



저작자표시-비영리-동일조건변경허락 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.
- 이차적 저작물을 작성할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



동일조건변경허락. 귀하가 이 저작물을 개작, 변형 또는 가공했을 경우에는, 이 저작물과 동일한 이용허락조건하에서만 배포할 수 있습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공 학 석 사 학 위 논 문

페이지 마커를 이용한 문서 인식 및 증강



2015년 2월

부경대학교 대학원

전자공학과

류호섭

공학석사학위논문

페이지 마커를 이용한 문서 인식 및 증강

지도교수 박한훈

이 논문을 공학석사 학위논문으로 제출함.

2015년 2월

부경대학교 대학원

전자공학과

류호섭

류호섭의 공학석사 학위논문을 인준함.

2015년 2월 27일



위원장 공학박사 문광석 (인)

위원 공학박사 권기룡 (인)

위원 공학박사 박한훈 (인)

목 차

그림 목 차	iii
표 목 차	v
I 서 론	1
1.1 연구 배경	1
1.2 연구 목적 및 필요성	2
II 관련 이론 및 연구	4
2.1 증강현실의 개념	4
2.2 증강현실의 활용사례	6
2.2.1 방송, 광고 분야	6
2.2.2 의료 분야	6
2.2.3 산업 분야	7
2.2.4 교육 분야	7
2.3 증강현실의 구현	9
2.3.1 마커 기반 증강현실	9
2.3.1.1 마커 검출	10
2.3.1.2 마커의 종류 및 ID 인식 방법	11
2.3.1.3 가상 콘텐츠 증강	14
2.3.2 비마커 기반 증강현실	20
2.3.2.1 에지 기반	21
2.3.2.2 광류 기반	22
2.3.2.3 텍스처 기반	23
2.3.2.4 비모델 기반	25
2.4 문서 인식 기술	25
III 문서 인식 및 증강 설계	29
3.1 시스템 개요	29

3.2 문서 검출	30
3.2.1 이진화	30
3.2.2 문서 후보 영역 검출	32
3.3 페이지 마커	39
3.4 페이지 마커 ID 분석	40
3.5 성능 개선	41
IV 실험결과 및 고찰	44
4.1 실험 환경	44
4.2 실험 과정	44
4.3 결과 및 고찰	49
4.4 추가 실험	53
4.4.1 서브영역 개수(L)에 따른 인식률	53
4.4.2 인덱스(index)를 이용한 페이지 마커 ID 분석	53
4.4.3 특징점 기반 페이지 ID 분석	55
4.4.4 그림 문서에 대한 실험	57
V 결 론	60
VI 향후 계획	61
참 고 문 헌	62

그림 목 차

그림 2.1. 현실세계와 가상세계의 연속성.	4
그림 2.2. 증강현실과 증강가상.	5
그림 2.3. 방송 분야의 증강현실.	6
그림 2.4. 의료 분야의 증강현실.	7
그림 2.5. 산업 분야의 증강현실.	8
그림 2.6. 교육 분야의 증강현실.	8
그림 2.7. 마커 기반 증강현실 처리 과정.	10
그림 2.8. ARToolkit의 마커 검출.	11
그림 2.9. 템플릿 마커.	12
그림 2.10. 2D 바코드 마커.	13
그림 2.11. 원형마커, Tri code와 Fourier tags.	14
그림 2.12. 마커로부터 카메라 위치 정의.	15
그림 2.13. 실세계 좌표계와 카메라 좌표계의 변환 관계.	16
그림 2.14. 핀홀 카메라의 구조.	17
그림 2.15. 실세계 좌표계와 영상 평면 좌표계의 변환 관계.	18
그림 2.16. ARToolkit을 이용한 마커 기반 증강현실 구현.	20
그림 2.17. 점 샘플링 기반 객체 추적.	21
그림 2.18. 명시적 에지 기반 증강현실.	22
그림 2.19. 광류 기반 객체 추적.	23
그림 2.20. 템플릿 매칭 기반 객체 추적.	24
그림 2.21. 특징점 기반 객체 추적.	24
그림 2.22. SfM 기반 객체 추적.	25
그림 2.23. OCR 기반 문자 인식.	26
그림 2.24. 단어들의 문자 개수를 통한 문서 인식.	27
그림 2.25. 단어들의 기하학적 매칭을 통한 문서 인식.	27
그림 2.26. 기존의 문서 인식 방법을 통한 증강현실 구현.	28
그림 3.1. 제안하는 시스템의 문서 마커화.	29
그림 3.2. 일반 이진화와 적응형 이진화 비교.	30

그림 3.3. 적응형 이진화 결과.	31
그림 3.4. 적응형 이진화의 임계값에 따른 결과.	31
그림 3.5. 외곽선 다각화.	32
그림 3.6. 허프 변환 및 직선 통합.	33
그림 3.7. $\rho-\theta$ 좌표계로 표현된 직선.	34
그림 3.8. 교점 검출 결과.	35
그림 3.9 교점 보정.	36
그림 3.10. 중형비 및 불록함 계산.	37
그림 3.11. 문서 후보 영역 검출.	38
그림 3.12. 입력 영상의 마커화 결과.	39
그림 3.13. 패턴 매칭을 위한 영상 분할.	40
그림 3.14. 유사도 d_i 의 분포.	42
그림 3.15. 마커화한 최종 사각형 후보 영역 및 인식 결과.	43
그림 3.16. 중복 꼭짓점 제거 결과.	43
그림 4.1. 데이터베이스 영상 구성.	44
그림 4.2. 입력 영상 구성.	45
그림 4.3. 제안하는 증강현실 시스템의 처리 과정.	48
그림 4.4. 거리 및 카메라 각도에 따른 ID 인식률.	49
그림 4.5. 거리에 따른 블록 모양 변화.	50
그림 4.6. 여러 가지 상황에서의 인식 결과.	51
그림 4.7. 문서 검출 및 인식이 어려운 경우.	52
그림 4.8. 시간에 따른 블록 모양 변화.	54
그림 4.9. 인덱스를 사용한 코드화.	54
그림 4.10. 특징점 기반 ID 인식률 및 처리 시간.	56
그림 4.11. 참조 평면 내 특징점 사이의 매칭 및 호모그래피 추정.	56
그림 4.12. SURF와 RANSAC을 이용한 문서 영역 호모그래피 추정.	57
그림 4.13. 그림 위주의 문서 데이터베이스 영상 구성.	58
그림 4.14. 그림 위주의 문서에서 발생하는 문제점.	59

표 목 차

표 4.1. 처리 과정별 소요 시간	46
표 4.2. 제안하는 증강현실 시스템의 각종 파라미터	47
표 4.3. 서버영역 개수에 따른 제안된 방법의 인식률	53



Ho-Sub Ryu

Department of Electronic Engineering, The Graduate School,
Pukyong National University

Abstract

Augmented reality (AR) have grown very popular because they provide users informative virtual contents in an intuitive way and interactivity to the virtual contents, thereby, attracting the interest of users. AR books, which is the main concern of this paper, is a representative one of AR applications. Thanks to their positive effects in education or entertainment, a number of AR books have been reported in the literature. However, almost all the existing AR books are available only on picture-based books or documents because the existing feature tracking methods are not useful for objects with poor texture such as text-based books or documents. Although use of fiducial markers or physical sensors is an easy way to implement AR on text-based books, the presence of markers reduces the users' immersion and the sensors require users to hold or wear additional physical equipment.

In this paper, we aim to implement AR on occluded text-based documents or books, especially distant ones where individual words or characters cannot be recognized. For this purpose, we propose a new paper-based AR system that converts documents or book pages to 2D binary markers, and identify them, estimate their relative 3D poses to the camera, and augment them with virtual contents. The system requires no fiducial markers or physical sensors. Experimental results demonstrate that the proposed system can recognize and augment occluded text-based documents or book pages with high accuracy and in real-time.

I 서론

1.1 연구 배경

정보통신기술이 발전함에 따라 사용자들의 새로운 디지털콘텐츠에 대한 요구가 증가해왔으며, 이로 인해 유비쿼터스 컴퓨팅, 사물인터넷, 3D 및 가상/증강현실 등 새로운 디지털콘텐츠 기반 기술에 대한 관심이 크게 증가하고 있다. 그 중 증강현실은 MIT Technology Review나 Gartner 그룹 등에서 대표적인 유망 미래 기술로 소개된 이후[1], 지속적으로 요소 기술의 고도화 및 관련 시장 규모의 확대가 이루어지고 있다.

증강현실이란 현실세계에 상호작용이 가능한 가상의 사물이나 정보를 실시간으로 정합하여 보여주는 컴퓨터 비전/그래픽스 기술로서 모든 정보를 3차원 가상정보로 제공하는 가상현실 기술과는 구별된다[2]. 증강현실은 방송, 광고, 게임, 의료 등의 여러 산업분야에 걸쳐 활용되고 있으며, 다양한 부가가치를 창출할 것으로 기대된다[3]. 최근 다양한 센서 및 향상된 카메라와 하드웨어를 지닌 스마트폰이 보편화되면서 모바일 기기를 통한 증강현실 시장의 규모가 급속히 확대되고 있으며, Juniper Research 그룹의 보고서에 의하면 2017년에는 약 52억 달러의 규모로 성장할 것으로 전망되고 있다[4]. 앞으로 스마트 안경 등과 같은 착용형(wearable) 기기들의 성능 향상과 더불어 그 규모는 더욱 확대될 것으로 보인다. 이전의 증강현실 기술이 새로운 콘텐츠에 대한 호기심으로 사용자들이 접하였다면, 최근에는 보다 실용적인 부분으로 사용자들에게 관심을 받고 있다. 특히, 교육 분야에서 증강현실은 직관적인 정보 제공을 통해 학습자의 흥미와 몰입감을 증가시키며 가상 정보와의 상호작용을 통한 다양한 체험형 학습이 가능하기 때문에 새로운 교육 매체로 주목받고 있다. 이와 관련하여, 현재 증강현실을 다양한 장르의 책에 접목

시킨 사례들이 소개되고 있다[5-10].

증강현실 기술이 접목된 책(이후 “증강현실 책(AR book)”이라고 함)은 책의 각 페이지에 가상 콘텐츠를 증강시키기 위해서 페이지를 찾고 구별하는 문서 인식 기술이 필요하다. 문서 인식은 카메라나 스캐너를 이용하여 문서 내 정보를 인식하는 패턴 인식의 한 분야이다. 문서란 문자나 그림 등으로 이루어진 정보 매체로서 아주 먼 옛날부터 의사소통의 한 도구로 사용되어 왔으며, 사람들은 실생활에서 문서나 문자 정보를 매우 빈번하게 접하며 살아간다. 그래서 문서 인식 기술은 패턴 인식 연구의 초창기부터 높은 관심과 함께 활발하게 연구되어 왔으며, 대부분 개별 문자나 단어를 인식하는 데 초점을 두고 있다[11-14].

1.2 연구 목적 및 필요성

증강현실 책은 책의 각 페이지에 내용과 관련된 부가적인 가상 콘텐츠를 증강시켜 보여주는 책으로 직관적인 정보를 제공함으로써 사용자에게 향상된 몰입감을 제공해 줄뿐 아니라 가상 콘텐츠와의 상호작용을 통한 체험효과를 제공한다. 증강현실 책은 교육 분야를 중심으로 다양한 장르의 책이 국내외로 출판되고 있으며, 최근 출시된 모바일 기기에는 증강현실 책을 위한 어플리케이션을 기본적으로 탑재할 정도로 증강현실 책은 대중화 되어있다. 증강현실 책은 교육 분야뿐만 아니라 엔터테인먼트, 광고 등 다양한 분야로 응용이 가능하기 때문에 관심과 수요는 더욱 증가할 것으로 예상된다[15-17]. 그렇기 때문에 책이나 문서에 증강현실을 효과적으로 구현하는 방법에 대한 요구가 크게 증가하고 있다.

시각적 추적(visual tracking)을 이용한 증강현실을 구현하는 대표적인 방법은 마커(marker) 기반과 비마커(markerless) 기반 방법으로 나뉜다. 마커 기반 증강현실은 페이지나 문서 내에 미리 고안된 마커를 부착하여 마커 종류에 따른 가상 정보를 증강시킨다[18-21]. 마커를 부

착하는 방법은 매우 간단하면서도 효율적인 방법이지만 문서의 내용과 동떨어진 색상 및 모양을 가지는 마커는 사용자에게 시각적 불편함을 주며 몰입감을 저하시킨다. 또한 일반적인 마커 기반 방법의 경우 마커의 일부분이 가려지거나 카메라 밖으로 벗어나는 경우 마커를 인식할 수 없기 때문에 증강현실을 구현할 수 없다. 반면, 비마커 기반의 증강현실은 영상으로부터 중요 정보를 담고 있는 특징점(feature)을 이용한 방법이 널리 사용되고 있다[22-26]. 특징점이란 주변과 밝기, 색상, 모양 등의 차이가 두드러지는 점들로서 페이지에서 검출된 특징점들의 매칭을 통하여 페이지를 구별하고 가상정보를 증강한다. 그러나 특징점 검출 및 매칭 방법은 흑과 백의 단순한 이진 정보인 글자(text) 위주로 작성된 문서에서는 사용이 어렵기 때문에 현재 그림이나 사진을 주 내용으로 포함하고 있는 책에서만 사용되고 있다. 그 외에도 단어를 이루는 글자의 개수나 단어들의 기하학적인 정보를 이용하여 문서 인식을 하는 방법들이 [27-29] 개발되어 있지만 이 방법들은 특징점 기반과는 반대로 글 외에 그림이나 사진이 포함된 문서에는 적용이 어려우며, 단어나 글자의 구별이 가능한 충분한 해상도가 요구되기 때문에 카메라의 움직임이 크게 제약적이다.

그러므로 미리 고안된 마커를 삽입하지 않고 증강현실을 구현함에 있어 글자나 그림뿐만 아니라 다양한 내용물을 포함하고 있는 문서의 인식을 위한 보다 효과적인 방법이 요구된다. 특히, 단어나 글자의 구별이 불가능한 상황이나 페이지나 문서의 일부분이 가려지는 상황에서도 동작이 가능한 문서 인식 방법이 필요하다.

II 관련 이론 및 연구

2.1 증강현실의 개념

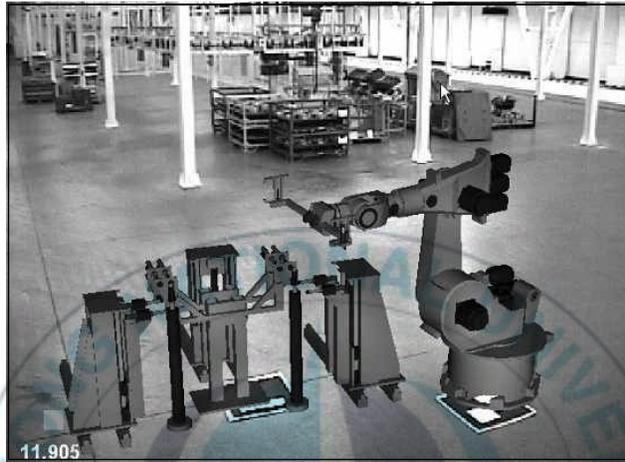
증강현실은 현실세계에 대한 카메라 영상에 상호작용이 가능한 가상의 콘텐츠를 정합시키는 기술이다. 증강현실은 1968년, Head Mounted Display(HMD)를 시초로 연구가 시작되었으며, 1992년 보잉사에서 항공기의 조립을 돕기 위해 실제 화면에 가상이미지를 중첩시켜서 설명하는 과정에서 증강현실이란 용어가 최초로 사용되었다. Milgram은 Mixed Reality(MR)이라는 용어를 통해 그림 2.1과 같이 현실세계와 가상세계의 경계를 증강현실과 증강가상이 추가된 4단계로 구분한다[2].



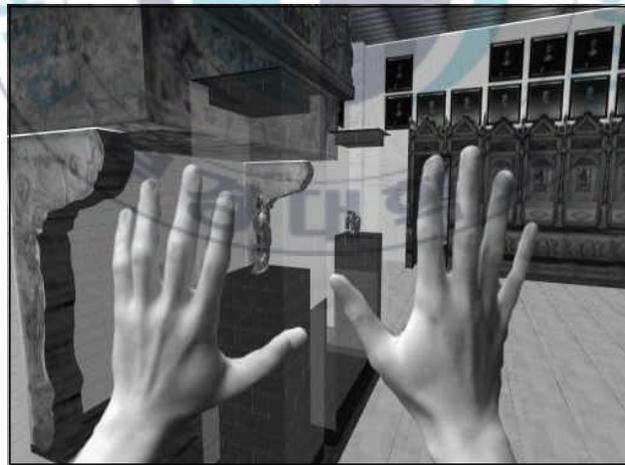
그림 2.1. 현실세계와 가상세계의 연속성.

그림 2.2-(a)는 실제 환경에 가상의 머신을 정합시켜 증강현실을 구현한 예를 보여주고 있으며, 그림 2.2-(b)는 가상현실로 구성된 공간에 실제 사용자의 손을 투영시켜 상호작용을 하는 증강가상을 보여준다. 증강가상은 가상세계 위에 현실정보를 보충해 몰입하게 하는 기술인 반면, 증강현실의 경우 현실세계의 정보 위에 가상정보를 정합함으로써 사용자의 오감을 향상시키기 때문에 가상현실과 증강가상 보다 현실성이 강조된다 [30-32]. 가상현실은 사용자의 현실세계 전체를 컴퓨터가 구축한 가상정

보로 대체하여 직접 체험하지 않고서도 그 환경에 들어와 있는 것처럼 보여주고 조작할 수 있게 해주는 기술로서 현실정보가 반영되는 증강현실과 증강가상과는 차이가 있다.



(a) 증강현실



(b) 증강가상

그림 2.2. 증강현실과 증강가상.

2.2 증강현실 활용사례

2.2.1 방송, 광고 분야

방송과 광고에서는 현재 증강현실 기술을 상당히 활용하고 있다. 시청자나 소비자는 원하는 정보를 이전보다 쉽게 취득할 수 있어 상업적 효과를 크게 증가시키고 있다. 방송에서는 스포츠 중계나 일기예보 등에서 증강현실 기술을 빠르게 도입하고 있다. 예를 들어 그림 2.3처럼 스포츠 중계 화면에 선수나 경기정보들을 필드 위에 직접 띄워주거나 위화감 없는 광고의 삽입으로 시청자들의 편의를 증진시키고 있다.



그림 2.3. 방송 분야의 증강현실.

2.2.2 의료 분야

증강현실은 의료 분야에서도 활발히 연구되고 있다. 예를 들자면, MRI나 CT 영상을 환자와 합성하여 보여주는 기술이 있다. 뮌헨대학교에서는 그림 2.4와 같은 맥락적 해부 묘사 방법을 통해 환자의 내부를 미리 살펴보는 방법을 연구하고 있다[33]. 이 연구에서는 환자의 얼굴에 IR

발광체를 붙이고 IR 카메라를 사용해서 정확한 얼굴의 자세를 추적한 뒤 기존의 MRI 데이터를 정합하여 환자의 얼굴 내부를 보는 듯이 진료가 가능하다.



그림 2.4. 의료분야의 증강현실.

