석 사 학 위 논 문

지역 Zernike 모멘트를 이용한 모델 기반 확률적 물체 인식

김 성 호 (金 聖 浩)

전자전산학과 전기및전자공학전공

한국과학기술원

2002

지역 Zernike 모멘트를 이용한
모델 기반 확률적 물체 인식

Probabilistic Model-based Object Recognition
using Local Zernike Moments

Probabilistic Model-based Object Recognition using Local Zernike Moments

Advisor : Professor In-So Kweon.

by

Sung-Ho Kim

Department of Electrical Engineering & Computer Science
[Division of Electrical Engineering]
Korea Advanced Institute of Science and Technology

A thesis submitted to the faculty of the Korea Advanced Institute of Science and Technology in partial fulfillment of the requirements for the degree of Master of Engineering in the Department of Electrical Engineering & Computer Science[Division of Electrical Engineering]

Daejon, Korea

2001. 12. 21.

Approved by

Professor In-So Kweon
Major Advisor

지역 Zernike 모멘트를 이용한 모델 기반 확률적 물체 인식

김 성 호

위 논문은 한국과학기술원 석사학위논문으로 학위논문심사위원회에서 심사 통과하였음.

2001 년 12 월 21 일

심사위원장 권 인 소 (인)

심사위원 김 성 대 (인)

심사위원 정 명 진 (인)

MEE 김 성 호, Kim, Sung-Ho.
20003089
Probabilistic Model-based Object Recognition using Local Zernike Moments, 지역 Zernike 모멘트를 이용한 모델 기반 확률적 물체 인식.
Department of Electrical Engineering and Computer Science, 2002, 98p
Advisor: Prof. In-So Kweon.

Abstract

Object recognition is one of the most important, yet the least understood, aspect of visual perception. The difficulties originate from the variations of objects such as view position, illumination changes, background clutter, occlusion and etc. So, the success of any recognition scheme will depend on its ability to cope with these variations. This thesis presents an object recognition paradigm robust to these variations using local Zernike moments and the probabilistic voting method.

The proposed method using Zernike moments is robust to rotation, scale changes, illumination changes, background clutter and occlusion. The original Zernike moments are normalized by (0,0) Zernike moment which makes the feature robust to scale illumination changes. The modified Zernike moments are calculated around corner points which are extracted from images represented in scale space. For object recognition, we have developed a probabilistic voting method which is an extension of simple voting method. The proposed probabilistic voting method is based on the stability of model Zernike moments and the similarity between the model Zernike moments and the input Zernike moments. This method is better than the simple voting which has high risk of misrecognition for similar objects.

The object recognition system is validated through various experiments. The experimental results show the robustness of the proposed object recognition system.

목 차

표 목차

그림 목차

1. 서 론

1.1. 연구 배경

1980년대 로봇 산업은 제조 위주의 산업용 로봇을 중심으로 발전하였고, 1990년대는 건설, 의료/복지, 농업 등의 비제조업용 로봇 위주로 개발되었다. 그러나, 21세기 차세대 로봇은 인간 공존형 개인용 로봇 위주로 발전할 것이다[1]. 현재 국내에서도 정보기술(IT)과 생명공학기술(BT)과 함께 로봇기술(RT)이 21세기 핵심 전략 기술로 떠오르고 있다. 앞으로 생활 패턴이 점점 네트워크화, 정보화 됨에 따라 고통스러운 가사 노동으로부터의 해방을 필요로 하게 될 것이다. 또한 사회 구조가 노령화, 개인화 됨에 따라 개인용 로봇의 필요성은 더욱 증대하게 된다. 설거지, 청소, 요리, 세탁 등을 하는 가사용 로봇이나, 실버 사회를 대비해 노인의 수발을 들어주는 보조 로봇, 환자 간병을 위한 간호 로봇, 장애인을 위한 안내 로봇, 여가를 즐기기 위한 엔터테인먼트 로봇 등 매우 다양한 분야에서 개인용 로봇을 필요로 할 것이다.

산업용 로봇이나 차세대 로봇에서 시각 기술은 필수적이다. 특히, 차세대 로봇에서 시각을 통한 인식 기능은 로봇이 다양한 환경에서 동작해야 하기 때문에 과거 시각 인식 기술보다 더욱 지능적이고 다양한 환경에서 강인하게 동작하는 특성을 가져야 한다. 산업용 로봇에서 시각 인식 기술이 제한된 환경에서 간단한 기계 부품을 인식하고 로봇 팔을 이용하여 조립하는 수준이었다면, 차세대 로봇에서 인식 기술은 집안에 있는 다양한 물건들이나, 사람, 자동차, 건물 등을 인식하고 어떤 물체가 어디에 있는지 다양한 환경에서 동작할 수 있는 강인성을 필요로 한다. 따라서 차세대 개인용 로봇에 물체 인식 능력을 부여하는 것은 매우 중요한 기반 기술이 된다.

본론에 들어가기에 앞서, 그 동안 물체 인식에 관해 연구해 온 결과를 토대로 물체

---

[1] 25회 21C 경제 포럼, 2000년 미래학회 보고서, 독일 지멘스 로크매너 연구소 보고서, 2000년 IFR UN/ECE 보고서

인식이란 엄밀하게 뭘 뜻하는지, 왜 어려운지, 그 동안 접근한 방법이 무엇이고 그 한계성은 무엇인지 차례로 살펴본 후, 본 논문에서 제안하는 방법에 대해서 간단히 소개하도록 한다.

1.2. 물체 인식이란

우리는 물체 인식이라는 말을 종종 접하지만, 물체 인식의 개념을 간결하고 명확하게 정의 내리기는 쉽지 않다. 가장 쉽게 대답할 수 있는 것은 물체에 이름을 부여하는 것이다 [1]. 아이 보고 시계를 보여주면, 그냥 시계라고 대답할 것이다. 하지만 이 답도 명확한 것은 아니다. 어떤 때에는 누구 소유의 시계인지 알고 싶을 때도 있고, 또 어떤 때는 이 시계가 손목 시계인지, 탁상 시계인지 혹은 아날로그 시계인지, 전자 시계인지 궁금할 때가 있다. 이러한 경우에는 적당한 관점에서 물체 인식을 정의할 수 있다. 한편, 뭐가 보이냐고 물을 경우에는 시계의 각 부품을 열거할 수도 있다.

위의 대답들을 정리해보면, 일반적으로 물체인식이란 물체와 관련된 정보를 뽑아내는 능력으로 볼 수 있다. 여기서 말하는 정보는 물체의 이름, 소유주, 용도, 크기, 모양, 색깔, 위치, 자세 등이 될 수 있다.

인식 방법론 관점에서 물체 인식은 다음과 같이 정의 내릴 수 있다. 일반 물체에 대한 사전 정보(general feature)를 알고 있을 때, 입력 영상에서 어떤 물체가 어디에 있는지 파악(matching)하는 것으로 볼 수 있다. 즉, 물체 인식은 일반적인 특징량의 매칭(general feature matching)이라고 정의할 수 있다.

1.3. 왜 물체 인식이 어려운가

앞 절에서 물체에 대한 사전 정보를 갖고 있을 때, 입력 영상에 어떤 물체가 있는지, 그리고 그 물체가 어디에 어떤 자세로 있는지 알아내는 것으로 물체 인식을 정의하였다. 물체 인식은 새로운 문제도 아니고 그렇다고 다 풀린 문제도 아니다. 그 동안 물체 인식

분야에 많은 발전이 있었지만, 현재 매우 높은 인식률을 자랑하는 물체 인식 시스템도 잘 제어된 환경에서만 물체를 인식하는 수준에 머무르고 있다. 그것은 다음과 같은 물체의 다양성이 있기 때문이다 [1][2].

물체 인식을 어렵게 하는 가장 큰 요인은 물체를 바라보는 위치에 따라서 물체가 다르게 보인다는 사실이다. 일반적으로 물체에 대한 카메라의 위치는 6 자유도를 갖는다. x, y, z 방향으로의 평행 이동과 각 축으로의 회전이 있다. 3차원 물체는 매우 다양한 시점(방향, 거리)에서 바라볼 수 있기 때문에 한 물체에 대해서 크기도 다르고 자세도 다른 다양한 영상을 얻게 되어 물체 인식을 어렵게 한다.

일반적으로 물체는 독립적으로 분할되어 존재하지 않는다. 물체를 구성하기에 따라서 물체 뒤에 복잡한 배경이 있을 수도 있고(background clutter) 혹은 다른 물체에 의해 물체가 가려지기도 한다(occlusion). 이러한 현상은 물체 인식을 어렵게 하는 요인이 된다.

다음으로 조명에 의해 물체의 색깔이 변한다는 사실이다. 낮에 본 물체가 초저녁에 본 물체와 동일하지 않다. 물체의 색깔은 조명의 위치나 분포, 조명의 파장, 다른 물체에 의한 반사, 그림자, 대기 조건 등에 크게 의존하므로 이러한 현상들이 물체 인식을 어렵게 한다.

한편, 사람의 신체나 동물 같이 살아 있는 물체는 움직임에 따라 모양이 바뀌게 된다. 가위와 같이 딱딱한 요소들로 구성된 경우도 있고, 사람 얼굴처럼 표정이 변할 때 매우 다양한 모습을 나타내는 경우도 있다. 이러한 물체 모양의 변화는 물체 인식을 어렵게 만든다.

마지막으로 영상 형성 과정이 완벽하지 않은 요인을 들 수 있다. 영상은 픽셀값으로 표현되는 셀로 이산화 된다. 비록 픽셀값이 평균값으로 표현되지만, 정확한 위치 정보를 잃게 된다. 또한 밝기값도 이산화 되므로 부정확도는 더 높아진다. 한편, 렌즈의 불완전성 때문에 영상에 왜곡이 발생하기도 하고 전자 장치에 의해 고주파 잡음이 첨가되기도 한다. 이렇게 카메라에서 발생한 잡음은 하위 레벨(low level) 프로세싱(에지, 선, 코너 추출 등)을 부정확하게 하여 상위 레벨(high level) 인식을 어렵게 만든다.

실제로 위와 같은 다양성을 갖게 하는 요인들이 물체에 복합적으로 나타나기 때문에 물체 인식이 더욱 어려운 것이다.

