

# 색기반 이진화를 이용한 장면 텍스트 추출과 써포트 벡터머신을 이용한 텍스트 영역 검증

장 대근\*, 김 의 정\*\*  
특허청 정보통신심사본부\*,  
공주대학교 컴퓨터교육과 (교신저자)\*\*

Scene Text Detection Using Color-Based Binarization  
and Text Region Verification Using Support Vector Machine  
The Korean Intellectual Property Office  
Department of Computer Education Kongju National University\*\*

## 요 약

기존의 텍스트 추출을 위한 이진화 방법은 입력 이미지를 명도 이미지로 변환한 뒤 이진화 하는 방법을 사용하였다. 이러한 방법은 칼라 이미지에서는 극명히 구분되는 색이라 할지라도 명도 이미지로 변환하는 과정에서 같은 밝기를 갖게 되는 경우(예를 들어, 배경은 붉은색, 텍스트는 초록색), 텍스트를 추출하는 데 어려움이 있다.

본 논문에서는 이러한 문제를 해결하기 위해 입력 이미지를 R, G, B로 분리하고 각각을 이진화 하여 텍스트를 추출하고 다해상도 웨이블릿(Wavelet) 변환을 이용하여 텍스트의 획 특징을 추출하여 추출된 특징들을 SVM(Support Vector Machine) 분류기로 검증하여 최종 텍스트 영역을 확정한다. 제안한 방법을 적용함으로써 명도 정보만으로는 추출하기 어려웠던 텍스트 영역을 효과적으로 추출하고 텍스트와 구별하기 어려운 영역을 획수준으로 검증할 수 있었다.

## 1. 서 론

이미지나 동영상에 인위적(Artificial)으로 삽입되거나 자연적으로(Natural) 포함된 텍스트들은 이미지의 내용을 함축적이고 구체적으로 표현하는 중요한 정보들이다. 이러한 정보들을 실시간에 추출하여 시각 장애인 보행안내, 로봇 자동주행 등과 같은 분야에서 활용할 수 있다. 문서 영상에서의 문자 추출은 많은 연구가 수행되어 이미 상용화된 제품들이 있는 반면, 일반적으로 자연영상 내에 포함되는 문자들은 해상도가 낮고 글자 형태와 크기, 색상 등이 다양하기 때문에 이를 추출하고 인식하는 일은 어려운 문제로 알려져 있다. 최근에는 다양한 자동화 시스템 개발의 활성화로 복잡한 배경을 갖는 자연영상에서 문자추출의 필요성이 대두되어 텍스트의 크기, 색, 폰트, 방향 등이 다양하게 나타나는 장면 텍스트를 추출하는 연구가 활발하게 진행되고 있다.

장면텍스트 추출 연구로서 Ohya 등은[1] 실외 자연이미지의 텍스트 영역을 추출하기 위해서 대부분

의 텍스트는 명도 값이 일정하다는 특징을 이용하였다. 지역적 이진화(Local Thresholding) 방법을 적용하여 텍스트 요소를 검출하고, 텍스트 요소들의 유사성을 계산한 후 완화(Relaxational) 방법을 이용하여 텍스트 요소들을 병합시켰다. 이정표, 자동차 번호판, 간판이미지를 대상으로 실험한 결과 단순한 이미지에서는 좋은 결과를 얻었지만, 배경이 복잡하거나 조명 변화가 있는 이미지에서는 오류가 많이 발생하는 단점이 있었다.

Gu 등은[2] Differential top-hats 연산자를 이용한 추출 방법을 제안하였으며, 문자 획을 크기에 따라 8단계로 나누고, 백색의 문자와 검은색의 문자로 구분하여 연산자를 적용하였다. 제안한 알고리즘은 조명의 변화가 있는 오전과 저녁 이미지에서는 오류가 많이 발생하였으며, 또한 처리 시간이 오래 걸리는 단점이 있다.