2.2.3 산업 분야

산업 분야에서는 비행기나 자동차와 같은 복잡한 구조물의 설계, 제작 시 수많은 정보를 직관적으로 표시하기 위해 증강현실 기술을 이용해 왔다. 그림 2.5는 미 해병대에서 증강현실을 이용하여 고장난 부분을 수리하는 장면이다[34]. 사용자는 HMD(Head Mounted Display)를 착용하고 각 부분에 대한 설명이나 교체가 필요한 부분에 대한 정보를 볼 수 있게 된다. 이러한 방식은 전문성이 요구되는 분야에 일반인이 손쉽게 따라서 할 수 있는 도와주는 역할로 사용된다.

2.2.4 교육 분야

증강현실 기술의 교육적 활용에 대한 효과는 예전부터 다양한 연구를 통해 입증되어오고 있다. 증강현실은 3차원 가상 정보를 통하여 복잡한 개념을 직관적으로 보여줌으로써 학습자의 이해 및 적용 능력을 높일 수

있으며, 높은 현존감 및 상호작용을 바탕으로 훈련 등과 같은 목적에서 가상의 도구를 이용하여 손쉽게 경험이 가능하다. 또한, 실제 학습 환경에 가상의 정보를 추가해줌으로 해서 학습자에게 보다 향상된 몰입감과 흥미를 유발한다. 가상 정보가 실제와 유사하고 현존감이 높을수록 학습 효과도 높아지며, 원리 및 과정에 대한 정확한 이해가 가능하다[35].

뉴질랜드의 HITLab에서 개발된 MagicBook(그림 2.6)은 책 페이지에 맞게 가상 콘텐츠를 정합하여 보여주는 방식으로 기존의 책에서는 볼 수 없는 입체적인 정보를 제공, 사용자에게 몰입감 및 현재감을 준다[36]. 그 외에도 AR Volcano 및 AR Planet을 개발하여 화산 폭발 과정이나 태양계의 행성을 입체적으로 관찰할 수 있게 하였다.



그림 2.5. 산업 분야의 증강현실.



그림 2.6. 교육 분야의 증강현실.

2.3 증강현실 구현

증강현실을 구현함에 있어서 가상 콘텐츠를 영상 위에 원하는 자리에 정확히 위치시키기 위해서는 카메라의 위치와 방향을 결정해야한다. 증강현실에서 추적(tracking)이란 실시간으로 카메라의 위치와 방향을 계산하는 것으로 사용되는 장비에 따라 센서 추적, 시각적(visual) 추적, 하이브리드 추적 방법들이 있다. 현재 대부분의 증강현실은 카메라를 사용하기 때문에 시각적 추적 방법은 증강현실에서 큰 비중을 차지하고 있으며, 크게 마커 기반 방법과 비마커 기반 방법으로 나눌 수 있다.

2.3.1 마커 기반 증강현실

시각적 추적 방법은 카메라가 현재 바라보고 있는 영상을 바탕으로 카메라 위치를 추정한다. 하지만 입력 영상 외에 아무런 정보가 없다면, 카메라 위치를 추정하는 것은 상당히 어렵다. 이를 극복하기 위한 하나의 방법은 영상으로부터 쉽게 검출할 수 있는 마커(marker)를 삽입하고, 그 마커를 찾는 것이다. 이러한 방법을 마커 기반 추적이라 하며, 증강현실에서 널리 사용되고 있는 방법이다.

마커 기반 증강현실은 간단하게 구현이 가능하며 성능도 뛰어난 방법으로서 잘 알려진 마커 기반 방법으로는 ARToolkit[19], ALVAR[20], ARTag[21]가 있다. 마커 기반 증강현실을 위해서는 마커 검출, 마커 ID(identification) 인식, 카메라 위치 계산이 요구된다. 그림 2.7은 마커 기반 증강현실의 대략적인 처리 과정을 나타낸다. 수행 순서나, 세부 알고리즘 차이는 있지만 대부분 비슷한 과정으로 진행된다.

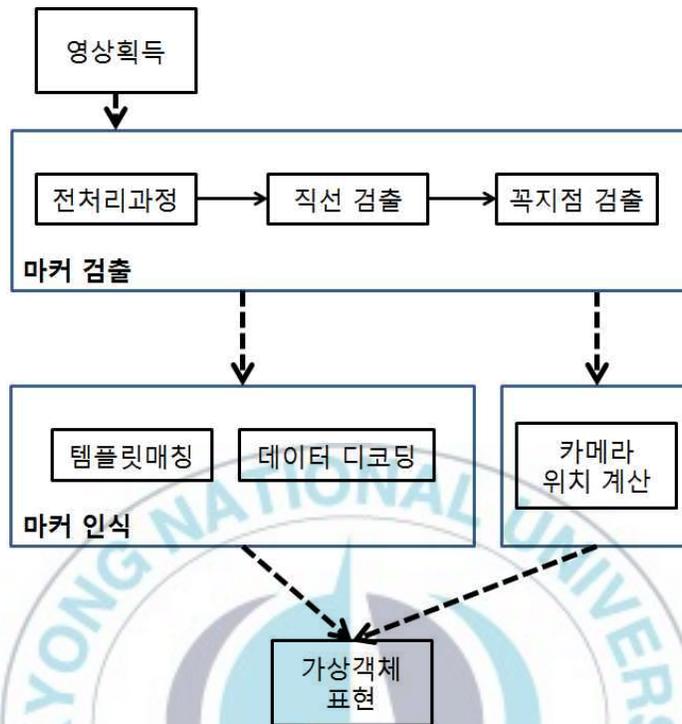
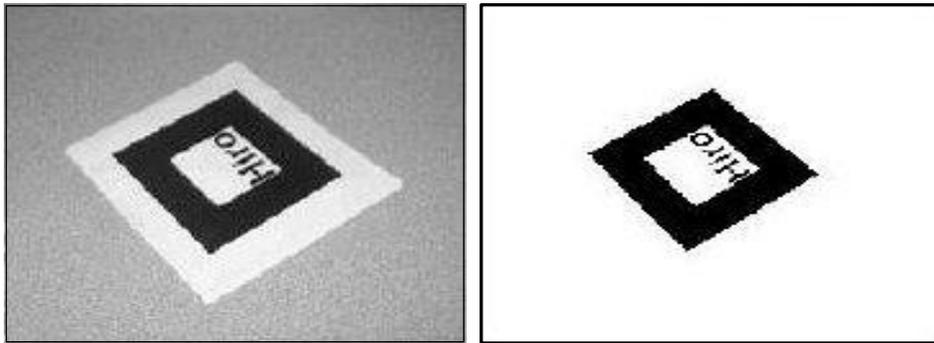


그림 2.7. 마커 기반 증강현실 처리 과정.

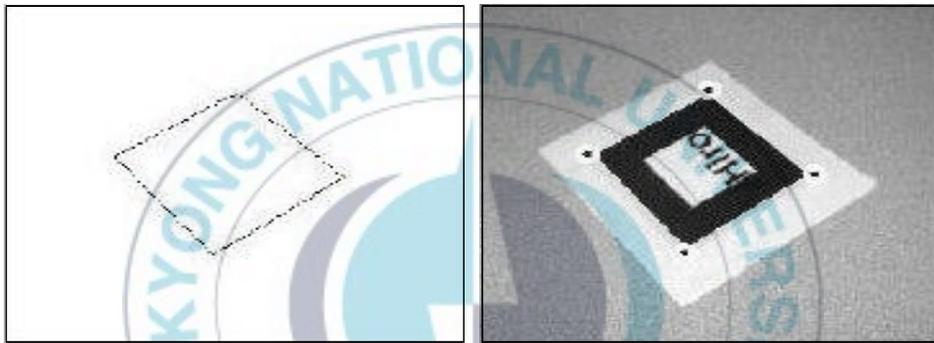
2.3.1.1 마커 검출

실세계 좌표계(world coordinates)로부터 위치를 알고 있는 4개 이상의 점이 있다면, 실세계 좌표와 영상 내의 좌표 사이의 관계를 추정할 수 있다. 4개의 점이 사각형일 때 가장 간단하게 계산이 가능하기 때문에 주로 사각 마커를 사용한다. 마커 검출 과정의 첫 번째 단계는 잠재적 마커 후보를 찾는 것과 마커의 꼭짓점 위치를 추정하는 것이다. 먼저, 전처리 과정으로 영상 내의 물체들을 배경과 분리시키기 위해 이진화 방법을 적용하며, 부분적 조명 변화에 대처하기 위해 일반적으로 적응형 이진화(adaptive threshold)를 사용한다. 분리된 물체들로부터 경계를 추출하여, 경계가 4개의 직선과 꼭짓점을 가지는 즉, 사각형의 형태를 가지는 물체를 잠재적 마커 후보로 본다.



(a) 입력 영상

(b) 이진화



(c) 윤곽선 검출

(d) 마커 및 꼭짓점 검출

그림 2.8. ARToolkit의 마커 검출.

2.3.1.2 마커의 종류 및 ID 인식 방법

좋은 마커란 모든 상황에서 쉽고 신뢰성 있게 검출이 되어야 한다. 물체의 색상은 보는 위치와 조명 등의 주변 환경에 영향을 많이 받기 때문에 물체를 검출함에 있어서 색상의 변화가 밝기에 대한 변화보다 민감하게 작용을 한다. 그리고 마커가 사각형 형태일 때 위치를 쉽고 정확하게 추정 가능하기 때문에 일반적으로 대비가 큰 흑백의 사각형 마커가 일반적으로 사용된다.

마커 기반 증강현실은 현재 수많은 종류의 마커가 개발되어 있지만, 크게 템플릿 매칭을 이용하는 템플릿 마커(template marker)와 디코딩

(decoding)을 이용하는 데이터 마커(data marker)로 구분된다. 템플릿 매칭의 경우 데이터베이스 내의 모든 마커들과 비교를 통해 ID를 인식해야 하지만 디코딩은 마커에 데이터를 인코딩(encoding)이 가능하므로 별다른 비교 과정이 요구되지 않는다.

템플릿 마커

템플릿 마커는 그림 2.9와 같이 내부에 간단한 그림을 사용하며, ARToolkit[19] 방법에서 처음으로 사용하였다. 템플릿 마커는 각 픽셀들의 밝기 값이나 미리 설정해둔 크기를 가지는 격자 내의 픽셀 밝기 값들의 평균을 SSD(the sum of squared differences)나 교차상관관계(cross-correlation) 방법을 통해 차례대로 비교하는 템플릿 매칭을 이용해 데이터베이스로부터 검출된 마커의 내부 그림과 가장 유사한 그림을 가진 마커를 찾아낸다. 템플릿 매칭은 자체적으로도 느리지만 회전에 대한 불변성(invariant)도 가지고 있지 않아 4방향에 대해 총 4번의 매칭 수행이 필요해 처리 속도가 굉장히 느리다는 단점이 있다. 또한, 사람의 눈으로 보았을 때 확연히 구분이 가능하더라도 템플릿 매칭으로는 구분이 어려운 경우가 있어 마커 내부의 그림에도 제약이 따른다.

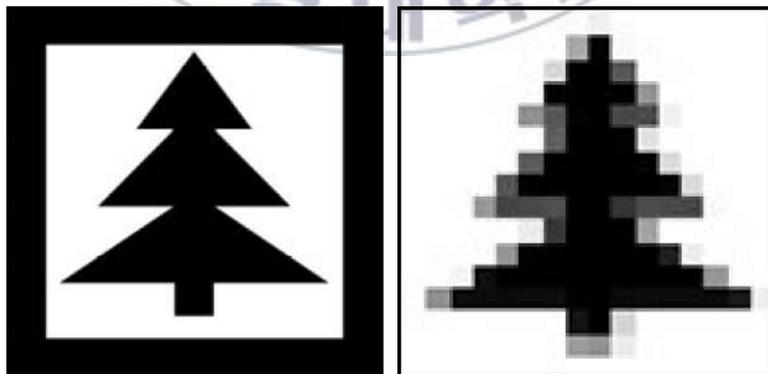


그림 2.9. 템플릿 마커.

데이터 마커

데이터마커는 내부의 흑백으로 표현된 이진 디지털 코드로 구성된다. 디지털 코드는 ID뿐만 아니라 데이터 인코딩이 가능하다. 그림 2.10은 2D 바코드 마커의 대표적인 예이다. 그림 2.10의 좌영상은 간단한 코드로 구성되어 ID 정보만을 담고 있다. 하지만 우영상과 같이 격자수에 따라 ID 외에 추가적인 데이터 인코딩이 가능하며, 기존의 에러검출코드인 CRC(cyclical redundancy check)나 FEC(forward error correction) 등을 통해 에러 탐지 및 수정이 가능해 낮은 긍정오류(false positive error)를 보인다.



그림 2.10. 2D 바코드 마커.

그 외 마커들

사각 마커는 위치 추정을 위해 4개의 꼭짓점을 찾지만 원형 마커는 둘레를 이용하기 때문에 사각 마커보다 검출은 어렵지만 마커 중심에 대한 위치 추정이 가능하다. 하지만 카메라 위치를 파악하기엔 충분하지 않아 실시간 및 카메라 위치의 자동 추적이 필요한 증강현실에서는 원형 마커는 거의 사용되지 않는다[37,38]. 그리고 이미 잘 알려진 QR code[39]도 2D 바코드 마커의 대표적인 방법이며, 마커 내부에 8개의 방향을 가지는 삼각형을 삽입하여, 삼각형 방향을 디지털 코드로 인코딩

하여 사용하는 Tri code[40]와 내부 패턴을 주파수 성분으로 구성한 Fourier tags[41] 등 다양한 형태의 마커가 연구되고 있다(그림 2.11). 하지만 사각형 형태의 마커가 카메라의 위치 추적에 가장 용이하여, 증강현실에서 가장 많이 사용되고 있다.

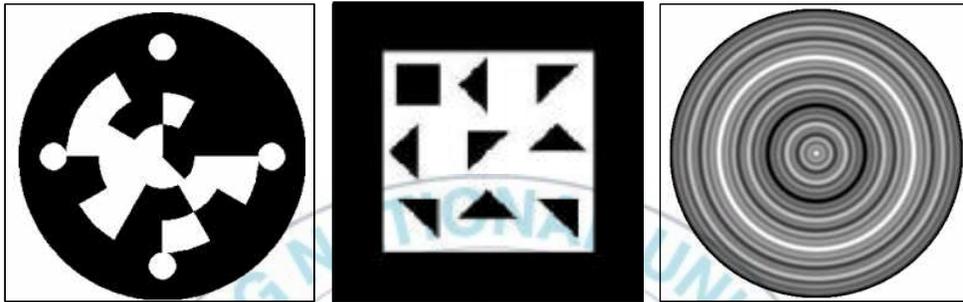


그림 2.11. 원형마커, Tri code와 Fourier tags.

2.3.1.3 가상 콘텐츠 증강

증강현실은 현실 정보 위에 가상의 콘텐츠를 정합시켜 보여주는 기술이다. 하지만 현실과 가상 콘텐츠는 3차원 좌표로 이루어져 있지만 정합된 영상은 카메라를 기준으로 한 2차원으로 이루어진 영상이다. 그렇기 때문에 영상 내의 원하는 위치에 가상 콘텐츠를 정합시키기 위해서는 실세계 좌표(world coordinates)부터 카메라를 중심으로 한 영상의 2차원 좌표와의 관계를 알아야 한다. 카메라 파라미터들을 이용하여 영상과의 관계를 파악할 수 있으며, 영상 속에서의 위치를 파악하게 되면 그 부분에 가상 콘텐츠를 정합시켜 증강현실을 구현할 수가 있다. 하지만 3차원 좌표는 한 대의 카메라로는 계산할 수가 없기 때문에 마커를 이용하여 마커를 중심으로 하는 실세계 좌표를 구성한다. 마커로부터 가상 콘텐츠의 실세계 좌표를 계산하여 2차원 영상으로 정합시켜 증강현실을 구현할 수가 있다.

카메라 외부 파라미터

카메라 외부 파라미터는 카메라 좌표계와 실세계 좌표계 사이의 변환 관계를 설명하는 파라미터이다. 카메라 외부 파라미터는 카메라 고유의 파라미터가 아니라 설치된 카메라의 위치와 방향에 따라 다르며, 실세계 좌표계를 어떻게 정의하였는지에 따라 달라진다. 마커 기반 증강현실에서는 실세계 좌표계를 마커의 중심으로 정의시켜 비교적 간단하게 외부 파라미터의 계산이 가능하다. 마커로부터 카메라 위치는 그림 2.12와 같이 회전각(α, β, γ)과 좌표 이동(x, y, z)을 가진 6자유도(6DOF)로 표현된다.

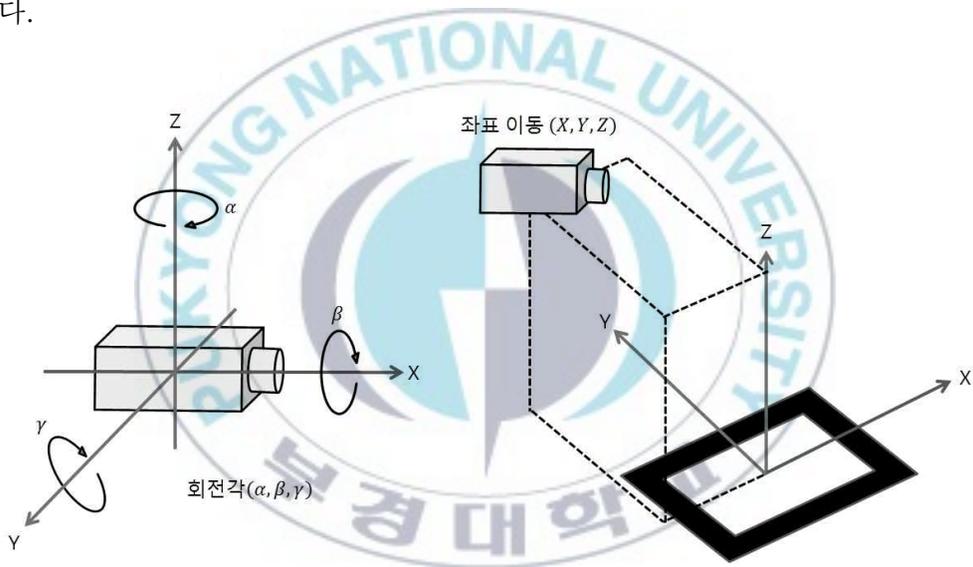


그림 2.12. 마커로부터 카메라 위치 정의.

마커를 중심으로 하는 실세계 좌표계(world coordinates) X 와 카메라를 중심으로 하는 카메라 좌표계 x_c 사이의 변환 관계는 이동 벡터 t 와 회전 행렬 R 로 구성된 카메라 외부 파라미터 T 로 정의된다(그림 2.13).

$$x_c = TX, T = [R | t]. \quad (2-1)$$

회전 행렬 R 는 3개의 파라미터 (α, β, γ) 를 가지며 3×3 행렬로 표현되며, 이동 벡터 t 는 3개의 스칼라 값으로 표현된다. 카메라 좌표계와 실세계 좌표계의 관계를 풀어서 나타내면 식(2-2)와 같다.

