

Automatic Segmentation of Product Bottle Label Based on GrabCut Algorithm

In Seop Na, Yan Juan Chen, Soo Hyung Kim*

School of Electronics and Computer Engineering

Chonnam National University, 77 Yongbong-ro, Buk-gu, Gwangju 500-757, Korea

ABSTRACT

In this paper, we propose a method to build an accurate initial trimap for the GrabCut algorithm without the need for human interaction. First, we identify a rough candidate for the label region of a bottle by applying a saliency map to find a salient area from the image. Then, the Hough Transformation method is used to detect the left and right borders of the label region, and the k-means algorithm is used to localize the upper and lower borders of the label of the bottle. These four borders are used to build an initial trimap for the GrabCut method. Finally, GrabCut segments accurate regions for the label. The experimental results for 130 wine bottle images demonstrated that the saliency map extracted a rough label region with an accuracy of 97.69% while also removing the complex background. The Hough transform and projection method accurately drew the outline of the label from the saliency area, and then the outline was used to build an initial trimap for GrabCut. Finally, the GrabCut algorithm successfully segmented the bottle label with an average accuracy of 92.31%. Therefore, we believe that our method is suitable for product label recognition systems that automatically segment product labels. Although our method achieved encouraging results, it has some limitations in that unreliable results are produced under conditions with varying illumination and reflections. Therefore, we are in the process of developing preprocessing algorithms to improve the proposed method to take into account variations in illumination and reflections.

Key words: GrabCut, Segmentation, Bottle label, Initial trimap, Human interaction.

1. INTRODUCTION

1.1 Background

A smart phone is a high-end mobile phone built on a mobile computing platform, with high computing performance and connectivity, at the same time, it not only combine kinds of digital devices such as PDA(personal digital assistant), a mobile phone, camera phone, PMP(portable media players), GPS and so on, but also provide high-resolution touchscreens and high-speed data access via Wi-Fi. That is why it can gradually play an important role. Due to its multi-function, it attracts the attention to not only ordinary users but also many researchers. The researchers attempt to develop more convenient technologies by using devices for more user friendly applications. The camera is one of convenient tool, and pattern recognition skills are also like object recognition, flower recognition, face recognition and etc. Most of them are based on the first step of ROI(Region Of Interest) extraction. Product is necessity for everyone's life. However, they just provide some basic information to introduce themselves (e.g. product name, composition of materials, production time and place), they are insufficient to satisfy consumers' needs. However, it is possible to provide the user with more

information on the camera based smart phone application. For example, wine label recognition can provide the win type and how to drink the win.

We pay the attention on the product bottle as our experimental object. Prior to access the required information, the most of important step is the object segmentation that is to separate ROI(Region Of Interest) object of interest from the input images which is captured by a high-definition smart phone, and then following steps can be go on such as image analysis, recognition, retrieval stage and so on.

This paper is organized as follows. Section 2 represents some state-of-art object segmentation approaches (GrabCut algorithm). The preprocessing of system is described in section 3, and we present the procedure of initial trimap building for object segmentation, following by the graph-based method, GrabCut method, is used to separate the product label region of interest from the complicated background. The experimental results are shown in section 4 and at last we discuss the conclusion in section 5.

1.2 Related Work

Image segmentation is the first step of image processing and the basic of computer vision. It is a key step in image processing and analysis. It not only reduces the volume of data but also extract the image structure features for advanced processing steps, image analysis and recognition. Due to the segmentation errors are propagated to the high-level processing

* Corresponding author, Email:shkim@chonnam.ac.kr

Manuscript received Apr. 28, 2014; revised Dec. 10, 2014; accepted Dec. 18, 2014

step, the segmentation step is critical and important. Image segmentation is an important technology, and it sometimes has other names: object delineation, thresholding, and target recognition, target detection, target recognition, target tracking technology, and so on. Core of these technologies or itself is image segmentation.

Image segmentation is widely used in image processing involving in various types of image. For example, in remote sensing image, target segmentation of synthetic aperture radar images; in medical image, GM, WM, CSF and NB segmentation of Brain MR images and so on; on the traffic image analysis, car or license plate is separated from background. On the image compression and image retrieval, image is segregated into several different object regions. Since image segmentation directly affects the following advanced processing stages, it is great significance for image processing.

So far, there are a large number of method have been developed for image segmentation for instance, the simplest method, thresholding, based on a binary image. It is a simple and effective method to separate the image into several regions by a threshold value. The most well-known method is Otsu's method [3]. The clustering techniques is also used for image segmentation based on distance between a pixels and a cluster center, such as k-means [4], Mean Shift Algorithm [5], [6]. Alternatively, there are many region based techniques for image segmentation, such as region growing method, split-and-merge methods and so on. The region growing method is the seeded region growing method, it takes a set of seeds as input along with the image, the region are interactivity grown by comparing all unallocated neighboring pixels to the region. The main limitation of many region based algorithms is the permeation to area where the boundaries between the objects are weak or blurry.

