

장면 이미지로부터 문자-에지 맵 특징을 이용한 텍스트 추출

박종천, 황동국, 이우람, 권교현, 전병민*
충북대학교*

simplepjc@hanmail.net, dkh@kebi.com, ninayosi@nate.com
khgamja@nate.com, bmjun@cbnu.ac.kr

Text Extraction using Character-Edge Map Feature From Scene Images

Jong-Cheon Park, Dong-GukHwang, Woo-Ram Lee,
Kyo-Hyun Kwon, Byoung-Min Jun*
School of Electronical & Computer Engineering, Chungbuk
Univ.*

요 약

본 연구는 장면 이미지로부터 텍스트에 존재하는 문자-에지 특징을 이용하여 텍스트를 추출하는 방법을 제안한다. 캐니(Canny)에지 연산자를 이용하여 장면 이미지로부터 에지를 추출하고, 추출된 에지로부터 16종류의 에지-맵 생성한다. 생성된 에지 맵을 재구성하여 문자 특징을 갖는 8종류의 문자-에지 맵을 만든다. 텍스트는 배경과 잘 분리되는 특징이 있으므로 텍스트에 존재하는 '문자-에지 맵'의 특징을 이용하여 텍스트를 추출한다. 텍스트 영역에 대한 검증은 문자-에지 맵의 분포와 텍스트에 존재하는 글자간의 공백 특징으로 한다. 제안한 방법은 다양한 종류의 장면 이미지를 실험대상으로 하였고, 텍스트는 적어도 2글자 이상으로 구성된다는 제한조건과 너무 크거나 작은 텍스트는 텍스트 추출에서 제외하였다. 실험결과 텍스트 영역 추출률은 약 83%를 얻었다.

1. 서 론

디지털 카메라 기술의 발전에 따라 카메라 보급률이 높아지고, 크기도 소형화되어 모바일 장비에 부착된 형태로 이용되고 있다. 따라서 카메라를 이용한 응용 프로그램의 욕구가 증가하고 있다.

최근에는 시각장애인을 위한 시각보조 시스템이 개발 중에 있다[1]. 이 시스템은 휴대장비에 부착된 카메라로부터 획득한 장면 이미지에서 문자정보를 추출하여 이를 음성으로 문자정보를 전달해주는 시스템이다. 또한 외국인이 여행 중에 휴대용 장비를 통해서 외국어로 된 지리정보와 관광정보 등을 문자 추출과 변환 기술로 외국인들이 이해할 수 있는 문자로 변환 가능하도록 하는 시스템도 개발 중이다[2].

그러나 이러한 시스템에 있어서 필수적으로 수반되는 텍스트 영역 추출 방법은 장면 이미지를 대상으로 하고 있다. 따라서 장면 이미지에서 텍스트가 갖은 다

양한 특성과 복잡한 배경 그리고 다양한 컬러 분포 등으로 인해서 해결해야 할 많은 문제점을 내포하고 있다.

1. 기존 연구방법

텍스트 영역 추출에 관한 수많은 연구가 진행되어 왔으며, 다양한 연구 방법이 제시되었다. 텍스트 영역 추출 방법들은 크게 두 가지 방법으로 분류될 수 있다 [3]. 첫 번째는 연결요소 분석방법(Connected Component based method)[4,5]이고, 두 번째는 에지-기반 분석 방법(Edge-based method)[6,7,8,9,10]이다.

연결요소 분석 방법은 텍스트 영역 추출 속도가 빠른 반면 복잡한 이미지 배경에 영향을 많이 받는다. 그러나 에지-기반 분석 방법은 텍스트 영역의 위치가 조금 정밀하지 못한 단점이 있다. 따라서 본 논문에서는 에지-기반 분석방법에서 발생하는 정확한 텍스트 영역을 찾지 못하는 문제점을 개선하여 문자-에지 맵 특

정에 대한 구조적 분석 방법을 이용하여 좀 더 정확한 문자영역 추출을 목적으로 한다.

