

뇌신경정보학연구사업

Brain Neuroinformatics Research Program

**이미지 및 동영상에 포함된 텍스트 정보 추출에
기반한 지능적 시각 응용시스템 구현에 관한 연구**

**Intelligent Visual System Design Based on Text
Extraction in Natural Scene Images**

숙명여자대학교

과 학 기 술 부

제 출 문

과학기술부 장관 귀하

본 보고서를 “이미지 및 동영상에 포함된 텍스트 정보 추출에 기반한 지능적 시각 응용시스템 구현에 관한 연구”의 보고서로 제출합니다.

2004. 5. 31

주관연구기관명 : 숙명여자대학교

주관연구책임자 : 최 영

연 구 원 : 송 영 자

” : 김 길 천

” : 노 명 철

” : 배 경 숙

” : 조 연 희

” : 고 혜 선

보고서 초록

과제관리번호	MI-0102-00-0009	해당단계 연구기간	2001. 08. 01 ~ 2004. 05. 31	단계 구분	2단계 / 총3단계
연구사업명	중 사업명	특정연구개발사업(국책연구개발사업)			
	세부사업명	뇌신경정보학연구사업			
연구과제명	중 과제명				
	세부(단위)과제명	이미지 및 동영상에 포함된 텍스트 정보 추출에 기반한 지능적 시각 응용시스템 구현에 관한 연구			
연구책임자	최 영 우	해당단계 참여연구원수	총 : 7 명 내부 : 0 명 외부 : 7 명	해당단계 연구비	정부: 128,600 천원 기업: 천원 계: 128,600 천원
연구기관명 및 소속부서명	숙명여자대학교 컴퓨터과학과		참여기업명		
국제공동연구	상대국명 :		상대국연구기관명 :		
위탁연구	연구기관명 :		연구책임자 :		
요 약				보고서 면수	85
<p>정지된 이미지나 동영상에 포함되어 있는 텍스트는 중요한 정보를 함축하고 있다. 이러한 정보를 실시간에 추출하여 인식할 수 있다면 다양한 분야에서 활용할 수 있다. 본 연구개발의 목표는 실내외에서 자연스럽게 얻어진 이미지 및 동영상에 포함되어 있는 장면 텍스트를 실시간에 추출하여 로봇 자동주행, 지능형 단말기의 정보 번역시스템과 같은 응용시스템에 활용하는 것이다. 본 연구에서 제안하는 텍스트 추출 방법은 텍스트의 색 연속성, 밝기 변화 및 색 변화와 같은 낮은 수준의 이미지 특징을 사용하여 텍스트 후보 영역을 찾은 후, 다해상도 웨이블릿 변환을 이용하여 높은 수준의 특징인 텍스트 획의 구성 여부로 검증하는 계층적인 구조를 제안한다. 색 연속성 특징은 대부분의 텍스트는 동일한 색으로 구성된다는 점을 이용하는 것이고, 밝기 변화 특징은 텍스트 영역은 주변과의 밝기 변화가 존재하며 에지 밀도가 높다는 점을 이용한다. 또한, 색 변화 특징은 텍스트 영역은 주변 배경과의 색 변화가 존재하며, 밝기 변화보다 민감한 색 분산 값으로 표현할 수 있다는 장점을 이용한다. 높은 수준의 텍스트 특징으로서 다해상도 웨이블릿 변환을 이용하여 텍스트 획의 방향성 정보를 추출하고, 추출된 정보를 SVM 분류기로 검증하여 최종 영역을 확정한다. 본 연구에서 제안한 방법을 다양한 종류의 자연이미지에 적용한 결과 배경이 복잡한 이미지에 대해서도 비교적 안정적으로 텍스트 영역을 찾는 것을 확인할 수 있었다. 또한, 3차년 연구에서는 제안한 방법을 지능형 단말기의 영상 인식시스템에 탑재하여 활용 가능성을 확인할 수 있었다.</p>					
색 인 어 (각 5개 이상)	한 글	텍스트 영역 추출, 계층적 특징, 색 병합, 색 분산, 다해상도 웨이블릿 변환, SVM			
	영 어	Text region extraction, Hierarchical features, Color clustering, Color variance, Multi-resolution wavelet transforms, Support Vector Machines			

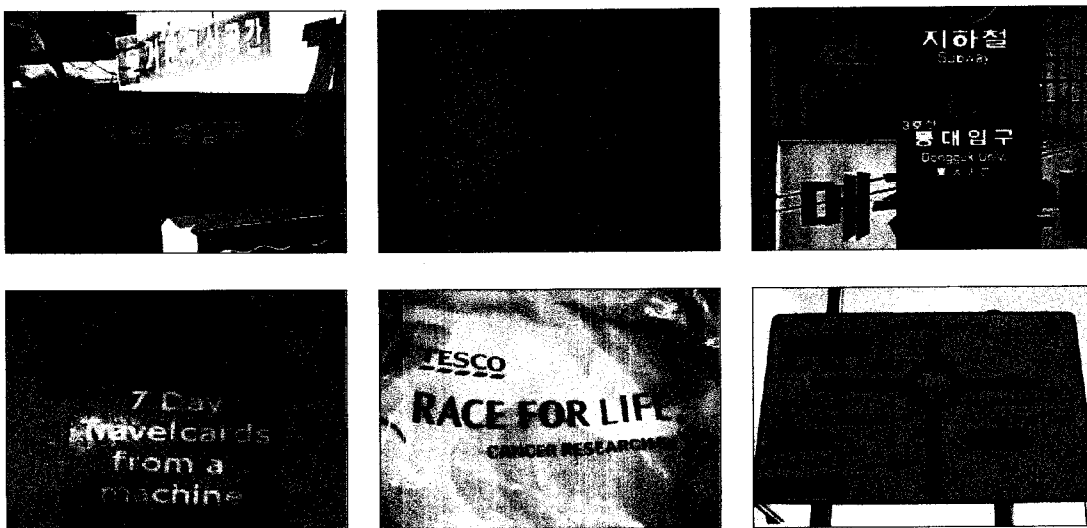
요 약 문

I. 제 목 : 이미지 및 동영상에 포함된 텍스트 정보 추출에 기반한 지능적 시각 응용시스템 구현에 관한 연구
(Intelligent Visual System Design Based on Text Extraction in Natural Scene Images)

II. 연구개발의 목적 및 필요성

정지된 이미지나 동영상에 인위적으로(Artificial) 또는 자연적으로(Natural) 포함되어 있는 텍스트는 중요한 정보를 함축하고 있다. 이러한 정보를 실시간에 추출하여 인식할 수 있다면 다양한 분야에서 활용될 수 있다. 그러나 어떠한 제약도 없이 자연스럽게 취득한 이미지에 포함되어 있는 장면(Scene) 텍스트를 정확히 찾아내는 것은 기술적으로 아주 어려운 일이다. 따라서 지금까지의 대부분의 관련 연구는 제한된 환경에서 취득한 이미지 또는 동영상에 인위적으로 삽입한 자막(Caption) 텍스트 추출에 초점이 맞추어져 있어서 폭넓은 분야에 적용하기에는 많은 문제점을 갖고 있다.

본 연구의 최종 목표는 자연이미지에 포함되어 있는 장면 텍스트 추출 및 인식으로서 폰트의 종류, 위치, 방향, 크기, 색, 배경 등에 비교적 무관한 추출 방법을 제안하고, 사용자의 불편을 최소화시킬 수 있는 수준의 속도 및 정확성으로 응용시스템에 활용하는 것이다. 따라서 1차년 연구에서는 영상처리 방법을 이용한 텍스트 영역 추출 방법을 제안하고, 2차년 연구에서는 인간의 시각적 주의 기능을 이용한 정보 결합 및 검증 방법을 개발하여 추출의 정확성을 향상시킨다. 3차년 연구에서는 극한 환경에서 취득한 이미지에 포함되어 있는 텍스트 영역의 추출로서 텍스트의 크기 변화, 조명 변화 및 시각 변화에 비교적 무관한 추출 방법을 개발하여 응용시스템에 적용한다. 본 연구개발 대상 이미지의 예는 그림 1과 같다.



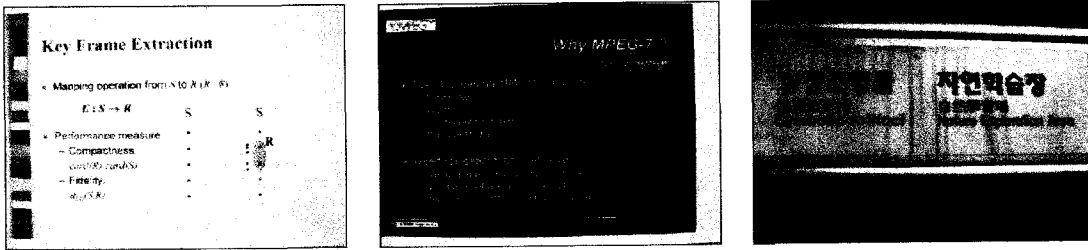


그림 1. 연구개발 대상 이미지의 예

III. 연구개발의 내용 및 범위

본 연구개발은 실내외에서 디지털 카메라, CCD 카메라 등을 이용해서 취득한 자연이미지에 포함되어 있는 장면 텍스트를 추출하고 응용하는 것으로서 연차별 연구개발 목표 및 내용을 정리하면 아래의 표와 같다.

구분	연구개발목표	연구개발내용 및 범위
1차년도 (2001.8-2002.5)	텍스트 영역 추출 방법 개발	<ul style="list-style-type: none"> - 실험 및 평가 이미지 DB 구축: 다양한 종류의 실내외 이미지 데이터 수집, 이미지 난이도에 따른 분류 - 색 정보를 이용한 영역 분리 방법 개발: 색 병합 및 추출 방법 개발 - 명도 변화를 이용한 영역 분리 방법 개발: 긴 선 추출 및 제거, 반복적 RLS 적용
2차년도 (2002.6-2003.5)	정보 선별 및 추출 성능 개선	<ul style="list-style-type: none"> - 정보 선별 방법 개발: 저 수준의 이미지 특징 검출 및 개선, 색 변화 특징 개발, 색 병합 및 선 처리 방법 개선, 특징 결합 및 검증 방법의 개발 - 추출 성능 개선: 영역의 기울어짐/원근 보정 방법 개발, 영역 이미지 개선 방법 개발, 에지 추출 시간 개선 - 평가 데이터 보완: 다양한 종류의 평가 이미지 수집
3차년도 (2003.6-2004.5)	극한 환경에서의 텍스트 정보 추출 및 응용시스템 구현	<ul style="list-style-type: none"> - 크기, 시각, 조명 변화에 무관한 추출 방법의 개발, 영역 결합 및 검증 방법 개선 - 지능정보 단말용 카메라로 취득한 색 문서이미지에서 텍스트 영역 추출 및 인식, 광범위한 학습 및 실험 DB 구축

IV. 연구개발결과

연 차	연구 내용	연구 결과 (요약)
1차년	색 정보를 이용한 텍스트 영역 추출	<ul style="list-style-type: none"> - 화소 병합, 보색 정보를 이용한 색 병합 방법의 개발 - 연결요소/외곽사각형의 단계적 제거, 결합에 의한 후보 영역 추출 및 검증 방법의 개발 - 정확성: 단순 이미지 Precision 58%, Recall 76%, 복잡 이미지 Precision 64%, Recall 71% - 처리 속도: 평균 173.6 msec/이미지
	명도 정보를 이용한 텍스트 영역 추출	<ul style="list-style-type: none"> - 긴 선 추출 및 제거, 반복적 RLS를 이용한 영역 추출, 검증 방법 개발 - 정확성: 단순 이미지 Precision 56%, Recall 88%, 복잡 이미지 Precision 48%, Recall 88% - 처리 속도: 평균 834.6 msec/이미지
	실험 및 평가 이미지 구축	<ul style="list-style-type: none"> - 다양한 종류의 실내외 자연이미지 수집, 복잡도에 따른 이미지 분류
2차년	정보 선별 방법 개발	<ul style="list-style-type: none"> - 색 병합 방법의 개선으로 안정적인 색 병합 결과를 얻음 - 명도이미지의 선 처리, 후보 영역의 강조 및 결합 방법의 개선으로 수직 문자열 추출 가능 - 색 변화를 고려한 색 분산 특징의 개발 - 색 연속성, 색 분산, 밝기 변화 특징 결과의 결합 방법 개발로 정보 선별 능력 향상 - SVM 및 다해상도 웨이블릿 특징을 이용한 검증 방법의 개발로 영역 추출의 정확성 및 신뢰성 향상
	추출 성능 개선	<ul style="list-style-type: none"> - 텍스트 주변의 긴 선들의 위치 및 방향 정보를 이용한 영역의 기울어짐 및 원근 보정 방법 개발 - 이미지 해상도 확대, 필터링에 의한 잡영 제거, 이진화 방법 개발 등의 단계적 이미지 개선으로 텍스트 영역의 이미지 품질을 인식 가능한 수준으로 향상시킴 - Canny 에지 추출 시간의 개선: Non-maximum suppression 과정을 개선하여 에지 추출 시간을 1/5에서 1/10 이상 감축시켰으며, 에지 이미지의 수준은 Gaussian 필터의 편차 조절로 유지함
	평가 데이터 보완	<ul style="list-style-type: none"> - 다양한 종류의 간판 이미지 추가, ICDAR 2003 학회의 텍스트 추출 경연 대회에서 사용하는 이미지를 평가 데이터로 보완

연 차	연구 내용	연구 결과 (요약)
3차년	극한 환경에서의 텍스트 정보 추출	<ul style="list-style-type: none"> - 이미지 레벨 및 특징 레벨에서의 다해상도 처리를 이용한 크기 변화에 무관한 영역 추출 방법 제안 - Template 매칭 및 영역 보정 방법을 이용해서 시각 변화에 따라 비선형적으로 변화, 이동하는 영역 추적 방법의 개발 - 색 변화 특징을 이용한 추출 방법 개발, 연세대에서 개발한 Color constancy 방법의 적용으로 조명 변화에 비교적 무관한 추출 방법 개발
	응용시스템 구현	<ul style="list-style-type: none"> - 지능정보단말용 카메라로 취득한 색 문서이미지 인식을 위한 전처리 방법 개발 및 적용 - 광범위한 학습 및 실험 DB 구축

V. 연구개발결과의 활용계획

연구개발 결과는 로봇 자동주행 및 감시시스템, 시각 장애인 보행 안내시스템, 공장 자동화, 자동 감시, 이미지/비디오 색인 및 검색 등에 다양하게 활용할 수 있다. 특히 본 연구의 3차년 연구에서는 한국전자통신연구원에서 개발하고 있는 지능형 단말기의 영상인식시스템에 텍스트 추출 모듈을 성공적으로 탑재하였다. 이 과정에서 지능정보 단말용 카메라로 취득한 발표 자료, 도서, 회의 자료 등과 같은 다양한 종류의 문서이미지에 대한 학습 및 실험 데이터베이스를 구축하였으며, 이러한 이미지들로부터 텍스트 영역을 추출하고 분리하는 방법을 개발하였다. 또한, 현재 개발된 소프트웨어 모듈이 산업체에서 활용할 수 있는 수준으로 판단하여 영상인식 관련 산업체에서의 다양한 활용을 추진하고 있다.

S U M M A R Y

Artificially or naturally contained texts in the natural scene images have significant and detailed information about the scenes. Thus, in this research, we develop a method that can extract and recognize those texts in real-time, and to apply the method to many important applications such as an intelligent robot navigation and an intelligent translation system on PDA. The proposed method extracts the text areas in the natural scene images using low-level image features of color continuity, gray-level variation and color variance and verifies the extracted candidate regions by using 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 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. Also, we have successfully applied the developed method to the visual interpretation applications of an intelligent PDA system.

CONTENTS

Chap. 1. Project Overview	10
Chap. 2. Current Status of the Technology	11
Chap. 3. Contents and Results	13
Sect. 1. Text Region Extraction	13
1. Proposed Method	13
2. Verification	34
3. Text Extraction in Various Environments	39
4. Experimental Results	44
Sect. 2. Text Region Image Processing	50
1. Skew and Perspective Correction	50
2. Text Segmentation	53
Sect. 3. Application System Design	59
1. Introduction	59
2. Database Construction	60
3. Text Region Extraction	62
4. Experimental Results	67
제 4 절. Conclusion	69
Chap. 4. Achievements and Contributions	70
Chap. 5. Application Plans	72
Chap. 6. Obtained Overseas Scientific Information	73
Chap. 7. References	74

목 차

제 1 장. 연구개발과제의 개요	10
제 2 장. 국내외 기술개발 현황	11
제 3 장. 연구개발수행 내용 및 결과	13
제 1 절. 텍스트 영역 추출	13
1. 제안하는 방법	13
2. 검증	34
3. 다양한 환경에서의 추출	39
4. 실험 및 결과	44
제 2 절. 텍스트 영역 처리	50
1. 기울어짐 및 원근 보정	50
2. 텍스트 분리	53
제 3 절. 응용시스템 구현	59
1. 개요	59
2. 데이터베이스 구축	60
3. 영역 추출	62
4. 실험 및 결과	67
제 4 절. 결론	69
제 4 장. 목표달성도 및 관련분야에의 기여도	70
제 5 장. 연구개발결과의 활용계획	72
제 6 장. 연구개발과정에서 수집한 해외과학기술정보	73
제 7 장. 참고문헌	74

제 1 장. 연구개발과제의 개요

정지된 이미지나 동영상에 인위적으로 또는 자연적으로 포함되어 있는 텍스트에는 중요한 정보가 함축되어 있다. 이러한 정보를 실시간에 추출하여 인식할 수 있다면 다양한 분야에 활용할 수 있다[1]. 그러나 어떠한 제약도 없이 자연스럽게 취득한 이미지에 포함되어 있는 장면 텍스트를 정확히 찾아내는 것은 기술적으로 어려운 일이다. 따라서 지금까지의 대부분의 관련 연구는 제한된 환경에서 취득한 이미지 또는 동영상에 인위적으로 삽입한 자막 추출에 초점이 맞추어져 있어서 폭넓은 분야에 적용하기에는 많은 문제점을 지니고 있었다.

본 연구의 최종 목표는 자연이미지에 포함되어 있는 장면 텍스트 추출 및 인식으로서 폰트의 종류, 위치, 방향, 크기, 색, 배경 등에 비교적 무관한 추출 방법을 제안하고, 사용자의 불편을 최소화시킬 수 있는 수준의 속도 및 정확성으로 응용시스템에 활용하는 것이다. 따라서 1차년 연구에서는 영상처리 방법을 이용한 텍스트 영역 추출 방법을 제안하고, 2차년 연구에서는 인간의 시각 기능을 모방한 정보 결합 및 검증 방법을 개발하여 추출의 정확성을 향상시킨다. 3차년 연구에서는 극한 환경에서 취득한 이미지에 포함되어 있는 텍스트 영역의 추출로서 텍스트의 크기 변화, 조명 변화 및 시각 변화에 무관한 추출 방법을 개발하여 응용시스템에 적용한다.

본 연구의 1차년 연구에서는 텍스트 영역 추출 방법의 개발로서 다양한 종류의 실험 및 평가 이미지 DB를 수집하고 분류하였으며, 색 연속성 특징과 밝기 변화 특징을 이용한 영역 추출 방법을 개발하였다. 2차년 연구에서는 정보 선별 방법의 개발로 저 수준의 이미지 특징을 검토하고 개선하였다. 저 수준의 특징으로 사용된 색 연속성 및 명도 변화 특징의 적용에서 색 병합 방법과 선 처리 방법을 개선하였고, 색 변화 특징을 추가로 개발하였다. 추출 성능을 개선하기 위해서 영역의 기울어짐 및 원근을 보정하는 방법을 제안하였으며, 단계적 영역 이미지 개선 방법을 개발하였다. 또한, 영역 추출 시간을 단축하기 위해서 에지 추출 방법을 개선하였다. 3차년 연구에서는 극한 환경에서의 텍스트 정보 추출 및 응용시스템 구축으로서 크기, 시각, 조명 변화에 무관한 추출 방법을 개발, 적용하였으며, 영역 결합 및 검증 방법을 개선하였다. 또한, 현재까지 개발한 방법을 적용한 응용시스템으로서 지능정보 단말용 카메라로 취득한 색 문서이미지에서의 텍스트 영역 추출 및 인식 시스템을 개발하였다.

본 연구에서 제안한 텍스트 추출 방법을 다양한 종류의 자연이미지에 대하여 실험한 결과 비교적 배경의 복잡도에 무관한 높은 영역 추출률을 얻을 수 있었다. 또한, 제안한 방법을 지능형 단말기의 영상인식시스템에 탑재하여 지능정보 단말용 카메라로 취득한 발표자료, 도서, 회의 자료, 자연이미지 등의 이미지에 대하여 실험한 결과 우수한 추출 성능을 보임을 알 수 있었다. 현재 개발된 소프트웨어 모듈이 산업체에서 활용할 수 있는 수준으로 판단하여 영상 인식 관련 산업체 및 연구센터 등과 활용 방안을 추진하는 중이다.

제 2 장. 국내외 기술개발 현황

국내외의 텍스트 영역 추출 연구는 흑백 문서이미지에서 영역 추출 작업이 주가 되어왔으나, 최근에는 책 표지, 비디오 프레임 등과 같은 색 이미지에서도 텍스트 영역을 추출하는 방향으로 진행되고 있다[2-7].

국내의 연구로서 자연적인 이미지에 모폴로지를 적용하여 텍스트 영역을 추출한 연구는[8] 텍스트 영역의 대략적인 위치와 최소 텍스트 후보영역을 찾아낸 후 텍스트와 비텍스트를 분류하고, 후처리 과정을 수행하여 텍스트의 위치를 찾았다. 디지털 카메라로 취득한 자연적인 이미지를 대상으로 실험하여 텍스트 정보는 비교적 잘 찾았으나, 고립된 텍스트는 비텍스트로 파악하는 오류가 발생하였다. 최근에는 복잡한 형태의 색 이미지를 처리하는 방법에 대한 연구가 활발히 진행되고 있다. 이 방법들은 대부분 RGB, HSI 등의 다양한 색 공간을 이용하여 색 병합하는 방법들을 제안하고 있다. 이는 대부분이 색 정보를 줄이는 색 양자화를 먼저 수행한 후 각각의 색 공간에서 영역을 추출한다[9].

Pyeong Kee Kim은[10] 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 2비트만을 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. 이 방법은 지역적 양자화의 처리 시간이 오래 걸리고 기울어진 텍스트 영역을 찾기 어려운 단점이 있다.

H. K. Kim은[11] 비디오 프레임으로부터 자동으로 텍스트 영역을 추출하기 위해서 텍스트들이 수평 방향으로 놓여 있고, 균일한 색과 일정한 크기를 갖는다는 가정으로 색 연속성 특징을 이용한 방법을 제안하였다. 제안한 알고리즘은 색 히스토그램 양자화를 수행하여 색 이미지를 분할한 후 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 텍스트 영역과 크기가 작은 텍스트 영역에 대해서는 정확성이 떨어지며, 16개의 경험적 임계값을 정해야 하는 단점이 있다.

시각 장애인을 위한 착용형 컴퓨터 시각 기술의 현황 및 전망에서는 시각 장애인용 보행 안내시스템인 OpenEyes를 소개하고 있다[12]. 이 시스템은 현재 건물 안의 제한된 환경에서 동작하며 사람의 얼굴, 표정 등도 인식하도록 설계되었다. 본 연구개발에서 추진하는 텍스트 추출 및 인식 기능을 OpenEyes에 보장한다면 바람직한 보행 안내시스템을 기대할 수 있다.

국의 연구에서도 제한된 환경에서의 연구가 주를 이루고 있으며, 개발된 응용 시스템으로는 이미지 검색시스템인 QBIC[13], 상표 검색시스템[14], 미술/역사 자료 검색시스템[15] 등이 있다. Jain 등은[16] 이진 이미지, 웹 이미지, 색 이미지 및 비디오 프레임의 네 종류의 이미지에서 텍스트를 추출하는 방법을 제안하였다. 이진 및 웹 이미지에 대해서는 텍스트의 밝기 값

이 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 다른 종류의 이미지에 대해서 서로 다른 특징과 임계값을 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[17] 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 추출하여 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.

Ohya 등은[18] 실외 자연이미지의 텍스트 영역을 추출하기 위해서 대부분의 텍스트는 명도 값이 일정하다는 특징을 이용하였다. 이 논문에서의 텍스트 영역은 기울어짐이 없고 배경과 뚜렷하게 구분된다는 가정을 전제하였다. 지역적 이진화 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[19] 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 서로 다른 시간대별로 획득한 이미지에 적용한 결과 정오와 오후에 대한 이미지에서는 정확한 추출을 보였지만 조명 변화가 있는 오전과 저녁 이미지에서는 오류가 많이 발생하였다. 또한, 처리 시간이 오래 걸리는 단점이 있다.

현재까지 진행된 국내의 관련 연구들은 책, CD 표지, 웹 이미지, 비디오 자막 이미지 등과 같이 제한된 환경에서 얻어진 이미지에서의 텍스트 정보 추출 연구가 대부분이며 자연적인 이미지에서의 텍스트 처리 또한 제한된 환경에서만 수행되었기 때문에 로봇 자동 주행, 지능정보 단말기의 영상인식 등과 같은 응용에 사용되기에는 많은 어려움이 있었다. 그 동안 텍스트 영역 추출은 문서이미지의 자동인식, 번호판 자동인식, 비디오 자막 색인 및 검색 등과 같이 어느 정도 정형화 되어있는 이미지에서의 추출로 진행되어왔다. 그러나 앞으로는 기술적으로 난이도 높고 광범위하게 활용할 수 있는 자연이미지에서의 텍스트 영역 추출로 연구가 진행될 것이며, 이러한 연구개발 결과는 다양한 분야에서 중요하게 활용될 수 있을 것이다.

제 3 장. 연구개발수행 내용 및 결과

본 장의 연구개발 내용은 텍스트 영역 추출, 텍스트 영역 처리, 응용시스템 구현 및 결론으로 기술한다. 제1절 텍스트 영역 추출에서는 영역 추출 및 검증 방법, 다양한 환경에서의 추출 방법과 실험 결과를 기술한다. 제2절 텍스트 영역 처리는 추출된 영역의 기울어짐과 원근 보정 방법을 설명하고, 추출된 영역 이미지를 인식할 만한 수준으로 개선하는 방법을 제안한다. 제3절의 응용시스템 구현에서는 지능정보 단말용 카메라로 취득한 색 문서이미지에서 텍스트 영역 추출 방법을 기술하며, 제4절에서 결론을 맺는다.

제 1 절. 텍스트 영역 추출

1. 제안하는 방법

본 연구에서 제안하는 방법은 그림 2와 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 획의 구성 여부를 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 연구에서는 문자인식의 전 단계인 획 구성 여부까지를 검증하여 텍스트 영역을 결정하고, 결정된 영역의 이진화 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 위한 것이며, 밝기 변화는 명도이미지에서 텍스트 영역은 밝기 변화가 심하여 조밀한 에지 분포를 보인다는 점을 이용한다. 색 변화는 텍스트는 배경과의 색 변화가 존재한다는 점을 이용하는 것이다. 위의 각 특징들을 이용하여 추출한 텍스트 후보 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, 안정적인 색 병합 방법을 제안한다. 각각의 색 면에 대한 연결요소를 분석하여 텍스트 후보영역을 찾고, 각 면에서 찾은 결과를 결합하여 텍스트 후보 영역으로 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 사용한다. 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 결합하여 후보영역을 선별한다. 선별된 후보 영역에서 각 부영역의 획 특징을 추출하고 SVM 분류기를 이용하여 검증한다.

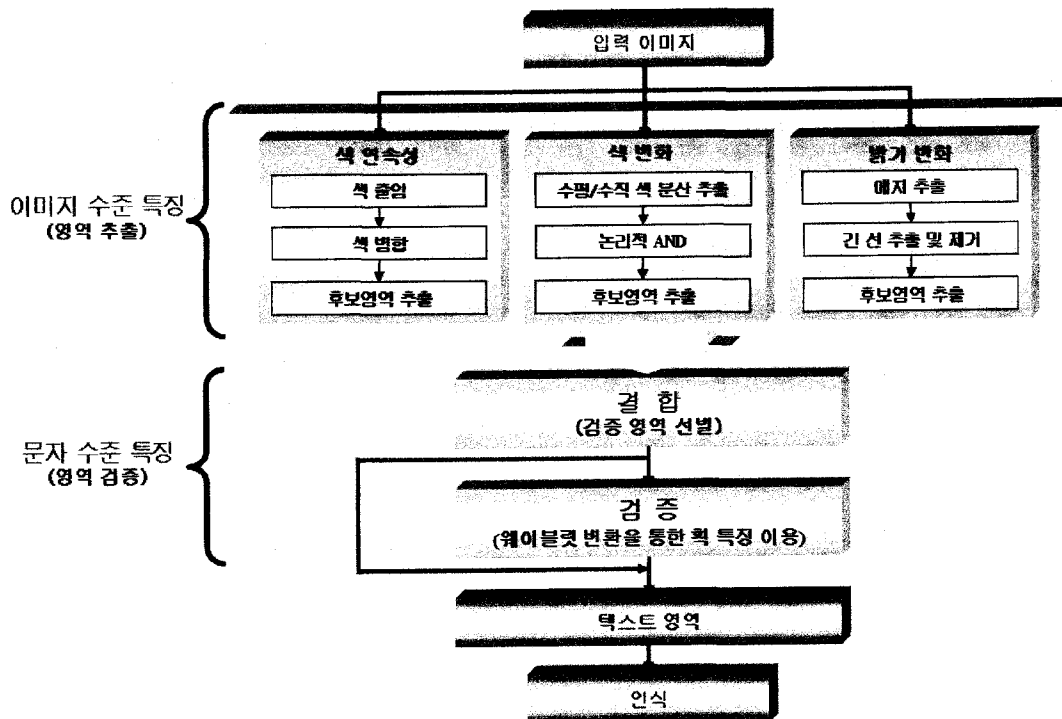


그림 2. 제안한 방법

가. 색 연속성을 이용한 추출

색 연속성을 이용한 추출은 그림 3과 같이 우선 색 면의 개수를 줄이기 위한 색 줄임을 수행하고, 조명 변화와 복잡한 배경에서의 영역 추출을 위한 색 병합 방법을 제안한다. 각각의 색 면에 대한 연결요소를 분석하여 텍스트 후보영역을 찾고, 각 면에서 찾은 결과를 결합함으로써 최종적인 후보영역을 확정한다.

