

계층적 특징 결합 및 검증을 이용한 자연이미지에서의 장면 텍스트 추출

(Scene Text Extraction in Natural Images using Hierarchical
Feature Combination and Verification)

최영우[†] 김길천^{**} 송영자^{***} 배경숙^{***}
(Yeongwoo Choi) (Kilcheon Kim) (Youngja Song) (Kyungsook Bae)

조연희^{***} 노명철^{****} 이성환^{****} 변혜란^{*****}
(Younhee Cho) (Myungchoel Roh) (Seongwhan Lee) (Hyeran Byun)

요약 이미지에 인위적 또는 자연적으로 포함된 텍스트는 이미지의 내용을 함축적이고 구체적으로 표현하는 중요한 정보이다. 이러한 정보를 실시간에 추출하여 정확히 인식할 수 있다면 다양한 분야에서 활용될 수 있다. 본 논문에서는 자연이미지에 포함된 장면 텍스트를 추출하는 방법으로서 텍스트의 색 연속성, 밝기 변화 및 색 변화와 같은 낮은 수준의 이미지 특징으로 텍스트 후보 영역을 찾고, 다해상도(Multi-resolution) 웨이블릿(Wavelet) 변환을 이용하여 높은 수준의 텍스트 특징인 획의 구성 여부로 검증하는 계층적인 구조를 제안한다. 색 연속성 특징은 대부분의 텍스트는 동일한 색으로 구성된다는 특징을 이용하는 것이고, 밝기 변화 특징은 텍스트 영역은 주변과의 밝기 변화가 존재하며 에지 밀도가 높은 특징을 이용한다. 또한, 색 변화 특징은 텍스트 영역은 주변 배경과의 색 변화가 존재하며, 밝기 변화보다 민감한 색 분산값으로 표현할 수 있다는 장점을 이용한다. 높은 수준의 텍스트 특징으로서 다해상도 웨이블릿 변환을 이용하여 텍스트 획의 방향성 정보를 추출하고, 추출된 정보를 SVM(Support Vector Machine) 분류기로 검증하여 최종 영역을 확정한다. 제안한 방법을 다양한 종류의 이미지에 적용한 결과 배경이 복잡해도 비교적 안정적으로 텍스트 영역을 추출하는 것을 확인할 수 있었다.

키워드 : 텍스트 추출, 계층적 특징 결합, 색 병합, 밝기 변화, 색 분산, 다해상도 웨이블릿 변환

Abstract Artificially or naturally contained texts in the natural images have significant and detailed information about the scenes. If we develop a method that can extract and recognize those texts in real-time, the method can be applied to many important applications. In this paper, we suggest a new method that extracts the text areas in the natural images using the low-level image features of color continuity, gray-level variation and color variance and that verifies the extracted candidate regions by using the high-level text feature such as stroke. And the two level features are combined hierarchically. The color continuity is used since most of the characters in the same text region have the same color, and the gray-level variation is used since the text strokes are distinctive in their gray-values to the background. Also, the color variance is used since the text strokes are distinctive in their color values to the background, and this value is more sensitive than the gray-level variations. The text level stroke features are extracted using a multi-resolution wavelet transforms on the local image areas and the feature vectors are input to a SVM(Support Vector Machine) classifier for the

· 본 연구는 한국과학기술연구원(KISTEP)에서 지원하는 뇌신경정보학사업의 지원을 받아 수행되었습니다. (M1-0102-00-0009).

† 정 회 원 : 숙명여자대학교 정보과학부 교수
ywchoi@sookmyung.ac.kr

** 비 회 원 : 연세대학교 컴퓨터과학과
kimkch@cs.yonsei.ac.kr

*** 비 회 원 : 숙명여자대학교 컴퓨터과학과
soyo303@sookmyung.ac.kr
encrypt78@hotmail.com

tangcyh@hotmail.com

**** 비 회 원 : 고려대학교 컴퓨터과학과

mcroh@image.korea.ac.kr

***** 종신회원 : 고려대학교 컴퓨터과학과 교수
swlee@image.korea.ac.kr

***** 종신회원 : 연세대학교 컴퓨터과학과 교수
hrbyun@cs.yonsei.ac.kr

논문접수 : 2003년 7월 15일

심사완료 : 2004년 1월 5일

verification. We have tested the proposed method using various kinds of the natural images and have confirmed that the extraction rates are very high even in complex background images.

Key words : Text extraction, Hierarchical feature combining, Color clustering, Gray-level variation, Color variance, Multi-resolution wavelet transforms

1. 서론

카메라로 취득한 이미지나 동영상에 인위적으로 삽입되거나 자연적으로 포함된 텍스트들은 이미지의 내용을 함축적이고 구체적으로 표현하는 중요한 정보들이다. 이러한 정보들을 실시간에 추출하여 인식한다면 시각장애인을 위한 보행안내시스템, 지능 로봇의 자동주행 등과 같은 분야에서 다양하게 활용될 수 있다[1]. 텍스트 추출 연구는 그래픽(Graphic) 텍스트 추출과 장면(Scene) 텍스트 추출 연구로 나누어 진행되고 있다. 그래픽 텍스트 추출은 대부분 이미지의 질이 높은 문서이미지나 비디오 프레임을 대상으로 영역을 추출하였지만[2-7], 최근에는 복잡한 배경을 갖는 자연이미지에서 텍스트의 크기, 색, 폰트 종류, 방향 등이 다양하게 나타나는 장면 텍스트를 추출하는 연구가 응용을 고려한 제한적인 환경에서 활발하게 진행되고 있다[8,9].

그래픽 텍스트 추출에 관한 연구로서 Jain 등은[10] 이진 이미지, 웹 이미지, 색 이미지 및 비디오 프레임의 네 가지 종류의 이미지에서 텍스트를 추출하는 방법을 제안하였다. 이진 및 웹 이미지에 대해서는 텍스트의 밝기 값이 균일하다는 특징을 이용하였고, 색 이미지 및 비디오 프레임에 대해서는 색 연속성을 이용하였다. 밝기 값이 균일하다는 특징을 이용하여 다중 값 분해(Multi-valued Decomposition)를 통해 전경과 배경을 분리한 후 연결요소를 분석하여 영역을 추출하였으며, 색 연속성은 색 줄임을 수행한 후 역시 다중 값 분해를 통해 전경과 배경을 분리한 연결요소를 분석하여 영역을 추출하였다. 이 방법은 네 가지 종류에 대해 서로 다른 특징과 임계 값을 적용하기 때문에 종류별로 수동적인 실험을 수행하였다. 실험결과 이진 이미지, 웹 이미지, 비디오 프레임에서는 높은 추출률을 보인 반면, 색 이미지에 대해서는 낮은 추출률을 보였다. 또한 대부분의 실험 이미지가 비디오 프레임에 국한되어 있는 단점이 있다.

Zhong 등은[11] 텍스트의 색은 일정하며, 명도이미지에서 텍스트 영역은 공간적 분산(Spatial Variance) 값이 크다는 특징을 이용하고, 두 방법을 순차적으로 결합한 방법을 제안하였다. 색을 이용한 방법에서는 색 양자화를 수행하여 색의 개수를 줄이고 각 색 면에 대한 연결요소를 분석하여 영역을 추출하였다. 명도이미지를 이용한 방법에서는 공간적 분산을 적용한 후 에지를 추출

하여 서로 반대 방향을 가지는 에지 쌍을 찾음으로써 텍스트 영역을 추출하였다. 끝으로 두 방법을 순차적으로 결합하여 영역을 확정하였다. 다양한 종류의 스캔된 이미지로 제안한 방법을 실험한 결과 길이가 짧거나 색이 일정하지 않은 문자열, 수직 방향 또는 필기된 문자열 등에서 오류가 발생하는 단점이 있었다.

H. K. Kim은[12] 비디오 프레임으로부터 자동으로 텍스트 영역을 추출하기 위해서 문자들이 수평 방향으로 놓여져 있고, 균일한 색과 일정한 크기를 갖는다는 가정으로 색 연속성 특징을 이용한 방법을 제안하였다. 알고리즘은 크게 두 부분으로서 우선 색 히스토그램 양자화에 의해 색 이미지를 분할하고, 연결요소의 길이와 X/Y 시그니처를 이용하여 각 색 면에서 비텍스트 요소들을 제거하였다. 50개의 비디오 프레임으로 제안된 알고리즘을 실험한 결과 86%의 추출률을 보였으나, 색의 대비가 크지 않은 텍스트와 크기가 작은 텍스트에 대해서는 정확히 찾지 못하며, 16개의 경험적 임계 값을 정해야 하는 단점이 있다.

P. K. Kim은[13] 색 이미지를 대상으로 텍스트의 색 연속성 특징을 이용한 추출 방법을 제안하였다. 제안한 방법은 24비트(Bit) 색을 모두 이용하는 전역적 양자화의 단점과 텍스트와 배경 영역이 잘 분리되지 않고 시간이 많이 소요되는 단점을 보완하기 위해서 R , G , B 각각의 상위 2비트를 이용하여 총 6비트를 이용하는 지역적 양자화 방법을 적용하였다. 지역적 양자화를 수행한 후 경계선추적(Contour Following)을 적용하여 얻어진 연결요소를 분석하여 텍스트 영역을 추출하며, 책 표지 이미지를 대상으로 실험하였다. 지역적 양자화의 처리 시간이 오래 걸리고 기울어진 텍스트 영역을 추출하기 어려운 단점이 있다.

장면텍스트 추출 연구로서 Ohya 등은[14] 실외 자연 이미지의 텍스트 영역을 추출하기 위해서 대부분의 텍스트는 명도 값이 일정하다는 특징을 이용하였다. 이 논문에서의 텍스트 영역은 기울어짐이 없고 배경과 뚜렷하게 구분된다는 가정을 전제하였다. 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 얻었지

만, 배경이 복잡하거나 조명 변화가 있는 이미지에서는 오류가 많이 발생하는 단점이 있었다.

Gu 등은[15] 미분 탑-햇츠(Differential Top-hats) 연산자를 이용한 추출 방법을 제안하였으며, 문자 획을 크기에 따라 8단계로 나누고 흰 문자와 검은 문자로 구분하여 연산자를 적용하였다. 제안한 알고리즘을 실외에서 오전, 정오, 오후, 저녁 등 네 가지 시간대별로 획득한 이미지에 적용한 결과 정오와 오후에 대한 이미지에서는 정확한 추출을 보였지만 조명의 변화가 있는 오전과 저녁 이미지에서는 오류가 많이 발생하였다. 또한, 처리 시간이 오래 걸리는 단점이 있다.

