

실세계 영상에서 적응적 에지 강화 기반의 MSER을 이용한 글자 영역 추출 기법

박영목*, 박순화**, 서영건**

요약

일반 생활 속에서 우리 인간의 눈으로 정보를 인식하고 그 정보를 이용하는 것에는 한계가 없을 만큼 다양하고 방대하다. 그러나 인공지능이 발달한 현재의 기술로도, 인간의 시각 처리 능력에 비하면 턱없이 능력이 부족하다. 그럼에도 불구하고 많은 연구자들은 실생활 속에서 정보를 얻고자 하고 있고, 특히 글자로 된 정보를 인식하는데 많은 노력을 기울이고 있다. 글자를 인식하는 분야에서 일반적인 문서에서 글자를 추출하는 것은 일부 정보처리 분야에서 이용되고 있지만, 실영상에서 문자를 추출하고 인식하는 부분은 아직도 많이 부족하다. 그 이유는 실영상에서는 색깔, 크기, 방향, 공통점 등에서 다양한 특징을 갖고 있기 때문이다. 본 논문에서는 이런 다양한 환경에서 문자 영역을 추출하기 위하여 적응적 에지 강화 기반의 MSER을 적용하여 장면 텍스트 추출을 시도하고, 비교적 좋은 방법임을 실험으로 보인다.

키워드: 적응적 에지, 텍스트 추출, MSER, 텍스트 영역, 자연 영상

An Extracting Text Area Using Adaptive Edge Enhanced MSER in Real World Image

Youngmok Park*, Sunhwa Park**, Yeong Geon Seo**

Abstract

In our general life, what we recognize information with our human eyes and use it is diverse and massive. But even the current technologies improved by artificial intelligence are exorbitantly deficient comparing to human visual processing ability. Nevertheless, many researchers are trying to get information in everyday life, especially concentrate effort on recognizing information consisted of text. In the fields of recognizing text, to extract the text from the general document is used in some information processing fields, but to extract and recognize the text from real image is deficient too much yet. It is because the real images have many properties like color, size, orientation and something in common. In this paper, we applies an adaptive edge enhanced MSER(Maximally Stable Extremal Regions) to extract the text area in those diverse environments and the scene text, and show that the proposed method is a comparatively nice method with experiments.

Keywords : Adaptive Edge, Text Extraction, MSER, Text Area, Natural Image

1. 서론

※ Corresponding Author : Yeong Geon Seo

Received : June 18, 2016

Revised : August 20, 2016

Accepted : August 24, 2016

* GeongNam Nat'l Univ. of Science and Technology, Institute of Computer Information

** Gyeongsang Univ. Dept. of Computer Science and Graduate School of CCBM

요즘 같은 스마트폰 시대에서 카메라는 빠질 수 없는 중요한 요소로 여러 가지 사진을 찍을 수 있다. 특히 텍스트를 포함하는 사진은 텍스트를 추출하는 것으로 중요한 정보가 되기도 한다. 이런 이유는 영상 내부의 텍스트 영역을 추출하고 인식하는 기술들은 꾸준히 발전하고 있으며

Tel: +82-55-772-1392

email: young@gnu.ac.kr

이렇게 텍스트를 추출하는 것을 장면 텍스트 추출이라고 한다. 장면 텍스트 추출 기술은 크게 텍스처를 기반으로 하는 방법과 연결요소 방법, 그리고 이 둘을 적절히 혼합하는 방법들로 크게 구분할 수 있다[1-4].