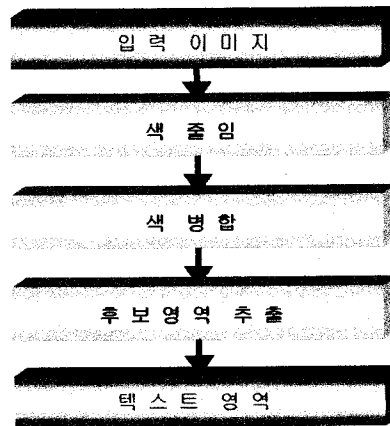


그림 3. 색 연속성을 이용한 텍스트 영역 추출

(1) 색 줄임

다양한 크기의 24비트 색 이미지를 320x240 크기로 먼저 축소시킨다. 24비트로 표현된 이미지의 각 화소(Pixel)를 사용하는 것은 많은 계산량을 필요로 하기 때문에 색 줄임을 수행한다. 색 줄임은 R, G, B 각 요소의 하위 6비트를 제거하여 상위 2비트만을 남긴다. 색 줄임을 수행한 결과 최대 64개의 색 면이 만들어 질 수 있다. 그림 4는 입력 이미지에 대한 색 줄임 결과를 보여준다.

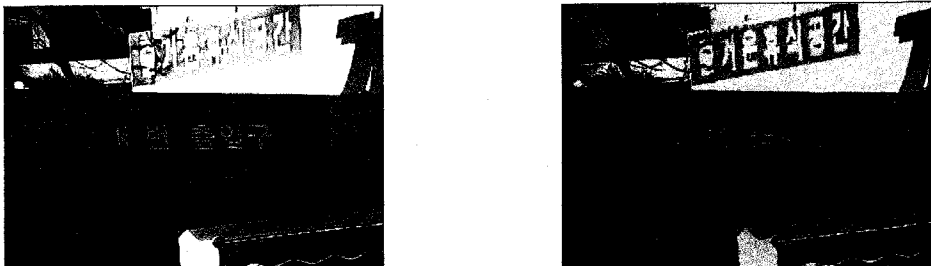


그림 4. 입력 이미지 및 색 줄임 결과

(2) 색 병합

RGB 로 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 찾는다면 처리시간이 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. 3차원 RGB 공간에서 유클리디언(Euclidean) 거리가 1 이하인 경우를 유사한 색으로 판단하여 병합한다. 색 병합은 그림 5와 같이 1차 병합과 반복적인 2차 병합으로 나누어 수행된다.

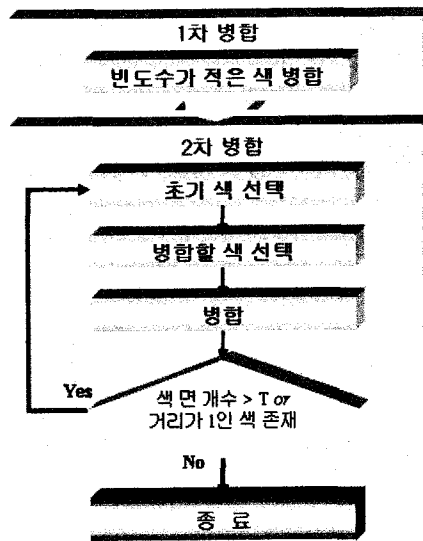
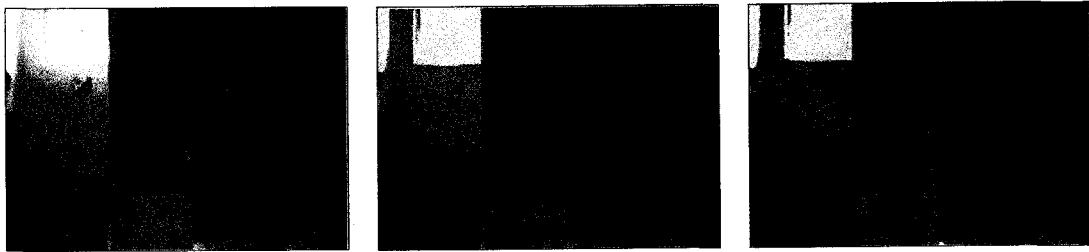


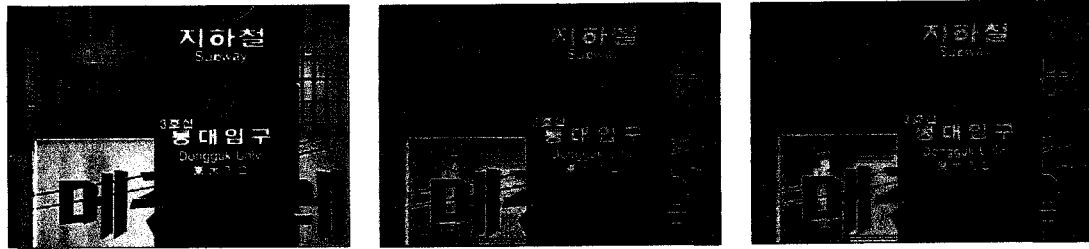
그림 5. 색 병합

색 줄임 결과 이미지는 색마다 다양한 빈도수를 갖는다. 1차 병합은 눈으로 식별하기 어려울 정도의 적은 빈도수를 갖는 색들을 미리 제거함으로써 클러스터(Cluster)들을 고립시켜 색 병합 결과를 향상시키기 위한 것이다. 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, B와 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, B로 구성된 영역과 C, M, Y로 구성된 영역간의 연결성이 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 6(a)은 색 줄임과 색 병합 과정을 거쳐 색의 수가 줄어들면서 R 계열의 색과 Y 계열의 색이 분리되어 배경과 뚜렷이 구분되는 결과를 보여준다.



(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 6. 색 줄임 및 색 병합 결과 예

(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 문제를 해결하기 위해서 텍스트 영역의 연결요소는 밀도가 높다는 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 7은 각 색 면의 연결요소와 연결요소의 외곽사각형을 보여준다.

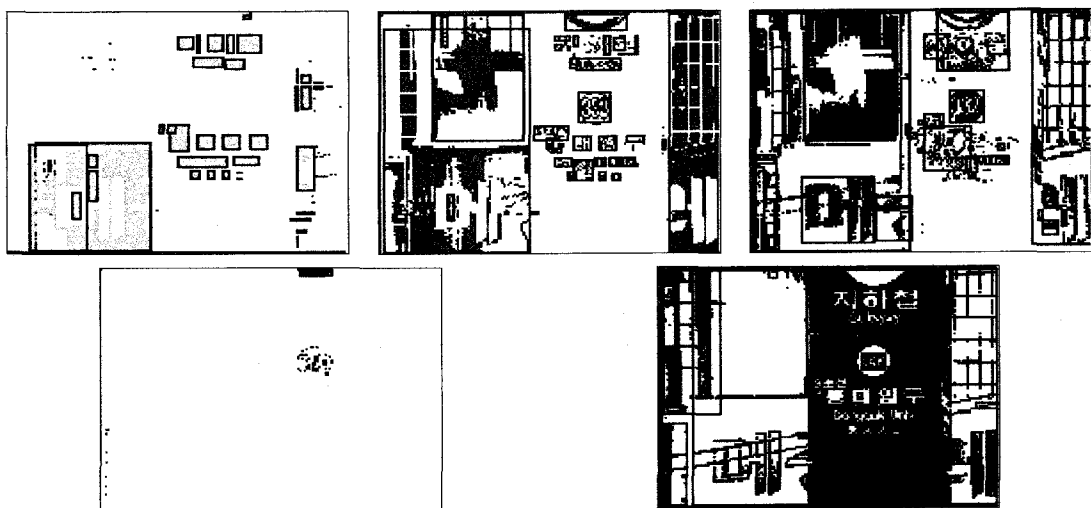


그림 7. 각 색 면의 외곽사각형 추출(5개의 색 면)

각 색 면에서 후보영역의 수를 줄이기 위해서 외곽사각형의 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 8은 그림 7의 각 색 면에 위의 규칙을 적용한 결과이다.

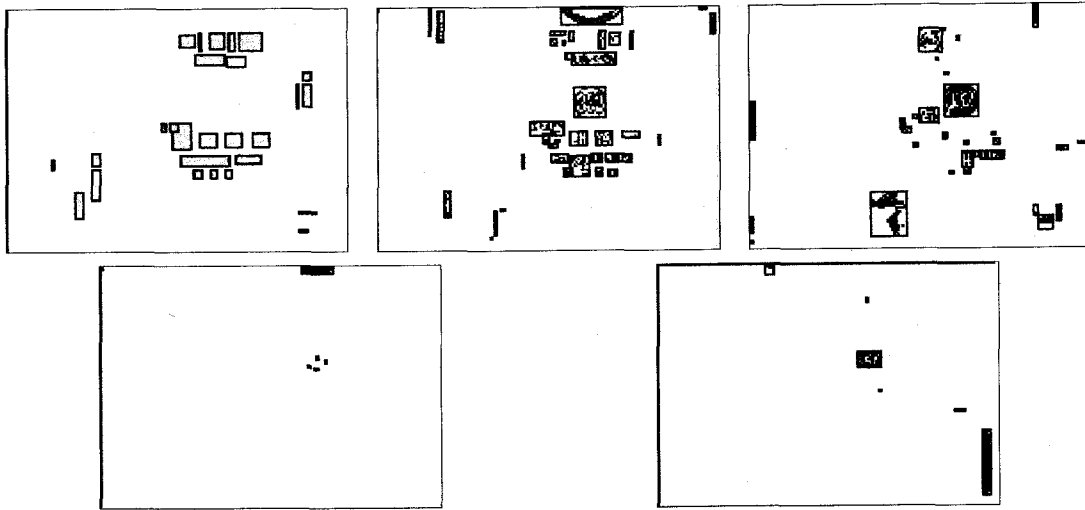


그림 8. 연결요소의 1차 제거

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 크거나 외곽사각형 사이의 거리가 10화소 이하이면 결합한다. 이 과정으로 의미 있는 문자 단위의 외곽사각형을 구성할 수 있다. 그림 9는 결합 결과를 보여준다.

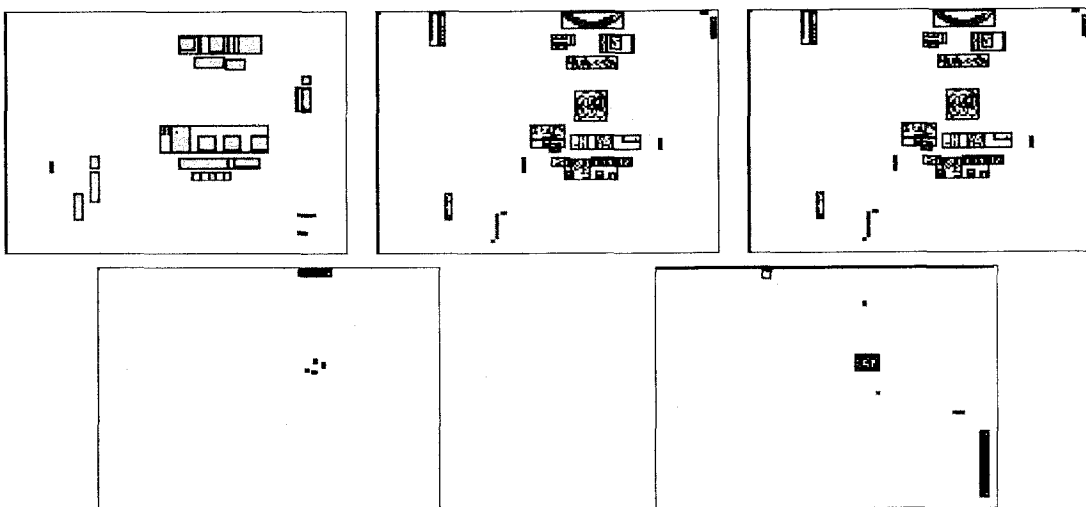


그림 9. 인접한 외곽사각형의 결합

처음 단계에서의 연결요소 제거는 텍스트의 획이 제거되는 것을 방지하기 위해서 명확하게

제거할 만한 것들만 제거하기 위해서 엄격한 기준으로 제거하였지만, 외곽사각형들이 1차로 결합된 후에는 획들이 인근 연결요소에 결합되었기 때문에 보다 완화된 기준으로 외곽사각형을 추가로 제거할 수 있다. 제거 규칙은 1차 제거와 같지만 임계값은 다르게 설정한다. 획과 같은 텍스트의 일부를 구성하는 연결요소들은 전 단계에서 문자 단위의 외곽사각형에 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, 외곽사각형을 구성하는 연결요소의 비율이 임계값 이하이면 제거하여 텍스트의 아주 작은 부분을 포함하는 연결요소는 제외시킨다. 그림 10은 2차 제거 조건으로 제거한 결과이다.

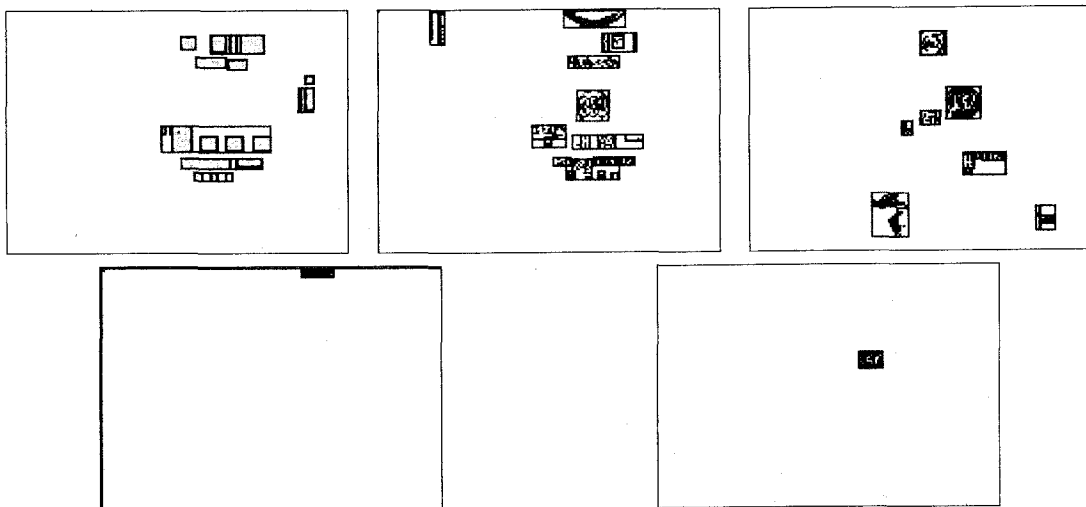


그림 10. 텍스트가 아닌 것으로 판단되는 외곽사각형의 2차 제거

끝으로 각 색 면에서 찾아진 후보영역들을 결합하여 하나의 후보영역 외곽사각형을 구성한다. 각 색 면에서 찾아진 후보영역들의 외곽사각형의 겹침 정도를 파악하여 80% 이상으로 서로 겹쳐졌으면 결합하여 하나의 외곽사각형으로 만든다. 그림 11은 각 색 면에서 추출된 영역들의 후보영역의 결합 전과 결합 후의 결과를 보여준다.

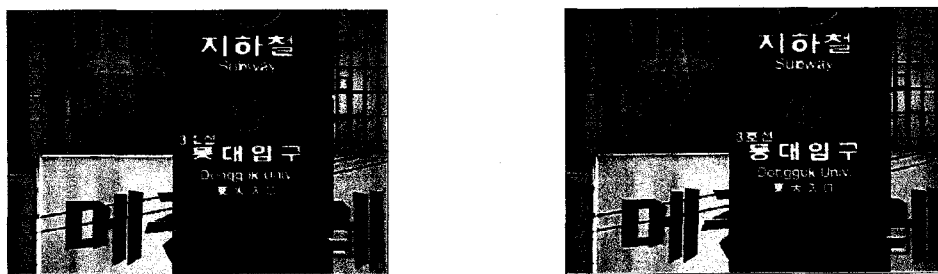


그림 11. 각 색 면에서 추출된 후보영역의 결합 전과 결합 후

나. 밝기 변화를 이용한 추출

밝기 변화를 이용한 추출은 그림 12와 같이 먼저 에지 이미지를 구한다. 에지 분포를 분석하여 이미지에 포함되어 있는 다양한 종류의 긴 선들을 제거하여 처리할 이미지를 단순하게 만든다. 긴 선 정보들은 텍스트 후보영역의 기울어짐과 원근을 보정하는데 사용한다.

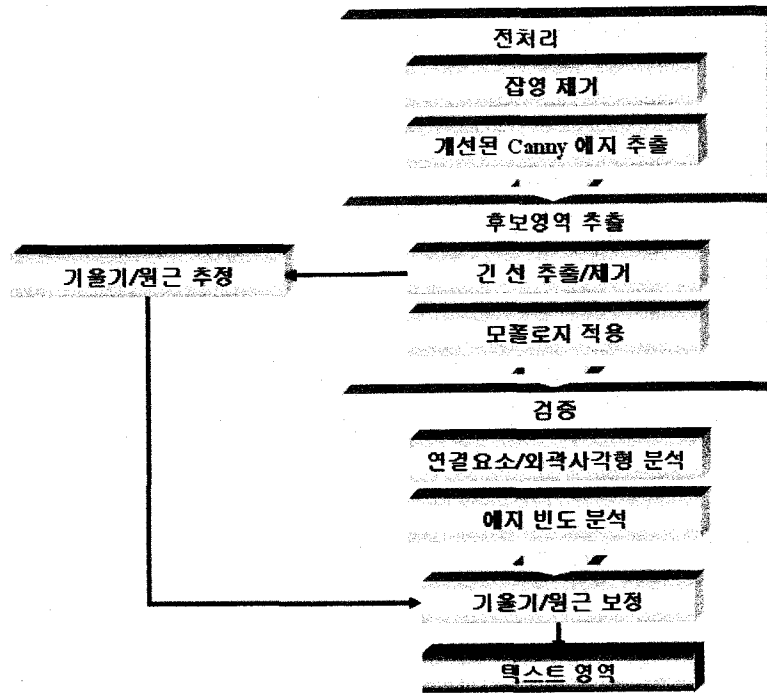


그림 12. 밝기 변화를 이용한 텍스트 영역 추출

(1) 전처리

색 연속성을 이용한 추출에서 사용한 320x240 크기의 축소된 이미지를 먼저 명도이미지 (Gray-level)로 바꾸고, 3x3 크기의 미디언(Median) 필터를 적용하여 잡영의 일부를 제거한 후, Canny 에지[20] 검출을 수행한다. 그림 13은 미디언 필터 및 Canny 에지 검출 결과이다.

Canny 에지 검출은 에지를 매우 정교하게 찾아주는 장점이 있지만, 연산 시간이 오래 걸리는 단점이 있다. 본 연구에서는 처리시간이 오래 걸리는 단점을 개선하기 위해서 Canny 에지 추출 과정에서 시간이 가장 많이 소요되는 억제(Suppression) 단계에서 주변의 모든 화소로부터 융기(Ridge)점을 찾지 않고, 경사(Gradient) 값이 큰 화소만을 대상으로 융기점을 찾음으로써 비교 횟수를 줄여 처리 시간을 감소시키는 방법을 제안한다.

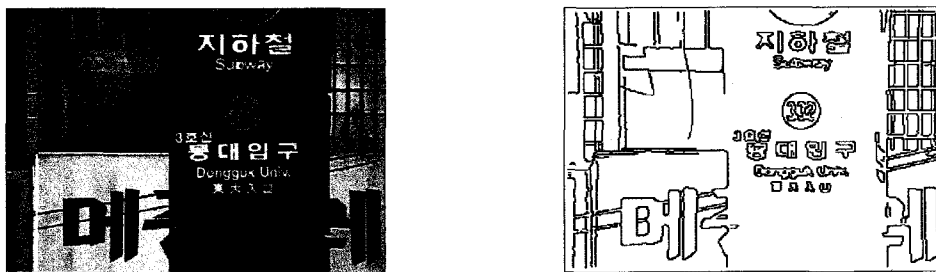


그림 13. 원 이미지의 전처리로 개선된 Canny 에지 이미지

Canny 에지 추출은 다음의 네 단계로 수행된다. 1단계에서는 가우시안 필터링으로 입력 이미지의 잡영을 제거하고, 2단계에서는 잡영이 제거된 이미지에 1차 미분 연산을 수행하여 행

과 열 방향의 경사 값을 각각 구한다. 3단계에서는 에지를 세선화하기 위해 단계 2에서 추출된 용기들을 추적하면서 실제로 용기의 최대값이 아닌 화소는 억제시킨다. 마지막 단계에서는 두 임계값 $T1$, $T2$ ($T1 > T2$)을 이용해서 연결성을 고려하여 이미지를 이진화시킨다. 이 가운데 단계 3이 Canny 연산 처리 시간의 대부분을 차지한다. 단계 3은 '하나의 에지에 대응한 결과(Response to single edge)' 조건을 만족시키기 위해서 8방향 이웃 화소들 보다 경사 값이 클 때에만 현재 화소를 에지가 되도록 하는 것이다. 경사의 방향이 단순히 수평 또는 수직 방향이라면 화소의 상, 하, 좌, 우 화소의 경사 값과 비교하면 되지만, 대부분 경사 방향은 유동적이어서 그림 14와 같이 두 화소 사이를 가리키는 경우가 쉽게 발생한다. 여기서의 경사 값은 식 (1)과 같이 이웃하는 화소들로부터 선형보간법을 사용하여 구한다.

$$Norm(P) = (By - Py)Norm(C) + (Py - Cy)Norm(B) \quad (1)$$

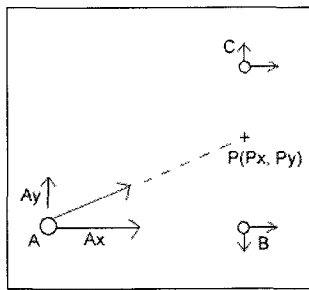


그림 14. 화소의 경사의 방향

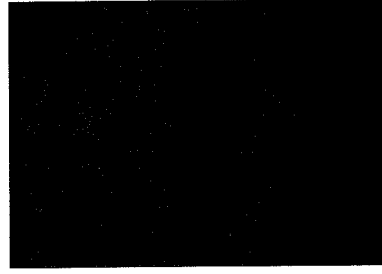
한 화소가 억제되어야 하는지를 결정하기 위해서 적어도 네 번의 이웃하는 화소와 경사 값을 비교해야 한다. 입력 이미지의 크기가 320x240 이라면 최소한 4x76,800번 경사 값을 계산해야 하므로 에지 추출의 처리 속도는 단계 3에서 크게 늦어진다. 실제로 용기의 봉우리가 전체 이미지에서 차지하는 비율이 낮기 때문에 이 단계에서 남겨지는 화소들의 수는 적다. 따라서 열과 행 방향 모두 경사 값이 매우 작아서 용기의 봉우리가 될 가능성이 없는 화소들과는 비교하지 않는다면 처리시간을 대폭 줄일 수 있다.

표 1. 복잡한 이미지와 단순한 이미지에서 경사 값에 따른 처리 시간 비교(단위: msec)

	0	0.01	5	15
복잡한 이미지	895	437	282	109
단순한 이미지	860	343	94	63



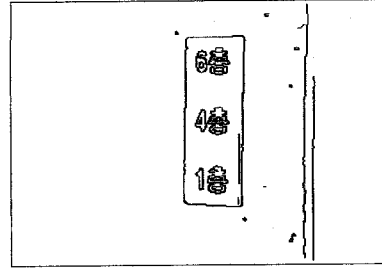
(a) 복잡한 이미지



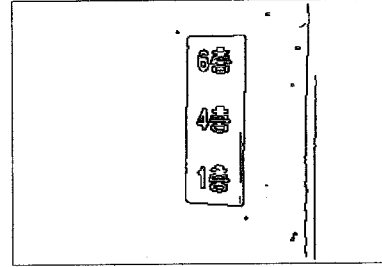
(b) 단순한 이미지



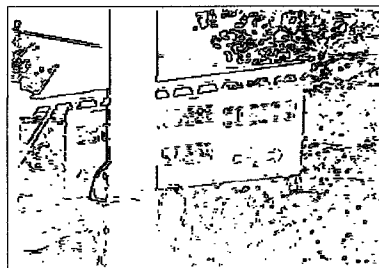
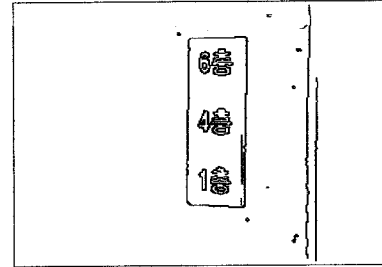
(c) 모든 화소를 고려한 결과



(d) 경사 값이 0.01이하이면 억제 대상에서 제외



(e) 경사 값의 크기가 5이하 제외



(f) 경사 값의 크기가 15이하 제외

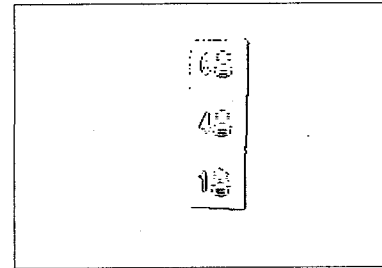


그림 15. 경사 값에 따른 Canny 에지 추출 결과

표 1은 경사 값을 비교하지 않은 화소의 수가 증가함에 따라 처리 시간은 감소하지만 일정 수준이 넘어서면 그림 15와 같이 에지 이미지의 질이 손상되는 것을 볼 수 있다. 실험을 통하여 성능 대비 이미지의 질이 최대가 되는 수준은 경사 값이 5이하인 화소를 고려하지 않았을

때로 나타났다. 보완된 Canny 에지 추출기는 1차년 연구와 비교할 때 대부분의 이미지에서 약 5배에서 10배 이상의 빨라진 처리 속도를 보이며, 특히 단순한 이미지일수록 개선 정도가 크다.

(2) 후보영역 추출

이미지에 포함되어 있는 텍스트 주변에는 일반적으로 텍스트를 둘러싸고 있는 긴 선들이 존재하며 텍스트 영역은 에지 밀도가 높은 특징이 있다. 이러한 특징을 이용하여 텍스트 주변의 긴 선들을 추출하여 기울어짐 및 원근을 추정한 후 긴 선들을 제거하여 모폴로지 적용 시 텍스트 후보영역과 연결되는 경우를 방지한다.

1차년 연구와 비교하여 긴 선 추출 및 제거 과정에서 8방향 히스토그램과 4방향 히스토그램을 분석하는 과정에서 중복되고 불필요한 조건들을 제거하고, 추가적으로 에지의 방향변화 정도와 X, Y의 최대/최소 점을 이용하였으며, 텍스트 에지와 연결된 긴 선들을 제거하기 위해 X/Y축 투영(Projection) 방법을 이용하여 긴 선 제거 방법을 보완하였다. 아래 그림 16은 1차년 선 제거 결과와 2차년 결과를 비교한 것이다.

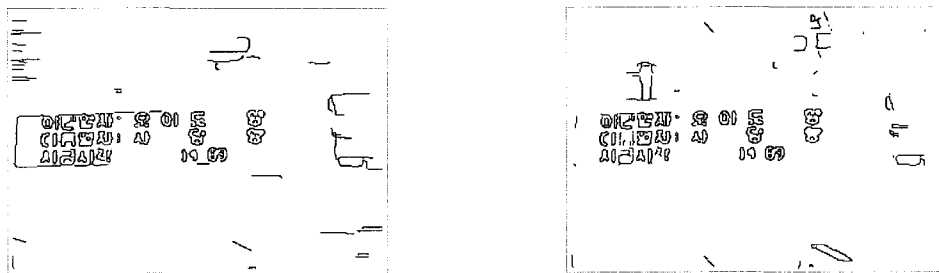


그림 16. 긴 선 제거 결과 비교: 1차년(좌) 및 2차년 결과(우)

후보영역을 강조하는 방법에서는 1차년에서는 반복적인 RLS(Run-Length Smearing)를 적용하여 단어 단위로 영역을 강조하였으나, 긴 선 제거 과정에서 제거되지 않고 남은 많은 에지들로 인하여 텍스트 영역과 남아 있는 에지들과 서로 연결되는 경우가 빈번하게 발생하였다. 이러한 오류를 개선하기 위해서 2차년 연구에서는 모폴로지를 적용하여 문자 단위의 텍스트를 추출한 후 다시 의미 있는 단어 단위로 결합한다. 그림 17은 RLS를 사용한 결과와 모폴로지를 사용한 결과를 비교한 예이다. 두 이미지를 비교할 때 (a)이미지는 RLS를 적용했을 때 텍스트 영역과 제거되지 않고 남아 주변의 원과 화살표 모양이 결합되는 결과를 보여주며, (b)이미지는 이러한 오류들이 긴 선 제거 과정의 개선과 모폴로지의 이용으로 개선된 것을 보여준다.

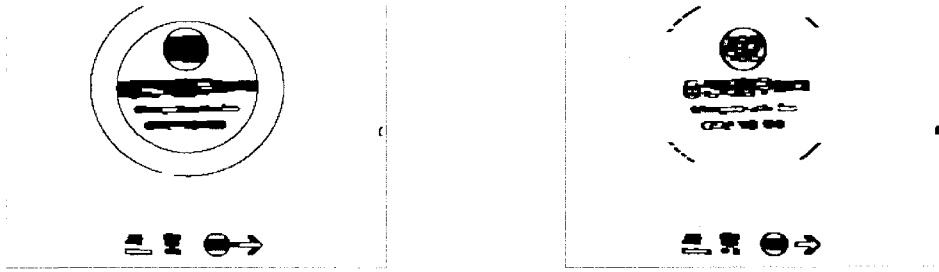


그림 17. 후보영역 강조: 반복적인 RLS를 적용한 결과(좌)와 모폴로지를 적용한 결과(우)

자연이미지의 텍스트 주변에는 대부분 사각형 또는 다른 긴 선들이 포함되어 있다. 이러한 긴 선들은 텍스트 영역을 추출하는데 방해가 되기 때문에 제거하는 것이 필요하다. 긴 선들을 추출하기 위해서 그림 18(a)과 같은 8방향의 인접 화소를 고려한 에지 추적(Following)을 수행하여 각 방향에 대한 히스토그램을 추출한다. 그러나 그림 18(b)과 같이 에지 추적의 시작점의 위치에 따라 한 방향의 긴 선 임에도 불구하고 그림 18(c)과 같이 서로 다른 방향으로 히스토그램이 구성되는 경우가 발생한다. 이를 해결하기 위해 8방향의 각 대칭되는 방향을 서로 합한 그림 18(a)의 4방향에 대한 그림 18(d)과 같은 히스토그램을 보완적으로 고려한다.