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} r_1 & r_2 & r_3 & t_{x_c} \\ r_4 & r_5 & r_6 & t_{y_c} \\ r_7 & r_8 & r_9 & t_{z_c} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \quad (2-2)$$

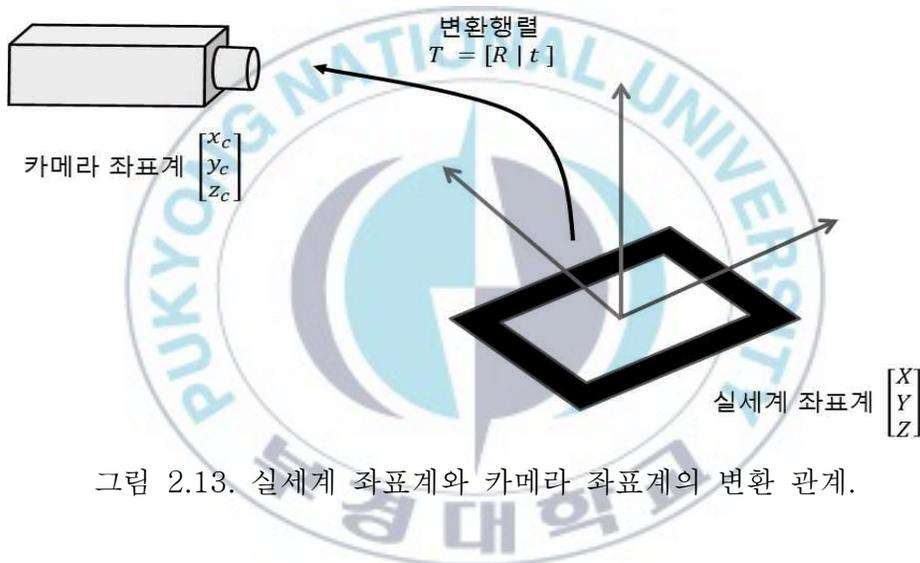


그림 2.13. 실세계 좌표계와 카메라 좌표계의 변환 관계.

카메라 내부 파라미터

그림 2.14는 이상적인 핀홀 카메라의 구조이다. 이상적이 핀홀 카메라에서, 빛은 카메라의 중심을 지나서 무한대로 뻗어가 카메라 내부의 영상 평면(image plane)에 맺힌다. 하지만 관찰되는 물체는 3차원으로 이루어져있고, 영상은 영상 평면에 2차원으로 얻어진다. 그래서 카메라 좌표계를 중심으로 한 물체의 3차원 좌표를 영상 평면의 2차원 좌표계로 변환이 필요하다. 이때의 변환 관계를 카메라 내부 파라미터라고 부른다. 카메라 좌표계 x_c 와 영상 평면 좌표계 x_p 사이의 관계는 카메라 내부 파라미터 K 를 이용해서 정의 할 수 있다.

K 는 평면 격자 패턴을 이용한 카메라 보정(camera calibration)[42]을 통해 계산이 할 수 있으며 최종적인 변환 관계는 식(2-4)로 표현된다.

$$\begin{bmatrix} x_p \\ y_p \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & p_{x_p} & 0 \\ 0 & f & p_{y_p} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix}. \quad (2-4)$$

위치 계산

카메라 외부 파라미터는 실세계 좌표와 카메라 좌표 사이의 관계를 나타내며, 내부 파라미터는 카메라 좌표와 영상 평면 좌표 사이의 관계를 나타낸다. 이 두 파라미터를 이용하면 실세계 좌표계 X 로부터 영상 평면 좌표계 x_p 로의 변환관계를 구할 수가 있으며, 이는 식(2-5)와 그림 2.15으로 표현된다.

$$x_p = KTX, \quad T = [R | t], \quad K = \begin{bmatrix} f_x & s & p_{x_p} & 0 \\ 0 & f_y & p_{y_p} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \quad (2-5)$$

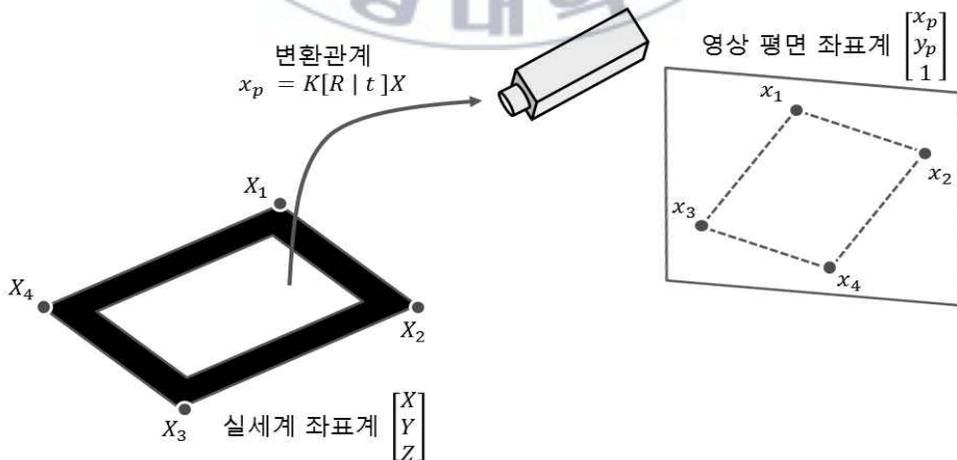


그림 2.15. 실세계 좌표계와 영상 평면 좌표계의 변환 관계.

식(2-5)를 풀어쓰면 다음과 같다.

$$\begin{bmatrix} x_p \\ y_p \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & p_{x_p} & 0 \\ 0 & f & p_{y_p} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} r_1 & r_2 & r_3 & t_{x_c} \\ r_4 & r_5 & r_6 & t_{y_c} \\ r_7 & r_8 & r_9 & t_{z_c} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \quad (2-6)$$

영상 내의 마커의 네 꼭짓점 좌표 x_i 알고 있으며, 크기를 알고 있는 마커로부터 실세계의 마커의 꼭짓점 좌표 X_i 를 알 수 있다. $KT = M$ 으로 두고 표현하면, 식(2-6)이 된다.

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \begin{bmatrix} m_1 & m_2 & m_3 & m_4 \\ m_5 & m_6 & m_7 & m_8 \\ m_9 & m_{10} & m_{11} & m_{12} \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix}. \quad (2-6)$$

$$x_p = MX_{ar}. \quad (2-7)$$

마커는 평면이기 때문에 Z_i 는 무조건 0의 값을 가지게 되고, M 는 실제적으로 3×3 호모그래피(homography)행렬이 된다. 식(2-6)으로부터 실세계의 좌표계의 마커와 영상 내의 마커의 변환 관계 M 을 계산할 수가 있으며, M 을 계산하면 식(2-7)을 통해 마커를 중심으로 한 가상 콘텐츠의 3차원 좌표 X_{ar} 를 영상 평면 좌표계 x_p 로 변환시켜 그림 2.16과 같이 마커 위에 가상의 콘텐츠를 증강시킬 수 있다.



그림 2.16. ARToolkit을 이용한 마커 기반 증강현실 구현.

2.3.2 비마커 기반 증강현실

비마커 기반 증강현실은 크게 모델 기반 방식과 Structure from Motion(SfM) 기반 방식으로 분류할 수 있다. 모델 기반 방식에서는 객체 추적이 이루어지기 전에 실세계로부터 정보를 획득하여 객체의 3D모델을 저장하고, 카메라의 방향을 추정한다. 이에 반하여 SfM 기반 방식은 카메라로부터 입력되는 여러 장의 프레임을 통해서 사전 정보 없이 카메라 위치를 추정할 수 있다. 모델 기반의 방식은 데이터베이스에 저장된 객체만이 추적된다는 단점이 있지만 SfM 기반 방식에 비해 간단하고, 가상 객체와의 상호 작용을 쉽게 구현 가능하기 때문에 현재 가장 많은 연구가 진행되고 있는 분야이다. 모델 기반 방식은 인위적인 마커를 삽입하지 않고 영상에 존재하는 에지(edges), 광류(optical flow), 텍스처(texture)와 같은 다양한 정보를 이용한 객체 추적을 통하여 가상 콘텐츠를 투영하는 방법을 사용한다.

2.3.2.1 에지 기반

카메라의 방향은 영상의 에지 정보와 객체의 에지를 정합함으로써 추정할 수 있다. 이러한 객체 정합은 객체의 에지 정보와 카메라로부터 입력 받은 객체의 에지 정보간의 차이가 최소가 되는 형태로 이루어진다. 이러한 객체 정합을 하기 위해서는 초기에 객체의 위치와 방향에 대한 사전 정보가 있어야 한다. 이러한 초기화 작업은 수작업으로 이루어진다. 초기에 결정된 객체의 위치와 방향 정보는 다음 프레임의 모델정보로 사용한다. 에지 기반 객체 추적은 복잡도가 낮아 계산상으로 효율적이고 구현하기가 비교적 쉬운 장점이 있다. 그렇지만 에지 기반 방식은 잡음 및 영상의 해상도에 따른 성능이 떨어지는 단점이 있다. 에지 기반 추적 방식은 두 가지로 나누어 볼 수 있는데 첫 번째는 그림 2.17에서 보는 바와 같이 에지를 추출함 없이 이미지에서 강한 변화를 갖는 부분을 찾아 객체 추적에 사용하는 점 샘플링(point sampling) 방법이다[43]. 점 샘플링 방법은 그림에서 보는 바와 같이 투영된 객체 모델의 통제점(control point)들과 이미지 객체의 대응점을 찾아 그 차이가 최소가 되는 방향으로 점 추적이 이루어진다.

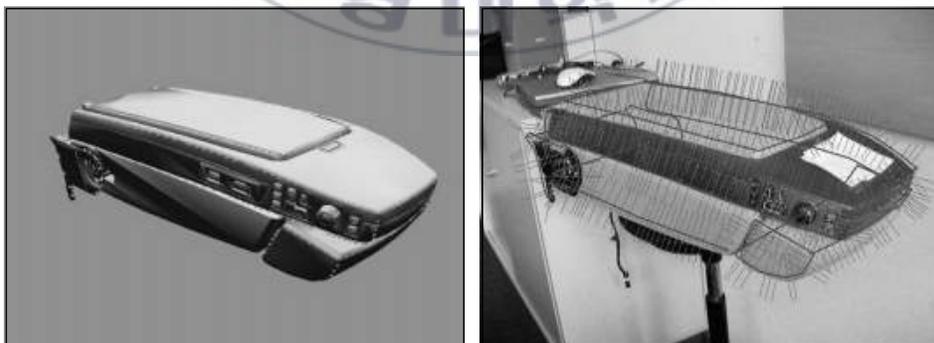


그림 2.17. 점 샘플링 기반 객체 추적.

명시적 에지 추출방법은 허프 변환[44]과 같은 선분 검출 연산자를 이

용하여 에지를 추출한다[45]. 추출된 에지와 영상의 에지를 정합하여 현재 카메라의 방향을 계산한다. 그림 2.18는 기둥으로부터 추출된 에지와 저장된 모델의 에지와의 정합을 통해 가상의 콘텐츠를 삽입한 모습이다. 에지 추출방식은 점 샘플링 방식보다 더 강력하나, 선분을 사용함으로써 다각형 객체에만 적용 가능하여 일반성이 미흡하여, 점 샘플링 방식이 에지 추출 방식보다는 효율적이다.



그림 2.18. 명시적 에지 기반 증강현실.

2.3.2.2 광류 기반

광류는 물리적인 점으로부터 영상으로 투영된 점의 움직임으로서, 영상과 모델의 정합에 의해 추출되는 공간정보에 의존하는 에지 기반의 방식과는 달리 광류 기반의 추적 방식은 시간적 정보를 추출한다[46]. 이는 영상에 투영된 객체의 상대적 이동으로부터 추출된다. 수동으로 초기화한 후에 t 와 $t+1$ 에 캡처된 프레임 사이의 광류가 계산된다. 그리고 시각 t 에서 투영된 모델의 점들 중 어느 점이 시각 $t+1$ 에서 여전히 나타날지를 알고리즘이 결정한다. 이 점들의 시간에 따른 이동량, 즉 속도를 이용하여 카메라의 이동을 추정한다. 광류 기반은 낮은 처리 시간을 요구하지만 조도의 변화와 카메라의 급격한 이동에는 강인하지 못하여 추적 오차를 유발하며, 오차 누적으로 인한 정확도 저하로 인하여 잦은 초기화를 해야

하는 단점이 있다. 그림 2.19은 광류 추적을 통한 증강현실 구현을 보여 준다[47].

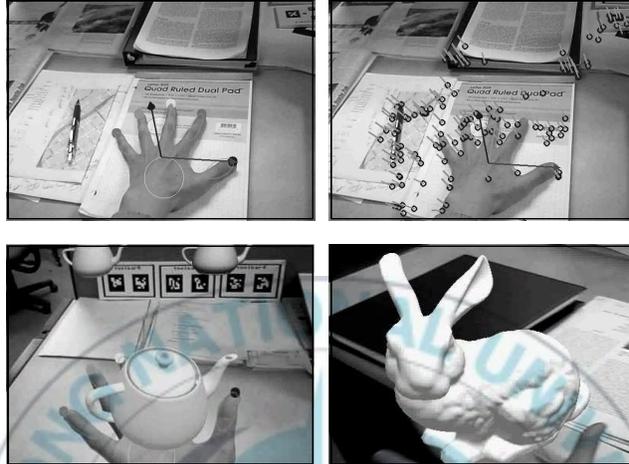


그림 2.19. 광류 기반 증강현실.

2.3.2.3 텍스처기반

텍스처 기반 객체 추적방법은 이미지 위에 나타난 텍스처 정보를 이용하는 방법으로 템플릿 매칭 방법[48]과 특징점 기반 방법[49,50] 두 가지가 있다. 템플릿 매칭 방법은 객체들의 고정된(rigid) 이동을 얻기 위해 적용된 방법으로 전역적 정보를 이용한다. 마커 기반 증강현실의 템플릿 마커의 ID 인식과 마찬가지로 방법으로 최소 제곱 오차방법을 적용하여 영상 영역과 기준 템플릿과의 정합 정도를 구해 객체 인식 및 추적에 사용한다. 이 방법은 지역 특징들에 의해서 모델화하기 어려운 복잡한 패턴을 다룰 수 있는 능력이 장점이라 할 수 있으나, 조도 변화와 가려짐에 약한 단점이 있고, 3D 추적에서도 좋은 결과를 얻지 못했다.

특징점 기반 방법은 템플릿 매칭 방법처럼 전역적 탐색 방법을 이용하는 것이 아니라 지역 특징정보를 고려한다. 광류 기반 방법과 유사하게 지역 특징점 정보를 이용하지만 탐색과 추적을 돕기 위해 텍스처 정보를

사용한다. 장점을 가지고 있으며 특징점 기반의 객체 인식 및 추적 방법들은 관심 연산자(interest operator)를 사용하여 이미지의 일정부분만을 가지고도 매칭이 가능하며, 템플릿 매칭보다 조도 변화에 강인하다는 장점이 있다. 특징점 기반의 매칭은 좋은 연산자 및 특징점을 찾는 것이 무엇보다 중요하다. 특징이 올바르게 선택이 되었다면 객체 추적 시간을 줄이고 정확도를 높일 수 있다. 그림 2.20은 템플릿 매칭 기반으로 영상 내에서 객체를 추적한 예이며, 그림 2.21은 영상 내 특징점으로부터 카메라 위치를 추정한 뒤 가상의 콘텐츠를 삽입한 것을 보여준다.

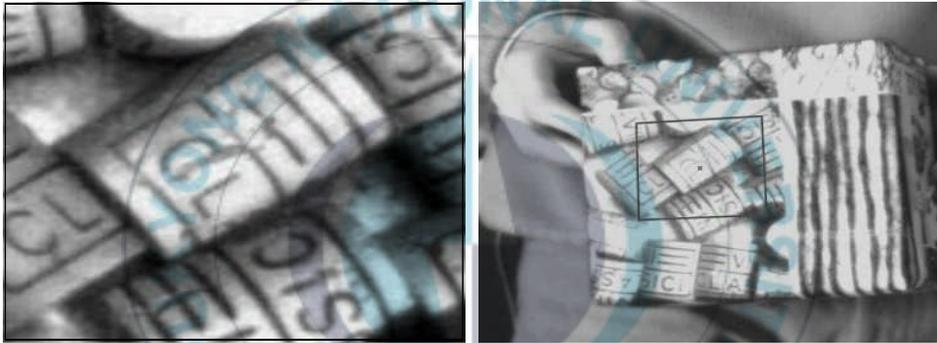


그림 2.20. 템플릿 매칭 기반 객체 추적.

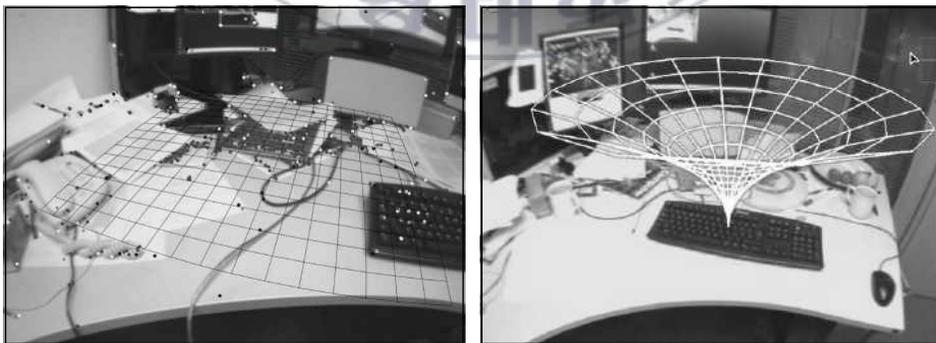


그림 2.21. 특징점 기반 증강현실.

2.3.2.4 비 모델 기반

비 모델 기반 방법은 추적되는 영상에 대해 사전 3D모델 정보를 이용
함 없이 실시간으로 영상의 기하학적 구조를 복원한다. 이러한 방법은 온
라인 방식이어서 오프라인으로 학습하는 과정없이 미지의 환경을 재구축
할 수 있지만 매우 복잡하다는 단점이 있다. 비 모델 기반 방법은 카메라
의 모션과 영상의 기하학적 구조를 추정할 수 있는데 실시간으로 영상화
면의 3D 구조를 임의의 좌표 체계로 표현할 수 있는 방법으로 정형화된
SLAM(Simultaneous Localization and Mapping) 의 MonoSLAM 방법
[51]과 또한 실시간으로 3D 재구성을 하기 위한 SfM(Structure from
Motion)의 Real-time SfM[52,53]방식이 있다. 그림 2.22은 같은 대상
을 찍은 여러 프레임으로 Real-time SfM방식을 이용하여 카메라 위치를
추정한 것을 보여준다.

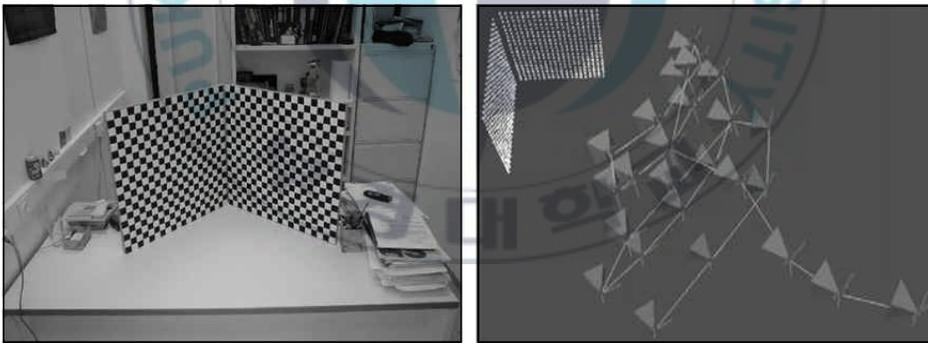


그림 2.22. SfM 기반 카메라 위치 추정.