텍스처를 기반으로 하는 방법은 영상의 색상, 명도 등의 정보를 이용하여 텍스트가 비텍스트 요소와는 다른 값을 갖는다는 전제하에 텍스트를 찾아 추출한다. 이에 반해 연결요소 방법은 영상의 각 픽셀마다 인접해 있는 유사 픽셀을 연결요소로 그룹지어 기하학적인 특성을 이용하여 텍스트인지 아닌지 판별한다. 이런 연결요소 방법의 큰 획을 그은 이론이 있다면 스트로크 폭 변환 (Stroke Width Transform, SWT)와 MSER일 것이다. 스트로크 폭 변환은 글자가 있는 이미지에서 글자의 폭을 찾는 방법으로, 이는 글자 인식에서 같은 영역에 있는 글자의 폭은 일정하다는 점에서 비롯되었다. MSER은 영상에서 명도 정보를 이용한 이진화를 통해 영역을 찾아내어 모든 이진화 값에 대해 연결영역을 찾아내는데 각 이진화 값의 변화에 따라 연결영역의 변화가 적은 곳을 찾는다. 이 결과를 바탕으로 변화가 적은 곳을 하나의 영역으로 보는 방법이다.

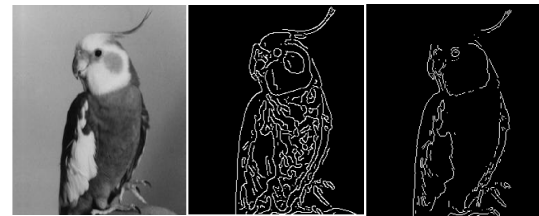
텍스트 추출을 할 때에는 텍스트 경계선 또한 중요하다. 기본적인 경계선 추출은 필터링을 통하여 경계가 되는 부분에서 색이 다른 경우에 그 부분을 경계로 뽑아내는 것이다[5, 6]. 대표적인 경계선 추출의 알고리즘으로는 캐니 연산자가 있다. 캐니 연산자는 잡음을 없애기 위한 블러링, 기본적인 경계선 검출, 비최대값 제거, 두 개의 임계값으로 크기를 구분, 검출된 경계선들의 연결로 다섯 단계를 거쳐 경계선을 추출한다[5]. 또한 연결요소 방법과 경계선 추출을 혼합한 연구도 존재한다. Chen 외 5명은 MSER의 결과영상과 경계선을 추출한 영상을 혼합한 Edge-Enhanced MSER 방법을 제안했다[7]. 하지만 캐니 연산자는 임계값에 따라 결과가 상이하게 나온다. 상이한 결과가 아닌 평균적으로 좋은 결과를 얻기 위한 방법이 여러 가지가 존재할 것이다.

2. 관련 연구

2.1 경계선 추출

Upadhyay 외 2인은 FIS(Fuzzy Inference System)을 이용한 경계선 추출을 제안하였다[8]. 영상의 명도 히스토그램을 이용하여 각 픽셀들을 명도에 따른 그룹으로 나눈다. 각 그룹마다 경계선을 추출할 때 사용할 임계값을 추론하는데 3가지 통계치를 이용한다. 이 후에 18개의 룰이 정의된 FIS에 입력되어 멤버십 함수에 의해 정도가 표현되고 이 정도를 이용하여 경계선을 추출한다.

(그림 1) 회색조 영상(좌), 캐니 연산자 영상(중), FIS에 의한 경계 영상(우)



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

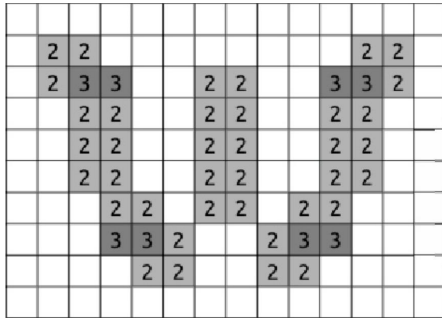
(그림 1)은 [8]에서 새의 영상을 두고 캐니 연산자를 이용한 영상과 FIS를 이용해 경계를 추출한 영상을 비교하기 위해 사용한 것이다. 캐니 연산자는 강력한 성능을 자랑하기 때문에 윤곽선뿐만 아니라 내부에서 경계로 삼을 만한 것까지 경계로 추출하는 반면 FIS에 의한 경계 추출은 윤곽선을 중심으로 경계를 추출한 것을 알 수 있다. 하지만 장면 텍스트 추출에서 상대적으로 크기가 작은 텍스트들의 경계는 룰에 의해 표현되지 않을 수가 있다.