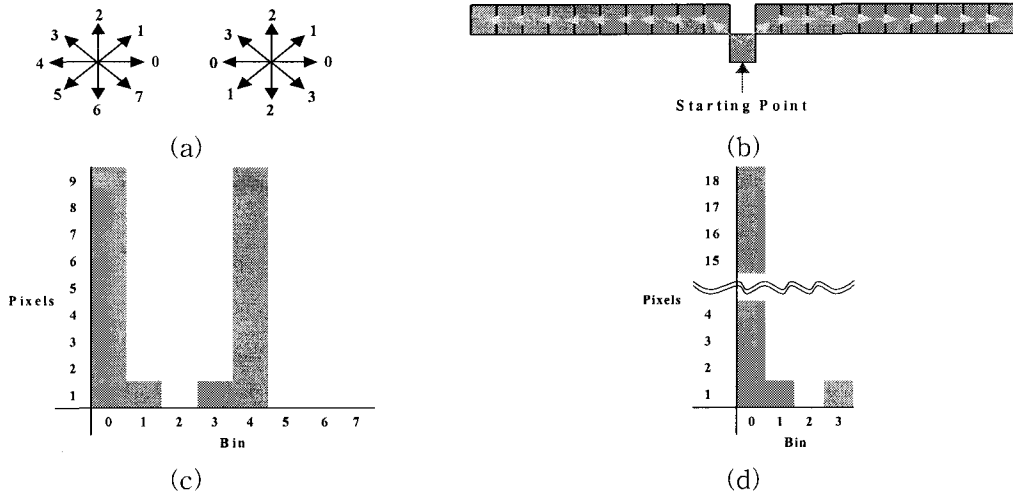


그림 18. 긴 선 제거 방법: (a) 8방향과 4방향 성분, (b) 시작점에 따른 방향성 분석, (c) 8방향 히스토그램, (d) 4방향 히스토그램

그림 19와 같이 긴 선이 텍스트와 연결되어 있는 경우가 발생하는데 이러한 문제를 해결하기 위하여 X/Y 투영을 이용하여 긴 선 부분만 제거한다. 또한, 제거 시 크기에 제한을 두어 텍스트의 획과 같은 부분이 제거되지 않도록 한다.

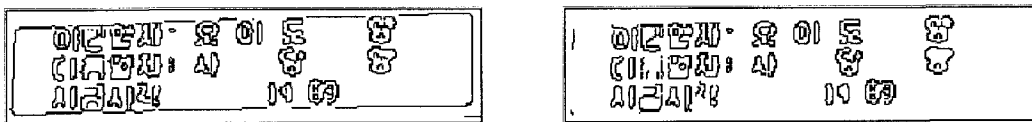


그림 19. 텍스트와 긴 선이 연결되어 있는 경우와(좌) 긴 선 제거 규칙과 X/Y 투영을 이용한 긴 선 제거 결과(우)

에지 추적을 통해 X/Y축의 최대, 최소 좌표와 선의 방향변화 빈도 등에 대한 정보를 추가

적으로 추출한다. 추출된 각 선의 히스토그램을 분석하면 긴 선들은 4방향 히스토그램의 최대 빈 값이 다른 빈들에 비해 크게 나타나고, 최대 X/Y와 최소 X/Y의 차가 크게 나타나며 방향의 변화 정도는 작게 나타난다. 사각형 모양의 경우 4방향 히스토그램의 상위 두 방향 빈의 합이 전체 히스토그램의 합과 유사하게 나타난다. 이러한 특징을 이용하여 긴 선, 사각형 형태 및 길고 꺾이는 부분이 포함된 선들을 제거하였다. 긴 선들이 제거된 결과 이미지는 그림 20과 같다.

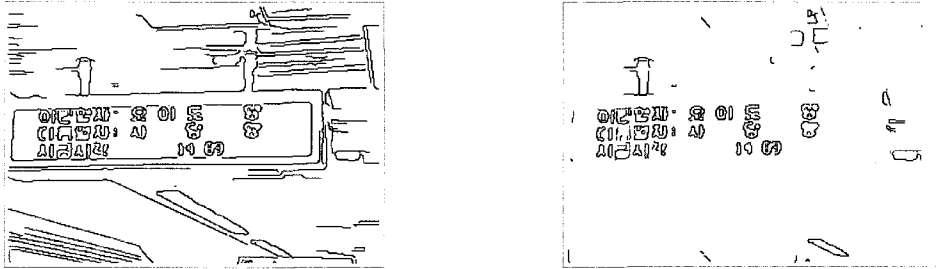


그림 20. 에지 이미지와 긴 선 제거 결과

긴 선들이 제거된 이미지에서 텍스트 영역은 경험적으로 에지 밀도가 높은 위치에 있다. 에지 밀도의 조밀함을 강조하기 위해서 모폴로지 연산을 적용한다. 텍스트의 크기 차이, 간격 등에 따라 에지 간격이 다양하기 때문에 모폴로지의 불림(Dilation)과 지움(Erosion)연산에 대해 그림 21과 같은 구조 연산자를 사용한다. 다음으로 수식 (2)의 닫힘 연산을 수행하여 수평 방향 텍스트는 많이 강조하고, 수직 방향으로는 약간 강조한다.

$$Closing = (I \oplus S) \ominus S \quad (2)$$

- I : 입력 이미지
- S : 구조 연산자
- \oplus : 불림 연산
- \ominus : 지움 연산

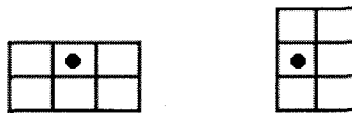


그림 21. 불림(좌) 및 지움 구조 연산자(우)

모폴로지에 의한 강조는 대부분 글자 단위로 강조하기 때문에 수직 문자열에 대해서도 글자 단위의 추출이 가능해진다. 이러한 모폴로지 연산을 수행하여 텍스트와 제거되지 않은 텍스트 주변의 에지들이 서로 연결되는 경우를 많이 줄일 수 있었다. 그림 22는 모폴로지를 적용한 결과이다.

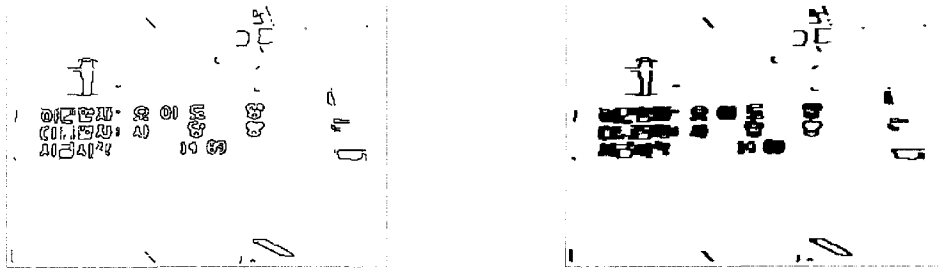


그림 22. 긴 선 제거 결과와 모폴로지를 이용한 텍스트 영역 강조

(3) 후보영역 검증

모폴로지를 적용하여 강조한 텍스트 후보영역 중에는 연결요소의 화소수가 너무 크거나, 너무 작은 것은 텍스트가 아닌 경우가 많기 때문에 제거하는 것이 필요하다. 연결요소의 화소수가 너무 큰 경우는 나무나 타일 등과 같이 에지 밀도가 조밀한 배경에서 발생된 경우가 대부분이고, 연결요소의 화소수가 너무 작은 경우는 비록 텍스트 영역이라 하더라도 인식이 어렵기 때문에 제거한다.

연결요소의 외곽사각형 정보도 이용하여 검증한다. 외곽사각형내의 연결요소가 그 사각형을 차지하는 비율이 정해진 임계값 이하로 작으면 텍스트 후보영역에서 제외한다. 이 때 크기가 큰 텍스트의 경우에는 모폴로지를 이용하여 강조해도 비율이 낮은 경우가 있기 때문에 이를 보완하기 위하여 연결요소의 크기에 반비례하게 임계값을 적용하였다. 즉, 연결요소의 크기가 크면 비율에 대한 임계값을 작게 하고, 연결요소의 크기가 작으면 비율에 대한 임계값을 크게 한다. 또한, 외곽사각형의 가로/세로 비율, 폭 또는 높이가 아주 작은 것들을 추가로 제거한다.

다음으로 텍스트 영역은 에지 밀도가 높다는 특징을 이용하여 텍스트 후보영역에 대하여 수평 및 수직 에지 빈도를 이용하여 검증한다. 에지 빈도는 그림 23과 같이 후보영역에 대한 수평 방향의 에지 빈도의 평균과 수직 방향의 에지 빈도의 평균을 이용하여 구한다.

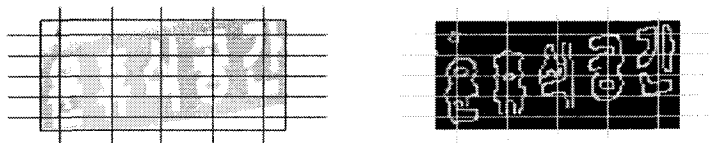


그림 23. 에지 빈도를 이용한 후보영역 검증: 후보영역과 해당 영역에 대한 에지 이미지

텍스트 후보영역 중에 텍스트의 크기가 크다면 에지 빈도가 낮을 수 있는 문제를 해결하기 위해 외곽사각형에 대한 연결요소의 밀도를 이용하는 방법과 같이 영역의 크기가 크면 임계값을 작게 하고, 영역의 크기가 작으면 임계값을 크게 하는 적응적 방법을 적용하였다. 이러한 방법으로 에지 빈도가 높은 영역만을 텍스트 후보영역으로 추출한다. 그림 24는 에지 빈도에 의해 제거된 결과를 보여주며, 그림에서 흰색 외곽사각형이 에지 빈도에 의해 검증되어 제거된 외곽사각형을 나타내고, 빨간색 외곽사각형이 검증되고 남은 결과를 나타낸다. 빨간색 원은 제거된 위치의 에지를 보여준다.

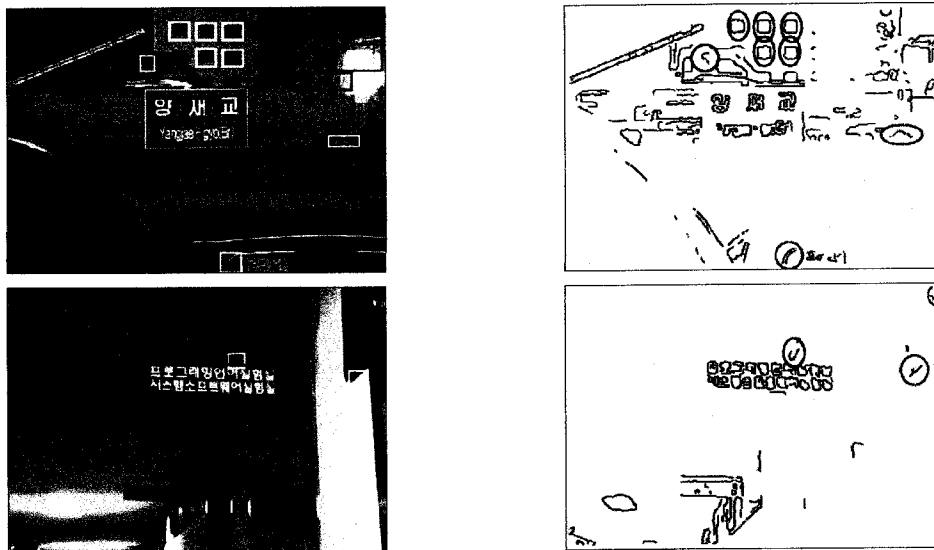


그림 24. 추출된 텍스트 후보영역과 에지 빈도에 의해 검증된 결과

그림 25는 연결요소에 대한 정보와 외곽사각형의 정보, 그리고 에지 빈도에 의해 검증되어 추출된 후보영역을 보여준다.

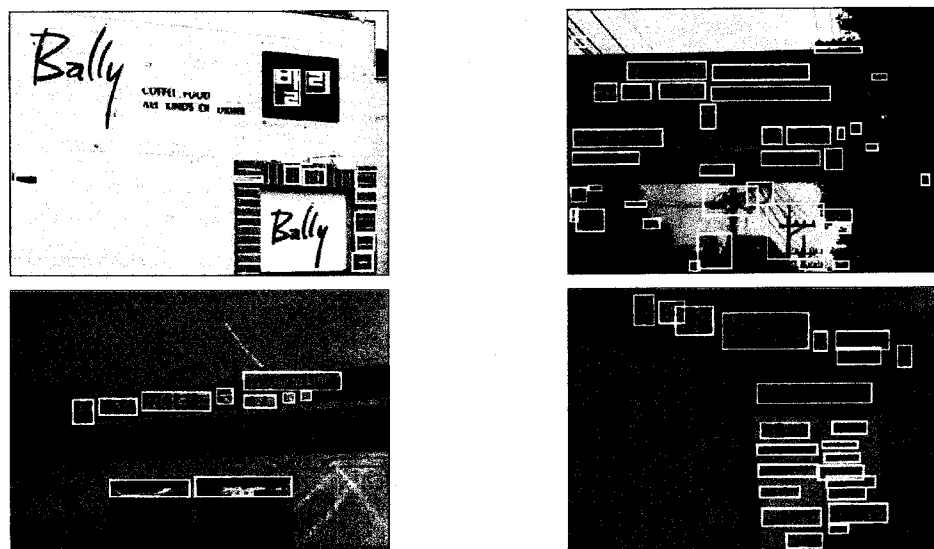


그림 25. 추출된 텍스트 후보 영역: 실내의 단순 이미지와 복잡 이미지

(4) 후보영역 결합

추출된 텍스트 후보 영역들의 일부는 문자열이 하나의 외곽사각형으로 존재하지만 일부는 글자 단위로 존재하기 때문에 최종적으로 인식에 필요한 정도로 결합하는 것이 필요하다. 본 연구에서는 3가지 방법으로 외곽사각형을 결합한다. 즉 외곽사각형들의 포함관계, 외곽사각형의 겹침 정도 및 외곽사각형 사이의 거리를 이용하여 결합한다.

첫째는 외곽사각형의 포함 관계에 의해 큰 외곽사각형 안에 완전히 포함되는 작은 외곽사각형을 큰 외곽사각형으로 대체한다. 이 경우 대부분이 그림 26과 같이 한글 모음 중에 "ㅁ", "

ㅂ", "ㅇ", "ㅎ", 그리고 영어 중에 "D", "O", "B" 등과 같이 큰 텍스트일 경우 예지가 서로 분리되어 큰 부분과 그 곳에 포함된 작은 부분이 발생한다. 이런 경우 반드시 하나의 외곽사각형으로 결합되어야 하기 때문에 큰 외곽사각형으로 결합한다.

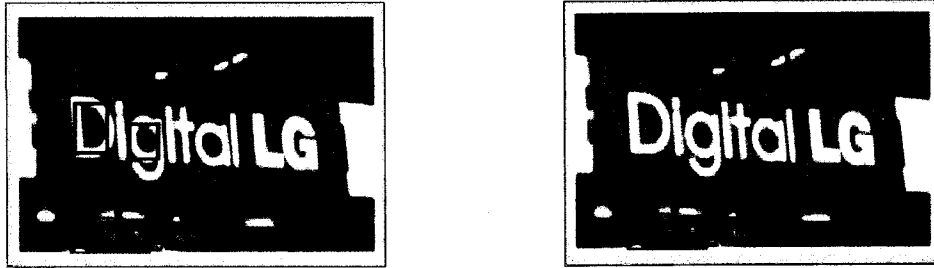


그림 26. 외곽사각형의 포함관계에 의한 결합: 결합 전과 결합 후 결과

둘째는 두 외곽사각형의 겹침 정도를 이용한 결합이다. 외곽사각형의 겹침 정도에 따라서 그림 27과 같이 가로 및 세로 방향에 대해 각각 4가지 경우가 존재한다.

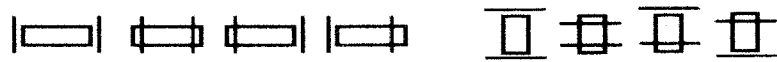


그림 27. 겹침 분포: 수평 방향의 겹침(좌), 수직 방향의 겹침(우)

수평과 수직 네 가지 경우에 대해 겹쳐지는 넓이를 이용한다. 구해진 넓이가 임계값 이상일 경우에만 두 외곽사각형을 결합한다. 결합된 외곽사각형에 대해서는 다시 반복하여 외곽사각형의 겹침 정도가 임계값을 넘는 외곽사각형이 없을 때까지 계속 반복한다. 그림 28은 겹침 정도에 의해 결합된 결과 이미지를 보여주는데, (a)의 노란색 외곽사각형은 겹침이 있는 것을 나타내고, (b)의 파란색 외곽사각형은 (a)의 노란색 외곽사각형이 결합된 결과를 보여준다.

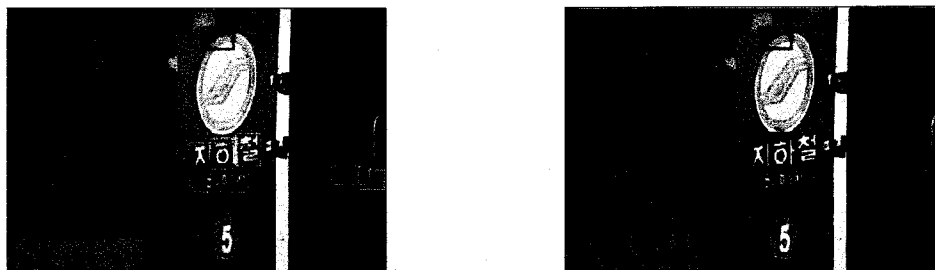


그림 28. 겹침에 의한 외곽사각형 결합: 결합 전(좌), 결합 후(우)

마지막으로 외곽사각형의 최단 거리를 구해 서로 가까운 외곽사각형을 결합한다. 이 경우 최단 거리를 구할 때 그림 29(a)와 같이 외곽사각형의 네 모서리를 벗어나는 외곽사각형은 거리가 가깝다 하더라도 의미 있는 텍스트의 경우 대부분 수평 방향 또는 수직 방향으로 매우 인접해 있으므로 결합 대상에서 제외시킨다. 또한, 거리를 구하는 대상에 존재하는 외곽사각형 중에서도 그림 29(b)와 같이 상하 또는 좌우의 연장선상에 겹쳐지는 넓이를 구해서 겹침 정도가 임계값 이상인 경우에만 결합한다. 대부분의 텍스트 열은 동일한 높이를 가지는 텍스트들이 모여서 이루어지기 때문에 겹침 정도가 작다면 서로 다른 의미의 텍스트 이거나, 텍스

트 이외의 잡음일 가능성이 높다.

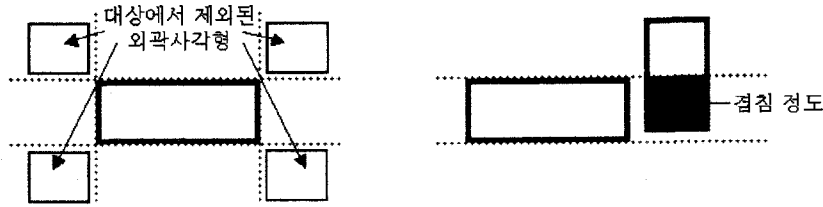


그림 29. 거리 결합 대상에서 제외된 외곽사각형(좌), 거리 결합 대상의 외곽사각형(우)

거리에 의한 결합은 우선 수평 방향으로 실시한다. 앞에서의 모폴로지를 이용한 텍스트 후보영역 강조에서도 수평 방향 위주로 텍스트를 강조하였기 때문에 결합에서도 수평 방향 위주의 텍스트 열을 추출한다. 이를 위해서 본 연구에서는 수직 방향의 거리에는 벌점(Penalty)을 두어 아주 가까운 외곽사각형이 아닌 경우 결합하지 않는다.

그림 30은 외곽사각형의 거리에 의한 결합 결과를 보여준다. 여기서 왼쪽 이미지는 첫 번째 포함 관계에 의한 결합과 두 번째 겹침 정도에 의한 결합까지 완료된 상태이며, 오른쪽은 최종적으로 결합된 결과이다. 그림 30의 경우 "eigen"과 "post"라는 의미 있는 두 개의 텍스트가 존재하는데 앞의 "eigen"의 경우 텍스트 하나하나가 서로 가깝기 때문에 결합되었으나, "post"는 각각이 너무 멀리 위치하기 때문에 결합되지 않았다. 또한, 수직 방향으로도 결합 거리의 차가 크기 때문에 결합되지 않았다.

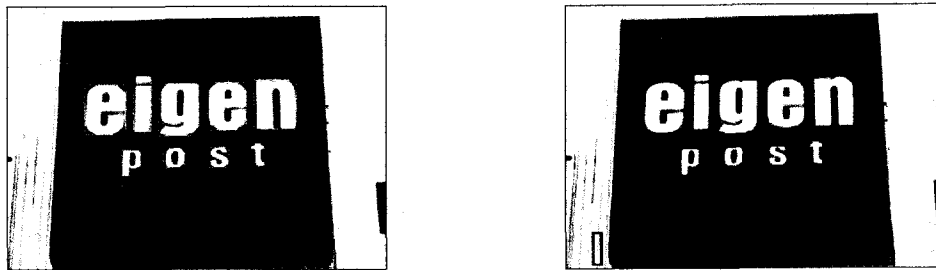


그림 30. 외곽사각형 사이의 거리에 의한 결합: 결합 전(좌), 결합 후(우)

1차년과 2차년 결과를 비교하면 다음과 같다. 그림 31은 1차년 연구에서 긴 선 제거가 잘 되지 않아 반복적 RLS 적용 시 텍스트와 예지들이 연결되는 오류가 발생하여 최종적으로 텍스트를 찾지 못하는 오류를 보여준다. 이러한 오류를 개선하기 위해서 긴 선 제거 방법을 개선하였으며, 1차년 연구에서 텍스트인데도 찾지 못한 영역에 대해서는 모폴로지를 적용하여 문자 단위로 추출한 후 다시 의미 있는 단어 단위로 결합하였다. 따라서 그림 31과 같이 1차년 결과에 비교하여 보다 정교한 텍스트 영역 추출이 가능하다.

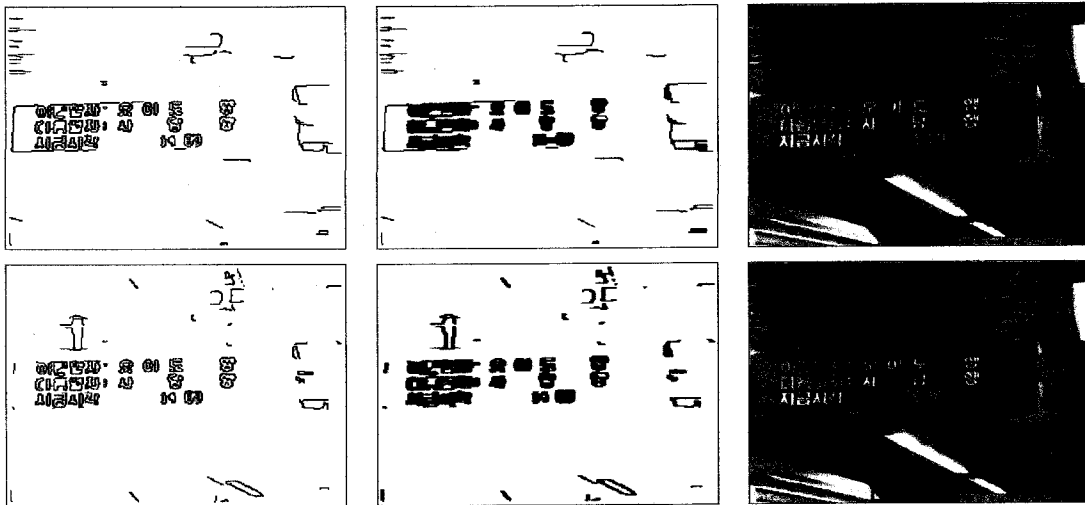
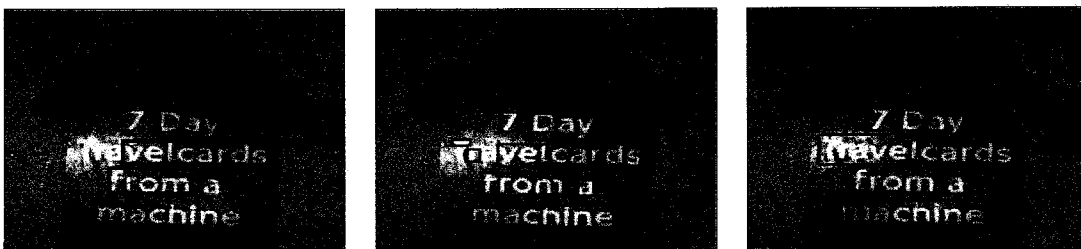


그림 31. 1차년 연구(상): 긴 선 제거, 반복적 RLS 적용, 영역 추출, 2차년 연구(하): 긴 선 제거, 모폴로지 적용, 영역 추출

다. 색 변화를 이용한 추출

실내외에서 취득한 자연이미지는 빛과 조명의 영향으로 또는 인위적인 의도에 의해서 텍스트와 배경이 명확하게 구분되지 않는 경우가 빈번하게 발생한다. 이러한 경우 색 변화를 특징으로 이용하면 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 32(a)는 주어진 이미지에 대하여 색 연속성을 이용한 추출 결과로서 텍스트의 색이 흐리고 조명 변화로 인해서 색 병합에서 오류가 발생하여 영역을 정확하게 찾지 못한 경우이다. 그림 32(b)는 밝기 변화를 이용한 추출 결과로서 역시 조명 변화로 인해서 에지가 잘 찾아지지 않아서 발생한 오류를 보여준다. 그림 32(c)는 색 변화를 이용한 추출 결과로서 앞의 오류가 해결된 것을 볼 수 있다. 이와 같이 색 변화를 이용한다면, 아주 작은 색의 변화도 쉽게 찾을 수 있는 장점이 있다.

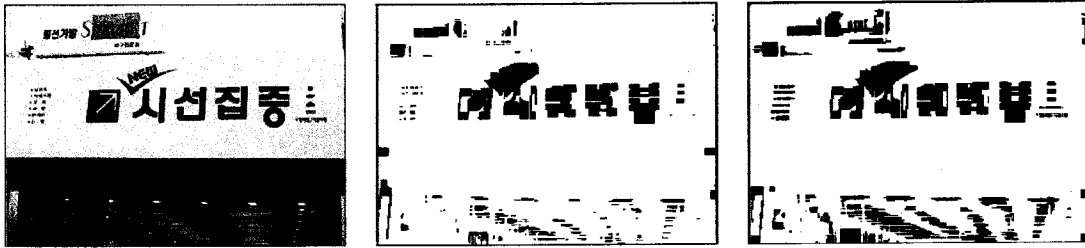


(a) 색 연속성 이용 (b) 밝기 변화 이용 (c) 색 변화 이용

그림 32. 세 종류의 특징을 이용한 영역 추출 비교

색 변화 정도는 분산으로 표현할 수 있는데, 분산은 명도 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 분산

으로 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 명도이미지에서 분산보다 크며 변화 정도를 보다 정확하게 측정할 수 있는 장점이 있다. 그림 33(b)은 원 이미지에서 밝기 값으로 분산을 계산하여 표현한 이미지이며, 그림 33(c)은 3차원 RGB 색 값으로 분산을 계산하여 표현한 이미지이다. 그림에서 글자와 배경의 변화가 보다 명확하게 표현된 것을 확인할 수 있다.



(a) 원 이미지

(b) 명도 분산 결과

(c) 색 분산 결과

그림 33. 명도 및 색 분산 비교

(1) 색 분산 이미지 생성

입력된 색 이미지에 대해 RGB 각각을 이용하여 수평 방향으로 1x21 크기의 윈도우를 사용하며, 수직 방향으로 17x1 크기의 윈도우를 사용하여, 식 (3)과 같이 RGB 각각에 대한 평균과 식 (4)의 분산을 구한다.

$$\begin{aligned} \mu_{HR} &= \frac{1}{21} \sum_{i=0}^{20} R_i, & \mu_{HG} &= \frac{1}{21} \sum_{i=0}^{20} G_i, & \mu_{HB} &= \frac{1}{21} \sum_{i=0}^{20} B_i, \\ \mu_{VR} &= \frac{1}{17} \sum_{i=0}^6 R_i, & \mu_{VG} &= \frac{1}{17} \sum_{i=0}^6 G_i, & \mu_{VB} &= \frac{1}{17} \sum_{i=0}^6 B_i \end{aligned} \quad (3)$$

$$\begin{aligned} v_H &= \frac{1}{21} \sum_{i=0}^{20} ((R_i - \mu_{HR})^2 + (G_i - \mu_{HG})^2 + (B_i - \mu_{HB})^2) \\ v_V &= \frac{1}{17} \sum_{i=0}^6 ((R_i - \mu_{VR})^2 + (G_i - \mu_{VG})^2 + (B_i - \mu_{VB})^2) \end{aligned} \quad (4)$$

μ_{HR} : 수평방향 R의 평균 μ_{VR} : 수평방향 R의 평균 v_H : 수평방향의 분산
 μ_{HG} : 수평방향 G의 평균 μ_{VG} : 수평방향 G의 평균 v_V : 수직방향의 분산
 μ_{HB} : 수평방향 B의 평균 μ_{VB} : 수평방향 B의 평균

구해진 분산값은 대부분 255를 넘는 큰 값이기 때문에 분산값에 대한 정규화가 필요하다. 정규화는 우선 분산값에 식 (5)의 제곱근 연산을 수행하여 값의 범위를 줄이고, 식 (6)을 이용한 최대최소 정규화를 수행하여 분산값들이 0과 255 사이의 값으로 분포하게 만든다.

$$v_{new} = \sqrt{v} \quad (5)$$

$$v_{normalized} = \left(\frac{v_{new} - min_1}{max_1 - min_1} \right) (max_2 - min_2) + min_2 \quad (6)$$

여기서 max_1 과 min_1 은 분산값 즉, v_{new} 의 최대, 최소의 밝기 값이며 max_2 와 min_2 는 새로운 범위의 밝기값을 의미한다. 그림 34는 수평 및 수직 방향의 분산 이미지를 보여주며, 색 변화가 크게 나타나는 부분은 밝기값이 크게 나타나고 변화가 작은 부분은 어둡게 나타난다. 이러한 분산 이미지를 분석하면 대부분의 텍스트 영역에서는 수평 및 수직의 분산값이 크게 나타나는 특징이 있다.

