Journal of Digital Contents Society Vol. 16 No. 6 Dec. 2015(pp. 951-959)
http://dx.doi.org/10.9728/dcs.2015.16.6.951

장면 텍스트 추출을 위한 캐니 연산자의 적응적 임계값을 이용한 AEMSER

박순화*, 김동현**, 임현수**, 김홍훈**, 백재경**, 박재홍*, 서영건*

요 약

장면 텍스트 추출은 현대 스마트 시대에서 쏟아져 나오는 다양한 영상 기반 응용에 중요한 정보를 제공하기 때문에 중요하다. 기본적인 MSER(Maximally Stable Extremal Regions) 추출 후에 캐니 연산자를 이용하여 경계를 강화시키는 Edge-Enhanced MSER은 텍스트 추출 측면에서 뛰어난 성능을 보인다. 하지만 캐니 연산자의 임계값 설정에 따라 Edge-Enhanced MSER의 결과영상이 다르게 나타나므로 임계값 설정을 계산하는 방법이 필요하다. 본 논문에서는 캐니 연산자의 임계값을 설정하는 방법 중 히스토그램의 중앙값을 이용하여 경계를 추출하고 이를 Edge-Enhanced MSER에 적용한 AEMSER(Adaptive Edge-enhanced MSER)을 제안한다. 이 방법은 명확한 경계에 대해서만 영역을 추출하기 때문에 기존의 방법보다 더 좋은 결과영상을 얻을 수 있다.

키워드: MSER, 캐니 자동 임계값, 장면 텍스트 추출, 적응적 임계값, 임계값 계산

AEMSER Using Adaptive Threshold Of Canny Operator To Extract Scene Text

Sunhwa Park*, Donghyun Kim, Hyunsoo Im**, Honghoon Kim**, Jaegyung Paek**, Jaeheung Park**, Yeong Geon Seo***

Abstract

Scene text extraction is important because it offers some important information on different image based applications pouring in current smart generation. Edge-Enhanced MSER(Maximally Stable Extremal Regions) which enhances the boundaries using the canny operator after extracting the basic MSER shows excellent performance in terms of text extraction. But according to setting the threshold of the canny operator, the result images using Edge-Enhanced MSER are different, so there needs a method figuring out the threshold. In this paper, we propose a AEMSER(Adaptive Edge-enhanced MSER) that applies the method extracting the boundary using the middle value of histogram to Edge-Enhanced MSER to get the canny operator's threshold. The proposed method can acquire better result images than the existing methods because it extracts the area only for the obvious boundaries.

Keywords : MSER, Canny Auto Threshold, Scene Text Extraction, Adaptive Threshold, Threshold Computation

1. Introduction

As a camera allows us to take some photos, it is indispensable major element on these days, such as smart phone generations. Especially, it can become important information as extracting text from the photo that includes it. For this reason, extracting and recognizing the text area from image technique have developed steadily, and these ways are called as a Extract scene text.

※ Corresponding Author : Yeong Geon Seo
Received : Nov. 18, 2015
Revised : Dec. 28, 2015
Accepted : Dec. 31, 2015
* Gyeongsang Univ. Dept. of Computer Science and Graduate School of CCBM
** Gyeongsang Univ. Dept. of Computer Science

Extract scene text technology can be classified into texture-based, connected component-based, and mixing of the both as appropriate method[1-4].

Assuming that text and non-text have different values, a texture-based method finds and extracts text to use color, brightness, and other information. On the other hand, using geometric property, such as creating similar adjacent pixels for each pixel of image in connected components, a connected component-based method determines whether it is a text or not. If something drew a large stroke in the history of this connected component-based method, this would be SWT(Stroke Width Transform) and MSER(Maximally Stable Extremal Regions). As SWT is the method that finds the width of character from image including text, it has been derived from what the text width in the same area is constant in text recognition. MSER finds all of the connected components against all of the threshold values through binarization using brightness information within the image, depending on variation of each binarization value and it finds the area which variation of connected component is a smallest area. For this result, it is the way to see a small change area in one area.

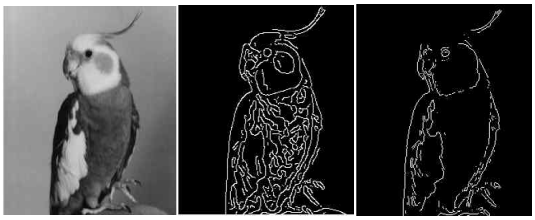
When extracting the text, the text boundary is also important. Through the filtering, the basic edge extraction extracts the portion as the boundary if the colors of the center area are different[5, 6]. An exemplary algorithm of edge extraction is Canny operator. Canny operator extracts the edges through the five steps, such as blurring to eliminate noise, basic edge extraction, removing non-maximum value, being separated the size by two threshold values, and connecting extracted edges[5]. Also, the hybrid research is exist, such as combining connected component and edge extraction. Chen et al proposed Edge-Enhanced MSER that combines extracted edge image and the result of MSER[7]. However, Canny operator comes to the result is different depending on the threshold value. It would be a number of methods for getting good results on average, not different results. As the reason, this paper propose the method that computes automatically the threshold value of Canny operator.

