applied sciences

MDPI

Article

Automatic Salient Object Extraction Based on Locally Adaptive Thresholding to Generate Tactile Graphics

Akmalbek Abdusalomov [1], Mukhriddin Mukhiddinov [2], Oybek Djuraev [2], Utkir Khamdamov [2] and Taeg Keun Whangbo [3,*]

1 Department of IT Convergence Engineering, Gachon University, Sujeong-Gu, Seongnam-Si, Gyeonggi-Do 461-701, Korea; akmalbekabdusalomov@gmail.com
2 Department of Hardware and Software of Management Systems in Telecommunications, Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Tashkent 100200, Uzbekistan; mmuhriddinm@gmail.com (M.M.); odjuraev@gmail.com (O.D.); utkir.hamdamov@mail.ru (U.K.)
3 Department of Computer Science, Gachon University, Sujeong-Gu, Seongnam-Si, Gyeonggi-Do 461-701, Korea
* Correspondence: tkwhangbo@gachon.ac.kr

Received: 10 March 2020; Accepted: 8 May 2020; Published: 12 May 2020

Abstract: Automatic extraction of salient regions is beneficial for various computer vision applications, such as image segmentation and object recognition. The salient visual information across images is very useful and plays a significant role for the visually impaired in identifying tactile information. In this paper, we introduce a novel saliency cuts method using local adaptive thresholding to obtain four regions from a given saliency map. First, we produced four regions for image segmentation using a saliency map as an input image and local adaptive thresholding. Second, the four regions were used to initialize an iterative version of the GrabCuts algorithm and to produce a robust and high-quality binary mask with a full resolution. Finally, salient objects' outer boundaries and inner edges were detected using the solution from our previous research. Experimental results showed that local adaptive thresholding using integral images can produce a more robust binary mask compared to the results from previous works that make use of global thresholding techniques for salient object segmentation. The proposed method can extract salient objects with a low-quality saliency map, achieving a promising performance compared to existing methods. The proposed method has advantages in extracting salient objects and generating simple, important edges from natural scene images efficiently for delivering visually salient information to the visually impaired.

Keywords: integral images; local adaptive thresholding; salient object extraction; saliency map; saliency cuts; visually impaired; tactile graphics

1. Introduction

Human beings have an incredible ability to visually capture relevant targets quickly and accurately, which is called the focus of attention or saliency. Recognizing these conspicuous, or salient, ranges in visual fields empowers one to apportion restricted perceptual resources in a proficient way [1]. This visual saliency mechanism makes some objects in a scene stand out from the surroundings, a focus that raises plenty of research interest. In general, accurately detecting the most visually significant foreground object in a scene, referred to as salient object detection, may be goal-driven or stimulus-driven, corresponding to the top-down or bottom-up process in human visual perception, respectively. Bottom-up approaches are based on low-level image attributes such as color, intensities, gradient, edges, and boundaries, and are usually fast, involuntary, and more effective in detecting fine details rather than the information pertaining to the overall shape. In contrast, top-down saliency

models are based on representative features from samples while being slow, task-driven, voluntary, and closed-loop. Recently, visual attention modeling has been extended to highlight objects uniformly and widely used as a first step in the computer vision field for its applications in image and video compression [2], image retargeting [3–5], image retrieval [6,7], object-of-interest image segmentation [8], object recognition [9–11], image classification [12], and advertising evaluation [13]. The recent success of deep learning in object recognition and classification has brought about a revolution in computer vision [14]. Although significant progress has been made in the research area of visual attention, it remains not only one of the most important fields but also a highly challenging issue in image analysis, pattern recognition, and computer vision [15].

Salient object detection mainly consists of two phases: saliency map generation and saliency cuts. Saliency map generation concentrates on producing a pixel-level or region-level map, in which each pixel or region is assigned a value proportional to its saliency. Saliency cuts focus on providing a binary mask of a salient object. There have been a variety of saliency cuts methods proposed [16,17], and these methods use global thresholding techniques for calculating a threshold value by disintegrating the pixels in a saliency map into the foreground and background regions according to their saliency values. The results of these methods are usually undermined by incorrect saliency maps as well as low saliency values for determining background regions in complex natural scene images. To increase the accuracy of a saliency map, some approaches use the original input image along with the saliency map in the binarization process [16–18]. Extracting salient objects from a scene and implementing tactile representation can assist the visually impaired to clearly understand important information contained in images. In general, the tactile representation of salient objects in a natural scene image should be designed as simple as possible and easily understood by the visually impaired. Moreover, a wide range of studies are being conducted to find optimal methods for extracting important objects from natural scene images [15]. Generally, tactile graphic information that visually impaired people can perceive by touching, is mainly obtained from the natural scene images through two steps: salient object extraction and tactile graphic translation. We focused on the automatic object segmentation approach for integrating local adaptive thresholding using integral images and the GrabCuts algorithm.

In this paper, on the basis of the fact that local adaptive thresholding estimates a different threshold value for each pixel according to the grayscale information of neighboring pixels [19], we proposed a novel saliency cuts method based on locally adaptive triple-thresholding using integral images, in order to assist the image recognition process of the visually impaired. Firstly, we obtained a saliency map using our previous saliency detection method [15]. Secondly, the binary saliency mask was obtained as the image of a four-region seeds by using adaptively triple-thresholding. Then, the seeds were fed to the GrabCuts [20] method and a robust binary mask was generated.

The main contribution of the proposed method was as follows:

- Salient object extraction based on locally adaptive triple thresholding using an integral image.
- We combined the GrabCuts algorithm with the generated four-region seeds to refine the segmentation results.
- One of the applications of salient object extraction is detected outer boundary, and inner edges of the salient object were illustrated on tactile graphics to facilitate the learning process of visually impaired.

The proposed method cuts the salient object and detects the outer contours and inner edges of salient objects so that the visually impaired can easily understand the content of an image scene. Figure 1 demonstrates the proposed method: (a) input image, (b) saliency map generated using our previous approach [15], (c) the proposed saliency cuts, and (d) salient objects extracted using the mask.

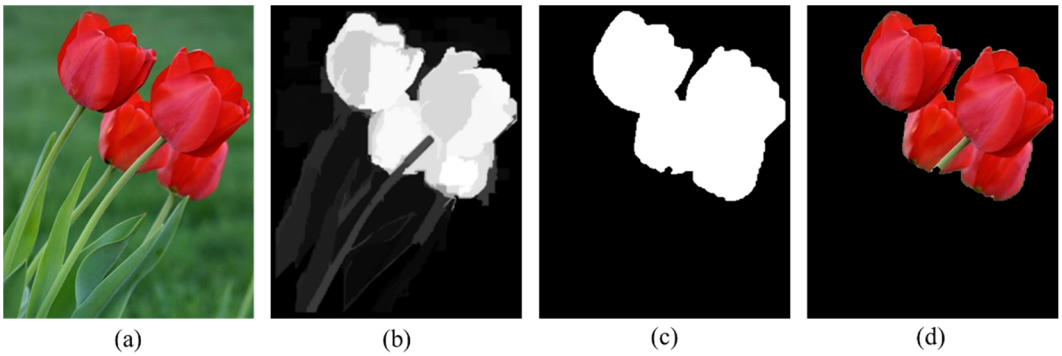

Figure 1. Example of the proposed method: (**a**) input image, (**b**) saliency map, (**c**) saliency cuts, (**d**) Salient objects.

The remainder of the paper is organized as follows: Section 2 reviews conventional methods for salient object detection and saliency cuts. Section 3 presents the saliency cuts approach in detail. The experimental results based on popular benchmark databases are shown and discussed in Section 4. Section 5 highlights some limitations of the proposed method. Finally, Section 6 concludes the paper by summarizing our findings.

2. Related Works

According to the objective and the technical component of this paper, we broadly reviewed previous academic work related to several research areas: salient region detection, saliency cuts, and tactile graphic generation.

Salient region detection is further extended to include objectless estimation in object recognition. Both are important and benefit different applications in high-level scene analysis [21]. Saliency models based on bottom-up methods convert natural scene images into saliency maps, where each pixel/superpixel or region is assigned a saliency value or probability. These methods initially apply image segmentation techniques (e.g., graph-based [22], mean shift [23], or superpixels [24]) to the input image and segment homogeneous regions to extract feature statistics from each segmented region to detect salient regions. Cheng et al. [17] demonstrated two saliency models: histogram-based contrast (HC), which assigns a pixel-wise saliency value, as well as region-based contrast (RC), which incorporates spatial relations at the cost of reduced computational efficiency. Zhou [25] proposed an object-based attention model that automatically identifies a series of regions far away from the image center as background prototypes. Zhou et al. [26] combined widely used contrast measurements, namely, center-surround, corner-surround, and global contrast, to detect visual saliency. Han et al. [27] developed a framework for saliency detection by first modeling the background and then separating salient objects from the background. Color contrast and spatial distribution were used to obtain pixel-accurate saliency maps.