위에서 살펴본 것처럼 그래픽 텍스트를 추출하기 위한 대부분의 방법들은 색 이미지에서의 색 연속성 특징을 이용한 색 병합 방법의 시도와 명도이미지에서는 명도 값이 균일하다는 특징을 이용하기 위한 이진화 방법을 시도하였다. 장면 텍스트 추출 연구에서도 유사한 방법을 시도하였으나, 그래픽 텍스트 추출에 비해 색, 방향, 크기, 폰트, 기울어짐 등에 제약을 두었다. 본 논문에서는 장면 텍스트 추출에서의 제약을 어느 정도 해제하며 기존 연구들의 문제점들을 해결하기 위해서 수준별 특징을 2단계 계층으로 구성하여 추출하고 검증하는 방법을 제안한다. 본 논문에서 제안하는 방법의 장점은 다양한 종류의 자연 이미지에 대하여 비교적 정확한 추출 결과를 얻을 수 있다는 점이며, 특징은 세 종류의 낮은 수준의 이미지 특징 선택으로 텍스트 후보 영역을 추출하며 높은 수준의 텍스트 획 특징을 SVM 분류기를 이용하여 검증하는 계층적인 구조를 제안한 것이다.

2. 제안 방법

본 논문에서 제안하는 방법은 그림 1과 같이 낮은 수준의 이미지 특징을 이용하여 텍스트 후보 영역을 추출하고 높은 수준의 텍스트 획 특징으로 검증하는 계층적인 구조로 이루어져 있다. 낮은 수준의 이미지 특징은 지역적인 영역에서의 밝기 및 색 변화와 색 연속성이 존재하는지를 파악하기 위한 것이며, 높은 수준의 텍스트 특징은 낮은 수준의 특징으로 찾은 후보 영역들의 각 부분들이 문자 구성의 기본 단위인 획을 구성할 수 있는가를 확인하기 위한 것이다. 검증된 획 요소들이 조합되어 자음, 모음, 영어 알파벳, 숫자 등의 의미 있는 문자를 구성할 것이며, 문자 단위의 검증은 찾아진 영역에 대한 최종 인식으로 확인할 수 있다. 따라서 본 논문에서는 문자인식의 전 단계인 획 구성 여부까지를 검증하여 텍스트 영역을 결정하는 방법을 제안한다.

낮은 수준의 이미지 특징으로서 색 연속성, 밝기 변화 및 색 변화를 사용한다. 색 연속성은 동일한 영역의 텍스트는 대부분 같은 색으로 구성된다는 점을 이용하기

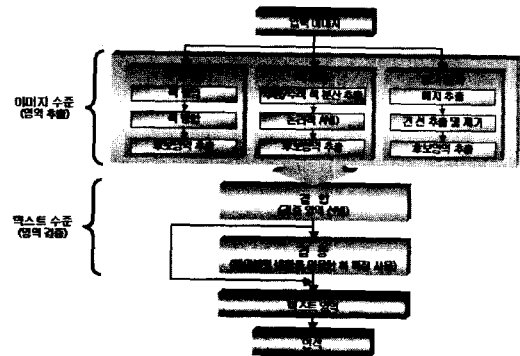


그림 1 제안한 방법

위한 것이며, 밝기 변화는 명도이미지에서의 텍스트 영역은 밝기 변화가 심하여 조밀한 에지 분포를 보인다는 점을 이용한다. 색 변화는 텍스트는 배경과의 색 변화가 존재한다는 점을 이용하기 위한 것이다. 위의 각 특징들을 이용하여 추출한 텍스트 영역을 적절히 결합한다면 조명 변화, 방향 변화, 기울어짐 존재 및 복잡한 배경을 갖는 이미지에서도 비교적 정확한 추출이 가능하다. 높은 수준의 텍스트 특징으로서 다해상도 웨이블릿 변환을 이용한 8-방향 획 특징을 사용한다. 텍스트 후보영역을 일정한 크기로 나눈 후 각 영역이 텍스트 획을 구성하는지를 SVM을 이용하여 검증한다. 여기서 다해상도에서 웨이블릿을 추출한 것은 다양한 크기의 문자에 대해서도 정확한 획 특징을 추출하기 위함이다.

각 특징으로 텍스트 영역을 찾는 과정을 살펴본다면 색 연속성을 이용한 추출은 우선 색 면의 개수를 줄이기 위해서 간단한 색 줄임을 수행하고, 조명 변화 및 다양한 배경에서도 안정적인 색 병합 방법을 제안한다. 각각의 색 면에 대한 연결요소를 분석하여 텍스트 후보영역을 찾고, 각 면에서 찾은 결과를 결합하여 텍스트 후보영역으로 결정한다. 밝기 변화를 이용한 추출은 입력된 이미지에서 에지를 구하고, 에지 분포를 분석하며 이미지에 포함되어 있는 다양한 모양의 긴 선들을 제거한다. 모폴로지(Morphology)를 이용하여 각 연결요소를 강조한 후 연결요소를 분석, 검증하여 후보영역을 결정한다. 이 과정에서 얻어진 긴 선 정보들은 텍스트 후보영역의 기울어짐과 원근 보정에 사용될 수 있다. 색 변화를 이용한 추출은 우선 수평 및 수직 방향의 색 분산 이미지를 구하는데, 이는 텍스트 영역은 두 방향의 분산 값이 모두 크게 나타나기 때문이다. 다음으로 두 결과의 논리적 AND 연산을 수행하여 선 또는 잡영과 같이 한 방향으로만 분산 값이 큰 부분들을 제거한 후 연결요소를 분석, 검증하여 후보영역을 결정한다. 각 특징에서 추출된 후보영역의 외곽사각형의 겹침 정도에 따라 결

합하여 후보영역을 선별한다. 선별된 후보 영역에서 각 부영역의 획 특징을 추출하고 SVM 분류기를 이용하여 검증한다.

3. 텍스트 영역 추출

3.1 색 연속성을 이용한 추출

색 연속성을 이용한 추출은 우선 색 면의 개수를 줄이기 위해서 간단한 색 줄임을 실시하고, 조명 변화와 복잡한 배경을 처리하기 위해서 색 병합 방법을 제안한다. 각각의 색 면에 대한 연결요소를 분석하여 텍스트 후보영역을 찾고, 각 면에서 찾은 결과를 결합함으로써 최종적인 후보영역을 확정한다.

3.1.1 색 줄임

입력된 이미지는 640x480 크기의 24비트 색 이미지로써 320x240 크기로 먼저 축소시킨다. 24비트로 표현된 이미지에서 색 면을 분리한다면 많은 계산 시간이 필요하기 때문에 색 줄임을 먼저 수행한다. 색 줄임은 화소(Pixel)의 *R*, *G*, *B* 각 요소의 하위 6비트를 제거하여 상위 2비트만을 남기는 Bit Dropping 방법을 이용한다. 색 줄임 결과 이미지는 최대 64개의 색으로 표현되며, 그림 2는 비트 제거에 의한 색 줄임 결과를 보여준다.

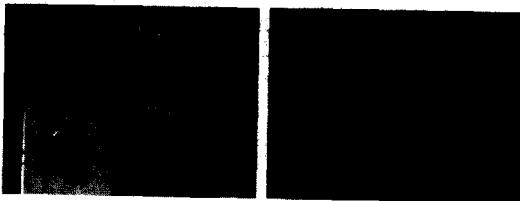


그림 2 원본 이미지 및 색 줄임 결과

3.1.2 색 병합(Color Clustering)

RGB로 표현된 색 이미지는 빛과 조명에 따라 동일한 색으로 구성된 화소들이 서로 다른 색으로 분리되는 경우가 빈번하게 발생한다. 또한, 색 줄임으로 만들어진 색 면 모두를 대상으로 영역을 찾는 것이 불필요하며, 처리시간도 오래 걸리기 때문에 색 병합을 수행하여 처리할 색 면의 개수를 줄이는 것이 필요하다. 64개의 색으로 표현되는 RGB 공간에서 색 사이의 거리를 이용한 색 병합을 수행한다. 색이 유사하다고 판단하여 병합되는 RGB 공간에서의 3차원 유클리디언(Euclidean) 거리가 1 이하인 경우로 설정한다. 색 병합은 그림 3과 같이 1차 병합과 반복적인 2차 병합으로 나누어 수행된다.

색 줄임 결과 이미지는 색마다 다양한 빈도수를 갖는다. 1차 병합은 눈으로 식별하기 어려운 정도의 적은 빈도수의 색들을 미리 제거함으로써 클러스터(Cluster)들을 고립시켜 색 병합 결과를 향상시키기 위한 것이다.

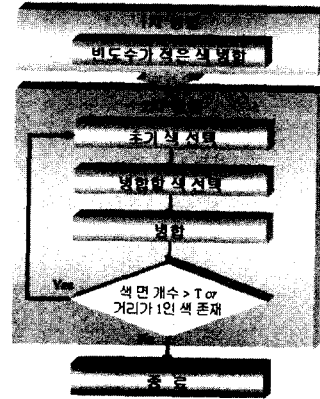


그림 3 색 병합

색 줄임 결과 이미지의 각 색의 히스토그램을 구한 후 히스토그램 수가 일정한 수 이하인 색들을 선택하여 그 색과 유사하며 빈도수가 가장 높은 색으로 병합시킨다.

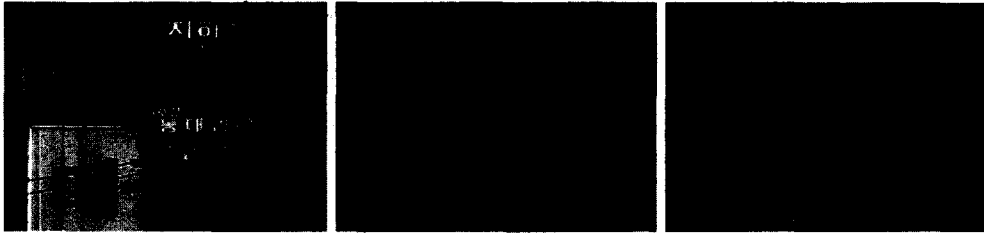
자연이미지에 나타나는 텍스트는 그 영역을 강조하기 위해서 대부분 배경과 보색 관계에 있거나 RGB 공간에서 서로 거리가 먼 색들로 구성되는 경향이 있다. 2차 병합은 유사도, 보색 관계 및 경험적인(Heuristic) 정보를 활용한다. 1차 병합 후에 남은 색 중에서 RGB 공간에서 각 모서리에 위치하며 빈도수가 가장 높은 색 또는 모서리와 가장 가까우면서 빈도수가 가장 높은 색을 병합할 초기 색으로 선택한다. 다음으로 초기 색과 유사하며 빈도수가 가장 큰 색을 선택하여 병합한다. 이 때 병합되는 색들은 대부분 경험적으로 색 공간에서 회색 계열의 색이거나 회색 계열의 색 근처에 위치한다. 그러나 *R*, *G*, *B*와 *C*(Cyan), *M*(Magenta), *Y*(Yellow)는 뚜렷이 구분되어야 하는 색임에도 불구하고 색 공간에서 회색 계열의 색과 같은 거리에 존재하여 하나의 회색 계열 색으로 병합되는 경우가 발생한다. 따라서 선택된 두 색을 병합할 때 두 색 중 하나 이상이 회색 계열의 색이라면 빈도수가 큰 회색 계열의 색으로 병합하여 *R*, *G*, *B*로 구성된 영역과 *C*, *M*, *Y*로 구성된 영역간의 연결이 끊어지도록 한다. 그러나 만약 회색 계열의 색이 없다면 빈도수가 높은 색으로 병합한다. 이러한 과정은 남아있는 색의 개수가 2가 되거나, 유사한 색이 더 이상 존재하지 않을 때까지 반복한다. 그림 4(a)는 색 줄임과 색 병합 과정을 거쳐 색의 수가 줄어들면서 *R* 계열의 색과 *Y* 계열의 색이 분리되어 배경과 뚜렷이 구분되는 결과를 보여준다.