1.4. 관련 연구

지난 30여년간 비전 그룹에서는 위와 같은 물체의 다양성 문제를 해결하고자 여러 가지 방법을 제안하였다. 기존 접근 방법을 인식 방법에 따라 분류해보면 다음과 같이 세 가지로 나눠볼 수 있다 [1][3].

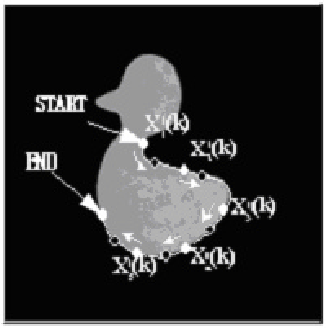
1) 불변성을 이용한 방법
2) 3D CAD 모델에 기반한 방법
3) 2D 영상에 기반한 방법

그럼 지금부터 각 방법의 특성 및 한계점을 하나씩 살펴보자.

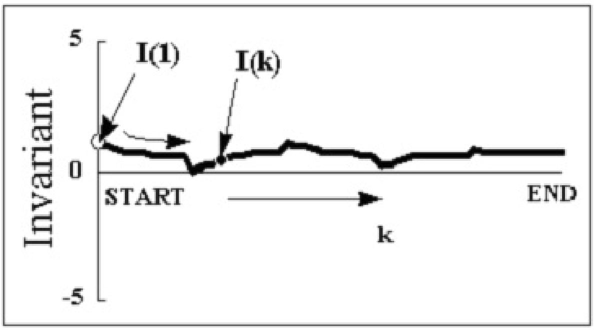
1.4.1. 불변성(invariance)

시점의 변화나 조명 변화에 대해 불변 특성을 갖는 특징량을 이용해 물체 인식을 할 수 있다. 물체의 실루엣은 조명이 바뀌더라도 거의 변하지 않고, 평행한 에지는 넓은 시점에 대해 불변 특성을 갖는다.

그림 (1.1)은 3차원 물체의 실루엣에 대한 불변성을 보여 주고 있다 [4]. 5개의 점을 이용하여 불변량을 정의하고 물체의 윤곽선에 대해 얻은 불변량을 이용하여 인식을 수행하게 된다.



(a) 분할된 입력 영상



(b) 불변량 그래프

그림 1.1 불변성의 예

비록 불변 특성이 물체 인식에 중요한 단서를 제공하지만, 3차원 물체에 대해 전방향 불변성을 제공하지 못하는 한계가 있다.

1.4.2. 3D 모델 기반(3D explicit model-based)

시점(view position)에 따른 물체의 변화에 대응하기 위해 앞서 설명한 불변 특성을 이용하지 않고 3D CAD 모델을 사용하는 방법이 있다. 인식할 때 물체의 모습을 예측하기 위해서 물체의 형성 과정을 이용한다. 결국 이 방법을 이용한 물체 인식의 성공 여부는 물체의 모양, 조명, 표면 반사, 광학적 투영을 어떻게 잘 모델링 하는가에 달려있다. 그림 (1.2)는 3D CAD 모델을 거리(range) 영상, 가시광 영상, 적외선 영상에 각각 정렬(align) 시켜서 차량을 인식한 결과를 보여 주고 있다 [5].

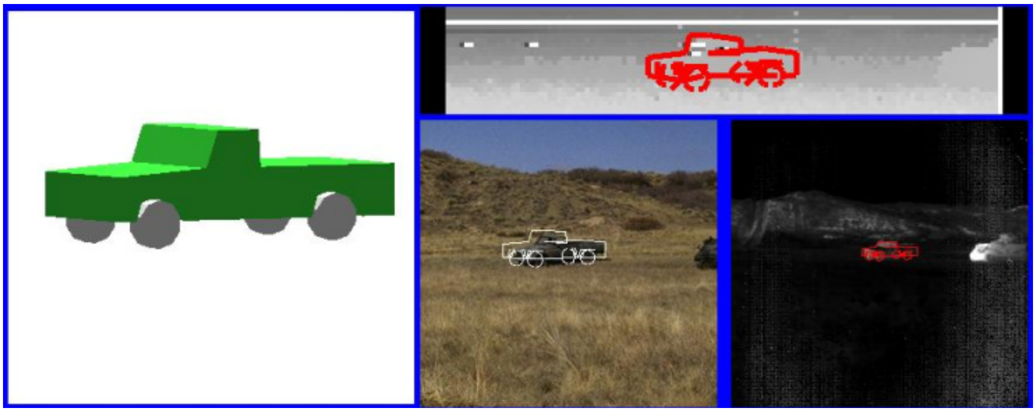
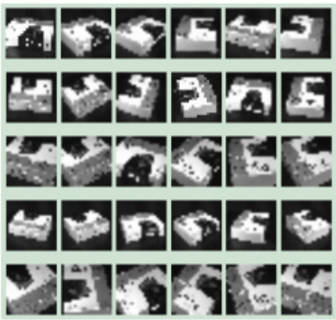


그림 1.2 CAD 모델을 이용한 인식

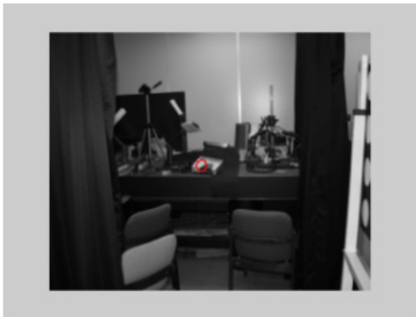
하지만 이러한 모델은 단순한 물체를 모델링 하는데 국한돼 있고, 물체 모양이나 영상 형성 과정을 모두 설명해주는 모델은 실제로 만들기 어렵다.

1.4.3. 2D 영상 기반(2D appearance-based)

다양한 조건에서 보이는 물체를 여러 장의 영상으로 모델을 만들 수 있다. 영상 기반 인식 시스템은 모든 가능한 물체 영상을 이용하여 물체 인식을 수행한다. 그림 (1.3) 은 최근에 제안된 ANTI-FACE 기법을 이용하여 물체를 찾는 것을 보여준다 [6]. 다양한 시점에서 찍은 500장의 모델 영상으로부터 템플릿과 모델간의 내적이 작고, 다른 배경과의 내적이 크다는 통계적 특성을 이용하여 템플릿 매칭을 수행한 결과이다. 이 방법의 장점은 모델이 물체의 실제 물체 모습을 예측할 필요가 없다는 점이다. 하지만, 물체를 모델링하기 위해 매우 다양환 환경에서 물체 영상을 얻어야 한다.



(a) 물체 모델 DB



(a) 인식 결과

그림 1.3 영상기반 물체 인식(Anti-face 방법)

앞서 언급한 불변량, 3D 모델, 2D 영상 기반으로 물체의 다양성을 고려할 수 없는 것은 잡음 범주로 생각할 수 있다.

위에서처럼 대표적인 방법론 관점에서 물체 인식 기법을 분류할 수 있다. 지금까지의 제안된 방법들을 살펴보면, 물체 인식이 특징량 사이의 매칭이라는 공통점이 있다. 물체 인식의 정의에서도 언급하였듯이, 물체의 사전 정보(feature)를 알고 있을 때, 영상에서 어떤 물체가 어디에 있는지 파악(matching)하는 것은 대부분의 인식 방법론들이 갖고 있는 공통점이다. 따라서 물체 인식은 물체의 특징량을 표현하는 부분과 이 특징량을 매칭하는 부분으로 구성된다. 하위 레벨 특징량에는 코너(corner), 에지(edge), 라인(line),

접합점(junction), 커브(curve) 등이 있고, 상위 레벨 특징량에는 모양(shape), 영역(region), 칼라(color), 디스크립터(descriptor) 등이 있고, 매칭 방법에는 수동적 방법, 상관관계, MMSE, 해시(hash), 정렬(alignment), MRF(Markov Random Field), HMM(hidden Markov model), PCA, anti-face, 유전자 알고리즘, 신경망 등이 있다. 다음은 1995년부터 2001년까지 저자별로 사용한 특징량과 매칭 방법을 간단히 정리한 것이다.

표 1.1 저자별 사용한 특징량 및 매칭 기법

저자명	Feature	Matching	년도
A. R. Pope et al. [7]	edge	alignment	1995
S. Z. Li el al. [8]	curve	MRF	1995
A. T. S. Au el al. [9]	line	hash	1996
G. R. Ji et al. [10]	moment	RBF	1997
S. Lin et al. [11]	color	PCA	1997
F. Jurie [12]	line	hash	1997
D. W. Jacob et al. [13]	region	hand	1997
M. R. Stevens et al. [5]	line	genetic algorithm	1998
A. E. Johnson et al. [14]	shape	correlation	1998
B. Cui et al. [15]	color	HMM	2000
D. Roobaert et al. [16]	shape	SVM	2000
H. Zha et al. [17]	region	Hopfield	2000
Y. H. Cu et al. [18]	curve	MMSE	2000
K. C. Lo et al. [19]	curve	MMSE	2001
D. Keren [6]	region	anti-face	2001

특징량 중에서 영역이나 모양, 커브, 에지, 라인 등은 물체가 이상적일 때 잘 추출될 수 있다. 그러나 실제로 복잡한 배경(background clutter)이나 다른 물체에 의한 가려짐

현상(occlusion)이 발생하는 경우에는 물체 고유의 특성을 잘 표현할 수 있어야 하고 효율적으로 추출되어야 한다. 이러한 관점에서 물체를 표현하는 방법은 지역 특징량을 이용하는 것이 적합하다. 지역 특징량을 이용할 경우, 물체의 영역을 분할하지 않고 물체 인식을 할 수 있는 장점이 있다. 또한 물체 외관이 바뀌더라도 일부 특징량만 바뀌기 때문에 매우 안정적으로 물체를 표현할 수 있고, 부분 가려짐 현상도 일부 특징량에만 영향을 주기 때문에 매우 강인하게 물체를 인식을 할 수 있다. 따라서 지역 특징량을 이용한 인식 기법이 매우 효과적인 것을 알 수 있다.