In recent years, curve propagation is a popular technique on image analysis for object segmentation, object tracking, and stereo reconstruction. It is to evolve a curve towards the lowest potential of a cost function, where its definition reflects the task to be addressed and imposes certain smoothness constraints. The level set method [7]-[13] can be used to efficiently address the problem of curve propagation in an implicit manner. To the similar intensity or concave of the region, it is difficult to segment. These methods described above relay on the assumption that the regions approximate normal distributions. They often get a relatively worse segmentation results for bottle label captured by smart phones due to complexity of label region. Therefore, another energy minimization method has to be mentioned, graph based method. It can be effectively used for image segmentation based on a cost function. In these methods, the image is modeled as a weighted, undirected graph. Good examples of these methods are normalized cut [14], minimum cut/max flow cut [15] and others. Moreover, the graph-cut based method is widely used to image segmentation. It has been applied to medical image (gray-level images), nature image (color images) for multiple objects segmentation by using region and boundary information. However, the segmentation result depends on the choice of rectangle (background and foreground) by users. Fig. 1 shows the graph cut-based (GrabCut method, that extends graph cut to color images) segmentation result with different rectangles by users.



Fig. 1 Graph-based method segmentation with different rectangles (a) and (c) input image with rectangles, (b) and (d) segmentation results respect to different rectangles

1.3 Challenging task

This research is the prior step for the product label recognition as smart phone application. Our final goal is to retrieval the information which we need such as shown in Fig. 2. Our experimental object, product bottle, is captured by a smart phone in the indoor environment. It always follows complex background such as uneven illumination, strong reflection (especial product bottle), and other unexpected factors.



Fig. 2. Label recognition system

In this paper, saliency map and Hough Transformation method are used to automatically obtain the initial rectangle that labeled background region and foreground region for product label region segmentation based on graph-based method (GrabCut algorithm). The product label region can be accurately separated from the complicated background without user interaction by our proposed method.

2. GRABCUT ALGORITHM

2.1 GrabCut Algorithm

GrabCut [16]-[19] is an iterative image segmentation technique based on the Graph Cut algorithm developed by C. Rother et al. It is widely used to address the problem of efficient, interactive extraction of a foreground object in a complex environment whose background cannot be trivially subtracted.

GrabCut extends Graph Cut algorithm which is proposed by Boykov and Jolly to color information and to complete

trimaps. User interaction is simplified to drawing a rectangle around the desired foreground. The basic procedures for GrabCut algorithm is as follows. Each procedure will be explained in more details later on.

● **GrabCut algorithm**

- A. User creates an initial trimap by selecting a rectangle around the ROI (object of interest). The trimap is composed of the foreground, background and the unknown part of image which can be either foreground or background. Pixels inside the rectangle will be marked as unknown. Pixels outside this rectangle will be marked as known background.
- B. The computer creates an initial image segmentation, where the unknown pixels are tentatively placed in the foreground class and all known background pixels are placed in the background class.
- C. Two Gaussian Mixture Models (GMMs) are created for initial foreground and background classes.
- D. Every pixel in the foreground class is assigned to the most corresponding Gaussian component in the foreground GMMs. The same process is done with the pixels in the background but for component of the background GMMs.
- E. New GMMs are learned from the pixel sets create in the previous set.
- F. A graph is built and Graph Cut is used to find a new pixel classification of foreground and background.
- G. Repeat steps 4-6 until the classification converges.

2.2 Methodology of GrabCut

2.2.1 Modeling foreground and background

The initial information provided about background and foreground is given by user as a rectangular region selection around the ROI (region of interest). Pixels inside the rectangle are marked as unknown and pixels outside are be marked as known background. According to this information, a model is attempt to be created that could be used to determine the unknown pixels are either background or foreground. In the GrabCut algorithm, the image is taken to consist of pixels Z_n in RGB color space; this is done by creating K components of multivariate GMMs (Gaussian Mixture Models) for two regions. K components for known background and K components for the region which could be foreground, giving a total $2K$ components. In GMM each cluster is mathematically represented by a parametric Gaussian distribution. Each component is a Gaussian distribution parameterized by μ_k, Σ_k , denote the date by $X, X \in R^d$. The density of component k is:

$$p_k(x) = \frac{1}{(2\pi)^d |\Sigma_k|} \exp\left(\frac{-(x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k)}{2}\right) \quad (2.1)$$

The prior probability (weight) of component k is π_k . The mixture density is:

$$p(x) = \sum_{k=1}^K \frac{\pi_k}{(2\pi)^d |\Sigma_k|} \exp\left(\frac{-(x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k)}{2}\right) \quad (2.2)$$

The parameters of GMM are estimated by ML (the maximum likelihood) based on K-means algorithm. Fig. 3 shows an image and its GMM labeling.



Fig. 3. Gaussian Mixture Model labeling (a) Input image and (b) segmentation result by GMM

In GrabCut algorithm, the Gaussian clustering algorithm consist of two steps such as above, and in the second step, once the pixels have been clustered, we have to throw away the current Gaussian components and create new ones for each component.

2.2.2 Segmentation by energy minimization based on Graph Cut

Graph cuts can be used to efficiently solve a wide area of low-level problem in computer vision as described by Boykov and Jolly that can be formulated in terms of energy minimization. The energy minimization problem can be translated as maximum flow problem in a graph. Fig. 4 shows the graph cuts for segmentation.

