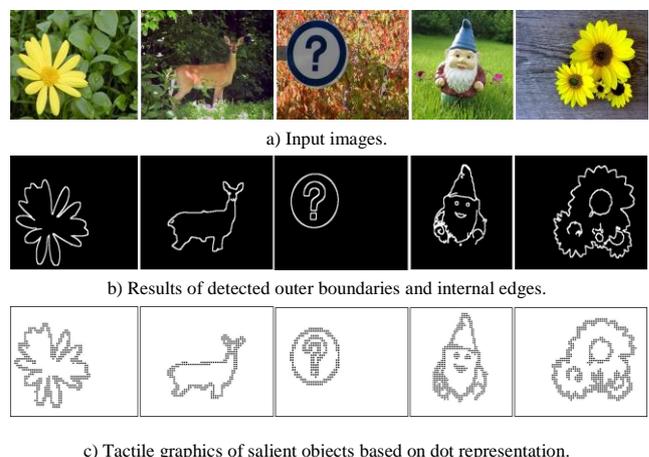


Figure 9. Generation of tactile graphics of salient objects.

## 5. Conclusions

We presented a saliency cuts method based on local adaptive thresholding using integral images. The proposed method is fully automatic and generates three-level thresholds to divide a saliency map into four regions. The seeds from the four regions are fed into the Grab Cut algorithm to obtain high-quality binary masks of salient objects. Based on the proposed method, we extracted salient objects in a full-color space. We then applied bilateral filtering and canny edge operator to produce outer boundaries and internal edges. Experimental results show that the proposed method is robust and performs better than other methods. The objective of this study was to extract salient objects from natural scene images, and finally to create tactile graphics based on dot representation for the visually impaired to better perceive and understand the natural scene images. We plan to conduct further research based on deep learning techniques for detecting and isolating multiple salient objects from natural scene images.

## Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) a grant funded by the Korea government Ministry of Science and ICT (MSIT). (No. 2019R1F1A1057757) and a grant from the Rand D Program of the Korea Railroad Research Institute, Republic of Korea.

## References

- [1] Achanta R., Shaji A., Smith K., Lucchi A., Fua P., and Süsstrunk S., "SLIC Superpixels Compared to State-of-the-Art Superpixel Methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274-2282, 2012.
- [2] AlSaeed D., Bouridane A., and El-Zaart A., "A Novel Fast Otsu Digital Image Segmentation Method," *The International Arab Journal of Information Technology*, vol. 13, no. 4, pp. 427-433, 2016.
- [3] Aytakin Ç., Ozan E., Kiranyaz S., and Gabbouj M., "Visual Saliency By Extended Quantum Cuts," in *Proceedings of IEEE International Conference on Image Processing*, Quebec City, pp. 1692-1696, 2015.
- [4] Benny D. and Soumya K., "New Local Adaptive Thresholding and Dynamic Self Organizing Feature Map Techniques for Handwritten Character Recognizer," in *Proceedings of International Conference on Power and Computing Technologies*, Nagercoil, pp. 1-4, 2015.
- [5] Bradley D. and Roth G., "Adaptive Thresholding Using the Integral Image," *Journal of Graphics Tools*, vol. 12, no. 2, pp. 13-21, 2007.
- [6] Chen J. and Takagi N., "A Pattern Recognition Method for Automating Tactile Graphics Translation from Hand Drawn Maps," in *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, Manchester, pp. 4173-4178, 2013.
- [7] Cheng M., Mitra N., Huang X., Torr P., and Hu S., "Global Contrast Based Salient Region Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 569-582, 2015.
- [8] Comaniciu D. and Meer P., "Mean Shift: A Robust Approach toward Feature Space Analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603-619, 2002.
- [9] Felzenszwalb P. and Huttenlocher D., "Efficient Graph-Based Image Segmentation," *International Journal of Computer Vision*, vol. 59, no. 2, pp. 167-181, 2004.
- [10] Fu Y., Cheng J., Li Z., and Lu H., "Saliency Cuts: an Automatic Approach to Object Segmentation," in *Proceedings of 19<sup>th</sup> International Conference on Pattern Recognition*, Tampa, pp. 1-4, 2018.
- [11] Jungil J., Hongchan Y., Hyelim L., and Jinsoo C., "Graphic Haptic Electronic Board-Based Education Assistive Technology System for Blind People," in *Proceedings of IEEE International Conference on Consumer Electronics*, Las Vegas, pp. 364-365, 2015.
- [12] Lee G., Tai Y., and Kim J., "Deep Saliency with Encoded Low Level Distance Map and High Level Features," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, pp. 660-668, 2016.
- [13] Li S., Ju R., Ren T., and Wu G., "Saliency Cuts Based on Adaptive Triple Thresholding," in *Proceedings of IEEE International Conference on Image Processing*, Quebec City, pp. 4609-4613, 2015.
- [14] Liu N. and Han J., "Dhsnet: Deep Hierarchical Saliency Network for Salient Object Detection," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, pp. 678-686, 2016.
- [15] Manipoonchelvi P. and Karuppiyah M., "Saliency Detection for Content Aware Computer Vision Applications," *The International Arab Journal of Information Technology*, vol. 14, no. 4, pp. 528-533, 2017.
- [16] Mehrani P. and Veksler O., "Saliency Segmentation Based on Learning and Graph Cut Refinement," in *Proceedings of British Machine Vision Conference*, Aberystwyth, pp. 1-12, 2010.
- [17] Mukhiddinov M., Salient Region Extraction based on Global Contrast Enhancement and Saliency Cut for Image Information Recognition