3.1.3 후보영역 추출

색 병합 결과에서 동일한 색으로 분리되어야 할 영역들이 조명의 영향 또는 색 줄임으로 인해서 한 개 이상의 다른 색으로 분리되는 경우가 발생한다. 이러한 문제



(a) R 계열의 색과 Y 계열의 색이 잘 분리된 경우: 색 줄임(25색), 색 병합(6색)



(b) 일반적인 경우: 색 줄임(30색), 색 병합(5색)

그림 4 색 줄임 및 색 병합 결과 예

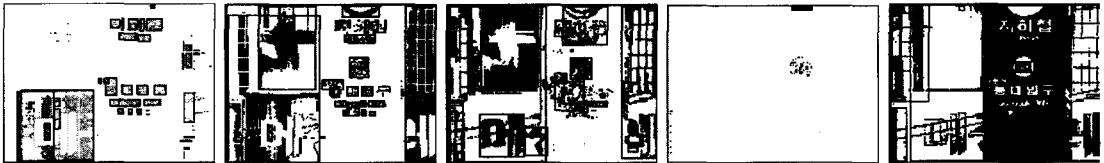


그림 5 각 색 면의 외곽사각형 추출(5개의 색 면)

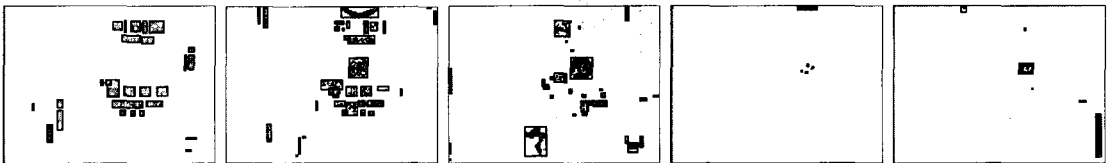


그림 6 연결요소의 1차 제거

를 해결하기 위해서 텍스트 영역의 연결요소는 밀도가 높은 특징을 이용하여, 색 병합된 이미지의 각 색 면에 3x3 크기의 구조 연산자(Structuring Element)를 이용한 모폴로지 닫힘(Closing) 연산을 수행하여 인접한 요소들을 연결시킨다. 벽면의 타일과 같이 좁은 간격으로 넓게 분포되어 있는 연결요소들이 하나로 뭉쳐지는 경우도 발생하지만, 연결요소의 크기를 제한함으로써 이러한 것들을 부분적으로 제거할 수 있다. 각 색 면에서 연결요소와 외곽사각형을 구하고, 연결요소의 크기와 외곽사각형의 위치, 크기, 종횡비(Aspect Ratio)를 추출한다. 그림 5는 각 색 면의 연결요소와 연결요소의 외곽사각형을 보여준다.

각 색 면에서 후보영역의 수를 줄이기 위해서 외곽사

각형의 크기가 임계값 이상으로 큰 것과 너무 작아서 식별할 수 없는 후보영역이 생기지 않도록 가로 또는 세로가 한 화소로 구성되거나 임계값 이하의 넓이를 갖는 외곽사각형을 제거한다. 또한, 외곽사각형의 세로/가로, 외곽사각형을 차지하는 연결요소의 비율에 제한을 두어 기둥이나 벽면 전체 또는 일부로 구성된 연결요소를 제거한다. 그림 6은 그림 5의 각 색 면에 위의 규칙을 적용한 결과이다.

다음으로 문자 단위의 의미 있는 외곽사각형을 구성하기 위해서 외곽사각형 사이의 거리와 X 및 Y축 방향으로 투영의 겹침 정도를 이용하여 연결요소를 결합한다. 각 색 면을 구성하고 있는 외곽사각형 사이의 시작과 끝점을 사용하여 가로, 세로 방향으로의 겹침 정도가

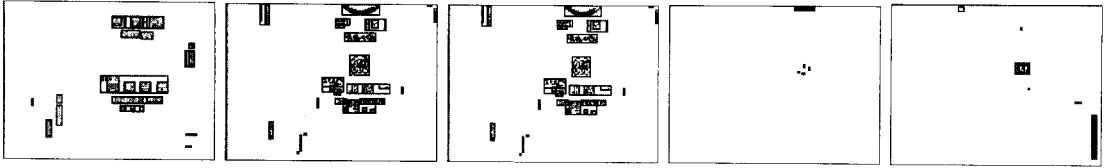


그림 7 인접한 외과사각형의 결합

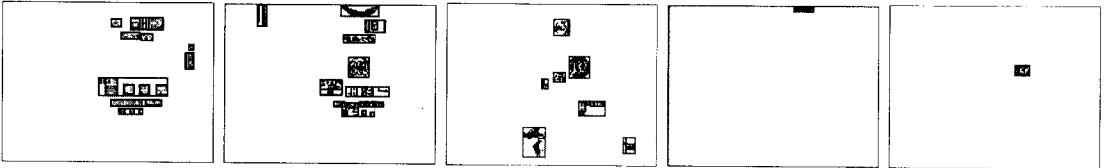


그림 8 텍스트가 아닌 것으로 판단되는 외과사각형의 2차 제거

크거나 외과사각형 사이의 거리가 10화소 이하이면 결합한다. 이 과정으로 의미 있는 문자 단위의 외과사각형을 구성할 수 있다. 그림 7은 결합 결과를 보여준다.

처음 단계에서의 연결요소 제거는 텍스트의 획이 제거되는 것을 방지하기 위해서 명확하게 제거할 만한 것들만 제거하기 위해서 엄격한 기준으로 제거하였지만, 외과사각형들이 1차로 결합된 후에는 획들이 인근 연결요소에 결합되었기 때문에 보다 완화된 기준으로 외과사각형을 추가로 제거할 수 있다. 제거 규칙은 1차 제거와 같지만 임계값은 다르게 설정한다. 획과 같은 텍스트의 일부를 구성하는 연결요소들은 전 단계에서 문자 단위의 외과사각형에 포함되기 때문에 높이와 폭을 제한하는 임계값을 크게 하여 작은 연결요소를 제거하고, 세로/가로 비율을 제한하는 임계값도 크게 하여 가로 또는 세로로 치우친 연결요소를 제거한다. 또한, 외과사각형을 구성하는 연결요소의 비율이 임계값 이하이면 제거하여 텍스트의 아주 작은 부분을 포함하는 연결요소는 제외시킨다. 그림 8은 2차 제거 조건으로 제거한 결과이다.

끝으로 각 색 면에서 찾아진 후보영역들을 결합하여 하나의 후보영역 외과사각형을 구성한다. 각 색 면에서 찾아진 후보영역들의 외과사각형의 겹침 정도를 파악하

여 80% 이상으로 서로 겹쳐졌으면 결합하여 하나의 외과사각형으로 만든다. 그림 9는 각 색 면에서 추출된 영역들의 후보영역의 결합 전과 결합 후의 결과를 보여준다.

3.2 밝기 변화를 이용한 추출

밝기 변화를 이용한 추출은 먼저 이미지의 에지를 구한다. 에지 분포를 분석하여 이미지에 포함되어 있는 다양한 종류의 긴 선들을 제거하여 처리할 이미지를 단순하게 만든다. 긴 선 정보들은 텍스트 후보영역의 기울어짐과 원근을 보정하는데 사용될 수 있다.

3.2.1 전처리

색 연속성을 이용한 추출에서 사용한 320x240 크기의 축소된 이미지를 먼저 명도이미지(Gray-level)로 바꾼다. 이 이미지에 3x3 크기의 미디언(Median) 필터를 적용하여 잡영의 일부를 제거한 후, Canny 에지[16] 검출을 수행한다. 그림 10은 미디언 필터 및 Canny 에지 검출 결과를 보여준다. Canny 에지 검출은 에지를 매우 정교하게 찾아주는 장점이 있지만, 연산 시간이 오래 걸리는 단점이 있다. 본 논문에서는 처리시간이 오래 걸리는 단점을 개선하기 위해서 Canny 에지 추출 과정에서 시간이 가장 많이 소요되는 억제(Suppression) 단계에서 주변의 모든 화소로부터 용기(Ridge)점을 찾지 않고,

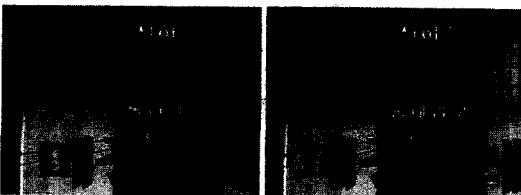


그림 9 각 색 면에서 추출된 후보영역의 결합 전과 결합 후

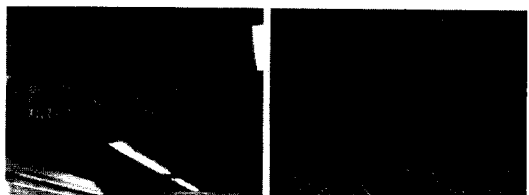


그림 10 원 이미지에 대한 전처리 수행 후 에지 추출 결과

경사(Gradient) 값이 큰 화소만을 대상으로 용기점을 찾음으로써 비교 횟수를 줄여 처리 시간을 대폭 단축시킬 수 있었다.

3.2.2 후보영역 추출

자연이미지에 포함되어 있는 텍스트 주변에는 텍스트를 둘러싸고 있는 다양한 종류의 긴 선들이 대부분 존재한다. 따라서 텍스트 주변의 긴 선들을 추출하여 기울어짐 및 원근 추정에 필요한 정보를 추출한 후 제거하면 이미지를 단순하게 만들 수 있다. 긴 선들이 제거된 단순한 이미지에서 텍스트 영역은 에지 밀도가 높은 특징이 나타난다. 에지 밀도가 높은 영역에 모폴로지를 적용한다면 텍스트 영역을 더욱 강조할 수 있으며, 밀도가 낮은 잡영의 일부를 제거할 수 있다. 강조된 이미지에 대해서는 연결요소, 외곽사각형 정보 및 에지 빈도를 이용하여 검증한다.

긴 선 추출 및 제거: 선을 찾기 위해서 8-방향의 인접 화소를 고려한 에지 추적을 수행하여 각 방향의 히스토그램을 구한다. 에지 추적의 시작점 위치에 따라 같은 방향의 선임에도 불구하고 서로 다른 방향으로 히스토그램이 구성될 수 있기 때문에 이를 해결하기 위해서 8-방향에서 서로 대칭인 방향을 합한 4-방향에 대한 히스토그램도 함께 고려한다. 또한, 그림 11과 같이 긴 선이 텍스트와 접촉된 경우도 빈번히 발생하는데 이러한 문제를 해결하기 위해서 X, Y 투영을 이용하여 긴 선 부분만을 선택적으로 제거한다. 이 과정에서 선의 크기를 제한하여 텍스트의 획과 같이 작은 부분이 제거되지 않도록 한다.