Another popular approach is saliency object detection methods based on deep convolutional neural networks (CNNs). Recently employed, these methods achieve substantially better results than traditional approaches because CNNs are typically pre-trained on datasets for visual recognition tasks. Gayoung Lee et al. [28] introduced 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 in evaluating similarities between different parts of an image with precise boundaries among superpixels. Guanbin Li et al. [29] proposed a neural network architecture, which has fully connected layers on top of CNNs responsible for feature extraction at three different scales. In addition, they created a novel and challenging dataset, HKU-IS, for saliency model research and evaluation. Nian Liu et al. [30] demonstrated an end-to-end saliency detection model, the DNSNet, to detect salient objects with a new hierarchical refinement model, the HRCNN, which can refine saliency maps hierarchically and progressively to recover image details

by integrating local context information without using over-segmentation methods. Yuan et al. [31] proposed a framework that performs both dense and sparse labeling (DSL) with multi-dimensional features for saliency detection. DSL consists of three major steps: dense labeling (DL), sparse labeling (SL), and deep convolutional (DC) networking. Li et al. [32] proposed a deep network, which consists of a fully convolutional stream at the pixel level and a spatial pooling stream at the segment level. A multi-scale fully convolutional network (MS-FCN) takes a raw image as input then directly produces a saliency map with pixel-level accuracy.

In recent years, approaches using saliency cuts have been widely explored because 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 GrabCuts algorithm [20], we can extract a precise image mask [16–18,33] Chul Ko et al. [34] proposed the object-of-interest (OOI) segmentation algorithm from natural scenes. They use a support vector machine (SVM) to select a salient region clustered into the OOI using a region merging technique. Jiang et al. [35] introduced an automatic salient object segmentation method, which integrates both bottom-up salient stimuli and object-level shape prior. Fu et al. [18] modified the graph cut method by exploring the effects of labels for graph-based segmentation. Aytekin et al. [36] proposed 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. Winn et al. [37] introduced Learning Object Classes with Unsupervised Segmentation (LOCUS), which uses a generative probabilistic model to combine bottom-up cues of colors and edges with top-down cues of shape. Shi et al. [38] introduced image segmentation as an issue associated with graph partitioning and extracting the global impression of an image, rather than focusing on local features. Peng et al. [39] demonstrated a saliency-aware stereo image segmentation approach using the disparity map and statistical information of stereo images to enrich high-order potentials. Grady et al. [40] treated an image as a purely discrete object, and each edge was assigned a real-valued weight corresponding to the likelihood that a random walker will cross that edge. Chew et al. [41] improved normalized cuts to allow both sets of constraints to be handled in a soft manner, enabling the user to tune the degree to which the constraints are satisfied. Mehrani et al. [33] exploited the standard features often used in vision-based factors such as color and texture. Properly normalized, these simple features yielded performance superior to the methods based on hand-crafted features specifically designed for saliency detection. Han et al. [42] developed a generic framework to automatically extract the viewer's attention upon objects, on the basis of human visual attention mechanisms. Without the full semantic understanding of image content, the model formulated attention objects as a Markov random field (MRF) by integrating computational visual attention mechanisms with attention object growing techniques. Li et al. [16] proposed a saliency cuts approach using the Otsu thresholding technique and the GrabCuts algorithm. They improved the Otsu algorithm to calculate three-level thresholds, and the saliency map was further split into four regions using these three thresholds.

We also surveyed relevant literature on global and local adaptive thresholding. Otsu et al. [43] proposed an unsupervised method of automatic thresholding, which is one of the most popular global techniques. In contrast, local adaptive thresholding used for binarization can account for variations in illumination. Wellner [44] introduced quick adaptive thresholding, which calculates the moving average of the last s pixels. This adaptive local thresholding technique simply compares every pixel to the average of neighboring pixels. Sauvola et al. [45] demonstrated adaptive document image thresholding, in which a page is considered a collection of subcomponents such as text, background, and picture. Bradley et al. [46] proposed real-time adaptive thresholding using an integral image of the input image. Peuwnuan et al. [47] proposed an improved version of Bradley's method by means of adaptive thresholding using integral images. Another extension of Bradley's algorithm, a new local adaptive thresholding technique, was proposed by Benny et al. [48]. They proposed a handwritten character recognition system using a local thresholding method for binarization and a

dynamic self-organizing feature map (DSOFM) technique for classification of extracted feature vectors of characters. Biswas et al. [49] used a local thresholding technique to binarize degraded document images. First, they blurred an input image using the Gaussian filter and computed a global threshold value using the Canny edge, and each pixel with a gray value greater than the threshold was labeled as a background pixel. Second, local threshold values classified non-labeled pixels into the background and foreground pixels. Singh et al. [19] proposed a local thresholding technique using a local contrast and a local arithmetic mean for grayscale images.

Visual information generation based on tactile graphics is one of the most ambitious research areas because these techniques require important information from an input image. Chen et al. [50] presented a method of mathematical graph recognition for extracting and classifying broken line elements from mathematical graphs. This method is a specific part of the mathematical graph recognition technique introduced for developing a computer-aided system to extract and classify broken lines. Jungil et al. [51] focused on an education assistive technology system based on a graphic haptic electronic board. This system enables authoring, automatic conversion, and real-time distribution of education materials for low-vision and blind students. Chen et al. [52] suggested 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 et al. [53] introduced a method for extracting character strings from scene images using edge detection, a morphology operator, and a fuzzy inference technique. This framework helps the visually impaired to walk more independently on the street.

The methods discussed above, such as salient object detection and extraction, salient region detection for the visually impaired, and tactile graphics generation, have some limitations. Figure 2 shows our scientific methodology to clarify the purpose of research work and its results. Initially, we broadly analyzed methods of digital image processing areas such as salient object detection, salient object extraction, contour detection, and tactile graphics generation. Then, on the basis of the results of the analysis, salient object extraction using adaptive triple thresholding was proposed, and edge detection technique was used. Finally, computer vision tools and libraries were used to develop salient object detection, extraction, and contour detection software. This developed software can be utilized in assistive technologies and software packages for visually impaired individuals, such as tactile graphics and displays. Generally, contrast-based salient object detection methods can be local and global contrast-based techniques. The local contrast-based approaches detect a salient region using local neighborhoods of the pixels. These approaches suffer from local noises when calculating complex pattern images. The principle system of the global-contrast approach processes the object's saliency by the calculation of the color contrast between every one of the pixels and the mean estimation of a whole picture. Despite the fact that the global-contrast approach is successful in identifying salient areas of straightforward pattern pictures, these models have a restriction in a poor global-contrast and a complex pattern picture. To overcome these limitations, we applied the global contrast enhancement method using histogram equalization to input images [15].

Salient object extraction approaches use an input image and saliency map to provide a binary mask of salient objects. These approaches apply threshold value to divide images into the foreground or background. The problem of this method is to compute the best threshold values that can accurately determine foreground and background regions in complex pattern images. We propose locally adaptive triple thresholding using integral images by calculating three-level thresholds instead of only one level along with other improvements.

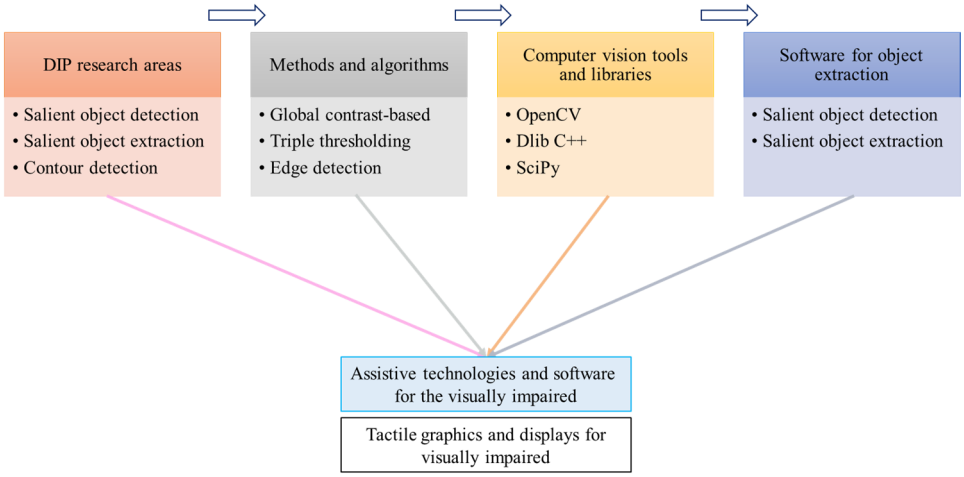

Figure 2. The methodology of research work.

3. Proposed Method

This section describes the proposed method for salient object extraction. Firstly, we provide an overview of the proposed method. Then, saliency cuts using the local adaptive thresholding method is detailed, which is used to produce a binary image and boundary of salient objects. Lastly, outer and inner edge detection, as well as the generation of tactile graphics process is discussed. We produced tactile graphics of salient objects by applying the suggested approach and the assistive technology software to help the visually impaired to better perceive and identify natural scene photographs, as presented in the flowchart of the proposed method.

3.1. Overview

In this sub-section, we give an overview of the proposed method. Figure 3 illustrates the main stages of the proposed saliency cuts. We first generated a saliency map from an input natural scene image, which we used in our previous work [15], as shown in the first and second parts of Figure 3. Next, as seen in the third part of Figure 3, we calculated an integral image in the first move through the saliency map grayscale image. In a second move, we identified the local window ($S \times S$), which was the image's width $W/2$, using an integral image for every pixel in constant time, and further obtained local adaptive three-level threshold values on the basis of the comparison. These three thresholds marked the saliency map with four different regions, demonstrating four types of seeds for the next step. The seeds were assigned an iterative variant of the GrabCuts method, alluded to as Saliency cuts, and a high-quality binary mask of a full resolution was produced. We were able to improve local adaptive thresholding using the integral images method by calculating three threshold levels instead of one [46]. The purpose of the improvement was to use local adaptive thresholding for computing three threshold levels and classifying each pixel into one of four categories: certain background, probable background, probable foreground, and certain foreground. Finally, we detected the boundary and the inner edges of salient objects to refine the recognition of visually salient information for the visually impaired and converted it to tactile graphics, as shown in the last part of Figure 3. We describe the proposed saliency cuts approach in the following sub-sections.