II. 텍스트 영역 추출

텍스트 영역 추출은 그림 2와 같은 처리과정에 따라서 수행된다. 기본적으로 문자-에지 맵 특징을 이용하여 텍스트의 상단과 하단을 그리고 좌우 영역을 결정한다. 이와같은 방법으로 추출된 텍스트 후보로부터 텍스트 영역의 분포 및 텍스트 영역의 글자간의 간격 등의 특성을 이용하여 텍스트 영역을 검증한다.

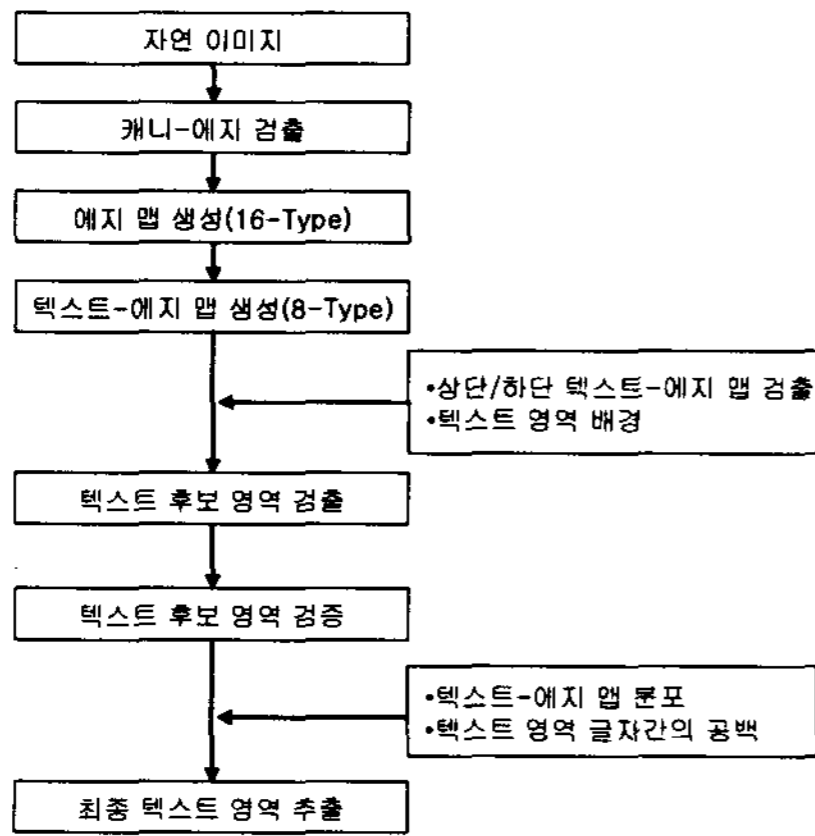


그림 1. 텍스트 영역 추출 절차

1. 에지 검출 및 에지 맵 생성

일반적으로 텍스트 영역은 배경영역과 대비가 강하다는 특징을 갖고 있다. 그러므로 대비가 강한 특징을 추출하기 위해서 에지 검출을 수행한다. 에지검출은 캐니(Canny) 에지 연산자를 이용하였다. 추출된 에지를 조합하여 16 종류의 에지 맵을 만든다 에지 맵의 분류 기준은 김수겸, 장유정[12]의 논문에서 제시한 존재 가능한 정확한 에지의 형태에 따라 선정하였고, 그림 2는 16종류 에지 맵을 나타낸다.