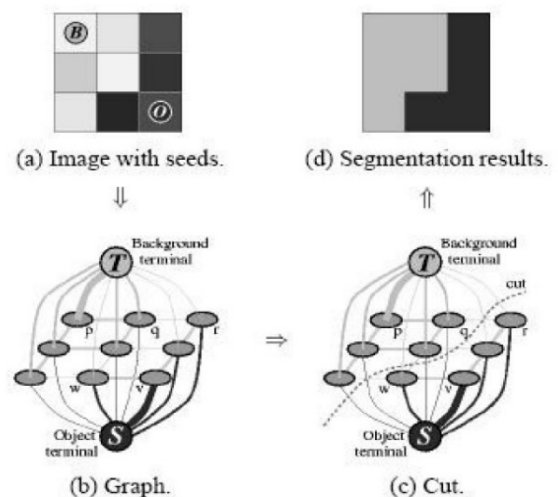


Fig. 4. Graph-Cut algorithm Segmentation Technique

For an image, it is an array $Z = (z_1, \dots, z_n, \dots, z_N)$ of grey values in Graph cut algorithm and this array is used to describe an image in the color space, indexed by n. The segmentation of

the image is expressed as an array of "opacity" value $\alpha = (\alpha_1, \dots, \alpha_N)$ at each pixel. For hard segmentation, $\alpha_N \in (0,1)$, with 0 being for background and 1 being for foreground. θ presents the $K = (k_1, \dots, k_n, \dots, k_N)$ parameters of GMMs containing μ_k, Σ_k, τ_k . In order to deal with the GMM, an additional vector is introduced, with k_n assigning to each pixel, a unique GMM component, one component either from the background or the foreground model, according as $\alpha_n = 0$ or 1.

The energy function of GrabCut algorithm can be described by Gibbs energy function with data term D and smoothness term V as following:

The Gibbs energy for segmentation is defined as:

$$E(\underline{\alpha}, K, \underline{\theta}, Z) = U(\underline{\alpha}, K, \underline{\theta}, Z) + V(\underline{\alpha}, Z) \quad (2.3)$$

The data term is defined according to color GMM models as:

$$U(\underline{\alpha}, K, \underline{\theta}, Z) = \sum_{n=1}^N D(\alpha_n, \theta_n, k_n, z_n) \quad (2.4)$$

where

$$D(\alpha_n, \theta_n, k_n, z_n) = -\log p(z_n | \alpha_n, k_n, \underline{\theta}) - \log \pi(\alpha_n, k_n) \quad (2.5)$$

and $p(\bullet)$ is a Gaussian probability distribution, and $\pi(\bullet)$ are mixture weighting coefficients, so that

$$D(\alpha_n, \theta_n, k_n, z_n) = \frac{1}{2} \log \det \sum (a_n, k_n) - \log \pi(\alpha_n, k_n) + \frac{1}{2} [z_n - \mu(\alpha_n, k_n)]^T \sum (a_n, k_n)^{-1} [z_n - \mu(\alpha_n, k_n)] \quad (2.6)$$

The smoothness Term V is calculated by using Euclidean distance in color space:

$$V(\underline{\alpha}, Z) = \gamma \sum_{(m,n) \in C} dis(m, n)^{-1} [\alpha_m \neq \alpha_n] \exp^{-\beta} \|z_m - z_n\|^2 \quad (2.7)$$

where $[\bullet]$ denotes the indicator function taking values 0, 1 for predicate, C is the set of pairs of neighboring pixels in the 8-neighborhood. This is energy encouraging coherence in region of similar color value. By optimizing performance the constant γ was defined 50 and β is chosen to be:

$$\beta = \left(2 \left\langle (z_m - z_n)^2 \right\rangle \right)^{-1} \quad (2.8)$$

the segmentation can be estimated as a global minimum:

$$\hat{\underline{a}} = \arg \min E(\underline{\alpha}, K, \underline{\theta}, Z) \quad (2.9)$$

Minimization is done using a standard minimum cut algorithm.

2.2.3 Comparison and Disadvantage of Graph-Cut Algorithm

GrabCut extends Graph Cut algorithm to color information and to complete trimaps. They both described as an energy minimization algorithm, however, they are many differences. On the energy equation, the monochrome image model is replaces for color by Gaussian Mixture Model (GMM) in place of histogram, and Graph cut algorithm is previous one-shot algorithm while GrabCut algorithm is an iterative procedure that alternates between estimation and parameter learning. On the provide information by users, Graph cut algorithm need users provide two kinds of information(background and foreground), while GrabCut algorithm only need the user specifies background region information for building the trimap, and this can be done by drawing a rectangle around the ROI(region of interest).

Generally, GrabCut algorithm consists of hard segmentation and soft segmentation. The hard segmentation portion is defined by the energy function $E(\underline{\alpha}, K, \underline{\theta}, Z)$. And the soft segmentation is to produce continuous alpha values based on a given matting tool. It can be augmented by "border matting", that is, It is used to smooth the object boundaries to obtain a better segmentation result comparing with hard segmentation result. In this paper, we focus on the hard segmentation. GrabCut algorithm is a way to perform object segmentation that is very user friendly. On the bright side, we can continue to modify the segmentation result by defined the background and foreground until the ideal result obtained. That is means that it cannot automatically separate the object from background, as we known, different initial trimaps can get different segmentation results. Therefore, we propose an approach to detect out candidate label region to build initial trimap (background and foreground) for GrabCut segmentation algorithm.

3. PROPOSED METHOD

Our proposed method is used to automatically separate the product label region from the background by using GrabCut algorithm without user interaction. The proposed method flow-chart is shown in following Fig. 5.