2. Related Works

2.1 Boundary Extraction

Upadhyay and etc proposed a boundary extraction method using FIS(Fuzzy Inference System)[8]. This method divides each pixel into different groups according to the brightness using the brightness histogram of the image and uses 3 statistics types to estimate the threshold to use to extract a boundary. After then, 18 rules are inputted into the defined FIS and the degree is represented by membership functions. The method extracts a boundary using the degree.

(Figure 1) Gray image(left), Canny operator(mid), Edge of FIS(right)



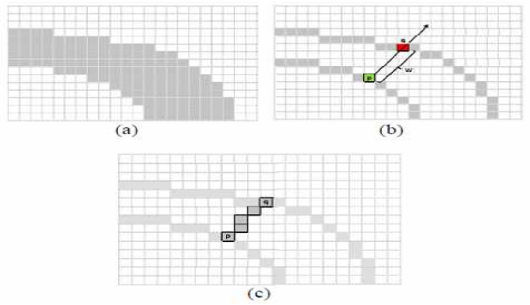
(Figure 1) is used to compare the image using Canny operator to the image that extracted the boundary using FIS for a bird image. As Canny operator has a strong performance, it can extract the boundary which is not only the edges but also something affordable to the boundary inside, but it says that FIS extracts only the boundary around the edges. In scene extract,

the boundary of the small texts comparatively can not be presented by the rule.

2.2 SWT

Epshtein and etc has proposed a method for extracting a text area [9]. To extract the border of the letters in the image, which uses a Canny operator. It shoots the rays from the border pixels in the directions which the slope is a biggest. The slope is known as x and y partial differentiation of the image. Using these values, when shooting the rays in the opposite letters and the two rays faces each other, this means that they are the stroke crossing the letters and filled with an area as the length. If the areas for every letter are filled using this method, a letter must be filled as similar width.

(Figure 2) Principles of SWT



(Figure 2) represents the figure which helps understanding the SWT principle. In the image having SW(Stroke Width) values, it extracts the letter object using CCs(Connected Components). The standard of filtering the letters are followings. Each object has a trend having low variance of SW values. It constraints that the aspect ratio is between 0.1 and 10 and the letters of the same area can be over 3 times of the width of the letter, and supposes that the letters have the same colors, the area of the letters is decided. But according to the font size the letters can not always have the same width, so it can not be decided when the letters have the different letter widths.

2.3 Edge-Enhanced MSER

Chen and etc a method, Edge-Enhanced MSER which extract the final image after mixing MSER and Canny operator and reenforcing the mixed boundary[7]. The method can get the middle image through crossover operating the final image of Canny operator with the final image of the existing MSER and crossover operates again the image after reversing the final one of the existing MSER with the middle image

(Figure 3) Basic principle of Edge-Enhanced MSER

(Figure 3) shows the strengthened boundary of the area after removing the blurring effect into the extracted boundary region using Canny operator. The red color area is the area got from Canny operator and the inner area of the red boundary is the area extracted from MSER. After expanding the boundary, it removes non-text area with SW information and geometric process using the CCs. The effect of Canny operator was already proved, but it can make different results according to setting the thresholds.

The effects of the existing studies were already proved, they have some problems yet in their own ways. In this paper, we propose an method, AEMSER(Adaptive EEMSER) which applies the adaptive threshold of Canny operator to EEMSER based on the problem of EEMSER and approve that it is excellent

comparing to the existing methods.

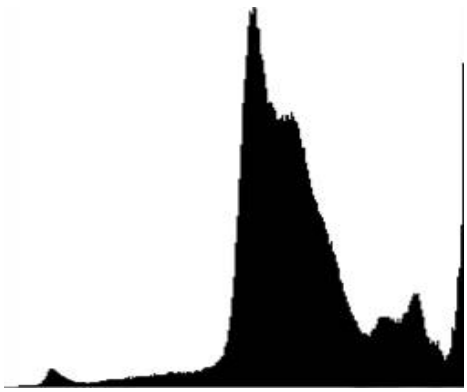
3. Threshold of Canny Operator for AEMSER

This chapter describes AEMSER applying the adaptive method for determining Canny operator to ger the edge threshold of EEMSER.