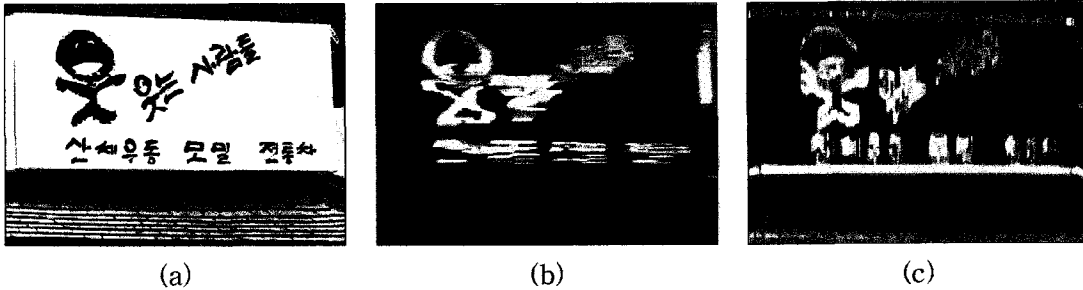


그림 34. (a) 원 이미지, (b) 정규화된 수평 방향 분산 이미지, (c) 정규화된 수직 방향 분산 이미지

(2) 논리적 AND 연산을 이용한 잡영 제거

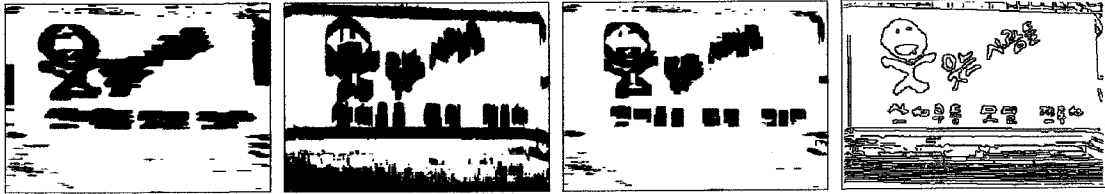
수평과 수직에 대한 분산 이미지의 특징을 이용하여 수평 및 수직 분산 이미지를 이진화하여 논리적으로 AND 결합하면 간단하게 텍스트 후보 영역을 찾을 수 있다. 그림 34(b)와 같이 수평 방향 분산 이미지는 가로 방향 에지 부분에서는 색 변화가 거의 없기 때문에 수평으로 긴 선 부분에서는 분산값이 아주 작게 나타나고, 그림 34(c)와 같이 수직 방향 분산 이미지는 수평 방향 분산 이미지와 반대로 가로 방향에서는 뚜렷하게 분산값이 큰 반면 세로 방향의 에지 부분에서는 색 변화가 거의 없기 때문에 분산값이 작게 나타나는 것을 볼 수 있다. 그러나 텍스트 영역들은 수평 및 수직 방향 모두 분산값이 높게 나타나기 때문에 이 두 이미지를 AND 결합하면 텍스트를 제외한 불필요한 영역들을 쉽게 제거할 수 있는 장점이 있다.

수평 및 수직 분산 이미지의 논리적 AND를 수행하기 전에 이진화를 먼저 수행한다. 본 연구에서는 임계값에 의한 이진화 과정에 정규화된 분산 이미지의 평균을 이용하여 임계값 T 를 결정한다. 임계값 T 는 다음 식 (7)과 같은 방법으로 모든 이미지에 대해 적응적으로 대응되도록 하였고, 여기에 경험적인 값인 상수 c 를 더해서 분산값이 높은 영역만을 강조하여 이진화한다.

$$T = \mu + c \quad (7)$$

μ : 정규화된 분산 이미지의 평균
 c : 실험적인 상수

그림 35는 두 이미지의 논리적 AND 결과와 명도이미지에서 에지 추출 결과를 비교한 것이다. (c)와 (d)를 비교한다면 명도이미지의 에지 추출에서 텍스트 주변에 분포하는 다양한 종류의 선들이 AND 연산으로 제거된 것을 볼 수 있다.



(a) 수평 분산이미지 (b) 수직 분산이미지 (c) AND 연산 (d) 에지 추출

그림 35. 수평 및 수직 분산이미지의 AND 연산 결과와 에지 추출 결과 비교

(3) 후보영역 추출

논리적 AND 연산을 수행한 결과 이미지는 분산값이 높으면서 색 변화가 큰 영역들이 주로 남게 된다. 이러한 결과를 가지고 모폴로지 연산을 수행하여 단순한 잡영은 제거시키고 후보영역은 강조한다. 모폴로지 연산시 수평 방향 위주로 강조하기 위해서 그림 36과 같은 구조 연산자를 적용하며, 수식 (8)과 같이 열림 연산과 닫힘 연산을 반복적으로 적용하였다.



그림 36. 불림 구조 연산자(좌), 지움 구조 연산자(우)

$$I = Closing + Opening + Opening \quad (8)$$

모폴로지를 적용하면 그림 37과 같이 AND 연산 결과 이미지와 비교하였을 때 잡영은 제거되고 후보영역이 강조되는 것을 볼 수 있다. 이렇게 강조된 영역에 대하여 연결요소를 분석하여 후보영역을 추출한다.

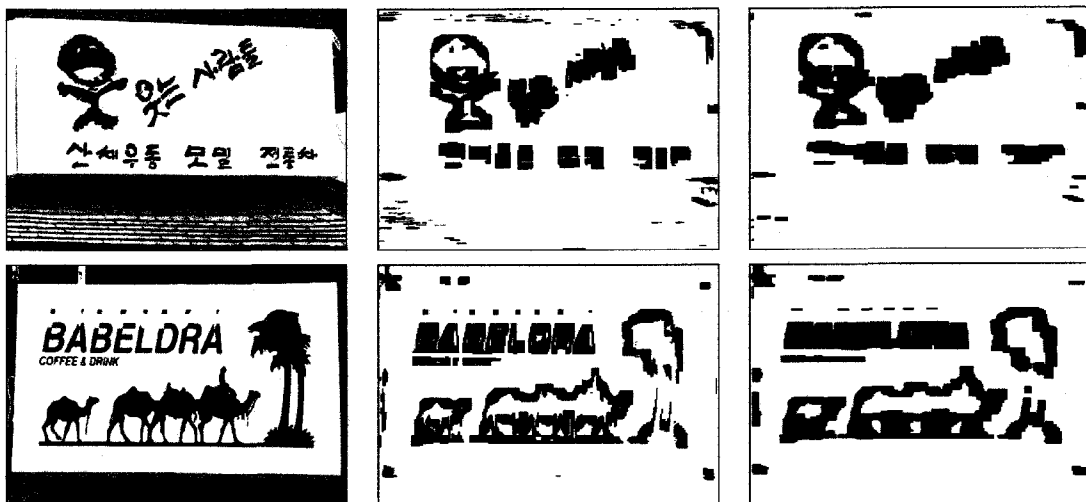


그림 37. 원 이미지, 논리적 AND 연산 결과, 모폴로지 적용 결과

후보영역의 검증은 밝기 정보를 이용한 방법과 동일한 방법을 사용한다. 우선 연결요소의 크기를 계산하여 너무 크거나 작은 것들을 제거하고, 연결요소와 연결요소의 외곽사각형 정보를 이용하여 외곽사각형내의 연결요소가 그 사각형을 차지하는 비율과 외곽사각형의 가로/세로 비율, 폭 또는 높이가 아주 작은 것들을 추가로 제거한다. 그리고 마지막으로 에지를 이용하여 제거한다. 그림 38은 최종적으로 추출된 텍스트 영역을 보여준다.

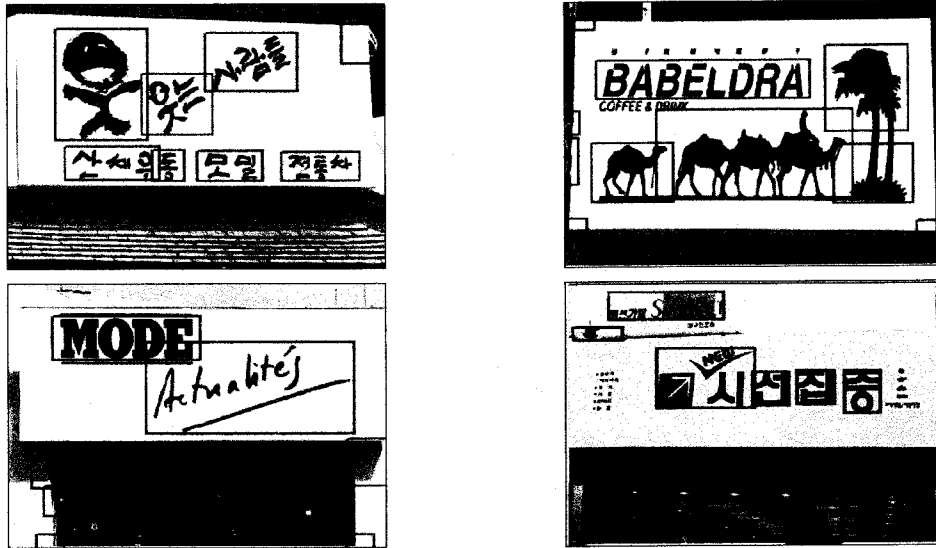


그림 38. 색 분산을 이용한 텍스트 추출 예

2. 검증

가. 영역 결합 및 선택

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 연구에서는 그림 39와 같이 각 방법에서 찾은 영역의 겹침 정도를 서로 비교해서, 검증으로 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 않기 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 에지를 정확하게 찾지 못하고, 배경이 복잡할 경우에는 텍스트 주변에 나뭇가지나 타일 등과 같이 많은 에지가 존재하여 오류가 발생하는 경우가 많다. 색 변화를 이용한 방법은 기울어진 텍스트와 배경이 복잡한 이미지에 대한 오류가 발생하기 쉽다. 이와 같이 각 방법의 단점을 파악하여 상호 보완적으로 결합을 한다면 보다 정확한 추출이 가능하다.

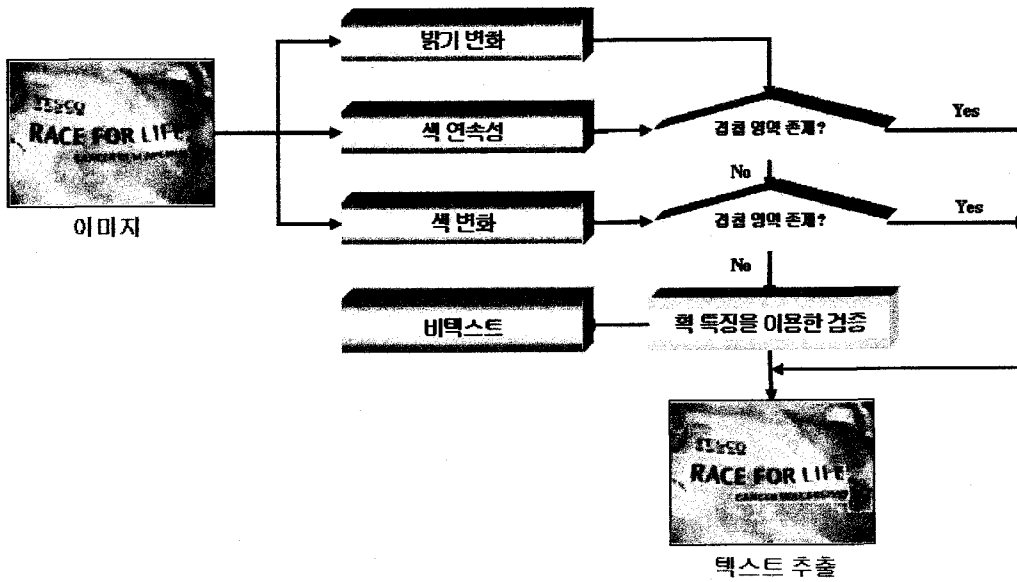


그림 39. 결합

결합은 각 특징에서 추출한 영역의 상호 겹침 정도를 비교하여 두 영역이 80% 이상 겹친 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 40은 결합한 결과의 예를 보여준다. 그림 40(e)의 ①번 영역은 세 특징 결과에서 모두 존재하기 때문에 검증 없이 텍스트 영역으로 확정되지만, 그림 40(d)의 ㉠-㉦ 영역은 하나의 특징에서만 존재하는 영역으로서 SVM 분류기로 검증되어 비텍스트 영역으로 분류되어 제거된 영역이다.

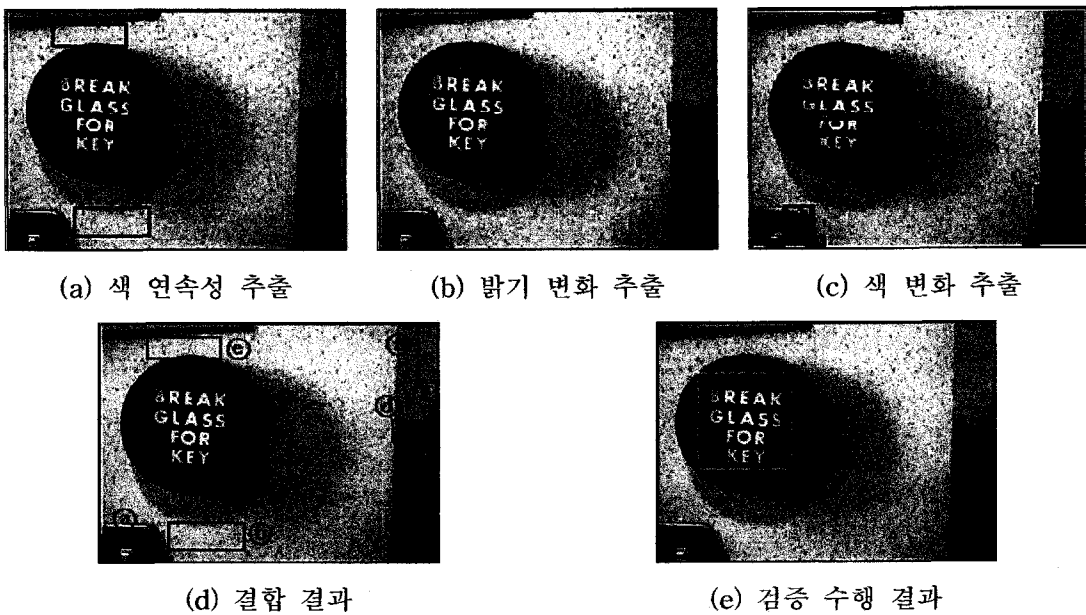


그림 40. 결합 및 검증 결과

나. SVM을 이용한 영역 검증

본 연구에서는 다해상도(Multi-resolution) 웨이블릿(Wavelet) 변환을 수행하여 문자 구성의 기본 단위인 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[21]. 또한, 자연이미지 내의 장면 텍스트는 다양한 크기로 존재하기 때문에 다해상도의 웨이블릿 변환을 이용하여 크기 변화에도 강인하게 만든다.

(1) 웨이블릿

웨이블릿 변환은 이미지에 수평, 수직 두 방향으로 필터를 적용하여, 이미지를 고주파 성분을 갖는 LH, HL, HH와 저주파 성분을 갖는 LL 등 4개의 부밴드(Subband)로 분해한다. 각 부밴드들은 서로 다른 특성을 나타내는데, LL은 원 이미지의 축소된 이미지이고, LH는 수평 성분, HL은 수직성분, HH는 대각성분에서 큰 웨이블릿 계수를 갖는다. 이러한 특성은 텍스트 영역의 수평, 수직, 대각선 방향의 획들이 고주파 부밴드에서 찾아지게 된다. 그림 41은 웨이블릿 변환의 예이다.



그림 41. 입력 이미지와 1차 웨이블릿 적용 결과

(2) 특징 추출

16x16 크기의 이미지 영역을 반복적으로 해상도를 줄이면서 웨이블릿 변환하여 특징 벡터를 추출한다. 특징 벡터는 웨이블릿 변환으로 생겨난 부밴드들의 평균(M), 2차(μ_2), 3차(μ_3) central moment를 사용한다[22]. 크기가 $N \times N$ 인 부밴드 I 에 대해 특징들은 식 (9)와 같이 계산된다.

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

$$\mu_2(I) = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (I(i, j) - M(I))^2 \quad (9)$$

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

이때의 웨이블릿 변환은 부밴드가 한 개의 화소가 될 때까지 반복한다. 따라서 16x16 크기의 영역은 3-레벨 웨이블릿 변환이 수행되어 12개의 부밴드들이 생겨서 36차원의 특징 벡터가 만들어진다.

(3) 특징 선택

구해진 특징 벡터들은 SVM에 입력되어 학습된다. 이 과정에서 특징의 차원이 너무 크면 많은 수의 학습 데이터와 학습 시간을 필요로 하기 때문에 특징 벡터의 차원을 줄일 필요가 있다. 특징 벡터의 차원을 줄이기 위해서 16x16 크기의 텍스트 데이터 1,000개와 텍스트가 아닌 데이터 1,000개를 웨이블릿 변환해서 36차원의 특징을 추출한 뒤, 각 500개는 학습 집단으로 나머지 500개는 테스트 집단으로 사용하여 Bayes 에러율을 구하고, 에러율이 작은 순서대로 12개를 선택하였다. 여기서 텍스트 집단은 색 연속성 특징과 밝기 변화 특징을 이용하여 찾은 영역 중 실제로 텍스트인 영역을 사용하였으며, 텍스트가 아닌 집단은 텍스트가 아님에도 불구하고 텍스트로 틀리게 찾은 영역을 사용하였다. 그림 42는 두 집단이 잘 분리되어 변별력이 높은 특징 차수와 두 집단이 잘 분리되지 않는 변별력이 낮은 특징 차수의 분포를 보여준다.

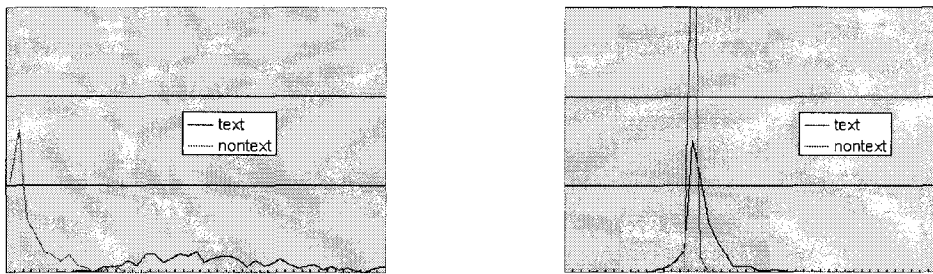


그림 42. 변별력이 높은 특징 차수(좌), 변별력이 낮은 특징 차수(우)

(4) SVM을 이용한 특징벡터 학습

SVM 학습에는 앞서 Bayes 에러율을 구하는데 사용되었던 텍스트 집단과 텍스트가 아닌 집단의 데이터를 웨이블릿 변환하여 얻은 12차원 특징벡터를 입력으로 사용한다. 특징들은 -1에서 1사이의 값을 갖도록 정규화하였다. 표 2는 학습 데이터의 수와 비율에 따른 인식률을 나타낸다. 인식률 측정에 사용된 테스트 데이터는 텍스트 데이터와 텍스트가 아닌 데이터를 각각 500개씩 사용하였다. 실험을 통해 학습 데이터는 텍스트 데이터 1,000개, 텍스트가 아닌

데이터 500개를 2:1 비율로 사용하였을 때 가장 좋은 결과를 만들었다.

표 2. 학습데이터의 수와 비율에 따른 블록(16x16크기) 단위 인식률

	텍스트 학습 데이터	텍스트가 아닌 학습 데이터	SV 개수	인식률 (%)	
				텍스트	텍스트 아님
1:1	500	500	446	66.6	79.8
2:1	1000	500	547	97.4	68.0
1:2	500	1000	460	87.6	86.6

(5) 검증 수행

SVM을 이용한 검증의 목표는 색 연속성, 색 변화, 밝기 변화 정보를 이용하여 텍스트라고 추출한 영역을 확인하는 것이다. 즉, 한 영역을 구성하는 여러 단위 블록(16x16) 각각에 대해 텍스트인지 아닌지를 결정하고, 그 결과를 통합하여 영역 전체에 대한 결정을 내린다. 색 연속성과 밝기 변화를 이용하여 추출된 텍스트 영역에 16x16 크기의 윈도우를 띄우고 웨이블릿 변환하여 12차원 특징 벡터를 구한 뒤 SVM을 이용한 학습에서 구해진 SV(Support Vector)들과의 거리를 계산하여 텍스트 여부를 판단한다. 윈도우 영역이 텍스트로 분류되면 윈도우 안의 모든 화소들을 텍스트로 레이블링하고 4 화소 이동한다. 이동한 위치에서도 같은 방법으로 반복한다. 색 정보와 명도 정보를 이용해서 텍스트로 찾아진 모든 영역에 대해서 위와 같이 수행한 후, 텍스트 윈도우와 겹치지 않는 화소들은 텍스트가 아닌 영역으로 레이블링한다. 최종적으로 텍스트로 레이블링 된 영역이 전체 영역의 2/3 가 넘으면 그 영역을 텍스트 영역으로 판단하고, 그렇지 않으면 텍스트가 아닌 영역으로 판단한다. 그림 43은 색 연속성과 밝기 변화를 이용해서 찾아진 영역을 검증한 결과이다.

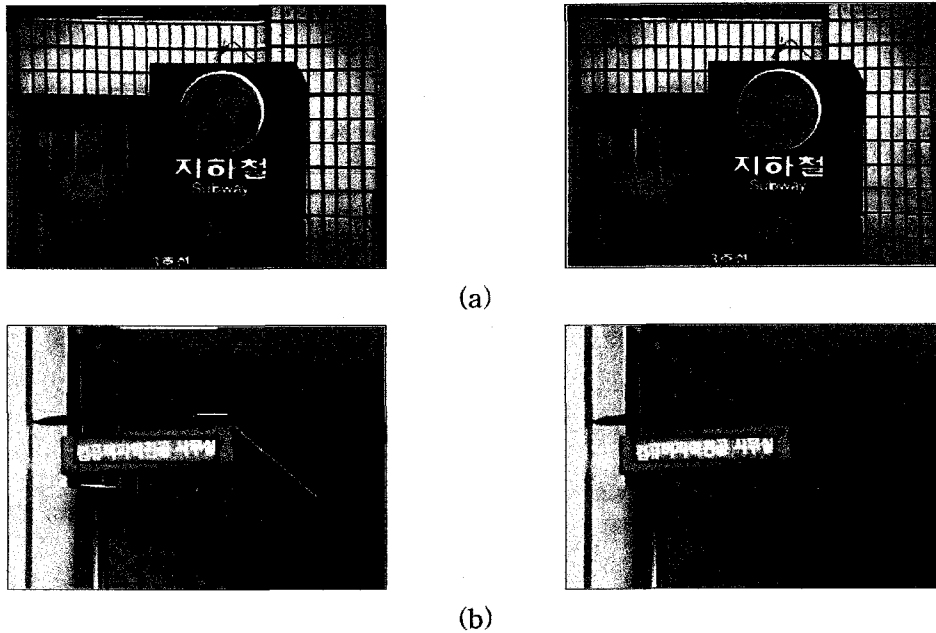


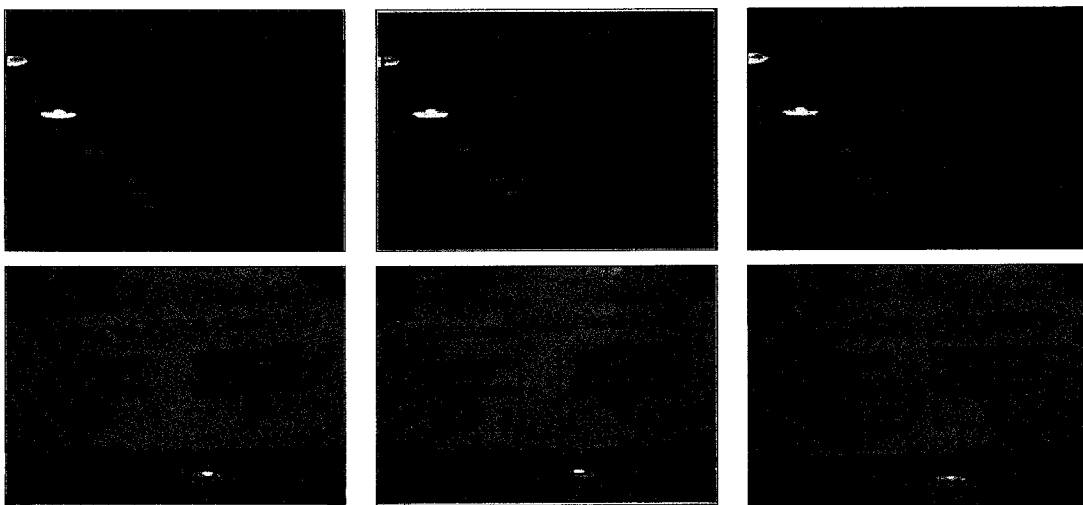
그림 43. (a) 색 연속성의 추출 결과(좌) 및 SVM 검증 결과(우) (b) 밝기 변화 추출 결과(좌) 및 SVM 검증 결과(우)

3. 다양한 환경에서의 추출

가. 크기 변화에 무관한 추출

자연이미지에는 다양한 크기의 텍스트들이 포함된다. 본 연구에서는 다양한 크기의 텍스트 영역을 추출하기 위해서 색 변화 특징을 이용한 추출과 높은 수준의 희 특징을 고려한 검증 과정에서 크기 변화에 무관한 추출 방법을 고려한다. 색 변화 특징에서만 고려한 것은 색 분산 이미지를 생성하는 과정에서 다양한 크기의 윈도우를 이용할 수 있고, 수평과 수직 분산 이미지를 AND 연산하는 과정에서 텍스트 정보의 손실을 최소화할 수 있기 때문이다.

색 변화를 이용한 추출에서는 320x240 크기의 이미지를 사용하여 텍스트 영역을 추출하기 때문에 텍스트의 크기가 전체 이미지의 1/3 이상을 차지할 정도로 크다면 연결요소와 외곽사각형을 분석하는 과정에서 잡영으로 판단되어 제거되는 경우가 발생한다. 따라서 본 연구에서는 320x240 크기의 1/2 축소 이미지인 160x120 크기의 이미지를 함께 사용하여 각각에서 앞에서 제안한 방법과 동일한 방법으로 후보영역을 추출하고, 두 결과를 결합하여 색 변화를 이용한 텍스트 추출 결과로 확정한다. 그림 44(a)는 320x240 크기의 이미지를 사용하였을 때의 크기에 대한 오류를 보여준다. 그림과 같이 대부분의 오류는 320x240 크기의 이미지에서 텍스트가 차지하는 비율이 1/5~1/3 이상인 경우이다. 과정은 320x240과 160x120 크기의 입력 이미지 각각에 대하여 분산 이미지 생성, 논리적 AND 연산, 후보영역 추출 단계를 수행한다. 그림 44(a)는 320x240 크기의 이미지에서의 결과이고, 160x120 크기의 이미지에서의 결과는 그림 44(b)와 같다. 그림과 같이 큰 텍스트도 추출됨을 확인할 수 있다. 2단계 이미지에서의 결과를 각각 320x240 크기로 변환하여 포함 및 겹침 정도를 고려하여 OR 결합한 결과가 그림 44(c)에 보여준다.



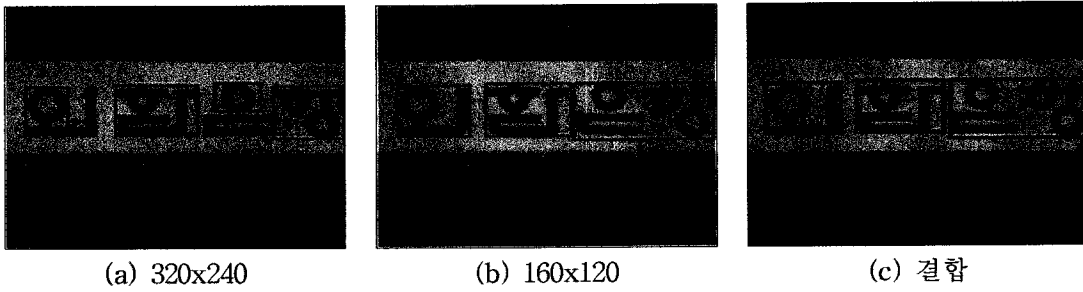


그림 44. 2단계 해상도에서의 추출 및 결합 결과

다음으로 검증 과정에서의 크기 변화에 대한 고려이다. 간판이미지와 ICDAR 이미지의 경우 검증 과정에서 이미지에 포함된 텍스트의 크기가 16x16 보다 큰 경우가 많기 때문에 추출된 후보영역이라 할지라도 검증과정에서 텍스트가 아니라고 판단되는 경우가 많다. 그림 45는 간판이미지와 ICDAR 이미지에서 텍스트의 크기로 인해서 발생한 오류를 보여준다.