한편, 아직까지 인간의 인식 방법이 확실하게 정립되지 않았지만, 인식에 있어서 물체의 외관이 매우 중요한 역할을 하는 것은 분명하다. 최근 인간의 시각 기능에 대해 연구한 결과에 의하면, HVS(human visual system)는 3D 모델에 기반하여 물체 인식을 하기 보다는 2D 영상을 기반으로 물체를 인식한다 [20]. 이 논문에서 저자는 2D 영상으로부터 3D 물체 인식을 수행했다. 복잡한 3D CAD 모델을 만들어 인식을 하는 대신, 단순히 2D 영상만으로 인식 모델을 만들었다. 이렇게 물체의 외관은 물체 인식에 있어서 매우 중요한 부분을 차지한다. 외관을 표현하는 기술자에는 PCA를 이용하는 방법등 여러 가지가 있지만, 모멘트를 이용한 방법에는 불변 특성이 있기 때문에 영상 기술자로 많이 사용된다.

최근에 이와 관련하여 Schmid [21]는 지역 미분 불변량(local differential invariant)을 이용하여 매우 높은 성공률로 영상 검색 시스템을 만들었다. 9차원의 디스크립터 벡터(descriptor vector)를 만들어 대응점을 찾았다. 공간적인 조건(Spatial constraint)을 적용하여 잘못된 매칭을 줄였다. 한편 Lowe [22]는 SIFT(scale invariant feature transform)라는 새로운 디스크립터(descriptor)를 이용하여 크기, 이동, 회전에 불변인 물체 인식을 수행하였다. Ikeuchi [23]는 고유 창(eigen window)에서 구한 지역 PCA를 이용하여 가려짐에 강한 인식기를 제안하였다. 하지만 회전이나 크기에는 강인하지 않은 문제점이 있다. 이와 유사하게 Jugessur [24]도 대칭변환(symmetry transform)과 FFT, 지역 PCA를 이용하여 이동, 회전에 강한 인식시스템을 만들었다. 하지만 크기에 민감한 문제점이 있다.

1.5. 문제 제기

기존 연구를 통해 물체 인식의 기본 구조가 특징량과 이들의 매칭으로 구성된다는 것을 알 수 있었다. 또한 특징량을 선택할 때 전역 특징량보다는 지역 특징량을 쓰는 것이 여러 환경(배경, 가려짐 등)에도 강인하게 작용하는 것도 알 수 있었다. 이러한 방법에 기초하여 최근에 많은 연구가 있었다 [21, 22, 23, 24]. 그러나 물체 인식 관점에서 크기, 회전, 조명 변화, 배경 잡음, 가려짐 현상이 있을 때 인식을 정량적으로 분석하지 못한 부분이 있다. 한편, 특징량 관점에서 보면, 미분 불변량을 사용한 방법은 잡음에 매우 민감하다는 단점이 있고, SIFT는 사람의 시각 특성을 모사해 불변 특징량을 만들 수 있지만 계산량이 많다는 문제가 있다. 또한 지역 PCA는 크기, 회전에 매우 민감하다는 문제가 있다. 따라서 본 논문에서는 물체의 지역 외관을 잘 표현할 수 있는 모멘트에 관심을 갖게 되었고, 그 중에서 회전 불변 특성을 가진 Zernike 모멘트를 특징량으로 선택하였다.

한편, 고차원 특징량 매칭 기법에는 Mahalanobis 거리를 이용한 단순 보팅 기법이 주로 사용되어왔다. Schmid [21]는 공간 조건을 이용하여 효과적인 보팅을 수행하였다. Lowe [22]는 best-bin-first search 방법을 이용하여 특징량을 찾고, Hough 변환을 통해 물체의 자세를 추정하였다. Ikeuchi [23]는 고유창 사이의 거리를 이용해 매칭을 수행하였다. 본 논문에서는 기존의 보팅 기법의 문제점을 보완한 확률적 보팅 기법을 제안한다. 특징점의 위치에 따른 Zernike 모멘트의 안정도, 특징량 사이의 유사도 개념을 확률적 보팅 기법에 반영하여 강인한 인식 시스템을 구현하였다.

1.6. 시스템 overview

제안된 물체 인식 시스템은 그림 (1.4)와 같이 전체적으로 두 부분, 즉 오프 라인 부분과 온 라인 부분으로 구성된다. 오프 라인 부분은 물체의 특징량 DB를 구축하는 부분으로써 DB 생성시 상당한 시간이 걸린다. 물체의 대표적인 영상들을 스케일 스페이스(scale space)로 바꾸어 각 스케일 스페이스마다 특징점을 추출한다. 다음, 이렇게 추출된 특징점

주변에 대해 Zernike 모멘트를 사용하여 각 점의 고유 특징량을 추출한다. 한편, 온 라인 부분에서는 물체와 배경이 포함된 임의의 물체 영상에 대해 특징점을 추출한다. 이 때, 스케일 스페이스를 사용하지 않고 입력 영상에서만 특징점을 구한다. 다음 각 점에 대해서 Zernike 모멘트로 특징량을 구한다. 이렇게 추출한 특징량으로부터 확률적 물체 인식 과정을 통해 물체 모델을 찾게 된다. 마지막으로 물체 인식이 제대로 되었는지 검증을 함으로써 인식을 종료하게 된다.

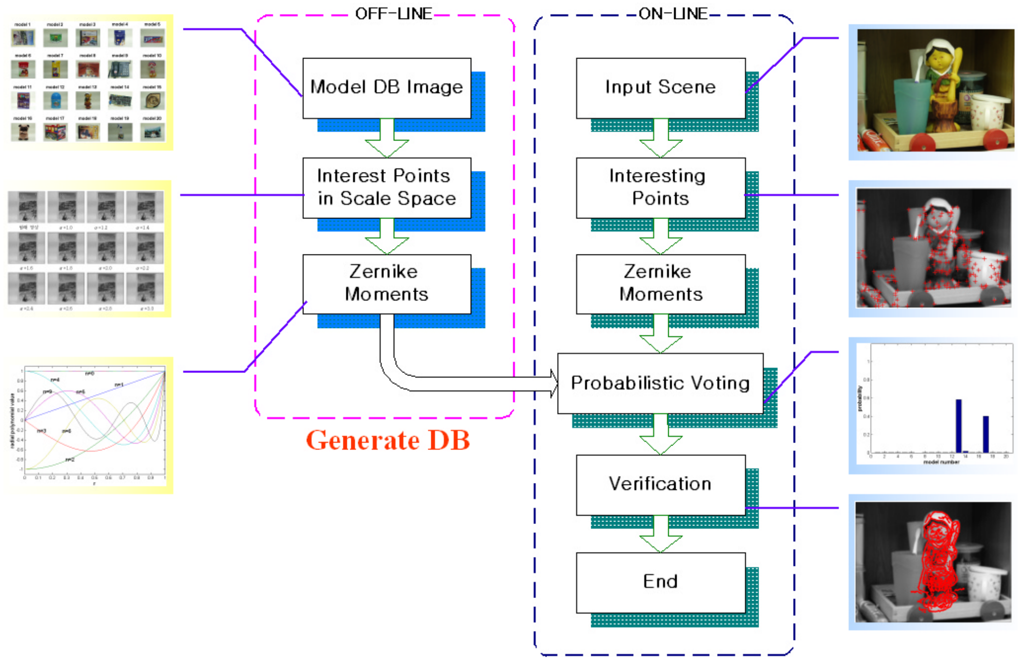


그림 1.4 물체 인식 시스템 구성도

1.7. 논문의 구성

본 논문의 2장에서는 강인 특성을 갖는 특징량에 대해서 다루고, 3장에서는 확률적 보팅에 기초한 물체 인식 기법에 대해 살펴본다. 4장에서는 20개의 모델에 대한 다양한 실험을 통해 제안한 시스템의 강인성을 검증하도록 한다. 5장에서는 결론 및 향후 과제에 대해서 살펴본다.

2. 지역 Zernike 모멘트

지역 Zernike 모멘트는 물체의 특징점에 기초한다. 먼저 특징점 추출기의 특성에 대해서 간단히 살펴본 후 Zernike 모멘트의 특성에 대해서 살펴보자.

2.1. 특징점 (interesting point)

제안한 인식 방법론은 물체의 지역 특징량에 기초한 인식 기법이다. 본 논문에서 물체의 지역 정보를 이용하는 것은 생물학적인 현상에 근거한다. 그 동안 인간 시각에 대해 연구하는 사람들은 뇌의 물체 인식 원리를 직접적으로 알 수 없었기 때문에 간접적인 현상에 기초하여 물체 인식 원리를 알고자 하였다. 최근 연구 결과를 종합해 보면, 시각 정보 처리 과정에 대한 간단한 결론을 얻을 수 있다. 두 가지 측면으로 요약할 수 있는데, 하나는 시야에 들어온 영상 정보가 동시에 자동(특정 지역에 의식적으로 초점을 맞추지 않는 것)으로 처리된다는 사실이고, 다른 하나는 의식적으로 초점이 맞추어진 상태에서 시야를 차례로 보면서 영상정보가 처리된다는 사실이다 [25]. 일반적으로 사람들은 특정 물체를 인식하고자 할 때에는 물체의 특징이 되는 부분에 집중하여 본 후 종합하여 물체를 인식한다. 또한, Marr는 영상정보처리가 시각 수용체에 들어온 빛으로부터 라인(line), 스팟(spot), 에지(edge) 그리고 이들의 위치, 방향, 칼라 등의 정보로부터 시작된다고 했다. 즉, 특징점 그리고 이러한 점을 찾는 특징점 추출 연산자는 이러한 인간의 인식 과정에서 비롯된다고 보았다.

위 사실에 기초하여, 특징점을 잘 추출하기 위해서는 무엇보다도 성능이 우수한 특징점 추출 연산자가 필요하다. 특징점이란 영상에서 2차원적으로 변하는 특정한 점을 의미하고 특징점 추출 연산자는 특징점을 구하는 연산자를 뜻한다. 물체 인식 관점에서 특징점이 갖춰야 할 요건은 다음과 같이 세가지로 정리 할 수 있다.

1) 어파인 변환(affine transformation: x, y축으로 스케일, 회전, 이동)에 대해 물체

영상과 모델 영상 모두에서 동일한 점을 찾을 수 있어야 한다.

2) 조명 변화에 대해서도 물체 영상, 모델 영상에 대해 모두 동일한 점을 찾을 수 있어야 한다.

3) 잡음에 무관하게 물체 영상, 모델 영상에 대해 모두 동일한 점을 찾을 수 있어야 한다.

