자연영상에서 한글문자 영역 검출에 관한 연구

박종천*, 권교현*, 전병민*
*충북대학교 전자정보대학 컴퓨터공학부
E-mail:simplepjc@hanmail.net

Study on Hangul Character Region Detection in Natural Images

Jong-Cheon Bak*, Gyo-Hyeon Gwon*,
Byeong-Min Jeon*
*School of Electrical&Computer Engineering,
Chungbuk National University

요 약

최근 모바일 기기로 획득된 영상을 이용한 다양한 분야의 연구가 활발히 진행되고 있으며, 스마트폰의 보급이 확대되면서 증강현실을 지원하고자 자연영상으로부터 문자정보를 추출 및 인식하여 이미지 검색을 가능하도록 하는 많은 연구도 진행되고 있다. 자연영상에서 한글문자 영역 검출은 한글문자 인식을 위한 전단계로서 다양한 환경에 노출된 문자영역을 정확히 검출하는 것이 인식 성능을 결정함으로 중요한 전처리 단계이다. 본 연구는 한글문자 영역의 에지 및 지역적 연결요소 성분 특징을 이용하여 한글문자 영역을 검출하는 방법을 제안한다. 에지 및 연결요소 성분의 특징을 검출하고, 그 결과를 레이블화하고 이를 분석함으로서 한글문자 후보 영역을 검출한다. 검출된 후보영역은 검증과정을 수행하여 최종적인 한글문자 영역을 추출한다. 제안한 방법은 다양한 환경에서 얻어진 자연영상을 대상으로 실험한 결과, 에지 및 연결요소 성분의 두 가지 특징을 결합함으로서 한글 문자영역 검출의 정확도를 향상하였다.

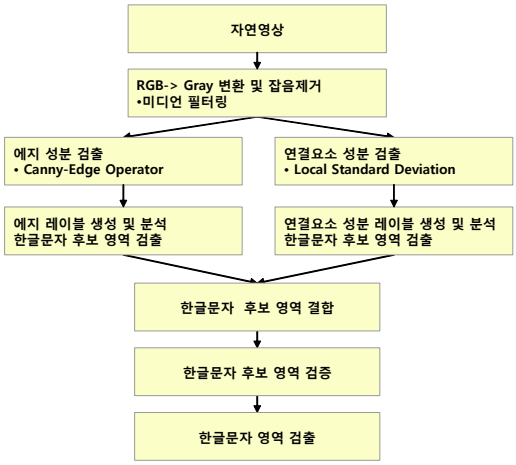
1. 서 론

모바일 기기의 발전과 더불어 모바일 응용 소프트웨어의 중요성이 부각되고 있다. 모바일 기기에 장착된 카메라, GPS, 중력감지 센서, Wi-Fi, 터치 센서 등을 기반으로 다양한 응용 프로그램 개발이 급속도로 개발되고 있고 이들 응용프로그램의 활용함으로서 보다 더 편리한 생활을 영위하고 있다. 영상처리 및 컴퓨터 비전 분야에서는 카메라를 이용한 다양한 응용 프로그램을 개발 및 연구 중에 있다.

외국인이 시내에서 관광할 경우, 그 나라의 문자를 모르면 어려움이 많다. 길거리의 이정표와 간판들을 인식하지 못함으로 원하는 곳에서 식사하거나, 쇼핑을 하기가 힘들다. 이런 상황에서 카메라가 달린 PDA로 간판과 이정표들을 촬영하면, 영상중의 문자 영역을 찾아내어 인식을 수행한 후에, 외국인이 이해할 수 있는 언어로 번역하여 표시하는 시스템이 있다면 관광이 편리할 것이다. 또한 비디오 영상에서 자막의 추출 및 인식은 아날로그로 저장되어 있는 동영상들을 인덱싱 하는데 사용될 수 있다. 최근 연구 중에는 시각장애인을 위한 시각보조 시스템[1]이 개발 중에 있고, 외국인이 여행 중에 휴대용 장비로 외국어로 된 지리정보와 관광정보를 문자영역 검출과 변환기술로 외국어로 변환 가능하도록 하는 시스템[2]도 개발 중이다.