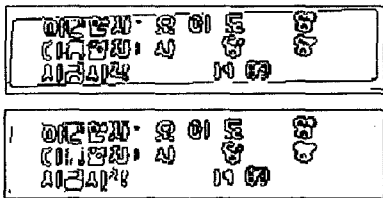


그림 11 텍스트와 긴 선이 접촉된 경우의 긴 선 제거 결과

에지 추적에서 X, Y축의 최대, 최소 좌표와 선의 방향 변화 빈도를 찾는다. 추출된 각 선의 히스토그램을 분석하면 긴 선들은 4-방향 히스토그램의 최대 빈(Bin)의 값이 다른 빈들에 비해 크며, 최대 X, Y와 최소 X, Y의 차가 크게 나타나며 방향의 변화 정도는 작게 나타난다. 사각형인 경우 4-방향 히스토그램의 상위 두 방향 빈의 합이 전체 히스토그램 합과 유사하게 나타난다. 이러한 특징을 이용하여 긴 선, 사각형의 테두리 선 및 길고 꺾이는 부분이 포함된 선 등의 다양한 모양의 선

들을 제거한다. 그림 12는 선 요소들의 히스토그램과 X, Y 투영을 이용하여 긴 선들을 제거한 결과이다.

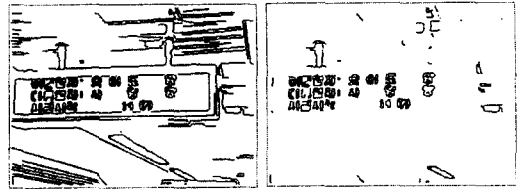


그림 12 에지 이미지와 긴 선 제거 결과

후보 영역 강조: 긴 선들이 제거된 이미지에서의 텍스트 영역은 경험적으로 에지 밀도가 높게 나타난다. 에지 밀도는 조밀하지만 문자의 크기, 간격 등에 따라 에지 간격이 다양하기 때문에 모폴로지 연산을 적용하여 문자 영역을 강조시킨다. 모폴로지 연산은 그림 13과 같은 2x5 크기의 불림(Dilation) 연산자와 3x2 크기의 녹임(Erosion) 연산자를 구성하여 닫힘(Closing) 연산을 적용한다.



그림 13 불림(좌) 및 녹임(우) 연산을 위한 구조 연산자

모폴로지를 이용한 강조는 대부분 글자 단위로 강조되기 때문에 수직 문자열에 대해서도 글자 단위의 추출이 가능해진다. 이러한 모폴로지 연산을 수행하여 텍스트와 제거되지 않은 텍스트 주변의 에지들이 서로 연결되는 경우를 줄일 수 있다. 그림 14는 모폴로지를 적용한 결과로서 텍스트 영역이 강조된 것을 보여준다.

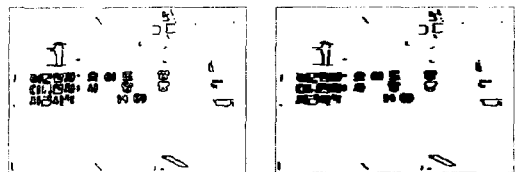


그림 14 긴 선이 제거된 결과에 모폴로지를 적용한 결과

3.2.3 후보영역 검증

모폴로지를 적용한 결과에서 연결요소의 화소수가 너무 크거나 작은 것은 텍스트가 아닌 경우가 대부분이기 때문에 제거한다. 연결요소의 화소수가 아주 큰 경우는 나무나 타일 등과 같이 에지 밀도가 조밀한 배경에서 발생된 경우가 대부분이고, 연결요소의 화소수가 아주

작은 경우는 비록 텍스트 영역이라 하더라도 사람도 구분하기 어렵기 때문에 제거한다. 또한, 외곽사각형에서 연결요소가 차지하는 비율, 외곽사각형의 세로/가로 비율, 폭 또는 높이 등이 아주 작은 것들도 추가로 제거한다. 또한, 그림 15와 같이 각 영역에 대하여 수평 및 수직 방향의 에지 빈도의 평균값을 확인하여 검증한다.

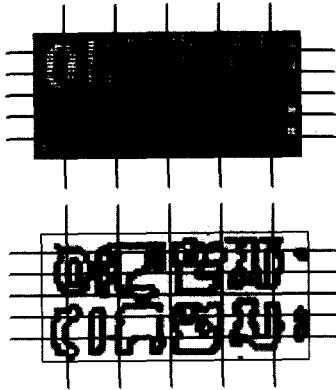


그림 15 에지 빈도를 이용한 후보영역 검증: 후보영역 및 해당 영역의 에지 이미지

그림 16은 에지 빈도를 이용한 검증 결과를 보여주며, 왼쪽 그림에서 흰색 사각형은 제거된 외곽사각형을 나타내며, 빨간색 사각형은 검증되고 남은 결과를 나타낸다. 오른쪽 그림에서 원으로 표시된 부분들은 제거된 사각형의 에지를 보여준다.

검증되고 남은 후보영역들을 의미 있는 문자 단위의

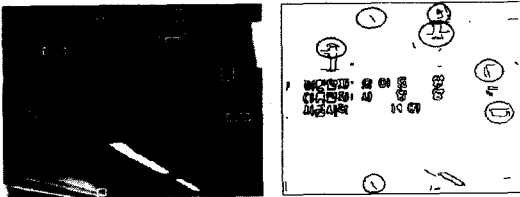


그림 16 텍스트 후보영역의 에지 빈도를 이용한 검증 결과

텍스트 열로 구성하기 위해서 외곽사각형의 겹침 정도와 외곽사각형 사이의 거리를 이용하여 하나의 외곽사각형으로 결합시킨다. 이 때 대부분의 텍스트 열이 수평 방향으로 존재하기 때문에 수평방향 위주로 결합을 수행한다. 그림 17은 겹침 정도는 83%, 거리는 12 화소로 설정하여 결합한 결과의 후보영역들을 보여준다.

3.3 색 변화를 이용한 추출

실내외에서 취득한 자연이미지는 빛과 조명의 영향으로 또는 인위적인 의도에 의해서 텍스트와 배경이 명확하게 구분되지 않는 경우도 빈번하게 발생한다. 이러한 경우 색 변화를 특징으로 이용한다면 보다 정확한 영역 추출을 기대할 수 있다. 이 방법은 색 연속성과 밝기 변화를 이용한 텍스트 영역 추출 방법이 조명의 변화에 따라 색 병합의 오류와 에지 추출의 오류가 쉽게 발생하여 생기는 오검출의 빈도수를 줄일 수 있다. 그림 18(a)은 주어진 이미지에 대하여 색 연속성을 이용한 추출 결과로서 텍스트의 색이 흐리고 조명 변화로 인해서 색 병합에서 오류가 발생하여 영역을 정확하게 찾지 못한 경우이다. 그림 18(b)은 밝기 변화를 이용한 추출 결과로서 역시 조명 변화로 인해서 에지가 잘 찾아지지 않아서 발생한 오류를 보여준다. 그림 18(c)은 색 변화를 이용한 추출 결과로서 앞의 오류가 해결된 것을 볼 수 있다. 이와 같이 색 변화를 이용한다면, 아주 작은 색의 변화도 찾을 수 있는 장점이 있다.

색 변화 정도는 분산으로 표현할 수 있는데, 분산은 명도(Gray) 값 또는 색 값에서 구할 수 있다. 명도이미지의 밝기 값으로 분산을 구한다면 문서이미지와 같이 텍스트와 배경의 구분이 뚜렷한 경우 분산이 크게 나타나지만, 밝기 값의 변화가 작은 경우 분산도 작게 표현된다. 그러나 RGB로 표현된 색 정보는 3차원 RGB 공간에서 화소 값이 표현되어 색 사이의 거리를 분산으로 정확하게 표현할 수 있다. 1차원 명도이미지에서 분산이 작은 이유는 3차원 공간에서 거리가 먼 화소들의 위치가 동일 선상에 선형으로 투영되어 색 변화 값이 축소되기 때문이다. 따라서 색 이미지에서의 분산은 명도이미지에서의 분산보다 크며 변화를 정확하게 측정할 수 있는 장점이 있다. 그림 19(b)는 원 이미지에서 밝기 값

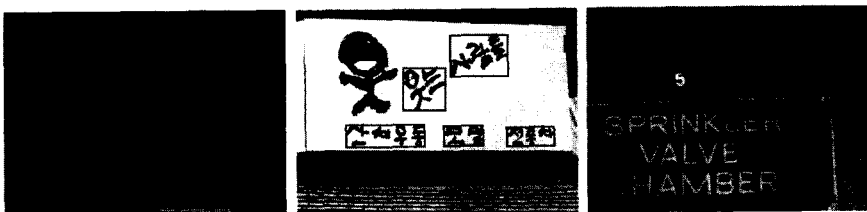
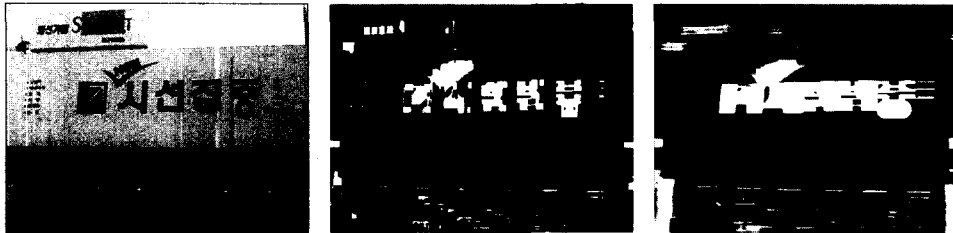


그림 17 밝기 변화를 이용한 텍스트 추출 결과 예



(a) 색 연속성 이용 (b) 밝기 변화 이용 (c) 색 변화 이용
그림 18 세 종류의 특징을 이용한 영역 추출 비교



(a) 원 이미지 (b) 명도 분산 결과 (c) 색 분산 결과
그림 19 명도 및 색 분산 비교



(a) 원 이미지 (b) 수평 분산이미지 (c) 수직 분산이미지
그림 20 수평 및 수직 분산이미지

으로 분산을 계산하여 표현한 이미지이며, 그림 19(c)는 3차원 RGB 색 값으로 분산을 계산하여 표현한 이미지이다. 그림에서 글자와 배경의 변화가 보다 명확하게 표현된 것을 확인할 수 있다.