3.1 Getting the Threshold Using Histogram

The proposed AEMER in this paper follows the method of (Figure 3) and modifies the method to get the final image from Canny operator. In the forth step of Canny operator process, two threshold values may be inputted arbitrary by users to divide their sizes. But these two values are very important and there exist a method to get the proper value using the histogram[10].

(Figure 4) Histogram for brightness



The histogram of image shows distribution about a special component and this paper uses one about the brightness because it can easily get maximum, minimum, mean values and median value. The value to use for compute the threshold is median or mean one, and this paper uses the former. The process to compute the mean value of histogram is following algorithm.

```
input : brightness histogram of image(hist)
output: median value of histogram
max = 0 ; min = 255 ;
For i=0 to 255
  If max<hist[i] Then
    max←hist[i]
    maxIndex←i
  End
  If min>hist[i] Then
    min←hist[i]
    minIndex←i
  End
End
If maxIndex>minIndex Then
  medianValue←(maxIndex-minIndex)/2
Else
  medianValue←(minIndex-maxIndex)/2
End
```

In the algorithm, the limited value, 255 of for statement is generally the brightness is between 0 and 255 and it searches only the scope of x-axis of histogram. On seeking it saves the minimum and maximum indexes. The median value for output is the one divided by 2 the summation of those indexes. Also, it can save a median value with the similar method like this. After getting the median value the next equation determines two threshold values.

$$\begin{aligned} ow \quad hreshold &= median \times 0.66 \\ high\,Threshold &= median \times 1.33 \end{aligned} \tag{1}$$

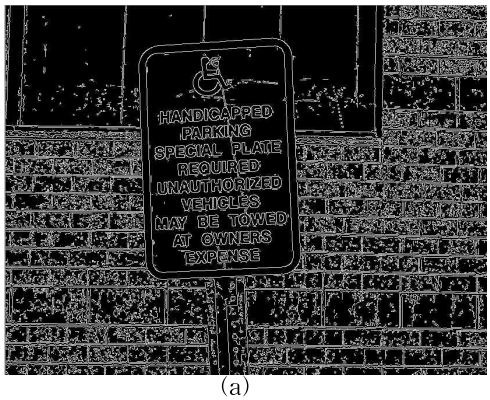
Equation (1) was introduced by Kerry[10]

and it means that each multiplied constant is determined by heuristic method.

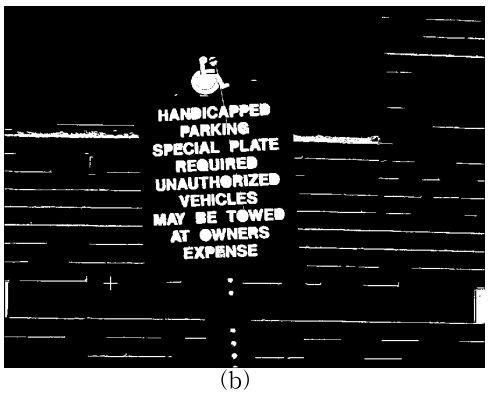
3.2 Extracting the Text Area

(a) of (Figure 5) shows the image from applying two thresholds got from equation (1) to Canny operator and (b) is the one of the existing MSER.

(Figure 5) The result of applying the thresholds computed by the proposed method(a), the result of applying the existing MSER(b)



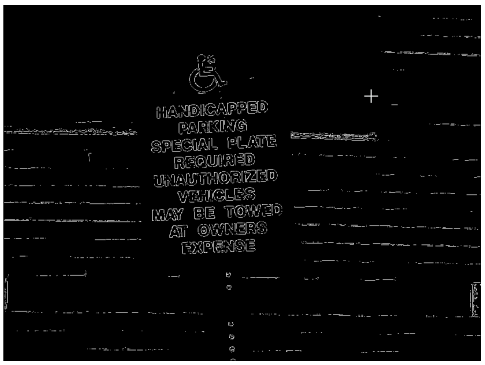
(a)



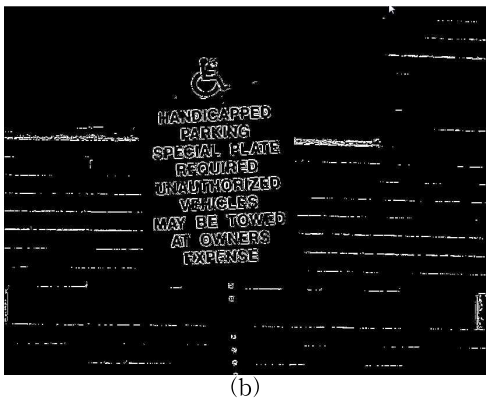
(b)

Cross operating (a) and (b) of (Figure 5) gets (a) of (Figure 6). Expanding the edge of (a) of (Figure 5) gets (b) of (Figure 6). The crossing method forces AND operating the brightness values which the pixels have comparing each pixel of (a) and (b) of (Figure 6).