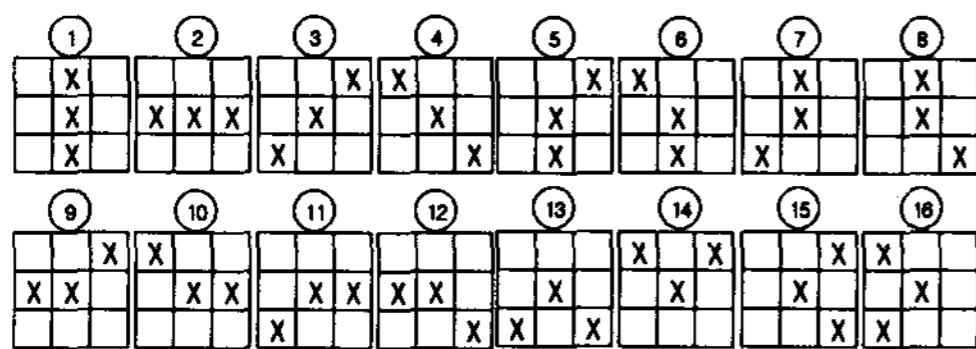


그림 2. 16종류의 에지 맵

16 종류의 에지 맵을 조합하여 텍스트 추출을 위해서 텍스트의 기본 단위인 문자의 특징을 추출하기 위해서 문자의 구조적 특징을 잘 반영하는 6종류의 문자-에지 맵의 조합을 생성하여 텍스트 영역 추출 과정에서 특징으로 사용한다.

2. 문자-에지 맵 추출

텍스트 영역 추출을 위해서 16종류의 에지 맵을 분석함으로써 텍스트 후보 영역을 검출할 수 있다. 그림 3은 텍스트 영역의 에지구조를 보여준다.

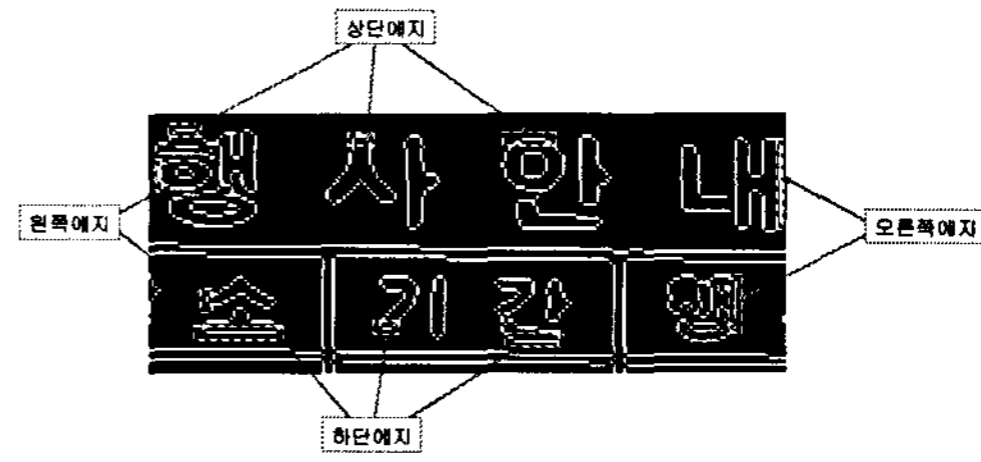


그림 3. 텍스트 영역 에지 구조

그림 2에서 텍스트 영역에 존재하는 여러 가지 에지 구조를 볼 수 있다. 따라서 텍스트 영역을 추출하기 위해서 이러한 텍스트 영역에 존재하는 6종류의 문자-에지 맵의 조합을 추출하였다. 그림 4는 6종류의 문자-에지 맵을 보여준다.