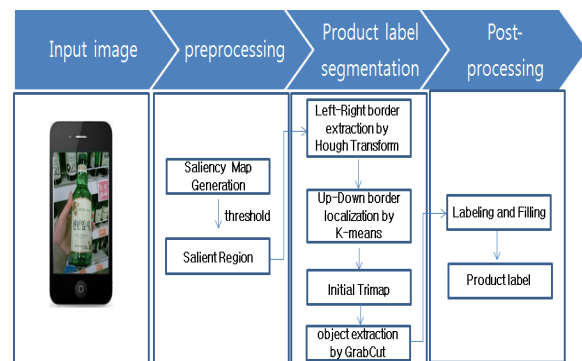


Fig. 5. The flow-chart of our proposed method

A smart phone is used to capture experimental images by users. The saliency map method has advantage of obtaining the

salient region as the preprocessing stage. The candidate label region as initial trimap for GrabCut algorithm is obtained from the salient region by combing Hough Transformation and K-means clustering method. Finally, the segmentation results are improved by labeling and filling in the post processing stage.

3.1 Pre-processing

The saliency map [20]-[25] is designed as input to the control mechanism for covert selective attention. A method to location the most salient area is proposed by Itti et al [21] based on computing the position of maximum in this map by the Winner-Take-All mechanism. Due to the experimental images are captured in the indoor environment, it often is with complicate background containing uninteresting objects, cabinets and hands holding the product bottle, it is not trivial task to segment an interesting object in our experimental images taken in the supermarkets. To take out the complicated background, we first generate a saliency map on the input image Fig. 3.2 and then apply a threshold value to the image in order to draw the boundary of the interest region or salient area.

3.1.1 Saliency Map Generation

The human attention mechanism plays an important role in biological vision and human cognition. Therefore, a lot of researches are being done in the vision community on how to accurately model human attention in order to detect ROI (region of interest). Several models of visual attention have been proposed in the literature which can be classified into (a) bottom-up method, (b) top-down method, and (c) Bayesian or hybrid model.

Bottom-up models of visual attention [21]-[25], use local features in a given image locations which are considerably different from their neighbors. These methods normally designed in three steps. (1) Feature Extraction, (2) Activation, (3) Normalization and Combination. In the present study, the saliency feature maps are drawn from the complementary work of Itti et al[21], including down sampling ratio, Color(C), Intensity(I), Orientation(R). A dynamic Gaussian pyramid $I(\sigma)$ is created based on intensity image i , where $\sigma \in [0...8]$ is the scale. Feature vectors are calculated by using linear "center-surround" operation.

Center-surround differences between a center fine scale C and surround coarser scale s yield the feature maps. The center is the pixel at scale $c \in (2,3,4)$, and surround is the corresponding pixel at scale $s = c + \sigma$, with $\sigma \in (3,4)$:

The first set with six feature maps is calculated for Intensity:

$$I(c,s) = |I(c) - I(s)| \quad (3.1)$$

Similarly, the second sets with 24 feature maps are generated by using color double opponent such as red-green (Eq. 3.2) and blue-yellow (Eq. 3.2) for the color channels:

$$RG(c,s) = |(R(c) - G(s)) - (G(c) - R(s))| \quad (3.2)$$

$$BY(c,s) = |(B(c) - Y(s)) - (Y(c) - B(s))| \quad (3.3)$$

Local orientation information is obtained from i using oriented Gabor pyramids $O(\sigma, \theta)$, the Gabor filters tuned to $\theta \in (0^\circ, 45^\circ, 90^\circ, 135^\circ)$:

$$O(c,s,\theta) = |O(c,\theta) - O(s,\theta)| \quad (3.4)$$

In total, 42 feature maps are calculated including six intensity maps, 12 color maps and 24 orientation feature maps. These feature maps are normalized and linearly added to generate respective conspicuity maps such as three feature maps (color, intensity orientation feature maps) in Fig. 3.2. And these conspicuity maps are further normalized and linearly added to the final saliency map.

3.1.2 Saliency Region Generation

In order to obtain the interesting region from the input images, we set a thresholding value 55% to detect out these regions with accuracy of 97.7%(only 3 is failure in total images).Fig. 6 show the process of saliency map generation.

3.2 Product Label Segmentation

3.2.1 Candidate Label Region Generation

a) Left – Right Border Detection

The edge detector can describe the boundary information of an interesting object. Hence, we use the canny edge detector to detect straight lines in the interesting region detected by saliency map such as show in Fig. 6. However, other factors such as discontinuous edge feature or human's hand can produce the unexpected results. Especially, the boundary of the label on the product bottle forms mostly straight lines. Therefore, we use a Hough Transformation detect straight lines in the restricted area. For example, the transformation is applied to all gradient pixels on the edge to convert to Hough space. The Hough Transformation [26], [27] is a technique that can be used to isolate feature of a particular shape within an image. It is most commonly used for the detection of regular curves such as lines, circles, ellipses and so on.

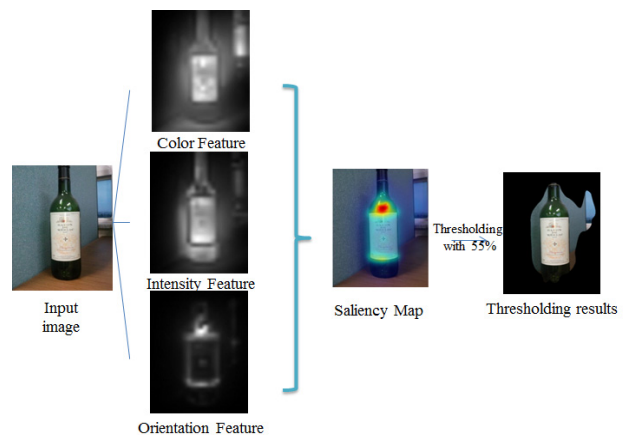


Fig. 6. Saliency Map and Region Generation Procedure

In this paper, we just consider Hough Transformation method for straight line detection. We can analytically describe a line segment in a number of forms such as below equations with a pairs of parameter (a,b) , the slope parameter a and the intercept parameter b respectively; however this representation fails in case of vertical lines. In [26], R. Duda and P. Hart explored the fact that any line in the xy plane, shown in Fig. 7. Can be described by a following convenient equation with parametric and normal notion:

$$\begin{aligned} \rho(\theta) &= x \cdot \cos(\theta) + y \sin(\theta) \\ y &= a \cdot x + b \end{aligned} \quad (3.5)$$

where $\rho(\theta)$ is the length of a normal from the origin to this line, and θ is the normal orientation of $\rho(\theta)$ with respect to the x-axis; for any point (x,y) on this line, $\rho(\theta)$ and θ are constant. Applying Hough Transformation for each point will produce a sinusoidal curve in the $\rho-\theta$ space, called parameter space shown in Fig. 8.

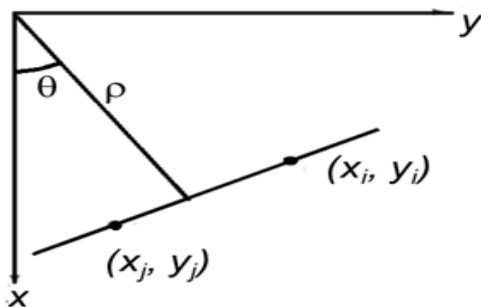


Fig. 7. Parametric description of a straight line

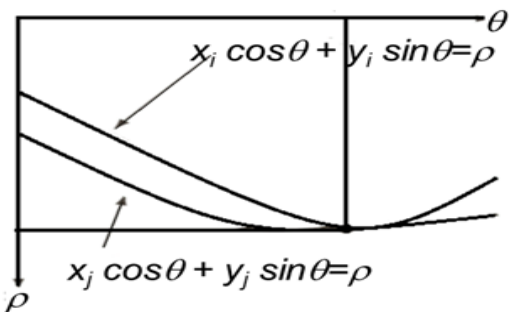


Fig. 8. The result of applying Hough Transform for (x_j, y_j) and (x_i, y_i)

However, Not only two straight lines are detected but also some other lines are detected in the image due to complex background or the light reflection. Hence, we have to select only two straight lines based on the observation of the two straight lines on the left and the right boundary of the label that is somehow similar in intensity value as the left and right borders of the label region such as Fig. 9. Therefore, the user's hand could be removed by the selected straight lines.



Fig. 9. Comparison of the results of the line detection (a) without and (b) with applying the saliency map, respectively and (c) the final straight line detection results after selecting two similar lines

b) Up – Down Localization

Due to the complicated background, edge information is insufficient to accurately locate the up-down boundary of label region. Hence, the k-means algorithm with 3 classes ($K = 3$) is used to segment the rough candidate label region defined by left-right border (Fig. 10(a) and Fig. 10(b)).

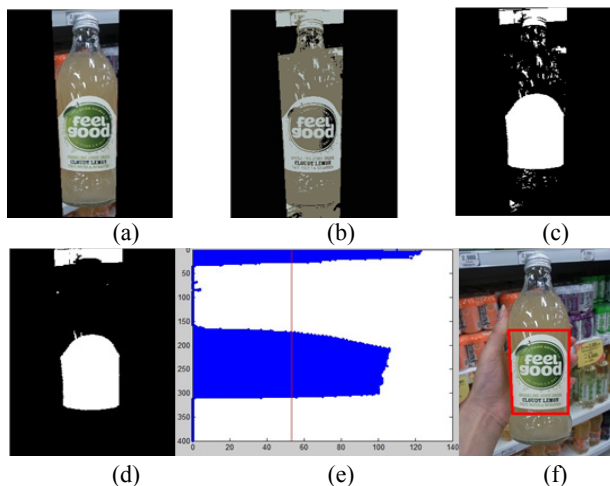


Fig. 10. The process of up-down border localization (a) input image (b) segmentation by k-means algorithm with 3 classes (c) result of region filling, (d) result of region removing, (e) profile of (d) and (f) the candidate label region as initial trimap.

From Fig. 10(b), many small regions intersected between different components, and affect the statistical data to detect boundary of label region. Therefore, two kinds of operations are used to provide us more accuracy data, (1) Region Filling, (2) Region Labeling. Generally, product label region consist of

multi-components such as the comprehensive text, decorative maps. They are often separated as different component with label region component. The Region Filling operation is used to fill all of these regions in order to change them into label region components such as shown in Fig. 10(c). According to the result of region filling (Fig. 10(b)), we can also find many the same component with label region distribute in the un-label region. Many of them are produced by the strong reflection. Therefore, we labeled all of them and try to remove these region with small area to gain a relatively clean map such as shown in Fig. 10(c).

Finally, the horizontal profile is calculated according to the result after removing operation by region labeling. The most area is selected as the label region to determine the up-down boundary (denoted by X and Y). In the Fig. 10(f) show the result of candidate label region by red line.