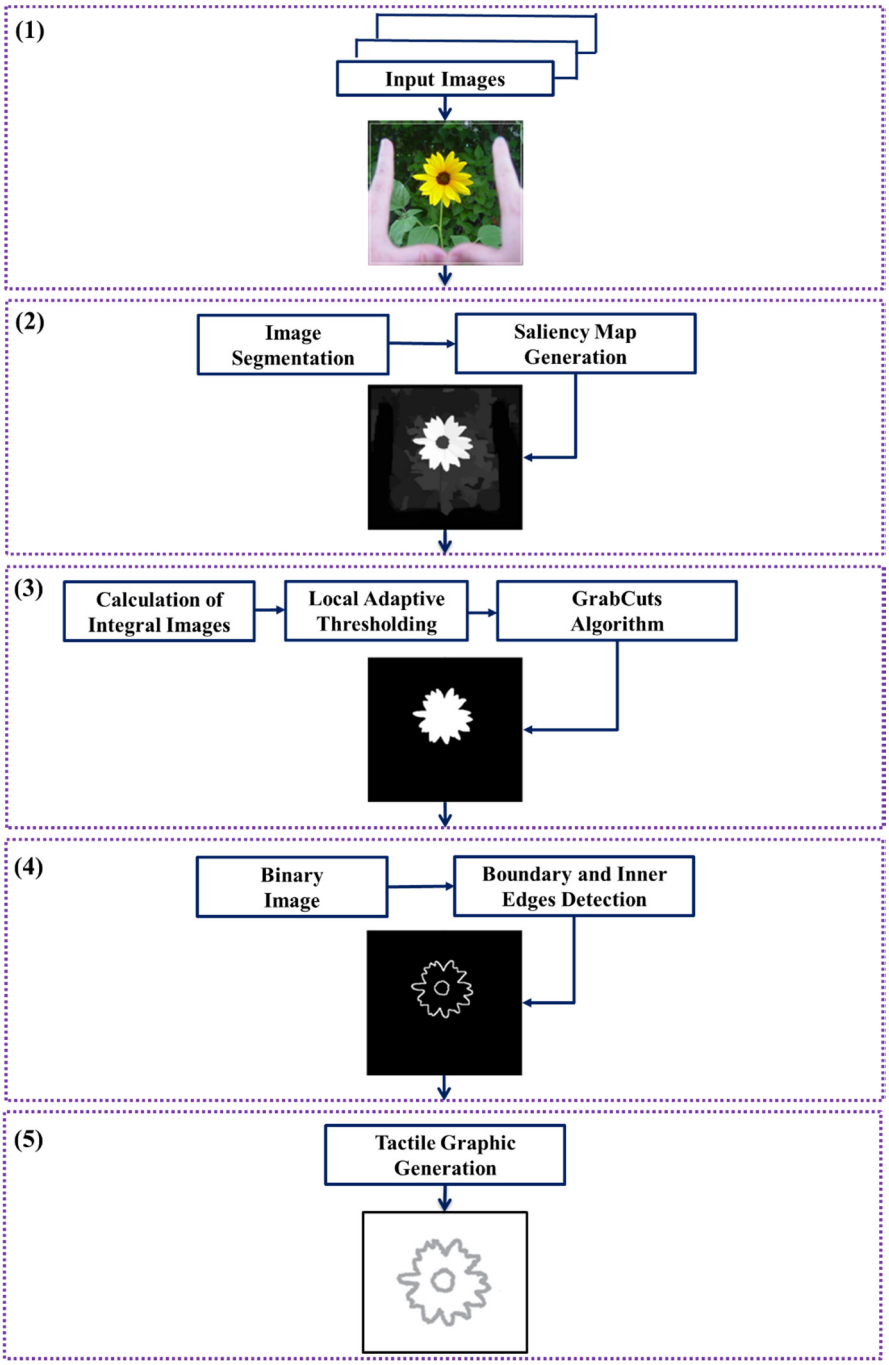

Figure 3. A flowchart of the proposed method: (1) Input natural scene image, (2) saliency map generation using our previous approach, (3) saliency cuts using local adaptive thresholding, (4) outer boundary and inner edge detection, (5) tactile graphics for visually impaired.

3.2. Saliency Cuts Using Local Adaptive Thresholding

This sub-section details salient object extraction using local adaptive thresholding. First, we provide a brief description of choosing the local thresholding method over a global one. Then, an integral image is calculated using the given saliency map image. Third, local adaptive triple thresholds are obtained, and afterwards these thresholds separate the saliency map image into four regions. Last, we acquire a binary image using the GrabCuts algorithm and detect a boundary of extracted objects.

3.2.1. Local Adaptive and Global Thresholding

A number of thresholding techniques have been proposed using global and local techniques. Global techniques assign one threshold to the entire image, whereas local adaptive thresholding techniques assign a varying threshold value to each pixel/region determined by neighboring pixels.

Global thresholding techniques are very fast and yield reliable results for typical images. Many popular global thresholding techniques use different approaches, such as the fixed thresholding technique, to perform binarization with respect to a specified threshold value. Cheng et al. [17] used global fixed thresholding to binarize saliency map images. This approach works well when the background and foreground intensities are clearly distinct and uniform throughout the image. Li et al. [16] proposed a saliency map binarization technique using the Otsu algorithm to calculate adaptive triple thresholding. The Otsu algorithm is an outstanding automatic thresholding algorithm for image segmentation, thanks to its simplicity and good performance. On the basis of clustering, the Otsu algorithm automatically obtains an optimal threshold from the gray histogram of an original gray image and separates it into the foreground and background. For many years, binarization of natural images has been performed on the basis of global thresholding approaches, which are more appropriate for uniformly illuminated images than diversely illuminated images. 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 different information contained in the given image. This is also known as dynamic thresholding and can be divided into different approaches, such as the water flow model, background subtraction, mean and standard deviation of pixel values, illumination model, and local image contrast. Although local adaptive thresholding approaches generally achieve better results, they often rely on individual parameters and require a substantially higher computational cost than global thresholding. In this paper, to reduce the computational cost, we used adaptive thresholding based on integral images, which is explained in the next sub-section. Wellner [44] proposed fast-adaptive thresholding, which calculates the moving average of the pixels last seen to be the local threshold. D. Bradley simply extended Wellner's method by using integral images to provide a better representation of the surrounding pixels than the moving average by sacrificing one additional iteration through an image [46]. This method is clear and straightforward. Both Derek's and Wellner's methods focus on the application of document images. These methods can also be applied to natural scene images with some improvements.

3.2.2. Integral Images

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

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

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). We computed integral images and locally adaptive triple-thresholding as shown in Figures A1 and A2 (Appendix A section). Figure 4 shows an example of a 4×3 image that demonstrates the calculation process of the integral image. As shown in Figure 5, $I(2, 3)$ is the sum of the intensity values of the saliency maps in the upper-left area. The integral image or summed-area table $I(x, y)$ can be quickly calculated using Equation (2) [46].

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

where $i(x,y)$ is the original pixel value from the image, $I(x-1,y)$ values are directly above this pixel, and $I(x,y-1)$ values are directly left to this pixel from the Integral image. Finally, we subtracted the value directly top-left of $i(x,y)$ from the Integral image, that is, $I(x-1,y-1)$. We gave a value of 0 to $I(x-1,y)$, $I(x,y-1)$, and $I(x-1,y-1)$ if $x-1$ and $y-1$ were outside of the image bounds.

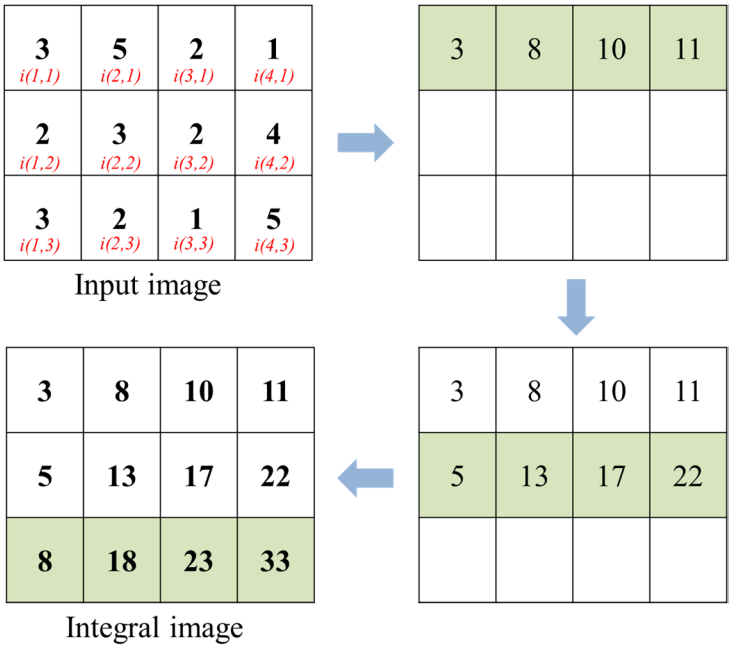

Figure 4. Steps for calculating the Integral image.