첫번째 특성은 모델 영상과 입력 물체 영상 사이에 시점 차이가 특정 범위 있기 때문에 어파인 변환으로 가정하였다. 따라서 시점 변화에 대해서 특징점의 위치가 동일 해야한다. 두 번째 특성은 특징점이 조명 변화에 강인하다는 것, 마지막으로, 특징점이 잡음에 덜 민감해야 한다는 것을 뜻한다.

특징점 추출 연산자는 특징점을 구하는 방법에 따라 크게 외곽선에 기초한 방법(contour-based), 픽셀 밝기 분포에 기초한 방법(intensity-based) 그리고 대칭성에 기초한 방법(symmetry-based)으로 나눌 수 있다. 외곽선에 기초한 방법에서는 곡률이 큰 부분이 특징점이 되고, 밝기에 기초한 방법은 밝기값이 크게 변하는 점이 특징점이 된다. 마지막으로 대칭성에 기초한 방법은 물체의 대칭 되는 곳을 특징점으로 한다.

본 논문에서는 빠르고 효과적인 픽셀 밝기 분포에 기초한 방법에 초점을 두도록 한다. 이 방법에는 여러 가지가 있지만, 본 논문에서는 잘 알려진 Harris 코너 검출기와 COP 코너 검출기, KLT 코너 검출기 그리고 SUSAN 코너 검출기에 대해서 각 특성을 비교한다. 특히, 물체 인식에 필요한 연산자 관점에서 각 코너 검출기 특성을 살펴 보기로 한다. 먼저 각 코너 검출기 별 연산법을 간단히 소개하기로 한다.

2.1.1. 특징점 추출 연산자

a) Harris 코너 추출 연산자

Harris [26]는 식 (2.1)과 같이 자기상관행렬(auto-correlation matrix) **A**를 사용하여 이산적인 방향(discrete direction)과 이산적인 이동(discrete shift)을 배제하는 대신 윈도

우 내에서 가우시안으로 미분을 콘볼루션(convolution) 했다. 특징점은 $\det(\mathbf{A}) - a\ \mathrm{trace}(\mathbf{A})$의 문턱치를 넘는 점으로 한다.

$$\mathbf{A}(x_k, y_k) = \begin{bmatrix} \sum_{(x_k, y_k)\in W} (I_x(x_k, y_k))^2 & \sum_{(x_k, y_k)\in W} I_x(x_k, y_k) I_y(x_k, y_k) \\ \sum_{(x_k, y_k)\in W} I_x(x_k, y_k) I_y(x_k, y_k) & \sum_{(x_k, y_k)\in W} (I_y(x_k, y_k))^2 \end{bmatrix} \tag{2.1}$$

기존 Harris 코너에서 Sobel 에지 검출기를 사용하는 대신 σ가 1인 derivative of Gaussian을 사용했다. 이렇게 함으로써 잡음을 잘 제거하면서 효과적으로 에지를 찾을 수 있다. 필터 크기는 7로 하고 a 값은 0.06으로 하여 물체의 경계선 쪽에는 코너를 가능하면 적게 찾게 하였다. 또한 특징적인 코너점만 선택하기 위해 8-neighborhood에서 가장 큰 값을 선택하여 두드러진 코너만 찾도록 하였다. 그림 (2.1)은 Harris 코너 검출기가 실제 영상에서 어떤 점을 찾는지 보여 주고 있다. 매우 정확하게 코너점을 찾는 것을 알 수 있다.

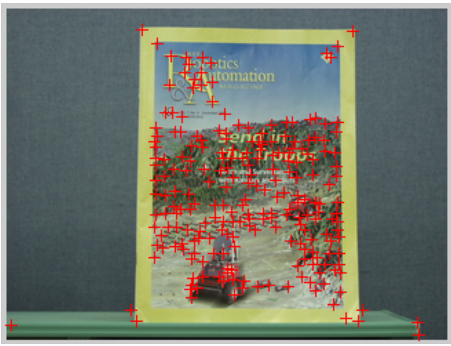


그림 2.1 Harris 코너 검출기 결과

b) SUSAN 코너 추출 연산자

SUSAN [29]은 smallest univalue segment assimilating nucleus의 약자로, USAN이라

는 원형 마스크를 사용하여 코너를 찾는 방법이다. 37개의 픽셀로 구성된 원형 마스크 안에 있는 픽셀값들과 가운데 있는 픽셀값과의 유사도를 이용하여 코너를 검출한다. 먼저, 원형 마스크 내에서 가운데 있는 픽셀값과 유사한 픽셀의 개수를 센다. 개선된 코너 영상을 얻기 위해 기하학적 문턱치에서 USAN 크기를 뺀다. 다음, USAN 영역에 대해 모멘트를 계산하여 방향을 구하고, NMS(non-maximal suppression)을 통해 최종적으로 코너점을 구한다.

그림 (2.2)는 SUSAN 코너 검출기를 이용하여 코너를 찾은 결과를 나타낸다.

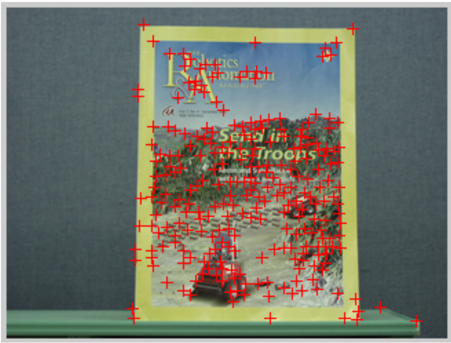


그림 2.2 SUSAN 코너 검출기 결과

c) COP (crosses as an oriented pair) 코너 추출 연산자

COP [27]는 그림 (2.3)과 같은 방향성 있는 두 개의 십자가 연산자를 통해 의미 있는 코너점을 찾는 방법이다. 왼쪽 연산자는 45도 에지에 강한 값을 내고, 오른쪽 연산자는 수평/수직 성분의 에지에 강한 값을 내는 특성을 갖는다. 코너점은 COP를 이용하여 inverted USAN 영역을 얻고 이 영역에 대해서 후보 코너점을 구한다. 이 후보 점에 대해 밝기 영상의 2계 미분값을 구하여 탐색 영역 내 최대점을 코너점으로 한다. 그림 (2.4)에 코너 추출 결과를 보여 준다.

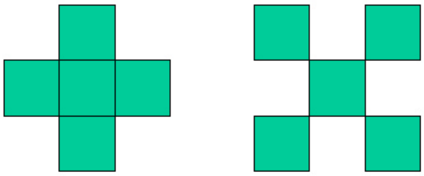


그림 2.3 2개의 COP 연산자

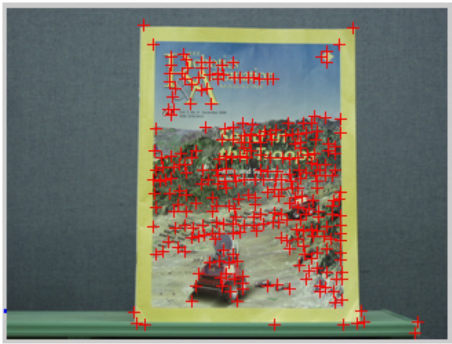


그림 2.4 COP 코너 검출기 결과

d) KLT 코너 추출 연산자

KLT 코너 [28]는 Kanade, Lucas, Tomasi가 제안한 연산자로, 추적 기술에 기초하여 특징점을 정의하였다. 추적이 잘 되는 점을 좋은 특징점으로 정의하고, 식 (2.2)와 같이 고유값이 큰 점을 특징점으로 한다. g_x는 x축 방향으로의 미분을 나타낸다. 그림 (2.5)는 KLT 코너 검출기의 결과를 보여 준다.

$$\begin{aligned}&\min(l_1, l_2) > \text{threshold}, \quad l_1, l_2 : \text{eigenvalues of } \mathbf{G}\\&\text{where}, \ \mathbf{G} = \int_W \mathbf{g}\mathbf{g}^T w d\mathbf{A}, \ \mathbf{g} = [g_x \quad g_y]^T\end{aligned} \tag{2.2}$$

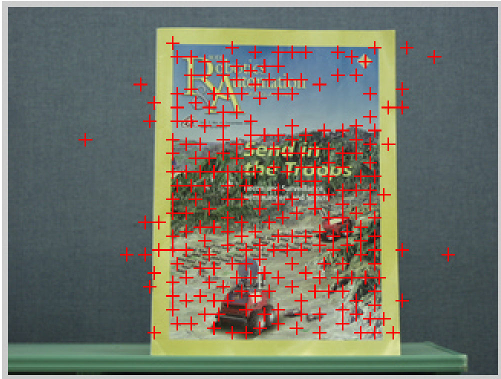


그림 2.5 KLT 코너 검출기 결과

2.2. 특징점 반복성(repeatability)

본 논문에서 제안하는 인식 기법의 기본 전제는 반복성이다 [30]. 모델에 있던 점이 입력 영상의 물체에도 똑같은 위치에 있어야 한다. 따라서, 3차원 공간 상의 점이 물체 영상에 대응되고 이 점이 모델에도 대응되는 반복성(repeatability)이 높을수록 좋은 특징점 추출 연산자라고 정의할 수 있다. 반복성은 식 (2.3)처럼 원래 영상에서 추출한 코너 수와 변환된 영상에서 반복된 코너 수의 백분율로 표현된다.

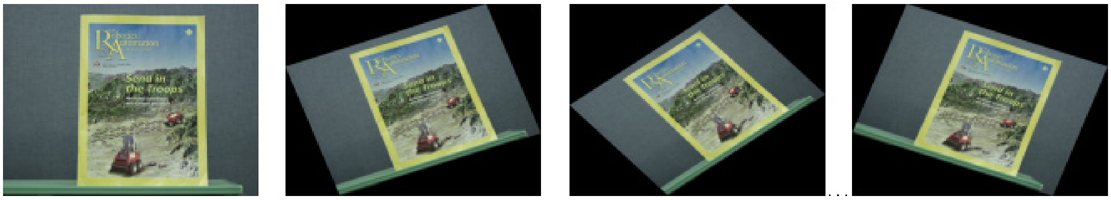
$$R = \frac{N_R}{N_O} \times 100[\%] \tag{2.3}$$

여기서 R은 반복성을 나타내고 N_O는 원래 영상에서 추출한 특징점 수를 나타내고 N_R은 변환된 영상에서 반복된 특징점 수를 나타낸다.