본 논문은 기존의 에지기반 연구 방법[3-6]과 연결요소 분석방법을 결합하여 문자영역을 검출하는 방법을 제안한다. 본 연구는 기존의 에지기반 방법은 복잡한 배경에서 문자영역을 검출하기 어려운 단점이 있고, 연결요소 성분 기반 방법은 빛 또는 조명에 민감한 단점이 있으므로 이러한 단점들을 상호 보완하기 위해서 각각의 방법을 결합함으로서 한글 문자영역 검출에 있어서 정확도를 향상할 수 있었다.

2. 한글문자 후보 영역 검출



[그림 1] 제안한 문자영역 검출과정

한글 문자영역 검출은 그림 1과 같은 과정을 수행한다. 명도정보를 이용한 에지추출은 캐니에지 검출기[7]를 이용하였고, 실험결과 임계값은 영상 전체의 표준편차이용하여 영상에 적응적으로 임계값을 설정하도록 하였다. 그리고 가우시안 필터 처리를 위한 시그마(sigma)값은 1로서 9-by-9 마스크 영역에 가우시안 계수를 설정하여 처리한다.

$$S=\left(\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2\right)^2 \quad 단,\quad \bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i \qquad (1)$$

연결요소는 3-by-3 영역의 식 1과 같은 표준편차를 적용하여 영상에 전체의 지역적 표준편차를 구하고 임계값을 적용하여 문자영역과 배경영역을 구분하게 된다. 실험결과 임계값은 '8'로 하였고, 영상전체의 표준편차가 임계값보다 작은 경우에는 임계값을 '5'로 설정한다. 에지 및 표준편차를 이용하여 검출한 결과를 그림 2에 제시하였다.

[그림 2] 에지 및 연결요소 성분 검출 결과

그림 2는 에지 및 지역적 표준편차를 이용하여 한글문자 영역의 특징을 검출한 결과로서 에지성분을 이용한 경우 에지가 문자영역 주변에 이중으로 검출되는 경우도 있고 한글문자 초성 및 중성 또는 중성 및 종성의 에지가 연결되어 문자영역 후보를 검출하는 알고리즘이 복잡해지는 문제점이 있고, 표준편차를 이용한 한글문자 영역 특징 검출은 문자영역을 초성 및 중성 그리고 종성이 분리되어 검출되고 있으나 영상의 대비가 낮은 경우 임계값의 영향으로 문자영역을 검출하지 못하는 경우도 있고 문자의 크기가 작은 경우에는 그 영역이 너무 작게 검출되어 문자의 형태를 잃어버림으로서 문자영역 검출을 어렵게하는 단점이 존재한다. 따라서 이러한 두 방법의 장단점을 효과적으로 결합하여 한글문자 영역을 검출하는 방법이 필요하다.

2.1 레이블 생성 및 분석

에지 및 표준편차로 검출된 각각의 성분은 레이블링 과정을 수행하여 연결요소 성분을 생성한다. 이렇게 생성된 연결요소 성분은 한글문자 영역 후보 특징을 기준으로 후보 영역을 검출한다. 연결요소 영역의 크기가 30 픽셀 미만이면 문자를 구성할 수 없으므로 제거하였다. 연결요소 영역의 구조적 특징은 표 2에 제시하였다. 표 2에 제시한 조건은 문자영역 에지는 문자 에지 구조의 특징으로 최소 2번 이상 존재해야만 한다는 것을 반영한 것이다. 연결요소 결과로 얻은 이진 영상으로 문자영역이 갖는 구조적 및 통계적 특징을 얻기 위해서 매트랩 함수 regionprops을 수행한다. regionprops는 연결요소 성분의 다양한 픽셀 분포 및 모양정보를 제공한다.

[표 1] 텍스트 레이블 영역의 구조적 조건

nr는 영상의 높이, nc는 영상의 너비 width는 레이블 영역의 너비 height는 레이블 영역의 높이 Number_Nested_Label은 레이블 영역에 포함된 다른 레이블의 수 1) Max_Width_Height=max(width,height); Max_Width_Height < 10 2) width > height*3 or (width <= 4 and height >= width*5) or height >= width*10 3) Number_Nested_Label >= 10 4) width < (nc*0.9) and height < (nr*0.9)

위의 조건 1)은 너비와 높이 중에서 가장 긴 것을 기준으로 최소 10픽셀 미만의 작은 텍스트 영역을 제외하는 것이고 조건 2)는 텍스트의 문자가 갖는 가로와 세로의 비율이고, 조건 3)은 잡음 영역을 제외하는 것이고, 조건 4)는 너무 크거나 작은 텍스트 영역을 제외하는 것이다.