3.3.1 수평 및 수직 분산이미지 생성

320x240 크기의 입력된 색 이미지에 대해서 1x21 크기의 수평 방향 윈도우와 17x1 크기의 수직 방향 윈도우를 각각 설정하여 식 (1)을 이용하여 분산을 구한다. 여기에서 v_H 와 v_V 는 수평 및 수직 방향의 분산을 의미하며, R, G, B 는 화소에 대한 색 영역의 값이고, v_{HR} 은 R 색상의 수평 방향 평균을 의미한다. 분산을 구하는 윈도우의 크기는 실험에 의해서 결정하였다. 구해진 분산은 식 (2)와 (3)을 이용하여 값의 범위를 0과 255 사이의 값으로 정규화 시킨다. 우선 분산이 일정 수준 이상으로 큰 값들은 하나의 큰 값으로 대체해 줄 수 있도록 식 (2)를 사용하여 값의 범위를 줄이고, 식 (3)을 사용하여 최대최소 정규화(Min-max Normalization)

를 수행한다. 여기에서 max_1 과 min_1 은 분산 v_{*w} 의 최대, 최소 값이며 max_2 와 min_2 는 새로운 범위의 명도 값을 의미한다.

$$v_H = \frac{1}{21} ((R_i - \mu_{HR})^2 + (G_i - \mu_{HG})^2 + (B_i - \mu_{HB})^2) \quad (1)$$

$$v_V = \frac{1}{17} ((R_i - \mu_{VR})^2 + (G_i - \mu_{VG})^2 + (B_i - \mu_{VB})^2) \quad (2)$$

$$v_{new} = \sqrt{v} \quad (2)$$

$$v_{normalized} = \left(\frac{v_{new} - \min_1}{max_1 - \min_1} \right) (max_2 - \min_2) + \min_2 \quad (3)$$

그림 20에서 (b)와 (c)는 정규화된 수평 및 수직 방향 분산 이미지이다. 원 이미지에서 색 변화가 큰 부분은 밝게 표현되고, 색 변화가 작은 부분은 어둡게 나타난다. 또한, 텍스트 영역은 수평 및 수직 방향의 분산이 모두 크게 나타나는 것을 확인할 수 있다.

3.3.2 AND 연산을 이용한 잡음 제거

텍스트 영역은 수평 및 수직 방향 분산이 모두 크게 나타나지만, 단순한 선들과 같은 불필요한 요소들은 수

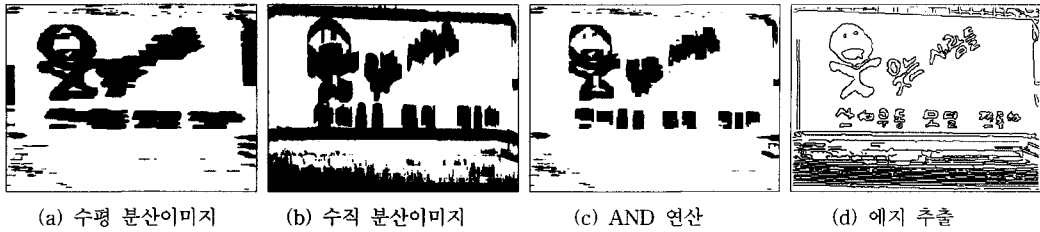


그림 21 수평 및 수직 분산이미지의 AND 연산 결과와 에지 추출 결과 비교

평 또는 수직의 한 이미지에서만 분산 값이 크게 나타난다. 따라서 두 분산이미지의 논리적 AND 연산을 수행한다면 선과 같이 불필요한 요소들을 쉽게 제거할 수 있다. 논리적 AND 연산을 수행하기 위해서 먼저 이진화를 수행하며, 이진화는 정규화된 분산이미지의 평균을 임계치로 사용하였다. 그림 21은 두 이미지의 논리적 AND 결과와 명도이미지에서의 에지 추출 결과를 비교한 것이다. (c)와 (d)를 비교한다면 명도이미지의 에지 추출에서 텍스트 주변에 분포하는 다양한 종류의 선들이 AND 연산으로 제거된 결과를 볼 수 있다.

3.3.3 영역 추출 및 검증

논리적 AND 연산을 수행한 결과 이미지에는 분산이 큰 색 변화가 심한 영역들이 남게 되지만 불필요한 잡영도 부분적으로 포함된다. 따라서 텍스트 영역의 처리를 쉽게 하고 간단한 잡영들을 제거시키기 위해서 모폴로지 연산을 수행한다. 모폴로지 연산은 수평 방향의 텍스트를 주로 강조하기 위해서 그림 22와 같은 2x5 크기의 불림 연산자와 3x3 크기의 녹임 연산자를 구성하여, 식 (4)와 같이 열림(Opening)과 닫힘(Closing)연산을 반복적으로 적용한다.

$$I = Closing + Opening + Opening \quad (4)$$



그림 22 불림 구조 연산자(좌), 지움 구조 연산자(우)

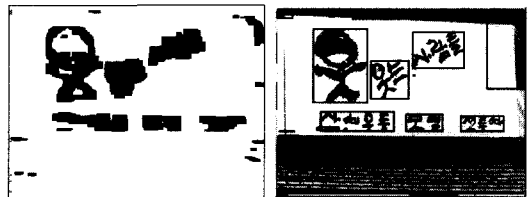


그림 23 (a) 모폴로지 적용, (b) 연결요소의 외곽사각형 추출

그림 23(a)은 그림 21(c)의 AND 연산 결과에 제한한 모폴로지 연산자를 적용한 결과로서 부분적으로 잡영이 제거되고 텍스트 영역은 강조된 것을 확인할 수 있다. 그림 23(b)는 연결요소와 연결요소의 외곽사각형을 구한 결과이며, 외곽사각형의 위치, 크기, 종횡비(Aspect ratio)와 연결요소의 크기 등이 저장된다.

후보영역의 검증과 결합은 밝기 변화를 이용한 방법과 동일한 방법으로 연결요소와 외곽사각형을 분석하여 검증하고, 겹침 정도와 외곽사각형 사이의 거리를 이용하여 외곽사각형을 결합한다. 그림 24는 최종적으로 추출된 텍스트 영역을 보여준다.

3.4 추출된 영역의 결합

색 연속성, 밝기 변화 및 색 변화 특징을 이용한 추출은 상호 보완적인 역할을 할 수 있기 때문에 각 방법에서 찾은 영역들을 결합하면 텍스트 영역 추출의 정확성을 향상시킬 수 있다. 본 논문에서는 그림 25와 같이 각 방법에서 찾은 영역의 겹침 정도를 서로 비교해서, 검증

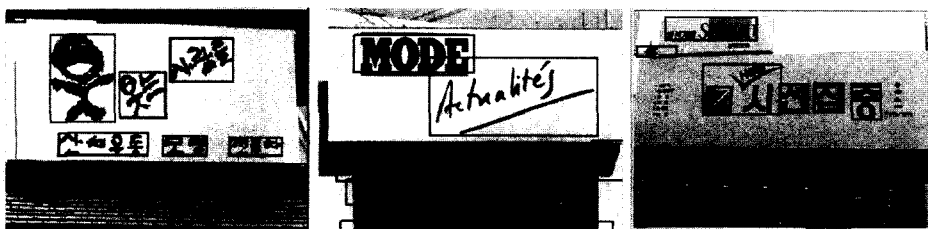


그림 24 색 변화를 이용한 텍스트 추출 결과

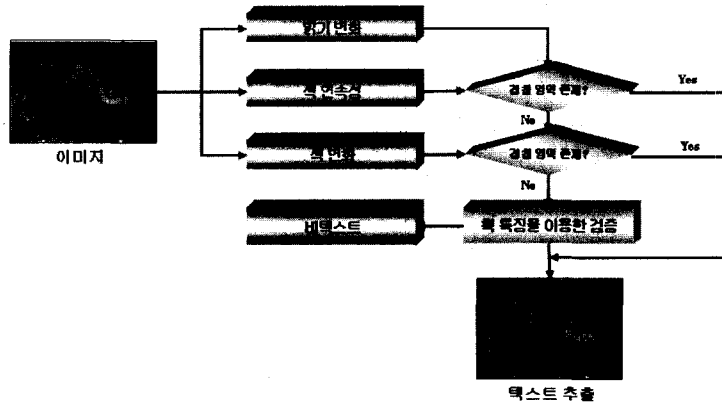


그림 25 결합

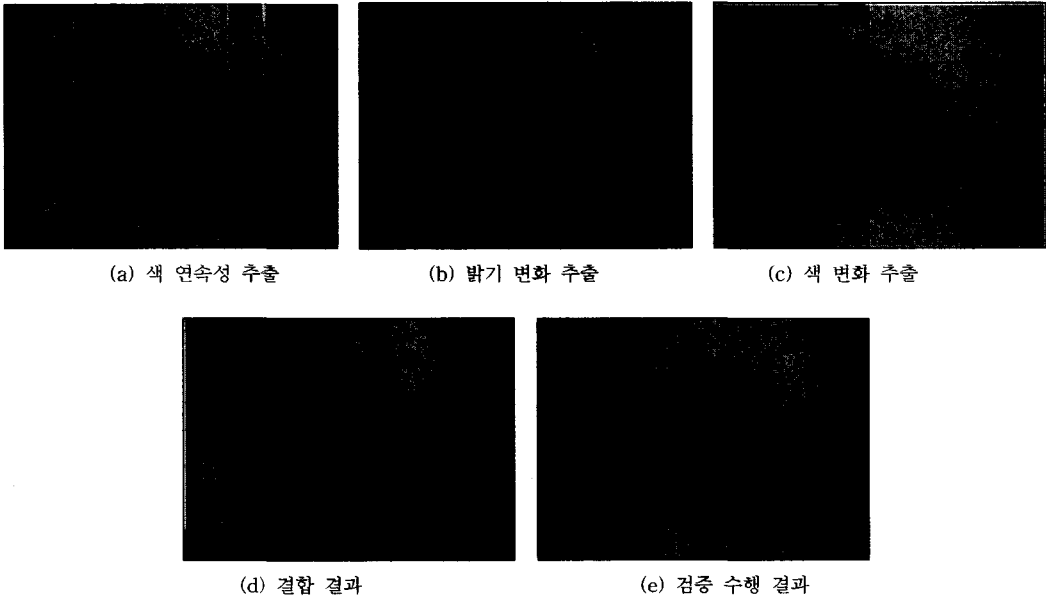


그림 26 결합 및 검증 결과

으로 확인해야 하는 후보영역과 검증을 생략하고 텍스트 영역으로 최종 확정하는 후보영역으로 구분한다.

각 특징을 이용한 텍스트 추출 결과를 분석하면 색 연속성을 이용한 방법은 빛 또는 조명에 민감하여 색 병합이 잘 이루어지지 않지 때문에 틀리게 추출된 영역(False Accept)과 찾지 못한 영역(False Reject)이 존재한다. 밝기 변화를 이용한 방법은 조명 변화로 인해서 텍스트와 배경 사이에 명도 대비가 크지 않아 예지를 정확하게 찾지 못하고, 배경이 복잡할 경우에는 텍스트 주변에 나뭇가지나 타일 등과 같이 많은 예지가 존재하여 오류가 발생하는 경우가 많다. 색 변화를 이용한 방

법은 기울어진 텍스트와 배경이 복잡한 이미지에 대한 오류가 발생하기 쉽다. 이와 같이 각 방법의 단점을 파악하여 상호 보완적으로 결합을 한다면 보다 정확한 추출이 가능하다.