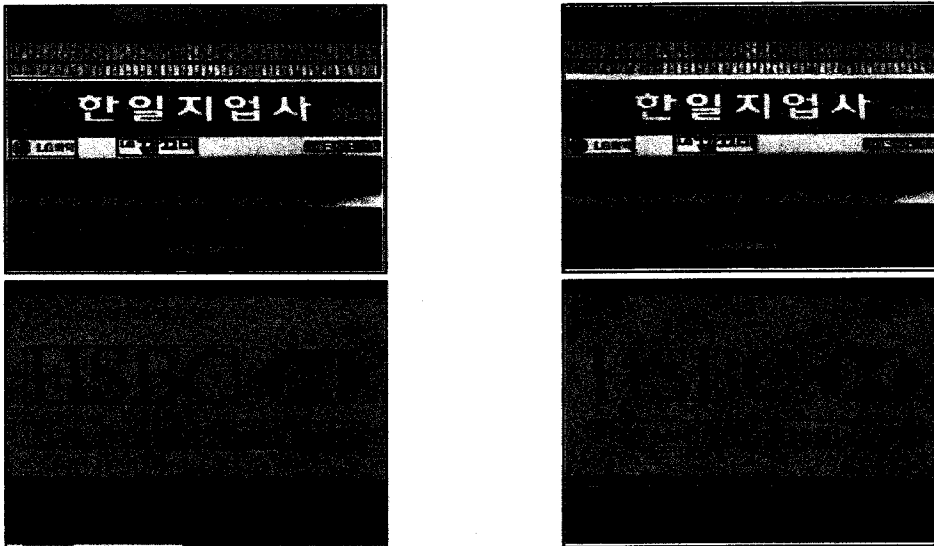
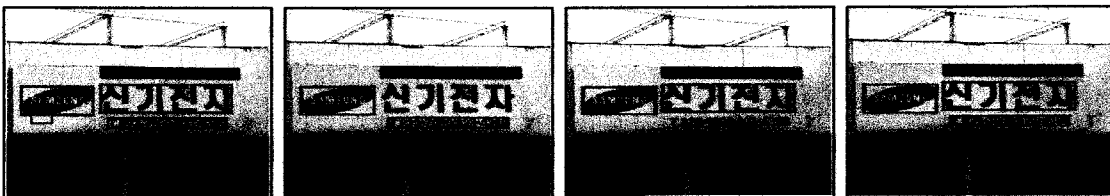
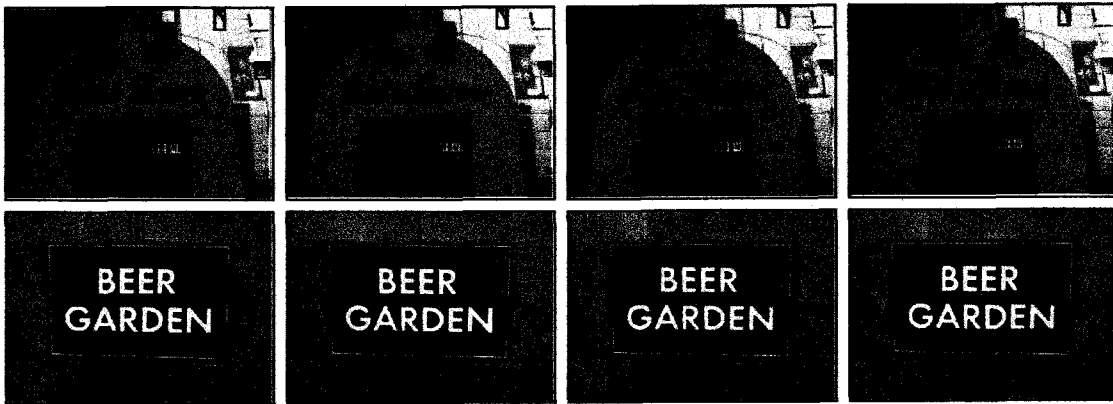


그림 45. 텍스트의 크기로 인한 오류 발생의 예

이러한 오류를 줄이기 위해서 160x120으로 축소된 이미지를 추가적으로 이용하고, 높이와 너비가 16 화소 이상인 외곽사각형에 대해서 같은 방법으로 검증을 수행하여 320x240의 검증 결과와 OR시킨다. 축소된 이미지를 사용함으로써 이미지에 포함된 크기가 큰 텍스트는 16x16 윈도우에 적합한 크기를 갖게 되어 검증 단계에서 텍스트로 분류할 수 있다. 그림 46은 텍스트의 크기가 커서 텍스트가 아니라고 검증된 오류가 발생한 320x240 크기의 이미지를 160x120 크기의 이미지에서 검증을 수행하여 텍스트로 분류한 결과이다.





(a) 추출된 후보영역 (b) 320x240에서 검증 (c) 160x120에서 검증 (d) OR 결합
그림 46. 크기를 고려한 검증 결과

나. 시각 변화에 무관한 추출

본 연구결과를 다양한 응용 분야에서 활용하기 위해서 동영상에서의 장면 텍스트를 효율적으로 추출하기 위한 텍스트 영역 추적 방법을 제안한다. 대부분의 동영상에는 다양한 시각에서 다양한 방향으로 장면 텍스트가 존재한다. 이러한 장면 텍스트에 대해 색 연속성, 밝기 변화 및 색 분산 정보를 이용하여 찾아진 텍스트 영역에 대해 추적이 가능하다면 시간과 노력을 절약하고 보다 정확하고 실시간으로 텍스트 영역을 추출할 수 있다. 본 연구에서 제안하는 방법은 크게 템플릿 매칭(Template Matching)을 이용한 영역 추적과 영역 보정으로 나눌 수 있다. 이전 프레임과 현재 프레임의 이동 방향을 예측하여 대략적인 텍스트 영역의 이동된 위치를 찾고, 텍스트의 크기, 기울기, 원근 변화에 따른 텍스트 영역의 변형을 보정한다.

(1) 텍스트 영역 추적

이전 두 개 이상의 프레임($n, n-1, \dots$)에서의 텍스트의 이동 방향을 찾아서 다음 프레임($n+1$)에서의 텍스트의 이동 방향을 예측한다. 템플릿 매칭을 하기 위해서 탐색할 관심 영역(ROI, Region of Interest)의 방향을 예측된 방향으로 옮김으로써, 모든 방향으로 탐색하는 것보다 계산량을 많이 줄일 수 있다. 예를 들어 x 방향으로의 예측 방향은 식 (10)과 같이 계산되며, y 방향으로의 예측 방향도 같은 방법으로 계산된다.

$$x_{n+1} = x_n + (x_n - x_{n-1}) \quad (10)$$

계산된 x_{n+1}, y_{n+1} 을 이용하여 탐색할 영역의 범위는 식 (11)과 같이 선택된다. W_n, H_n 는 이전 프레임(n)에서 추적된 영역의 넓이와 높이이다.

$$\begin{aligned} x_{n+1} - 10 &\leq ROI_x \leq x_{n+1} + W_n + 10 \\ y_{n+1} - 10 &\leq ROI_y \leq y_{n+1} + H_n + 10 \end{aligned} \quad (11)$$

검색할 윈도우 내에서, 이전 프레임($n-1$)에서 찾아진 텍스트 영역의 명도 이미지를 템플릿

으로 하여 템플릿 매칭을 실시한다. 템플릿 매칭 방법은 SSD(Sum of Squared Difference)를 최소화하는 영역을 찾는 것으로, 식 (12)의 ϵ 을 최소화하는 영역이 현재 프레임에서 추적된 문자영역이 된다. J, I 는 각각 탐색 영역과 템플릿의 명도값이고, W 는 탐색 영역을 의미한다.

$$\epsilon = \iint_W [J(\vec{v} + d) - I(\vec{v})]^2 d\vec{v} \quad (12)$$

실생활에서는 카메라의 이동 등에 의해서 텍스트의 크기, 기울기, 원근이 변화하게 된다. 이러한 여러 변화에서도 정확한 텍스트 영역을 적합(Fitting)시켜야 다음 프레임에서의 텍스트 추적의 신뢰성을 높일 수 있다.

(2) 영역 보정

추적에서 찾아진 영역을 초기 영역으로 하여 텍스트 영역의 변화에 따른 영역의 확장 또는 축소를 수행한다. 텍스트는 배경과 구분되는 색, 명도를 갖기 때문에 초기 영역을 이진화하여 텍스트와 배경을 구분한다. 이런 특징을 이용하여 초기 영역에서의 경계(Boundary) 화소 연결성으로 문자 영역의 확장 또는 축소를 결정한다.

이진화할 때의 임계값을 T 라고 하고, 텍스트 영역의 명도값이 T 보다 작으면 전체 흐름도는 그림 47과 같다.

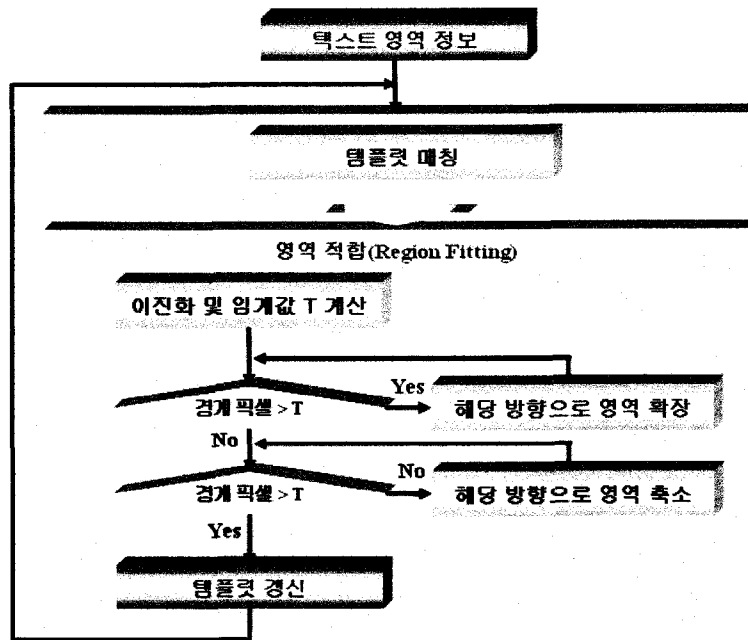


그림 47. 동영상에서 텍스트 영역 추적

텍스트 영역의 변화로 초기 영역보다 영역이 커진 경우 영역을 확장한다. 초기 영역의 경계 화소들의 명도 값을 T 와 비교하여 T 보다 작고, 연결된(Connected) 경계 안쪽 화소가 존재하면 현재 방향 바깥쪽으로 1화소 크기를 확장한다. 4 방향에 대해서 반복 실시함으로써 확장되

어려야할 방향으로 영역을 1화소 확장한다.

반대로 텍스트 영역의 변화로 초기 영역보다 작아졌을 경우에는 영역을 축소한다. 초기 영역의 경계 화소들의 명도 값을 T 와 비교하여 T 보다 작지만, 연결된 경계 안쪽 화소가 존재하지 않으면 현재 방향 안쪽으로 1화소 크기 축소한다. 4 방향에 대해서 반복하여 실시함으로써 축소되어야할 방향으로 영역을 1화소 축소한다.

초기 영역이 더 이상 확장되거나 축소되지 않을 때까지 영역 확장과 축소를 반복하여 영역을 보정한다. 이 방법을 이용하여, 텍스트 영역 추적 시에 기울기, 확대, 축소 등의 변화에도 정확한 템플릿을 유지할 수 있다. 추적에 사용되는 템플릿은 매 프레임마다 보정을 통해서 갱신되고, 그 결과로 보다 좋은 추적이 가능하다. 그림 48은 위의 방법을 이용하여 추적하고, 영역을 보정한 결과이다.

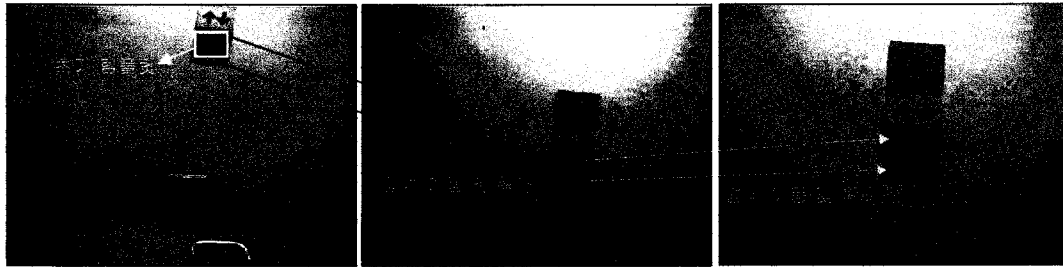
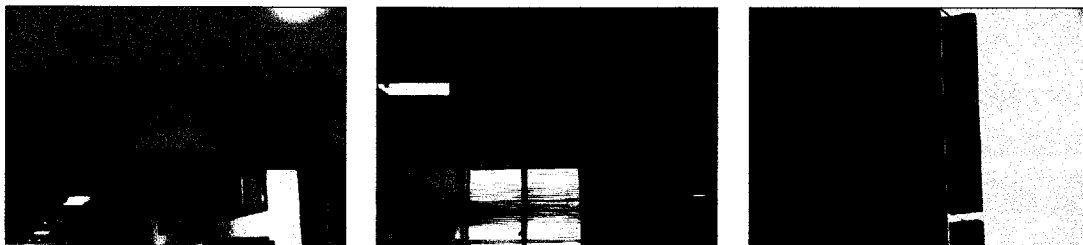


그림 48. 영역 추적 및 보정 결과

본 연구에서는 간단하고 빠른 방법을 이용하여 찾아진 영역을 추적하는 방법을 제안한다. 템플릿 매칭으로 대략의 영역을 찾고 영역을 보정한다. 보정된 영역은 다시 다음 추적의 템플릿으로 사용하여 영역의 변화에도 강인한 추적이 가능하다.

다. 조명 변화에 무관한 추출

본 연구에서는 조명 변화에 무관한 추출을 위해서 뇌신경정보학연구사업을 함께 참여하고 있는 연세대학교 정찬섭 교수팀에서 개발한 Color constancy 방법을 적용한다. 그림 49는 원 이미지에 대해 Color constancy 방법을 적용하여 복원된 이미지를 보여주며, 보다 정확한 텍스트 영역의 추출 결과를 기대할 수 있다.



(a) 원 이미지



(b) 이미지 복원 결과

그림 49. Color constancy 적용 결과

4. 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 실내외 환경에서 취득한 120개의 자연이미지, 다양한 크기의 200개의 실외 간판이미지 및 ICDAR에서 제공한 367개의 자연이미지를 사용하였다. 텍스트 영역의 Ground Truth 기준으로서 수평 방향으로 글자 간격이 너무 떨어진 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, 기울어짐과 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 3과 같으며, *Total*은 전체 실험 대상 이미지에서의 텍스트 영역의 개수이며, *Correct*는 정확히 찾은 텍스트 영역의 개수, *Partial*은 텍스트 영역의 일부만이 찾아진 개수, *Missing*은 텍스트 영역을 찾지 못한 개수이며, *False*는 텍스트가 아닌 영역을 텍스트로 찾은 개수이다. 성능을 측정하는 단위로서 식 (13)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[23].

$$Precision = (Total\ number\ of\ correct\ texts) / (Number\ of\ detected\ texts) \quad (13)$$

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

표 3. 각 실험 이미지에 대한 추출 결과

			<i>Total</i>	<i>Correct</i>	<i>Partial</i>	<i>Missing</i>	<i>False</i>
실험실 구축 자연이미지	단순 이미지	영역 개수	256	239	2	15	78
		정확률 = 75.5%		재현률 = 94.1%			
	복잡 이미지	영역 개수	328	291	13	24	148
		정확률 = 67.2%		재현률 = 92.6%			
간판 이미지		영역 개수	839	645	50	144	395
		정확률 = 63.7%		재현률 = 82.8%			
ICDAR 컨테스트 이미지	학습 이미지	영역 개수	378	235	18	125	142
		정확률 = 64.0%		재현률 = 66.9%			
	테스트 이미지	영역 개수	687	474	29	184	346
		정확률 = 59.2%		재현률 = 73.2%			

표 3의 실험실에서 구축한 자연이미지는 배경의 복잡도에 따른 추출 성능을 보기 위한 것으로서 단순한 이미지와 복잡한 이미지 모두 재현률이 높은 좋은 결과를 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 50은 자연이미지에 대한 영역 추출 결과의 예를 보여준다.

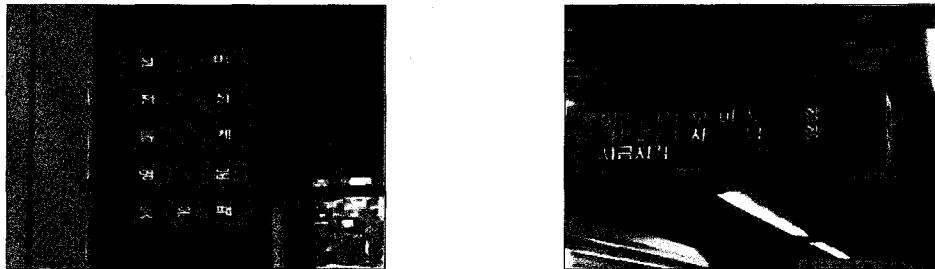
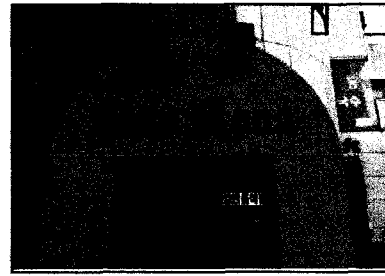
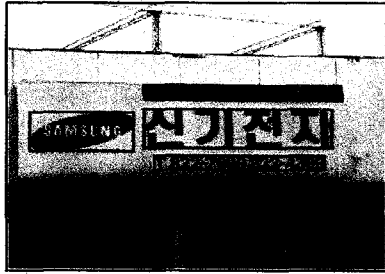


그림 50. 자연이미지에 대한 영역 추출 결과 예

표 3에서 간판이미지는 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 제한하여 크기가 적당한 텍스트에 대해서만 추출하였다. 그림 51(a)은 크기와 폰트 스타일 등에 의해 오류가 발생한 예를 보여준다. 본 연구의 3차년에서는 크기에 대한 문제를 해결하기 위해서 2 단계 크기의 이미지를 고려함으로써 그림 51과 같은 오류를 해결할 수 있었다. 그림 51(b)은 색 변화 특징에서 크기 변화에 무관하게 추출한 결과를 보여준다. 그림에서 볼 수 있듯이 크기가 큰 텍스트에 대한 오류가 사라진 것을 볼 수 있다.



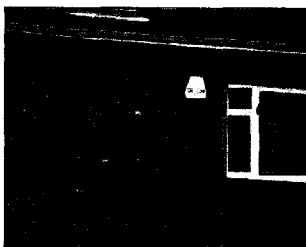
(a) 간판이미지에서 발생한 오류의 예



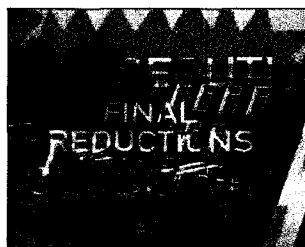
(b) 크기 오류에 대한 보완 결과

그림 51. 간판이미지에서 발생한 오류의 예

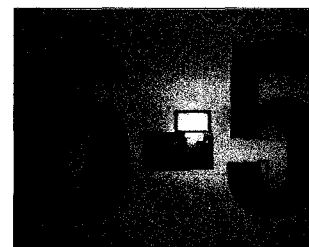
ICDAR 컨테스트 이미지는 텍스트 추출 경연대회에서 사용하기 위해서 116개의 학습용 이미지와 251개의 테스트용 이미지로 나누어 구성되어 있으며, 본 연구에서는 모두 테스트용으로 사용하였다. ICDAR 이미지에 포함되어 있는 텍스트는 모두 영문자와 숫자이며, 간판 이미지와 같이 폰트의 크기, 모양, 조명 변화 등이 다양하다. ICDAR 이미지 테스트 결과에 대한 다양한 종류의 오류 예를 그림 52에서 볼 수 있다. (a)는 주로 SVM 검증에서 걸러내지 못해서 발생한 오류들로서, SVM 분류기를 학습시킬 때 포함되지 않은 새로운 패턴들로 인해서 발생한 것으로 판단된다. (b)는 복잡한 배경으로 인해서 발생한 오류이며, (c)는 간판이미지와 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.



(a) SVM 검증 오류



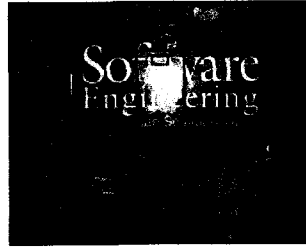
(b) 복잡한 배경에 의한 오류



(c) 텍스트 크기에 의한 오류



(d) 양각 효과에 의한 오류



(e) 심한 조명에 의한 오류

그림 52. ICDAR 이미지에서 발생한 오류의 예

텍스트 영역의 추출 시간은 배경의 복잡도에 따라 단순한 이미지와 복잡한 이미지로 분류하여 측정하였다. 표 4는 이미지 수준의 세 가지 특징에서의 처리 시간과 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 차이가 있음을 확인할 수 있다. 처리 시간을 더 단축시키기 위한 각 모듈에 대한 최적화 과정이 필요하다.

표 4. 배경의 복잡도에 따른 평균 처리 시간(단위: msec)

	175	210	631	138	1,154
	190	382	651	246	1,469

그림 53부터 그림 57은 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 52에서 언급한 문제점에 대한 제안한 방법의 한계로서 향후 연구에서 고려하고자 한다.

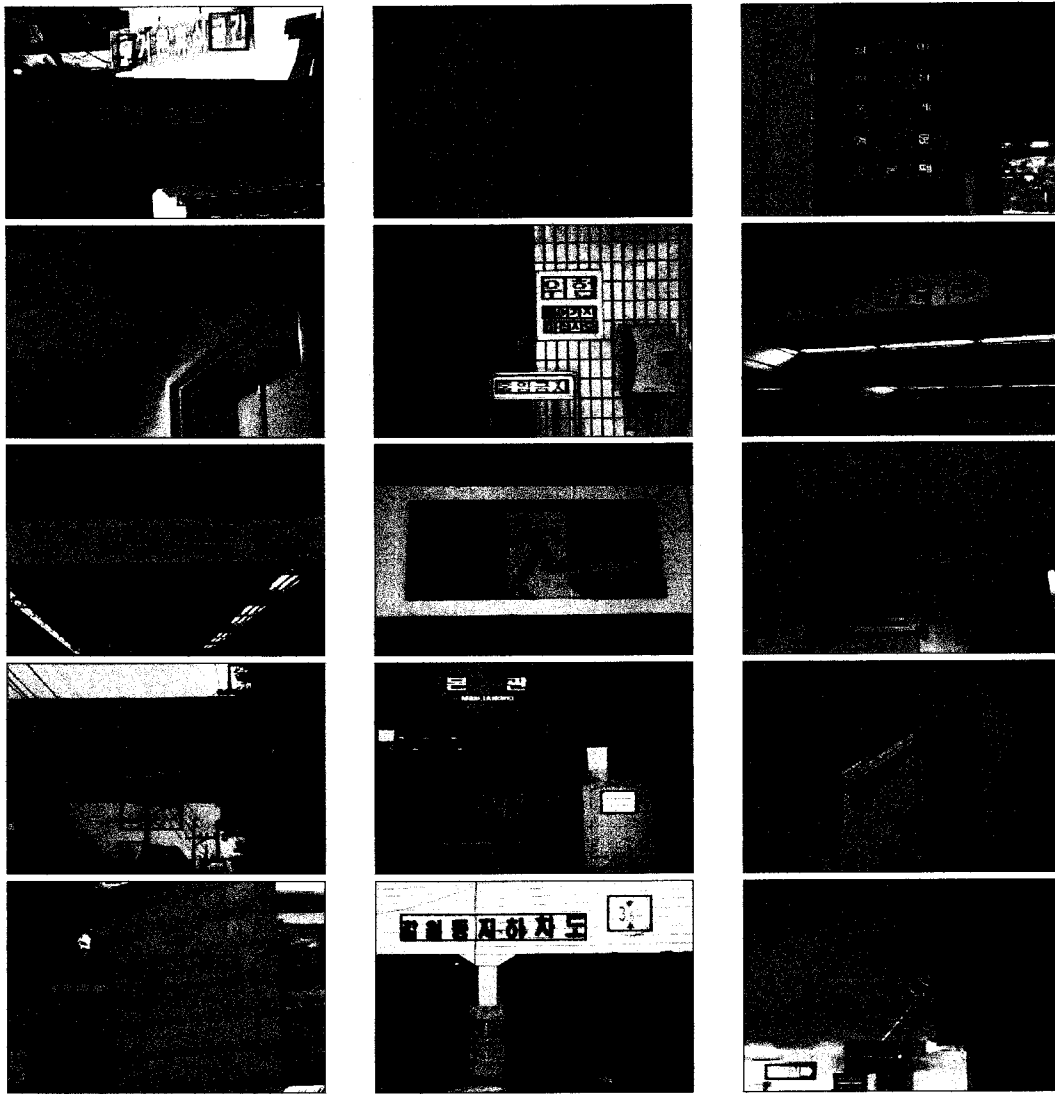


그림 53. 실험실 구축 자연이미지의 추출 결과

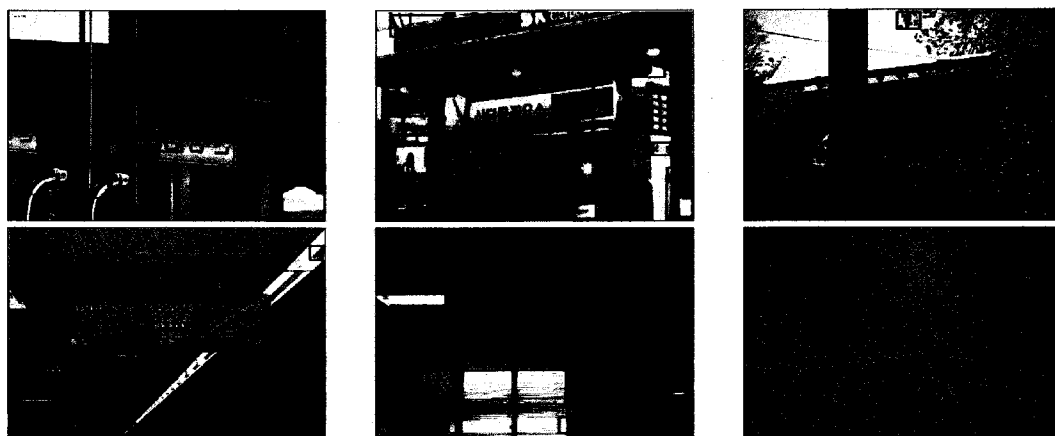


그림 54. 실험실 구축 자연이미지의 오류의 예



그림 55. 간판이미지 추출 결과

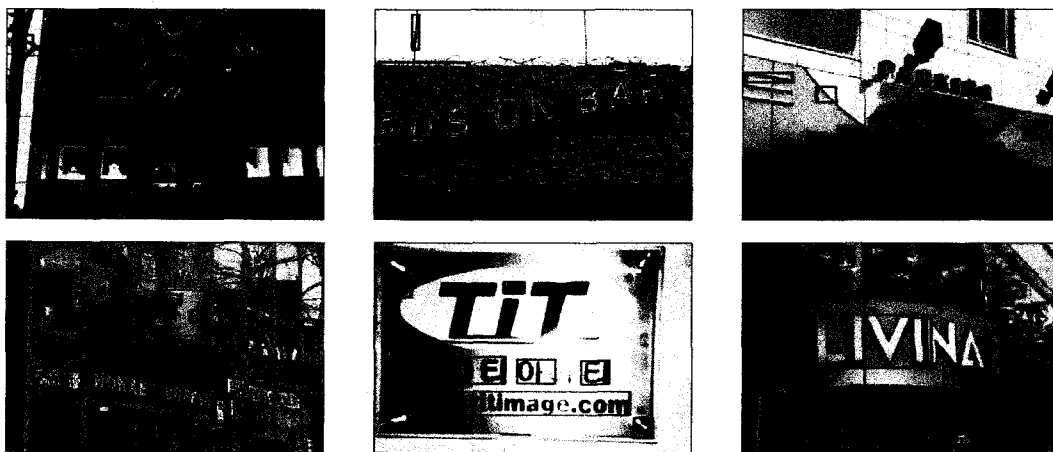


그림 56. 간판이미지 오류 예

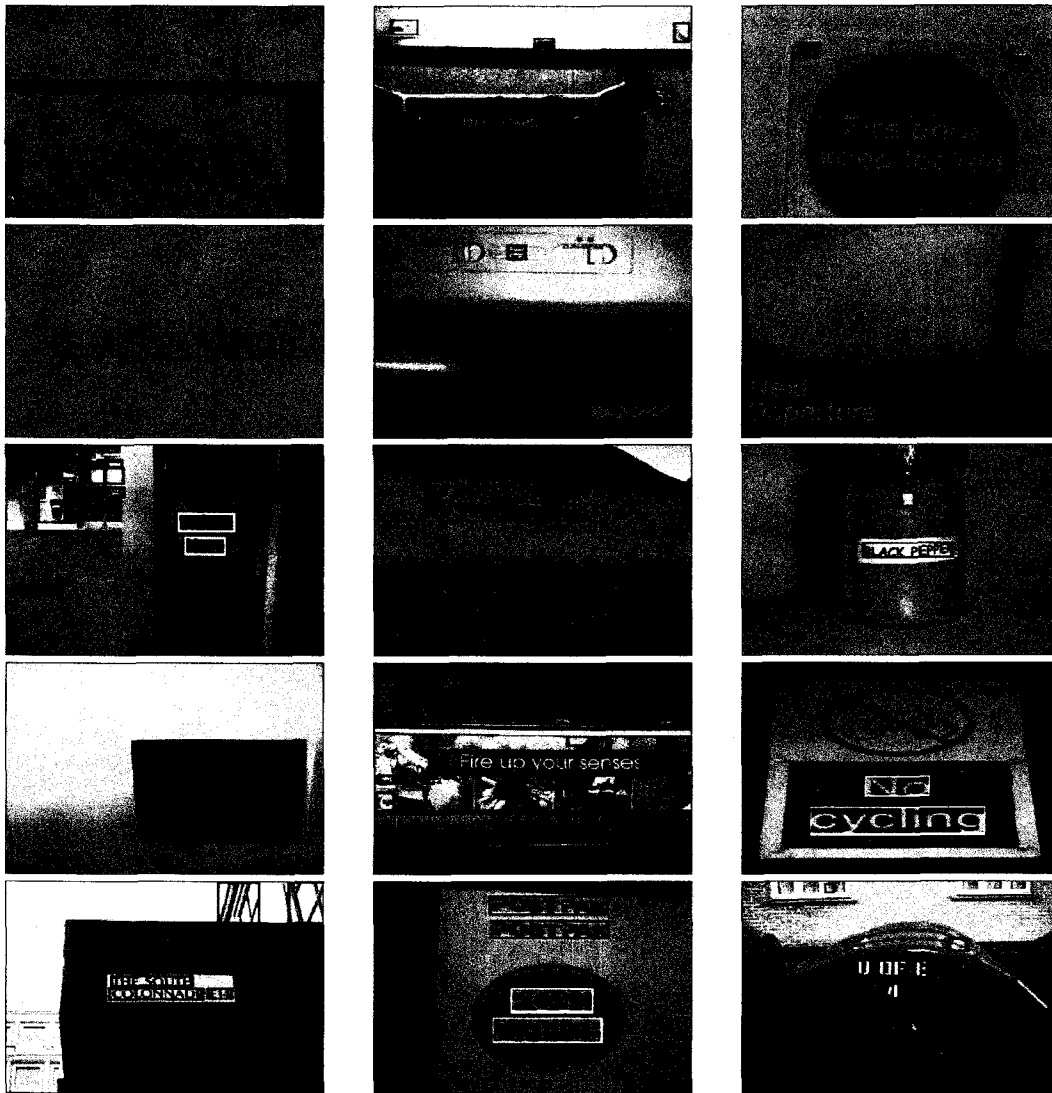


그림 57. ICDAR 컨테스트 이미지 추출 결과

제 2 절. 텍스트 영역 처리

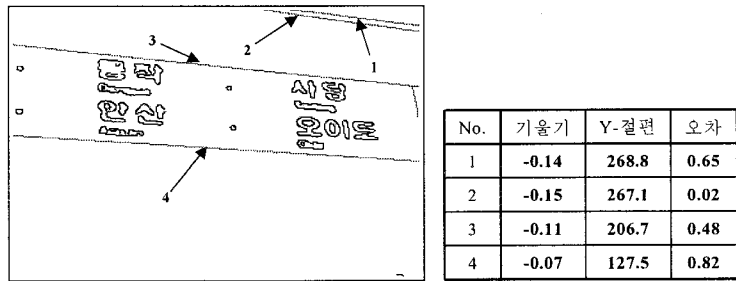
1. 기울어짐 및 원근 보정

추출된 텍스트 영역을 정확히 인식하기 위해서는 텍스트 영역의 기울어짐(Skew)과 원근(Perspective) 보정이 필요하다. 그림 58은 이미지 내의 긴 선들을 이용하여 기울어짐과 원근을 추정하는 것을 보여준다. 추출한 긴 선들에 최소자승법(Least Mean Square) 방법을 적용하여 가장 근사한 직선의 식을 구하고, 텍스트 영역과의 위치를 고려하여 기울어짐과 원근을 추정한다.