2.4 문서 인식 기술

문서란 아주 먼 옛날부터 사용이 되어온 문자나 그림 등의 정보를 담
고 있는 정보 매체이다. 현재까지도 문자나 문서 정보가 생활에서 사용
되는 빈도는 굉장히 높으며 패턴 인식 분야에서도 오래 전부터 문자 및

문서 인식에 대한 연구를 활발히 진행해오고 있다.

현재 개발된 문서 인식 기술의 대부분은 정형화된 형태로 스캔된 영상에서 글자의 구조적인 특징인 굴곡점, 분기점, 교차점, 종단점, 끝점, 연결점 등을 이용하는 OCR(optical character recognition)방법을 통해 그림 2.23처럼 글자를 먼저 검출하고, 검출된 글자들의 배열을 가지고 문서를 인식하고 있다[11-14]. 일반적으로 사람이 문서를 인식할 때 단순히 문서 내의 문자만을 인식하는 것이 아니라 내용물의 구조를 토대로 문서를 인식하게 된다. 그래서 사람과 비슷한 방법으로 문서의 기하학적인 정보를 이용해서 문서를 인식하는 몇몇 방법들도 개발되어 있다. 그림 2.24는 문서 내 단어들을 구성하는 문자들의 개수를 수직과 수평방향으로 분석하여 문서를 인식을 하는 방법이며[27], 그림 2.25는 단어들의 중심을 계산하여 중심으로부터 단어들의 기하학적 매칭을 통하여 문서를 인식하는 방법을 보여준다.



그림 2.23. OCR 기반 문자 인식.

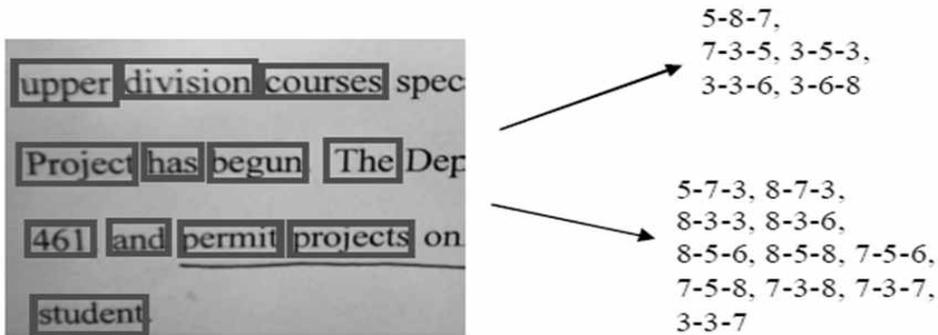


그림 2.24. 단어들의 개수를 통한 문서 인식.

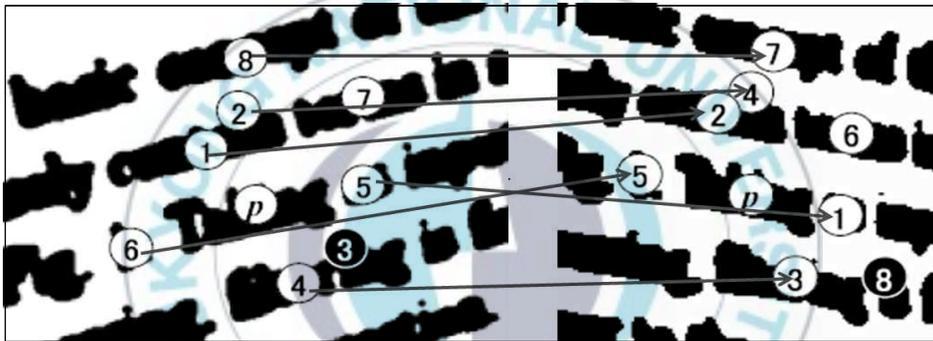


그림 2.25. 단어들의 기하학적 매칭을 통한 문서 인식.

마커를 사용하지 않고 문서 위의 정확한 위치에 가상 콘텐츠를 증강시키기 위해서는 문서 인식뿐만 아니라 카메라의 위치도 알아야한다. 하지만 문자를 직접 인식하는 OCR 방법이나 단어들의 기하학적인 정보를 이용하는 문서 인식 방법을 이용해서 카메라의 3차원 위치 추적하기는 매우 어렵다. 그래서 현재 그림 2.26과 같이 단순히 2차원 정보만 보여주는 기술로 많이 사용되고 있다. 또한, 문자나 단어를 인식하기 위해서는 고해상도의 입력 영상이 요구되며 특징점 기반을 통한 증강현실 방법들[22-26]과 반대로 글자(text) 위주의 문서가 아닌 그림이나 사진이 혼합된 문서는 인식이 어렵다는 단점이 있다.

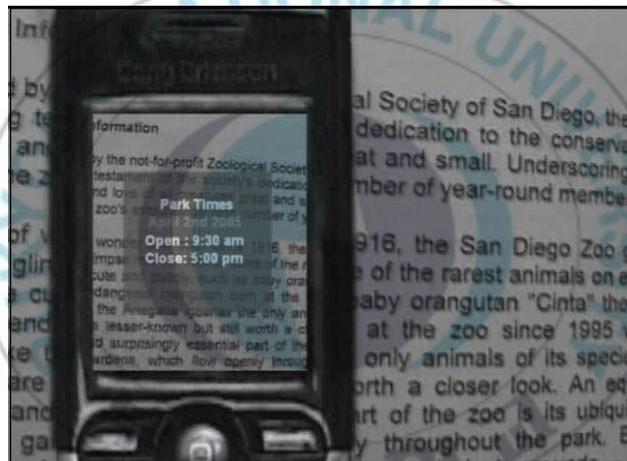
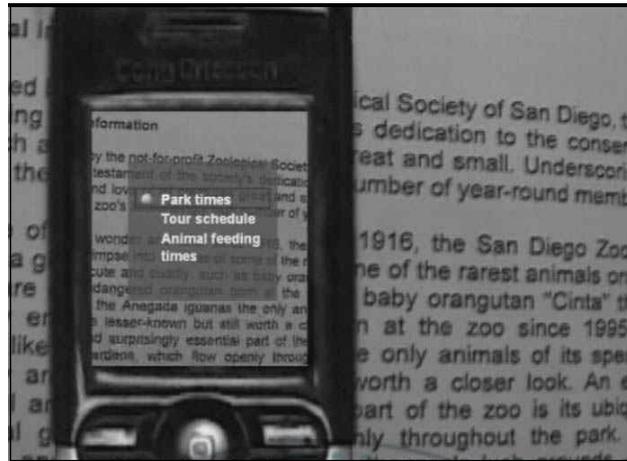


그림 2.26. 기존의 문서 인식 방법을 통한 증강현실 구현.

Ⅲ 문서 인식 및 증강 설계

3.1 시스템 개요

제안하는 시스템은 그림 3.1과 같이 일반적인 문서 내부를 기존의 사각 마커의 패턴과 유사하게 변환시켜 기존의 마커 기반과 동일한 방법을 통해 문서에 증강 현실을 구현하는 것이다. 먼저, 영상을 획득하면 이진화를 통해 문서의 경계선(boundary)을 분리해 내고 경계선으로부터 문서 후보 영역을 검출해낸다. 문서 후보 영역을 검출하였으면 그 내부를 모폴로지(morphology) 기법을 사용하여 잡음 제거 및 내용물을 블록화시켜 마커화한다. 제안하는 시스템에선 마커화된 문서를 페이지 마커라 칭하며, 검출된 각 페이지 마커를 데이터베이스 영상과 패턴비교를 통해 문서 ID를 분석, ID에 맞는 가상 정보를 증강시킨다. 자세한 방법은 아래에서 설명하겠다.

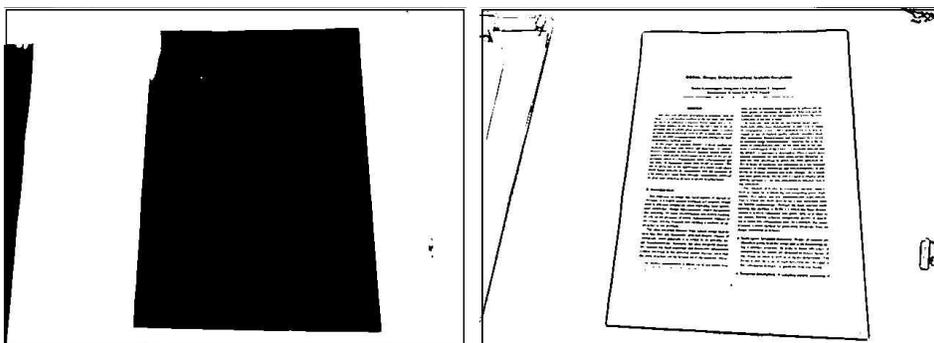


그림 3.1. 제안하는 시스템의 문서 마커화.

3.2 문서 검출

3.2.1 이진화

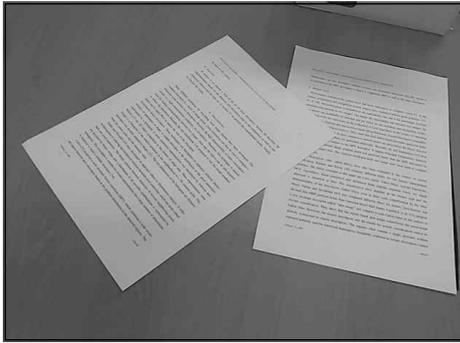
카메라로부터 입력 영상이 들어오면 배경으로부터 문서 경계 영역을 찾기 위해 이진화 과정을 거친다. 하지만 그림 3.2-(a)와 같이 일반적인 이진화를 적용시켰을 경우 페이지의 굴곡에 의한 음영, 조명의 밝기 등의 요소로 인하여 페이지를 배경으로부터 정확하게 분리하기란 매우 제약적이다. 그렇기 때문에 ARToolkit[21]에서 사용하는 적응형 이진화 (adaptive thresholding) 방법이나 Canny[54]을 이용하여 이진화를 한다. 그리고 전처리 과정으로 그림 3.3-(b)와 같이 가우시안 스무딩과 모폴로지 연산을 통하여 잡음 및 텍스처 성분을 제거 한 후 이진화를 적용한다. 하지만 전처리 과정으로 인하여 적응형 이진화의 경우 상대적으로 굵은 경계선이 검출되며 그림 3.4와 같이 배경의 색상 및 조명에 따라 결과 차이가 있기에 적절한 임계값을 설정할 필요가 있다. Canny는 우수한 성능으로 현재도 널리 사용되고 있지만 복잡한 계산으로 인하여 속도면에서는 적응형 이진화보다 느리다는 단점이 있다. 문서 후보 영역 검출 방법에 따라 적절한 이진화 방법을 결정한다.



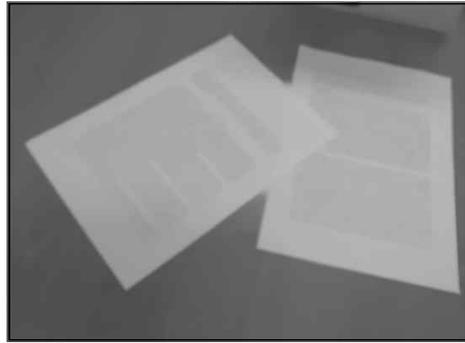
(a) 일반 이진화

(b) 적응형 이진화

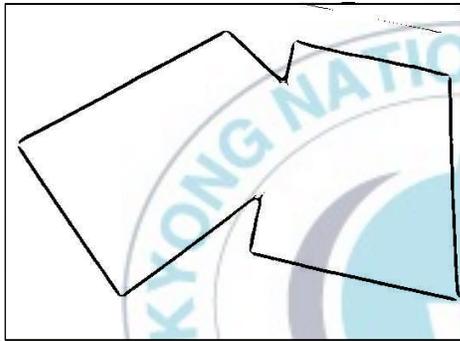
그림 3.2. 일반 이진화와 적응형 이진화 비교.



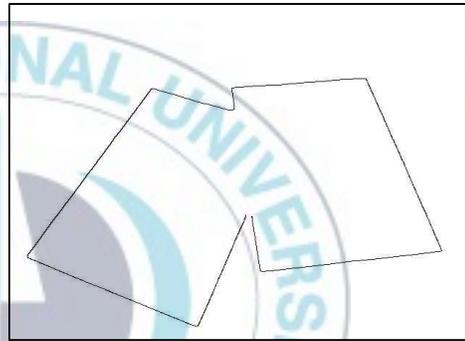
(a) 입력 영상



(b) 스무딩 영상



(c) 적응형 이진화



(d) Canny 이진화

그림 3.3. 적응형 이진화 결과.

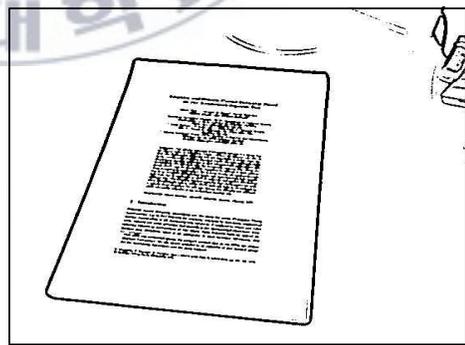
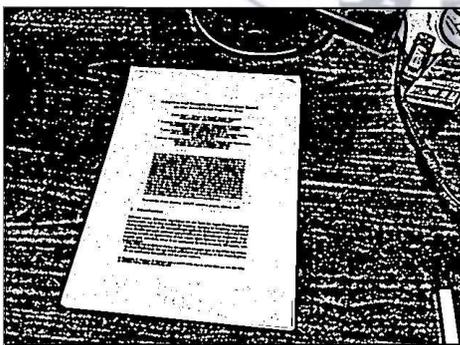


그림 3.4. 적응형 이진화의 임계값에 따른 결과.

3.2.2 문서 후보 영역 검출

이번 단계는 이진화 영상으로부터 경계를 추출, 문서 후보영역을 찾는 과정이다. 일반적으로 문서는 사각형 형태를 띄며, 제안하는 시스템에서도 기존의 마커 기반 증강현실 기술[19-23]과 마찬가지로 영상으로부터 문서 후보로 사각형 영역을 먼저 검출 한다. 사각형 검출 방법으로는 외곽선을 이용한 방법과 허프변환[44]을 이용한 방법을 사용하며, 방법에 따라 장단점이 있기 때문에 상황에 맞는 방법을 결정해야한다.

외곽선 다각화를 이용한 사각형 검출

이진화 과정을 거친 후, 이진화 영상으로부터 폐곡선으로 이루어진 외곽선 영역을 추출한다. 찾아진 여러 외곽선들의 굴곡 부분을 바탕으로 그림 3.5와 같이 외곽선을 다각형으로 근사화한 후 꼭짓점 4개를 가지는 다각형을 사각형으로 검출한다.



그림 3.5. 외곽선 다각화.

외곽선을 이용하는 방법은 간단하면서도 강인하게 사각형 검출이 가능하다는 장점이 있지만 외곽선의 일부분만 가려져도 사각형으로 검출이 불가능하다. 영상 내 문서가 가려짐(occlusion)이 없는 상황에선 매우 효과적으로 사각형 검출이 가능하며, 이진화 방법은 비교적 속도가 빠른 적

응형 이진화 방법을 사용한다.

허프변환을 이용한 사각형 검출

문서가 가려진 경우 외곽선을 이용한 방법은 사각형을 검출이 불가능하다. 가려진 경우에도 사각형을 검출하기 위하여 문서의 경계로부터 허프변환을 사용하여 직선을 추출한 후 직선들의 관계로부터 사각형 후보를 판단한다.

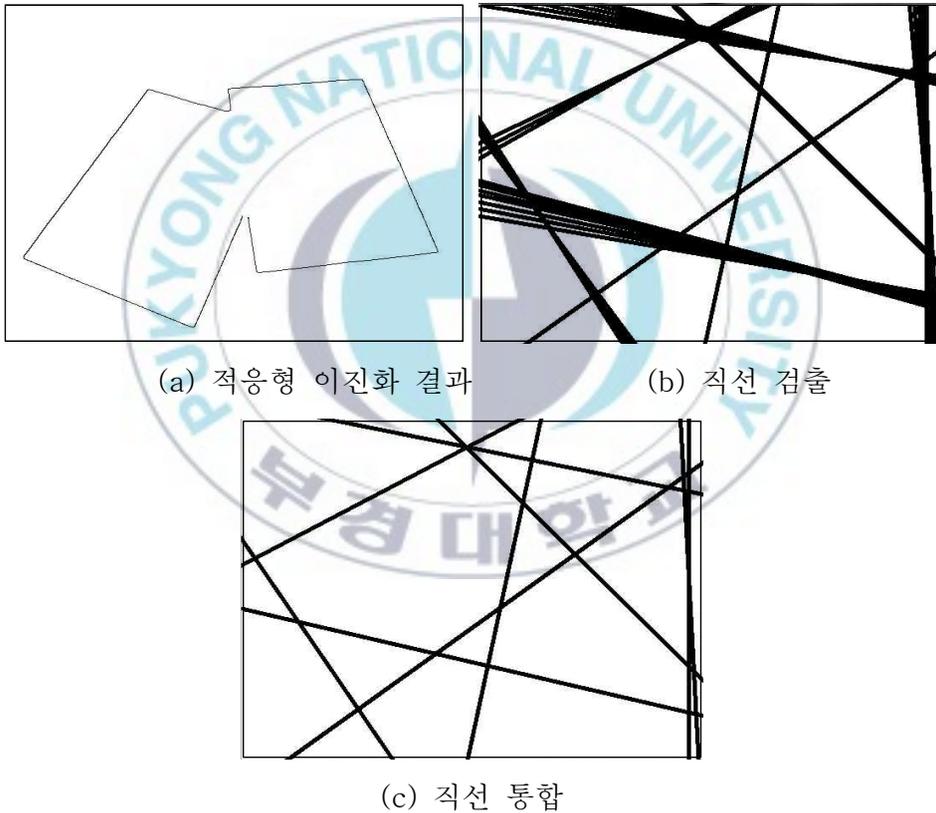


그림 3.6. 허프변환 및 직선 통합.

그림 3.6-(b)는 이진화 영상으로부터 허프변환을 이용하여 검출된 직선들이다. 이진 이진화 단계에서 잡음 제거 및 텍스처 제거를 하기 위해

사용된 가우시안 스무딩 및 모폴로지 연산으로 인하여 문서 경계로부터 하나가 아닌 여러 개의 직선이 검출된다. 이진화 단계에서 적응형 이진화를 사용하는 경우 문서의 경계는 Canny를 사용할 때와 비교하였을 시 상대적으로 굵은 경계를 가지기 때문에 검출되는 직선도 더 많다. 그래서 허프변환을 사용하여 사각형 영역을 검출할 때에는 적응형 이진화 보다 Canny가 적합하다.