결합은 각 특징에서 추출한 영역의 상호 겹침 정도를 비교하여 두 영역이 80% 이상 겹침 경우 동일한 영역으로 판단하여 검증 과정 없이 텍스트 영역으로 확정한다. 겹침 정도가 80% 이하이거나 겹침이 없는 경우에는 검증한다. 검증은 획 기반 특징으로 학습된 SVM 분류기를 이용하여 텍스트 영역과 비텍스트 영역으로 구분한다. 그림 26은 결합한 결과의 예를 보여준다. 그림

26(e)의 ①번 영역은 세 특징 결과에서 모두 존재하기 때문에 검증 없이 텍스트 영역으로 확정되지만, 그림 26(d)의 ④-⑥ 영역은 하나의 특징에서만 존재하는 영역으로서 SVM 분류기로 검증되어 비텍스트 영역으로 분류되어 제거된 영역이다. 검증 과정은 다음의 4절에서 설명한다.

4. 검증

본 논문에서는 다해상도(Multi-resolution) 웨이블릿(Wavelet) 변환을 수행하여 텍스트 구성단위인 획을 추출한 후 학습된 SVM 분류기를 이용하여 검증하는 방법을 제안한다. 문자의 획은 문자를 표현하는 기본 단위로서 한글, 영어, 숫자 등이 모두 획의 조합으로 표현된다. 작은 이미지 영역에서의 획의 구성 여부를 웨이블릿 변환을 이용한 다차원 특징벡터로 표현하고, SVM의 학습 과정에서 생성된 SV(Support Vector)와 비교함으로써 검증한다. SVM은 간단한 수식의 표현으로 복잡한 패턴인식 문제를 해결하는 방법으로서 텍스트와 비텍스트를 구분하는 것과 같은 이진 분류 문제에 적합하다 [17]. 또한, 자연이미지 내의 장면 텍스트는 다양한 크기로 존재하기 때문에 다해상도의 웨이블릿 변환을 이용하여 크기 변화에도 강인하게 만들었다.

4.1 웨이블릿 변환

웨이블릿 변환은 이미지에 수평 및 수직 방향으로 필터를 적용하여 고주파 성분을 갖는 LH, HL, HH 부밴드(Subband)와 저주파 성분의 LL 부밴드로 이미지를 분해한다. 각 부밴드들은 서로 다른 특성을 나타내는데 LL은 원 이미지를 축소한 것과 같으며, LH는 수평 성분, HL은 수직 성분, HH는 대각 성분에서 웨이블릿 계수 값이 크게 나타나며, 이는 수평, 수직 및 대각선 방

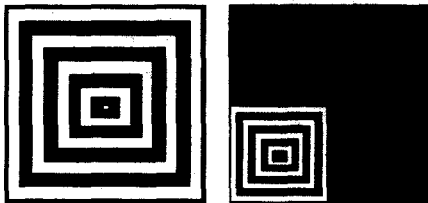


그림 27 입력 이미지의 1차 웨이블릿 변환

향의 획의 존재 여부를 파악할 수 있게 해준다. 그림 27은 1차 웨이블릿 변환으로 얻어진 4개의 부밴드를 보여준다.

4.2 특징벡터 추출

16x16 크기의 작은 이미지 영역을 대상으로 해상도를 반복적으로 1/2로 축소하면서 웨이블릿 변환을 적용하여 특징벡터를 추출한다. 특징벡터는 웨이블릿 변환으로 생겨난 부밴드들의 평균(M), 2차(μ_2), 3차(μ_3) 중앙 모멘트(Central Moment)를 사용한다[18]. NxN 크기의 부밴드 I 에 대한 특징벡터들은 식 (5)로 계산된다. 16x16 크기의 영역은 3-레벨 웨이블릿 변환이 수행되어 12개의 부밴드를 만들며, 전체 36차원의 특징벡터가 생성된다.

$$M(I) = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} I(i, j)$$

$$\mu_2(I) = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (I(i, j) - M(I))^2 \quad (5)$$

$$\mu_3(I) = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (I(i, j) - M(I))^3$$

4.3 특징 선택

다해상도 웨이블릿 변환으로 구한 36차원의 특징벡터의 차수를 줄이는데, 이는 처리 시간의 감소와 변별력이 없는 특징을 제거하기 위한 것이다. 특징을 선택하기 위해서 16x16 크기의 텍스트 영역 이미지 데이터 1,000개와 비텍스트 영역 이미지 데이터 1,000개를 웨이블릿 변환하여 36차원의 특징을 추출한다. 텍스트 영역과 비텍스트 영역 각 500개의 데이터를 학습 데이터로 사용하고, 나머지 500개는 테스트 데이터로 사용한다. 36개 각각의 특징에 대한 학습 및 테스트 데이터를 이용하여 Bayes 에러율을 구하고 에러율이 작은 12개의 특징을 선택하였다. 비텍스트 영역의 학습 및 테스트 데이터에는 텍스트 영역과 유사한 영역들이 많이 사용되었다.

4.4 SVM 학습

12차원의 특징벡터를 사용하여 SVM 분류기를 학습하며, 각 특징은 -1과 1사이로 정규화 시킨다. 표 1은 학습 데이터의 수에 따른 인식률을 나타내며, 텍스트 및 비텍스트 영역의 데이터 비율을 1:1로 같게 하였다. 인식률을 측정하기 위한 테스트 데이터는 텍스트 영역 데이터 1,000개와 비텍스트 영역 데이터 2,000개를 각각

표 1 학습데이터의 수에 따른 블록(16x6크기) 단위의 인식률

학습데이터 수	Support Vector 수		인식률(%)		
	텍스트	비텍스트	텍스트	비텍스트	평균
300	85	86	64.9	81.8	73.4
500	107	84	81.4	86.7	84.0
1,500	256	253	86.1	87.4	86.8
2,000	346	343	85.8	88.6	87.2

사용하였다. 학습 데이터의 수가 증가하면 인식률은 증가하지만, SV(Support Vector)의 개수도 증가한다. 본 연구에서는 학습 데이터의 개수를 1,000개로 정하여, 그때 만들어진 SV를 사용하여 분류한다.

4.5 텍스트 후보 영역 검증

결합 과정에서 하나의 특징 결과에서만 존재하는 후보 영역을 다해상도 웨이블릿 특징과 SVM 분류기로 검증한다. 검증이 필요하다고 판단된 영역을 여러 개의 단위 블록으로(16x16) 나누어, 각 블록의 텍스트 영역 여부를 판단하고 각 블록의 결과들을 종합하여 전체 영역을 판단한다. 검증할 영역에 대해 16x16 크기의 윈도우를 띄우고 12차원의 특징벡터를 추출하여 SVM 분류기로 판단한 후 윈도우를 오른쪽 또는 아래로 네 화소를 이동하여 같은 방법으로 판단한다. 텍스트 블록으로 판정된 영역이 검증하고자 하는 영역의 2/3 이상이면 검증 영역 전체를 텍스트 영역으로 판단한다. 그림 28은 검증한 결과의 예를 보여준다.

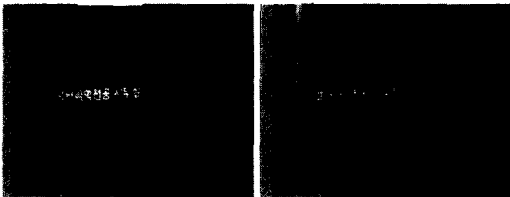


그림 28 영역 추출 결합 및 SVM을 이용한 검증 결과

5. 실험 및 결과

본 연구에서 제안한 방법은 Windows 2000에서 Visual C++ 6.0과 펜티엄 IV 1.8GHz 하드웨어를 사용하였다. 제안한 방법의 평가를 위해서 디지털 카메라를 사용하여 640x480 해상도로 학교, 병원, 지하철 역, 도로 등의 실내의 환경에서 취득한 120개의 자연이미지, 200개의 실외 간판이미지 및 ICDAR에서 제공한 367개

의 자연이미지를 사용하였다. 텍스트 영역의 Ground Truth 기준으로서 수평 방향으로 글자 간격이 너무 멀어진 텍스트 영역과 수직 방향의 텍스트 영역은 각 글자를 구분하여 텍스트 영역으로 설정하였다. 또한, 개발자의 눈으로 인식 가능한 크기의 텍스트만을 추출 대상으로 설정하여 16x16 크기 이상의 텍스트 영역만을 고려하였다.

실험실에서 구축한 이미지에 대해서는 배경의 복잡도에 따른 성능을 비교하기 위해서 단순한 이미지와 복잡한 이미지로 분류하였다. 단순한 이미지는 적당한 크기의 텍스트를 포함하며 배경과의 구분이 명확하며, 기술어짐과 원근이 비교적 적은 것들이며, 복잡한 이미지는 단순한 이미지에 속하지 않는 나머지 이미지들이다. 간판 이미지와 ICDAR 이미지는 오류 분석을 위해 주로 테스트 데이터로 사용하였으며, ICDAR 이미지 데이터는 학습 데이터와 테스트 데이터로 분류되어 있다.

실험 결과는 표 2와 같으며, Total는 전체 실험 대상 이미지에서의 텍스트 영역의 개수이며, Correct는 정확히 찾은 텍스트 영역의 개수, Partial은 텍스트 영역의 일부만이 찾아진 개수, Missing은 텍스트 영역을 찾지 못한 개수이며, False는 텍스트가 아닌 영역을 텍스트로 찾은 개수이다. 성능을 측정하는 단위로서 식 (6)의 정확률(Precision)은 전체 찾아진 영역 중에서 정확히 찾은 텍스트 영역의 비율로서 정확성을 표현하며, 재현률(Recall)은 텍스트 영역의 개수 중에서 맞게 찾아진 영역의 비율로서 찾지 못하는 것을 파악하기 위한 것이다[19].

$$Precision = (Total\ number\ of\ correct\ texts) / (Number\ of\ detected\ texts)$$

$$Recall = (Total\ number\ of\ correct\ texts) / (Total\ number\ of\ texts) \quad (6)$$

표 2의 실험실에서 구축한 자연이미지는 배경의 복잡도에 따른 추출 성능을 보기 위한 것으로서 단순한 이미지와 복잡한 이미지 모두 재현률이 높은 좋은 결과를

표 2 각 실험 이미지에 대한 추출 결과

		영역 개수	256	239	2	Missing	15	False	78
복잡한 구조 의 이미지	간판 이미지	영역 개수	256	239	2	15	78		
		정확률 = 75.5%				재현률 = 94.1%			
간판 이미지	복잡 이미지	영역 개수	328	291	13	24	148		
		정확률 = 67.2%				재현률 = 92.6%			
간판 이미지	간판 이미지	영역 개수	839	645	50	144	395		
		정확률 = 63.7%				재현률 = 82.8%			
ICDAR 문서도 이미지	간판 이미지	영역 개수	378	235	18	125	142		
		정확률 = 64.0%				재현률 = 66.9%			
ICDAR 문서도 이미지	텍스트 이미지	영역 개수	687	474	29	184	346		
		정확률 = 59.2%				재현률 = 73.2%			

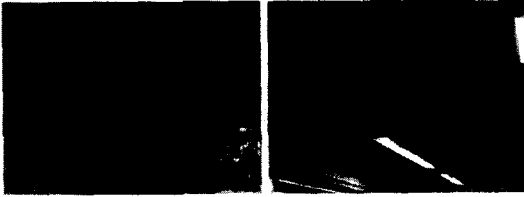


그림 29 자연이미지에 대한 영역 추출 결과 예

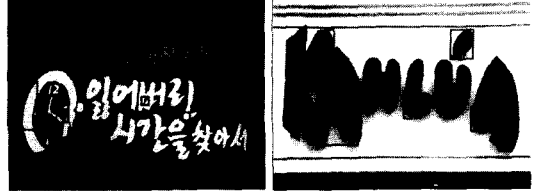


그림 30 간판이미지에서 발생한 오류의 예

확인할 수 있다. 결합 단계에서 각 특징에서 추출된 *False* 영역이 두 개 이상의 특징 결과에 함께 존재하여 검증 과정 없이 텍스트로 인정되는 영역이 많아 전반적으로 *False* 개수가 증가하여 추후에 결합 과정의 보완이 필요하다. 그림 29는 자연이미지에 대한 영역 추출 결과의 예를 보여준다.