반복성은 세 가지 측면 즉, 물체의 회전 변화, 크기 변화, 조명 변화에 대해서 성능을 비교해 보았다. 반복된 코너는 영상간에 변환 함수를 알고 있을 때, 변환을 통해서 얻은 참값과 코너 연산자를 이용하여 구한 값과의 유클리드 거리가 2픽셀 이내인 경우만을 고려하였다.

2.2.1. 회전 변화에 대한 반복성 비교

물체를 보는 방향(평면 회전)에 따라 특징점이 동일한 점에 위치하는지 각 코너 검출기 별 특성을 살펴본다. 반복성을 계산하기 위해 수학적인 영상 변환을 한다. 그림 (2.6)은 원래 영상과 회전 변환된 영상을 나타낸다.



0 도 20 도 40 도 340 도

그림 2.6 회전 변환된 물체 영상들

그림 (2.7)은 20도 간격으로 회전 변환된 영상의 반복성을 검출기 별로 나타낸 결과이다. 모든 회전에 대해서 Harris 코너 검출기의 반복성이 85%를 항상 넘는다는 것을 알 수 있다.

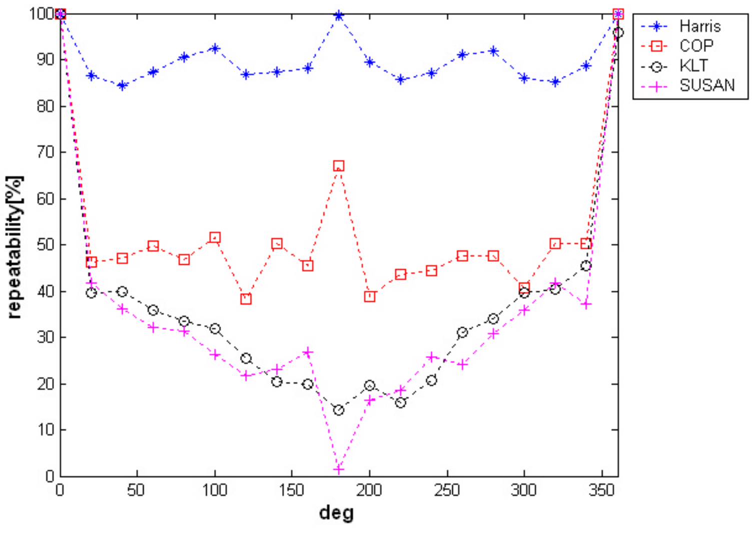
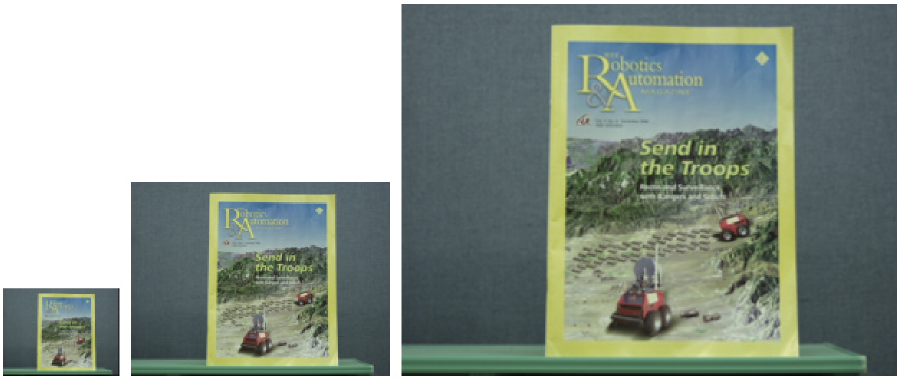


그림 2.7 회전 변화에 대한 반복성 비교

2.2.2 크기 변화에 대한 반복성 비교

물체의 크기 변화에 대해서 같은 점을 찾는 것은 매우 중요하다. 실제 물체 인식을 할 때 크기가 가장 많이 변하기 때문이다. 물체 크기는 반복성 계산을 위해 리샘플링(resampling)과 인터폴레이션(interpolation) 기법으로 변화 시켰다. 그림 (2.8) 은 이렇게 하여 얻은 물체 영상들이다.



0.5 배　　1 배　　2 배

그림 2.8 크기 변환된 물체 영상들

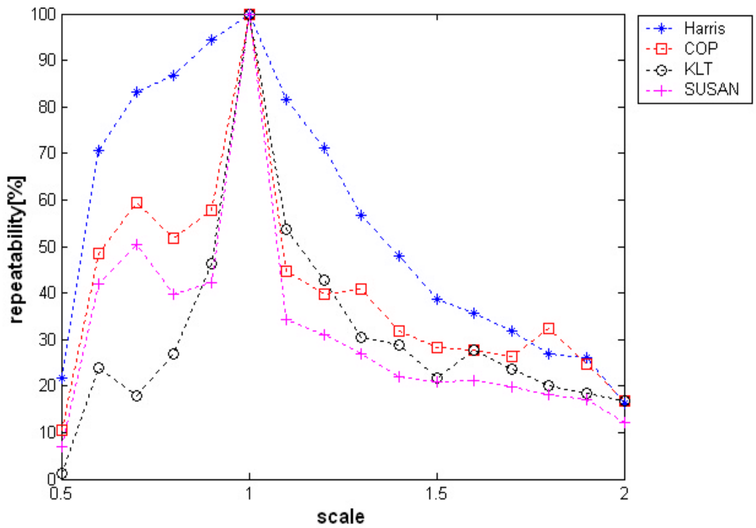


그림 2.9 크기 변화에 대한 반복성 비교

그림 (2.9)는 크기에 따른 각 검출기 별 반복성 특성을 나타낸다. Harris 코너 검출기의 반복성이 가장 높지만, 크기가 1.3배를 넘어서게 되면 반복성이 50% 이하로 떨어지는 문제가 있다. 이러한 문제를 해결하기 위해 다음 장에 스케일 스페이스 이론을 이용하여 해결한다.

2.2.3 조명 변화에 대한 반복성 비교

시간과 장소가 바뀜에 따라 조명이 변하게 된다. 햇빛에 있을 때와 그늘에 있을 때 물체의 밝기 분포가 바뀌게 된다. 이렇게 변할 때에도 물체를 올바르게 인식하기 위해서는 모델과 물체가 있는 영상 간에 동일한 코너 점을 찾을 수 있어야 한다. 그림 (2.10)은 네 가지 조명에 대해서 촬영한 물체 영상들을 나타낸다.

평균밝기: 107　평균밝기: 132　평균밝기: 165　평균밝기:193

그림 2.10 조명에 따른 물체 영상들

그림 (2.11)은 조명 변화에 대한 각 검출기의 반복성 성능을 나타낸다. 반복성 계산은 가장 밝은 영상을 기준 영상으로 하여 계산하였다. 반복성 결과를 보면, SUSAN 코너 검출기와 Harris 코너 검출기가 가장 우수한 반복성 특성을 나타내고, COP 코너 검출기가 조명 변화에 대해서 가장 낮은 반복성 특성을 보인다.

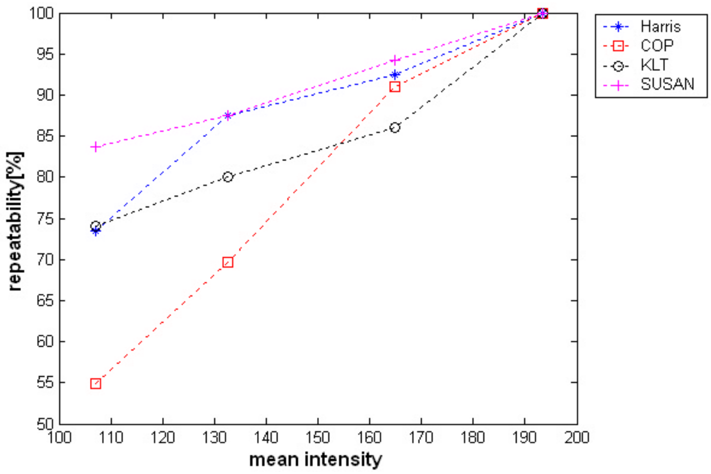


그림 2.11 조명 변화에 대한 반복성 비교

2.2.4 특징점 추출기 결론

Harris 코너 검출기, COP 코너 검출기, KLT 코너 검출기 그리고 SUSAN 코너 검출기에 대한 다양한 실험(회전 변화, 크기 변화, 조명 변화) 결과, 종합적인 반복성 성능이 Harris 코너 검출기, COP 코너 검출기, KLT 코너 검출기, SUSAN 코너 검출기 순으로 우수하다는 결론을 얻었다. 따라서 본 논문에서는 반복성 특성이 전반적으로 우수한 Harris 코너 검출기만을 사용한다.

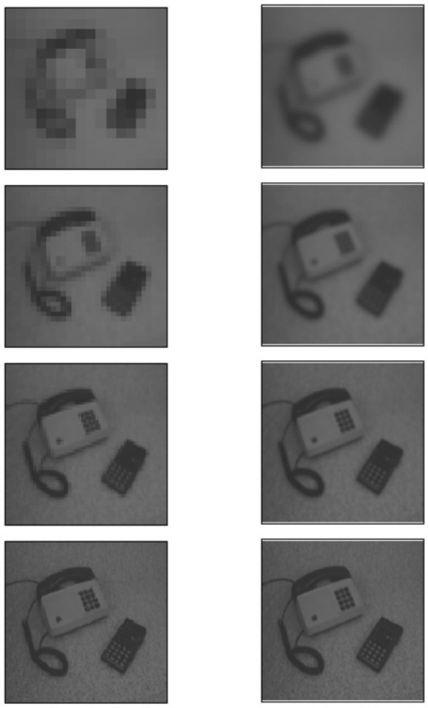
2.3. 스케일 스페이스 표현 (scale space)

특징점 검출기는 물체의 크기에 대해 불변 특성을 갖지 못한다. 앞 절 실험 결과에서 볼 수 있듯이, 크기가 원래 영상의 1.3배 이상이 되면 반복성(repeatability)이 50%이하로 떨어져 물체 인식기가 성능을 제대로 발휘하지 못하는 문제가 발생한다. 물체 크기에 대한 반복성 문제를 해결하고자 스케일 스페이스 이론을 도입한다 [31]. 다음 절에 구체적으로

설명하겠지만, 스케일 스페이스를 적용할 경우, 크기가 다른 물체에 대해 매우 정확하게 코너점을 추출할 수 있어 크기가 변할 때 반복성이 떨어지는 문제점을 보완한다. 스케일 스페이스는 모델 영상에만 적용된다.