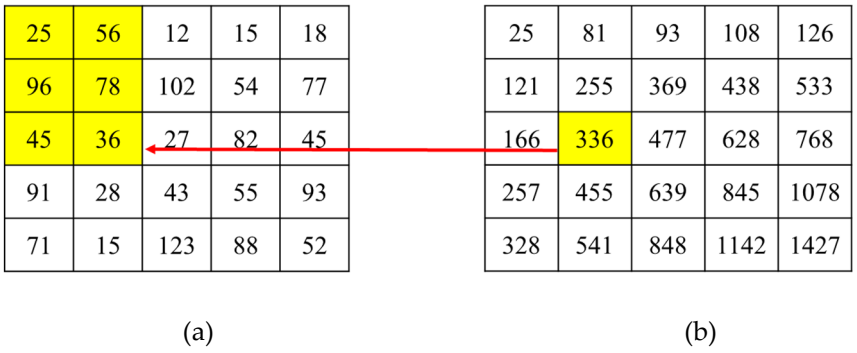

Figure 5. Example of integral image calculation: (**a**) grayscale image, (**b**) integral image.

3.2.3. Local Adaptive Triple Thresholding

This step involves a fast approach for computing local thresholds without compromising the performance of local thresholding using the integral sum image as a prior process for finding the local mean of neighboring pixels in a window, irrespective of window size. Using this technique, we can accomplish a binarization speed close to those of global binarization methods. Next, we categorize each pixel as follows:

(1) Sum of pixel values R_s over a rectangle R within a moving window with a size $S \times S$ was defined so that the window size depended on the image's width W. The window size is very important in local adaptive thresholding methods. D. Bradley and Wellner used a window size one-eighth of the image width to calculate the average value. We used one-half of the image width for the window size

S due to the saliency object size in natural scene images. As shown in Figure 6, shadow area R is on the grayscale image.

$$R_S = I(x_2, y_2) - I(x_2, y_1 - 1) - I(x_1 - 1, y_2) + I(x_1 - 1, y_1 - 1) \tag{3}$$

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

(2) The number of pixels in R can be calculated as

$$N = (x_2 - x_1) \times (y_2 - y_1) \tag{4}$$

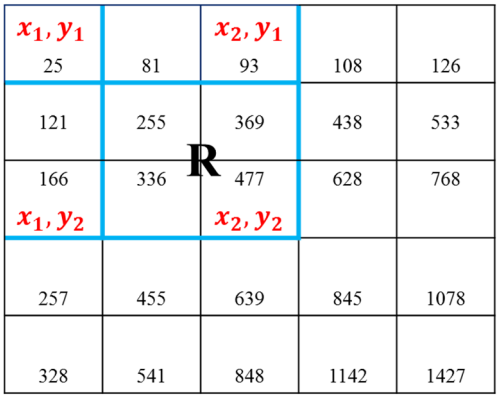

Figure 6. Sum over rectangle R in grayscale image.

Wellner and D. Bradley calculated the optimal threshold value on the basis of comparison [44,46]—if the value of the current pixel was 15 percent less than the average, it was set to black; otherwise, it was set to white for document images. In the proposed method, we improved this threshold value by assigning it to the local arithmetic mean value using Equations (5) and (6). Because a 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 yielded the best results for a variety of images.

Let us assume that the scale of a saliency map is [0; L], and that the pixel number of each bin in its histogram is $B = \{b_0, \ldots b_i, \ldots b_L\}$, where i represents the saliency value. We can improve the Derek Bradley algorithm to obtain three-level thresholds, including t_l, t_m, and t_h $0 < t_l < t_l < t_l < L$, and decompose the histogram of a saliency map into four parts: certain background T_{cb} = [0; t_l], probable background T_{pb}= [t_l + 1; t_m], probable foreground T_{pf} = [t_m + 1; t_h], and certain foreground T_{cf} = [t_h + 1; L]. We first calculated t_l, then divided the saliency histogram into T_b= [0, t_l] (certain background) and T_u = [t_m +1, L] (unknown). The unknown regions were initially used to train foreground color models, thus helping the algorithm to identify foreground pixels. We then calculated t_m and t_h, and further divided the saliency histogram into T_{cb}, T_{pb}, T_{pf}, and T_{cf}. The numbers of pixels with saliency values in T_b and T_u were p_b and p_u, respectively, and the number of pixels in the entire image was $P = p_b + p_u$. We could then calculate threshold t_l by averaging the pixels within the window of size $S \times S$.

(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 could be calculated as

$$m(x, y) = R_s/N \tag{5}$$

$$t_l = m(x, y) \tag{6}$$

(4) Finally, t_m and t_h could be calculated as follows:

$$t_h = t_l + (L - t_l)/2 \tag{7}$$

where L is maximum pixel value in the image and

$$t_m = t_l + (t_h - t_l)/2 \tag{8}$$

Given a saliency map, initially found t_l maximized the difference between the object region and the background region. We then performed the calculation for t_m and t_h. According to the triple thresholding saliency histogram, a saliency map could also be decomposed into four regions: certain background (pixels belong to T_{cb}), probable background (pixels belong to T_{pb}), probable foreground (pixels belong to T_{pf}), and certain foreground (pixels belong to T_{cf}). Certain background regions were retained, whereas other regions may have been changed during GrabCuts optimization.

3.2.4. GrabCuts with Auto-Generated Seeds

For image binarization, we used a four-region seeds image. With adaptive three-level threshold values, we already acquired the seeds of four kinds for the masking requirement of the GrabCuts method [20]. For an image pixel value greater than t_m in Equation (8), the largest connected region was considered the initial candidate region of the most dominant salient object. This candidate region was marked as probable foreground and certain foreground, whereas other regions were marked as probable background and certain background.

We could obtain the contour of the salient object using the binary mask. Figure 7 shows that the boundary of a salient object was detected using binary images and Canny's edge detection technique [54]. These outer boundaries can translate into tactile graphics to provide information about a natural scene image to the visually impaired. In some cases, however, the visually impaired may not be confident about an object just by touching its boundary. Therefore, we also demonstrated a technique for detecting the inner edges of an object from the results of the proposed saliency cuts in natural scene images using our previous technique [15].

Figure 7. Results of boundary detection: (**a**) input image, (**b**) saliency map, (**c**) the proposed saliency cuts approach, (**d**) outer boundary of a salient object.

3.2.5. Boundary and Inner Edge Detection

It is very important that visually impaired individuals fully identify a salient object in a natural scene. Therefore, we obtained the internal edges of a salient object using the binary mask, which was the result of the proposed saliency cuts approach. In the first step of post-processing, we produced a salient object using its binary mask by creating a matrix with a size and type identical to those of the input image to achieve the desired output image. After that, we copied the non-zero elements of the binary mask that indicated the elements of the original input image matrix.

$$S_0 = B_m(x, y) * I_i(x, y) \tag{9}$$

where S_0 is the salient object, $B_m(x, y)$ is the binary mask, and $I_i(x, y)$ is the input image. Thus, we procured a full-color space salient object. Figure 8 shows an example of the masking method using the proposed binary mask.

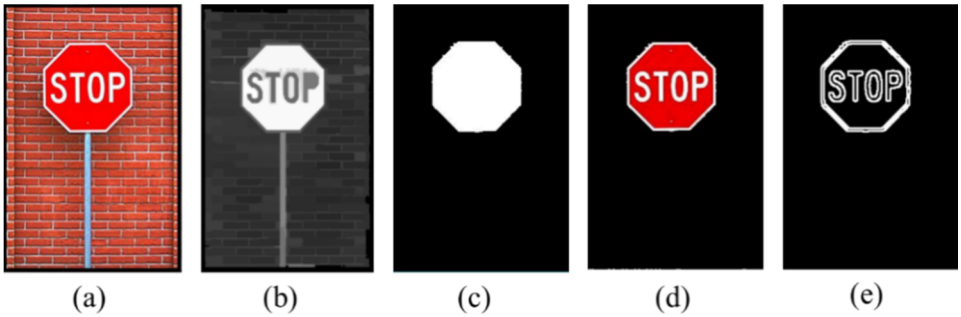

Figure 8. Example of using the masking method: (**a**) input image, (**b**) saliency map, (**c**) the proposed saliency cuts, (**d**) full-color space salient objects, (**e**) outer and inner edges.

In the end, we were able to produce the boundary and the inner edges of a salient object with valuable visual information for the visually impaired to identify the content of a natural scene. Furthermore, we could translate these contours and key visual information into tactile graphics for assistive technology systems.

4. Experiment Results and Analysis

In order to evaluate the performance of the proposed saliency cuts method, we conducted experiments using the C++ language and a PC with 3.60 GHz CPU and 4 GB of RAM. We used an MSRA 10k dataset [17], which included 10,000 natural scene images and human manually labelled ground truth, which is exact and full salient object(s) in the given image. To the best of our knowledge, the database is the largest of its kind, and has pixel-level ground truth in the form of accurate human-marked labels for salient regions. We selected 6000 natural scene images with its corresponding ground truth images that have various types of single and multiple objects from MSRA 10K dataset for efficient evaluation. We performed qualitative and quantitative comparisons between the proposed method and other techniques, and also subjectively evaluated the proposed algorithm.