2.2 스트로크 폭 변환

Epshtein 외 3명은 스트로크 폭 변환을 이용하여 글자 영역을 추출하는 방법을 제안했다[9]. 이미지에서 글자의 테두리를 추출하는데 이는 캐니 연산자를 이용한다. 테두리의 한 픽셀에서 기울기가 큰 방향으로 광선을 쏜다. 이는 영상의 x, y방향에 대한 각각 편미분 값으로 알 수 있다. 이 값을 통하여 반대편 글자 픽셀에서도 광

선을 봤을 때 서로 마주보게 될 시에 이는 글자를 가로지르는 획으로 간주하고 선을 거리만큼 채우면 된다. 이러한 방식으로 모든 글자 영역을 채우게 되면 한 글자는 비슷한 폭으로 채워지게 된다.

(그림 2) 문자 W를 SWT 처리한 결과



(Figure 2) An example output image from SWT for character W

(그림 2)는 [9]에서 스트로크 폭 변환의 원리를 이해하는 것을 돕기 위해 제시한 그림이다. SW(Stroke Width)값들이 있는 영상에서 CCs(Connected Components)를 사용하여 글자 객체를 추출한다. 글자 필터 기준은 다음과 같다. 각각의 객체는 SW값의 분산이 낮은 경향이 있다. 중형비를 0.1과 10 사이의 값으로 제한하여 같은 영역의 글자는 자간이 글자의 폭의 3배를 넘을 수 없고 글자는 같은 색을 가지고 있다고 가정하여 글자의 영역을 판단한다. 하지만 폰트에 따라 글자가 항상 같은 폭을 갖지는 않기 때문에 글자가 다른 폭을 갖는 형태라면 글자로 판단하지 못할 수 있다. 그림에서 2, 3은 상하좌우로 스캔하여 최소의 픽셀을 의미하며, 글자의 폭을 의미한다.

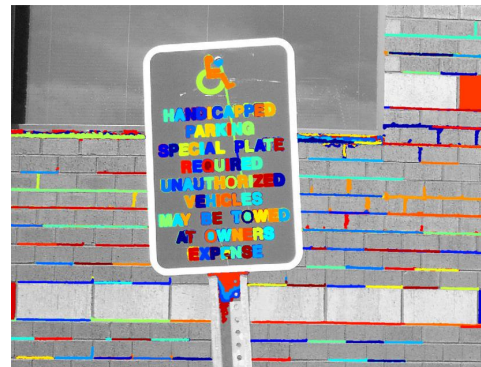
2.3 에지 강화

Chen 외 5명은 MSER과 캐니 연산자를 혼합하고 혼합한 경계선을 강화하여 결과영상을 추출하는 방법인 Edge-Enhanced MSER을 제안했다[7]. 기존의 MSER 결과영상에서 캐니 연산자의 결과영상을 교차 연산하여 중간 영상을 구할 수 있고 기존 MSER 결과를 반전시켜 중간 영상과 다시 교차 연산한다.

(그림 3)은 MSER을 통해 검출된 텍스트 영역

을 표시하고 있다. 각 독립된 영역 별로 다른 색깔로 표시하였으며, MSER을 이용하여 충분히 후보 영역을 검출할 수 있지만, 에지 강화를 이용하면 훨씬 더 좋은 후보 영역을 얻을 수 있게 되며, 그 기술을 적용하여 실험한 결과를 본 연구에서 보이게 된다. 에지 강화를 위해서 캐니 연산자를 이용하여 검출된 경계 영역으로 번짐 효과를 제거하고 영역의 경계를 강화시키는 방법을 이용한다. 경계를 늘린 후에는 연결요소를 이용하여 기하학 처리와 SW 정보를 가지고 비텍스트 영역을 제거한다. 캐니 연산자의 효과는 이미 입증된 바 있으나 사용되는 임계값의 설정에 따라 다른 결과를 얻을 수 있다. 기존의 연구들은 효과가 입증은 되었으나 나름대로의 문제를 가지고 있다. 본 연구에서는 Edge-Enhanced MSER이 가지고 있는 문제점을 바탕으로 캐니 연산자에서 필요한 임계값을 계산하여 Edge-Enhanced MSER에 적용시킨 적응적 에지 강화기반의 MSER 방법을 제안하고 실험을 통하여 우수성을 증명한다.