2.3.1. 스케일 스페이스 이론

스케일 스페이스에서 영상을 표현하는 것은 다양한 해상도에서 영상을 표현하는 것을 뜻한다. 기존에 제안된 피라미드 구조도 다양한 해상도에서 영상을 표현하는 하나의 방법이다 [31]. 그림 (2.12)의 (a)에서 보듯이 피라미드 구조는 단순히 서브 샘플링(subsampling)과 스무딩(smoothing)을 반복하기 때문에 크기 감소로 인해 계산량을 줄이고 메모리 절감 효과를 가져온다. 그러나, 스케일 방향으로의 양자화가 크고 서브 샘플링에 의존하기 때문에 피라미드 내 영상 구조가 깨지는 문제가 발생한다.



(a) 동일 크기로 한 피라미드 영상 (b) 스케일 스페이스 영상

그림 2.12 피라미드 영상과 스케일 스페이스 영상 [31]

그러나 스케일 스페이스는 식 (2.4)와 같이 한 영상에 대해 가우시안 커널을 이용한 연속 스케일 파라미터로 표현할 수 있기 때문에 영상 내 물체의 구조를 보존할 수 있다. 이것은 그림 (2.12)의 (b)를 보면 알 수 있다. 가우시안 커널을 이용하여 스케일 스페이스로 표현하는 것은 인간의 시각 기능(receptive field profile)과 유사하다는 특성이 있다. 스케일 스페이스의 가장 큰 특징은 특징점을 추출할 경우, 각 스케일에서 매우 정확하게 대응 특징점을 찾을 수 있다는 사실이다. 이것은 간단한 실험을 통해서 검증 하도록 한다.

스케일 스페이스에서 다양한 해상도를 갖는 영상은 가우시안 커널과 콘볼루션(convolution)함으로써 얻을 수 있다. s가 스케일 인자이고, s 가 가우시안 커널의 표준편차이고, $\mathbf{X}=(x,y)$가 픽셀 좌표값, $\mathbf{G}$가 가우시안 커널, I가 영상을 의미할 때 s 레벨에서의 스케일 스페이스 영상은 식 (2.4)와 같이 표현된다.

$$L(\mathbf{X},s)=\mathbf{G}(ss)*I(\mathbf{X}) \tag{2.4}$$

식 (2.4)를 이용하여 구한 특정 스케일의 영상에 대해서 미분연산자, 가령, Harris 코너 검출기를 적용하면 이 스케일에서의 특징점을 구할 수 있다. 문제는 스케일이 커짐에 따라서 공간 미분의 크기가 감소하는 경향이 있다는 사실이다. 스케일에 대해서 Harris 코너 검출기가 불변 특성을 갖기 위해서 식 (2.5)처럼 행렬 $\mathbf{A}$를 구하기 위해 주어진 스케일 인자 s로 정규화 하여야 한다.

$$\mathbf{A}(\mathbf{X},ss,s\tilde{s})=s^2\mathbf{G}(s\tilde{s})*\begin{bmatrix} I^2{}_x(\mathbf{X},ss) & I_xI_y(\mathbf{X},ss) \\ I_xI_y(\mathbf{X},ss) & I^2{}_y(\mathbf{X},ss) \end{bmatrix} \tag{2.5}$$

여기서 $I_x(\mathbf{X},ss)$ 는 $I(\mathbf{X})*\mathbf{G}_x(\mathbf{X},ss)$ 를, $I_y(\mathbf{X},ss)$ 는 $I(\mathbf{X})*\mathbf{G}_y(\mathbf{X},ss)$ 를 나타낸다. $\mathbf{G}(s\tilde{s})$는 편차 $\tilde{s}$ 인 가우시안 커널로 행렬에 가중치를 부여한다. Harris 코너 검출기는 2.1 절에서 언급하였듯이 $\det(\mathbf{A})-a\,\mathrm{trace}(\mathbf{A})$이 문턱치를 넘는 점을 코너점으로 한다.

2.3.2. 스케일 스페이스 구현

가우시안 커널의 s 를 1로 두고 스케일 인자를 1.0부터 3.0까지 0.2 간격으로 총 11

단계의 스케일 스페이스를 얻었다. 커널 크기는 스케일 크기에 비례하도록 하였다. 11 레벨에 대한 스케일 스페이스는 그림 (2.13) 과 같다. 스케일이 커질수록 물체의 세세한 정보가 사라지는 것을 알 수 있다. 스케일 인자와 물체 크기는 반비례 관계에 있다. 즉, 스케일 인자가 클수록 실제로 물체를 축소 시키는 것과 같다.

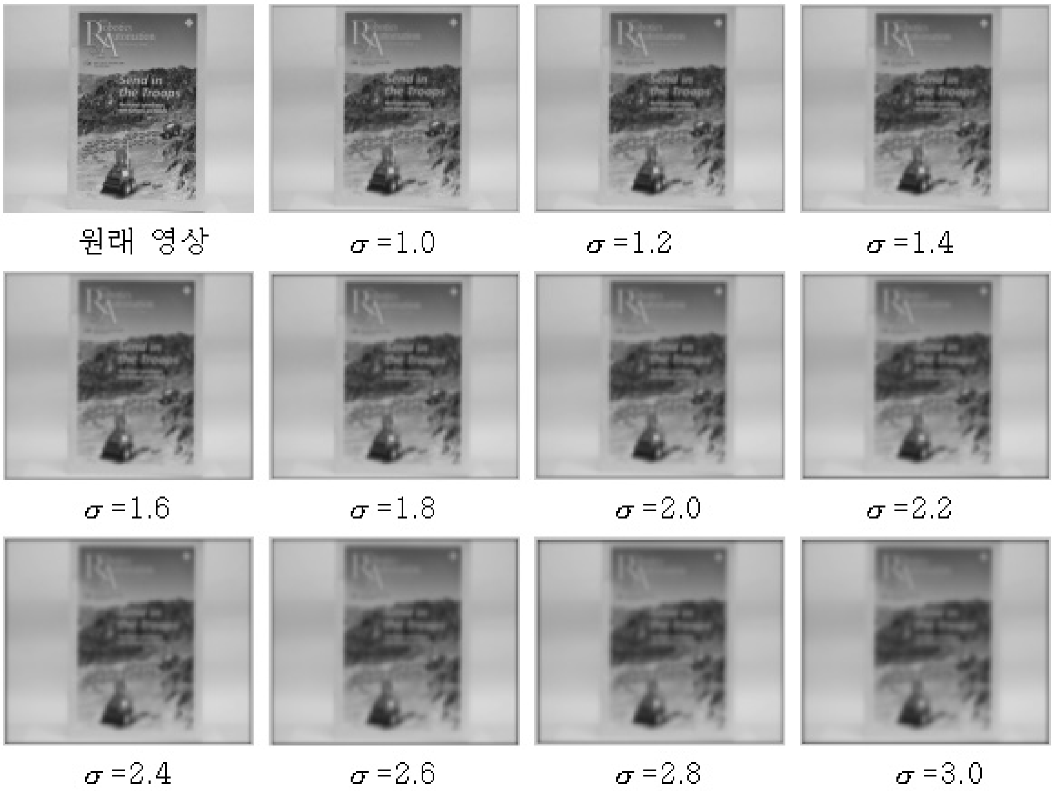
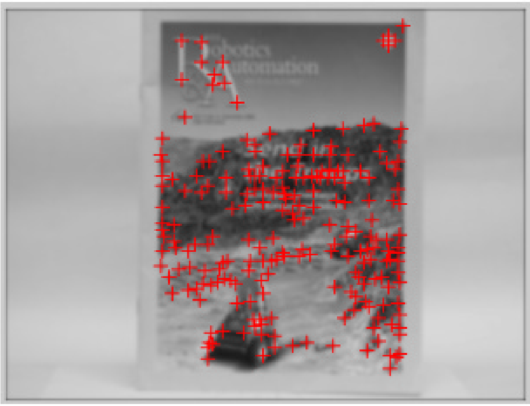


그림 2.13 스케일 스페이스 표현

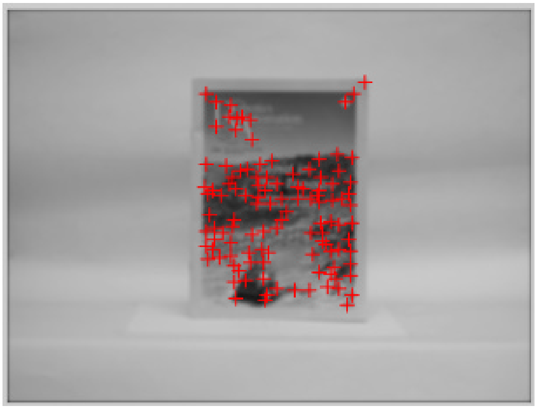
2.3.3. 스케일 스페이스에서의 특징점

스케일 스페이스에서 스케일 인자는 실제 영상에서 크기와 상관 관계가 있다. 그림 (2.14)의 (b)에서 s 가 1.4인 것은 원래 물체 크기를 1/1.4배 하는 것과 같은 효과이다.

두 경우 특징점이 대부분 동일한 위치에 있다는 것을 확인할 수 있다. 이 특성은 본 논문에서 제안하는 물체 인식 기법에서 매우 중요한 부분을 차지한다. 앞 절에서 알 수 있듯이 특징점 추출기는 물체 크기에 불변이지 않다. 물체의 해상도에 따라 코너점의 위치가 변하게 된다. 스케일 스페이스 도입으로 이러한 코너점의 불완전성을 해소할 수 있다. 마찬가지로 s 가 2일 때는 물체 크기가 1/2배 되는 것과 같은 효과를 갖는다.