3.2.2 Label region segmentation by GrabCut

In GrabCut algorithm, the energy function E is defined by "Gibbs" energy including the data term U and the smoothness term V . The data term U evaluate the fit of the opacity distribution α to the data z an array of image is defined by $Z = (z_1, \dots, z_n, \dots, z_N)$, and the smoothness term V is calculated using Euclidean distance in color space. This energy encourages coherence in regions of similar grey-level. GrabCut is a iterative minimization algorithm with user interaction for rough segmentation between foreground and background. Here, the obtained candidate label region is as the initial trimap to segment the product bottle images. In order to obtain the more accurate product label region, the region filling and labeling are modified the segmentation result by GrabCut algorithm.

4. EXPERIMENTAL RESULTS

4.1 Experimental Data

To our research, the high-definition smart phone was used to take the product bottle images in the supermarket. In this paper, a set of 130 images have around 300×400 , were collected to test and evaluate the performance of our proposed methods by using Matlab (7.7.0.471 R2008b, Mathworks).

4.2 Segmentation Results

The segmentation results based on our proposed method are shown in Table 1. and Table 2. Table 1 show the result of saliency map detection, and Table 2 show the final segmentation results. The results are classified into three groups: Success, Partial Success and Failure. Success indicates that the methods segment correctly the label regions. Partial Success means that the segmentation results contain the main label region, and we classified it into two sub-groups again: Over-segmentation and Under-segmentation. Over-segmentation is the results not only contain the label region correctly, but also contain small part of the un-label region; Under-segmentation is defined as the results are the main part of the whole label region that lost a small part. The Failure means that the results do not separate from the background because of the ambiguity and low contrast around boundaries.

Table 1. Results of Most Salient Area Detection

	Success	Partial Failure	Failure
Number	108	11	3
Accuracy (%)	89.23	8.46	2.31

Table 2. Results of product bottle label segmentation

	Total	Success	Partial Success		Failure
			Over	Under	
Number	130	108	7	5	10
Accuracy (%)	100	83.08	9.23		7.69

Although user's hand was connected to an object of interest in an input image, Hough Transformation was able to detect the straight lines on the left and right border of the product label as background shown in Fig. 11(b). Besides, the irregular shape or with and without shelf railing of the label region in Fig. 11(a), all of them have been detected successfully by the proposed method.



Fig. 11. Example of successful results (a) a set of original images and (b) the segmentation results

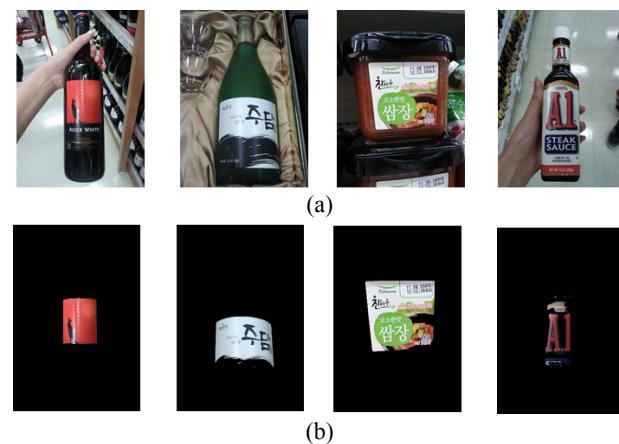


Fig. 12. Example of failure results (a) a set of original images and (b) the segmentation results.

Fig. 12 shows the example of failure results. The color component of the part of label region is similar or the label region has low contrast around boundaries with the un-label region, these part of region are often segmented as background component. Such as the fourth results of Fig.12, the main color component of label region is white where is similar as the ground of supermarket, the text with high contrast is only segmented out.

The segmentation results are compared with the region based active contour, grow cut and our proposed approach in Fig. 13. Region based active contour method is sensitive to the intensity. It has difficulty to segment the label region with uneven illumination such as the third result in Fig. 13. Grab cut algorithm is one of energy minimizing algorithms based on color information; it could obtain the relatively good segmentation results based on the selection of seeds (background and foreground) by user for the input images with relatively simple color. However, both of the region and boundary information with GMM model are used to separate the ROI (region of interest) from complex background.

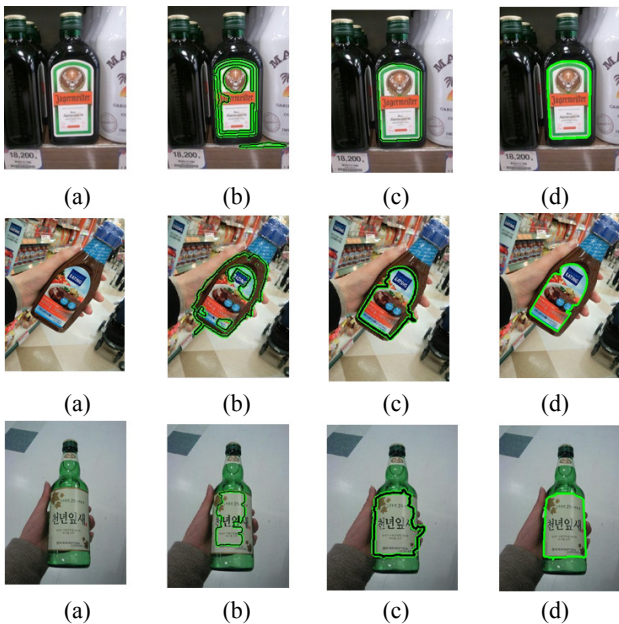


Fig. 13. Comparison with other methods (a) original images, (b) region based active contours, (c) grow cut method, (d) our proposed method

In order to assess the segmentation performance, the Receiver Operation Characteristic (ROC) (Duda et al., 2001) curve is analyzed Fig. 14, which is defined by *FPR* and *TPR* as *x* and *y* axes respectively. Since *TPR* is equivalent with sensitivity and *FPR* is equal to 1-specificity, the ROC graph is sometimes called the sensitivity vs (1-specificity) plot. It is defined as:

$$TPR = \frac{TP}{TP + FN} \tag{4.1}$$

$$FPR = \frac{FP}{FP + TN} \tag{4.2}$$

Where, *TP* is the number of true positives (pixels of the foreground correctly classified) *TN* is the number of true negatives (pixels of the background correctly classified) *FP* is the number of false positives or false alarms (pixels of the background classified as foreground) *FN* is the number of false negatives (pixels of the foreground classified as background).