(그림 3) MSER을 사용하여 문자 영역의 후보를 찾은 이미지



(Figure 3) Detecting candidate text regions using MSER

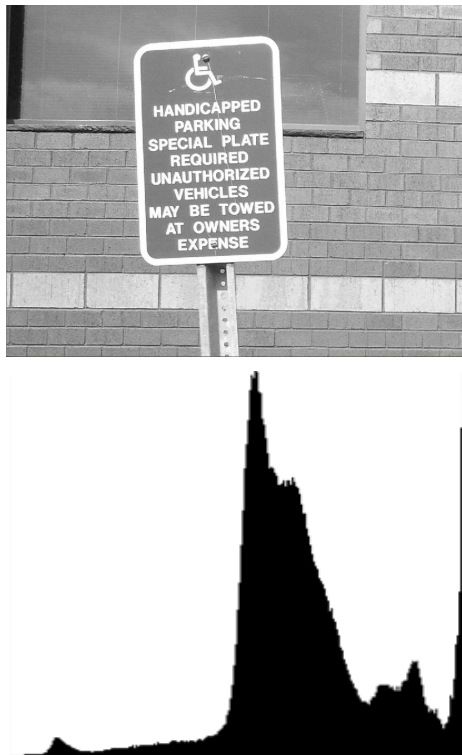
3. 에지 강화를 위한 캐니 연산자의 임계값 계산

3.1 히스토그램을 이용한 임계값 계산

본 연구에서 제안하는 에지 강화 기반의 임계값 계산은 기본적으로 (그림 3)의 방법을 따르고 캐니 연산자 부분에서 결과 영상을 구하는 방법

을 수정했다. 캐니 연산자 과정의 네 번째 단계인 두 개의 임계값으로 크기를 구분하는 부분에서 일반적으로 두 개의 임계값을 사용자 임의로 입력하는 경우가 많다. 하지만 이 두 개의 임계값을 계산할 수 있는 방법들이 있는데 이 방법들 중에서 히스토그램을 이용하는 방법이 존재한다[10-11].

(그림 4) 실영상에서 명도에 대한 히스토그램



(Figure 4) Histogram for brightness of a real scene image

영상의 히스토그램은 특정 요소에 관한 분포를 보이는데 본 연구에서는 명도에 대한 히스토그램을 사용한다. 이는 최솟값, 최댓값, 평균값, 그리고 중앙값 등을 쉽게 구할 수 있기 때문이다. 임계값 계산을 위해 사용될 것은 중앙값 혹은 평균값인데 본 연구는 중앙값을 이용하였다. 히스토그램의 중앙값을 계산하는 과정은 다음 알고리즘과 같다.

입력: 영상의 명도 히스토그램(hist)

출력: 히스토그램의 중앙값(medianValue)

```

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
    
```

알고리즘에서 for문의 제한횟수 255는 일반적으로 명도의 값이 0~255로 표현되기 때문에 히스토그램의 x축의 범위까지만 탐색하도록 하는 의미이다. 탐색을 하면서 최솟값의 인덱스와 최댓값의 인덱스를 구한다. 출력되는 중앙값은 최댓값 인덱스와 최솟값 인덱스의 차를 2로 나눈 것이다. 이와 비슷한 방법으로 평균값 또한 구할 수 있다. 중앙값을 구한 다음에는 다음의 식을 통하여 두 개의 임계값을 결정한다.