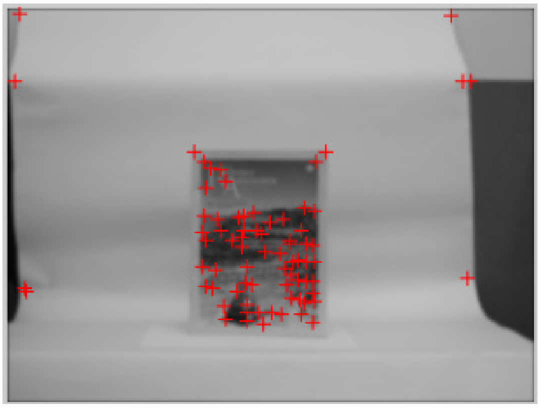


(a) σ=1.4



(b) 실제영상:원래 크기의 1/1.4

(c) σ=2.0



(d) 실제영상:원래 크기의 1/2

그림 2.14 스케일 스페이스와 실제 영상에서의 코너점

2.4. Zernike 모멘트

2.4.1. 이론적 배경

모멘트는 2D 영상에 대해 불변 특성을 제공하기 때문에 패턴 인식등에 주로 사용되어왔다. 모멘트에는 기하 모멘트, 르장드르(Legendre) 모멘트, 저니크(Zernike) 모멘트, 유사 저니크(pseudo-Zernike) 모멘트, 복수(complex) 모멘트 등이 있다. 이들 모멘트들을 잡음 특성, 정보 중복도(aspect of information redundancy), 영상 표현 능력(capability of image representation) 관점에서 특성을 비교 분석한 결과를 보면 [32], 저니크(Zernike)와 유사 저니크(pseudo-Zernike) 모멘트가 우수하다는 것을 알 수 있다.

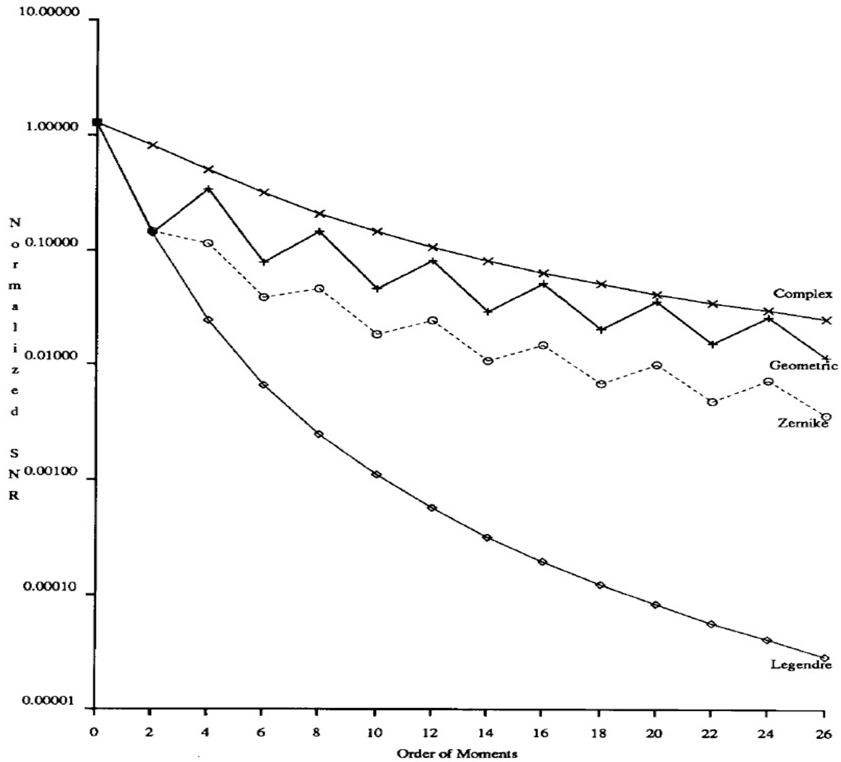


그림 2.15 차수에 따른 각 모멘트의 정규화된 SNR [32]

그림 (2.15)는 차수에 따른 각 모멘트의 잡음에 대한 민감도를 나타낸다. 일반적으로 차수가 높아질수록 잡음에 민감해지는 경향이 있지만 복수(complex), 기하, 저니크(Zernike) 모멘트가 르장드르(Legendre) 모멘트에 비해 잡음에 강한 것을 알 수 있다.

표 2.1 백색 잡음 프로세스에 대한 각 모멘트의 공분산 [32]

Type of Moment		Region of Integration	cov $\{N_{pq}, N_{rs}\}$, Assuming $\sigma_n^2 = 1$
GM	M_{pq}	Square $[-1 \le x, y \le 1]$	$\mathrm{cov}\{N_{pq}, N_{rs}\} = \dfrac{[1-(-1)^{p+r+1}][1-(-1)^{q+s+1}]}{(p+r+1)(q+s+1)}$
LM	λ_{mn}	Square $[-1 \le x, y \le 1]$	$\mathrm{cov}\{N_{mn}, N_{rs}\} = \dfrac{(2m+1)(2n+1)}{4}\delta_{mr}\delta_{ns}$
ZM, PM	A_{nl}	Unit Disk	$\mathrm{cov}\{N_{nl}, N_{mk}\} = \dfrac{n+1}{\pi}\delta_{mn}\delta_{lk}$
RM	D_{nl}	Unit Disk	$\mathrm{cov}\{N_{nl}, N_{mk}\} = \dfrac{2\pi}{n+m+2}\delta_{lk}$
CM	C_{pq}	Unit Disk	$\mathrm{cov}\{N_{pq}, N_{rs}\} = \dfrac{2\pi}{p+q+r+s+2}\delta_{p-q,r-s}$

표 (2.1)은 각 모멘트의 정보 중복도를 비교한 것이다. 정보 중복도는 백색 잡음 프로세스에 대해 각 모멘트의 공분산을 통해서 얻는다. 저니크(Zernike) 모멘트, 유사 저니크(pseudo-Zernike) 모멘트, 르장드르(Legendre) 모멘트만 직교(orthogonal) 모멘트이기 때문에 차수 간에 상관 관계가 전혀 없다. 따라서 정보 중복도가 다른 모멘트에 비해 가장 낮다.

그림 (2.16)은 각 모멘트의 영상 표현력을 비교한 그래프이다. 정보 표현력은 정해진 차수에 대해 재건 에러가 얼마나 작은 지로 비교한다. 그림 (2.16)의 결과에서 보면 모든 차수에 대해서 저니크(Zernike) 모멘트가 르장드르(Legendre) 모멘트보다 재건 오차가 작은 것을 알 수 있다. 따라서 영상 표현력은 저니크(Zernike) 모멘트가 르장드르

(Legendre) 모멘트보다 우수하다는 것을 알 수 있다.

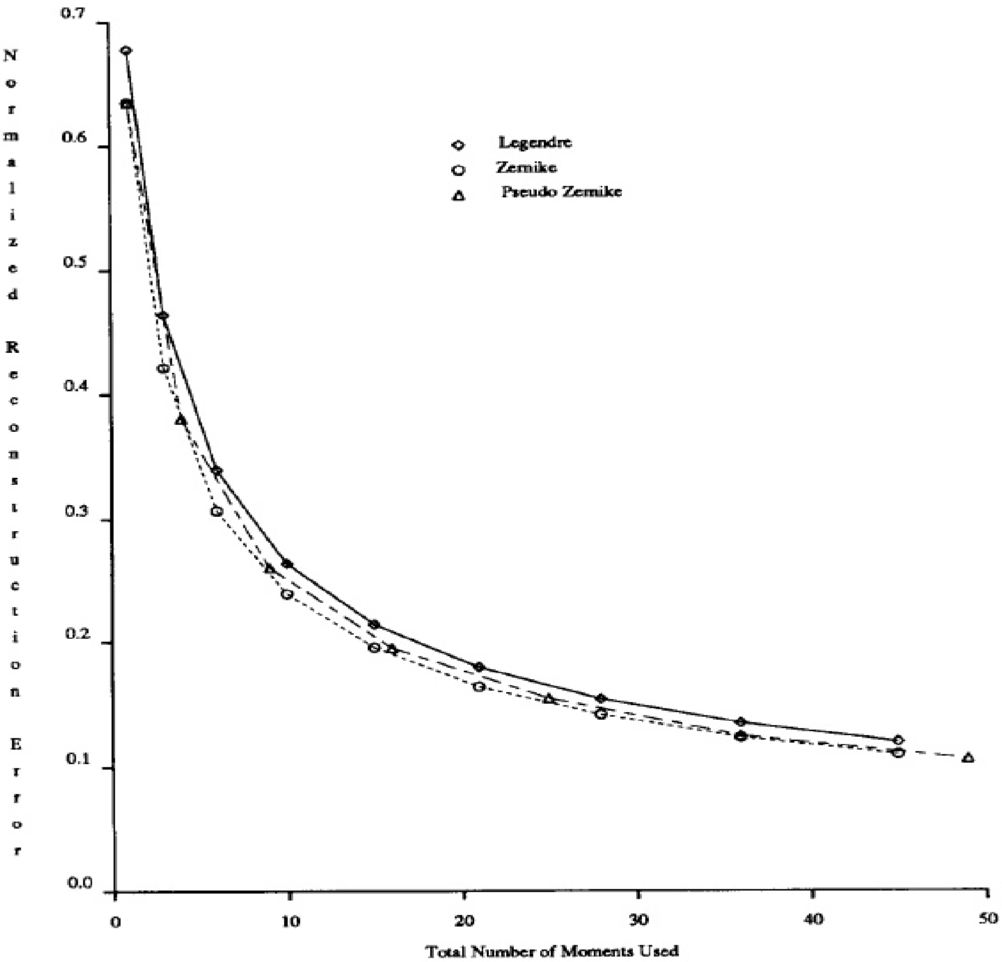


그림 2.16 각 모멘트별 정규화된 재건 에러 [32]

Zernike 모멘트는 1961년 Hu [33]가 소개하여 잘 알려지게 되었다. Reiss [34]가 어파인 변환에 불변인 특성을 갖도록 모멘트의 조합을 수정하였고, Liao와 Pawlak [35]은 Zernike 모멘트의 기하학적, 수치적 에러에 대해서 심도있게 다뤘다. 그들은 Zernike 모멘트가 단위 디스크에서 정의되기 때문에 정확성에 한계가 있다고 보았다. 최근에는 Bigorgne이 Zernike 모멘트를 이용하여 영상 검색을 하였고 [36], Mariani는 Zernike 모멘트를 사용하여 얼굴 인식을 하였다 [37].