(Figure 6) Crossed image of Canny image and MSER image(a), image with enlarged boundaries(b)



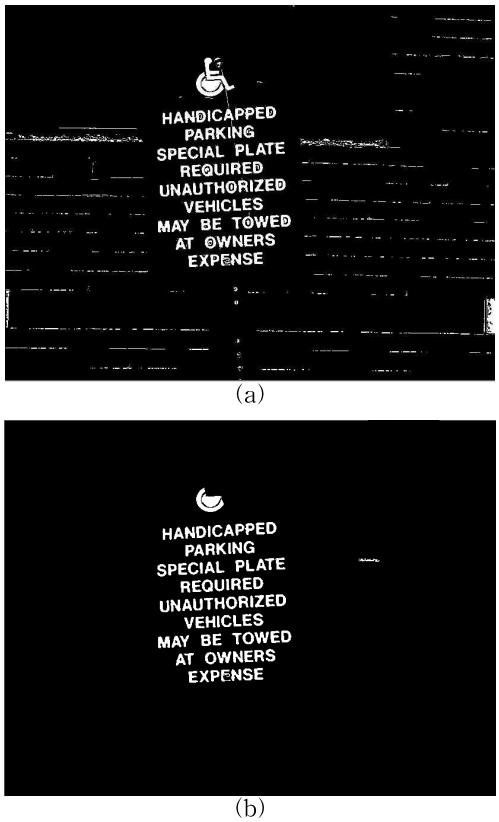
(a)



(b)

Crossing (b) of (Figure 5) and reversed (b) of (Figure 6) gets (a) of (Figure 7). Geometric operation of (a) of (Figure 7) gets (b) of (Figure 7). Reversing (b) of (Figure 6) forces that each pixel value is changed into the 2's complement numbers and AND operates them with each pixel of (Figure 5) (b). And then, after making CCs based on (Figure 7) (a), CCs which are judged with non-text area will be removed using SW information.

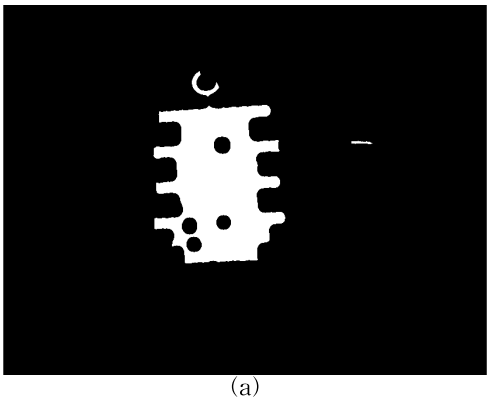
(Figure 7) Image crossing MSER image and reversed image(a), image adding geometric processing(b)



(a)

(b)

(Figure 8) The result to check out the text area(a), the text area extracted by the proposed AEMSER(b)



(a)

(b)

(Figure 8) (a) is the image which (Figure 7) (b) is connected with white color to determine where the text is as whole. This can find the highest values according to up down left right side of the white color and is displayed on (Figure 8) (b). When the proposed method is used, finally it can acquire the text area like (Figure 8) (b)

4. Experiments and Evaluation

4.1 Experimental Method

This paper uses the images of (Figure 9) and the next expression to extract the accuracy for the extracted text area .

$$= \frac{Extracted\,Region}{eal\,Text\,Region} \quad (2)$$

In expression 2, p means an accuracy which means a ratio of real text area and extracted area. We used the system having Windows 7 professional 64 bit operating system, 8GB memory.