4.1. Qualitative Evaluation

We visually compared the proposed saliency cuts method with two other state-of-the-art saliency cuts methods: adaptive triple thresholding for saliency cuts using Otsu automatic thresholding (ATT_Cuts) [16] and saliency cuts using fixed thresholding (RC_Cuts) [17]. In Figure 9, the first row displays typical input images such as people and objects. Ground truth images are shown in the second row. The third row shows the saliency maps of the natural scene images of the first row. The fourth and fifth rows are saliency cuts using the Otsu algorithm and global fixed thresholding, respectively. The last row shows the results of the proposed saliency cuts method. As shown in Figure 9, the results of the proposed method for saliency map binarization were close to ground truth images. The proposed saliency cuts technique extracted salient objects even when the background and foreground regions had very similar information. In comparison, the fixed thresholding method can extract only a single salient object from a given image and may fail to segment salient objects when pixels in the image had low saliency values. Saliency cuts using the Otsu method 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. Comparatively speaking, the proposed approach can reduce these drawbacks; it can extract salient objects with a lower amount of data about the saliency map without misclassifying them as background regions, as well as even being able to detect multiple salient objects.

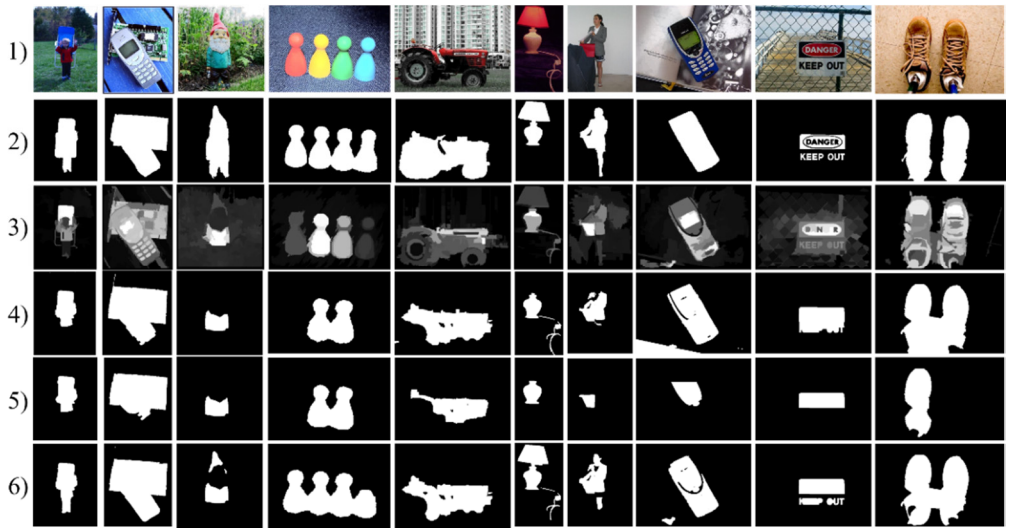

Figure 9. Visual comparison between saliency cuts methods. From top to bottom: (1) input images, (2) ground truth, (3) saliency maps, (4) ATT_Cuts [16], (5) RC_Cuts [17], (6) the proposed method.

The results of boundary detection together with a comparison between the proposed method and other well-known methods are shown in Figure 10. The first row displays given images, such as people, birds, and different kinds of objects. The second and third rows show the saliency maps and ground truth generated from the images in the first row, respectively. The other saliency cuts methods such as RC_Cuts and ATT_Cuts are illustrated in the fourth and fifth rows. The results of saliency cuts using the proposed method are shown in the sixth row, and the last row displays the boundaries of salient objects. Experimental results show that in many cases the proposed method extracted objects more accurately than other methods. Moreover, our approach worked effectively, even when there were multiple objects in natural scene images, as illustrated in Figure 10.

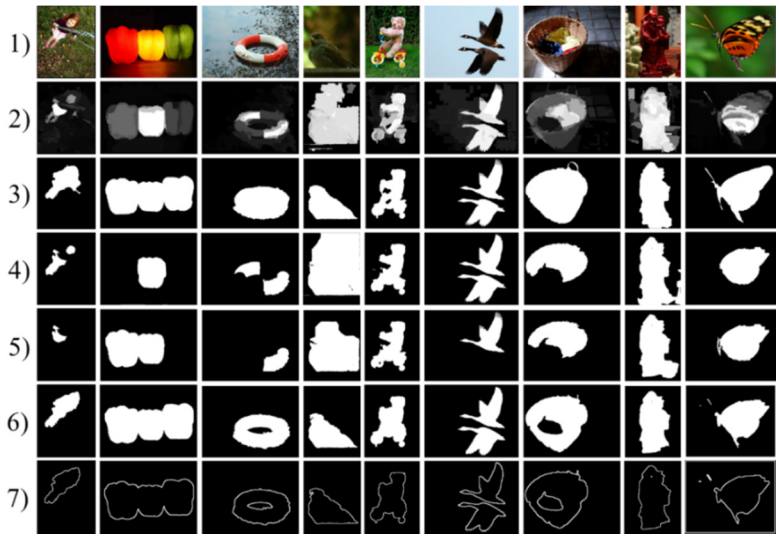

Figure 10. Experimental results of outer contour generation. From top to bottom rows: (1) input images, (2) saliency maps (grayscale images), (3) ground truth images, (4) ATT_Cuts results, (5) RC_Cuts results, (6) results of the proposed method, (7) boundary detection using the Canny edge technique.

However, the proposed method may fail to segment a salient object that has a very low saliency value and contains very similar foreground and background regions, as shown in Figure 11.

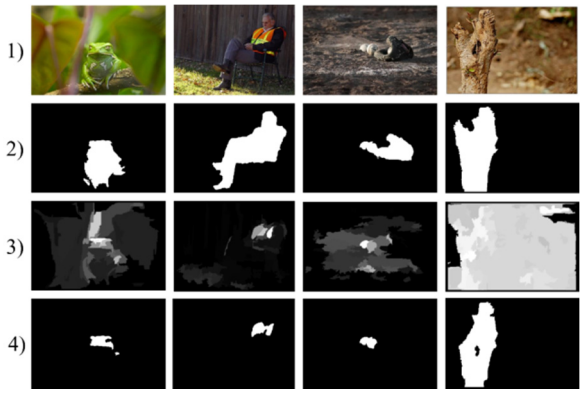

Figure 11. Failure cases: (1) input images, (2) ground truth, (3) saliency map, (4) saliency object segmentation using the proposed method.

4.2. Quantitative Evaluation

We performed quantitative analysis by averaging precision (P) and recall (R) rates along with F-measures (F_β). Precision and recall rates could be obtained by comparing pixel-level ground truth images with the results of the proposed method and calculated as follows:

$$P = \frac{T_p}{T_p + F_p} \tag{10}$$

$$R = \frac{T_p}{T_p + F_n} \tag{11}$$

where T_p denotes the number of true positives pixels that were obtained as salient by the proposed method and also labelled as salient in the ground truth image, F_p denotes the number of false-positive pixels that were obtained as salient by the proposed method but were not labelled as salient in the ground truth image, and F_n denotes the number of false-negative pixels that were not obtained as salient by the proposed method but were labelled as salient in the ground truth image. P is defined as the number of true-positive pixels over the number of true-positive pixels plus the number of false-positive pixels. R is defined as the number of true-positive pixels over the number of true-positive pixels plus the number of false-negative pixels. In other words, P is the fraction of salient pixels among all the obtained pixels by the proposed method, whereas R is the fraction of the total number of salient pixels that were actually obtained as salient by the proposed method. Worth noting are perfect precision (no false positive pixels) and perfect recall (no false negative pixels), as seen in Equations (10) and (11). The results showed that the proposed saliency cuts accomplished the highest precision rate of 0.94. RC_Cuts [17] and ATT_Cuts [16] yielded precision rates of 0.93 and 0.90, respectively. In addition, we computed the F-measure value, which balanced measurements between the mean of precision and recall rates. A higher F-measure meant higher performance and it was defined as follows:

$$F_\beta = \frac{(1 + \beta^2) P \times R}{\beta^2 \times P + R}. \tag{12}$$

where $\beta^2 = 0.3$ to weigh precision more than recall as in most of the existing methods [16,17,30–32], and others. The proposed approach achieved the highest F-measure value of 0.90, whereas the other two methods RC_Cuts and ATT_Cuts showed 0.89 and 0.88, recpectively. The results of quantitative comparison among the three algorithms are shown in Figure 12 and Table 1. The proposed method produced saliency objects with the highest pricision rate, whereas ATT_Cuts showed the highest recall rate. Although all three methods produced similar results, the proposed method proved to be most robust in complex images that contain foreground and background with similar appearances, as demonstrated in Figures 9 and 10.

