

명도 정보와 분할/합병 방법을 이용한 자연 영상에서의 텍스트 영역 추출

(Text Region Extraction of Natural Scene Images using
Gray-level Information and Split/Merge Method)

김 지 수 [†] 김 수 형 ^{**} 최 영 우 ^{***}
(Ji-Soo Kim) (Soo-Hyung Kim) (Yeong-Woo Choi)

요 약 본 논문에서는 자연 이미지에 포함되어 있는 텍스트를 추출하기 위해 명도 정보를 사용한 하 이브리드 분석 방법(HAM)을 제안하였다. 즉, 제안한 방법은 명도 정보 분석(Gray-intensity Information Analysis)과 분할/합병 분석(Split/Merge Analysis)을 결합하였다. 제안한 방법의 추출 결과를 보면 단순 한 영상과 복잡한 영상 모두에서 기존의 연구 결과보다 우수한 성능을 보임을 알 수 있었다.

키워드 : 장면 텍스트 영역 추출, 분할/합병 세그멘테이션, 명도 기반 영상처리

Abstract In this paper, we propose a hybrid analysis method(HAM) based on gray-intensity information from natural scene images. The HAM is composed of GIA(Gray-intensity Information Analysis) and SMA(Split/Merge Analysis). Our experimental results show that the proposed approach is superior to conventional methods both in simple and complex images.

Key words : Scene Text Extraction, Split/Merge Segmentation, Intensity-Based Image Processing

1. 서 론

최근 고도 정보화 사회로의 급진적 발전 추세에 따라 개인용 컴퓨터에 화상 카메라, 디지털 카메라, 디지털 캠코더, 카메라 폰 등의 부착 및 이들을 통한 정지 영상 및 동영상의 생성이 보편화되고 있다. 이러한 영상 내의 텍스트 정보들을 추출 및 인식 할 수 있다면 시각 장애인 보행 보조 시스템, 로봇 자동 주행 시스템, 비디오 자동 검색 및 색인 시스템, 텍스트 자동 번역 시스템, 이미지와 함께 포함되어 전송되는 스팸 메일 필터 등과 같은 다양한 분야에서 널리 활용될 수 있다.

지금까지의 문자 인식에 관련된 대부분의 연구들을 보면 문서 이미지에서의 텍스트 영역과 그림 영역 분리 및 제한된 환경에서 얻어진 이미지에서의 장면 텍스트 추출이 주류를 이루었다. 그러나 최근에는 복잡한 배경

과 여러 종류의 잡영들이 포함되어 있는 이미지에서의 장면 텍스트 추출에 관한 연구가 많이 진행되고 있다 [1-20].

장면 텍스트 추출에 관한 연구는 크게 명도 정보를 이용한 텍스트 추출 연구와 색 정보를 이용한 텍스트 추출 연구로 분류할 수 있다. 본 논문에서는 명도 정보를 이용한 장면 텍스트 추출에 관한 연구를 주 대상으로 한다. 기존의 명도 정보를 이용한 장면 텍스트 추출의 단점으로는 복잡한 배경 영상에서의 텍스트 추출률이 많이 낮다는 점을 들 수 있는데, 이에 본 논문에서는 3가지의 장면 텍스트 추출 방법을 제안하여 명도 정보를 이용한 장면 텍스트 추출의 단점을 개선하였다. 첫 번째 제안한 방법은 그레이 값 스트레칭과 전체 이미지 그레이 값의 평균값을 이용한 이치화로 복잡한 배경 영상에서의 불필요한 부분의 제거에 의한 장면 텍스트 영역 추출이며, 두 번째 제안한 방법은 이미지 세그멘테이션 방법 중의 하나로 널리 사용되고 있는 분할(Split)/합병(Merge) 방법을 적용한 장면 텍스트 추출 방법이다. 마지막 세 번째 방법으로는 위의 2가지 모델을 조합한 방법이다.

본 논문의 구성은 다음과 같다. 먼저 2절에서는 관련 된 기존 연구를 소개하며, 3절에서는 제안하는 방법을

· 본 연구는 2004년 한국학술진흥재단 선도연구자 지원사업(2004-041-D00631)에 의하여 수행되었음

† 비 회 원 : 전남대학교 전산학과
kimjisoo@iip.chonnam.ac.kr

** 종신회원 : 전남대학교 전자컴퓨터정보통신공학부 교수
shkim@chonnam.ac.kr

*** 비 회 원 : 숙명여자대학교 컴퓨터학과 교수
ywchoi@sookmyung.ac.kr

논문접수 : 2004년 1월 30일

심사완료 : 2005년 4월 8일

구체적으로 기술하며, 4절에서는 실험 환경 및 결과를 실었으며, 5절에서는 결론을 맺는다.

2. 관련 연구

장면 텍스트 추출에 관한 관련 연구를 살펴보면 크게 명도 정보를 이용한 장면 텍스트 추출과 컬러 정보를 이용한 장면 텍스트 추출로 나눌 수 있다. 명도 정보를 이용한 장면 텍스트 추출에 관한 연구를 살펴보면 다음과 같다. Jain[1] 등은 입력 이미지로 정지 영상, 웹 이미지, 동영상 사용하였다. 전처리로는 다중 값(Multi-valued) 이미지 분해에 의한 9단계의 지역적 편차 이미지를 구하게 되며, BAG 알고리즘과 연결 요소 분석을 통한 텍스트 영역을 추출하였다. Hoyal[2] 등은 입력 이미지로 그레이 정지 영상을 사용하였다. 전처리로는 8×8 크기의 윈도우(단위 : 픽셀)를 이용한 지역적 이치화 방법을 사용하였으며, 인접 영역의 명도 차를 고려하여 텍스트 영역을 추출하였으나, 이미지 전체에 나타나는 불필요한 에지 값들 때문에 다양한 종류의 연산과 시간이 많이 걸린다는 단점이 존재하였다. Lienhart[3] 등은 입력 이미지로 동영상 프레임에 입력 이미지로 사용하였다. 전처리로는 분할/병합 알고리즘을 적용하였으며, 블록 매칭과 대비(Contrast)분석을 이용한 텍스트 영역을 추출하였다. Messelodi[4] 등은 입력 이미지로 이진 정지 영상, 그레이 정지 영상을 사용하였다. 전처리로는 지형적 특징, 지역적 대비 특징 및 필터들을 사용하였다. 추출되어진 텍스트 영역 각각의 중심점과 투영된 프로파일을 이용하여 텍스트의 기울기를 추정하는 방법 등을 제안하였다. Zhong[5] 등은 입력 이미지로 그레이 정지 이미지를 사용하였다. 전처리로는 1×N 크기를 갖는 윈도우를 사용하여 지역 공간적 변화 값을 계산한다. 텍스트 영역은 명도 값의 공간적 변화가 큰데 전처리 단계에서 구해진 특징 값을 이용하여 수평 방향의 분산 이미지를 구하는 방법을 제안하였다. 이 방법은 수직 방향의 분산 이미지를 구하지 못하는 점 때문에 세로 방향의 문자열은 찾지 못하는 단점이 존재하였다. Choi[6] 등은 입력 이미지로 컬러 정지 영상을 사용하였다. 전처리로는 컬러 이미지를 그레이 이미지로 변환 한 후, Median 필터와 Canny 에지 필터를 적용하여 텍스트 후보 영역을 찾은 후, 긴 선 제거 및 모폴로지 연산을 적용한 후에 텍스트 영역 후보를 추출하였다.