of the Visually Impaired, Thesis Master's, Gachon University, 2017.

- [18] Rother C., Kolmogorov V., and Blake A., "Grabcut: Interactive Foreground Extraction Using Iterated Graph Cuts," *ACM Transactions on Graphics*, vol. 23, no. 8, pp. 309-314, 2004.
- [19] Sauvola J. and Pietikinen M., "Adaptive Document Image Binarization," *Pattern Recognition*, vol. 33, no. 2, pp. 225-236, 2000.
- [20] Singh O., Sinam T., James O., and Taiyenjam R., "Local Contrast and Mean Thresholding in Image Binarization," *International Journal of Computer Applications*, vol. 51, no. 6, pp. 4-10, 2012.
- [21] Takagi N. and Chen J., "Character String Extraction From Scene Images By Eliminating Noncharacter Elements," in *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, San Diego, pp. 3685-3690, 2014.
- [22] Zhou Q., Chen J., Ren S., Zhou Y., Chen J., and Liu W., "On Contrast Combinations for Visual Saliency Detection," in *Proceedings of IEEE International Conference Image Processing*, Melbourne, Australia, pp. 2665-2669, 2013.



**Mukhriddin Mukhiddinov** received the B.S. degree in Informatics and Information Technologies from Tashkent University of Information Technologies (TUIT), Uzbekistan in 2015 and the M.S. degree in IT Convergence Engineering from Gachon University, Korea in 2017. In 2020, he received the Ph.D. degree in Computer Engineering from TUIT, Uzbekistan. He is currently assistant professor in the Department of Hardware and Software Control Systems in Telecommunication, TUIT, Uzbekistan. His research interest includes object extraction, image processing, and pattern recognition.



**Rag-Gyo Jeong** received the B.S., M.S., and Ph.D. degrees, all in Electrical Engineering, from Inha University, Incheon, Korea, in 1991, 1999, and 2005, respectively. He joined KRRI (Korea Railroad Research Institute), Uiwang, South Korea, as a Senior Researcher in 1995. Currently, he is a Principal Researcher and team leader of On-demand Transit Research Team in the New Transportation Systems Research Center at KRRI. From 1990 to 1994, he was a Staff Engineer at Hanjin Heavy Industries Co., Ltd. His research interests include autonomous train control system, electric-powered transportation systems, PRT (Personal Rapid Transit), system engineering.



**Jinsoo Cho** received his B.S. in Electronic Engineering from Inha University in 1994, M.S. in Electrical Engineering from Columbia University in 1998, and Ph.D. in Electrical and Computer Engineering from the Georgia Institute of Technology in 2003. From 2004 to 2006, he was a senior research engineer in the D-TV development team of System LSI Division, Samsung Electronics Co., Ltd in Korea. He is currently an associate professor in the Department of Computer Engineering, College of IT, Gachon University, Korea. His current research interests include image/video processing, computer vision, and assistive technology for the visually impaired.