검출된 여러개의 직선들로부터 경계를 지나는 하나의 직선만을 얻기 위해 $\rho-\theta$ 좌표계로 표현된 직선(그림 3.7)으로부터 임계값 th_ρ, th_θ 을 설정, 서로 비슷한 ρ, θ 값을 가진 직선들을 평균하여 그림 3.6-(c)와 같이 하나의 직선으로 통합시킨다.

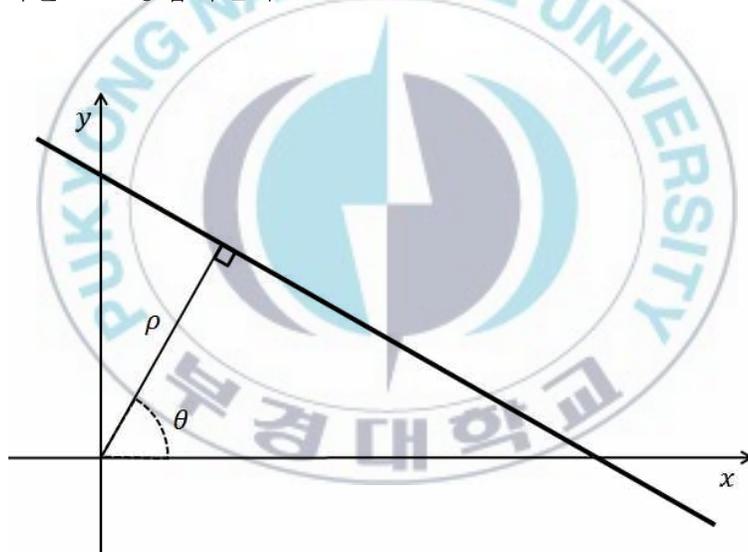
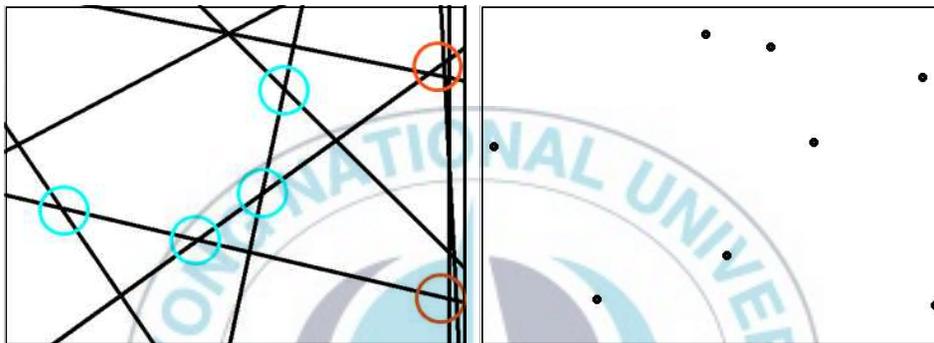


그림 3.7. $\rho-\theta$ 좌표계로 표현된 직선.

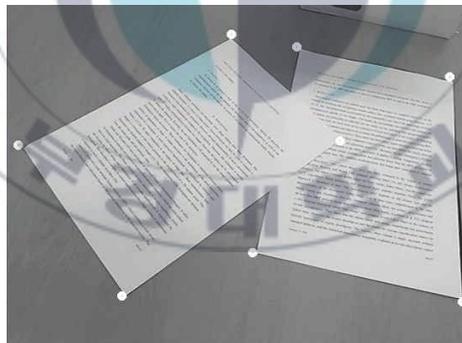
다음으로, 통합된 직선들로부터 교점을 찾는다. 문서 경계를 지나는 직선의 교점은 문서의 꼭짓점이 된다. 하지만 모든 교점을 검출하게 될 경우 문서의 꼭짓점이 아닌 교점들도 검출되며(그림 3.8-(a)의 빨간 원), 또한, 그림 3.8-(b)의 파란 원처럼 직선통합과정에서 통합이 되지 않는 직선들로부터 여러 개의 교점들이 검출이 된다. 하지만 문서의 꼭짓점이 아

닌 교점들은 불필요한 교점들이며, 이러한 교점들은 다음 단계인 문서 영역 검출에 어려움을 준다. 그래서 불필요한 교점들을 최대한 제거하기 위해 교점을 이루는 직선의 사잇각 θ_i 가 일정 범위의 값을 가질 때 ($th_a < \theta_i < th_b$)의 교점만 문서의 꼭짓점 후보로 검출하며, 검출된 교점들 역시 직선과 마찬가지로 거리에 대한 평균을 이용하여 통합시킨다.



(a) 직선 검출 및 통합

(b) 교점 계산



(c) 교점 검출 결과

그림 3.8. 교점 검출 결과.

검출된 꼭짓점들의 경우 직선과 교점 통합과정으로부터 문서의 실제 꼭짓점과 약간의 오차가 생기게 된다. 그래서 꼭짓점을 중심으로 $h(\text{너비}) \times h(\text{높이})$ 크기를 가지는 윈도우를 생성, 윈도우 내에서 해리스 코너 검출방법[55]을 사용하여 꼭짓점을 실제 문서의 꼭짓점으로 보정한

다. 하지만 꼭짓점이 다른 물체 등으로 가려지거나 주변부가 높은 텍스처(texture)를 가질 경우 그림 3.9-(b)와 같이 보정 결과가 더욱 나빠지는 결과가 나타난다. 그래서 윈도우 내부의 평균을 계산하여 평균보다 밝기가 큰 픽셀 수와 밝기가 작은 픽셀 수의 비율을 통하여 해리스 코너 검출 방법을 이용한 교점 보정 유무를 판단한다. 그림 3.9-(c),(d)는 그림 3.9-(a),(b)에 대한 밝기 히스토그램으로서 히스토그램을 보면 보정이 가능한 경우와 불가능한 경우는 서로 다른 분포를 띄고 있음을 알 수 있다.

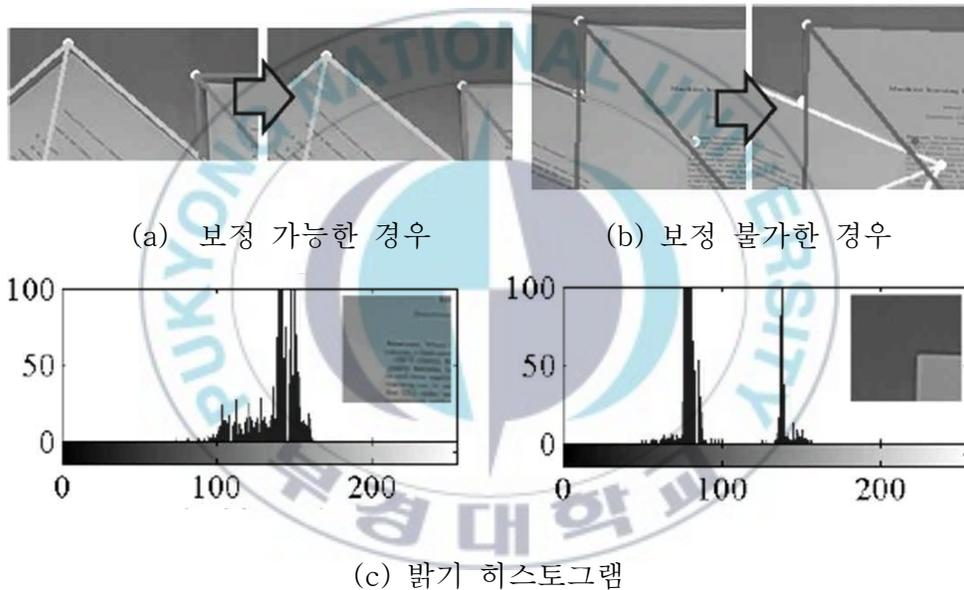


그림 3.9. 교점 보정.

마지막으로, 검출된 꼭짓점 교점들을 4개씩 조합하여 만들어 지는 사각형을 문서 후보 영역으로 판단한다. 하지만 단순히 교점을 조합을 하게 될 경우 그림 3.11-(b)와 같이 검출된 꼭짓점 수(N)에 대하여 굉장히 많은 후보 영역을 가지게 된다(${}_N C_4$). 만약 최종적으로 8개의 교점이 검출되었다더라도 70개의 문서 후보 영역이 검출된다(${}_8 C_4 = 70$). 그래서 다음 3단계의 과정을 거쳐 최종적인 후보영역을 가려낸다. 이는 외곽선을 이용

한 사각형 검출에서도 동일하게 적용한다.

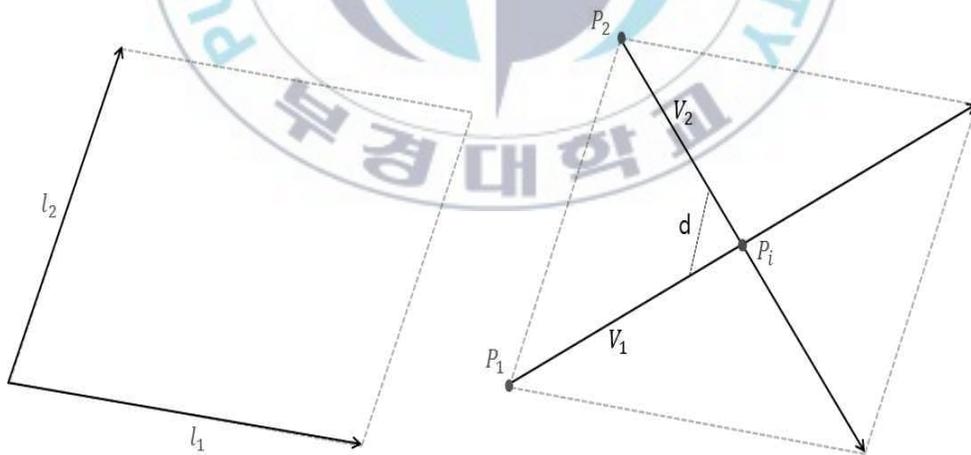
첫째, 검출된 사각 영역의 크기가 th_s 보다 커야한다

둘째, 가로·세로비율(aspect Ratio)가 th_r 보다 작아야한다. 그림 3.10-(a)로부터의 종횡비(aspect Ratio)는 식(3-1)과 같다.

$$Aspect\ Ratio = Max(l_1, l_2) / Min(l_1, l_2). \quad (3-1)$$

셋째, 대각선의 교점이 사각형의 가운데 오는지를 판단하여 볼록한(convex) 모양의 사각형만 가려낸다. 그림 3.10-(b)에서 볼록함(convexity) 계산을 위한 두 직선(= 대각선) l_1, l_2 에 대한 방정식은 다음과 같다.

$$l_1 : P_1 + tV_1, l_2 : P_2 + uV_2, 0 \leq t, u \ll 1. \quad (3-2)$$



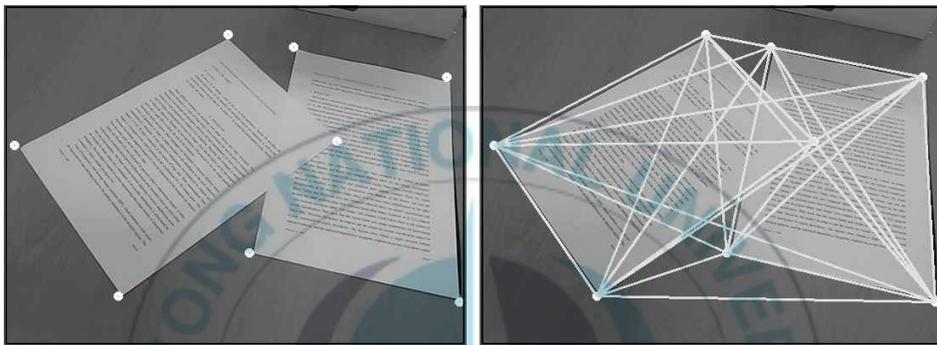
(a) 종횡비

(b) 볼록함

그림 3.10. 종횡비 및 볼록함 계산.

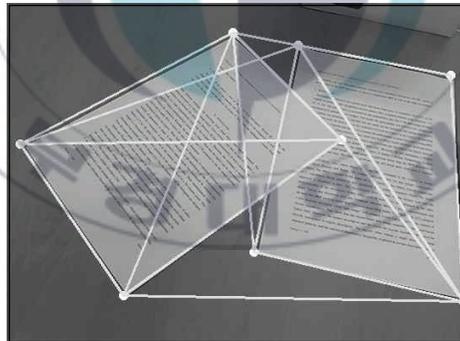
두 직선 위 임의의 점 사이의 거리 d 를 t 와 u 에 대한 함수로 보고, 이

를 최소화하는 t 와 u 값은 두 직선의 교점(P_i)가 된다. 구해진 t, u 의 값이 중점인 0.5에 가까워야하므로, 임계값 $th_c(|0.5 - t, u| > th_c)$ 로부터 불록함을 판단한다. 그림 3.11-(c)는 최종 후보영역의 결과를 보여준다. 실제 문서 영역만 후보로 추출할 순 없지만 상당수의 후보영역들이 없어진 것을 알 수 있다.



(a) 꼭짓점 검출

(b) 문서 후보 영역 검출



(c) 최종 문서 후보 영역

그림 3.11. 문서 후보 영역 검출.

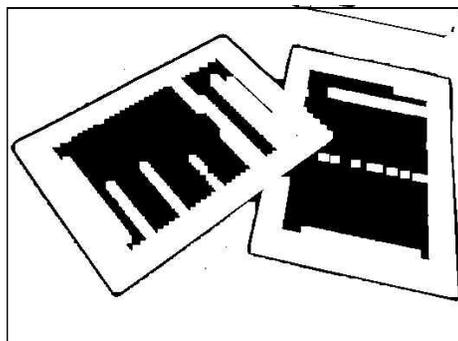
3.3 페이지 마커

최종적으로 문서 후보로 검출된 영역을 페이지 마커로 사용하기 위해 문서 자체를 블록화 시킨다. 먼저, 그림 3.12-(b)와 같이 카메라 입력 영상에 적응형 이진화를 이용하여 문서 내 내용물과 바탕을 분리시킨다. 이때 원활한 마커화를 위해 매우 큰 마스크(mask)를 사용한다. 다음으로 모폴로지 팽창연산과 침식연산의 반복을 통하여 잡음제거 및 글자 등의 내용물들은 두껍게 채워 기존의 마커기반 증강현실에서 사용하는 이진 마커와 유사한 모양의 페이지 마커를 생성한다. 모폴로지 팽창과 침식의 횟수(N_m)는 카메라와의 거리에 따른 유동적 적용이 필요하다.



(a) 입력 영상

(b) 이진화 영상



(c) 마커화

그림 3.12. 입력 영상의 마커화 결과.

3.4 페이지 마커 ID 분석

마커화된 후보 영역들을 먼저, $w(\text{너비}) \times w(\text{높이})$ 의 크기로 와핑을 한다. 그 후 내부 패턴들을 데이터베이스에 저장된 영상들과 비교를 통하여 그들의 ID를 분석한다. 데이터베이스 영상 역시 앞에 과정을 통하여 마커화된 영상을 같은 크기로 와핑한 영상이다. 패턴비교를 위해서 기존에는 템플릿매칭 기법이 많이 사용되었었다. 하지만 템플릿 매칭을 통한 픽셀단위 유사성을 비교하는 방법은 연산속도가 너무 느려 실시간으로 처리하기에는 무리가 있으며, 패턴이 가려질 경우 인식이 불가능하다는 단점이 있다. 그래서 제안된 시스템에서는 그림 3.13와 같이 마커화된 영역을 일정한 크기로 나누고 나뉜 영역안의 흰 영역과 검은 영역의 비율을 이용하여 매칭을 한다.

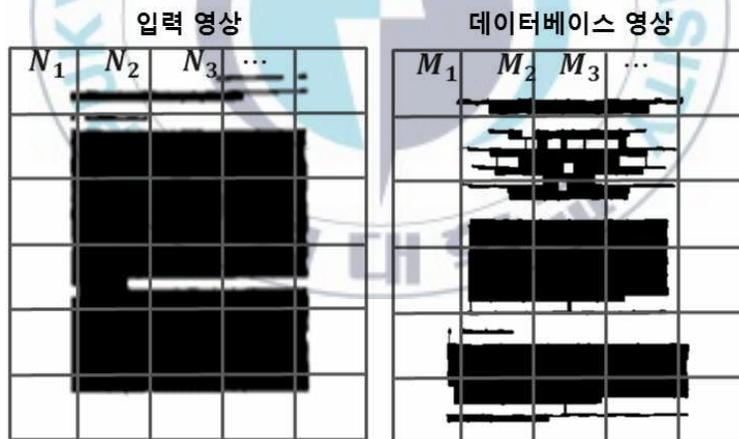


그림 3.13. 패턴 매칭을 위한 영상 분할.

먼저, 와핑된 마커화 영상을 L 개의 서브영역으로 나눈 후 각 영역의 검은 픽셀의 수를 계산한다(그림 3.13). 계산된 검은 픽셀의 수를 데이터베이스 영상과 순차적으로 영역별로 비교를 한다. ID 분석 과정은 다음과 같은 식으로 나타낼 수 있다.

$$k = \operatorname{argmin} \sum_{i=1}^L \{d_i = |N_i(k) - M_i(k)|\}, k = 1, 2, \dots, K. \quad (3-3)$$

K 는 데이터베이스 내의 ID, 즉 데이터베이스 영상의 번호를 뜻하며 $N_i(k)$ 와 $M_i(k)$ 는 k 번째 서브영역의 검은 픽셀의 수를 뜻한다. 하지만 위의 과정의 경우 잡음이나 가려짐 등은 d_i 의 합에 영향을 끼치고 그로 인하여 ID를 오인식하는 경우가 생길 수 있다. 그렇기 때문에 임계값 th_d 를 설정, d_i 의 값이 th_d 보다 클 경우 그 때의 d_i 의 값은 무시한 후 th_d 보다 작은 d_i 만의 합을 구한 후 평균으로 나타낸다(식(3-4)).

$$d_i = \begin{cases} 0, & \text{if } d_i > th_d \\ d_i/D, & \text{else.} \end{cases} \quad (3-4)$$

이 때 D 는 d_i 가 th_d 보다 작은 서브영역의 개수이다.

3.5 성능 개선

기존의 마커기반의 증가현실기술들은 연결요소(connected component)나 외곽선 다각화로부터 마커영역이나 사각영역을 검출한다. 하지만 이 방법들은 영상의 일부분이 가려지는 경우 사용할 수가 없다. 반면에 허프변환을 이용하는 방법은 직선의 교점을 문서의 꼭짓점으로 판단하여, 꼭짓점 조합을 통해 사각 영역 검출하기 때문에 가려지거나 훼손된 경우에도 사각 영역의 검출이 가능하다. 그러나 그림 3.12에서 보이는 것처럼 적은 수의 꼭짓점의 조합에서도 상당한 개수의 마커 후보 영역이 검출되고 있으며, 순차적으로 데이터베이스 영상과 비교를 하는 과정에서 많은 계산량이 요구되고, 실제 문서 영역이 아닌 후보 영역도 매칭이 되는 경

정오류(false positive)가 발생한다.