앞서 살펴본 것처럼 장면 텍스트를 추출하기 위한 대부분의 방법들은 색 이미지에서의 색 연속성 특징을 이용한 색 병합 방법의 시도와 명도 이미지

에서는 명도 값이 균일하다는 특징을 이용하기 위한 이진화 방법을 시도하였다.

본 논문에서는 기존의 명도 정보만을 이용했던 이진화 방법에 색 정보도 함께 고려하여 텍스트를 추출하는 방법을 제안한다. 본 논문에서 제안하는 방법은 다음과 같은 단계를 거친다. 컬러 이미지가 입력되면 컬러 이미지를 R, G, B 세 개의 이미지로 분리한다. 각각의 이미지에서 Otsu 연산을 수행하고, 이진화된 이미지의 반전된 영상도 각각 생성하여 총 6개의 이진 이미지를 얻는다. 6개의 이미지 각각에 모폴로지 불립연산을 적용하여 인접한 연결요소들을 연결시킨 후, 연결요소를 분석하고 검증하여 텍스트 후보 영역을 추출한다. 마지막으로 6개의 이미지 각각에서 추출된 후보영역을 적절히 결합하여 텍스트 영역을 최종 확정한다. 그림 1은 제안하는 방법의 개략도이다.

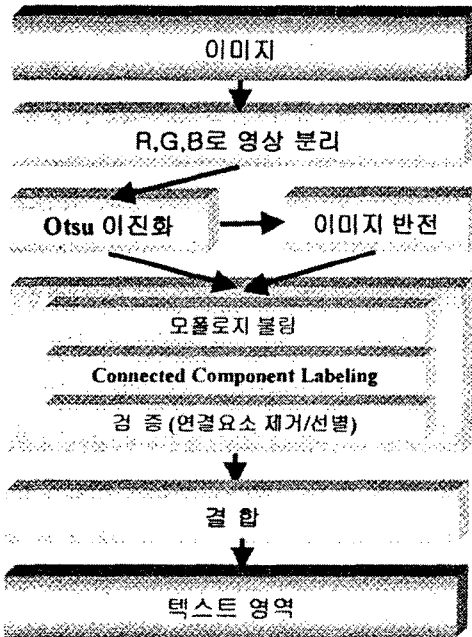


그림 1. 제안한 방법

## 2. 텍스트 추출

### 2.1 Otsu 이진화

입력된 이미지는 칼라 이미지로써 먼저 R, G, B 이미지로 분리한다. Red와 Green은 명도 이미지에서 서로 같은 밝기 값(intensity)을 가지므로 배경과 텍스트가 Red와 Green으로 구성되어 있는 이미

지의 경우, 명도 정보만으로는 텍스트를 추출하기 어렵다.

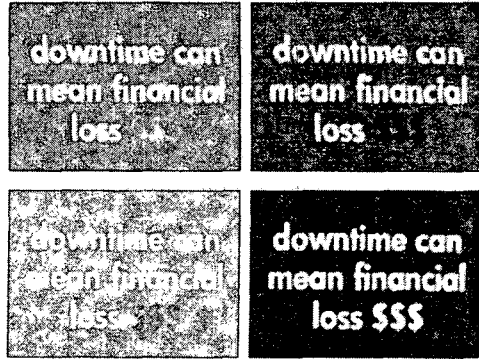


그림 2. (시계방향으로) 원이미지, R 이미지, G 이미지, B 이미지

전역적 이진화 방법으로는 히스토그램을 분석하여 임계값을 찾는 방법이 가장 보편적으로 이용되고 있으며, 그 중에서도 Otsu 등이 제안한 이진화 방법이 가장 좋은 성능을 보이는 것으로 알려져 있다. 본 논문에서도 분리된 각각의 R, G, B 이미지에 Otsu 이진화를 사용하였다. 그림 3은 Otsu 이진화의 결과이다. Green 배경의 Red 텍스트가 R, G 이미지에서는 손실되지만 B 이미지에서는 텍스트로 배경과 분리되어 이진화 되었음을 알 수 있다