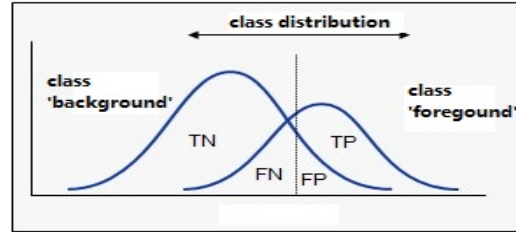


Fig 14. Class distribution

Fig. 15 shows that GrabCut converge to True Positive 1 faster than GrowCut or Region Based Active Contour.

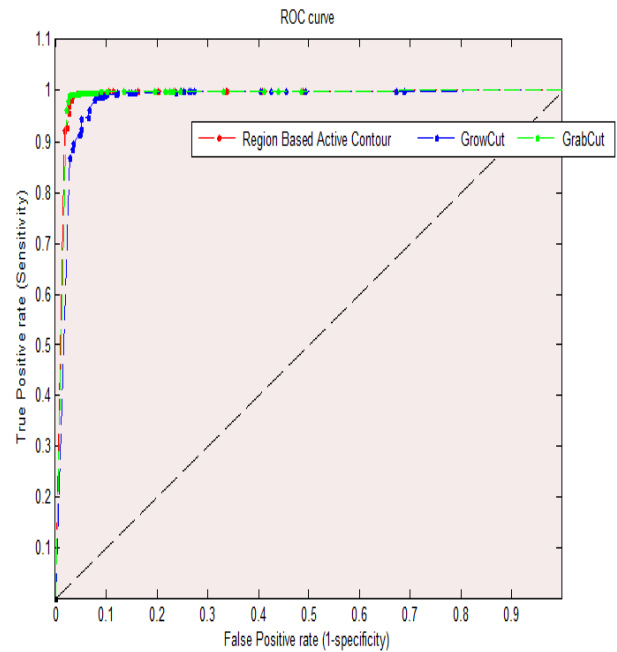


Fig 15. Comparison ROC curve (a) confusion Matric (b) ROC curve

5. CONCLUSION

Product bottle is captured by smart phone in normal indoor lighting environment. They are often with complex background. The object (product label region) segmentation is a challenge task. Generally, the object segmentation just uses one kind of information to separate the object from the relatively simple background. However, GrabCut algorithm is a kind of energy minimization algorithm with a cost function formed by combining the region and boundary information based on GMM in the color space for foreground detection. The drawback is that it need user set the initial trimap (background and foreground) manually. In this paper, we

focus on the product label region detection for the initial trimap for the GrabCut without user interaction. The saliency map and Hough Transformation method are used to find the fine candidate label region, the region is as input to automatically segment out the label region. Experimental results have shown that, our proposed method has been successfully applied to detection for product bottle images and achieved a high detection accuracy of 92.31% under the complex background. Although the results are encouraging our study has several limitations. The method hardly works on the label region similar to the background. In addition, strong reflection often negatively affects the segmentation results due to Hough Transformation detection. Therefore, we plan to extend our works to low contrast and light reflection.

ACKNOWLEDGEMENT

“This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science, and Technology (2014-014400).”