$$\begin{aligned}
 low\ Threshold &= median \times 0.66 \\
 high\ Threshold &= median \times 1.33
 \end{aligned}
 \tag{1}$$

식 (1)은 Kerry가 [10]에서 안내한 것으로 각각 곱해지는 상수는 경험적 법칙에 의해 결정됐다. 문자영역의 추출 정확도를 높이기 위해 모폴로지 연산을 사용하며 3x3 마스크를 사용한다. 기본적 연산은 침식, 팽창, 열림, 닫힘 네 가지 연산을 사용한다. 침식은 영상 내에서 객체의 영역을 갠 결과물을 얻고, 팽창은 영상 내에서 객체 영역을 확장하는 듯한 결과를 얻으며, 열림은 침식 연산 후 다시 팽창 연산을 수행하여 돌출 부분을 제거하고 잡음을 제거한다. 닫힘은 팽창 연산 후 다시 침식 연산을 수행하여 가늘게 패인 부분을 채워주는 역할을 한다. 이후에 문자 영역을 인식하게 되는데, 본 제안 방법에서는 문자는 가로로 나열 형이 많다는 특징과 영역의 외곽선만 있어도 구분이 가능하다는 것을 이용한다.

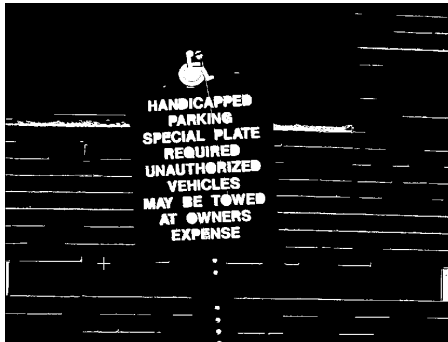
3.2 텍스트 추출을 위한 임계값 적용

(그림 5)의 (a)는 식 (1)에서 계산된 두 개의 임계값을 캐니 연산자에 적용시켜 나온 영상이고 (b)는 기존의 MSER을 거친 영상이다.

(그림 5) 제안 방법에 의해 계산된 임계값을 적용시킨 결과(a), 기존 MSER을 적용한 결과(b)



(a)



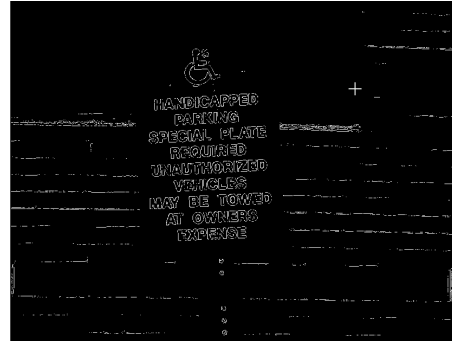
(b)

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

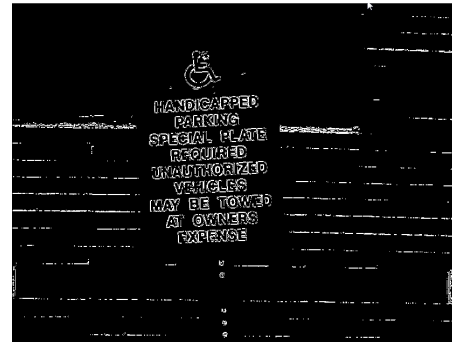
(그림 5)의 (a)와 (b)를 교차시키면 (c)를 얻을 수 있고 (c)의 경계를 늘리면 (d)를 얻을 수 있다. 교차 방법은 (c)와 (d)를 각 픽셀마다 비교하여 픽셀이 가지고 있는 명도 값을 AND 연산시키는 것이다. (그림 5)의 (b)영상과 (그림 6)의 (d) 영상을 반전시킨 영상을 다시 교차시키면 (그림 8)의 (e)를 얻을 수 있고 (e)에 기하학적인 처리를 가하면 (f)가 나오게 된다. (d)를 반전시키게 되면 각 픽셀들이 갖는 값들이 2의 보수형태로 바뀌고 이를 다시 (b)의 각 픽셀들과 AND 연산하는 것이다. 그 다음에 (e)를 기반으로 연결요소들을 만들어 이 연결요소들의 기하학적

특성과 SW 정보를 이용하여 비텍스트 영역이라고 판단되는 연결요소들을 제거한다.