색 정보를 이용한 장면 텍스트 추출에 관한 연구를 살펴보면 다음과 같다. Zhong[5] 등은 입력 이미지로 컬러 정지 영상(자연색)을 사용하였다. 전처리로는 컬러 히스토그램을 계산한 후에 비슷한 컬러 영역별로 분할/합병 알고리즘을 적용하였다. 또한 전처리 부분에서 지워져 버린 텍스트 영역을 복원한 후에 텍스트 영역을

추출하는 방법을 제안하였다. Choi[6] 등은 입력 이미지로 컬러 정지 영상을 사용하였다. 전처리로는 17×1 수직 윈도우 및 1×21 수평 윈도우를 사용하여 수직 및 수평 방향의 컬러 변화(Variation) 영상 2개를 구한 후에, 2개의 이미지를 논리 곱 연산 및 모폴로지 연산을 수행한 후에 텍스트 영역을 추출하였다. 제안한 다른 방법으로는 입력 이미지로 컬러 정지 영상을 사용하였다. 전처리로는 입력 이미지로 들어온 컬러 영상의 R·G·B 값 3바이트 영상(24비트 자연 색상)을 각 바이트(R·G·B)의 낮은 6비트를 제거하게 되면, 64 컬러를 갖는 군집화된(Clustering) 영상으로 바뀌게 된다. 다시 이 영상을 가지고 비슷한 컬러를 갖는 영역끼리 합병 한 후에 남아 있는 영역에서 텍스트 영역을 추출하였다. 장인영[7] 등은 입력 이미지로 컬러 정지 영상(자연색)을 사용하였다. 전처리로는 컬러 이미지를 그레이 이미지로 변환 후에 명암 대비 스트레칭과 적응적 임계값을 적용하였으며, 모폴로지노입(Erosion)을 적용한 영상과 모폴로지(열림 닫힘[OpenClose] + 닫힘열림[CloseOpen]) / 2가 적용된 영상 사이의 차이를 이용한 텍스트 영역을 추출하였다. Gu[8] 등은 입력 이미지로 컬러 정지 영상(자연색)을 사용하였다. 전처리로는 컬러 이미지를 그레이 이미지로 변환 후에 미분 탑 햇츠(Differential Top-hats)라는 모폴로지 분할 알고리즘과 방향성 필터(Directional Filter)를 사용하여 텍스트 영역과 잡영 영역을 분리하는 방법을 적용하여 텍스트를 영역을 추출하였다.

지금까지 살펴본 기존의 연구들은 대부분 배경이 단순한 이미지를 대상으로 실험하였으며, 텍스트의 크기나 위치, 기울기 등에 심한 제약을 받는 단점이 존재한다.

3. 제안 방법

본 절에서는 장면 텍스트 추출에 관한 3가지의 제안한 방법에 대해서 기술한다.

3.1 명도 정보 분석(Gray-level Information Analysis)

GIA에서 제안하는 명도 정보를 이용한 텍스트 추출 방법은 그림 1과 같이 크게 2단계로 구성되었다. 첫 번째 단계인 전처리 단계에서는 입력 이미지를 컬러 이미지로 받아들이고 후에 복잡한 배경 영상을 가지는 이미지의 영상으로부터 텍스트 에지 부분만을 남기고 배경 영상을 제거한다. 텍스트 후보 영역 추출 단계에서는 에지 이미지의 긴 선 요소들을 추출하여 제거한 후, 짧은 선 영역 및 전처리 단계에서 제거하지 못한 잡영들을 제거하게 된다. 마지막으로 연결 요소 분석 및 외곽 사각형 분석을 통해 워드 단위로 텍스트 영역을 추출하게 된다.

3.1.1 전처리

입력된 이미지(그림 2(a))는 다양한 사이즈의 컬러 이미지이다. 입력된 이미지는 먼저 식 (1)에 의해서 컬러

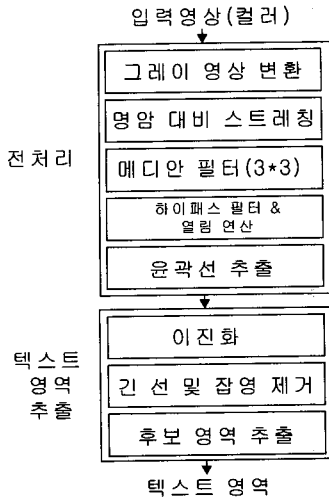


그림 1 GIA의 처리 과정

정보를 그레이 값으로 변환하게 된다[9].

$$Gray_Image = rgb.Red * 0.299 + rgb.Green * 0.587 + rgb.Blue * 0.114 \quad (1)$$

명도 정보를 통한 텍스트 추출의 단점은 복잡한 영상에서의 텍스트 추출이 어렵다는 것이다. 이러한 어려움을 극복하고자 변환된 그레이 이미지에 다시 이미지 스트레칭과 3x3의 Median 필터를 적용하여 원본 이미지에서 잡영이 일부 제거된 이미지(그림 2(b))를 만들게 된다. 이렇게 만들어진 이미지는 물체들의 에지 부분과

에지가 아닌 부분이 쉽게 구별되지 않는 단점이 있다. 이러한 점을 보완하기 위해 고주파 통과 필터를 적용하여 고주파 성분과 저주파 성분이 구별 될 수 있게 하였다(그림 2(c)). 또한 열림 연산을 수행하여 이미지 상에 존재하는 잡영들을 추가로 제거하였다(그림 2(d)). 이미지에서 에지 검출을 위해 라플라시안 에지 검출 방법을 적용하여 그림 2(e)와 같이 에지 이미지를 만든다.