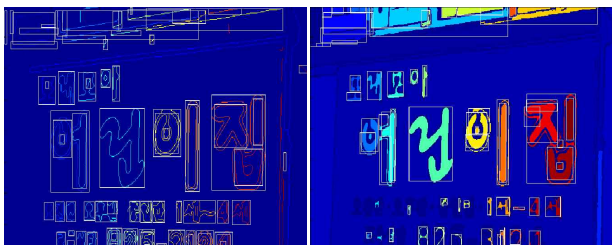
위와 같은 4가지 조건을 문자영역의 후보를 검출하

는 1차적인 조건이고, 표 2에 제시한 2차 조건으로 지역성과 구조적 유사도를 적용하여 만족하지 않는 것은 문자영역 후보에서 제외한다.

[표 2] 텍스트 지역성과 구조적 유사도

```
NB_Width = 인접한 후보 텍스트 영역의 너비
NB_Height = 인접한 후보 텍스트 영역의 높이
Similarity_Width = NB_Width/width;
Similarity_Height = NB_Height/height;
NB_Similarity =
  max(Similarity_Width, Similarity_Width);
If height >= width*5 and
  Similarity_Height >= 0.7 and
  Similarity_Height <= 1.5
else if Similarity_Height >= 0.7 and
      Similarity_Height <= 1.5 and
      Similarity_Width <= 2
```

그림 3은 각각의 방법으로 검출된 한글문자 영역 후보를 보여준다. 그림 3의 왼쪽은 에지 성분 추출결과에서 한글문자 영역 후보를 검출한 것이고, 오른쪽은 지역적 표준편차를 이용하여 한글문자 영역을 검출한 결과이다. 왼쪽의 에지 성분 추출결과를 보면 많은 문자영역 후보가 겹치는 것을 볼 수 있고, 반면 오른쪽 표준편차를 이용한 결과는 겹치는 영역은 상대적으로 적은 반면 작은 글자 영역은 검출하지 못하는 것을 볼 수 있다. 따라서 각각의 에지 및 표준편차를 이용한 연결요소 분석결과를 위에서 제시한 조건으로 최종적인 문자영역을 검출하게 된다.



[그림 4] 한글문자 후보 영역 검출결과

3. 한글문자 후보 영역 검증

문자 후보 영역 검증은 한글문자 영역의 지역적 군집화 특징을 이용하였다. 즉, 한글문자 영역은 일반적으로 일정한 방향으로 군집을 형성한다. 따라서 이러한 특징이 존재하면 최종적으로 한글문자 영역으로 확정한다. 군집화 방법은 검출된 문자 후보 영역의 레이블을 기준으로 주변영역에 다른 문자 후보 영역 레이블이 존재하는지를 검사하여, 존재하면 이들을 하나의 군집으로 간주하여 한글문자 영역을 검증하게 된다. 주변 문자 영역의 설정 기준은 두 가지로 설정하였다. 첫째 조건은 기준 한글문자 영역 레이블과 유사한 크기의 너비와 높이를 갖고 있어야 하며, 둘째 조건은 기준 한글문자 영역의 너비/높이의 크기 중 가장 큰 값을 기준으로 2배 이내의 거리에 이웃 한글문자 영역 레이블이 존재해야 한다.

[그림 3] 한글문자 영역 검출 결과

4. 실험 결과

실험대상 영상은 KAIST 인공지능 연구실에서 한글문자 영역 검출 데이터베이스로 제공하고 있는 SceneText DB를 대상으로 하였고 그 중에서 본 논문은 영상 중에서 한글문자를 포함하는 것을 대상으로 실험하였다. 한글문자 영역 검출 결과에 대한 성능평가[9]는 재현률(Recall), 정확률(Precision)를 평가요소를 사용하여 표 1과 같이 제시하였다.

