

실시간 글자 인식을 위한 안드로이드 기반의 글자 영역 추출 기술

*이규철 *이상용 *유지상

광운대학교

*gyucheol0116@gmail.com

A text region extraction algorithm based on Android for real-time text recognition

*Lee, Gyu-Cheol *Lee, Sangyong *Yoo, Jisang

*Kwangwoon University

요약

본 논문에서는 안드로이드 환경에서 글자 인식을 위한 전처리 과정으로 입력 영상에서 글자 영역만을 추출하는 기법을 제안한다. 대부분의 글자 인식 어플리케이션에서 글자를 인식하는 방법은 RoI(Region of Interest)에 인식하려는 글자를 위치시켜 놓고 사용자가 촬영함으로써 진행된다. 하지만 촬영된 영상 그대로를 인식에 사용하기 때문에 잡음 및 글자가 아닌 영역들을 글자로 인식하는 문제 등으로 인하여 인식률이 현저히 떨어진다. 제안하는 기법에서는 MSER(Maximally Stable Extremal Regions) 기법을 통해 각각의 글자를 추출한 후, 글자의 특성을 이용하여 글자 영역만을 추출한다. 기법의 성능 평가는 무료 OCR(Optical Character Recognition) 엔진인 Tesseract-OCR을 통해 글자 인식률을 비교하였으며, 제안하는 기법을 적용한 글자 인식 시스템이 적용하지 않은 시스템보다 글자의 인식률이 향상되는 것을 확인하였다.

1. 서론

매년 한국에 방문하는 외국인 관광객이 꾸준히 늘어감에 따라, 외래 관광객의 관광 지출을 확대하는 방안이 연구되고 있다. 특히 스마트폰으로 음식점이나 관광지 또는 시내에서 사용자가 번역을 원하는 글자를 카메라로 촬영하면 번역하여 제공하는 기술에 대한 연구가 활발히 진행 중이다. 영상 내의 글자를 인식하여 기계가 읽을 수 있는 문자로 변환하는 OCR(Optical Character Recognition) 기술은 위와 같은 연구에서 가장 핵심적인 기술로 주목받고 있다.

OCR 기술은 전처리, 인식 그리고 후처리 과정으로 나눌 수 있다 [1]. 전처리 과정은 인식 과정에서의 성능을 향상시키기 위해 입력된 영상을 영상 처리 알고리즘을 통해 보정하는 과정이다. 주로 사용되는 기술로는 입력 영상에서 발생할 수 있는 잡음을 제거하는 기술과 배경을 제거하여 인식 성능을 향상시키는 기술이 있다. 인식 과정은 전처리 과정을 거친 글자 영상을 컴퓨터가 읽을 수 있는 글자로 변환하는 단계이며 무료 글자 인식 엔진인 Tesseract-OCR이 주로 사용된다[2]. 후처리 과정은 인식 과정에서 발생한 에러를 교정하는 단계이다. 일반적으로 DB 사전을 만들어서 인식 과정의 결과와 DB간의 비교를 통해 결과를 보정하는 방법이 사용된다.

대부분의 문자 인식 어플리케이션은 문서를 대상으로 개발된다. 문서를 촬영한 영상은 비교적 균형이 잘 맞고 품질이 깨끗하여 인식이 잘 되는 편이다. 하지만 스마트폰으로 촬영한 영상의 경우 잡음이나 블러(blur)가 발생할 수 있으며 영상이 기울어질 수도 있다. 위와 같은 현상들은 전처리 과정에서 글자를 제대로 추출하지 못하여 글자 인식률을 떨어뜨리는 원인이 된다. 또한 스마트폰에서 사용되기 위해서는 어플리케이션이 실시간으로 동작되어야 하기 때문에 알고리즘의 복잡도

또한 고려되어야 한다.

본 논문에서는 스마트폰 카메라를 이용하여 글자를 촬영하면 배경 및 잡음들을 제거하고 글자 영역만을 추출하는 전처리 기술을 제안한다. 다운 샘플링(down-sampling)과 가우시안(Gaussian) 블러를 이용하여 영상의 잡음을 제거하고 MSER 기법[3]을 통해 글자라고 판단되는 후보 영역들을 추출한다. 마지막으로 글자의 특성을 이용하여 후보 영역들 중에서 글자라고 판단되는 영역을 최종적으로 결정한다.

본 논문은 다음과 같이 구성된다. 2장에서는 제안하는 기법에 대해 소개하고, 3장에서는 실험을 통해 성능을 평가한다. 마지막으로 4장에서는 결론을 맺는다.