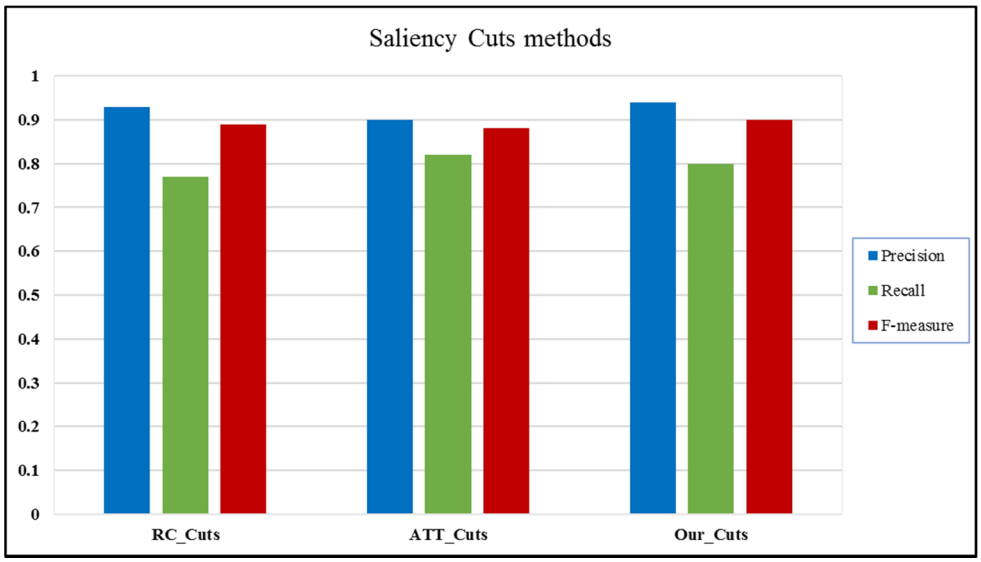

Figure 12. Quantitative comparisons of saliency cuts approach using MSRA 10k dataset.

Table 1. Quantitative analysis of three saliency cuts methods performed using MSRA 10k dataset.

Methods	RC_Cuts [17]	ATT_Cuts [16]	Proposed Method
Precision	0.93	0.90	**0.94**
Recall	0.77	**0.82**	0.80
F-measure	0.89	0.88	**0.90**

4.3. Subjective Evaluation

We conducted this experiment with 10 subjects (8 males, 2 females) with normal vision without visual disabilities. The goal of this subjective evaluation was to compare the visual inspection of the results of the three saliency cuts methods because every person has a unique visual system. It is a very important experiment in terms of generating tactile graphics for visually impaired because tactile graphics must be simple and understandable. We provided 100 various types of images that had single and multiple objects to 10 subjects. Each image file consisted of the result of the three saliency cuts methods and ground truth image. The subjects were asked to assess the results of the three saliency cuts methods with respect to ground truth images by providing scores on "foreground extraction accuracy", "clear edges of objects", "multiple object detection", "amount of noise introduced", and "false object detection". The subjects scored each item on a scale from 1 to 5—1 (very bad), 2 (bad), 3 (average), 4 (good), and 5 (very good) in 15 s overall, with 5 s per each of the saliency cuts methods. We informed the subjects that the answer to the "amount of noise introduced" and "false object detection" should be "very good" for only the best results. Furthermore, we prevented any help or discussion with each other by experimenting with the individuals one by one. As the results of subjective evaluation, Table 2 shows that foreground extraction accuracy of the proposed approach was very high. This was because locally adaptive triple-thresholding produced a reliable binary mask by assigning a different threshold value to each pixel. The proposed method and ATT_Cuts were able to detect multiple objects in a scene, whereas RC_Cuts failed to do so. In summary, the proposed method achieved the best overall results.

Table 2. Subjective evaluation of the proposed and other methods.

Methods	Foreground Extraction Accuracy	Clear Edge of Objects	Multiple Objects Detection	Noise Introduced	False Object Detection	Average
RC_Cuts	4	5	3	3	3	3.6
ATT_Cuts	4	5	4	4	4	4.2
Proposed method	5	5	5	4	4	4.6

4.4. Runtime Analysis

The runtime of RC_Cuts, ATT_Cuts, and the proposed method are shown in Table 3 over images of MSRA10K (typical image resolution of 400 × 300) using a PC with 3.60 GHz CPU and 4 GB of RAM. The RC_Cuts method here was the fastest (about 1.24 s per image), followed by the proposed method (about 1.33 s) and ATT_Cuts (about 1.86 s) method.

Table 3. The comparison of processing time obtained by each method.

	The Comparison of Processing Time		
Method	**RC_Cuts [17]**	ATT_Cuts [16]	Proposed Method
Times (s)	1.24	1.86	1.33
Code Type	C++	C++	C++

4.5. Implementation at the School for Visually Impaired

The work done within the study can be used for a variety of practical purposes. In many areas today, the problem of object detection and extraction in images remains relevant. In recent years, one of the most important works has been to create convenient conditions to meet the cultural needs of persons with disabilities using modern, highly effective information and communication technologies, assisting them in regular and independent education, along with improving their social support.

The proposed methods and algorithms of salient object extraction based on locally adaptive triple-thresholding have been implemented in the production and education processes in the training laboratory of the specialized boarding school for visually impaired No. 77 under the Ministry of Public Education of the Republic of Uzbekistan. Furthermore, it can be used to create tactile graphics from natural scene images and to organize other types of assistive technologies and software.

Implementation in practice involves proposed methods and software, uploading images on the computer, extracting salient objects, preparing and printing tactile graphics or tactile display from an object's contours. Figure 13 illustrates the experimental process of perceiving tactile graphics. The contour of the extracted salient object was printed as tactile graphics using Index Braille EVEREST-D V4 braille embosser. Fourteen visually impaired students of the boarding school were tested for their level of familiarity with produced tactile graphics by braille embosser. Tactile graphics were given to 14 blind students and the content of the objects on a tactile graphic was recognized in 60 s. In the experiment, 20 tactile graphics (the same objects for all) were used for each blind student. To assess whether the tactile graphics were understandable, blind students were given a time limit of 60 s and then asked their answers. Blind students were assessed on the basis of 5% (1 image) for each blind student to identify correctly in 60 s without any help, 2.5% (0.5 image) for correct identification in 120 s with little help, or 0% (0 image) for incorrect identification of complex cases. The results of the experimental process of perceiving tactile graphics are presented in Table 4. Experimental results showed that in some cases such as complex or multiple objects, visually impaired students faced some problems in perceiving and classifying objects. In complex cases, teachers gave some additional information to blind students for correct identification. In contrast, they did not fail and misclassify when the objects were single and simple. As a result, visually impaired students were able to clearly

identify 74% (73.93% rounded up) of the tactile graphics that were presented to them. The remaining 26% (26.07% rounded up) of the tactile graphics showed that students did not have an idea of the object in their life, or that the edges of the object were complex.

Figure 13. The visually impaired students of the boarding school who perceived visual information using tactile graphics.

Table 4. The results of the experimental process of perceiving tactile graphics.

Blind Students	Correct Identification		Incorrect Identification	
	Percentage %	Number of Images	Percentage %	Number of Images
Student 1	72.5%	14.5/20	27.5%	5.5/20
Student 2	75%	15/20	25%	5/20
Student 3	67.5%	13.5/20	32.5%	6.5/20
Student 4	77.5%	15.5/20	22.5%	4.5/20
Student 5	72.5%	14.5/20	27.5%	5.5/20
Student 6	75%	15/20	25%	5/20
Student 7	80%	16/20	20%	4/20
Student 8	77.5%	15.5/20	22.5%	4.5/20
Student 9	75%	15/20	25%	5/20
Student 10	70%	14/20	30%	6/20
Student 11	77.5%	15.5/20	22.5%	4.5/20
Student 12	70%	14/20	30%	6/20
Student 13	75%	15/20	25%	5/20
Student 14	70%	14/20	30%	6/20
Overall	**74%**		**26%**	

As a result of implementation of the research work in the specialized boarding school for the visually impaired, using salient object extraction method to access visual information, not only did blind people recognize geometric shapes that are included in the education process, but they also perceived the surrounding visual information using a tactile graphic and display, which can help them to adapt quickly to learning, independent movement, and social life.

5. Limitations

The proposed method may have made some errors in extracting the regions where the image pixels' values were very close to each other. Figure 11 shows these kinds of drawbacks. In addition,

the expert's or teacher's instructions and comments played an important role in improving the natural image recognition process of the visually impaired individuals and in explaining the complex tactile graphics, causing discomfort to the visually impaired individuals when walking alone on the street.

6. Conclusions and Future Work

We presented a saliency cuts method based on local adaptive thresholding using integral images. The proposed method is fully automatic and requires no manual interaction, generating three-level thresholds to divide a saliency map into four regions. The four kinds of seeds are fed into the GrabCuts algorithm to obtain high-quality binary masks of salient objects. On the basis of the proposed saliency cuts method, we performed full-color space salient object extraction. We then applied bilateral filtering and Canny edge detection to produce outer boundaries and internal edges. The experimental results showed that the proposed method was robust and performed better than other methods. In other words, on the basis of this study that extracted a salient object from a natural scene image, the final goal was to create a tactile graphic so that the visually impaired could perceive and understand the natural scene image well. The proposed methods and algorithms of salient object extraction were implemented in a specialized boarding school for visually impaired students who were able to clearly identify 74% of the tactile graphics that were presented to them. We intend to conduct further research on computer vision integrated with deep learning applications for detecting and recognizing multiple salient objects, as well as text information from natural scene images to generate a Text-to-Speech synthesizer for the visually impaired.

Author Contributions: This manuscript was designed and written by A.A., who also performed the experiments, and M.M. and O.D. wrote the original draft, while U.K. helped to revise and improve the manuscript. Theory and experiments were analyzed and commented upon by T.K.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the GRRC program of Gyeonggi province (GRRC-Gachon2017 (B03), Development of Personalized Digital Support Technology based on Artificial Intelligence).