표 2에서 간판 이미지는 실외에서 수집한 200개의 다양한 크기, 폰트, 방향의 문자열을 포함하며, 대부분 간판 텍스트에 초점을 맞추어 수집된 이미지들이다. 재현률이 높게 나타났지만, *Missing*과 *False* 개수가 많아 정확률은 낮게 나타났다. 이러한 결과는 주로 텍스트의 크기가 전체 이미지에서 차지하는 비율이 높고, 그래픽 폰트의 텍스트들이 많기 때문에 오류가 발생한 것으로 파악된다. 본 연구에서는 텍스트 크기에 제약을 두어 텍스트의 높이가 이미지 전체 높이의 1/3 이하로 제한하였으며, 이에 대한 보완도 필요하다. 그림 30은 크기와 폰트 스타일 등에 의해 오류가 발생한 예를 보여준다.

ICDAR 컨테스트 이미지는 텍스트 추출 경연대회에서 사용하기 위해서 116개의 학습용 이미지와 251개의

테스트용 이미지로 나누어 구성되어 있으며, 본 연구에서는 모두 테스트용으로 사용하였다. ICDAR 이미지에 포함되어 있는 텍스트는 모두 영문자와 숫자이며, 간판 이미지와 같이 폰트의 크기, 모양, 조명 변화 등이 다양하다. ICDAR 이미지 테스트 결과에 대한 다양한 종류의 오류 예를 그림 31에서 볼 수 있다. (a)는 주로 SVM 검증에서 걸러내지 못해서 발생한 오류들로서, SVM 분류기를 학습시킬 때 포함되지 않은 새로운 패턴들로 인해서 발생한 것으로 판단된다. (b)는 복잡한 배경으로 인해서 발생한 오류이며, (c)는 간판 이미지와 마찬가지로 문자의 크기가 너무 커서 찾지 못한 오류이다. (d)는 밝기 및 색 변화가 거의 없는 양각 효과로 인해서 특징이 제대로 찾아지지 않아서 발생한 오류이고, (e)는 심한 조명으로 인해서 발생한 오류의 예이다. 이와 같은 유형의 오류를 줄이기 위한 방법을 향후 연구에서 고려하고자 한다.

텍스트 영역의 추출 시간은 배경의 복잡도에 따라 단순한 이미지와 복잡한 이미지로 분류하여 측정하였다. 표 3은 이미지 수준의 세 가지 특징에서의 처리 시간과



(a) SVM 검증 오류

(b) 복잡한 배경으로 인한 오류

(c) 문자 크기에 의한 오류



(d) 양각 효과에 의한 오류

(e) 심한 조명에 의한 오류

그림 31 ICDAR 이미지에서 발생한 오류의 예

SVM을 이용하여 검증하는데 소요되는 평균 처리 시간을 보여준다. 색 연속성은 배경의 복잡도에 무관하게 비슷한 처리시간을 보였으나, 밝기 변화에서는 에지 추출 및 선 제거, 색 변화에서는 색 분산 이미지 생성 과정에서 배경에 복잡도에 따라 처리 시간이 차이남을 확인할 수 있다. 단순한 이미지와 복잡한 이미지 사이에 약 1,000msec 정도의 차이가 있었다. 처리 시간을 더 단축시키기 위한 각 모듈에 대한 최적화 과정이 필요하다.

그림 32부터 그림 36은 다양한 이미지들에 대한 제안 방법의 결과들이다. 텍스트 영역을 찾지 못하거나, 텍스트가 아닌 부분을 텍스트로 찾는 오류도 많이 발생했지만, 대부분의 경우 잘 찾은 결과를 볼 수 있다. 또한, 틀리거나 찾지 못한 영역에 대해서는 그림 31에서 언급한 문제점에 대한 제안한 방법의 한계로서 향후 연구에서 고려하고자 한다.

6. 결론

본 논문에서는 낮은 수준의 이미지 특징과 높은 수준의 텍스트 특징을 계층적으로 결합하고 검증하여 자연 이미지에 포함된 텍스트 영역을 정확하며 안정적으로 추출할 수 있는 방법을 제안하였다. 이미지 수준의 특징으로는 색 연속성, 밝기 변화, 색 변화를 이용하였고, 텍스트 특징으로는 웨이블릿 변환을 이용한 획 특징을 사용하였다. 이 과정에서 특징 개발 및 선택, 안정적인 색 병합 방법의 제안, 긴 선 추출 및 제거에 의한 이미지 단순화 방법 제안, 색 변화 특징을 효율적으로 사용하는 방법의 제안 및 결합 방법, 획 특징과 SVM을 이용한 효율적인 검증 방법 등을 새롭게 제안하였다.

제안한 방법을 다양한 종류의 자연이미지에 테스트한 결과 배경의 복잡도, 다양한 조명 여건, 폰트 종류 및 방향 등에 비교적 무관하게 높은 재현률을 보여서 응용 시스템에의 적용 가능성을 높였다. 그러나 현재까지의 연구로는 문자의 크기가 너무 크거나 조명 변화가 아주 심한 경우에는 텍스트 영역을 찾지 못하는 단점이 있다. 향후 연구에서는 이러한 오류의 원인을 정확히 분석하여 제안한 방법을 개선하고, 실시간 응용을 위한 처리 시간 단축 방법을 고려하고자 한다.

본 논문에 실려 있는 다양한 종류의 컬러 그림 파일을 확인하기 위한 논문 원본 PDF 파일이 "<http://cs.sookmyung.ac.kr/~ywchoi>"에 있다.

참고 문헌

- [1] J. C. Shim, C. Dorai, R. Bolle, "Automatic Text Extraction from Video for Content-Based Annotation and Retrieval," Proceedings of 14th International Conference of Pattern Recognition, Vol. 1, pp. 618-620, Brisbane, 1998.
- [2] 장인영, 고병철, 김길천, 변혜란, "모폴로지를 이용한 비디오 영상에서의 자동 문자 추출", 정보과학회 추계 학술대회 논문집, pp. 418-420, 2001.
- [3] M. Bertini, C. Colombo, A. Del Isimbo, "Automatic Caption Localization in Videos using Salient Points," IEEE International Conference on Multimedia and Expo, pp. 69-72, 2001.
- [4] Y. Zhong, H. Shang 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.
- [5] F. LeBourgeois, "Robust Multifont OCR System from Gray Level Images," International Conference on Document Analysis and Recognition, Vol. 1, pp. 1-5, 1997.
- [6] R. Lienhart, F. Stuber, "Automatic Text Recognition in Digital Videos," Image and Video Proceeding IV, The International Society for Optical Engineering(SPIE), 1996.
- [7] M. A. Smith, T. Kanade, "Video Skimming for Quick Browsing Base on Audio and Image Characterization," Technical Report CMU-CS-95-186, Carnegie Mellon University. July 1995.
- [8] Chuang Li, Xiaoqing Ding, Youshou Wu, "Automatic Text Location in Natural Scene Images," Document Analysis and Recognition, Proceedings Sixth International Conference, pp. 1069-1073, Sept 2001.
- [9] Hao Wang, "Automatic Character Location and Segmentation in Color Scene Images," Image Analysis and Processing, Proceedings 11th International Conference, pp. 2-7, 2001.
- [10] Anil K. Jain, Bin Yu, "Automatic Text Location in Images and Video Frames," Pattern Recognition, Vol. 31, No. 12, pp. 2055-2076, 1998.
- [11] Yu Zhong, Kalle Karu, Anil K. Jain, "Locating Text in Complex Images," Pattern Recognition, Vol. 28, No. 10, pp. 1523-1535, 1995.
- [12] H. K. Kim, "Efficient Automatic Text Location Method and Content-based Indexing and Structuring of Video Database," Journal of Visual Communications and Image Representation, Vol. 7, pp. 336-344, 1996.

표 3 배경의 복잡도에 따른 평균 처리 시간(단위: msec)

배경	175	421	726	410	1,732
단순한 이미지	175	421	726	410	1,732
복잡한 이미지	190	906	741	801	2,638

- [13] Pyeoung-Kee Kim, "Automatic Text Location in Complex Color Images using Local Color Quantization," TENCON 99. Proceedings of the IEEE Region 10 Technical Conference, Vol. 1, pp. 629-632, 1999.
- [14] 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.
- [15] 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.
- [16] John Canny, "A Computational Approach to Edge Detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI 8(6), pp. 679-698, 1986.
- [17] C. Cortes, V. Vapnik, "Support Vector Networks," In Proceedings of Machine Learning, Vol. 20, pp. 273-297, 1995.
- [18] Huiping Li, David Doermann, and Omid Kia, "Automatic Text Detection and Tracking in Digital Video," IEEE Transactions on Image Processing, Vol. 9, No. 1, pp. 147-156, 2000.
- [19] S. M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong and R. Young, "ICDAR 2003 Robust Reading Competition," International Conference on Document Analysis and Recognition(ICDAR), Vol. 2, pp.682-687, 2003.