(그림 6) 캐니 영상과 MSER 영상의 교차영상(c), 경계를 늘린 영상(d)



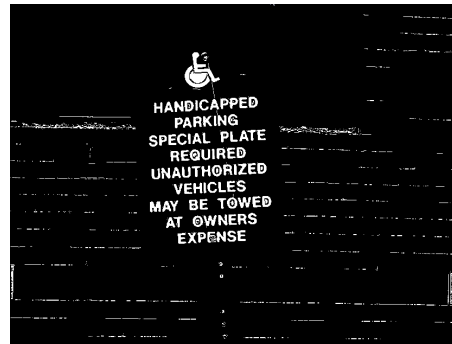
(c)



(d)

(Figure 6) Crossed image of Canny image and MSER image(c), one of enlarged boundaries(d)

(그림 7) 경계 강화 영상을 반전시켜 MSER 영상과 교차한 영상(e)



(e)

(Figure 7) Image crossing MSER and reversed (d) image(e)

(그림 8) (e)를 기하학적 처리한 영상(f)

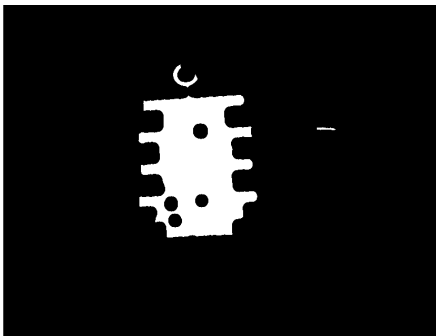


(f)

(Figure 8) Image adding geometric processing(f)

(그림 9)의 영상은 전체적으로 텍스트가 어디에 위치하는지 판단하기 위해 (f)영상을 모두 흰색으로 연결한 영상이다. 이를 통해 흰색의 상화좌우 별로 각각 가장 높은 수준의 값을 찾은 영상이다. 본 제안 방법을 이용하면 최종적으로 (f)와 같은 영상에서 각 분리된 객체 단위로 텍스트 영역으로 확정된다.

(그림 9) 텍스트 영역을 찾기 위해 연결



(Figure 9) The result to check out the text area(g)

4. 실험 및 평가

4.1 실험 방법

본 연구는 (그림 10)의 영상을 이용하여 실험하였으며 추출된 텍스트 영역에 대해서 정확도를 추출하기 위하여 다음과 같은 식을 사용했다.

$$p = \frac{Extracted\ Region}{Real\ Text\ Region} \quad (2)$$

식 (2)에서 p는 정확도를 의미하고 이는 추출된 영역과 실제 텍스트 영역의 비를 의미한다. 본 실험을 진행한 시스템 환경으로는 Windows 7 Professional 64비트 운영체제, 8GB 메모리를 사용하였다.

(그림 10) 실험에 사용한 영상



(Figure 10) Image used for experiments

4.2 성능 평가

제안된 방법의 성능을 평가하기 위한 척도는 기존의 방법과 제안된 방법이 영상에서 텍스트 영역을 얼마나 정확하게 찾아내는가를 정확도로 표현한다. (그림 11)은 (그림 10)에 대한 모폴로지 연산을 한 후의 결과를 나타내며, (그림 12)는 모폴로지 연산 후의 객체 단위의 레이블링을 한 것이다. 텍스트 영역을 표시한 결과이며 이를 바탕으로 기존의 Edge-Enhanced MSER과 MSER 방법과 비교하여 <표 1>에 보이고 있으며 제안 방법이 기존의 방법에 비하여 성능 향상을 보이고 있다. 이에 대한 평균을 차트로 나타낸 것이 (그림 11)이다. (그림 12)를 통해 제안 방법이 기존의 Edge-Enhanced MSER보다 5%, MSER 보다 20% 성능이 향상된 것을 알 수 있다.