이를 보완하기 위해 크기, 종횡비(aspect ratio), 볼록함(convex)의 조건을 이용하여 후보 영역의 개수를 감소시켰지만 여전히 상당수의 후보 영역들이 존재한다(그림 3.12). 이러한 문제점을 해결하기 위하여 후보 영역들을 순차적으로 데이터베이스 영상과 비교를 하되 매칭 유사도가 크게 떨어질 경우 매칭에서 제외시켜 여러 문서 후보 영역 중에서 실제 문서 영역만 매칭이 되게 하여 긍정오류를 줄일 수 있다. 그림 3.14은 입력 영상과 데이터베이스 사이의 유사도 d_i (식(3-4))의 분포이다. 같은 패턴을 가진, 즉 매칭이 되는 경우 d_i 값의 평균은 약 4.96으로 나타나며 매칭이 되지 않는 영상들 간에는 약 31.15의 평균값을 보인다. 이를 바탕으로 d_i 에 대한 임계값 th_{d_i} 로 매칭 유사도가 떨어지는 문서는 매칭에서 효과적으로 제거가 가능하다. 추가적으로 본 시스템에서 사용하는 서브영역의 크기는 14×14 로서 k (식(3-3))는 최대 196까지 나타날 수 있기에 이론적으로 16%(31.15/196) 이상의 가려짐이나 두 개의 패턴의 차이가 2.1%(4.16/196) 이하일 때는 인식률이 상당히 떨어지게 된다.

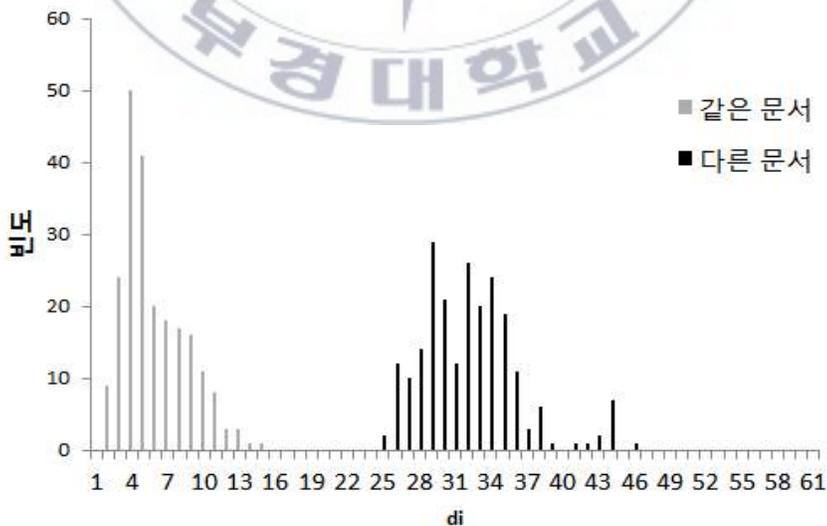


그림 3.14. 유사도 d_i 의 분포.

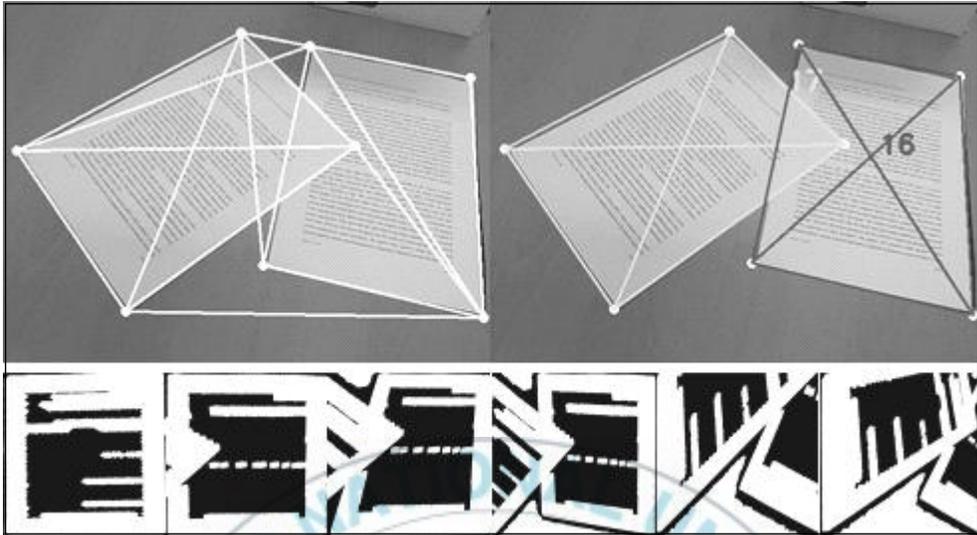


그림 3.15. 마커화한 최종 사각형 후보 영역 및 인식 결과.

그림 3.15는 총 6개의 최종 사각형 후보들을 마커화 한 영상으로부터 d_i 임계값을 통하여 문서 영역만 제대로 인식됨을 보여준다. 추가적으로 그림 3.16과 같이 문서 후보 영역들 중 하나가 데이터베이스 영상과 매칭이 된 경우 매칭된 후보 영역이 포함한 꼭짓점에서 조합이 가능한 모든 후보 영역을 제외시킴으로써 후보 영역들과 데이터베이스 영상과의 비교 횟수의 감소를 통한 속도 향상이 가능하다.

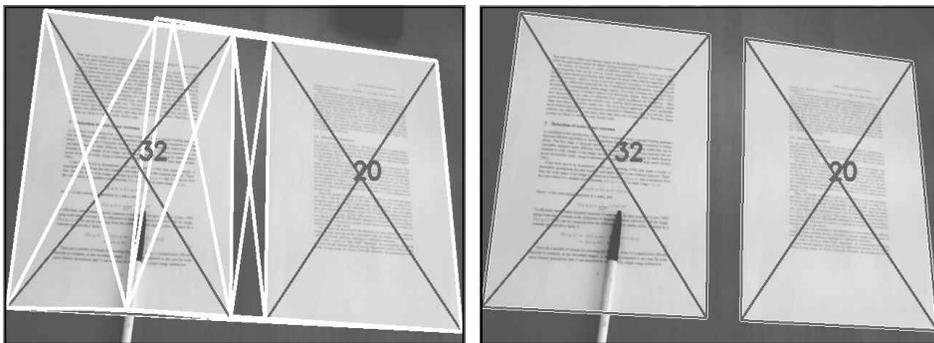


그림 3.16. 중복 꼭짓점 제거 결과.

IV 실험결과 및 고찰

4.1 실험 환경

제안하는 방법의 결과를 확인하기 위한 데이터베이스 영상은 마커화 후 와핑한 영상으로 이루어져 있으며, 본 방법은 회전에 불변하지 않기 때문에 각 문서는 $0^\circ, 90^\circ, 180^\circ, 270^\circ$ 의 4 방향에서 문서로부터 수직(90°)으로 45cm 떨어진 거리에서 획득, 총 227장(56×4)으로 구성되어 있다 (그림 4.1). 실험은 공개 라이브러리 OpenCV[47]와 OpenGL[56]를 사용해 구현하였으며, i5 2.5GHz CPU, 4GB RAM을 탑재한 랩탑 환경에서 수행하였다.

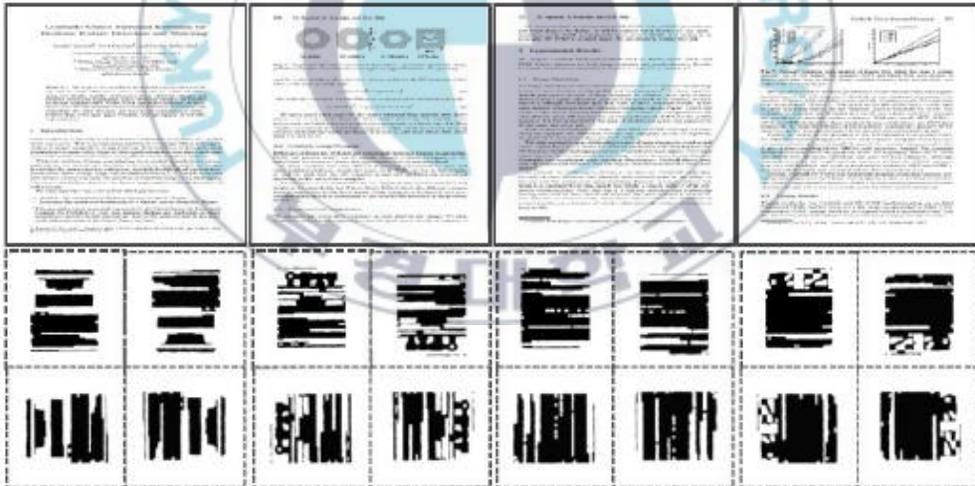
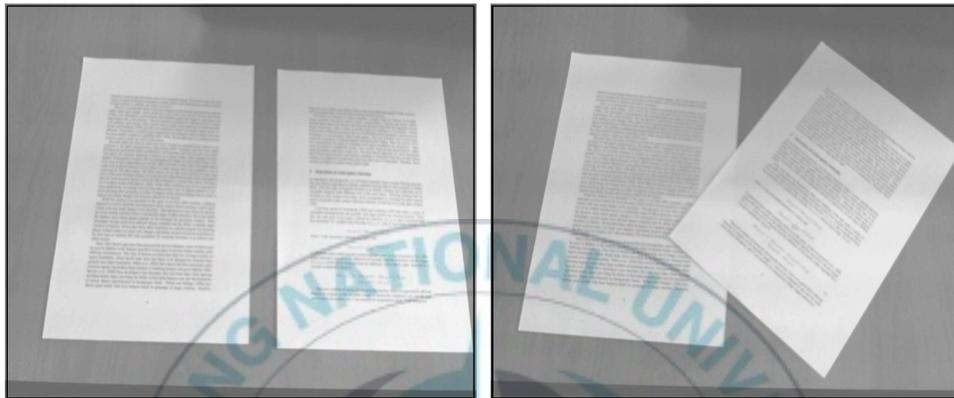


그림 4.1. 데이터베이스 영상 구성.

4.2 실험 과정

실험은 그림 4.2와 같이 영상 내 다수의 문서가 존재하거나, 문서의 경

계나 내부 내용물이 가려졌을 때 등 다양한 환경에 대하여 문서로부터 카메라의 거리와 각도를 각각 30-60cm, 수직으로부터 40°-90°로 변화시키면서 실험을 수행하였다.



(a) 다수 문서 영상

(b) 경계가 겹쳐진 문서 영상



(c) 내부가 가려진 문서 영상

그림 4.2. 입력 영상 구성.

표 4.1는 각 처리 과정 별 소요시간을 보여준다. 먼저, 마커 검출에서 스무딩은 이진화 전 전처리 과정인 가우시안 스무딩과 모폴로지 연산에 소요되는 시간이며, 문서 경계 검출은 Canny와 허프변환을 통하여 직선 검출 및 직선 통합에 소요되는 시간이다. 교점 검출 및 보정의 경우 검출

된 교점수 I^* 에 따라 선형적으로 소요 시간이 증가한다. 문서 후보 영역 검출은 교점을 4개씩 조합한 후 여러 가지 조건을 통하여 최종적인 문서 후보 영역을 결정하는데까지 소요되는 시간을 의미한다. 마커 ID 인식 과정에서 와핑은 검출된 문서 후보 영역을 데이터베이스와 비교하기 위해 와핑하는 과정으로 검출된 문서 후보 영역수 Q^* 에 따라 소요 시간이 결정되며, 패턴 비교는 영역수(L)가 100개 일 때 소요 시간이다.

제안하는 시스템에서 실험을 위해 사용된 각종 파라미터들은 표 4.2에 나타내었다. 적응형 이진화 마스크 크기는 전체 영상을 마커화의 전처리 과정으로 내부 텍스트를 효율적으로 블록화하기 위해 매우 큰 마스크 크기를 가진다. 몇몇 파라미터들은 카메라의 거리나 주변 환경의 영향을 받기 때문에 동적인 파라미터 설정이 필요하다. 실험 환경에서는 약간 제약을 두어 유동성이 필요한 파라미터들은 고정 후 실험하였다.

표 4.1. 처리 과정별 소요시간

처리 과정		시간 (ms)
마커 검출	스무딩	1.0
	문서 경계 검출	7.0
	교점 검출 및 보정	$0.1I^*$
	문서 후보 영역 검출	5.0
	마커화	11.5
마커 ID 인식	와핑	$0.2Q^*$
	패턴 비교	$0.1Q^*$

표 4.2. 제안하는 증강현실 시스템의 각종 파라미터

파라미터	설명	값
s	적응형 이진화 마스크 크기	71
th_ρ, th_θ	직선 통합시 임계값	10, 10
th_a, th_b	교점 검출시 두직선이 이루는 각 $th_a < \theta_i < th_b$	60, 120
th_s, th_r, th	문서 후보 영역 조건에 대한 임계값	20000, 0.35, 2
h	교점 보정시 해리스 코너 윈도우 크기	48
N_m	마커화시 모폴로지 팽창, 침식 연산 적용 횟수	3
w	ID 분석시 와핑 영상 크기	100
L	ID 분석시 서브 영역 수	100
th_d	ID 분석시 유사도 d_i 임계값	147
th_{d_i}	ID 분석시 매칭 유사도 임계값	20

그림 4.3은 제안하는 방법의 처리 과정을 순서대로 보여준다. 문서의 경계 검출에 있어서 잡음이나 텍스처가 허프변환 시 직선성분으로 검출되는 것을 막기 위하여 가우시안 스무딩 및 모폴로지 연산 후 문서 경계로부터 직선 및 교점을 검출한 후 평균을 이용하여 통합한다. 다음으로 검출된 교점들을 4개씩 조합을 하여 여러 가지 임계값을 통하여 문서 후보를 최종적으로 결정한 후 검출된 문서영역 후보들로부터 성공적으로 ID를 검출하여 ID의 따른 가상의 콘텐츠를 성공적으로 삽입한 결과 영상을 보여준다.

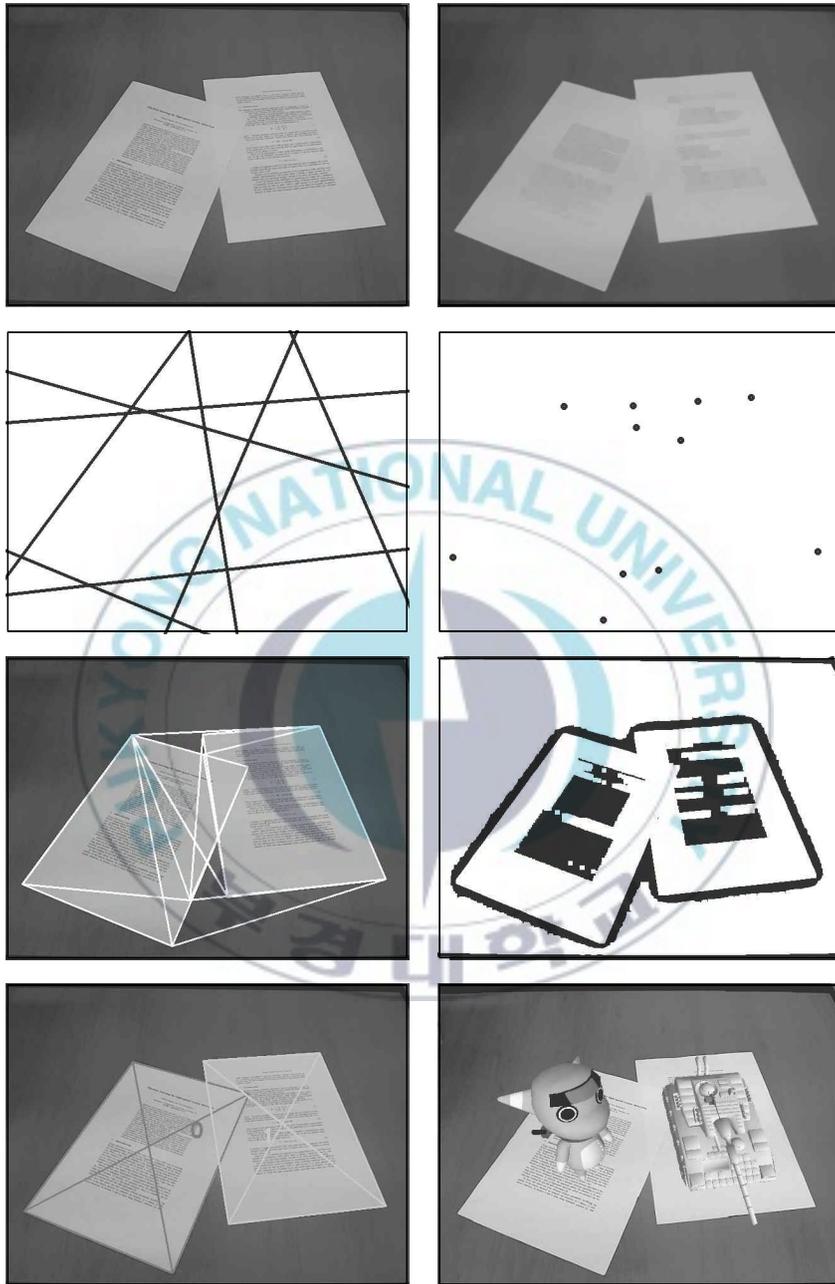


그림 4.3. 제안하는 증강현실 시스템의 처리 과정. 좌상부터 (a)입력 영상, (b)스무딩영상, (c)경계검출, (d)교점계산, (e)사각영역검출, (f)페이지 마커화, (g)마커 ID분석, (h)가상 콘텐츠 증강.

4.3 결과 및 고찰

그림 4.3은 제안하는 시스템에서 문서로부터 카메라 거리가 40cm와 60cm로 고정된 후 문서의 상하좌우에서 70°와 50°의 각도로 바로 보았을 때의 페이지 마커 ID 인식률의 결과이다. 데이터베이스 영상을 구성할 때의 카메라 위치(45cm, 90°)와 유사한 위치(40cm, 90°)에서는 100%의 인식률을 보이고 있으며 카메라 위치에 따라 최저 79%로 평균 91%의 인식률을 보인다.

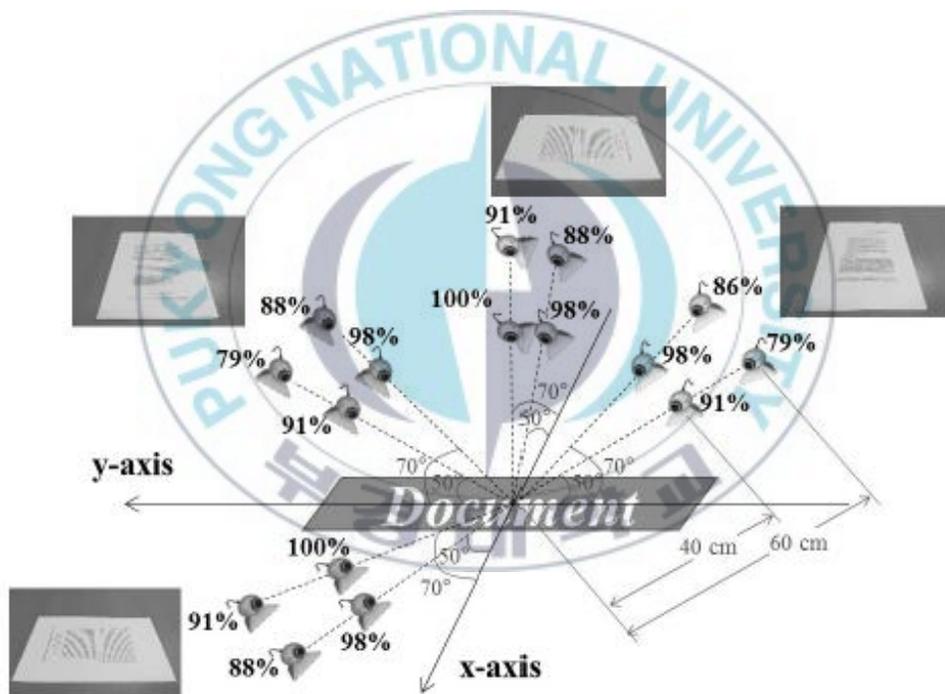


그림 4.4. 거리 및 카메라 각도에 따른 ID 인식률.