부 록

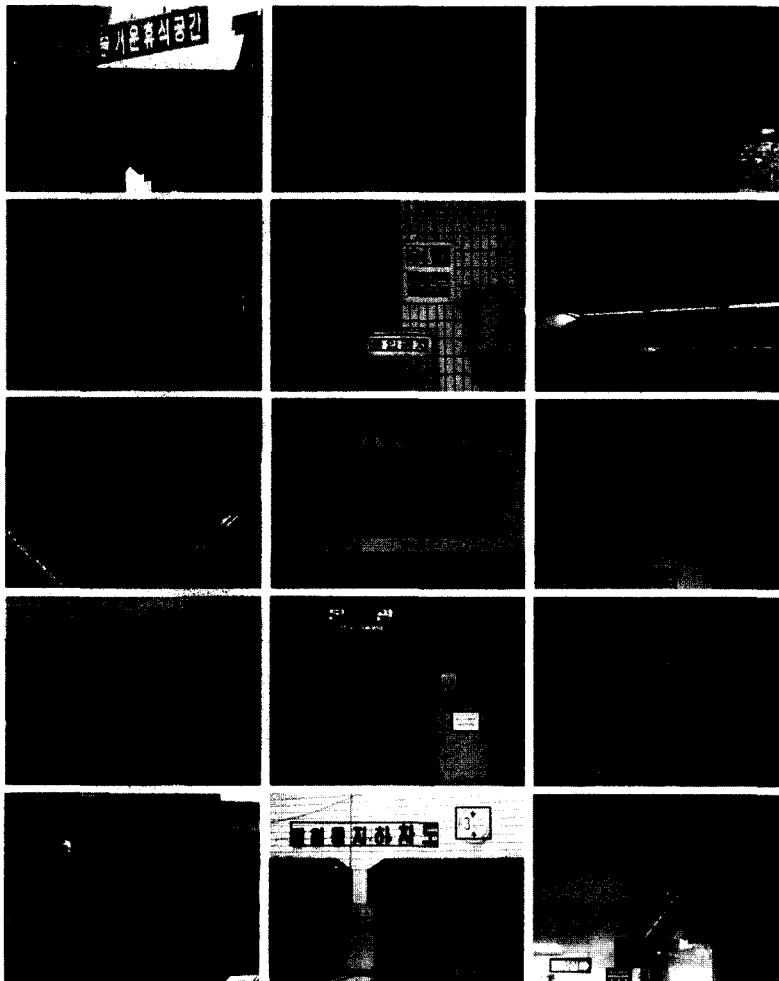


그림 32 실험실 구축 자연이미지의 추출 결과

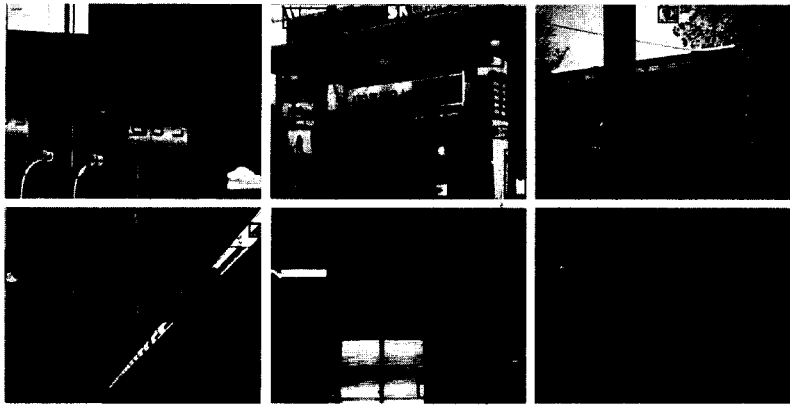


그림 33 실험실 구축 자연이미지의 오류의 예

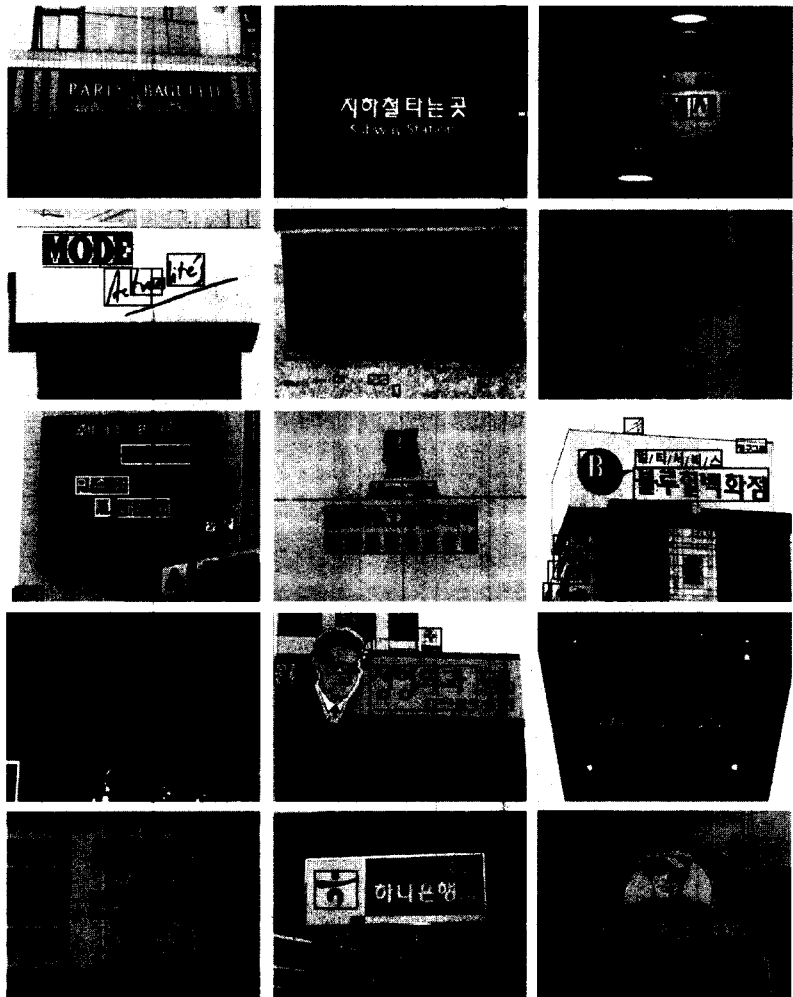


그림 34 간판이미지 추출 결과

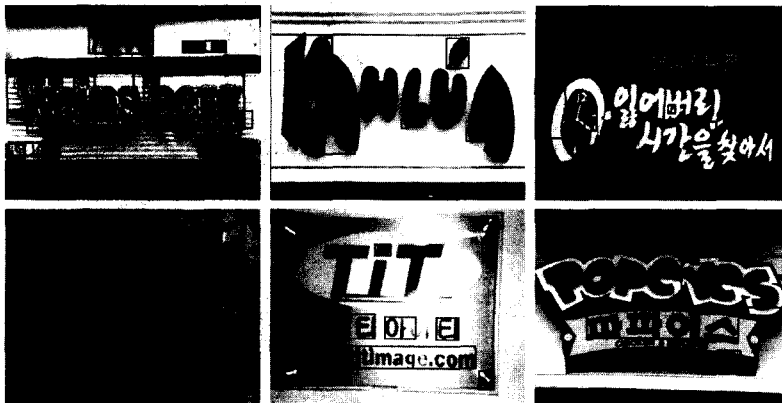


그림 35 간판이미지의 오류 예

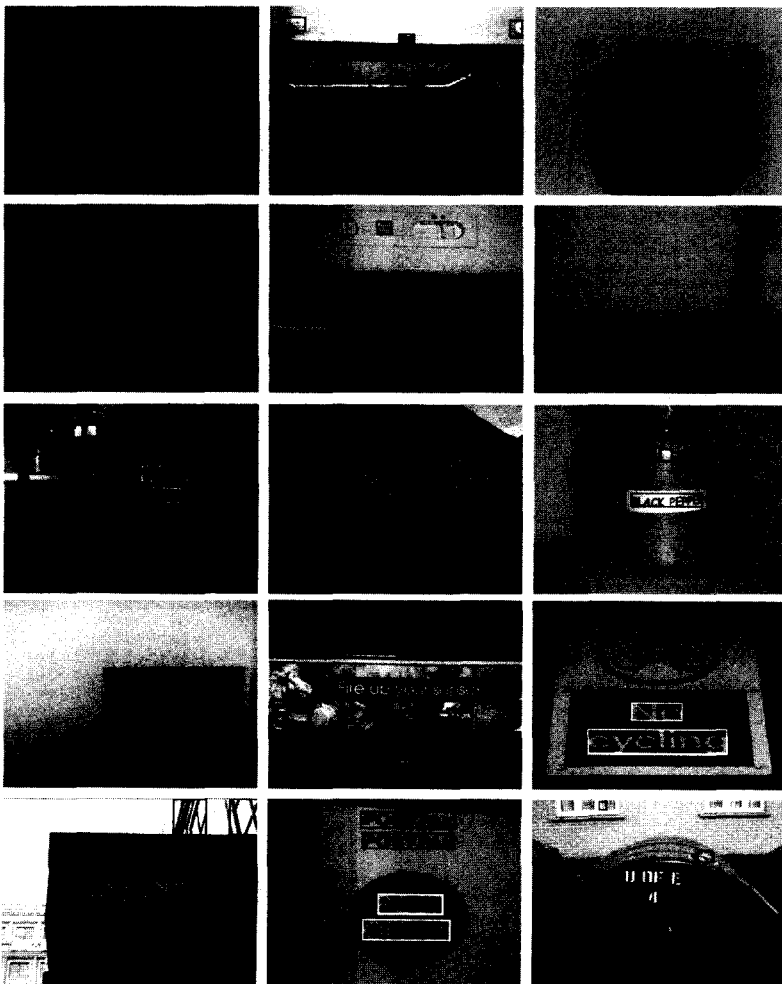


그림 36 ICDAR 컨테스트 이미지 추출 결과



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 정보과학부 조교수, 부교수
관심분야는 영상처리, 패턴인식, 문자인식 등



이 성 환

1989년~1994년 충북대학교 조교수. 1995년~2000년 고려대학교 부교수. 2000년~2001년 MIT 인공지능연구소 객원교수. 1997년~현재 고려대학교 인공지능각연구소 연구소장. 2001년~현재 고려대학교 정교수. 관심분야는 영상처리, 비디오처리, 패턴인식, 문자추출, 인공지능



김 길 천

1995년~2001년 경기대학교(이학사). 2001년~현재 연세대학교 대학원(석사과정) 관심분야는 영상처리, 비디오처리, 패턴인식, 문자추출, 인공지능



변 혜 란

1980년 연세대학교 수학과 졸업(이학사) 1983년 연세대학교 대학원 수학과 졸업(이학석사). 1997년 University of Illinois, Computer Science(M.S.). 1993년 Purdue University, Computer Science(Ph.D.) 1994년~1995년 한림대학교 정보공학과 조교수. 1995년~1998년 연세대학교 컴퓨터과학과 조교수 1998년~2003년 연세대학교 컴퓨터과학과 부교수. 2003년~현재 연세대학교 컴퓨터과학과 교수. 관심분야는 인공지능, 영상인식, 영상처리



송 영 자

1985년~1989년 경북대학교(이학사). 1989년~1991년 2월 경북대학교(공학석사). 2001년 3월~현재 숙명여자대학교(박사과정). 관심분야는 영상처리, 문자추출



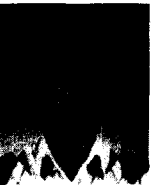
배 경 숙

1998년~2002년 2월 숙명여자대학교(이학사). 2002년 3월~2004년 2월 숙명여자대학교(공학석사). 관심분야는 영상처리, 문자추출



조 연 회

1998년~2002년 2월 숙명여자대학교(이학사). 2002년 3월~2004년 2월 숙명여자대학교(공학석사). 관심분야는 영상처리, 문자추출



노 명 철

1993년~2001년 강원대학교(공학사). 2001년~2003년 2월 고려대학교(공학석사). 2003년 3월~현재 고려대학교(박사과정) 관심분야는 영상처리, 비디오처리, 패턴인식, 문자추출, 인공지능