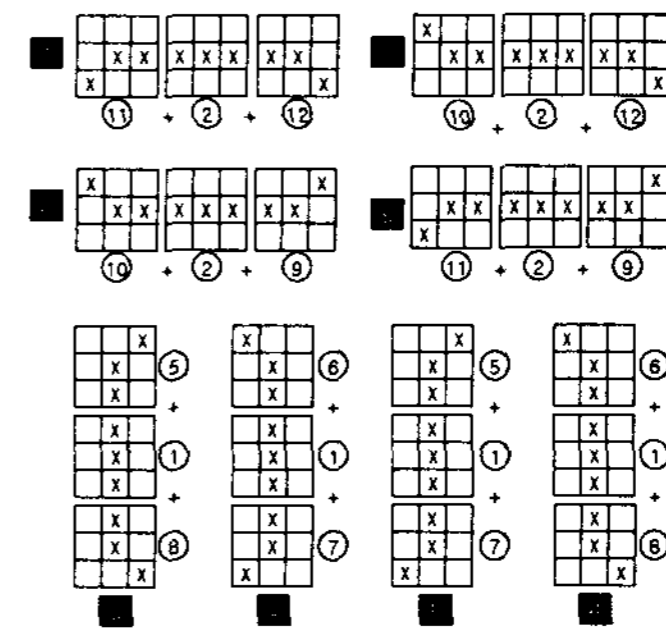


그림 4. 문자-에지 맵

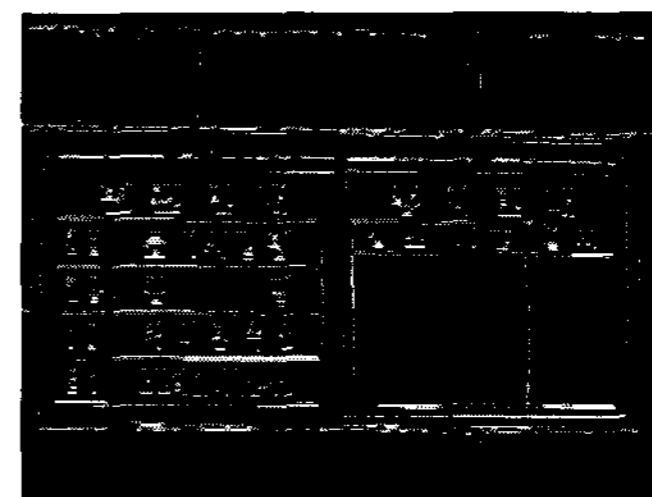


그림 5. 상하단 문자-에지 맵 추출

문자-에지 맵은 문자에지의 위치에 따라서 모두 "1" 부터 "8"까지의 번호를 부여하였다. 텍스트 상단에지는 텍스트 영역 상단부분에 존재하는 에지 형태로 각각의 글자에서 한번이상 검출되는 것을 볼 수 있다. 그러나 텍스트 영역이 아닌 영역에도 비슷한 텍스트 상단 및 하단 에지가 검출된다. 일반적으로 텍스트 영역에는 6종류의 문자-에지가 강한 연결성을 보여주고 있다. 그러나 텍스트 영역이 아닌 부분에서도 여러 종류의 텍스트 에지가 존재한다. 그러나 각각의 에지들 간의 연결성은 강하지 않다. 따라서 이러한 텍스트에

존재하는 에지들 간의 강한 연결성을 특징으로 텍스트 영역을 검출한다.

3. 텍스트 후보 영역 검출

텍스트 후보 영역을 검출하기 위해서 텍스트 영역의 상단 부분을 찾는다. 텍스트 영역의 상단 부분은 그림 7에서 문자-에지 맵 “1”번의 형태를 갖는다. 텍스트 형태에 따라 텍스트 상단에는 같은 수평 좌표에서 검출되지 않고, 연속적으로 몇 개의 수평좌표 영역에 걸쳐 나타나게 된다. 그러므로 아래와 같은 휴리스틱 조건을 설정하여 텍스트 상단 영역 검출 수행한다.

- 조건 1. 텍스트 상단에지 좌우 끝에서 수직에지와 연결성을 적어도 1개 이상 갖고 있어야 한다.
- 조건 2. 텍스트 상단에지의 최소 에지 수는 3이상이어야 한다.
- 조건 3. 텍스트 상단 에지 특징의 평균 개수가 2 이상이어야 한다.
- 조건 4. 여러 개의 텍스트 상단 에지 중에서 최상단 에지 위 부분 영역의 에지의 분포가 전체 이미지 너비의 10% 미만인 것으로 한다.