(Figure 9) Images used for experiments

(Figure 10) Text area processed by AEMSER

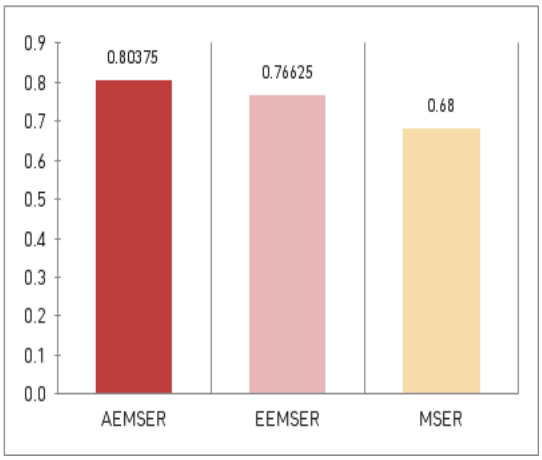
4.2 Performance Evaluation

The criterion to evaluate the performance of the proposed method how the existing methods and the proposed method find precisely the text area from the images. (Figure 10) shows the result of text area for (Figure 9) and <table 1> shows the compared results from MSER and EEMSER based on the (Figure 10). The proposed method in <table 1> shows better performance compared to the existing methods. (Figure 11) shows average chart for the performance. From (Figure 11), the proposed method is superior to the existing methods, 5% with EEMSER and 20% with MSER.

<Table 1> Comparisons for each experimental image

images	accuracy		
	AEMSER	EEMSER[7]	MSER
1	0.77	0.67	0.28
2	0.97	0.97	0.94
3	0.76	0.76	0.46
4	0.79	0.62	0.55
5	0.89	0.86	0.81
6	0.94	0.93	0.94
7	0.55	0.38	0.50
8	0.93	0.94	0.96

(Figure 11) Average precision rate to extract text area for each method



5. Conclusion

This paper modified the method computing the boundary from Canny operator in Edge-Enhanced MSER and proposed AEMSER. Edge-Enhanced MSER is superior to the existing MSER mixing Canny operator but the Canny operator acquired different images according to the threshold. We used the median value of histogram with this threshold and applied the value. So we could the image which the boundary was strengthened and find more accurate text area.

Reference

[1] Li, Yao, et al., "Characterness: an indicator of text in the wild.", Image Processing, IEEE Transactions on 23.4 : pp.1666-1677, 2014.

[2] Jung, Keechul, et al., "Text information extraction in images and video: a survey.", Pattern recognition 37.5, pp.977-997, 2004.

[3] Kang, Le, et al., "Orientation robust text line detection in natural images.", Computer Vision and Pattern Recognition (CVPR), IEEE, 2014.

[4] Sung, Myung-Chul, et al., "Scene Text Detection with Robust Character Candidate Extraction Operator.", 13th ICDAR, 2015.

[5] Canny, John, "A computational approach to edge detection." IEEE Trans. Pattern Anal. Mach. Intell., vol. 8, pp.679-698, 1986.

[6] Fang, Mei, et al., "The study on an application of otsu Operator in canny operator." International Symposium on Information Processing (ISIP). 2009.

[7] Chen, H., et al., "Robust text detection in natural images with edge-enhanced maximally stable extremal regions." In Image Processing (ICIP), 18th IEEE International Conference on, pp.2609-2612, Sep. 2011.

[8] Upadhyay, Nishchal Gyan, and Kamlesh Lakhwani, "Edge Detection Using Fuzzy Approach Involving Automatic Threshold Generation.", International Journal Of Scientific & Techonology Research Vol. 2, Iss. 7, pp.128-131, July 2013.

[9] Epshtein, Boris, et al., "Detecting text in natural scenes with stroke width transform.", Computer Vision and Pattern Recognition, IEEE, 2010.

[10] Kerry D. Wong, "Canny Edge Detection Auto Thresholding", http://www.kerrywong.com/2009/05/07/canny-edge-detection-auto-thresholding/, Sep. 2015.

박 순 화

1996년 : 진주산업대학교 전산과
2000년 : 경상대학교 컴퓨터과학과 석사
2015년~현재 : 경상대학교 대학원 문화융복합학과 박사과정 재학

관심분야 : 멀티미디어, Medical Imaging, IT융복합

김 동 현

2016년 : 경상대학교 컴퓨터과학과 학사
2016년 : 경상대학교 대학원 컴퓨터과학과 재학

임 현 수

2016년 : 경상대학교 컴퓨터과학과 학사

김 홍 훈

2016년 : 경상대학교 컴퓨터과학과 재학

백 재 경

2010년 : 호남대학교 게임에니메이션학과 학사
2015년 : 경상대학교 교육대학원 컴퓨터교육과 재학

박 재 홍

1978년 : 충북대학교 수학교육과
1989년 : 중앙대학교 전산과(박사)
1983년~현재 : 경상대학교 컴퓨터과학과 교수
관심분야 : 소프트웨어 공학

서 영 건

1987년 : 경상대학교 전산과 학사
1997년 : 숭실대학교 전산과 박사
1989년~1992년 : 삼보컴퓨터
1997년~현재 : 경상대학교 컴퓨터과학과 교수
2014년~현재 : 경상대학교대학원 문화융복합학과 교수
관심분야:Med. Image, IT융복합, Computer Network