인식률을 보면 거리가 멀어지거나 문서와 카메라의 각도가 작아질수록 나빠지는 것을 알 수 있다. 이는 그림 4.5과 같이 거리나 각도의 변화가 커질수록 문서 내 블록들이 뭉쳐지는 현상이 발생해 거리가 멀어질수록

점차적인 인식률 감소가 나타난다. 특히, x축보다 y축 방향으로의 변화에 대해 인식률이 더욱 민감함을 알 수 있는데 문서 내 텍스트 정보들은 주로 가로방향으로 삽입이 되어 있어 블록들의 뭉쳐짐 역시 y축 방향의 변화에 더욱 민감하기 때문이다. 거리나 각도에 따라 적응형 이진화의 마스크 크기나 모폴로지 팽창, 침식 연산의 횟수 등에 관한 파라미터를 유동적으로 적용한다면 인식 결과는 보다 좋아 질 것이다.

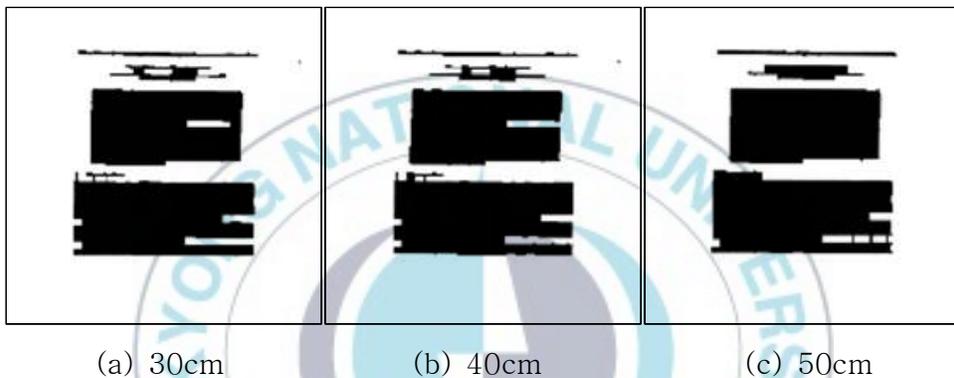
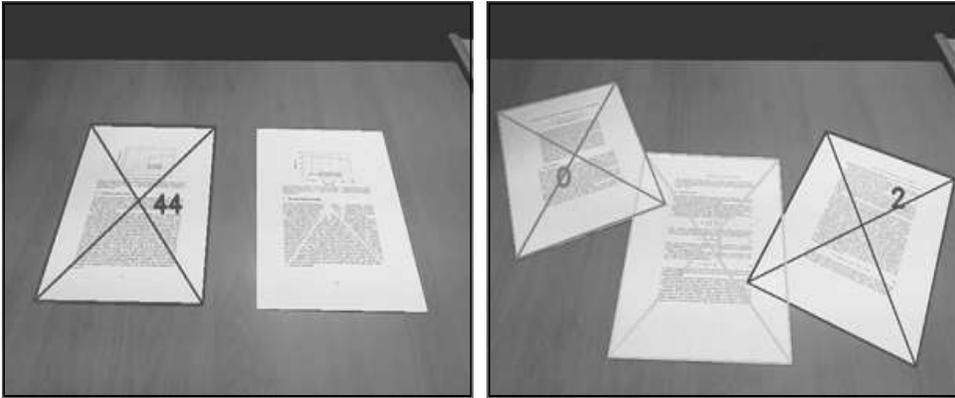
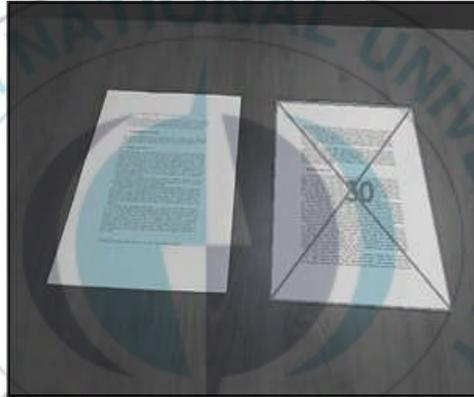


그림 4.5. 거리에 따른 블록 모양 변화.

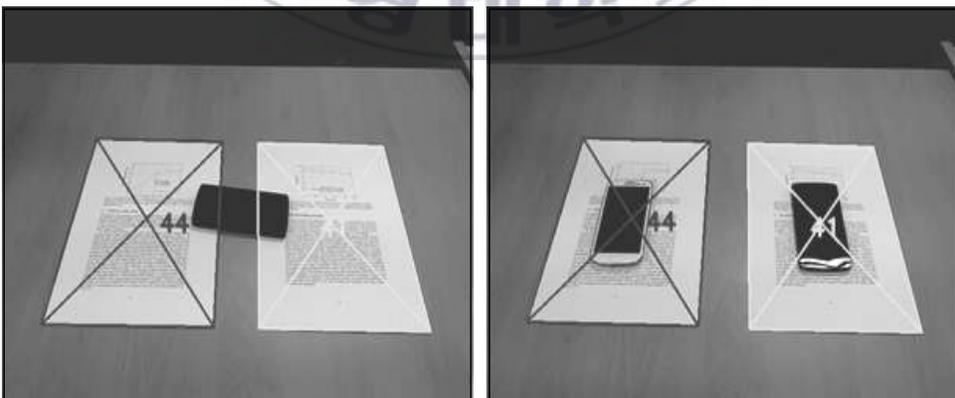
그림 4.6는 여러 가지 상황에서의 인식 결과이다. 그림 4.6-(b)의 왼쪽 문서는 데이터베이스에 저장되지 않은 문서로서 데이터베이스에 저장된 문서만 인식하는 것을 보여주며, 그림 4.6-(c)는 문서의 일부가 가려진 상황에서도 인식이 가능함을 보여준다. 반대로 그림 4.7는 인식이 어려운 상황에 대한 예시이다. 그림 4.7-(a)의 좌영상처럼 두 문서가 나란히 붙어 있어 하나의 문서로 판단되거나 우영상처럼 문서의 경계로부터 직선을 검출하지 못할 경우와 그림 4.6-(c)와 같이 많은 부분이 가려질 경우 제대로 인식이 어렵다. 또한, 그림 4.7-(d)처럼 전처리 과정에서 제거되지 않고 남은 내부 텍스처로 인하여 생긴 문서 후보 영역으로 인하여 중복 인식되는 경우가 있다.



(a) 다수 문서 인식

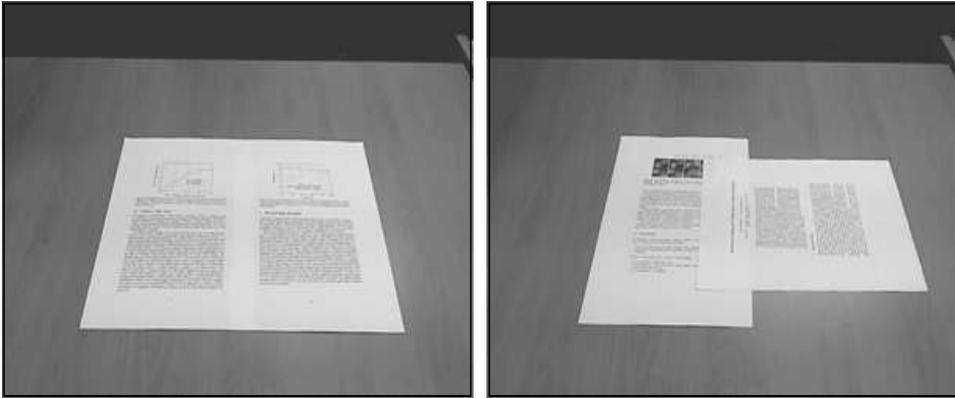


(b) 데이터베이스에 존재하는 문서만 인식



(c) 가려진 문서 인식

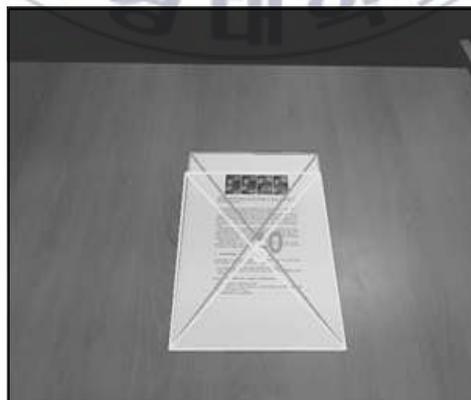
그림 4.6. 여러 가지 상황에서의 인식 결과.



(a) 문서 후보 영역 검출 불가



(c) 심한 가려짐



(d) 중복 인식

그림 4.7. 문서 검출 및 인식이 어려운 경우.

4.4 추가 실험

4.4.1 서브영역 개수(L)에 따른 인식률

표 3. 서브영역 개수에 따른 제안된 방법의 인식률

영역 수(L)	인식률(%)
64	88
100	95
144	97
225	97
400	97

제안하는 방법에서는 영상을 서브영역으로 나눠 각 서브영역의 비율 차이를 이용하여 마커 ID를 분석한다. 그렇기 때문에 나누는 서브영역의 개수가 적다면 비교에 소요되는 시간도 감소된다. 표 3은 서브영역의 개수에 따른 인식률 보여준다. 서브영역 개수가 증가함에 따라 인식률도 증가함을 보이지만 일정 개수 이상에서는 변화가 없음을 알 수 있다 [57]. 영역을 격자뿐만 아니라 다양한 형태로 구분하는 방법에 대한 추가 연구도 필요할 것으로 사료된다.

4.4.2 인덱스를 이용한 페이지 마커 ID 분석

제안하는 방법은 마커 ID 분석을 위해 마커화된 영역을 서브영역으로 나누고 영역안의 흰 영역과 검은 영역의 비율 차이를 이용하여 데이터베이스 영상과 매칭을 한다. 하지만 그림 4.8과 같이 잡음 등의 영향으로 시간에 따라 블록의 모양은 약간의 변화를 가진다. 이로 인한 서브영역간의 비율 차이의 영향을 줄이기 위해 인덱스(index)를 이용한 패턴비교 방법을 적용시켜 보았다[58]. 그림 4.9와 같이 마커화된 영역을 일정한 크기의 서브영역으로 나누고 각 서브영역 내의 흰 영역과 검은 영역의 비율

을 계산한 후, 미리 주어진 룩업(lookup)테이블을 참조하여 인덱스 값을 얻는다. 각 서브영역에서 얻어진 인덱스 값을 정렬함으로써 각 페이지는 서브 영역의 개수의 인덱스를 가지는 코드로 나타난다. 데이터베이스 영상과 코드를 비교함으로써, 각 페이지를 고속으로 분석할 수 있다. 인식률은 최대 75%로서 제안하는 방법보다 떨어지는데, 인덱스를 계산하는 영역의 크기와 룩업테이블의 인덱스 값의 범위에 따라 인식률 차이를 보인다. 예를 들어, 영역을 세밀하게 나눌수록 오매칭은 줄어드나 인식을 하지 못하는 경우는 증가한다. 추후, 인식률 향상을 위한 서브영역이나 룩업테이블을 보다 정교하게 생성하는 등의 추가 실험이 더 요구된다.

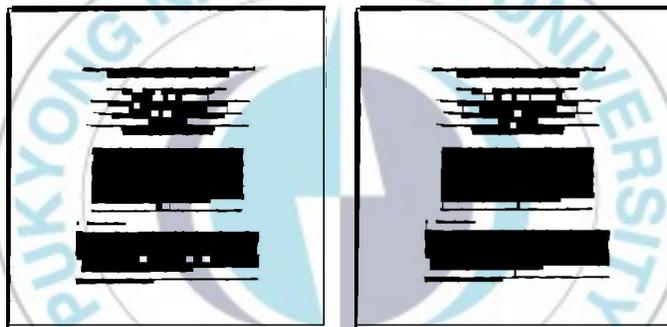


그림 4.8. 시간에 따른 블록 모양 변화.

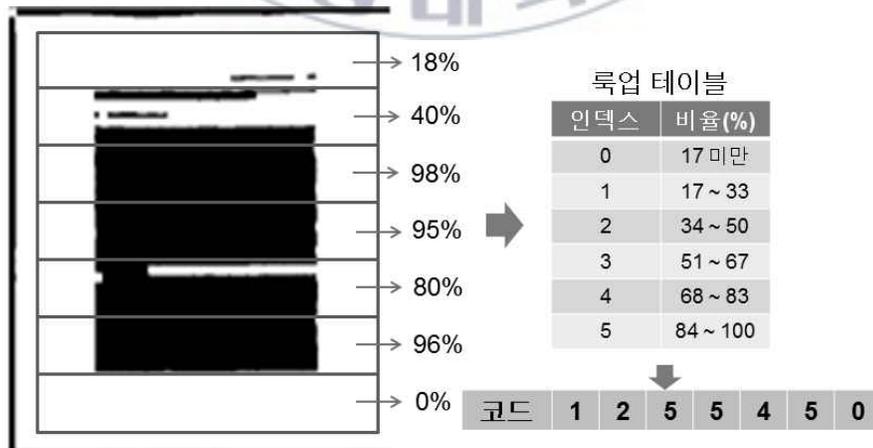


그림 4.9. 인덱스를 사용한 코드화.

4.4.3 특징점 기반 페이지 ID 분석

기존의 마커 기반 아닌 특징점 기반의 증강현실은 마커를 삽입하지 않고 증강현실을 구현할 수 있는 좋은 방법 중 하나이다. 하지만 텍스처(texture)가 열악한 환경에서 사용이 어렵기 때문에 그림이나 사진 위주의 문서에만 그 적용이 용이하다는 하다는 단점이 있다. 추가적인 실험으로 제안하는 시스템에서 사용한 데이터베이스 영상(그림 4.1)에 기존 특징점 기반 매칭방법들인 SURF[24],BRISK[25],ORB[59]을 이용하여 페이지의 ID를 분석해보았으며 그림 4.10는 그 인식률 결과이다. 인식률은 사용하는 특징점 수가 증가할수록 일정량까지는 지속적인 증가를 보이며, 인식률로만 보았을 시 최대 82%까지 나타난다. 하지만 특징점 수가 증가함에 따라 ID 인식 시간도 함께 증가하며, 기존의 빠른 특징점 매칭 방법으로 잘 알려진 BRISK[25] 방법 역시 제안하는 시스템에 비해선 굉장히 느린 처리시간을 보인다. 증강현실을 구현하기 위해서는 그림 4.11과 같이 참조평면과의 호모그래피 관계를 계산한 후, 2.3.1.3절에서 설명한 마커 기반과 마찬가지로 카메라 위치를 추정할 수 있어야한다. 하지만 그림 4.12을 보면 문서의 ID 인식은 가능할지라도 특징점으로부터 호모그래피 계산은 불가능해서 결과적으로 텍스트가 열악한 글자 위주의 문서에는 특징점 기반 증강현실이 적용은 어려움을 알 수 있다.

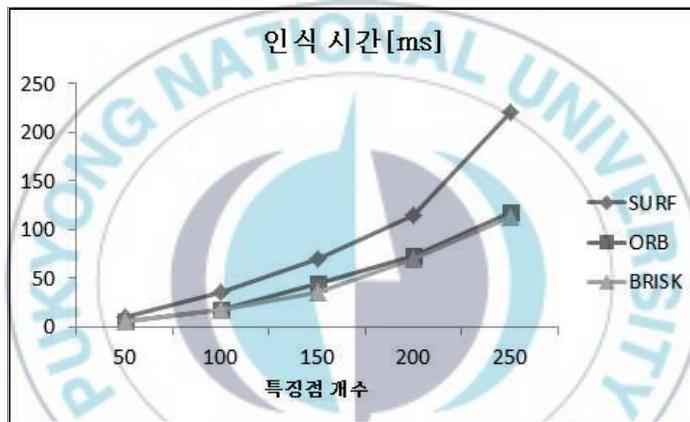
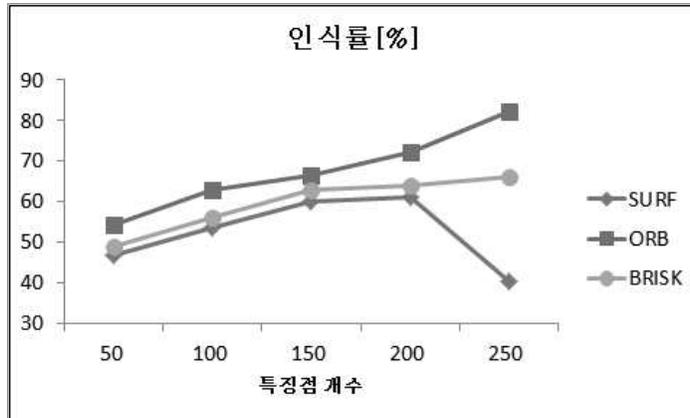


그림 4.10. 특징점 기반의 ID 인식률 및 처리 시간.



그림 4.111 참조 평면 내 특징점 사이의 매칭 및 호모그래피 추정.

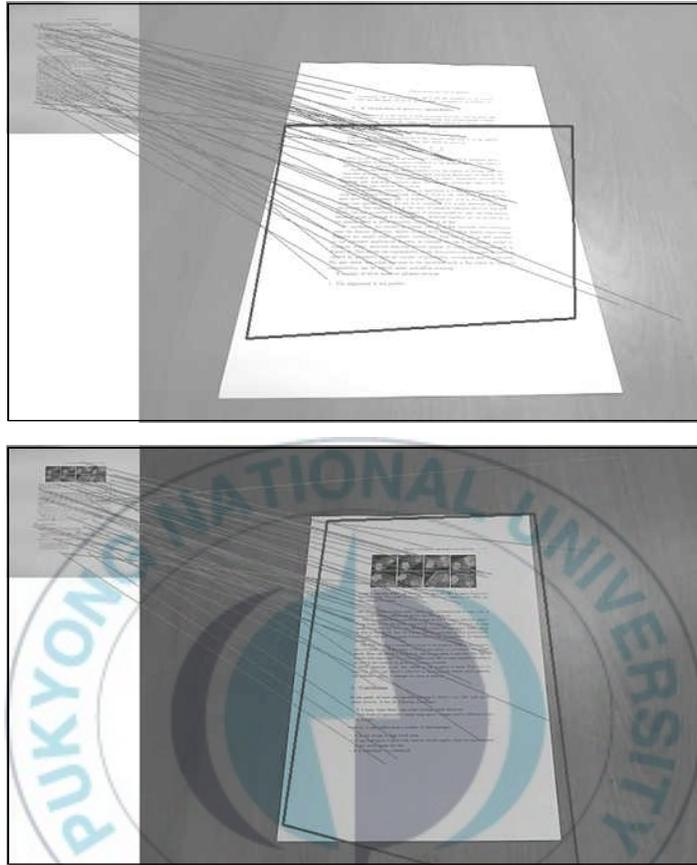


그림 4.12. SURF와 RANSAC을 이용한 문서 영역 호모그래피 추정.