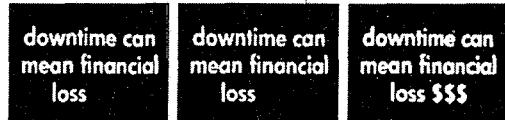


그림 3. (왼쪽부터) R, G, B Otsu 이진화 결과

다음 예제와 같이 텍스트가 배경보다 밝은 색의 픽셀인 경우, 이진화 과정에서 텍스트가 배경으로 잘못 인지되는 오류가 발생한다. 이러한 오류를 방지하기 위해 Otsu 이진화 이미지의 반전된 이미지를 생성하여 총 6개의 이진화 이미지를 얻게된다.



그림 4. 텍스트가 배경보다 밝기 값이 낮은 경우의 이진 이미지와 반전 이미지

## 2.2 연결요소 분석, 검증 및 텍스트 추출

이진화 이미지에서 텍스트 영역을 강조하기 위해 모폴로지 불립연산을 적용한다.

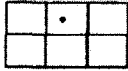


그림 5. 모폴로지 불립연산을 위한 구조연산자

불립연산을 적용한 이진 이미지에서 연결요소의 픽셀수가 매우 많거나 적은 것은 대부분 나뭇잎이나 타일 등과 같은 조밀한 배경인 경우가 대부분이며 아주 작은 경우는 비록 텍스트 영역이라 하더라도 사람 눈으로 구분하기 어렵기 때문에 제거한다. 또 외곽사각형에서 연결요소가 차지하는 비율, 외곽사각형의 세로/가로 비율, 폭 또는 높이 등이 아주 작은 것들도 추가로 제거한다.

## 2.3 결합

R, G, B 이미지와 각각의 반전된 이미지에서 추출된 텍스트 영역들을 적절히 결합하여 텍스트 추출의 정확성을 향상시키고자 한다. 본 논문에서는 각각의 이미지에서 추출된 영역들의 겹침 정도를 비교하여 텍스트 영역을 결정하였다. 그림 6의 바운딩 박스 영역은 최종적으로 텍스트로 추출된 영역이다.



그림 6. 텍스트 추출 결과

## 3. 결론 및 향후과제

본 논문에서는 기존의 명도 정보만을 이용했던 이진화 방법에 색 정보도 함께 고려하여 텍스트를 추출하는 방법을 제안하였다. 제안한 방법은 기존의 명도 정보만을 이용했던 텍스트 추출 방법보다

조명에 강한 결과를 보였다. 그러나 텍스트의 형태가 아니라 이진화 결과에만 의존하였으므로 복잡한 배경에서 텍스트가 아닌 영역을 텍스트로 추출하는 오류가 많았다. 따라서 향후에는 이진화 뿐만 아니라 텍스트의 에지 정보나 획 정보 등 다른 특징들을 이용한 방법을 같이 이용한 텍스트 추출성능의 향상이 필요하다.

## 참고문헌

- [1] J Ohya, A. Shio, S. Akamatsu, "Recognizing Characters in Scene images," IEEE Transactions Pattern Analysis and Machine Intelligence, PAMI-16(2), pp. 67-82, 1995.
- [2] Lixu Gu, Toyahisa Kaneko, "Robust Extraction of Characters from Color Scene Image Using Mathematical Morphology", Proceeding of 7th International Conference on Pattern Recognition, Vol. 2, pp. 1002-1004, 1998.
- [3] D. Doermann, J. Liang, and H. Li, "Process in camera-Based Document Images Analysis", Proc. of the ICDAR, 2003, pp359-363.
- [4] Yu Zhong, Kalle Karu, Anil K. Jain, "Locating Text in Complex Images," Pattern Recognition, Vol. 28, No. 10, pp. 1523-1535, 1995.