4. 텍스트 영역 추출

텍스트 영역은 그림 4에서 제시한 6종류의 문자-에지 맵으로 구성된다고 볼 수 있다. 그러므로 텍스트 영역에 존재하는 문자-에지 맵을 검출함으로써 텍스트 영역을 추출한다. 텍스트 영역 추출을 위한 휴리스틱을 아래와 같은 조건을 만족하는 것을 기준으로 하였다.

- 조건 1. 텍스트 영역의 가로/세로 비율에 따른 글자 수에 따른 빈 공간이 존재해야 한다.
- 조건 2. 텍스트 영역의 레이블 처리 결과 텍스트 영역에 존재하는 글자 수에 비례하여 레이블 수가 최소 임계값 보다 커야 한다.
- 조건 3. 텍스트 영역의 레이블 처리 결과 텍스트 영역에 존재하는 글자 수에 비례하여 레이블 수가 최대 임계값 보다 작아야 한다.

조건 1은 문자와 문자 사이에는 공백이 존재하는 특징을 반영한 것이고, 조건 2는 텍스트 영역은 에지의 조합에 의해서 형성되므로 에지의 수가 최소값 보다 커야 하고, 조건 3은 텍스트 영역의 레이블 수는 최대 임계값 보다 작아야 한다. 그러므로 텍스트 영역에 존재하는 공백과 레이블 수의 최소 및 최대값을 갖고 텍스트 영역을 검증할 수 있다.

III. 실험 및 결과분석

1. 실험 환경

본 연구에 사용된 장면 이미지는 올림푸스 CW-5060 디지털 카메라로 촬영된 이미지를 사용하였고, 실험에 사용된 이미지는 도로의 교통 표지판, 간판, 안내표지판 등 총 150개의 다양한 종류의 장면 이미지를 대상으로 하였다. 실험 이미지의 해상도는 640×480이고, 사용된 컴퓨터 환경은 펜티엄4-2.0GHz 이고 주기억 용량은 512MB이다.

2. 실험 및 성능 평가

성능평가 결과는 표 1과 같은 텍스트 영역 추출률을 나타내었다. 표 1 텍스트 추출률에서 Hit rate, Miss rate, False alarm 은 아래와 같이 정의하였다.

$$\text{Hit rate} = \frac{\text{검출된 텍스트 박스의 수}}{\text{이미지에 존재하는 텍스트의 박스의 수}} \times 100$$

$$\text{Miss rate} = 100 - \text{hit rate}$$

$$\text{False alarm rate} = \frac{\text{잘못 검출된 텍스트 박스의 수}}{\text{이미지에 존재하는 텍스트의 박스의 수}} \times 100$$

그림 7에서 텍스트 영역 추출 결과를 보여준다. 텍스트 영역 추출의 정확성을 떨어뜨리는 경우는 텍스트

표 1. 텍스트 영역 추출률

장면 이미지	Number of boxes in Images	Hit rate	False alarm rate	Miss rate
책 표지	10	85%	5%	15%
간판	50	75%	10%	25%
안내판	40	87%	10%	13%
교통표지판	50	85%	7%	15%
전체 이미지	150	83%	8%	17%

영역 주변에 에지가 복잡한 경우로 텍스트 영역의 정확한 세로 좌표를 구하지 못하지 때문이다. 그러나 텍스트 영역 주변이 단순한 경우에는 텍스트 영역을 정확히 추출한다. 전체적으로 본 연구에서 제안한 방법은 에지 주변의 에지들의 복잡도에 따라 영향을 받는다.

IV. 결론

본 연구에서는 장면 이미지로부터 텍스트 영역을 추출하는 방법을 제안하였다. 텍스트 영역에 존재하는 에지의 형태를 분석하고, 텍스트 영역에 존재하는 에지 맵의 조합을 생성하여 문자-에지 맵의 특징만을 사

용하여 텍스트 영역을 추출할 수 있었다. 따라서 본 논문에서 제안한 텍스트 영역 추출 기법은 에지의 구조적인 특징만 사용하여 텍스트 영역을 추출할 수 있다는 장점을 갖고 있다.