3.1.2 후보 영역 추출

전처리를 한 후의 이미지에 대해서 전체 픽셀에 대한 평균값을 구한 후에 평균값을 기준으로 이진화를 수행하고 텍스트 영역 주변에 존재하는 긴 선 제거 및 잡영 제거를 수행 한 후에 텍스트 영역에 대한 후보 영역 추출 과정을 통해 워드 단위로 텍스트 영역을 추출한다.

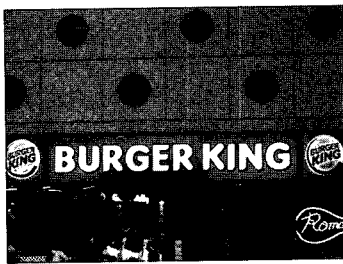
(1) 이진화

전처리를 하고 난 후의 이미지는 그레이 정보로 구성되어 있다. 따라서 긴 선 제거 및 짧은 선 제거 등 여러 가지 필터링 연산을 적용하기 위하여 그레이 이미지를 이진화된 이미지로 변경해야 한다. 그레이 이미지에서 이진화된 이미지로 변경하고자 할 때 전체 이미지의 임계값을 구하는 식은 식 (2)와 같다.

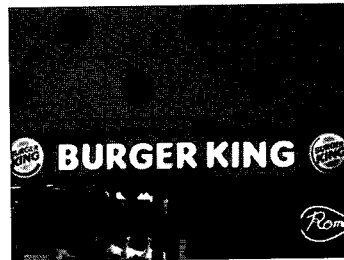
$$Mean = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} F(i, j) \quad (2)$$

(M = 이미지의 세로 크기, N = 이미지의 가로 크기) 식 (2)에서 구해진 임계값을 기준으로 전체 이미지의 픽셀을 이진화 한다.

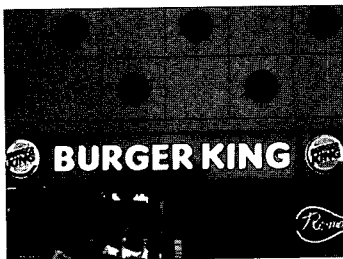
(2) 긴 선 및 잡영 제거



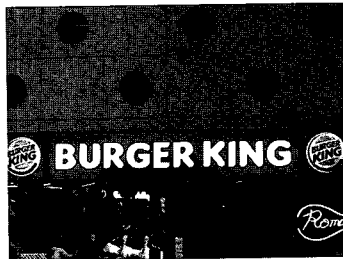
(a) 입력 이미지



(b) 스트레칭 및 미디언 필터 적용 후 이미지



(c) 고주파 필터 적용 후 이미지



(d) 열림 연산을 수행한 후의 이미지



(e) 에지 추출 후 이미지

그림 2 전처리 단계의 각 단계별 결과

이미지내의 텍스트 주변에는 텍스트를 둘러싸고 있는 네모난 박스나 긴 선 같은 요소들이 존재하는데, 이러한 특징들은 텍스트를 찾는데 많은 어려움을 주게 된다. 하지만 또 다른 면에서는 이미지 자체가 가지고 있는 기울기 및 원근 정보에 대한 정보를 취득할 수 있게 해준다. 앞서 나열한 특징들에서 필요한 정보를 취득 후에는 임계 값보다 큰 크기를 갖거나 작은 크기를 갖는 네모난 박스나 긴 선, 짧은 선, 꺾인 선 등의 잡영을 제거하고 나면 그림 3과 같다.

(3) 후보 영역 추출

긴 선과 잡영이 제거된 그림 3 이미지를 가지고 후보 영역에 대한 연결 요소 분석 및 외곽 사각형 분석을 수행한 후에 워드 단위의 텍스트 영역 추출을 수행하게 된다. 후보 영역을 추출하기 위해서는 먼저 가로 방향과 세로 방향에 대한 투영 프로파일에 대한 정보를 구하게 된다. 이 정보를 이용하여 연결 요소와 연결 요소의 외곽사각형의 위치, 크기, 중형비, 연결요소의 크기 등을 구하게 된다. 이때 이미지의 가로와 세로 길이에 비해 너무 큰 크기를 갖거나 작은 크기를 갖는 외곽 사각형들은 텍스트 후보 영역에서 제외하게 되며, 제외되고 남은 영역들은 텍스트 후보 영역으로 추출을 하게 된다. 다음으로는 문자 단위의 텍스트 후보 영역을 가지고 워드 단위의 텍스트 후보 영역을 만들기 위해서 외각 사각형 사이의 거리 및 외곽 사각형들의 포함 관계를 이용하여 워드 단위로 외각 사각형을 추출한다. 그림 4는 워드 단위로 추출된 텍스트 영역 이미지를 보여주고 있다.