REFERENCES

- [1] S. W. Hong and L. Choi, “Automatic Flowers Recognition Using Segmentation,” Korea Computer Congress, vol. 38, no. 1(A), 2011, pp. 463-465.
- [2] J. S. Lee, S. H. Kim, and J. H. Park, G. S. Lee, H. J. Yang, C. W. Lee, “Recognition of Text in Wine Label Images,” IEEE Pattern Recognition on Chinese Conference, 2009, pp. 1-5.
- [3] N. Otsu, “A Threshold Selection Method from Gray-Level Histograms,” IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, no. 1, 1979, pp. 62-66.
- [4] J. B. MacQueen, *Some Methods for Classification and Analysis of Multiplicate Observations*, Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, 1967, pp. 281-297.
- [5] D. Comaniciu and P. Meer, “Mean Shift: A Robust Approach Toward Feature Space Analysis,” IEEE Transactions on PAMI, vol. 24, no. 5, 2002, pp. 1-17.
- [6] Jong Hyun Park, Guee Sang Lee, and Soon Yong Park, “Color image segmentation using adaptive mean shift and statistical model-based method,” Computers and mathematics with Applications, vol. 57, issue 6, Mar. 2009, pp. 970-980.
- [7] T. F. Chan and L. A. Vese, “Active Contours Without Edge,” IEEE Transactions on Image Processing, vol. 10, no. 2, 2001, pp. 266-277.
- [8] A. Tasi, A. Yezzi, and A. Willsky, “Curve Evolution Implementation of the Mumford-Shah Functional for Image Segmentation, Depositing, Interpolation and Margination,” Institute of Electrical and Electronics Engineers Transactions on Image Processing, vol. 10, no. 8, 2001, pp. 1169-1186.
- [9] D. Cremers, “A multiphase level set framework for variational motion segmentation,” In Scale Space Methods in Computer Vision, vol. 2695, 2003, pp. 599-614.
- [10] C. Li, C. Xu, and M. D. Fox, “Level set evolution without reinitialization: A new variational formulation,” Institute of Electrical and Electronics Engineers Conference on Computer Vision and Pattern Recognition, vol. 1, 2005, pp. 430-436.
- [11] C. Li, C. Y. Kao, J. C. Gore, and Z. Ding, “Implicit Active Contours Driven by Local Binary Fitting Energy,” IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1-7.
- [12] V. Caselles, F. Catte, T. Coll, and F. Dibis, “A geometric model for active contours in image processing,” Numerische Mathematik, no. 66, 1993, pp. 1-31.
- [13] M. Rousson and R. Serricchio, “A Variational Framework for Active and Adapt active Segmentation of Vector Valued Images,” Proceeding of IEEE Workshop on Motion and Video Computing, 2002.
- [14] J. B. Shi and J. D. Malik, “Normalized Cuts and Image Segmentation,” IEEE Transactions on pattern analysis and machine intelligence, vol. 22, no. 8, 2000, pp. 888-905.
- [15] Y. Boykov and Vladimir Kolmogorov, “An Experimental Comparison of Min-Cut/Flow Algorithms for Energy Minimization in Vision,” IEEE Transactions on pattern analysis and machine intelligence, vol. 26, no. 9, 2004, pp. 1124-1137.
- [16] Y. Boycov and M. Jolly, “Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images,” In Proc. IEEE Int. Conf. on Computer Vision, In ICCV, 2001, pp. 105-112.
- [17] C. Rother, V. Kolmogorov, and A. Blake, “GrabCut: interactive for foreground using iterate graph cuts,” SCM Transaction on Graphics(TOG), 2004.
- [18] Justin F. Talbot and Xiaoqian Xu, “Implementing GrabCut,” Brigham Young University Revised, Apr. 7, 2006.
- [19] Peng Wang, “GrabCut-Interactive Foreground Extraction,” <https://mywebpage.wisc.edu/pwang6/CS766>.
- [20] B. C. Ko and J. Y. Nam, “Object-of-interest image segmentation based on human attention and semantic region,” Optical Society of Society of America, vol. 23, Oct. 2006, pp. 2462-2470.
- [21] L. Itti, C. Koch, and E. Niebur, “A Model of Saliency-Based Visual Attention for Rapid Scene Analysis,” Institute of Electrical and Electronics Engineers Transactions on Pattern Analysis and Machine Intelligence, vol. 20, 1998, pp. 1254-1259.
- [22] L. Itti, “Automatic Foveation for Video Compression Using a Neurobiological Model of Visual Attention,” Institute of Electrical and Electronics Engineers Transactions on Image Processing, vol. 13, no. 10, 2004, pp. 1304-1318.
- [23] L. Elazary and L. Itti, “Interesting objects are visually salient,” Journal of Vision, vol. 8, no. 33, 2008, pp. 1-15.
- [24] L. Itti and C. Koch, “Computational modeling of visual attention,” Nature Reviews Neuroscience, vol. 2, no. 3, 2001, pp. 194-203.

- [25] J. Harel, C. Koch, and P. Perona, *Graph-based visual saliency*, in *Advances in Neural Information Processing System 19*, Cambridge, MA: MIT Press, 2007, pp. 545-552.
- [26] D. Ballard, "Generalizing the Hough Transform to Detection Arbitrary Shape," *Pattern Recognition*, vol. 13, no. 2, 1981, pp. 111-122.
- [27] R. Duda and P. Hart, "Use of the Hough Transformation to Detect Line and Curves in pictures," *Communication of the ACM*, vol. 15, no. 1, Jan. 1972, pp. 11-15.
- [28] R. C. Gonzalez, E. W. and S. L. Eddins, *Digital Image Processing using MATLAB*, Publishing House of Electronics Industry, 2002.



In Seop Na

He received his B.S., M.S. and Ph.D. degree in Computer Science from Chonnam National University, Korea in 1997, 1999 and 2008, respectively. Since 2012, he has been a research professor in Department of Computer Science, Chonnam National University, Korea.

His research interests are image processing, pattern recognition, character recognition and digital library.



Yan Juan Chen

He received the B.S in Computer Science from Hanoi University of Science and Technology, Vietnam in 2010. And he recieved the M.E in the Department of Computer Science, Chonnam National University, Korea. In 2013. His main research interests include pattern

recognition, image processing, text recognition, object segmentation and object tracking. She received the B.S in computer science (Information Security) from Yangzhou University, China in 2010. And he recieved the M.E in the Department of Computer Science, Chonnam National University, Korea. In 2012. Her main research interests include pattern recognition, product image processing and medical image processing



Soo Hyung Kim

He received his B.S. degree in Computer Engineering from Seoul National University in 1986, and his M.S. and Ph.D degrees in Computer Science from Korea Advanced Institute of Science and Technology in 1988 and 1993, respectively. From 1990 to 1996, he was

a senior member of research staff in Multimedia Research Center of Samsung Electronics Co., Korea. Since 1997, he has been a professor in the Department of Computer Science, Chonnam National University, Korea. His research interests are pattern recognition, document image processing, medical image processing, and ubiquitous computing.