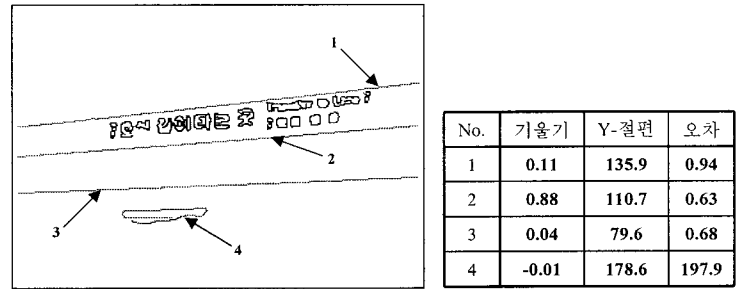
가. 긴 선 정보 추출

앞의 긴 선 제거 과정에서 추출한 선들을 이용하여 추출된 텍스트 영역의 기울어짐을 추정

한다. 이렇게 추출된 긴 선들 중에는 한 두 화소가 끊어져 있거나, 위 아래로 튀어나온 화소가 존재하는데 이러한 화소 때문에 최소자승법을 이용한 직선 추정시 에러값이 증가하는 경우가 발생한다. 이러한 경우를 보완하고 정확한 기울어짐을 추정하기 위해서 간단한 필터를 구성하여 선을 스무딩(Smoothing)한다. 스무딩된 선의 기울기와 위치를 나타내는 절편을 구하기 위해서 각각의 선에서 10개의 좌표를 추출한다. 이 과정에서 긴 선임에도 불구하고 그림 58(a)의 3번처럼 전혀 다른 방향으로 심하게 꺾이는 부분이 일부 포함될 수 있는데 이러한 꺾이는 부분이 직선의 방정식에 영향을 주지 못하도록 긴 선 정보 중에 4방향 히스토그램을 이용하여 히스토그램의 최대 빈(Bin)을 가지는 방향에 대해서만 좌표를 추출한다.



(a)



(b)

그림 58. 긴 선 요소의 기울기, Y-절편 및 오차

각 선 요소에서 추출된 10개의 좌표를 최소자승법에 적용하여 직선의 기울기와 절편을 구한다. 이 때 그림 58(b)의 4번처럼 최소자승 오차 값이 임계값 이상으로 크게 발생하면 대부분이 곡선이거나 부분적으로 심하게 꺾이는 선들이기 때문에 기울기 추정 대상에서 제외시킬 수 있다. 그림 58은 추출된 긴 선 요소들과 위의 방법으로 구해진 기울기, 절편 및 오차 값을 보여준다. 절편은 직선의 위치 정보를 나타내며, 이 정보는 텍스트 영역에 적용될 위치의 직선의 기울어짐과 원근을 추정하는데 사용된다.

나. 선 위치에 따른 기울기 병합

제거된 긴 선들은 다양한 기울기를 가지며 다양한 위치에 분포한다. 텍스트 영역의 기울어짐과 원근을 추정하기 위해서는 인접한 위치에 존재하는 기울기가 유사한 선들을 하나로 병합할 필요가 있다. 선을 병합하기 위해서 기울기와 절편 값을 이용한다. 최소자승 오차 값이 작은 선들에 대해서 기울기가 유사하면 다시 절편 값을 고려하여 병합한다. 이 과정을 거치면 그림 59(a)의 1, 2번 직선은 병합된다. 두 직선의 기울기가 유사하더라도 절편 값의 차이가 크

면 두 직선이 서로 다른 텍스트 영역에 인접해 있다고 판단 할 수 있기 때문에 그림 59(b)의 1, 2번은 병합하지 않는다. 병합된 직선들은 기울기의 평균값과 절편의 평균값을 최종적으로 해당 영역의 직선의 기울기와 절편으로 설정한다. 병합된 직선의 기울기와 절편을 이용하여 텍스트 영역의 기울어짐과 원근을 추정한다.

다. 영역에 따른 기울어짐/원근 추정

기울어짐 추정은 텍스트 영역의 중심에서 기울기 병합에 의해 남은 모든 직선으로의 수직 거리를 구한다. 이들 수직 거리 중 가장 가까운 직선의 기울기를 최종적으로 텍스트 영역의 기울어짐 정도로 추정한다. 이때 만약 음의 방향과 양의 방향에 서로 유사한 거리에 두 직선이 존재한다면 두 직선을 이용하여 원근이 존재하는 경우도 추정할 수 있다. 본 연구에서는 원근 정도를 두 직선의 좌측과 우측 절편의 차로 추정하고 있다. 또한, 두 직선의 기울기 차이가 없는 경우에는 기울어짐만을 추정한다. 예로 59(b)에서 ①번 영역의 중심에서 59(a)의 세 개의 직선까지의 수직거리를 구하면 59(a)의 1번 직선이 양의 방향으로 가장 가깝고, 2번 직선이 음의 방향으로 가장 가깝다는 것을 알 수 있다. 여기서 추출된 ①영역은 가장 가까운 직선 1에 의해 기울어짐을 추정할 수 있고, 서로 대립된 두 직선을 이용하여 원근정도를 추정할 수 있다. 그림 59(c)는 추출된 텍스트 영역에 대해 추정된 기울어짐과 원근을 보여준다.

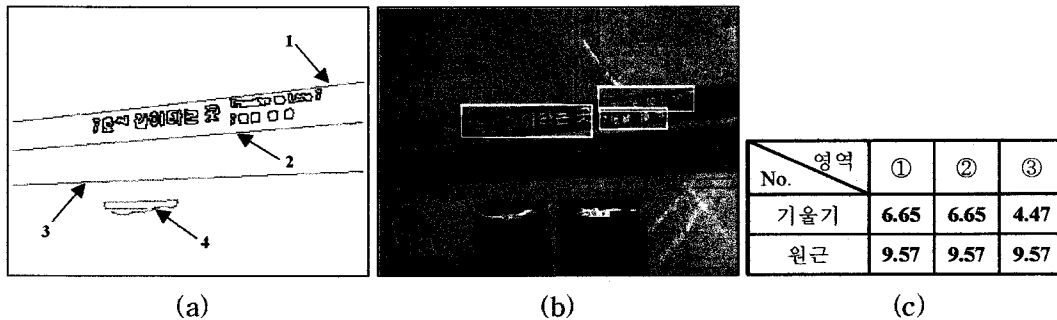


그림 59. 텍스트 영역의 기울어짐 및 원근 추정 결과, (a) 분류된 직선, 추출된 텍스트 영역, (c) 추정된 기울어짐 및 원근

라. 기울어짐과 원근 보정

기울어짐과 원근의 보정은 위에서 추정된 값에 의해 수행된다. 우선 기울어짐은 추정된 기울기 값을 이용하여 Affine 변환 중 Shearing 변환을 수행하여 보정한다. 원근의 경우 대부분이 기울어짐과 원근이 동시에 존재하기 때문에 기울어짐에 대한 보정 방법으로 우선 기울어짐을 보정한 후에 다시 소실점(Vanishing Point)을 이용하여 원근을 보정하는 방법을 적용하였다. 그림 60은 기울어짐과 원근을 보정한 결과 이미지를 보여 준다.

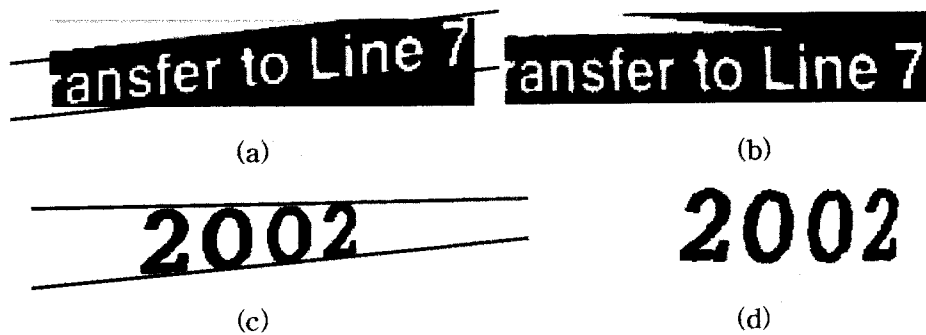


그림 60. 기울어짐과 원근 보정 결과: (a) 기울어진 이미지, (b) 보정된 이미지, (c) 원근을 갖는 이미지, (d) 보정된 이미지

2. 텍스트 분리

추출된 텍스트를 유용한 정보로 변환하기 위해서는 인식이 필요하다. 이진화는 텍스트 인식을 위한 첫 단계로 텍스트와 배경을 분리하는 것을 목표로 한다. 자연이미지에서 추출한 텍스트 영역들은 다양한 색, 크기, 해상도, 폰트 등으로 구성되기 때문에 다양한 환경에서 텍스트를 배경과 효과적으로 분리하기 위한 방법이 필요하다. 기존의 대부분의 연구에서는 텍스트를 분리하기 위하여 명도이미지를 이용한 이진화를 적용하였으나, 배경이 복잡하거나 배경과 텍스트 사이에 색 차이는 있으나 명도이미지로 변환되는 과정에서 이 차이가 손실되어 이진화가 어려워지는 텍스트 영역들도 존재한다. 본 과제에서는 이러한 단점을 보완하기 위하여 색 병합을 이용한 이진화 방법을 제안하고, 기존의 명도이미지를 이용한 전역적 방법과 지역적 방법을 함께 고려한다.

가. 명도이미지에서의 이진화

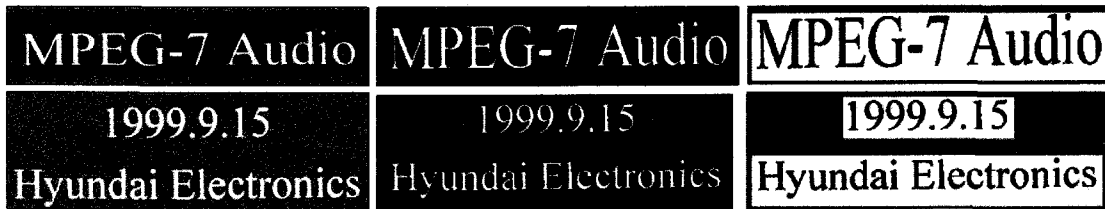
명도이미지에서의 이진화 방법은 크게 전역적인 방법과 지역적인 방법으로 구분한다. 전역적(Global) 이진화는 전경과 배경을 나누는 임계값을 결정하기 위해서 전체 화소에 대한 정보를 이용한다. 지역적(Local) 이진화는 임계값을 확인하기 위해 지역적인 인접 화소들의 정보를 이용한다. 본 시스템에서는 전역적 이진화 방법으로는 Otsu 등이[24] 제안한 방법을 이용하고, 지역적인 이진화 방법으로는 김계경 등이[25] 제안한 방법을 사용한다.

(1) 전역적 이진화

전역적 이진화 방법은 이미지 전체에 대해서 명도 값의 히스토그램을 분석하여 최적의 임계값을 찾아 이진화시키는 방법이다. 전역적 이진화 방법은 하나의 임계값으로 전체 이미지에 대해 적용하기 때문에 처리 시간이 빠른 장점이 있지만 전반적으로 성능이 저하될 수 있는 단점이 있다. 전역적 이진화 방법에서는 히스토그램을 분석하여 임계값을 찾는 방법을 가장 보편적으로 이용하고 있으며, 그 중에서도 Otsu 등이 제안한 이진화 방법이 가장 좋은 성능을 보여주고 있다.

Otsu 등은 이미지에 대한 히스토그램을 분석하여 최적의 임계값을 찾기 위해 텍스트와 배

경을 각 클래스로 가정하고, 두 클래스 사이의 분산값의 차를 최대로 하는 임계값을 찾는 방법을 제안하였다. Otsu 등이 제안한 방법을 본 과제에서는 전체 이미지에 대해서 적용하는 것이 아니라 텍스트 추출 방법에 의해 추출된 영역에 대해서만 적용하기 때문에 전체 이미지에서 임계값을 찾는 것 보다 안전하고, 정확한 임계값을 찾아 이진화함으로써 이진화 성능을 향상시킬 수 있다. 그림 61은 Otsu 방법을 적용하여 이진화한 결과를 보여준다. 그림 61에서 이미지 전체에 분포되어 있는 회색은 시각 효과를 높여주기 위하여 추출된 영역 이외의 영역을 표시한 것이다. 실제 이진화에서는 추출된 영역에 대해서만 관심 있는 객체(본 연구에서는 텍스트)는 검정색으로 나머지는 모두 흰색으로 표현한다. 명도이미지에서의 이진화는 텍스트 영역이 배경보다 밝은 경우 그림 61(b)과 같이 배경이 검정색으로 표현되는 것을 확인할 수 있다. 또한, 그림 61(d)과 같이 배경이 복잡한 영역에 대해서는 이진화 결과가 깨끗하지 못한 문제를 가지고 있다.



(a) 원 이미지 (b) Otsu 이진화 결과 (c) 반전 결과



(d) 배경이 복잡한 경우의 이진화 결과

그림 61. 추출된 영역에 대해 Otsu 방법을 이용한 이진화 결과

위와 같이 전경과 배경이 반전되어 나타나는 경우와 배경이 복잡할 경우 깨끗하게 이진화되지 못하는 두 가지 문제를 해결하기 위해서 본 연구에서는 배경이 텍스트 보다 어두운 명도값으로 분포되어 있을 경우 이진화된 결과를 분석하여 결과를 반전시킨다. 일반적으로 추출된 텍스트 영역의 이진화 결과를 살펴보면 배경이 텍스트보다 많은 영역을 차지하고 있는 특징이 있기 때문에 흰색과 검정색의 화소수를 측정하여 검정색의 화소가 흰색의 화소보다 많을 경우 결과를 반전시킨다. 그리고 배경이 복잡하여 텍스트가 정확하게 이진화되지 않는 문제를 해결하기 위해서 “색 병합을 이용한 이진화”에서 색 정보를 분석하여 텍스트에 해당하는 색을 찾아내는 방법을 제안하여 배경의 복잡도에 상관없이 텍스트가 분명한 색으로 구성되어 있을 경우 이진화시킬 수 있는 방법을 제안한다. 그림 62는 Otsu 방법을 다양한 이미지에 적용한 결과이다.

Shape Descriptors Shape Descriptors

비선형 신호처리연구실 비선형 신호처리연구실



loss

loss

(a) 원 이미지

(b) Otsu 방법 적용 결과

그림 62. 다양한 이미지에 Otsu 이진화 방법 적용

(2) 지역적 이진화

지역적 이진화 방법은 김계경 등이[25] 제안한 방법을 적용한다. 지역적 이진화는 일정 크기의 윈도우를 띄워서 그 영역에 대해서만 고려하기 때문에 보다 정확한 임계값을 찾을 수 있는 장점이 있다. 그러나 윈도우 크기가 작을 경우 처리 시간이 많이 소요되고, 윈도우의 크기가 클 경우에는 전역적 이진화와 마찬가지로 최적의 임계값을 찾는 데 어려움이 있다. 본 연구에서는 김계경 등이 개발한 이진화 방법에서 다양한 크기의 윈도우와 임계값의 가중치를 조절하면서 결과를 비교함으로써 본 연구와 가장 적합한 결과를 찾고자 한다. 그림 63은 윈도우 크기에 따른 결과이고, 그림 64는 최적의 윈도우 크기에 대해서 임계값의 가중치에 따른 결과를 비교한 것이다. 이 때 윈도우는 이진화의 임계값을 구하기 위한 지역적 영역을 의미하며, 가중치는 지역적 영역에서 구해진 임계값에 실제 이미지에서의 이진화 질을 향상시킬 수 있도록 실험적으로 정해진 값을 의미한다. 보통 0~1사이의 값으로 표현된다. 그림 63과 그림 64에서 볼 수 있듯이 대부분의 이미지에서 좋은 성능을 보임을 확인할 수 있었다. 실험을 통해 윈도우 크기와 가중치에 대한 성능을 비교한 결과 본 과제와 가장 적합한 윈도우 크기는 45x45이고, 가중치는 0.8임을 확인하였다.



(a) 원 이미지



(b) 윈도우 크기: 10x10, 가중치: 0.8



(c) 윈도우 크기: 45x45, 가중치: 0.8



(d) 윈도우 크기: 60x60, 가중치: 0.8

그림 63. 윈도우 크기에 따른 결과 비교

Texture Descriptor
in MPEG-7

in MPEG-7

(a) 윈도우 크기: 45x45, 가중치: 0.5

Texture Descriptor
in MPEG-7

in MPEG-7

(c) 윈도우 크기: 45x45, 가중치: 0.7

Texture Descriptor
in MPEG-7

in MPEG-7

(d) 윈도우 크기: 45x45, 가중치: 0.8

scriptor
in MPEG-7

in MPEG-7

(e) 윈도우 크기: 45x45, 가중치: 0.9

그림 64. 가중치 변화에 따른 결과 비교

나. 색 병합을 이용한 이진화

이미지의 배경이 여러 색으로 구성되거나, 배경과 전경의 명도 변화가 크지 않은 경우, 또한 조명 변화가 심한 경우에는 명도 이미지에서의 이진화로서 배경과 텍스트를 분리하는 것은 그림 65와 같이 아주 어려운 일이다. 따라서 본 연구에서는 추출된 영역으로부터 텍스트를 정확히 분리하기 위해서 그림 66과 같이 텍스트를 구성하는 색을 이용한 이진화 방법을 제안한다.



(a) 원 이미지

(b) 명도 이미지에서의 이진화

그림 65. 명도 이미지에서의 이진화 어려움

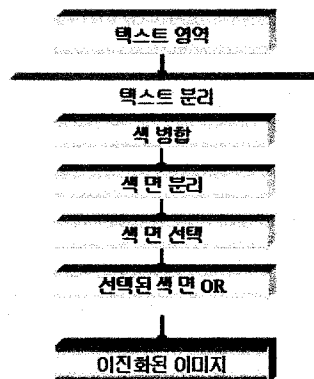


그림 66. 색 병합을 이용한 이진화 방법

(1) 색 병합

추출된 텍스트 영역 이미지의 크기는 다양하며 각 화소는 24비트로 표현된 색 이미지이다. 24비트를 모두 사용한 이미지에서 각 색을 분리하여 처리한다면, 많은 계산 시간을 필요로 하며 효율성이 낮기 때문에 색 줄임 및 병합이 필요하다. 빠르고 간단하게 사용되는 색 줄임 방법으로서 각 색 요소의 *Bit-dropping*을 이용하는 방법이 있다. 그림 67과 같이 추출된 부분 영역에 대해서 *R, G, B* 각 요소의 하위 6비트를 제거하고 상위 2비트만으로 표현된 이미지는 배경이 복잡하거나 배경과 텍스트의 색 차이가 적은 경우 정보의 손실이 많은 단점이 있다. 이러한 단점을 극복하기 위해서 본 연구에서는 추출된 부분 영역에 대한 *K-means* 방법을 이용하는 색 줄임 및 병합 방법을 제안한다.

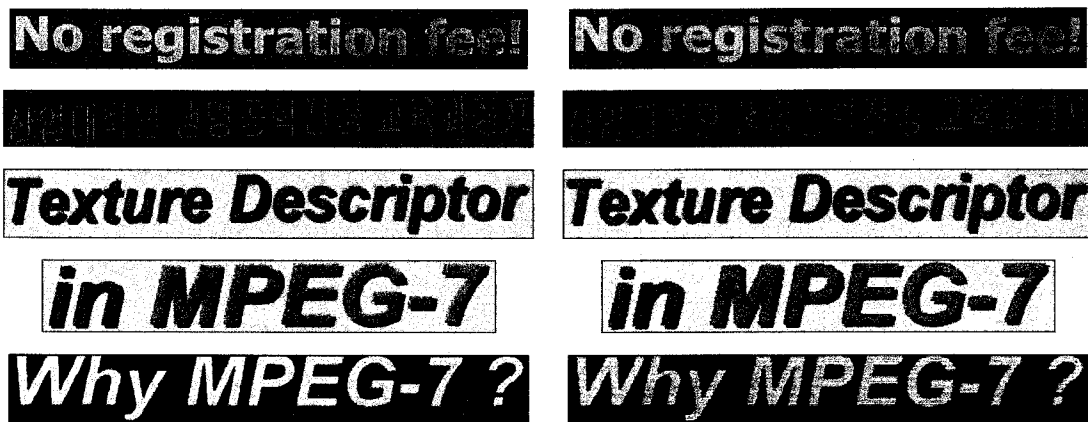


(a) 추출된 텍스트 영역

(b) *Bit-dropping* 색 줄임

그림 67. *Bit-dropping*을 이용한 색 줄임 결과 예

K-means 병합에서 클러스터(Cluster)의 개수인 *K*를 4로 결정하였으며, 이는 다양한 이미지를 대상으로 한 실험으로부터 결정하였다. 또한, 초기 클러스터 중심의 선택은 임의의 색으로 시작한다. 색 병합 과정에서 클러스터의 평균값과 각 색까지의 거리는 절대거리를 이용하였으며, 평균값의 변화가 없거나 평균값의 변화에 대한 임계값 *T*가 일정한 값보다 작을 때까지 반복하면서 병합한다. 그림 68은 추출된 영역에 대한 색 병합 결과로서 각 이미지는 네 개의 색으로 줄어든 결과이다. *Bit-dropping*을 이용한 결과보다 좋은 결과가 만들어짐을 그림에서 확인할 수 있다.



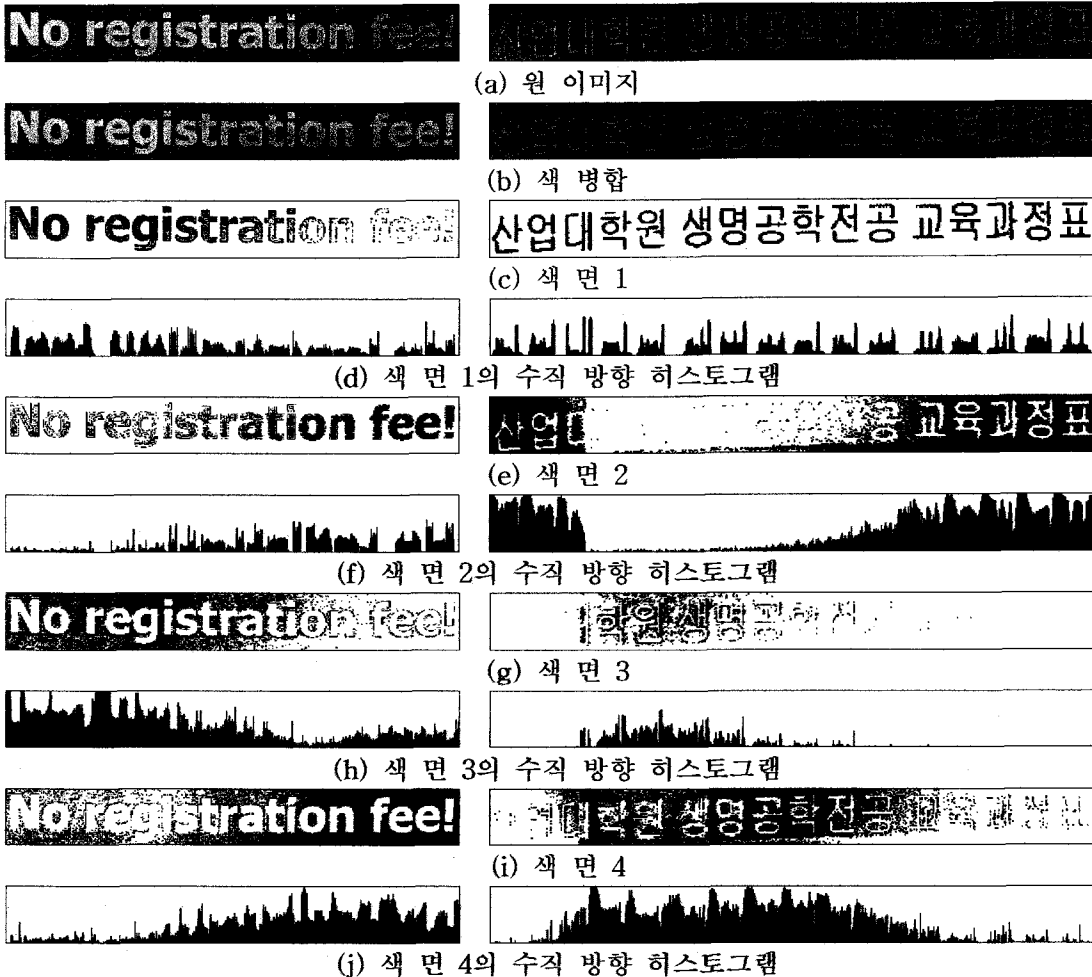
(a) 추출된 텍스트 영역

(b) *K-means* 색 병합 결과

그림 68. *K-means*를 이용한 색 병합 결과

(2) 색 면 선택

*K-means*를 이용한 색 병합으로 구해진 4개의 색 면을 분리한 후, 각 색 면들의 수직 방향 히스토그램 프로파일을 분석한다. 색 병합에 의해서 분리된 각 색 면들은 텍스트, 배경, 잡영, 텍스트 테두리 등의 객체가 포함될 수 있으며, 이로부터 텍스트 색 면을 선별한다. 그림 69는 색 병합한 결과 이미지에 대해서 각 색 면의 수직 히스토그램 프로파일을 보여준다. 텍스트를 포함하는 색 면은 배경에 비해서 프로파일의 빈도수가 작고, 값이 0인 히스토그램 프로파일이 존재하며, 영역의 높이와 같은 프로파일이 존재하지 않는 특징이 있다. 또한, 각 빈의 평균을 기준으로 0에서 1로 변하는 빈도수가 배경에 비해서 높게 나타난다. 이러한 특징을 나타내는 색 면들만을 선별하여 OR된 이진 이미지를 텍스트 색 면으로 결정한다. 그림 69의 배경이 단순한 이미지는 (c), (e)와 같이 색 면 1과 2에서 텍스트의 특징이 존재하여 두 색 면을 합하여 이진화 결과를 구했다. 그림 69의 배경이 복잡한 이미지의 경우 (c)에서만 텍스트의 특징이 존재하여 최종적으로 색 면 1만을 선택하여 이진화 결과를 구했다. 또한, 복잡한 이미지의 경우 그림 69(k)의 색 병합을 이용한 이진화 결과와 그림 69(l)의 명도 이미지에서의 이진화 결과를 비교하면 제안한 방법의 필요성과 우수성을 파악할 수 있다.



No registration fee!

산업대학원 생명공학전공 교육과정표

(k) 색 병합을 이용한 이진화 결과

No registration fee!

대학원 생명공학전공

(l) Otsu 방법을 이용한 이진화 결과

배경 색이 단순한 이미지 배경 색이 복잡한 이미지
그림 69. 색 병합을 이용한 이진화 및 Otsu 방법과의 비교

그림 70은 명도이미지에서 이진화 오류를 제안한 색 병합 및 선택 방법으로 개선된 또 하나의 결과를 보여준다. 그림 70(b)은 추출된 텍스트 영역 이미지에 색 병합을 적용한 결과이며, 70(c)은 원 이미지를 명도이미지로 바꾼 것이다. 그림 70(d), (e), (f), (g)는 색 병합된 결과의 각 색 면들이며, 그림 70(h), (i)는 각각 색 병합 결과에서 선택된 색 면과 Otsu 방법을 적용하여 얻은 이진화 결과이다. 명도이미지에 Otsu 방법을 적용한 그림 70(i)에서 텍스트 일부가 분리되지 못한 결과를 보였는데, 이는 명도이미지에서 배경과 텍스트 사이의 명도 값이 차이가 적어서 임계값의 설정에 어려움이 있었기 때문이다. 그러나 색 병합을 이용한 결과에서는 이러한 오류가 개선된 것을 확인할 수 있다.



(a) 원 이미지



(b) 색 병합 결과



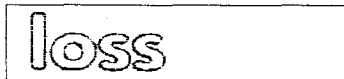
(c) 명도 이미지



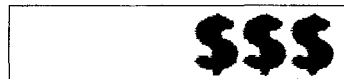
(d) 색 면 1



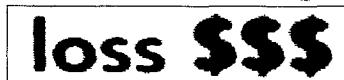
(e) 색 면 2



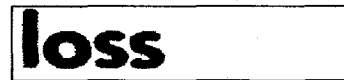
(f) 색 면 3



(g) 색 면 4



(h) 색 병합을 이용한 이진화



(i) Otsu 방법을 이용한 이진화

그림 70. 색 병합 및 Otsu 방법을 이용한 이진화 결과 비교

제 3 절. 응용시스템 구현

1. 개요

현대 사회의 정보화는 인간이 사용하기 편리하고, 효율적인 방향으로 컴퓨터가 응용되어 가고 있으며, 디지털 카메라, 휴대폰 등 다양한 종류의 주변기기 및 통신 장비들도 급속도로 발전하고 있다. 특히 디지털 카메라의 발전으로 인해서 다양한 종류의 이미지를 쉽게 취득할 수 있게 되었으며, 취득한 이미지로부터 사용자에게 보다 편리하고 정확한 정보를 전달하기 위한 연구가 활발하게 진행되고 있다. 본 연구개발 결과의 응용으로서 지능형 정보단말용 카메라를 이용하여 획득한 다양한 종류의 문서이미지에서 텍스트를 추출하고 인식하는 응용시스템을 고려한다. 이를 위해서 발표자료, 도서, 회의 자료, 자연이미지 등 다양한 종류의 문서를 카메라

라를 이용해서 학습 및 실험이미지 데이터베이스를 구축하고, 구축된 실험 이미지에 대해서 텍스트 영역을 추출하고 이진화하는 방법을 앞에서 제안한 방법을 기반으로 개발한다.