Acknowledgments: I would like to express my sincere gratitude and appreciation to the supervisors, Utkir Khamdamov (Tashkent University of Information Technologies), and Taeg Keun Whangbo (Gachon University) for their support, comments, remarks, and engagement over the period in which this manuscript was written.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

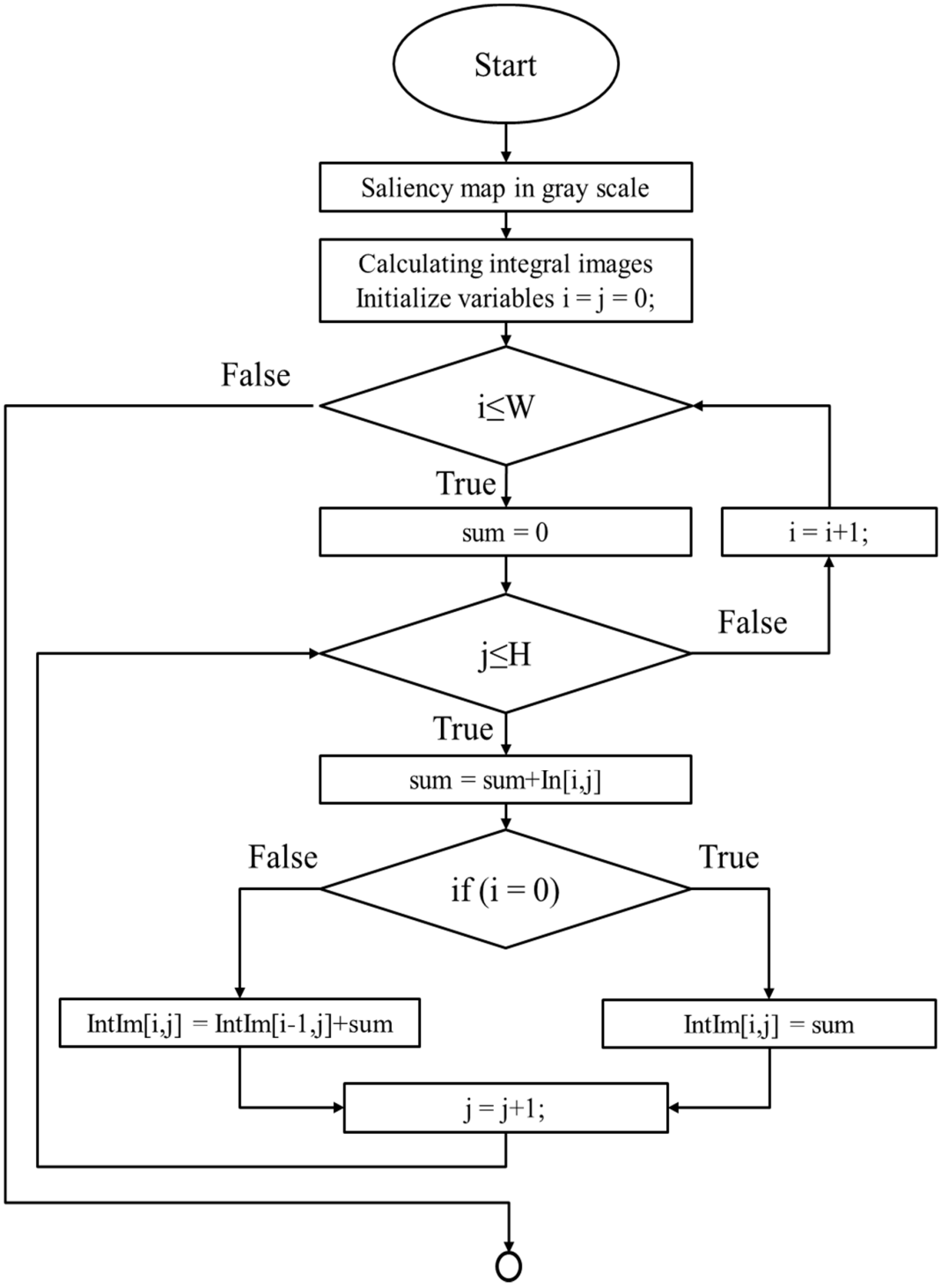

Figure A1. The block diagram of the proposed locally adaptive triple-thresholding using integral images for salient object extraction (one-part).

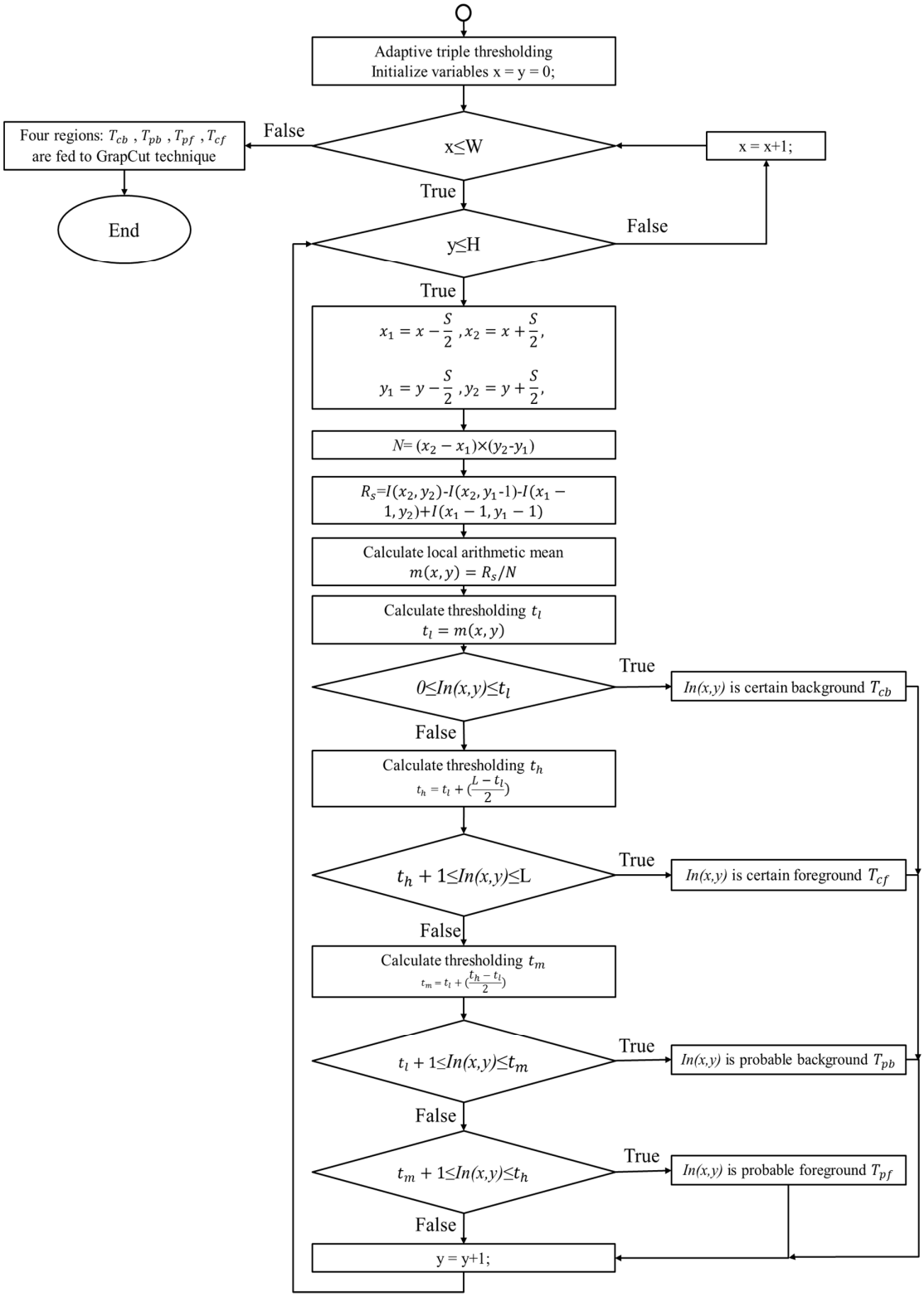

Figure A2. The block diagram of the proposed locally adaptive triple thresholding for salient object extraction (two-part).