향후 연구과제는 보다 정확한 텍스트 사각형 영역을 검출하기 위해서 문자-에지 맵을 일반화하고, 텍스트 영역을 정확히 검증할 수 있는 방법을 도출하는 것이다. 그리고 연결요소 성분 분석 방법과 효과적으로 병합하는 방법도 연구할 과제이다.

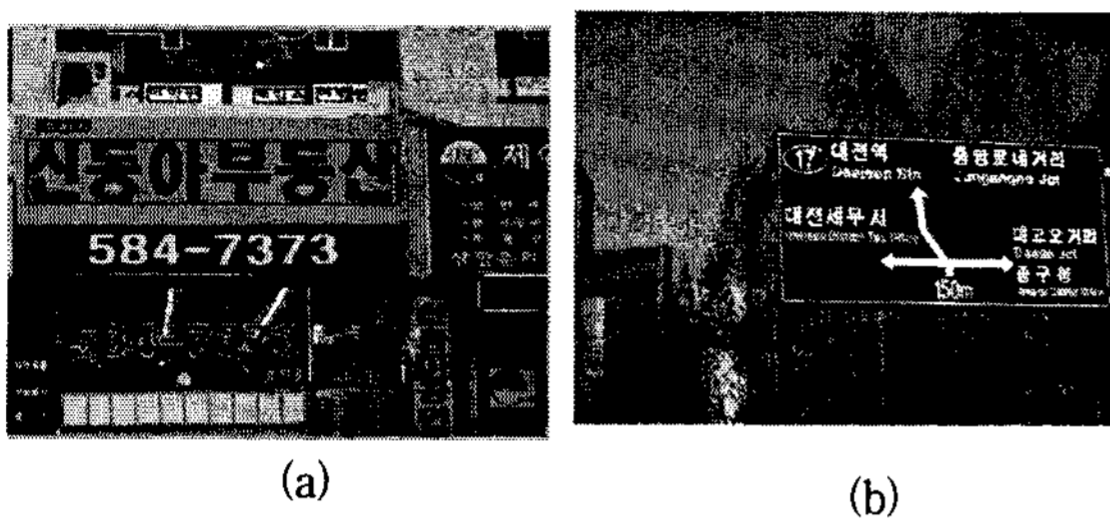


그림 6. 텍스트 추출 결과

참 고 문 헌

- [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, 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, 2101-2104, 2002.
- [3] K. Jung, K. I. Kim and A. K. Jain, "Text Information Extraction in Images and Video: A Survey", Pattern Recognition, Volume 37, Issue 5, 977-997, 2004.
- [4] A. K. Jain, B. Yu, "Automatic Text Location in Images and Video Frames," pattern Recognition, Vol. 31, No. 12, 2055-2076, 1998.
- [5] J. Ohya, A. Shio, S. Akamatsu, "Recognizing Characters in Scene images," IEEE Transactions on PAMI, Vol. 16, No. 2, 67-82, 1995.
- [6] 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, July 1995.
- [7] D. Chen, K. Shearer, and H. Bourlard, "Text Enhancement with Asymmetric Filter for Video OCR", Proc. of International Conference on Image Analysis and Processing, 192-197, 2001.
- [8] Y.M.Y. Hasan, L.J. Karam, "Morphological Text Extraction from Images", IEEE Transactions on Image Processing, Volume 9 issue(11), 1978-1983, 2000.
- [9] W. Jiang, Q. Shao-Lin, Q. Zhuo, W. Wen-Yuan, "Automatic text detection in complex color image", Machine Learning and Cybernetics, 2002. Proceedings. 2002 International Conference on Volume 3, 1167-1171, 2002.
- [10] 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, 498-507, 2004.
- [11] J. Canny, "A Computational approach to edge detection", IEEE Trans. Pattern Anal. and Machine Intelligence, 8(6) 679-698, 1986.
- [12] 김수겸, 장유정, "에지의 구조적 정보를 이용한 에지추출", 정보처리학회 논문지, Vol.3 No.5, 1337-1345, 1996.