학습 데이터베이스는 문서이미지에 존재할 수 있는 다양한 종류의 폰트와 크기를 고려하기 위해서 현재 가장 많이 사용되고 있는 ‘한글 2002’와 ‘MS-워드’ 응용프로그램에서 수집한다. 한글과 비한글로 구분하며, 카메라로 문서이미지를 취득할 때 주변 조명이 인식에 많은 영향을 줄 수 있기 때문에 간접 조명을 사용한다. 실험 데이터베이스는 다양한 종류의 문서이미지에 대해서 최적의 알고리즘을 구현할 수 있도록 다양하게 수집한다. 본 연구에서는 배경이 있는 도서, 각종 학회의 발표자료, 회의록, 자연이미지의 4종류로 분류하고, 3종류의 서로 다른 성능의 카메라를 이용한다.

본 연구에서 제안하는 현장 문서이미지에 대한 전처리 방법은 크게 텍스트 영역 추출과 텍스트 분리로 나눌 수 있다. 텍스트 영역 추출은 본 연구에서 제안한 방법을 이용하고, 텍스트 분리는 전역적 방법과 색 병합을 이용한 분리 방법을 이용한다.

2. 데이터베이스 구축

가. 학습 데이터베이스 구축

본 연구에서는 지능정보 단말용 카메라 문서이미지의 인식을 위하여 문서이미지 획득에 적합한 외부 환경을 마련하여 다중 인쇄 문서이미지의 인식에 적합한 학습 문서이미지 데이터베이스를 구축한다. 본 연구에서 사용한 카메라는 지능정보 단말기에 사용될 130만 화소의 PDA 카메라와 유사한 낮은 성능의 디지털 카메라를 사용함으로써 PDA 카메라의 촬영상의 불편함을 최소화한다.

지능정보 단말용 카메라로부터 최적의 문서이미지를 획득하기 위해서 카메라 고정 장치를 사용하여 기울어짐과 흔들림에 대한 질(質) 저하를 최소화하였고, 실내 간접 조명을 이용함으로써 시간대 별로 변하는 자연광의 영향을 최소화하였다. 다양한 폰트에 대한 정확한 학습을 위해서는 현재 가장 보편적으로 사용되고 있는 ‘한글 2002’와 ‘MS-워드’로 분류하여 촬영하여 같은 폰트에 대해서도 에디터에 따라 다르게 나타나는 문제를 해결하고자 하였다. 또한, 다양한 크기에 대해서도 각각의 크기마다 촬영 거리를 다르게 함으로써 인쇄된 문서의 텍스트 크기에 따라 초점 거리가 달라지는 문제를 해결하였다. 학습용 문서이미지의 촬영 과정에서 하나의 셋(Set)을 구성하기 위해 카메라 한 컷(Cut)에 초점이 가장 잘 맞으면서 최대한 촬영할 수 있는 텍스트의 수를 균일하게 하여 여러 장의 컷으로 구성된 셋에서 촬영하여 지역적으로 명도와 조명의 차이를 없애도록 노력하였다. 그림 71은 균일한 텍스트 수만큼 나누어진 컷 이미지를 보여주며, 이러한 이미지들이 결합하여 하나의 셋 이미지로 구성된다. 약 600 셋의 학습 이미지를 구축하였다.

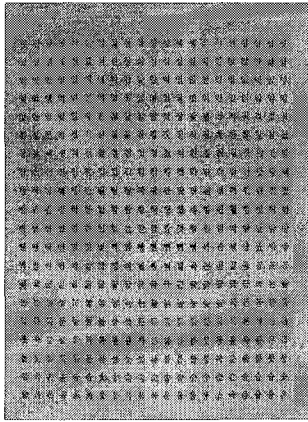
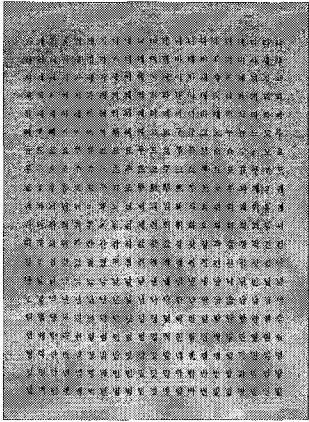
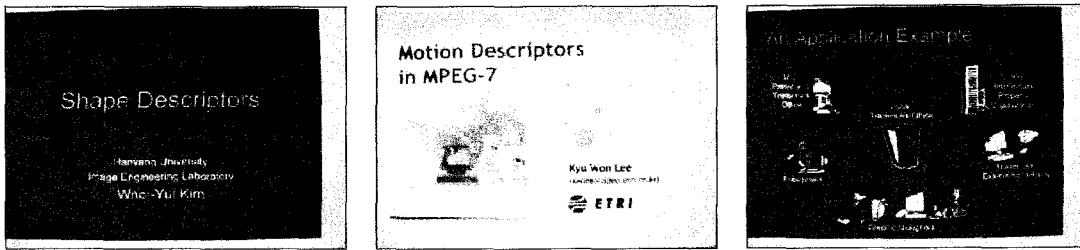
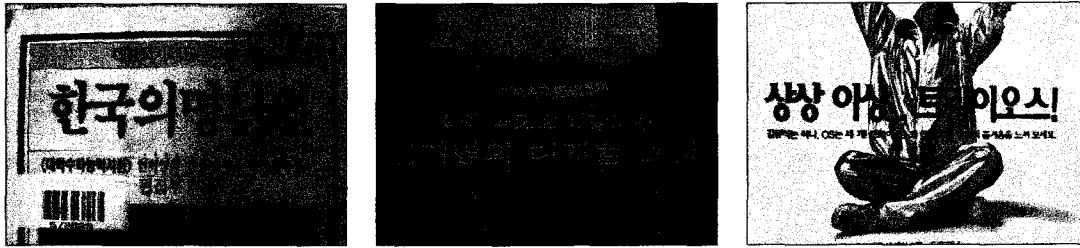


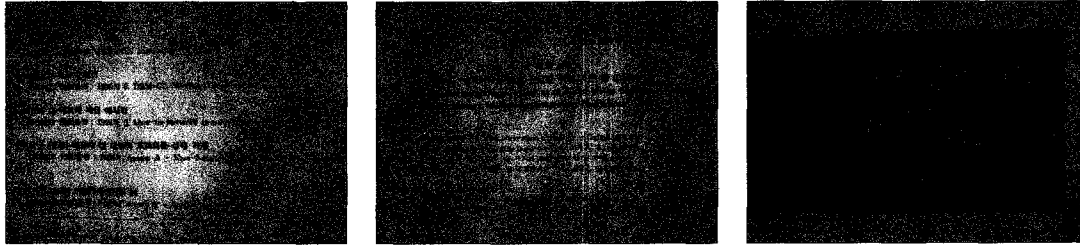
그림 71. 각각의 컷 이미지



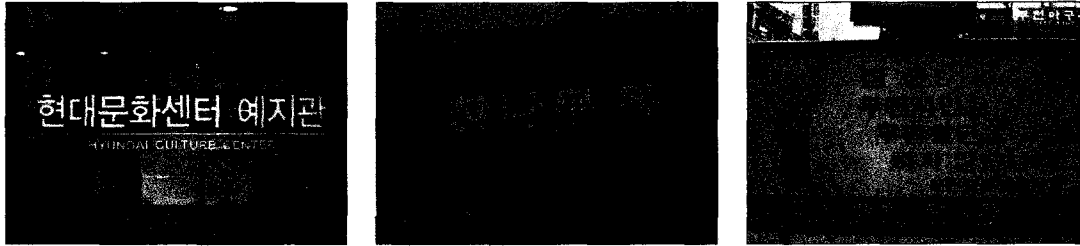
(a) 발표 자료



(b) 도서



(c) 회의 자료



(d) 실내의

그림 72. 실험 데이터의 종류별 분류

3. 영역 추출

현장 문서이미지에서 텍스트 영역을 추출하고 분리하는 방법을 제안한다. 제안하는 방법은 그림 73과 같이 1차년과 2차년 연구를 통해 제안된 계층적인 구조를 이용한다. 제안하는 텍스트 추출 방법은 지역적인 이미지 영역에서의 밝기 및 색 변화가 존재하는지를 파악하여 후보 영역을 추출하며, 추출된 후보영역의 각 부분들이 문자 구성의 기본 단위인 획을 구성할 수 있는가를 검증한다.

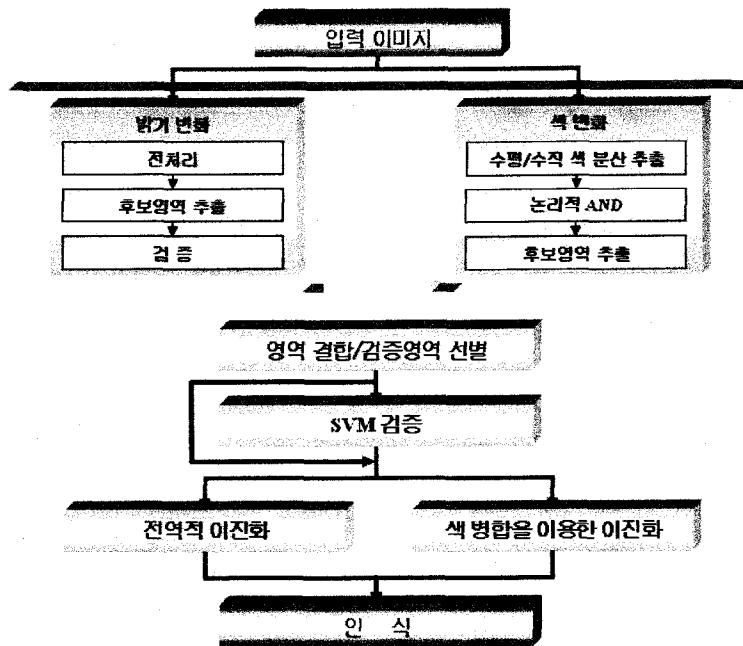


그림 73. 텍스트 영역 추출 방법

밝기 변화를 이용한 추출 방법은 1, 2차년 연구와 동일하게 입력된 이미지에서 에지를 구하고, 에지 분포를 분석하며 이미지에 포함되어 있는 다양한 모양의 긴 선들을 제거한다. 모폴로지를 이용하여 각 연결요소를 강조한 후 연결요소를 분석, 검증하여 후보영역을 결정한다. 색 변화를 이용한 추출은 2차년 연구와 같이 수평 및 수직 방향의 색 분산 이미지를 구하고, 두 결과의 논리적 AND 연산을 수행하여 선 또는 잡영과 같이 한 방향으로만 분산 값이 큰 부분들을 제거한 후 연결요소를 분석, 검증으로 후보영역을 결정한다. 두 특징에서 추출된 후보영역의 외곽사각형의 겹침 정도에 따라 결합하여 후보영역을 선별한다. 선별된 후보영역에서 각 부영역(Sub-region)의 획 특징을 추출하고, SVM 분류기로 검증한다. 이렇게 최종적으로 추출된 영역에 대해서는 간단한 Otsu 이진화 방법과 본 연구에서 제안한 색 병합을 이용한 이진화를 모두 적용한다. 두 이미지 모두를 인식하여 최종 인식결과를 선택한다.

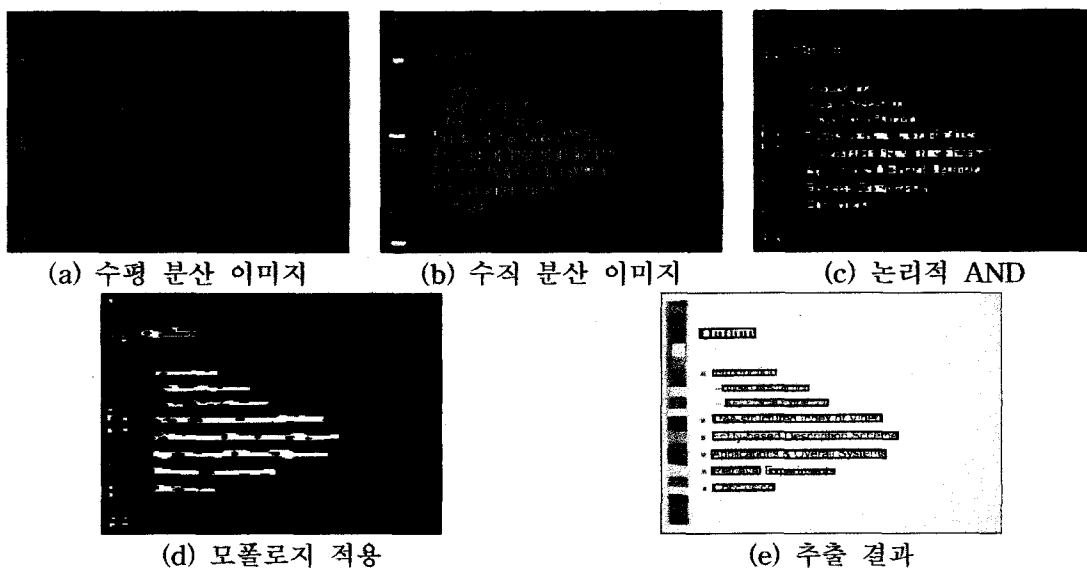
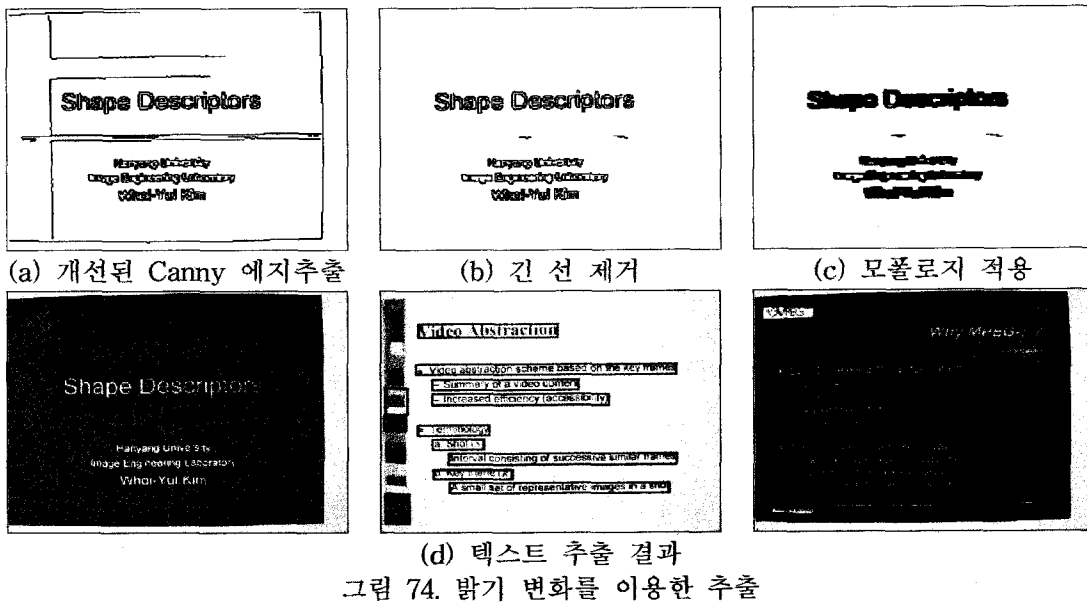
가. 이미지 축소

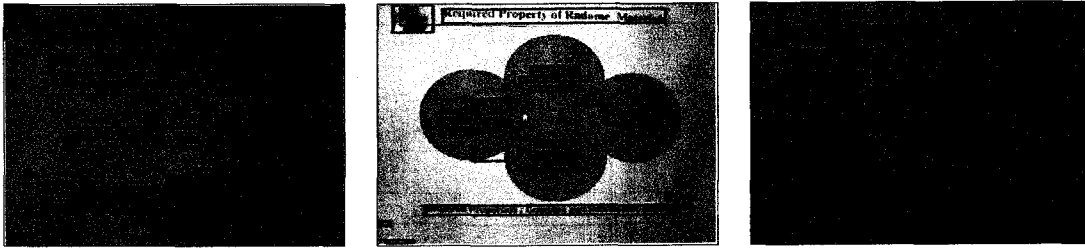
카메라로 획득한 이미지를 처리하기 위해서 다양한 크기로 입력되는 색 이미지를 320x240 크기의 색 이미지로 축소한다. 이는 실시간 응용을 목적으로 하는 본 과제의 특성을 위해 처리 시간을 줄이기 위한 것이다. 이미지 축소는 식 (14)를 이용하여 X축 방향으로의 축소와 Y축 방향으로의 축소를 차례로 수행한다. 이 때 V 는 계산된 화소 값이며 R, G, B 에 대해서 각각 계산한다. $Real$ 는 변환 비율에 따른 입력 이미지에서 고려 대상이 될 화소의 위치 값이며, Int 는 $Real$ 을 정수로 변환한 값이다. Low 는 이미지에서 Int 위치의 화소 값을 나타내며, $High$ 는 그 뒤쪽 화소값을 나타낸다.

$$V = Low + (High - Low) * (Real - Int) \quad (14)$$

나. 밝기 변화 및 색 변화를 이용한 추출

텍스트 영역을 추출하기 위해서 본 연구에서는 1, 2차년 연구에서 색 연속성, 밝기 변화, 색 변화 등 세 종류의 이미지 수준의 특징을 이용하였다. 세 가지의 특징을 이용한 결과 다양한 종류의 자연이미지에서 높은 추출률을 보일 수 있었다. 본 연구에서 제안하는 응용시스템을 구현하기 위해서 대상 이미지인 문서이미지에 대한 추출 성능과 실시간으로 텍스트 영역을 추출하기 위해서 세 가지의 특징 중에서 밝기 변화와 색 변화 특징만을 이용하였다. 문서이미지의 경우 대부분 배경이 자연이미지에 비해 비교적 단순하기 때문에 색 연속성 특징을 제외한 두 특징만으로도 높은 추출률을 보일 수 있었다. 그림 74는 밝기 변화 특징을 이용한 영역 추출의 단계적 결과를 보여준다. 그림 75는 색 변화 특징을 이용한 영역 추출의 단계적 결과를 보여준다.





(g) 기타 추출 결과
그림 75. 색 변화를 이용한 추출

다. 결합 및 SVM 검증

밝기 변화와 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 연구에서는 그림 76과 같이 각 방법에서 찾은 영역의 겹침 정도를 서로 비교해서, 검증으로 확인해야 하는 후보영역과 검증을 생략하고 텍스트 영역으로 최종 확정하는 후보영역으로 구분한다.

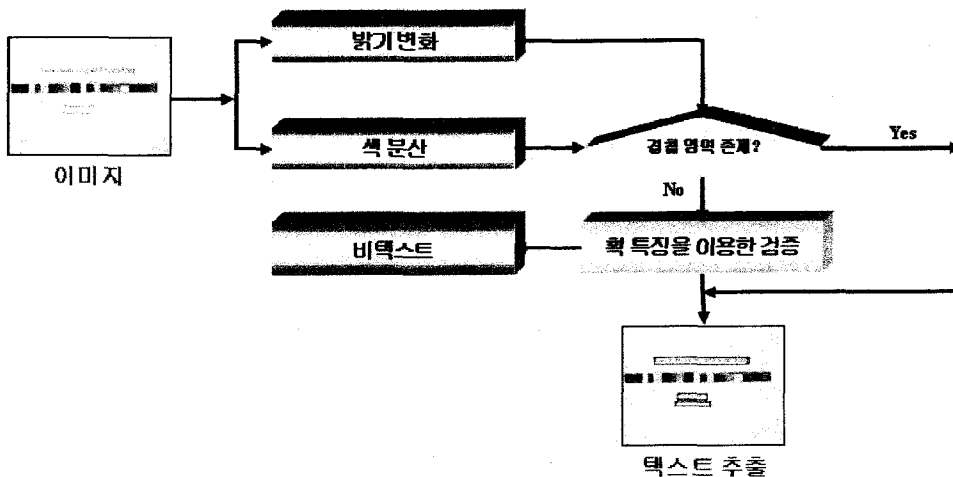


그림 76. 각 특징 결과의 결합

두 특징을 이용한 텍스트 추출 결과를 분석하면 밝기 변화를 이용한 방법은 조명 변화로 인해서 텍스트와 배경 사이에 명도 대비가 크지 않아 에지를 정확하게 찾지 못하고, 배경이 복잡할 경우에는 텍스트 주변에 그림이나 여러 종류의 다양한 그래픽 등과 같이 많은 에지가 존재하여 오류가 발생하는 경우가 많다. 색 변화를 이용한 방법은 기울어진 텍스트와 배경이 복잡한 이미지에 대한 오류가 발생하기 쉽다. 이와 같이 두 방법의 단점을 파악하여 상호 보완적으로 결합한다면 보다 정확한 추출이 가능하다.

결합은 밝기 변화와 색 변화 특징에서 추출된 영역의 겹침 정도를 비교하여 4가지 경우로 나누어 결합하여 검증할 영역을 선별하였다. 이 때 4가지 경우는 일치하는 경우, 전혀 일치하지 않는 경우, 한쪽에 포함되는 경우, 일부 겹치는 경우이다. 첫째, 일치하는 영역의 경우는 상호 겹침 정도가 80% 이상일 경우 일치한다고 판단하여 검증 과정 없이 텍스트 영역으로 확정한다. 둘째, 한쪽에 포함되는 경우는 포함관계에 있는 두 외곽사각형의 높이를 고려하여 만약

포함시키고 있는 외곽사각형의 높이 포함된 외곽사각형의 높이의 2배 이상일 경우에는 포함된 영역을 검증 없이 텍스트 영역으로 확정하였고, 포함시키고 있는 외곽사각형은 검증대상으로 분류하여 SVM 검증과정을 수행한다. 또한, 2배 이하의 경우에는 반대로 포함시키고 있는 외곽사각형을 검증 없이 텍스트 영역으로 확정하고 포함된 영역에 대해서는 검증 대상으로 분류한다. 셋째, 일부 겹치는 영역에 대해서는 두 외곽사각형을 하나의 큰 외곽사각형으로 결합하여 역시 검증 대상으로 분류한다. 넷째, 전혀 일치하지 않는 경우에도 일치하지 않는 모든 외곽사각형을 검증대상으로 분류한다. 이 때 검증은 획 기반 특징으로 학습된 SVM 분류기를 이용하여 텍스트 영역과 비텍스트 영역으로 구분한다.

검증 대상으로 분류된 후보영역에 대해서는 최종적으로 SVM을 이용하여 여러 단위 블록 (16x16) 각각에 대해 텍스트인지 아닌지를 결정하고 그 결과를 통합하여 영역 전체에 대한 결정을 내린다. 본 시스템에서는 1, 2차년과 동일한 방법으로 웨이블릿 변환을 이용하여 특징을 추출하고 선택하여 SVM을 이용하여 특징벡터를 학습시켜 SV구하여 각각의 SV들과의 거리를 계산하여 텍스트 여부를 판단한다. 윈도우 영역이 텍스트로 분류되면 윈도우 안의 모든 화소들을 텍스트로 레이블링하고 오른쪽으로 4 화소 이동한다. 이동한 위치에서도 같은 방법으로 반복한다. 색 정보와 명도 정보를 이용해서 텍스트로 찾아진 모든 영역에 대해서 위와 같이 수행한 후, 텍스트 윈도우와 겹치지 않는 화소들은 텍스트가 아닌 영역으로 레이블링한다. 최종적으로 텍스트로 레이블링 된 영역이 전체 영역의 2/3가 넘으면 그 영역을 텍스트 영역으로 판단하고, 그렇지 않으면 텍스트가 아닌 영역으로 판단한다. 이렇게 검증하는 과정에서 텍스트 영역의 크기가 크거나 작은 경우에는 16x16 크기의 윈도우로는 검증이 잘 안 되는 경우가 발생하는데 이를 해결하기 위해서 본 과제에서는 영역의 폭 또는 높이가 40 화소 이상인 경우에는 이미지 크기를 160x120으로 1/2로 축소된 이미지에서 검증하였고, 영역의 폭 또는 높이가 10 화소 이하인 경우에는 이미지 크기를 640x480으로 두 배 확대된 이미지에서 검증하였다. 그림 77은 밝기 변화와 색 변화를 이용해서 찾아진 영역을 검증한 결과로써 (b)는 검증되어 텍스트라고 확정된 영역을 보여준다.

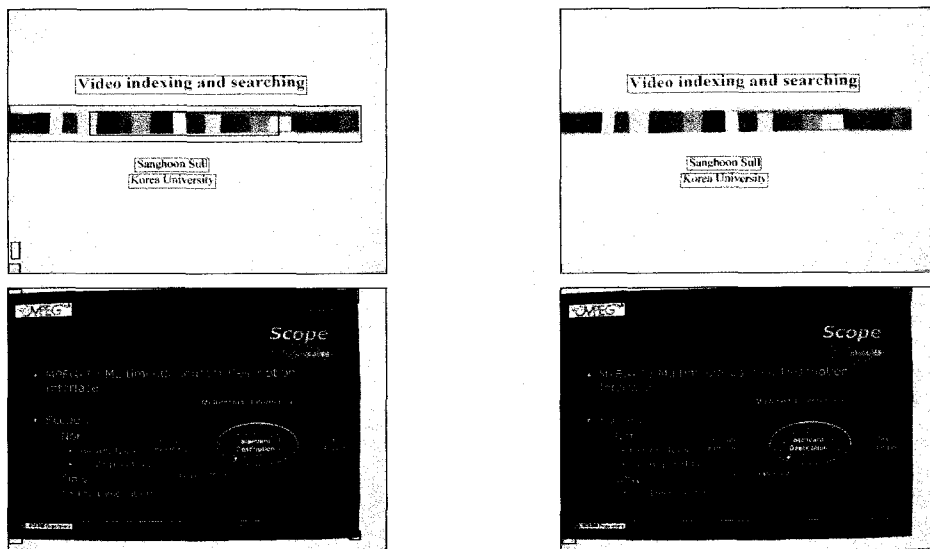


그림 77. 추출된 영역의 결합 결과(좌), SVM 검증 결과(우)

4. 실험 및 결과

본 연구에서 제안한 방법은 Windows XP에서 Visual C++ 6.0을 이용하여 Pentium IV 1.8GHz 하드웨어 상에서 구현하였다. 제안한 텍스트 영역 추출 방법의 평가를 위해서 디지털 카메라를 사용하여 구축된 실험 데이터베이스 중에서 발표 자료에 대해서 120개의 이미지를 선별하여 사용하였다. 실험 이미지에 대해서는 자연이미지에 대한 평가 기준과 동일한 방법으로 배경의 복잡도에 따른 성능을 비교하였다.

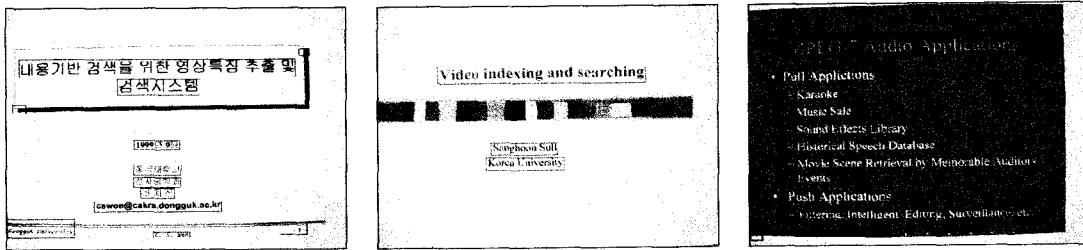
가. 텍스트 영역 추출

표 5는 제안한 방법의 밝기 변화 특징과 색 변화 특징 각각에 대한 결과와 두 특징을 결합한 결과를 보여준다.

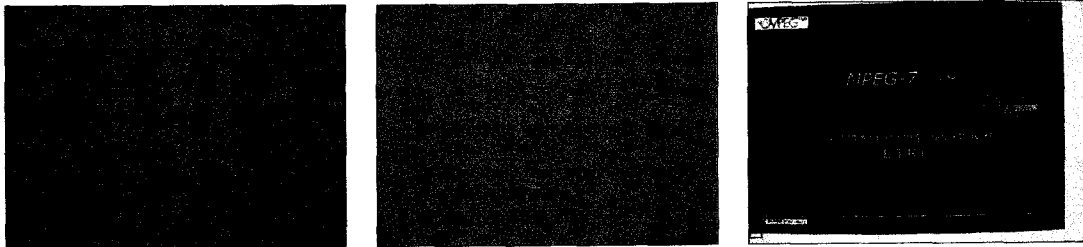
표 5. 실험 이미지에 대한 추출 결과

			<i>Total</i>	<i>Correct</i>	<i>Partial</i>	<i>Missing</i>	<i>False</i>
밝기 변화	단순 이미지	영역 개수	499	453	18	16	112
		정확률 = 77.1%	재현률 = 90.7%				
복잡 이미지	영역 개수	642	578	19	46	189	
		정확률 = 73.2%	재현률 = 90.7%				
색 변화	단순 이미지	영역 개수	499	487	4	8	203
		정확률 = 69.9%	재현률 = 97.6%				
복잡 이미지	영역 개수	642	596	9	37	241	
		정확률 = 69.3%	재현률 = 93.1%				
결 합	단순 이미지	영역 개수	499	482	8	9	84
		정확률 = 83.9%	재현률 = 96.5%				
복잡 이미지	영역 개수	642	601	6	18	98	
		정확률 = 85.2%	재현률 = 93.6%				

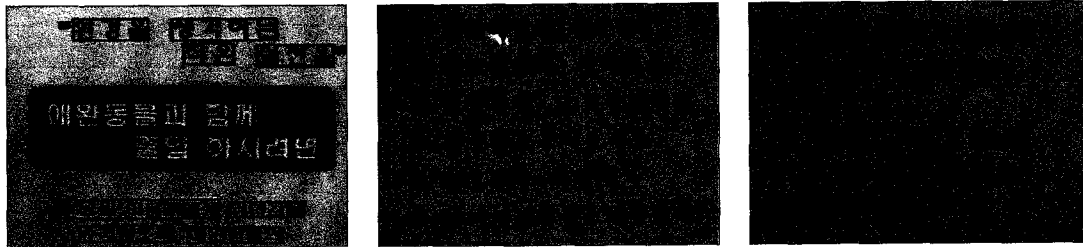
표 5와 같이 본 과제에서 제안한 방법을 실험 이미지에 적용한 결과 단순 이미지 및 복잡 이미지 모두에서 재현률(Recall)이 높은 것을 확인할 수 있다. 밝기 변화를 이용한 방법에서는 선 제거 과정에서 긴 선과 텍스트의 일부가 연결되어 긴 선과 함께 제거되면서 *Partial*의 개수가 색 변화 특징을 이용한 결과보다 상대적으로 높게 나타났다. 이는 선 제거 방법을 수정함으로써 개선이 가능하다. 또한, 색 변화 방법과 비교할 때 *False*가 적게 추출되었다. 이는 색 변화를 사용한 결과에서는 대각선 또는 원과 같은 곡선이 분산 이미지의 AND 연산으로 제거되지 않았기 때문이다. 두 특징을 결합하는 과정에서 SVM 검증을 필요로 하는 영역을 선별하여 검증함으로써 전체적으로 *False*의 개수를 줄여 정확도를 높일 수 있었다. 그림 78은 실험 이미지에 대한 추출 및 검증 결과를 보여준다.