Sum : 실험 대상 이미지 내에 존재하는 전체 문자 영역의 수 (True+Part+Error)
True : 정확히 검출한 문자영역의 수
Part : 문자영역의 일부분만을 검출한 수
Error : 검출하지 못한 문자영역의 수
False : 문자 영역이 아닌 영역을 문자영역으로 잘못 검출한 수
Precision : True / (True+Part+False)
Recall : True / Sum

[표 1] 한글문자 영역 검출률

실험 이미지 종류	Precision	Recall
안내판	89.7	72.8
간판	90.1	73.3
교통표지	87.1	75.4
번호판	90.0	66.9
교통안내	88.9	80.1
광고표지	91.0	94.8
전체	89.47	77.22

그림 4는 에지 및 표준편차를 이용한 연결요소 분석 방법을 결합한 최종적인 한글문자 검출 결과를 보여준다. 대부분의 한글문자 영역을 검출하였으나 상단의 장석이 있는 부분을 한글문자 모양과 비슷한 특징을 갖고 있으므로 한글문자가 아닌 영역을 한글로 검출하는 오류를 보여준다. 일반적으로 한글문자 영역 검출에서 이러한 경우가 자주 발생되고 있으며 이에 대한 대안으로 한글문자를 검증하는 방법에 대한 연구가 필요하다.

5. 결론

본 연구에서는 자연영상에서 한글문자 영역을 검출하는 방법으로 에지 및 연결요소 성분을 결합함으로서 상호간의 단점을 보완하는 방법을 제안하였다. 본 논문에서 제안한 한글문자 영역 검출 기법은 에지 정보만을 사용하여 한글문자 영역을 검출하는 경우, 배경이 복잡하거나 명도 대비가 낮은 이미지로부터 한글문자 영역을 잘못 검출하거나 검출하지 못하는 경우가 있었으나 연결성분을 이용한 방법을 결합함으로서 이러한 단점을 개선하였다.

향후 연구 과제는 한글문자 영역 검출을 위한 전처리 과정에서 이미지의 밝기 대비에 따른 에지 파라미터를 자동으로 설정하는 방법을 연구하고, 한글문자 영역의 연결요소 성분을 잘 검출할 수 있는 표준편차를 위한 임계값의 결정 방법을 도출하는 것이다. 또한 다른 물체와 겹쳐짐으로서 에지가 끊어짐으로서 발생되는 문제점을 해결하는 기법에 대한 연구가 필요하며, 마지막으로 한글문자 영역 검증을 위한 보다 일반적인 방법에 대한 연구가 필요하다.

참고문헌

[1] N. Ezaki, M. Bulacu, L. Schomaker, “Text detection from natural scene images: towards a system for visually impaired persons”, Pattern Recognition, ICPR 2004. Proceedings of the 17th International Conference on Volume 2, pp.683-686, 2004.

[2] J. Yang, X. Chen, J. Zhang, Y. Zhang, A. Waibel, “Automatic detection and translation of text from natural scenes”, Acoustics, Speech, and Signal Processing, IEEE International Conference on Volume 2, pp.2101-2104, 2002.

[3] M.A. Smith and T. Kanade, “Video Skimming for Quick Browsing Based on Audio and Image Characterization”, Carnegie Mellon University, Technical Report CMU-CS-95-186, 1995.

[4] D. Chen, K. Shearer, and H. Bourlard, “Text Enhancement with Asymmetric Filter for Video OCR”, Proc. of International Conference on Image Analysis and Processing, pp.192-197, 2001.

[5] W. Jiang , Q. Shao-Lin , Q. Zhuo, W. WenYuan , “Automatic text detection in complex color image”, Machine Learning and Cybernetics, 2002. Proceedings. 2002 International Conference on Volume 3, pp.1167-1171, 2002.

[6] H. Xian-Sheng, Z. Liu Wenyin, Hong-Jiang, “An Automatic Performance Evaluation Protocol for Video Text Detection Algorithms”, IEEE Transactions on Circuits and Systems for Video Technology, Vol 14, No. 4, pp.498-507, 2004.

[7] J. Canny, “A Computational Approach to Edge Detection”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI-8, No.6, pp.679-698, 1986.

[8] Luc Vincent, “Morphological Grayscale Reconstruction in Image Analysis:Applications and Efficient Algorithms”, Image Processing, IEEE Transactions on Volume 2, pp. 176-201, 1993.

[9] Vijay Raghavan, Peter Bollmann, Gwang S. Jung, “A critical investigation of recall and precision as measures of retrieval system performance”, ACM Transactions on Information Systems (TOIS), Vol.7, No.3, pp.205-229, 1989.