3.2 분할 / 합병 분석(Split and Merge Analysis)

본 논문에서 제안하는 분할/합병 분석을 이용한 텍스트 추출 방법인 SMA의 처리절차는 그림 5와 같다. 입력된 이미지를 그레이 이미지로 바꾸어 분할과 합병 단계를 거치고 나면 주어진 영상은 비슷한 명도를 갖는 이미지 영역으로 분할 되게 된다. 이렇게 분할된 이미지 영역들의 크기가 원본 이미지의 가로와 세로 길이에 비해 너무 큰 크기를 갖거나 작은 크기를 갖게 되면 비

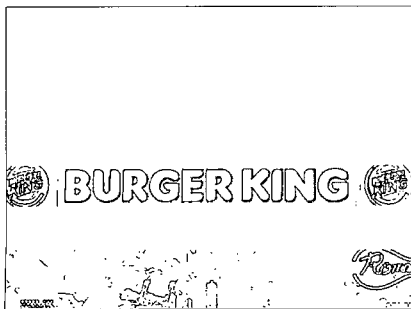


그림 3 긴 선 및 잡영이 제거된 이미지

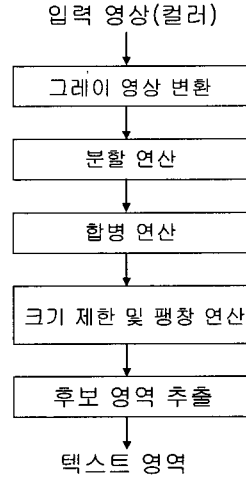


그림 5 SMA의 처리 과정

텍스트 영역으로 간주하고 지우게 된다. 이렇게 지우고 난 이미지를 이진화 및 채움 연산을 수행하고 연결영역 평균분석(Connected Region Mean Analysis)과 영역채움요소분석(Region Fill Factor Analysis) 연산을 수행하고 난 후에, 텍스트 영역 추출을 하게 된다.

3.2.1 분할 & 합병

분할/합병 연산은 영역 팽창 및 수축을 위해서 사용하는 대표적인 방법 중의 하나이다. 다음은 본 논문에서 사용한 분할/합병 과정을 설명한다.

(1) 분할

입력된 컬러 이미지는 식 (1)에 의해서 그레이 이미지로 변환하게 된다[9]. 변환된 그레이 이미지는 분할 과정에 의해서 전체의 이미지를 그림 6과 같이 4개의 작은 이미지로 반복적으로 나누게 된다. 다음은 본 논문에서 사용한 분할 알고리즘을 설명한다.

Step 1. Start with the entire image as a single region

Step 2. Pick a region R . If $H(R)$ is true, split



그림 4 워드 단위 추출된 텍스트 영역

the region into four subregions

Step 3. Repeat these steps until no further splits take place

$$H(R) : (R_{max} - R_{min}) > T$$

R_{max} : a maximum gray value in a region.

R_{min} : a minimum gray value in a region.

T : Threshold

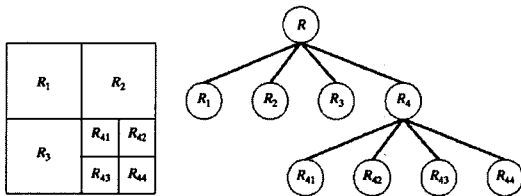


그림 6 분할 연산

(2) 합병

합병 연산의 입력은 분할 연산을 거치고 난 후의 이미지를 사용한다. 다음은 본 논문에서 사용한 합병 알고리즘을 설명한다.

Step 1. Consider any two or more neighboring subregions, R_1, R_2, \dots, R_n , in the image

Step 2. If $H(R_1 \cup R_2 \cup \dots \cup R_n)$ is true, merge the n regions into a single region

Step 3. Repeat these steps until no further merges take place

$$H(R_1 \cup R_2 \cup \dots \cup R_n) : H(R_1 \cup R_2 \cup \dots \cup R_n)_{max} - H(R_1 \cup R_2 \cup \dots \cup R_n)_{min} < T$$

그림 7은 입력 이미지의 예이고, 그림 8은 분할/합병 알고리즘을 적용한 후의 결과를 보여준다.



그림 7 입력 이미지



그림 8 분할/합병 연산 후의 이미지 모습

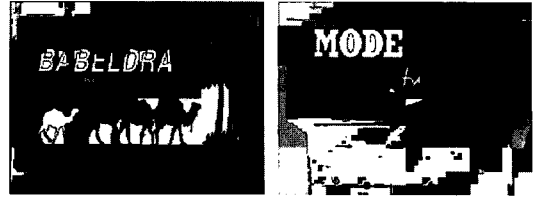


그림 9 크게 제한 연산 후의 이미지 모습

그림 8을 보면 분할과 합병 연산은 입력 이미지내의 화소들을 군집화하여 단순화하되, 밝기 값이 비슷한 영역들을 모으는 효과가 있는 것을 알 수가 있다. 이것은 텍스트와 배경이 서로 다른 대비를 갖고 있다는 특징을 이용하여 텍스트와 배경을 구별해 낼 수 있음을 의미한다. 반면, 텍스트와 배경이 서로 비슷한 색이거나 비슷한 그레이 값을 가진 영상에서는 분할과 합병을 수행해도 텍스트와 배경이 하나의 영역으로 합쳐져 버리는 문제점이 존재하기도 한다.

3.2.2 크기 제한 및 팽창 연산

분할과 합병 연산을 거친 후의 이미지는 비슷한 그레이 값을 갖는 영역으로 나누어져 있는데, 너무 큰 영역이나 너무 작은 영역은 글자 영역으로 볼 수가 없기 때문에 이를 제거할 수 있다. 대부분의 텍스트 영역의 특징은 비슷한 크기와 비슷한 색을 가지고 있기 때문인데, 분할된 이미지 영역들의 크기가 원본 이미지의 가로와 세로 길이를 기준으로 너무 큰 크기를 갖거나 작은 크기를 갖게 되면 잡영으로 생각하고 제거한다.

그림 9에서 보면 배경은 검정색이며, 그 이외의 명암 값은 텍스트 후보 영역에 해당한다고 할 수 있다. 남아 있는 텍스트 후보 영역에서 텍스트 영역과 배경의 분리는 2가지 경우로 생각하였다. 텍스트가 배경보다 더 밝은 명도를 갖는 경우와 배경이 텍스트 보다 더 밝은 명도를 갖는 경우로 나누었다. 다음은 본 논문에서 사용한 알고리즘을 설명한다.