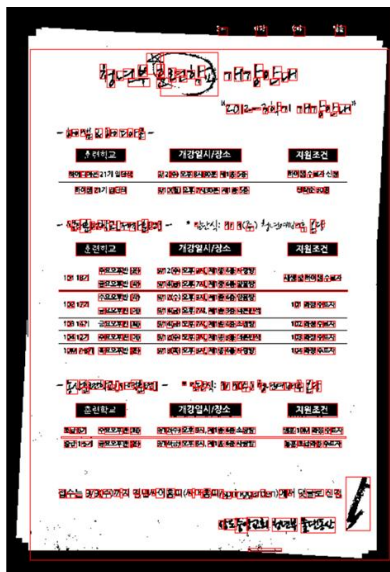
(그림 11) 네 가지 모폴로지 연산 후 결과



(Figure 11) Result from 4 step morphologies

(그림 12)에서 문자 영역을 얻기 위하여, 문자의 가로 세로 특성을 이용하여 한 묶음으로 처리할 수 있는 단위로 레이블링을 하면 (그림 13)과 같이 얻어지며 제안된 방법에서 얻고자 하는 것이다.

(그림 12) 독립된 객체 단위의 레이블링



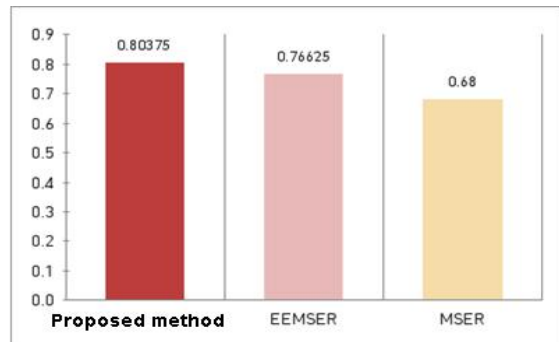
(Figure 12) Labeling of individual object unit

(그림 13) 입력 영상의 최종 텍스트 영역 레이블링



(Figure 13) Final text area labeling of the image

(그림 14) 각 방법을 이용한 텍스트 추출 평균 정확도



(Figure 14) Average precision rate to extract text area for each method

5. 결론

본 연구는 Edge-Enhanced MSER에서 캐니 연산자 결과를 구하는 부분 중 경계를 구하는 부분을 수정하여 적용하는 방법인 적응적 에지 강화 기법을 적용한 MSER을 제안하였다. Edge-Enhanced MSER은 기존의 MSER 영상과

캐니 연산자 결과영상을 혼합한 방법으로 성능이 우수하지만 캐니 연산자는 임계값 설정에 따라 전혀 다른 영상을 얻게 된다. 우리는 이 임계값을 히스토그램의 중앙값을 이용하여 계산하고 이를 적용하여 경계가 강화된 영상을 얻을 수 있었으며 더욱 정확한 텍스트 영역의 위치를 찾을 수 있었다.

References

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

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

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

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

[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 scen

es 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.

[11] S. H. Park and etc, "AEMSER Using Adaptive Threshold of Canny Operator To Extract Scene Text", J. of Digital Contents Society, Vol. 16, No. 6, pp. 953-961, 2015.

박 영 목



1989년 : 경상대학교 전산과 학사
 2003년 : 경상대학교 전산과 박사
 1996년~현재 : 경남과학기술대학교
 정보전산원 팀장

관심분야: 인공지능, 영상처리, 컴퓨터네트워크

박 순 화



2000년 : 경상대학교 컴퓨터학과 석사
 2015년~현재 : 경상대학교 대학원
 문화융복합학과 박사과정
 재학

관심분야 : 멀티미디어, IT융복합, Medical Imaging

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