4.4.4 그림 문서에 대한 실험

기존의 특징점 기반 방법으로는 텍스처가 열악한 텍스트 위주의 문서 인식은 효과적이지 않음을 앞의 추가실험을 통해 알 수 있다(4.4.3절). 하지만 제안하는 방법의 경우 문서의 내용물들을 블록화시켜 패턴을 비교하고 있으며, 문서 내부에 그래프나 사진 등의 텍스처를 가진 내용물들이 존재할 때도 높은 인식률을 보여주고 있다. 그러나, 텍스트 위주 문서 이외의 문서에 대한 제안된 방법의 성능을 명확하게 분석하기 위해 텍스처가 많이 포함된 그림 위주의 문서에 제안하는 방법을 적용시켜 그 결과를

살펴보았다.

그림 4.13는 실험에 사용된 데이터베이스 영상과 마커화 영상이다. 136(34*4)장의 그림 문서를 데이터베이스로 사용하였으며, 약 85%의 인식률을 보였다. 결과적으로 텍스트 위주 문서에 비해서는 약간 떨어지지만 여전히 높은 인식률을 보인다. 하지만 그림 4.14-(a)에서 문서의 우측 위 부분처럼 문서 내부의 텍스트처로 인해 제안하는 방법의 해리스 코너 검출 방법을 통한 교점 보정이 적용이 어려우며 그림 4.14-(b)와 같이 내부 텍스트처로 인하여 생기는 문서 후보 영역으로 인하여 중복 인식되는 경우가 기존의 텍스트 위주의 문서보다 더 잦게 나타난다. 이러한 문제점들을 보완한다면 텍스트 위주의 문서뿐만 아니라 문서의 종류에 제약없이 제안하는 방법을 통해 문서 인식이 가능할 것으로 보인다.

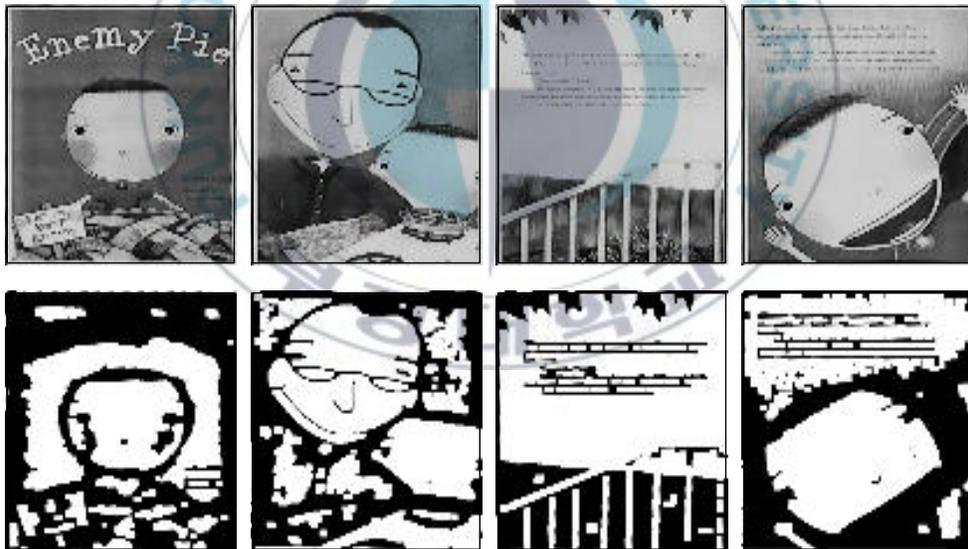
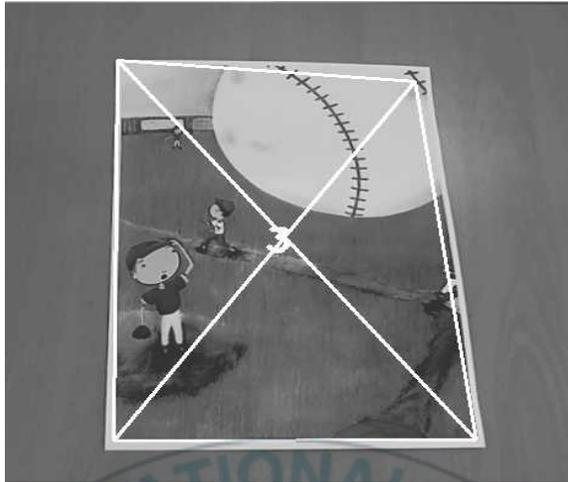


그림 4.13. 그림 위주의 문서 데이터베이스 영상 구성



(a) 해리스 코너 검출을 이용한 교점 보정 실패



(b) 문서 내 텍스처로 인한 중복 인식

그림 4.14. 그림 위주의 문서에서 발생하는 문제점.

V 결 론

증강현실 책을 구현하기 위한 일반적인 기술인 마커 사용 혹은 특징점 기반 방법은 제작이나 사용 상의 불편함, 글자(text) 위주의 책에는 적용 불가 등의 약점을 가진다. 본 논문에서는 증강현실 책을 구현하기 위한 새로운 기술로서, 페이지 자체를 마커화하여 증강현실을 구현하는 페이지 마커 기반 증강현실 방법을 제안하였다. 대상 문서로, 주로 글자 위주로 구성된 논문을 이용한 실험에서, 다양한 거리와 각도를 가지는 입력 영상으로부터 각 문서의 영역을 정확하게 검출하고, 약 91%의 인식률로 인식할 수 있음을 확인하였다. 제안된 방법은 문서 내의 글자나 단어를 개별적으로 인식할 수 없는 환경, 즉 기존의 문서 인식 및 글자 인식 방법의 적용이 불가능한 환경에서도 효과적이었다. 또한, 기존의 마커 기반 방법과 달리 페이지 마커의 일정 부분이 가려진 경우에도 성공적으로 증강현실 책을 구현하는 것이 가능했다.

제안하는 방법은 문서의 종류에 구애를 받지 않기에 글자 위주의 잡지나 교과서에 내용과 관련된 가상 콘텐츠를 삽입함으로써 독자의 이해도를 높이거나 소설책의 각 장면에 대한 애니메이션을 삽입하여 독자의 흥미와 몰입감을 상승시키는 등 기존의 증강현실 책의 활용 범위를 크게 확대할 수 있을 것으로 기대된다.

VI 향후 계획

본 논문에서 제안하는 방법의 문서 인식률은 약 91%로 높은 인식률을 보이나, 문서 후보 영역 검출 시 검출되는 교점이나 후보 영역의 개수에 따라 수행 시간은 선형적으로 증가하며 문서와의 거리 및 각도, 조명과 같은 주변 환경에 따라 일정하지 않은 블록화 영상을 보인다. 그렇기 때문에 보다 정확한 교점과 문서 후보 영역 검출을 통한 수행 속도 향상과 제안하는 방법에 사용되는 각종 파라미터 값들의 유동적 적용을 통한 인식률 향상 대한 연구가 필요하다. 또한, 본 논문에서는 문서 인식에 있어 문서를 평면으로 제약을 두고 실험을 하였으나, 실용성을 높이기 위해 실제 책과 같이 곡면의 페이지를 검출 및 인식하는 방법에 대한 연구를 할 계획이며, 4.4절에서 추가적으로 실험한 인덱스를 이용한 방법과 그림 위주의 문서에 제안하는 방법을 적용할 시 나타나는 문제점들에 대한 추가적인 연구가 필요하다.

참 고 문 헌

- [1] D. Bradley, “10 Emerging technologies 2007,” *MIT Technology Review*, 2007.
- [2] P. Milgram and F. Kishino, “A Taxonomy of mixed reality visual display,” *IEICE Trans. on Information and Systems*, vol. E77-D, no. 12, pp. 1321-1329, 1994.
- [3] R. Azuma, Y. Bailiot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, “Recent advances in augmented reality,” *Computer Graphics and Application*, vol. 21, pp. 34-47, 2001.
- [4] *Mobile Augmented Reality : Smartphones, Tablets and Smart Glasses 2013-2018*, Juniper Research, 2013.
- [5] M. Billinghamurst and A. Dunser, “Augmented reality in the classroom,” *Computer*, vol. 45, no. 7, pp. 56-63, 2007.
- [6] 정상현, 계보경, “증강현실(Augmented Reality) 콘텐츠의 교육적 적용,” *한국콘텐츠학회지*, vol. 5, no.2, pp. 79-85, 2007.
- [7] 김진숙, “교실수업 적용을 통한 증강현실 콘텐츠의 교육 효과 분석,” *대한전자공학회 하계종합학술대회 논문집*, pp. 808-809, 2009.
- [8] 김영옥, “증강 현실(AR) 기술이 접목된 그림책으로 보는 상상력의 진화,” *인문콘텐츠*, vol 19, pp. 465-489, 2010.
- [9] R. Grasset, M. Billinghamurst, A. Dünser, and H. Seichter, “The mixed reality book: a new multimedia reading experience,” *CHI Extended Abstracts*, pp. 1953-1958, 2007.
- [10] K. Kiyokawa, H. Iwasa, H. Takemura, and N. Yokoya, “Collaborative immersive workspace through a shared augmented environment,” *Proc. SPIE*, vol. 3517, pp. 2-13, 1998.

- [11] N Li, *An implementation of OCR system based on skeleton matching*, University of Kent at Canterbury, 1993.
- [12] Y. Lu. “Machine printed character segmentation – An overview,” *Pattern Recognition*, vol. 28 no. 1, pp. 67-80, 1995.
- [13] L. Neumann and J Matas, “Real-time scene text localization and recognition,” *Proc. of CVPR*, pp. 3538-3545, 2012.
- [14] A. Coates, B. Carpenter, C. Case, S. Satheesh, B. Suresh, T. Wang, D.J. Wu, and A.Y. Ng, “Text detection and character recognition in scene images with unsupervised feature learning,” *Proc. of ICDAR*, pp. 440-445, 2011.
- [15] <http://arbooks.co.kr>
- [16] <http://colarapp.com>
- [17] <http://www.gizmag.com/ikea-augmented-reality-catalog-app/28703>
- [18] H. Kato, M. Billinghurst, I. Poupyrev, K. Imamoto, and K. Tachibana, “Virtual object manipulation on a table-top ar environment,” *Proc. of ISAR*, pp. 111-119, 2000.
- [19] H Kato and M. Bilinghurst, *ARToolkit User Manual*, University of Washington, Human Interface Technology Lab, 2000.
- [20] M. Fiala, *Artag Revision 1, A Fiducial Marker System Using Digital Techniques*, National Research Council of Canada, 2004.
- [21] <http://virtual.vtt.fi/virtual/proj2/multimedia/alvar>
- [22] K. Kim, V. Lepetit, and W. Woo, “Scalable real-time planar targets tracking for digilog books,” *The Visual Computer*, vol. 26, pp. 1145-1154, 2010.
- [23] D.G. Lowe, “Distinctive image features from scale-invariant

- keypoints,” *Int. J. of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.
- [24] H. Bay, T. Tuytelaars, and L.V. Gool, “SURF: Speeded up robust features,” *Computer Vision and Image Understanding*, vol. 110, pp. 346-359, 2008.
- [25] S. Leutenegger, M. Chli, and R.Y. Siegwart, “BRISK: Binary robust invariant scalable keypoints,” *Proc. of ICCV*, pp. 2548-2555, 2011.
- [26] A. Alahi, R. Ortiz, and P. Vandergheynst, “FREAK: Fast retina keypoint,” *Proc. of CVPR*, pp. 510-517, 2012.
- [27] J.J. Hull, “Paper-based augmented reality,” *Proc. of ICAT*, pp. 205-209, 2007.
- [28] H. Peng, F. Long, Z. Chi, and W. Siu, “Document image recognition based on template matching of component block projections,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1188-1192, 2003.
- [29] H. Uchiyama and H. Saito, “Augmenting text document by on-line learning of local arrangement of keypoints,” *Proc. of ISMAR*, pp. 95-98, 2009.
- [30] F. Doil, W. Schreiber, T. Alt, and C. Patron, “Augmented Reality for manufacturing planning,” *Proc. of the Workshop on Virtual Environments*, pp. 71-76, 2003.
- [31] F. Steinicke, G. Bruder, K. Rothaus, and K.H. Hinrichs, “Poster: A virtual body for augmented virtuality by chroma-keying of egocentric videos,” *Proc. of 3DUI*, pp. 125-126, 2009.
- [32] K.T. Simsarian and K.P. Akesson, “Windows on the world: An example of augmented virtuality,” *Man-Machine Interaction*,

pp. 68-71, 1997.

- [33] <http://campar.in.tum.de/Chair/ResearchIssueMedAR>
- [34] S.J. Henderson and S.K. Feiner, *Augmented Reality for Maintenance and Repair*, University of Columbia, 2007.
- [35] 김희수, “웹기반 지구과학교육에서 가상현실 기술의 활용,” *한국지구과학회지*, vol. 23 no. 7, pp. 531-542, 2009.
- [36] M. Billingham, H. Kato, and I. Poupyrev, “The MagicBook - Moving seamlessly between reality and virtuality,” *IEEE Trans. on Comput. Graph. Appl.*, vol. 21, no. 3, pp. 6-8, 2001.
- [37] K. Johannes, A. Pagani, and D. Stricker, “Robust detection and identification of partially occluded circular markers,” *Proc. of the 3rd International Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, vol. 1, pp. 387-392, 2010.
- [38] L. Naimark and E. Foxlin, “Circular data matrix fiducial system and robust image processing for a wearable vision-inertial self-tracker,” *Proc. of ISMAR*, pp. 27-36, 2002.
- [39] *Information Technology - Automatic Identification and Data Capture Techniques - QR Code 2005 Bar Code Symbology Specification*, International Organization for Standardization, ISO/IEC 18004, 2006.
- [40] M. Jonathan, S. You, and U. Neumann, “Tricodes: A bar-code-like fiducial design for augmented reality media,” *Proc of ICME*, pp. 1301-1304, 2006.
- [41] J. Sattar, E. Bourque, P. Giguère, and G. Dudek, “Fourier tags: Smoothly degradable fiducial markers for use in human-robot interaction,” *Computer and Robot Vision*, pp. 165-174, 2007.

- [42] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330-1334, 2000.
- [43] H. Wuest, F. Vial, and D. Stricker, "Adaptive line tracking with multiple hypotheses for augmented reality," *Proc. of ISMAR*, pp 62-69, 2005.
- [44] R.O. Duda and P.E. Hart, "Use of the hough transformation to detect lines and curves in pictures," *Int. J. of CACM*, vol. 15, no. 1, pp. 11-15, 1972.
- [45] A.I. Comport, E. Marchand, and F. Chaumette, "A real-time tracker for markerless augmented reality," *Proc. of ISMAR*, pp. 36-45, 2003.
- [46] T. Lee, and T. Höllerer, "Hybrid feature tracking and user interaction for markerless augmented reality," *Proc. of Virtual Reality Conference*, pp. 145-152, 2008.
- [47] Intel Corporation. *Open Source Computer Vision Library referencemanual*. 2000.
- [48] S. Benhimane and E. Malis, "Real-time image-based tracking of planes using efficient second-order minimization," *Proc. of IROS*, vol. 1, pp. 943-948, 2004.
- [49] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," *Proc. of ISMAR*, pp. 1-10, 2007.
- [50] L. Vacchetti, V. Lepetit, and P. Fua, "Stable real-time 3D tracking using online and offline information," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.26, no.10, pp. 1385-1391, 2004.
- [51] A.J. Davison, I.D. Reid, N.D. Molton, and O. Stasse, "MonoSLAM : Real-time single camera SLAM," *IEEE Trans.*

- on Pattern Analysis and Machine Intelligence*, vol.29, no.6, pp. 1052-1067, 2007.
- [52] M.I. Lourakis, and A.A. Argyros, "Efficient, causal camera tracking in unprepared environments," *Computer Vision and Image Understanding*, vol. 99, no.2, pp. 259-290, 2005.
- [53] C. Arth, M. Klopschitz, G. Reitmayr, and D. Schmalstieg, "Real-time self-localization from panoramic images on mobile devices," *Proc. of ISMAR*, pp. 37-46, 2011.
- [54] J. Canny, "A Computational Approach To Edge Detection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679-698, 1986.
- [55] C. Harris and M. Stephens, "A combined corner and edge detector," *Proc. of AVC*, pp. 147-151, 1988.
- [56] <http://www.opengl.org>
- [57] 류호섭, 박한훈, 문광석, "페이지 마커를 위한 패턴 인식 방법에 관한 연구," *한국신호처리시스템학회 추계학술대회 논문집*, pp. 236-243, 2013.
- [58] 류호섭, 박한훈, 문광석, "증강현실 책을 위한 페이지 마커," *한국신호처리시스템학회 하계학술대회 논문집*, 2013.
- [59] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," *Proc. of ICCV*, pp. 2564-2571, 2011.
- [60] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed., Cambridge University Press, New York, NY, USA, 2003.

연구 실적

◎ 국내 전문학술지

1. 박한훈, 류호섭, 문광석, "페이지 마커를 이용한 텍스트 증강," *Telecommunications Review*, 제 24권, 1호, 2014년 2월.

◎ 해외 전문학술지

1. Ho-Sub Ryu and Hanhoon Park, "A system for supporting paper-based augmented reality," *Multimedia Tools and Applications*. (online published)

◎ 국내 학술발표대회

1. 류호섭, 박한훈, 문광석, "특징점과 이웃 특징점 사이의 각을 이용한 회전, 이동 불변인 영상매칭," *정보 및 제어 심포지엄 논문집*, pp. 135-136, 서울, 2013년 4월.
2. 박한훈, 류호섭, 문광석, "순차적 부분 비교를 통한 이진 특징 기술자의 고속 정합," *정보 및 제어 심포지엄 논문집*, pp. 157-158, 서울, 2013년 4월.
3. 류호섭, 박한훈, 문광석, "증강현실 책을 위한 페이지 마커," *한국신호처리시스템학회 하계학술대회 논문집*, 울산, 2013년 7월.

4. 류호섭, 박한훈, 문광석, "페이지 마커를 위한 패턴 인식 방법에 관한 연구," *한국신호처리시스템학회 추계학술대회 논문집*, pp. 236-243, 순천, 2013년 12월.

5. 류호섭, 박한훈, 문광석, "기하 정보를 이용한 특징 기술 기법에 관한 연구," *한국신호처리시스템학회 하계학술대회 논문집*, 대구, 2014년 6월.

◎ 수상 실적

1. 우수발표논문상 : 류호섭, 박한훈, 문광석, "증강현실 책을 위한 페이지 마커," *한국신호처리시스템학회 하계학술대회 논문집*, 울산, 2013년 7월.

◎ 특허

1. 박한훈, 류호섭, *텍스트 마커를 이용한 증강현실 시스템* (출원: 대한민국, 10-2013-0057593, 2013년 05월)