Step 1. if $B_i > Mean$ is true, a candidate Text region

else if $B_i < T_B$ is true, a candidate Text region

else non-candidate Text region

Step 2. Repeat these steps until N

$$Mean = \frac{1}{N} \sum_{i=0}^{N-1} B_i \quad (3)$$

(N = 이미지에 남아 있는 텍스트 후보 영역들의 개수, B_i = 이미지에 남아 있는 텍스트 후보 영역의 평균 밝기 값, T_B : 밝기 임계 값)

그림 10처럼 비 텍스트 영역 제거 연산을 하고 나면 불필요한 배경 영역이 상당히 제거되었음을 알 수가 있

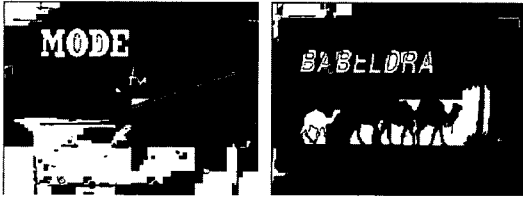


그림 10 이진화 결과 이미지

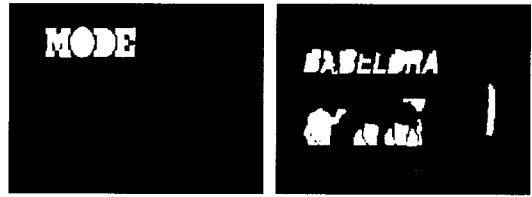


그림 12 분할된 최종 결과

다. 비 텍스트 영역 제거 연산을 수행하고 난 후의 이미지에서 화이트 노이즈를 제거하고 인접한 비슷한 영역끼리 합병을 하기 위해서 3×3 팽창 연산을 적용한 후, 크기 제한 연산을 다시 적용한 결과가 그림 11과 같다. 상당히 많은 잡영이 제거되었음을 알 수가 있다.



그림 11 팽창 및 크기 제한 연산 후의 모습

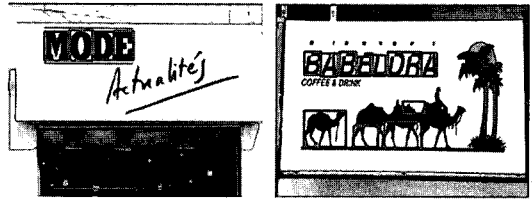


그림 13 추출된 최종 텍스트 결과

그림 13은 SMA에 의해서 추출된 텍스트 영역 이미지를 보여주고 있으며, 영상 내에서 찾지 못한 작은 크기를 갖는 텍스트나 초서체 형태의 텍스트는 이진화 및 크기제한에 의해서 제거되었다.

3.3 두 방법의 결합(Hybrid Analysis Method)

앞서 기술한 GIA와 SMA는 상호 보완적인 점이 존재한다. GIA시스템으로 장면 텍스트 추출을 실행한 결과 벽돌 무늬나 수직 수평 방향으로 무질서한 라인들이 많은 영상에서는 별로 좋지 못한 결과를 얻게 되었다. 그림 14는 GIA시스템에 의해서 얻은 실험 영상이다. 왼쪽에 있는 이미지는 텍스트의 일부만 찾은 경우이고, 오른쪽에 있는 이미지는 무질서한 라인들 때문에 텍스트 영역을 하나도 찾지 못한 경우를 보여주고 있다.

3.2.3 후보 영역 추출

비 텍스트 영역 제거 연산 및 팽창 연산을 수행하고 나면 그림 11과 같은 이미지 영역들만 남게 된다. 이 영역들을 대상으로 연결영역평균분석(Connected Region Mean Analysis) 과 영역채움요소분석(Region Fill Factor Analysis)을 거치고 나면 그림 12와 같은 최종 텍스트 후보 영역만 남게 된다. CRMA 방법은 앞서 구해진 이미지에 Blob Coloring[10]을 수행하여 각각 독립된 영역을 구한 후, 각 영역들의 평균값을 구하여 비슷한 평균값이 가로 방향이나 세로 방향으로 일직선상에 임계값 이상 놓여 있으면 일차 텍스트 영역 후보로 지정한다. 독립된 각 영역의 평균값을 구하는데 사용된 이미지는 합병 연산 직후에 구해진 그레이 값을 사용하였는데, 그 이유는 합병 연산 후에 나온 이미지가 비슷한 영역별로 가장 잘 분할되어 있기 때문이다. 2차 텍스트 후보 영역 검증에서는 RFFA 방법을 사용하였다. RFFA 방법은 CRMA 방법을 거친 1차 텍스트 후보 영역들의 최외곽사각형(Boundary Box)을 구한 후에 최외곽사각형 내에 검정색(픽셀값 0)과 하얀색(픽셀값 255)의 비율이 임계값 미만일 경우는 텍스트 후보 영역에서 제외한다. 이렇게 2단계의 장면 텍스트 영역 검증을 거친 최종 이미지의 모습이 그림 12와 같다. 그림에서 알 수 있듯이 SMA 방법에 의해서는 텍스트 후보 영역을 문자 단위로 추출할 수 있다.

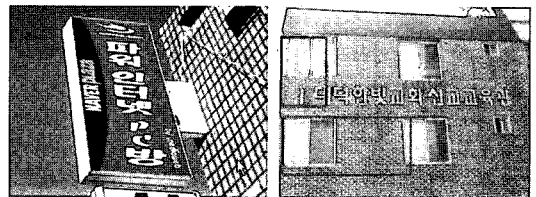


그림 14 GIA를 이용한 잘못된 추출된 예제

반면, SMA시스템에 의한 장면 텍스트 추출의 경우에는 텍스트 영역과 비 텍스트 영역의 색이 일치하거나 비슷한 경우에 텍스트 영역과 배경 영역이 합병 되어서 텍스트 추출에 어려움이 존재하였다. 그림 15은 SMA시스템에 의한 오류 이미지를 보여주고 있다. 아래 그림은 배경과 텍스트의 색이 거의 일치하는 경우를 보여주고 있다.