2. 본론

제안하는 기법은 스마트폰으로 촬영한 영상에서 글자 영역만을 추출하는 전처리 과정이다. 입력 영상으로 스마트폰으로 촬영된 전체 영상을 사용하면 알고리즘의 처리량이 많아지기 때문에 실시간으로 동작하기 어려워진다. 또한, 사용자의 사용 특성상 영상 전체가 아닌 일부 영역의 번역만을 필요로 하기 때문에 그림 1과 같은 UI를 가지는 촬영 시스템을 설계한다. RoI에 번역을 원하는 글자를 위치시키고 촬영하여 RoI 크기만큼의 영상을 입력 영상으로 사용한다.

촬영한 영상에 서류와 같이 글자만 존재한다면 글자는 잘 인식되는 편이다. 하지만 일상생활에서 촬영한 영상의 경우 글자가 아닌 것들이 불가피하게 포함된다. 그림 1은 '해물전골'이라는 글자를 인식하기 위해 RoI를 해당 글자에 위치시킨 상태이다. RoI 상단에는 음식 그림에 해당되는 부분과 좌측에는 인식하려는 글자가 아닌 다른 글자가 의도하지 않게 포함되는 것을 보여주고 있다. 이와 같은 영상에 대하여

글자 인식을 수행하면 그림에 대해서 인식 알고리즘이 어떠한 글자와 일치하는지 해마다가 결국 엉뚱한 글자를 결과로 출력하게 된다. 또한 RoI 좌측의 글자가 인식되어 사용자에게 올바른 정보를 제공하지 못한다. 따라서 RoI에서 글자 영역만을 추출하여 인식을 진행하여야 한다.



그림 1. 촬영 시스템의 UI

스마트폰에서 촬영한 영상은 인식률을 떨어뜨리는 여러 가지 요인이 존재한다. 영상에 잡음이 존재하면 인식 과정에서 글자로 잘못 인식할 가능성이 있다. 사용자의 손 흔들림으로 인하여 발생한 블러 현상은 글자의 경계를 흐릿하게 하여 글자의 오인식을 초래한다. 위와 같은 현상을 방지하기 위해 영상의 품질을 개선하는 과정을 진행한다. 먼저 입력 영상에 짝수 번 짤 열과 행을 제거하여 1/4 크기로 다운샘플링된 영상을 획득한다. 다운샘플링은 영상의 잡음을 제거하고 연산량을 줄여줌으로써 알고리즘이 실시간으로 동작하게 한다. 다운샘플링시 발생하는 엘리어싱(aliasing) 현상을 제거하기 위해 가우시안 블러를 적용한다. 위 과정을 통해 영상에서 잡음들은 어느 정도 제거되고 인식하려는 글자와 배경 및 인식 대상이 아닌 글자들이 남아있게 된다.

글자는 형이상적으로 블랍(blob)의 형태를 가지고 있다. 블랍은 주변 환경과는 밝기, 색상 등으로 구분 지을 수 있으며, 블랍 안에 있는 점들은 서로 비슷한 특성을 가지고 있다. 블랍을 검출하는 여러 가지 알고리즘들 중에서 MSER은 영상에 블러가 있거나 불균일한 조명에서 강인한 결과를 보여주기 때문에 제안하는 기법에서는 MSER을 사용하여 글자를 추출한다. MSER 기법은 영상에서 유사한 값들이 안정적으로 연결되어 있는 영역을 검출한다. 안정적이란 말은 임계값을 0부터 255까지 천천히 변화시켰을 때 임계값에 따라 유사한 값들로 이루어진 영역의 크기가 달라지는데, 영역의 변화량이 최소가 되는 것을 의미한다. 식 1은 MSER 기법을 수식으로 표현한 것이다.

$$m(i) = \arg \min_i \frac{|M_{i+\Delta} \setminus M_{i-\Delta}|}{|M_i|} \quad (1)$$

여기서 i 는 임계값이며, M 는 임계값 i 에 따라 유사한 값들로 이루어진 영역을 의미한다. Δ 는 사용자가 지정하는 파라미터이다.

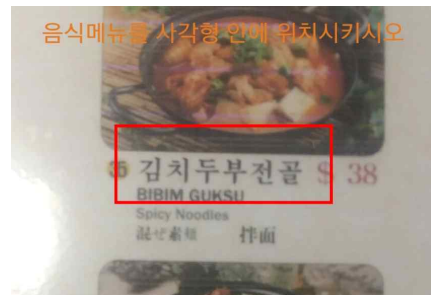
MSER 기법을 이용해 블랍을 추출하면 글자 영역과 더불어 배경의 일부도 블랍으로 판단하여 검출된다. 배경이 아닌 문자만을 추출하

기 위해 otsu 이진화 알고리즘[4]을 적용한 후 수평방향으로 모폴로지(morphology) 연산을 통해 같은 행에 있는 글자 영역들을 사각형 형태로 레이블링(labeling) 한다. 레이블링 한 영역들 중에서 넓이 대비 블랍이 차지하는 픽셀의 개수가 특정 값보다 큰 영역을 글자가 포함되어 있다고 판단한다.