References

1. Rahtu, E.; Kannala, J.; Salo, M.; Heikkila, J. Segmenting Salient Objects from Images and Videos. In Proceedings of the 11th European Conference on Computer Vision, Heraklion, Greece, 5–11 September 2010.
2. Guo, C.; Zhang, L. A Novel Multiresolution Spatiotemporal Saliency Detection Model and Its Applications in Image and Video Compression. *IEEE Trans. Image Process.* **2010**, *19*, 185–198. [PubMed]
3. Setlur, V.; Lechner, T.; Nienhaus, M.; Gooch, B. Retargeting Images and Video for Preserving Information Saliency. *IEEE Comput. Graph. Appl.* **2017**, *27*, 80–88. [CrossRef] [PubMed]
4. Fang, Y.; Chen, Z.; Lin, W.; Lin, C.-W. Saliency Detection in the Compressed Domain for Adaptive Image Retargeting. *IEEE Trans. Image Process.* **2012**, *21*, 3888–3901. [CrossRef]
5. Khamdamov, U.; Zaynidinov, H. Parallel Algorithms for Bitmap Image Processing Based on Daubechies Wavelets. In Proceedings of the 2018 10th International Conference on Communication Software and Networks (ICCSN), Chengdu, China, 6–9 July 2018; pp. 537–541.
6. Gao, Y.; Shi, M.; Tao, D.; Xu, C. Database Saliency for Fast Image Retrieval. *IEEE Trans. Multimed.* **2015**, *17*, 359–369. [CrossRef]
7. Wei, X.; Tao, Z.; Zhang, C.; Cao, X. Structured Saliency Fusion Based on Dempster–Shafer Theory. *IEEE Signal Process. Lett.* **2015**, *22*, 1345–1349. [CrossRef]
8. Donoser, M.; Urschler, M.; Hirzer, M.; Bischof, H. Saliency Driven Total Variation Segmentation. In Proceedings of the IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 29 September–2 October 2009.
9. Rutishauser, U.; Walther, D.; Koch, C.; Perona, P. Is bottom-up attention useful for object recognition? In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Washington, DC, USA, 27 June–2 July 2004.
10. Ren, Z.; Gao, S.; Chia, L.-T.; Tsang, I.W.-H. Region-Based Saliency Detection and Its Application in Object Recognition. *IEEE Trans. Circuits Syst. Video Technol.* **2014**, *24*, 769–779. [CrossRef]
11. Mukhiddinov, M.; Akmuradov, B.; Djuraev, O. Robust Text Recognition for Uzbek Language in Natural Scene Images. In Proceedings of the 2019 International Conference on Information Science and Communications Technologies (ICISCT), Tashkent, Uzbekistan, 4–6 November 2019; pp. 1–5.
12. Sharma, G.; Jurie, F.; Schmid, C. Discriminative Spatial Saliency for Image Classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012.
13. Ma, Z.; Qing, L.; Miao, J.; Chen, X. Advertisement evaluation using visual saliency based on foveated image. In Proceedings of the IEEE Conference on Multimedia and Expo, New York, NY, USA, 28 June–3 July 2009.
14. Ghariba, B.; Shehata, M.S.; McGuire, P. Visual Saliency Prediction Based on Deep Learning. *Information* **2019**, *10*, 257. [CrossRef]
15. Yoon, H.; Kim, B.-H.; Mukhriddin, M.; Cho, J. Salient Region Extraction based on Global Contrast Enhancement and Saliency Cut for Image Information Recognition of the Visually Impaired. *TIIS* **2018**, *12*, 2287–2312.
16. Li, S.; Ju, R.; Ren, T.; Wu, G. Saliency Cut based on adabtive triple thresholding. In Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 27–30 September 2015.
17. Cheng, M.-M.; Zhang, G.-X.; Mitra, N.J.; Huang, X.; Hu, S.-M. Global contrast based salient region detection. *TPAMI* **2015**, *37*, 409–416. [CrossRef]
18. Fu, Y.; Cheng, J.; Li, Z.; Lu, H. Saliency cuts: An automatic approach to object segmentation. In Proceedings of the 2008 19th International Conference on Pattern Recognition (ICPR), Tampa, FL, USA, 8–11 December 2008; pp. 1–4.
19. Singh, O.I.; Sinam, T.; James, O.; Singh, T.R. Local Contrast and Mean based Thresholding technique in Image Binarization. *Int. J. Comput. Appl.* **2012**, *51*, 6.
20. Rother, C.; Kolmogorov, V.; Blake, A. "GrabCut": Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.* **2004**, *23*, 309–314. [CrossRef]
21. Shi, J.; Yan, Q.; Xu, L.; Jia, J. Hierarchical Image Saliency Detection on Extended CSSD. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 717–729. [CrossRef] [PubMed]
22. Felzenszwalb, P.; Huttenlocher, D. Efficient graph-based image segmentation. *Int. J. Comput. Vis.* **2004**, *59*, 167–181. [CrossRef]

23. Comaniciu, D.; Meer, P. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 603–619. [CrossRef]
24. Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Susstrunk, S. *SLIC Superpixels*; EPFL Technical Report No. 149300; EPFL: Lausanne, Switzerland, June 2010.
25. Zhou, Q. Object-based attention: Saliency detection using contrast via background prototypes. *Electron. Lett.* **2014**, *50*, 997–999. [CrossRef]
26. Zhou, Q.; Chen, J.; Ren, S.; Zhou, Y.; Chen, J.; Liu, W. On Contrast Combinations for Visual Saliency Detection. In Proceedings of the 2013 IEEE International Conference on Image Processing (ICIP), Melbourne, Australia, 15–18 September 2013.
27. Han, J.; Zhang, D.; Hu, X.; Guo, L.; Ren, J.; Wu, F. Background Prior-Based Salient Object Detection via Deep Reconstruction Residual. *IEEE Trans. Circuits Syst. Video Technol.* **2015**, *25*, 1309–1321.
28. Lee, G.; Tai, Y.-W.; Kim, J. Deep saliency with encoded low level distance map and high level features. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 660–668.
29. Li, G.; Yu, Y. Visual Saliency Based on Multiscale Deep Features. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.
30. Liu, N.; Han, J. DHSNet: Deep Hierarchical Saliency Network for Salient Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
31. Yuan, Y.; Li, C.; Kim, J.; Cai, W.; Feng, D.D. Dense and Sparse Labelling with Multi-Dimensional Features for Saliency Detection. *IEEE Trans. Circuits Syst. Video Technol.* **2016**, *28*, 1130–1143. [CrossRef]
32. Li, G.; Yu, Y. Deep Contrast Learning for Salient Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
33. Mehrani, P.; Veksler, O. Saliency segmentation based on learning and graph cut refinement. In Proceedings of the British Machine Vision Conference, Aberystwyth, UK, 31 August–3 September 2010; pp. 1–12.
34. Ko, B.C.; Nam, J.-Y. Automatic Object-of-Interest segmentation from natural images. In Proceedings of the 18th International Conference on Pattern Recognition, Hong Kong, China, 20–24 August 2006.
35. Jiang, H.; Wang, J.; Yuan, Z.; Liu, T.; Zheng, N. Automatic Salient Object Segmentation based on Context and Shape Prior. In Proceedings of the British Machine Vision Conference, University of Dundee, Dundee, UK, 29 August–2 September 2011; pp. 110.1–110.12.
36. Aytekin, Ç.; Ozan, E.C.; Kiranyaz, S.; Gabbouj, M. Visual Saliency by Extended Quantum Cuts. In Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 27–30 September 2015.
37. Winn, J.; Jojic, N. LOCUS: Learning Object Classes with Unsupervised Segmentation. In Proceedings of the 10th IEEE International Conference on Computer Vision, Beijing, China, 17–21 October 2005.
38. Shi, J.; Malik, J. Normalized Cuts and Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 888–905.
39. Peng, J.; Shen, J.; Jia, Y. Saliency Cut in Stereo Images. In Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops, Sydney, Australia, 1–8 December 2013.
40. Grady, L. Random Walks for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 1768–1783. [CrossRef]
41. Chew, S.E.; Cahill, N.D. Semi-Supervised Normalized Cuts for Image Segmentation. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015.
42. Han, J.; Ngan, K.N.; Li, M.; Zhang, H.-J. Unsupervised Extraction of Visual Attention objects in Color Images. *IEEE Trans. Circuits Syst. Video Technol.* **2006**, *16*, 141–145. [CrossRef]
43. Otsu, N. A threshold selection method from gray-level histograms. *TSMC* **1979**, *9*, 62–66. [CrossRef]
44. Wellner, P.D. *Adaptive Thresholding for the DigitalDesk*; Rank Xerox Ltd.: Cambridge, UK, 1993.
45. Sauvola, J.; Pietikainen, M. Adaptive document image binarization. *Pattern Recognit.* **2000**, *33*, 225–236. [CrossRef]
46. Bradley, D.; Roth, G. Adaptive thresholding using the integral image. *J. Graph. Tools* **2007**, *12*, 13–21. [CrossRef]
47. Peuwnuan, K.; Woraratpanya, K.; Pasupa, K. Modified Adaptive Thresholding using integral image. In Proceedings of the 13th International Joint Conference on Computer Science and Softare Engineering, Khon Kaen, Thailand, 13–15 July 2016.

48. Benny, D.; Soumya, K.R. New Local Adaptive Thresholding and Dynamic Self-Organizing Feature Map Techniques for Handwritten Character Recognizer. In Proceedings of the International Conference on Circuit, Power and Computing Technologies, Nagercoil, India, 19–20 March 2015.
49. Biswas, B.; Bhattacharya, U.; Chaudhuri, B.B. A Global-to-Local Approach to Binarization of Degraded Document Images. In Proceedings of the 22nd International Conference on Pattern Recognition, Stockholm, Sweden, 24–28 August 2014.
50. Takagi, N.; Chen, J. A Broken Line Classification Method of Mathematical Graphs for Automating Translation into Scalable Vector Graphic. In Proceedings of the IEEE 43rd International Symposium on Multiple-Valued Logic, Toyama, Japan, 22–24 May 2013.
51. Jungil, J.; Hongchan, Y.; Hyelim, L.; Jinsoo, C. Graphic haptic electronic board-based education assistive technology system for blind people. In Proceedings of the 2015 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 9–12 January 2015.
52. Chen, J.; Takagi, N. A Pattern Recognition Method for Automating Tactile Graphics Translation from Hand-drawn Maps. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Manchester, UK, 13–16 October 2013.
53. Takagi, N.; Chen, J. Character string extraction from scene images by eliminating non-character elements. In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, San Diego, CA, USA, 5–8 October 2014.
54. Meng, Y.; Zhang, Z.; Yin, H.; Ma, T. Automatic detection of particle size distribution by image analysis based on local adaptive canny edge detection and modified circular Hough transform. *Micron* **2018**, *106*, 34–41. [CrossRef] [PubMed]