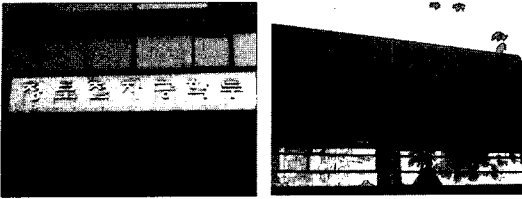


그림 15 SMA을 이용한 잘못 추출된 예제

2개 시스템의 문제점을 보완하기 위해 HAM에 의한 장면 텍스트 추출을 시도한 결과 상호 보완적인 장면 텍스트 추출 결과를 얻을 수 있었다. 그림 16은 HAM에 대한 전체 구성도를 나타낸다.

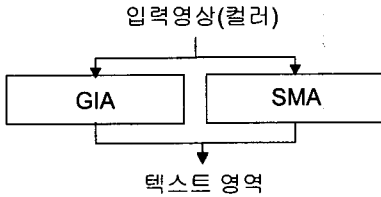


그림 16 HAM의 처리 과정

제안한 HAM시스템에서는 GIA와 SMA시스템이 개별적으로 텍스트 추출을 수행하고, 각각 추출된 결과에 대해서 OR 연산을 수행하여 장면 텍스트 영역을 추출한다. 그림 17은 HAM시스템에 의해 추출된 텍스트 추출 결과들을 보여준다. 그림 17내의 적색박스는 GIA시스템이 추출한 결과이며, 녹색박스는 SMA시스템이 추

출한 결과를 나타낸다. 그림 17에서도 알 수 있듯이 장면 텍스트 추출은 단일한 추출 방법을 사용하는 GIA 및 SMA 방법보다 복합적인 특징 추출 방법을 사용하는 HAM 방법이 더 좋은 추출 결과가 있음을 알 수 있다.

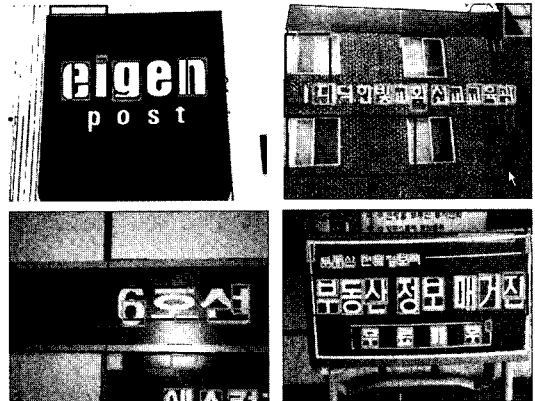
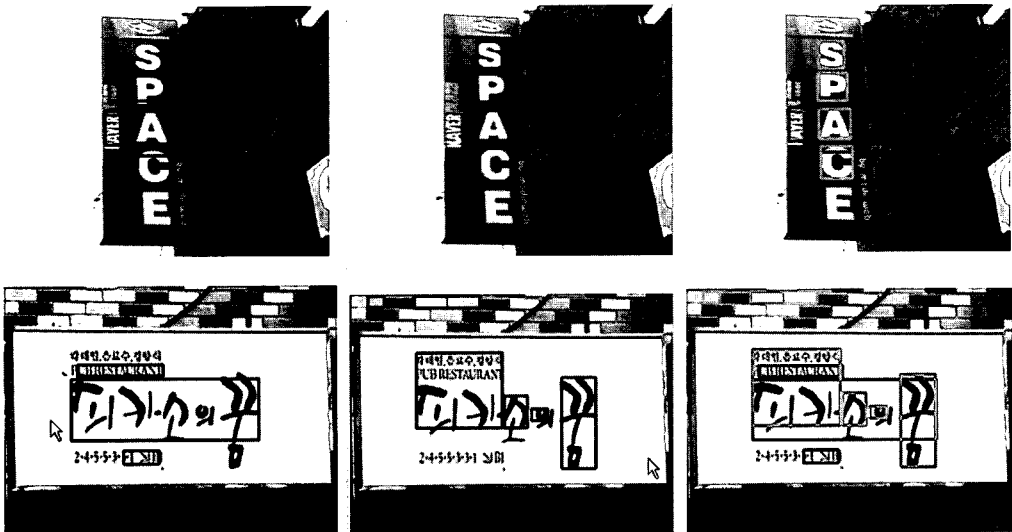


그림 17 추출된 최종 텍스트 결과(HAM 시스템 결과)

그림 18은 동일한 영상에 대해서 GIA, SMA 그리고 HAM 방법에 의한 추출 결과를 보여준다. 결과에서 보면 제안한 3가지 방법 중에서 HAM 방법이 가장 우수한 추출 결과를 보여주고 있다.

4. 실험 및 결과

장면 텍스트 추출 실험에 사용한 이미지는 학교, 병



(a) GIA 텍스트 추출 결과 (b) SMA 텍스트 추출 결과 (c) HAM 텍스트 추출 결과

그림 18 GIA, SMA and HAM 텍스트 추출 결과

원, 지하철역, 도로 간판, 상가 간판 및 현수막 등의 실내외에서 취득한 120개의 자연 이미지와 2003 ICDAR Contest Test-용 이미지 251개 총 371개 이미지를 대상으로 Visual C++ 6.0 환경에서 실험을 수행하였다. 입력된 이미지의 크기는 제한을 두지 않았으며, 취득한 이미지들은 단순한 영상과 복잡한 배경을 갖는 다양한 영상들을 실험에 사용하였다.

실험 데이터

- KAIST 및 숙명여대에서 수집한 자연 이미지 120개
- 2003 ICDAR Contest Test-용 이미지 251개

실험 결과 표 1, 2, 3에 있는 용어는 다음과 같다.