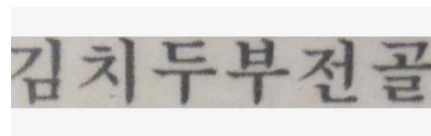
글자라고 판단된 후보들 중에서 사용자가 인식하려는 글자는 RoI의 가운데에 위치할 확률이 높다. 따라서 RoI의 중앙 수평선에 위치해 있는 글자 후보를 최종적으로 인식할 글자로 결정한다.

3. 실험 결과

본 논문에서 제안하는 글자 영역 추출 전처리 기술을 평가하기 위해 스마트폰 어플리케이션을 제작하였다. 개발환경의 운영체제는 Microsoft Windows 10이며 Android Studio를 이용하여 개발하였다. 실험에 진행한 스마트폰의 기종은 SAMSUNG Galaxy A5이며, 운영체제는 Android 4.0 Ice Cream Sandwich이다. 그림 2는 개발한 어플리케이션을 이용하여 영상을 촬영하고 최종적으로 검출된 글자 영역을 보여준다. RoI 상단의 배경과 좌우의 글자가 제거되고 인식하려는 글자만이 추출된 모습을 보여주고 있다.



(a)



(b)

그림 2. 글자 영역 추출 결과

성능을 평가하기 위해 제안하는 기법을 사용할 때와 그렇지 않았을 경우의 인식률을 비교하는 실험을 진행하였다. 글자 인식은 무료 OCR(Optical Character Recognition) 엔진인 Tesseract-OCR을 이용하였다. 표 1은 각 경우의 인식률을 보여준다.

표 1. 성능 평가 비교

	전처리 미적용	전처리 적용
인식률	48.4%	65.3%

성능 비교 결과 제안하는 글자 영역 추출 기술을 사용하였을 때의 인식률이 그렇지 않은 경우에 비해 인식률이 17%가량 높아지는 것을 확인할 수 있다. 배경 및 잡음 같은 요소들은 인식 대상에서 제외되고 글자 영역만을 추출하여 인식을 진행하기 때문에 인식률이 향상된다.

또한 실험에서는 후처리 알고리즘을 적용하지 않았기 때문에 60%대에 머무는 인식률을 보여준다. 후처리 과정을 추가한다면 더 높은 인식률을 보여줄 것으로 판단된다.

4. 결론

본 논문에서는 스마트폰 환경에서 OCR을 위한 전처리 기술로서 입력 영상에서 글자 영역만을 추출하는 기법을 제안하였다. 사용자의 사용 특성 및 시스템의 처리 속도를 고려하여 카메라 영상의 가장자리에 위치한 붉은색 사각형 모양의 RoI를 입력 영상으로 이용하였다. 다운샘플링과 가우시안 블러를 삽입하여 영상의 품질을 개선한 후에 글자 영역을 추출하기 위해 MSER 기법을 이용하여 영상 내의 모든 블랍을 추출하였다. 블랍들을 레이블링 한 이후 글자라고 판단되는 조건을 만족하지 않은 영역들을 제거하였다. 남아있는 영역들 중에서 RoI의 중앙 수평선에 걸쳐있는 영역을 인식할 글자로 최종 결정하였다.

제안한 기법의 성능을 평가하기 위해 Tesseract-OCR을 이용하여 글자 인식을 수행하였다. 실험 결과 제안한 글자 영역을 추출하는 전처리 기술을 사용하면 그렇지 않은 경우에 비해 인식률이 상승하는 것을 확인할 수 있었다.

참고 문헌

- [1] J. H. Roh and D. S. Choi, "Character recognition using a target string," Proceedings of Symposium of the Korean Institute of communications and Information Sciences, Jun. 2014.
- [2] Tesseract-OCR. Available: <https://github.com/tesseract-ocr>
- [3] M. Donoser and H. Bischof "Efficient maximally stable extremal region (MSER) tracking," IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2006, Jun. 2006.
- [4] N. Otsu, "A threshold selection method from gray-level histograms," IEEE Transactions on Systems, vol. 9, no. 1, pp. 62-66, Jan. 1979.