(a) 단순한 이미지에 대한 추출 결과



(b) 복잡한 이미지에 대한 추출 결과

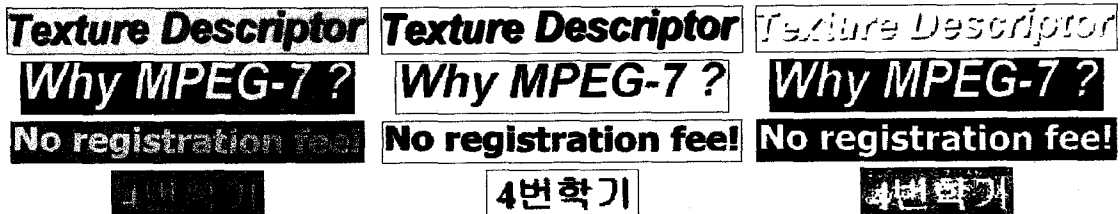


(c) 자연이미지에 대한 추출 결과

그림 78. 문서이미지 및 자연이미지에 대한 추출 결과

나. 텍스트 분리

추출된 텍스트 영역에 대해 *K-means* 색 병합 방법을 적용하여 네 가지의 색 면으로 분리한 후, 화소 수, 히스토그램 프로파일 정보, 히스토그램의 변화 빈도 등을 이용한 색 면 선택으로 텍스트를 분리할 수 있었다. 본 논문에서 제안하는 방법은 조명 변화, 명도 값, 윈도우 크기 등에 대해 민감한 결과를 얻는 명도 이미지에서의 이진화 결과의 단점을 극복할 수 있음을 보여준다. 그림 79는 다양한 추출 영역에 대한 색 병합을 이용한 이진화 결과와 명도 이미지를 이용한 이진화 결과를 비교한 것이다.



(a) 추출된 영역 이미지 (b) 색 병합을 이용한 이진화 (c) 명도이미지의 Otsu 이진화

그림 79. 색 병합 및 Otsu 방법을 이용한 이진화 결과 비교

제 4 절. 결론

본 연구에서는 낮은 수준의 이미지 특징과 높은 수준의 텍스트 특징을 계층적으로 결합하고, 검증하여 자연이미지에 포함된 텍스트 영역을 정확하며 안정적으로 추출할 수 있는 방법을 제안하였다. 이미지 수준의 특징으로는 색 연속성, 밝기 변화, 색 변화 특징을 이용하였으며, 텍스트 특징으로는 웨이블릿 변환을 이용한 획 특징을 사용하였다. 이 과정에서 특징 개발 및 선택, 안정적인 색 병합 방법, 긴 선 추출 및 제거에 의한 이미지 단순화 방법, 색 변화 특징의 개발 및 결합 방법, 획 특징과 SVM을 이용한 효율적인 검증 방법을 제안하였다. 제안한 텍스트 영역 추출 방법을 다양한 종류의 자연이미지에 적용한 결과 배경의 복잡도, 다양한 조명 상태, 폰트 종류 및 방향 등에 비교적 무관하게 높은 재현률을 보여서 응용시스템에의 활용 가능성을 높였다.

텍스트 영역 처리 방법으로서 다양한 환경에서 텍스트를 추출하기 위해서 크기 및 시각 변화에 무관한 추출 방법을 제안하였다. 크기 변화에 대해서는 2단계 해상도 이미지에서 텍스트 영역을 찾아서 결합하였으며, 시각 변화에 대해서는 Template 매칭 및 영역 보정 방법을 이용해서 비선형적으로 변화, 이동하는 영역을 추적하는 방법을 제안하였다. 조명 변화에 대해서는 연세대 정찬섭 교수팀에서 개발한 Color constancy 방법을 적용하였다. 제안한 영역 처리 방법을 적용함으로써 심한 조명 변화와 크기 변화가 존재하는 경우에도 텍스트 영역을 추출의 재현률과 정확률을 높일 수 있었다.

본 연구에서 제안한 방법들은 로봇 자동주행 및 감시시스템, 시각 장애인 보행 안내시스템, 공장 자동화, 자동 감시, 이미지/비디오 색인 및 검색 등에 다양하게 활용될 수 있다. 특히 본 연구에서는 한국전자통신연구원에서 개발하고 있는 지능형 단말기의 영상인식시스템에 본 연구에서 개발한 텍스트 추출 모듈을 성공적으로 탑재하였다. 지능정보 단말용 카메라로 입력한 다양한 종류의 문서이미지에 대한 학습 및 실험 데이터베이스를 구축하였으며, 발표자료, 도서, 회의 자료, 자연이미지에서 텍스트 영역을 추출하고 분리하는 방법을 개발하였다.

제 4 장. 목표달성도 및 관련분야에의 기여도

연 차	연구 내용	달성도	연구 결과 (요약)
1차년	색 정보를 이용한 텍스트 영역 추출	40%	<ul style="list-style-type: none"> - 화소 병합, 색 줄임, 보색 정보를 이용한 색 병합 방법 개발 - 연결요소/외곽사각형의 단계적 제거, 결합에 의한 후보 영역 추출 및 검증 방법 개발 - 정확성: 단순 이미지 Precision 58%, Recall 76%, 복잡 이미지 Precision 64%, Recall 71% - 처리 속도: 평균 173.6 msec/이미지
	명도 정보를 이용한 텍스트 영역 추출	40%	<ul style="list-style-type: none"> - 긴 선 추출 및 제거, 반복적 RLS를 이용한 영역 추출, 검증 방법 개발 - 정확성: 단순 이미지 Precision 56%, Recall 88%, 복잡 이미지 Precision 48%, Recall 88% - 처리 속도: 평균 834.6 msec/이미지
	실험 및 평가 이미지 구축	20%	<ul style="list-style-type: none"> - 다양한 종류의 실내외 자연이미지 수집, 복잡도에 따른 이미지 분류
2차년	정보 선별 방법 개발	50%	<ul style="list-style-type: none"> - 색 병합 방법의 개선으로 안정적인 색 병합 결과를 얻었으며, 영역 추출의 정확성 향상 - 명도 이미지의 선 처리, 후보 영역의 강조 및 결합 방법의 개선으로 수직 문자열 추출 가능, 추출 정확성 향상 - 색 변화를 고려한 색 분산 특징의 개발로 영역 추출의 정확성 및 신뢰성 향상 - 색 연속성, 색 분산, 밝기 변화 특징들의 결합 방법의 개발로 정보 선별 능력 향상 - SVM 및 다해상도 웨이블릿 특징을 이용한 검증 방법의 개발로 영역 추출 정확성 향상
	추출 성능 개선	40%	<ul style="list-style-type: none"> - 텍스트 주변의 긴 선들의 위치 및 방향 정보를 이용한 영역의 기울어짐 및 원근 보정 방법 개발 - 단계적 이미지 개선 방법 개발: 이미지 해상도 확대, 필터링에 의한 잡영 제거, 이진화 방법의 개발로 텍스트 영역의 이미지 품질을 인식 가능한 수준으로 향상 - Canny 에지 추출 시간의 개선: Non-maximum suppression 과정을 개선하여 에지 추출 시간을 1/5에서 1/10 이상 감축시켰으며, 에지 이미지의 수준은 Gaussian 필터의 편차 조절로 유지시킴
	평가 데이터 보완	10%	<ul style="list-style-type: none"> - 다양한 종류의 간판 이미지 추가, ICDAR 2003 학회의 텍스트 추출 경연대회에서 사용하는 이미지를 평가 데이터로 보완

연 차	연 구 내 용	달성도	연 구 결 과 (요약)
3차년	극한 환경에서의 텍스트 정보 추출	50%	<ul style="list-style-type: none"> - 크기 변화에 무관한 추출: 이미지 레벨 및 특징 레벨에서의 다해상도 처리를 이용하여 크기 변화에 무관한 영역 추출 방법 제안 - 시각 변화에 무관한 추출: Template 매칭 및 영역 보정 방법을 이용해서 시각 변화에 따라 비선형적으로 변화, 이동하는 영역 추적 방법의 개발 - 조명 변화에 무관한 추출: 색 변화 특징을 이용한 추출 방법 개발, 연세대에서 개발한 Color constancy 방법의 적용
	응용시스템 구현	50%	<ul style="list-style-type: none"> - 지능정보단말용 카메라로 취득한 색 문서이미지 인식을 위한 전처리 방법 개발 및 적용 - 광범위한 학습 및 실험 DB 구축

제 5 장. 연구개발결과의 활용계획

본 연구에서는 다양한 종류의 자연이미지에 포함되어 있는 텍스트 영역을 추출하는 방법을 제안하였다. 이 과정에서 이미지 수준의 색 연속성, 밝기 변화 및 색 변화 특징을 함께 이용하여 텍스트 후보 영역을 추출하여 각각의 특징에서 추출된 결과를 상호 보완해 주는 장점도 있지만, 처리 시간이 길어지는 단점도 있다. 추후 연구에서는 선택적 또는 순차적으로 특징을 적용하는 방법을 개발하여 추출의 정확성을 향상시키고 처리 시간을 단축시키고자 한다. 또한, 추출된 텍스트 영역의 색 병합 과정에서 현재의 색 화소의 히스토그램과 RGB 색 공간에서의 거리뿐만 아니라 화소간의 공간적인 거리를 함께 고려할 수 있도록 색 병합 방법을 개선하고자 한다.

본 연구에서 개발한 방법은 이미지 내의 텍스트 영역을 찾는 것으로 제한하고 있지만, 국부 영역에서의 보다 안정적인 색 병합 방법으로 개선한다면 이미지 내에 존재하는 다양한 종류의 객체 추출 및 추적에 활용할 수 있다. 특히 PDA 또는 핸드폰 카메라로 취득한 근거리 얼굴 이미지에서의 정확한 얼굴 영역 추출에 활용할 수 있다. 이러한 연구는 추후 응용시스템의 개발과 함께 진행할 예정이다. 또한, 본 연구에서 개발한 자연이미지에서의 텍스트 영역 추출은 자동차 번호판 인식을 이용한 물류 시스템의 효율적 운영과 무인 주차 시스템 관리, 공연장 또는 공원과 같은 대중 장소에서의 안내판 자동 번역 등에 활용될 수 있다.

제 6 장. 연구개발과정에서 수집한 해외과학기술정보

자연이미지에서 텍스트 영역 추출 및 인식에 관한 연구가 활발하게 진행되고 있다[26]. 이는 컴퓨터의 발전과 더불어 디지털 카메라가 급속도로 보급되면서 기존의 스캐너 기반의 이미지 획득에서 카메라 기반의 이미지 획득으로 대체됨에 따라 보다 쉽고, 편리하게 다양한 종류의 이미지를 획득할 수 있기 때문이다. 또한, 이미지 장치의 변화로 인해 텍스트 추출과 인식의 연구 방향 또한 스캐너를 통한 문서이미지로부터 카메라를 통한 문서이미지, 자연이미지, 웹이미지 등의 인식 연구로 확대되고 있다. 그러나 스캐너 기반의 이미지에 비해 카메라 기반의 이미지는 해상도, 조명, 원근, 렌즈 왜곡, 복잡한 배경, 줌과 초점(Zooming and Focusing), 센서 잡영(Sensor noise) 등에 의한 다양한 문제점을 가지고 있다. 본 연구에서는 해상도, 조명 변화, 원근, 복잡한 배경 등과 같은 문제점에 대해서는 어느 정도의 결과를 만들었지만, 나머지 다양한 문제점들에 대한 추가 연구가 필요하다.

2003년 ICDAR(International Conference on Document Analysis and Recognition)에서는 텍스트 영역 추출 경연대회가 개최되었다[23]. 116개의 학습 자연이미지와 251개의 실험 이미지에 대한 데이터베이스를 구성하여 5개 팀이 참가하였다. 경연대회의 평가 기준으로 본 연구에서와 같이 정확률(Precision), 재현률(Recall) 및 평균 속도를 측정하였으며, 5개 팀이 참가한 결과 아직 응용시스템으로의 적용에는 수준이 못 미치는 것으로 평가되었다. 본 연구팀이 개발한 텍스트 추출 모듈이 경연대회에 참가하지는 못했지만, 동일한 데이터를 사용하여 비교했을 때 경쟁력이 있다고 판단하여 2005년 ICDAR 경연대회에 참가하여 객관적인 평가를 받고자 추후 계속 연구하고자 한다.

제 7 장 . 참고문헌

- [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 추계학술대회 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] 최미화, 김희승, "자연영상에서 문자열 추출," 한국정보과학회 추계학술발표논문집, Vol. 27, pp. 389-391, 2001.
- [9] 박종현, 오일석, "클러스터링에 기반한 자연 영상의 텍스트 영역 분할," 제16회 영상처리 및 이해에 관한 워크샵 발표논문집, pp. 261-266, 2004.
- [10] 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.
- [11] 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.
- [12] 이성환, "시각장애인을 위한 착용형 컴퓨터 시각 기술의 현황 및 전망," 제13회 영상처리 및 이해에 관한 워크샵 발표논문집, pp. 3-8, 2001.
- [13] C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D. Petkovic and W.

- Equitz, "Efficient and effective querying by image content," *Journal of Intelligence Information Systems* Vol. 3, pp. 231-262, 1994.
- [14] G. Cortelazzo, G. A. Mian, G. Vezzi and P. Zamperoni, "Trademark shapes description by string-matching techniques," *Pattern Recognition*, Vol. 27, pp. 1005-1018, 1994.
- [15] B. Holt and L. Hartwick, "Visual image retrieval for applications in art and history," *Proceeding of SPIE* 2181, pp. 70-81, 1984.
- [16] Anil K. Jain, Bin Yu, "Automatic Text Location in Images and Video Frames," *Pattern Recognition*, Vol. 31, No. 12, pp. 2055-2076, 1998.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] John Canny, "A Computational Approach to Edge Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI 8(6), pp. 679-698, 1986.
- [21] C. Cortes, V. Vapnik, "Support Vector Networks," *In Proceedings of Machine Learning*, Vol. 20, pp. 273-297, 1995.
- [22] 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
- [23] S. M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong and R. Young, "ICDAR 2003 Robust Reading Competition," *ICDAR 2003* Vol. 2, pp.682-687, 2003.
- [24] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Systems, Man, and Cybernetics*, Vol. SMC-9, No. 1, pp. 62-66, 1979
- [25] 김계경, 지수영, 정연구, 박상규, "조명이 적은 카메라기반 문서 영상 인식시스템," *컴퓨터비전 및 패턴인식 연구회 워크샵*, pp.90-92, 2002.
- [26] David Doermann, Jian Liang, Huiping Li, "Progress in Camera-Based Document Image Analysis," *ICDAR 2003* Vol. 1, pp.606-616, 2003.

특정연구개발사업 연구결과 활용계획서

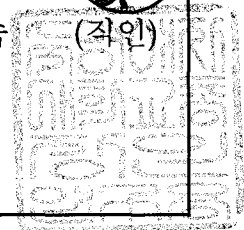
사업명	중사업명	특정연구개발사업(국책연구개발사업)		
	세부사업명	뇌신경정보학연구사업		
과제명	이미지 및 동영상에 포함된 텍스트 정보 추출에 기반한 지능적 시각 응용시스템 구현에 관한 연구			
연구기관	숙명여자대학교	연구책임자	최 영 우	
총연구기간	2001년 8월 1일 ~ 2004년 5월 31일 (34개월)			
총 연구비 (단위 : 천원)	정부출연금	민간부담금	합계	
	128,600		128,600	
기술분야	정보산업분야, 컴퓨터기술, 응용소프트웨어기술			
참여기업				
공동연구기관				
위탁연구기관				
연구결과활용 (해당항목에(√) 표시)	1. 기업화 ()	2. 기술이전(√)	3. 후속연구추진(√)	4. 타사업에 활용()
	5. 선행 및 기초연구(√)	6. 기타목적활용(교육연구)()	7. 활용중단(미활용)()	8. 기타()

특정연구개발사업 처리규정 제 31조(연구개발결과의 보고) 제 2항에 의거 연구결과 활용계획서를 제출합니다.

- 첨부 : 1. 연구결과 활용계획서 1부.
2. 기술요약서 1부

2004 년 5 월 31 일

연구책임자 : 최 영 우
연구기관장 : 이 경 숙



과학기술부장관 귀하

[첨부1]

연구결과 활용계획서

1. 연구목표 및 내용

본 연구의 목표는 자연이미지에 포함되어 있는 장면 텍스트 추출 및 인식으로서 폰트의 종류, 위치, 방향, 크기, 색, 배경 등에 비교적 무관한 추출 방법을 제안하고, 사용자의 불편을 최소화시킬 수 있는 수준의 속도 및 정확성으로 응용시스템에 활용하는 것이다. 따라서 1차년 연구에서는 영상처리 방법을 이용한 텍스트 영역 추출 방법을 제안하고, 2차년 연구에서는 인간의 시각 기능을 모방한 정보 결합 및 검증 방법을 개발하여 추출의 정확성을 향상시킨다. 3차년 연구에서는 극한 환경에서 취득한 이미지에 포함되어 있는 텍스트 영역의 추출로서 텍스트의 크기 변화, 조명 변화 및 시각 변화에 무관한 추출 방법을 개발하여 응용시스템에 적용한다.

2. 연구수행결과 현황

가. 특허(실용신안) 등 자료목록

발명명칭	특허공고번호 출원(등록)번호	공고일자 출원(등록)일자	발명자 (출원인)	출원국	비고
디지털 영상 신호의 텍스트 영역 추출 방법	P01-060925	2001. 9. 29 출원	최영우외 3인	한국, 미국	
디지털 영상 신호에서 추출된 텍스트 영역 이미지 향상 방법	P01-060926	2001. 9. 29 출원	최영우외 3인	한국, 미국	

나. 프로그램 등록목록

해당사항 없음.

다. 노하우 내역

해당사항 없음.

라. 발생품 및 시작품 내역

해당사항 없음.

다. 논문게재 및 발표 실적

○ 논문게재 실적

학술지 명칭	제목	게재연월일	호	발행기관	국명	SCI게재 여부
한국정보과학회 논문지	영화비디오 자막 추출 및 추출된 자막 이미지 향상 방법	2002년 4월	제29권 제4호	한국정보과학회	한국	
LNCS	Scene Text Extracion Using Color and Gray Information	2002년 8월	Vol. 2423	Springer	미국	○
LNCS	ext Extraction in Digital News Video Using Morphology	2002년 8월	Vol. 2423	Springer	미국	○
한국정보과학회 논문지	계층적 특징 결합 및 검증을 이용한 자연 이미지에서의 장면 텍스트 추출	2004년 4월	제31권 제4호	한국정보과학회	한국	
계 : 4건						

○ 학술회의 발표 실적

학술회의 명칭	제목	게재연월일	호	발행기관	국명
한국정보과학회 추계워크샵	명도이미지에서의 장면 텍스트 추출	2001년 11월		한국정보과학회	한국
한국정보과학회 추계워크샵	자연영상에 포함된 비디오 프레임에서의 장면 텍스트 추출	2001년 11월		한국정보과학회	한국
14회 영상처리 및 이해에 관한 워크샵	색상 및 명도 정보를 이용한 장면 텍스트 추출	2002년 1월		한국정보과학회	한국
SPIE Visual Communications and Image Processing 2002	Automatic Text Extraction in News Images using Morphology	2002년 1월		SPIE	미국
한국정보과학회 추계워크샵	장면 텍스트 추출 및 기울어짐/원근 보정	2002년 11월		한국정보과학회	한국
한국 뇌학술회의	이미지 및 동영상에 포함된 텍스트 정보 추출에 기반한 지능적 시각 응용시스템 구현	2003년 5월		한국뇌학회	한국
대한전자공학회 하계종합학술대회	색 변화 특징을 이용한 자연이미지에서의 장면 텍스트 추출	2003년 7월		대한전자공학회	한국
2nd KAIST-Tsinghua Joint Workshop on Pattern Recognition	Scene Text Extraction in Natural Images using Hierarchical Feature Combining and Verification	2003년 10월		한국과학기술원 / 청화대학교	한국
5th China-India-Japan-Korea Joint Workshop on Neurobiology and Neuroinformatics	Scene Text Extraction in Camera-captured Images	2003년 11월		한국뇌과학연구센터	한국

16회 영상처리 및 이해에 관한 워크샵	카메라로 취득한 이미지에서의 색 정보를 이용한 텍스트 영역 추출 및 텍스트 분리	2004년 1월		한국정보과학회	한국
16회 영상처리 및 이해에 관한 워크샵	다해상도 웨이블릿 변환과 써포트 벡터 머신을 이용한 자연영상에서의 문자 영역 검증	2004년 1월		한국정보과학회	한국
International Conference on Pattern Recognition	Scene Text Extraction in Natural Images using Hierarchical Feature Combining and Verification	2004년 8월		International Association of Pattern Recognition	영국
계: 12 건					

3. 연구성과

자연이미지에서의 안정적인 텍스트 영역 추출 방법의 개발과 이를 지능정보 단말용 응용시스템에 적용한 것으로 연구 성과를 요약할 수 있다. 자연이미지에서의 안정적인 텍스트 영역 추출은 대부분의 연구실 개발 결과와는 다르게 거의 제약이 없는 다양한 종류의 자연이미지에 대해서 폰트의 종류, 크기, 배경 등에 무관한 안정적인 영역 추출을 개발한 것으로서 기술적인 의미가 크다. 또한, 응용시스템 개발 및 활용으로서 한국전자통신연구원의 지능정보 단말용 영상인식시스템을 위한 텍스트 영역의 추출 및 분리 전처리 모듈을 개발하여 성공적으로 탑재하였으며, 이 결과는 본 연구개발 결과가 제한된 환경에서만 적용되는 결과가 아님을 입증한 것이다. 응용시스템의 성공을 위해서 고성능 및 저성능 디지털 카메라와 PDA 카메라를 각각 이용하여 수 천 장의 테스트 이미지 DB를 구축하였으며, 이미지내의 문자를 인식하는데 필요한 1,084자의 한글 및 비한글 학습 셋을 정의하여, 600셋 이상의 문자 DB 이미지를 구축한 것도 성공적인 연구 성과이다.

4. 기술이전 및 연구결과 활용계획

가. 당해연도 활용계획(6하원칙에 따라 구체적으로 작성)

2003년 12월 및 2004년 12월 두 차례에 걸쳐서 개발한 방법을 카메라로 취득한 색 문서이미지에서의 텍스트 영역 추출 및 분리 방법으로 개선하여 한국전자통신연구원에서 개발하고 있는 지능형 서비스 로봇 및 지능형 단말기의 영상 인식시스템에 활용하고 있다.

나. 활용방법

본 연구에서 개발한 자연이미지에서의 텍스트 영역 추출 방법을 우선 카메라로 취득한 색 문서이미지에서의 텍스트 추출 및 텍스트 분리 방법으로 개선하여 활용하고 있다. 또한, 다양한 해상도를 갖는 디지털 카메라, PDA 카메라, 핸드폰 카메라 등을 이용하여 광범위한 색 이미지 데이터베이스를 구축하고 제안한 방법을 개선하여 다양한 응용시스템에 활용하고자 한다.

다. 차년도이후 활용계획

본 연구에서 개발한 방법은 이미지에서 텍스트 영역을 찾는 것으로 제한하고 있지만, 국부 영역에서의 보다 안정적인 색 병합을 개발한다면 이미지 내에 존재하는 다양한 종류의 객체 추출 및 추적에 활용할 수 있다. 특히 PDA 또는 핸드폰 카메라로 취득한 근거리 얼굴 이미지에서의 정확한 얼굴 영역 추출에 활용할 수 있다. 이러한 연구는 추후 응용시스템의 개발과 함께 진행할 예정이다. 또한, 본 연구에서 개발한 자연이미지에서의 텍스트 영역 추출은 자동차 번호판 인식을 이용한 물류 시스템의 효율적 운영과 무인 주차 시스템 관리, 공연장 또는 공원과 같은 대중 장소에서의 안내판 자동 번역 등에 활용될 수 있다.

5. 기대효과

본 연구개발은 텍스트 정보 추출의 가장 어려운 문제인 실내외에서 자연스럽게 얻어진 이미지와 동영상에 자유롭게 위치한 장면 텍스트 정보를 추출한 것으로서 기술적인 의미가 크다. 이 기술은 컴퓨터비전 관련 연구의 핵심 요소 기술로서 다양한 분야의 자동 인식, 감시, 주행 등에 사용할 수 있다.

당해 기술의 성공적인 개발은 로봇 자동 주행 및 감시시스템, 시각장애인 보행 안내시스템과 같은 지능적 시각 응용시스템의 핵심 기술로 사용될 것이다. 또한, 다양한 분야의 공장 자동화, 이미지/비디오 색인 및 검색, 자동 번역, 웹 검색 등에 활용될 수 있기 때문에 사회 경제적인 면에서 중요하다.

6. 문제점 및 건의사항

해당사항 없음.

[첨부2]

기술 요약서

■ 기술의 명칭

자연이미지에 포함된 텍스트 영역 추출 기술

■ 기술을 도출한 과제현황

과제관리번호	M1-0102-00-0009			
과제명	이미지 및 동영상에 포함된 텍스트 정보 추출에 기반한 지능적 시각 응용시스템 구현에 관한 연구			
사업명	특정연구개발사업(국책연구개발사업)			
세부사업명	뇌신경정보학연구사업			
연구기관	숙명여자대학교	기관유형	學	
참여기관(기업)				
총연구기간	2001년 8월 1일 ~ 2004년 5월 31일 (34개월)			
총연구비	정부(128,600)천원 민간()천원 합계(128,600)천원			
연구책임자 1	성명	최 영 우	주민번호	
	근무기관 부서	숙명여자대학교	E-mail	ywchoi@sookmyung.ac.kr
	직위/직급	부교수	전화번호	(02) 710-9763
연구책임자 2	성명		주민번호	
	근무기관 부서		E-mail	
	직위/직급		전화번호	
실무연락책임자	성명	최 영 우	소속/부서	숙명여자대학교/컴퓨터과 학과
	직위/직급	부교수	E-mail	ywchoi@sookmyung.ac.kr
	전화번호	(02) 710-9763	FAX	(02) 710-9296
	주소	(140-742) 서울 용산구 청파동 2가 53-12 숙명여자대학교 컴퓨터과학과		

■ 기술의 주요내용

[기술의 개요]

본 연구는 자연이미지에 포함되어 있는 장면 텍스트 추출 기술의 개발로서 텍스트 폰트의 종류, 위치, 방향, 크기, 색, 배경 등에 비교적 무관한 추출 방법을 제안하고, 사용자의 불편을 최소화시킬 수 있는 수준의 속도 및 정확성으로 응용시스템에 활용하는 것이다. 제안한 방법은 이미지의 색 정보와 명도 정보를 함께 이용하여 영역을 추출하고, 텍스트의 획 정보를 SVM으로 학습하여 추출된 영역을 검증한다. 다양한 종류의 자연이미지로 제안한 방법을 실험한 결과 배경의 복잡도에 비교적 무관한 높은 추출률을 확인할 수 있었다. 또한, 제안한 방법을 지능형 단말기의 영상인식시스템에 탑재하여 지능정보 단말용 카메라로 수집한 발표자료, 도서, 회의 자료, 안내판 등의 문서 유형의 이미지에 대해서도 좋은 결과를 보임을 확인할 수 있었다.

<기술적 특징>

- (1) 낮은 수준의 이미지 특징과 높은 수준의 텍스트 특징을 계층적으로 결합한 텍스트 영역 추출 방법의 개발: 낮은 수준의 이미지 특징으로는 색 연속성, 밝기 변화 및 색 변화 특징을 사용하여 영역을 추출하였으며, 웨이블릿을 이용한 높은 수준의 문자 획 특징을 사용하여 검증하는 방법을 제안하였다. 두 종류의 특징을 계층적으로 결합하고 SVM으로 검증하는 방법을 함께 제안하였다.
- (2) 크기, 시각 및 조명 변화에 강인한 추출 방법의 제안: 다해상도 이미지 처리를 이용한 크기 변화를 고려한 추출 방법의 개발, 템플릿 매칭 방법을 이용한 시각 변화를 고려한 추적 방법의 개발 및 Color constancy를 이용한 조명 변화에 강인한 추출 방법을 함께 제안하였다.
- (3) 응용시스템 적용: 다양한 종류의 카메라로 취득한 문서 유형의 이미지에서의 텍스트 추출로 제안한 방법을 개선하여 지능형 정보단말기에 사용이 가능하도록 하였으며, 텍스트 영역으로부터 텍스트만을 분리하기 위한 색 정보를 이용한 이진화 방법도 개발하였다.

[용도 · 이용분야]

- (1) 시각장애인 및 외국인 안내시스템: 시각장애인을 위한 보행 보조 및 외국인을 위한 안내 시스템에 활용할 수 있으며, 이는 이미지 내의 텍스트 정보를 이해하여 음성으로 제공함으로써 가능하다.
- (2) 지능형 로봇: 지능 로봇의 시각 기능에 활용하여 로봇의 자동주행, 책 읽기, 안내 등에 활용할 수 있다.
- (3) 자동차 번호판 인식: 무인 주차 시스템과 물류 시스템 등에서 활용할 수 있다.

■ 본 기술과 관련하여 추가로 확보되었거나 개발중인 기술
해당사항 없음.

[기술개요]

기술명	
개발단계	<input type="checkbox"/> 연구개발 계획 <input type="checkbox"/> 연구개발 중 <input type="checkbox"/> 연구개발 완료
기술개요	

[기술을 도출한 과제현황]

과제관리번호			
과제명			
사업명			
세부사업명			
연구기관		기관유형	
참여기관(기업)			
총연구기간			
총연구비	합계 : ()백만원 - 정부 : ()백만원 민간 : ()백만원		
연구책임자	소속		성명
	전화번호		E-mail
연구개발 주요내용			