Sum : 실험 대상 이미지 내에 존재하는 전체 텍스트 개수 (True+Part+Error)

True : 정확히 찾은 텍스트 개수

Part : 텍스트의 일부분만을 찾은 개수

Error : 텍스트를 찾지 못한 개수

False : 텍스트 영역이 아닌 영역을 텍스트로 잘못 찾은 개수

Precision : $True / (True+Part+False)$

Recall : $True / Sum$

표 1은 자연 이미지 120개에 대한 3가지 제안 방법의 성능을 보여준다. 실험 결과에서 보면 제안한 3가지 방법 중에서 HAM 방법이 GIA와 SMA에 방법보다 더 우수한 결과를 보여줌을 알 수가 있다. 표 2는 제안한 방법 중에서 가장 우수한 성능을 보여준 HAM 방법과 동일한 데이터에 대한 참고문헌 [6]시스템의 성능을 보여준다. 표 2에서 두 시스템의 결과를 비교해 보면 True 개수에서 HAM 시스템은 923개이고, [6]시스템은 866개의 성능을 보였다. Error개수에서도 HAM은 88개이고, [6]시스템은 130개의 성능을 보였다. False 개수에서는 HAM 시스템은 254개를 보였으나 [6]시스템에서

표 1 자연 영상 텍스트 추출 결과(문자 단위)

이미지	시스템	Sum	True	Part	Error	False
자연영상 (120)	GIA	1045	829	54	162	222
		Precision = 75%		Recall = 79.3%		
	SMA	1045	759	50	236	284
		Precision = 69.4%		Recall = 72.6%		
	HAM	1045	923	34	88	254
		Precision = 76.2%		Recall = 88.3%		

표 2 자연 영상 텍스트 추출 결과(문자 단위)

이미지	시스템	Sum	True	Part	Error	False
자연영상 (120)	HAM	1045	923	34	88	254
		Precision = 76.2%		Recall = 88.3%		
	[6]	1045	866	49	130	450
		Precision = 63.4%		Recall = 82.8%		

표 3 ICDAR 영상 텍스트 추출 결과(워드 영역 단위)

이미지	시스템	Sum	True	Part	Error	False
ICDAR 영상 (251)	HAM	721	464	76	181	284
		Precision = 56.3%		Recall = 64.3%		
	[6]	687	474	29	184	346
		Precision = 55.8%		Recall = 68.9%		

표 4 [11]에서 인용한 ICDAR 영상 텍스트 추출 결과 (워드 영역 단위)

System	Precision	Recall
Ashida	55%	46%
HWDavid	44%	46%
Wolf	30%	44%
Todoran	19%	18%
[6]	55.8%	68.9%
HAM	56.3%	64.3%

는 450개의 성능을 보였다. False의 개수는 텍스트 영역이 아닌 영역을 텍스트로 잘못 추출한 경우의 수를 의미하는데, 이 부분에서도 HAM 시스템이 [6]시스템 보다 더 좋은 성능을 보여줬다. 표 1, 2는 문자 단위 텍스트 영역 추출 결과를 보여준다.

표 3, 표 4는 ICDAR Contest Test-용 이미지들에 대한 HAM과 참고문헌 [6], [11]에 대한 실험 결과로서 HAM에 의한 텍스트 추출 결과가 더 좋은 성능을 보여줌을 알 수 있다.

두 가지 종류의 데이터를 사용한 실험 결과를 통해서 단순한 이미지, 복잡한 이미지 모두에서 제안한 방법이 좀더 정확한 텍스트 추출이 가능함을 알 수가 있었다. 그러나 실험에서도 알 수 있듯이 2개 시스템을 결합한 HAM은 텍스트 추출률이 더 올라갔지만, 이와 반대로 비 텍스트 영역을 텍스트 영역으로 추출하는 False의 개수가 증가함을 알 수가 있다. 마찬가지로 참고문헌 [6]에서 개발한 장면 텍스트 추출 시스템에서도 이와 유사한 점을 보였다. 이 시스템은 명암정보, 색 연속성, 색 분산 3가지 연산을 수행하여 한 개의 결과 값으로 결합을 하기 때문에 비교적 단순한 영상에서는 비 텍스트 영역이 나타나지 않았으나 배경이 복잡한 영상에서는 비 텍스트 영역이 텍스트 영역의 2배 이상 나타나는 것을 알 수가 있었다. 그림 19는 HAM에 의해서 추출한 실험 영상 결과와 참고문헌 [6]에서 개발한 시스템의 결과를 보여주고 있다.

5. 결론

본 논문에서는 자연 이미지에 포함되어 있는 텍스트를 추출하기 위해 3가지의 텍스트 추출 방법을 제안하였다. 제안한 3가지의 알고리즘은 단순한 이미지와 복잡

HAM 실험 결과 이미지

[6] 실험 결과 이미지

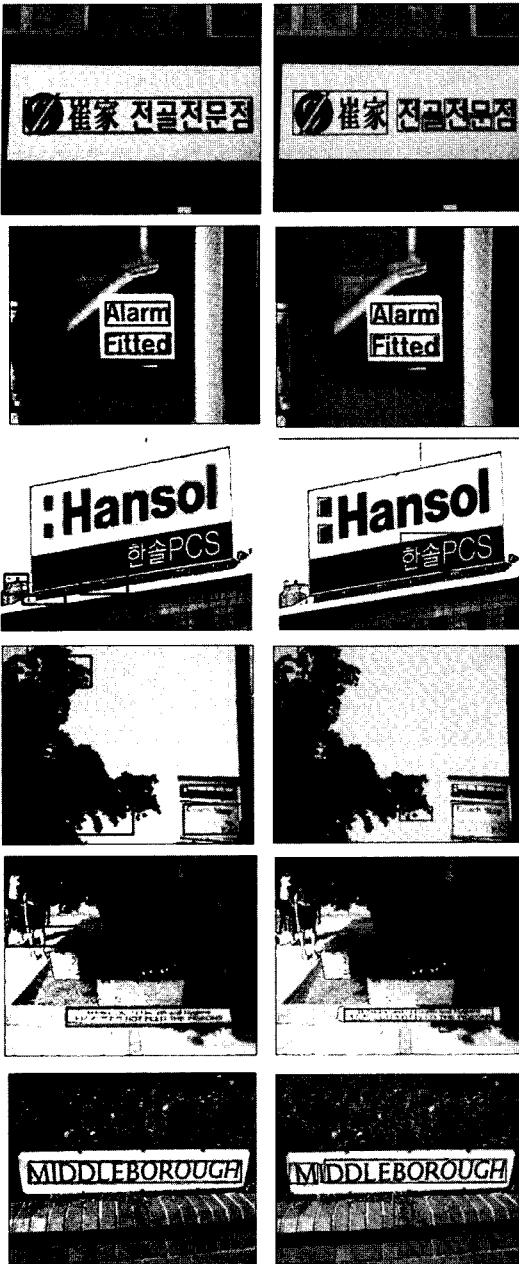


그림 19 추출된 텍스트 후보 영역

한 이미지 모두에서 상당히 정확하게 텍스트를 추출함을 보였다. 이러한 연구 결과는 앞으로 시각 장애인 보행 보조 시스템, 로봇 자동 주행 시스템, 비디오 자동 검색 및 색인 시스템, 텍스트 자동 번역 시스템, 이미지와 함께 포함되어 전송되는 스펙 메일 필터 등과 같은 다양한 분야에서 널리 활용될 수 있다.

향후 연구 과제로는 추출된 텍스트 후보 영역의 검증 을 위해 SVM, HMM 등을 이용한 다양한 검증 방법 개발과 추출된 텍스트 후보 영역에 대한 문자 인식 방법을 개발하는 것이다.

참 고 문 헌

[1] A. K. Jain, B. Yu, "Automatic Text Location in Images and Video Frames," *Pattern Recognition*, Vol. 31, No. 12, pp. 2055-2076, 1998.

[2] J. Hoya, A. Shio and S. Akamatsu, "Recognizing Characters in Scene Images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 16, No. 2, pp. 67-82, 1995.

[3] R. Lienhart, F. Stuber, "Automatic Text Recognition in Digital Videos," *Image and Video Processing IV*, SPIE, 1996.

[4] S. Messelodi and C. M. Modena, "Automatic Identification and Skew Estimation of Test Lines in Real Scene Images," *Pattern Recognition*, Vol. 32, No 5, pp. 701-810, 1999.

[5] Y. Zhong, K. Karu and A. K. Jain, "Locating Text in Complex Color Images," *Pattern Recognition*, Vol. 28, No. 10, pp. 1523-1535, 1995.

[6] Y. Choi, "Scene Text Extraction in Natural Images Using Hierarchical Feature Combining and Verification," *The 2nd KAIST-Tsinghua Joint Workshop on Pattern Recognition*, pp. 76-102, Daejeon, Korea, 2003.

[7] 장인영, 고병철, 변혜란, "모폴로지를 이용한 비디오 영상에서의 자동 문자 추출", 정보과학회 추계학술대회 논문집, pp. 418-420, 2001.

[8] L. Gu and T. Kaneko, "Robust Extraction of Characters from Color Scene Images Using Mathematical Morphology," *Proc. of the 14th International Conference on Pattern Recognition*, pp.1002-1004, 1998.

[9] R. Crane, *A simplified approach to Image Processing*, Prentice-Hall, 1997.

[10] D. H. Ballard and C. M. Brown, *Computer Vision*, Prentice-Hall, 1982.

[11] S. M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong and R. Young, "ICDAR 2003 Robust Reading Competitions," *Proc. of the 7th International Conference on Document Analysis and Recognition*, pp. 682-687, 2003.

[12] J. C. Shim, C. Dorai and R. Bolle, "Automatic Text Extraction from Video for Content-Based Annotation and Retrieval," *Proc. of the 14th International Conference of Pattern Recognition*, Vol. 1, pp. 618-620, Brisbane, 1998.

[13] Y. Zhong, H. Zhang and A. K. Jain, "Automatic Caption Localization in Compressed Video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 4, pp. 385-392, 2000.

- [14] K. Wang, J. A. Kangas and W. Li, "Character Segmentation of Color Images from Digital Camera Author," *Proc. of the 6th International Conference on Document Analysis and Recognition*, pp. 210-214, 2001.
- [15] 김형균, 최원호, "자연 영상에서의 문자 패턴 추출", 울산대학교 공학 연구 논문집, 제26권 2호, pp. 35-53. 1995.
- [16] 김의정, 정원일, "칼라 문서에서 문자 영역 추출을 위한 클러스터링 기법", 제14회 대전산업대학교 논문집, pp. 107-116, 1997.
- [17] 박영규, 김성국, 유원영, 김준철, 이준환, "MPEG-2 뉴스 영상에서 문자영역 추출 및 문자인식", 한국정보처리학회 논문지 제6권 제5호, pp. 1410-1418, 1999.
- [18] 노명철, 최영우, 이성환, "자연 영상에 대한 비디오 프레임에서의 장면 텍스트 추출", 컴퓨터비전 및 패턴인식 연구회 추계워크샵, pp. 161-162, 2001.
- [19] 김길천, 최영우, 변혜란, "장면(Scene) 텍스트 추출 및 기울기 / 원근 추정", 제 14회 영상처리 및 이해에 관한 워크샵 발표 논문집, pp. 277-282, 2002.
- [20] 김지수, 김수형, "명도 정보를 이용한 자연 영상에서의 텍스트 영역 추출", 한국정보처리학회 호남·제주지방 학술발표논문집, 제3권 제1호, pp. 127-132, 2003.



김 지 수

1996년 광주대학교 컴퓨터공학과 졸업(학사). 2003년 전남대학교 전산학과 졸업(이학석사). 2003년 3월~현재 전남대학교 전산학과 박사과정 재학 중. 관심분야는 인공지능, 패턴인식, 문자인식 등



김 수 형

1986년 서울대학교 컴퓨터공학과 졸업(학사). 1988년 한국과학기술원 전산학과 졸업(공학석사). 1993년 한국과학기술원 전산학과 졸업(공학박사). 1993년~1996년 삼성전자 멀티미디어연구소 선임연구원. 1997년~현재 전남대학교 전자컴퓨터 정보통신공학부 부교수. 관심분야는 인공지능, 패턴인식, 문서영상 정보검색, 유비쿼터스컴퓨팅



최 영 우

1985년 연세대학교 전자공학과 졸업(학사). 1986년 University of Southern California 컴퓨터공학과 졸업(석사). 1994년 University of Southern California 컴퓨터공학과 졸업(박사). 1994년 10월~1997년 2월 LG전자기술원 선임연구원. 1997년 3월~현재 숙명여자대학교 컴퓨터과학과 조교수, 부교수. 관심분야는 영상처리, 패턴인식, 문자인식 등