2.4.2. Zernike 모멘트의 불변 특성

Zernike 모멘트는 단위 디스크 $x^2 + y^2 \le 1$ 상에서 완전히 직교(orthogonal)하는 복소수 다항식 집합(complex polynomial set)으로 정의된다 [38]. 즉, Zernike 모멘트는 식 (2.6)과 같이 영상 $f(x,y)$ 를 직교기저함수(orthogonal basis function) $V_{nm}(x,y)$ 로 투영한 것으로 볼 수 있다.

$$A_{nm} = \frac{n+1}{p}\sum_{x=0}^{N-1}\sum_{y=0}^{M-1} f(x,y)[V_{nm}(x,y)]^* \tag{2.6}$$

Zernike 다항식(ZP)은 다음과 같이 쓸 수 있다.

$$ZP = \{V_{nm}(x,y) \mid x^2 + y^2 \le 1\} \tag{2.7}$$

Zernike 다항식은 차수가 n 이고 반복(repetition)이 m일 때,

$$V_{nm}(x,y) = R_{nm}(x,y)\exp(jm\tan^{-1}(y/x)) \tag{2.8}$$

이고, 방사 다항식(radial polynomial)은 식 (2.9)와 같이 정의된다.

$$R_{nm}(x,y) = \sum_{s=0}^{(n-|m|)/2} (-1)^s \frac{(n-s)!}{s!\left((\frac{n+|m|}{2}-s)!\right)\left((\frac{n-|m|}{2}-s)!\right)}(x^2+y^2)^{(n-2s)/2} \tag{2.9}$$

단, 다음 조건을 만족해야 한다.

$$n \in N^+, m \in N, (n-|m|): even, |m| \le n \tag{2.10}$$

식 (2.9)에서 $r = \sqrt{x^2 + y^2}$ 로 치환할 경우, 직교방사다항식(orthogonal radial polynomial) $R_{nm}(r)$ 을 살펴 볼 수 있다. 먼저, 차수에 따른 직교방사다항식의 예는 식 (2.11)과 같다.

$$
\begin{aligned}
&R_{00}(r) = 1 \\
&R_{11}(r) = r \\
&R_{20}(r) = 2r^2 - 1 \\
&R_{31}(r) = 3r^3 - 2r \\
&R_{40}(r) = 6r^4 - 6r^2 \\
&R_{51}(r) = 10r^5 - 12r^3 + 3r \\
&R_{60}(r) = 20r^6 - 30r^4 + 12r^2 - 1 \\
&R_{91}(4) = 126r^9 - 280r^7 + 210r^5 - 60r^3 + 5r
\end{aligned}
\tag{2.11}
$$

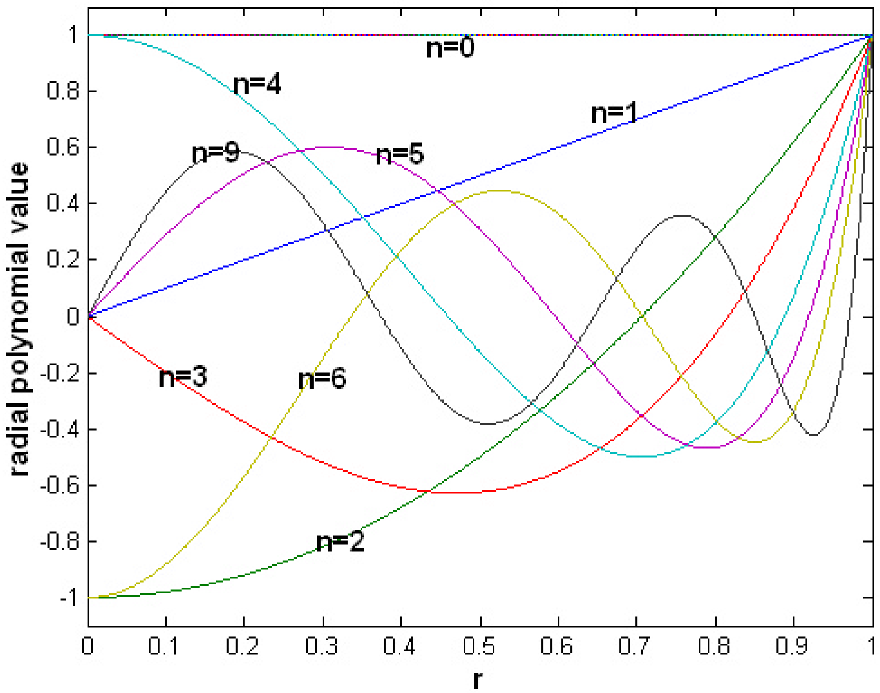


그림 2.17 차수(n)에 대한 직교방사다항식 그래프

그림 (2.17)은 차수가 높아질수록 직교방사다항식이 고주파 형태가 되는 것을 보여준다. 즉, 주어진 영상에 대해서 고차 모멘트를 구하는 것은 영상의 고주파 성분을 뽑아내는 것과 동일하다. 따라서 차수가 높아질수록 모멘트 값이 잡음에 민감해진다는 것도 쉽게 짐작할 수 있다.

한편, 특정 차수 예를 들면, n=9에 대한 반복성(m)은 식 (2.12)와 같다.

$$
\begin{aligned}
R_{91}(r) &= 126r^9 - 280r^7 + 210r^5 - 60r^3 + 5r \\
R_{93}(r) &= 84r^9 - 168r^7 + 105r^5 - 20r^3 \\
R_{95}(r) &= 36r^9 - 56r^7 + 21r^5 \\
R_{97}(r) &= 9r^9 - 8r^7 \\
R_{99}(r) &= r^9
\end{aligned}
\tag{2.12}
$$

마찬가지로 그래프로 나타내보면 그림 (2.18)과 같이 된다. 이 때에는 반복(m)이 커질수록 파형이 저주파 형태가 된다.

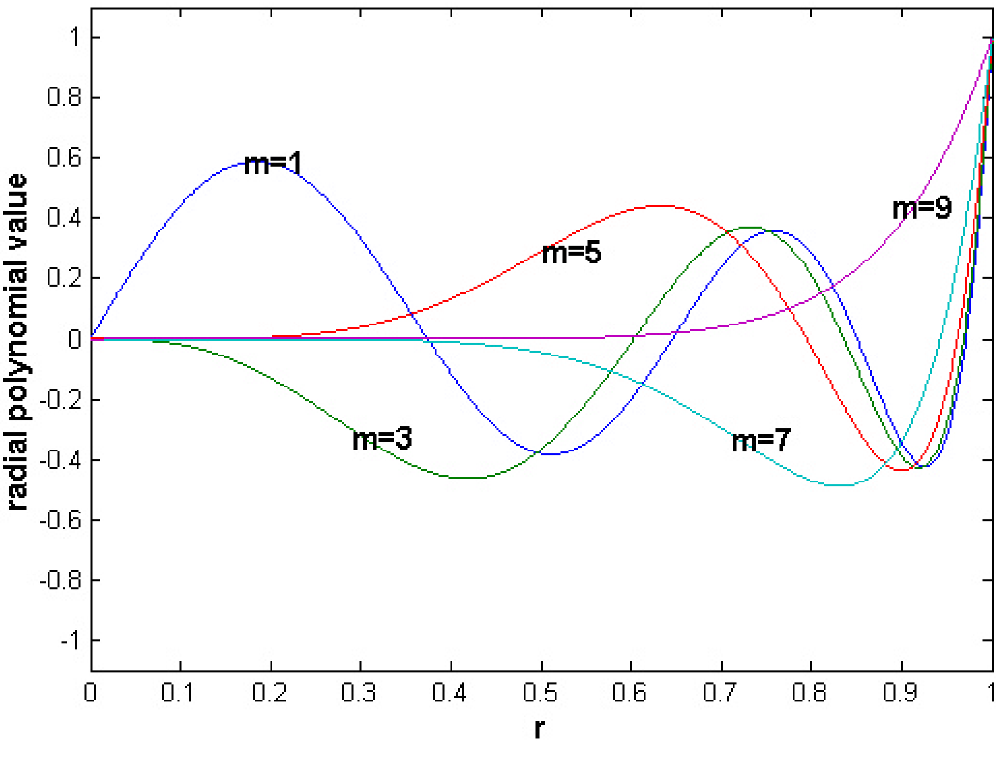


그림 2.18 반복(repetition:m)에 대한 직교방사다항식 그래프

a) 회전 불변 특성

Zernike 모멘트의 가장 큰 강점이 회전에 대해 불변 특성을 갖는다는 사실이다. 회전 불변 특성은 다음과 같이 증명할 수 있다.

$f(r,q)$ 가 원래 영상 $f(x,y)$ 를 극좌표계(polar coordinate)으로 바꾼 것일 때, $f_r(r,q) = f(r,q-a)$ 은 원래 영상을 α만큼 회전된 영상을 나타낸다. 이 영상에 대한 Zernike 모멘트는 연속 공간에서 식 (2.13)과 같이 된다.

$$\begin{aligned} A_{nm}^{r} &= \frac{n+1}{p}\int_0^{2p}\int_0^1 f_r(r,q)[V_{nm}(r,q)]^* r dr dq \\ &= \frac{n+1}{p}\int_0^{2p}\int_0^1 f_r(r,q) R_{nm}(r)\exp(-jmq) r dr dq \\ &= \frac{n+1}{p}\int_0^{2p}\int_0^1 f(r,q-a) R_{nm}(r)\exp(-jmq) r dr dq \end{aligned} \tag{2.13}$$

$b = q - a$ 로 변수를 치환하면,

$$\begin{aligned} A_{nm}^{r} &= \frac{n+1}{p}\int_0^{2p}\int_0^1 f(r,b) R_{nm}(r)\exp(-jm(b+a)) r dr db \\ &= \left[\frac{n+1}{p}\int_0^{2p}\int_0^1 f(r,b) R_{nm}(r)\exp(-jmb) r dr db\right]\exp(-jma) \\ &= A_{nm}\exp(-jma) \end{aligned} \tag{2.14}$$

따라서 $\left|A_{nm}^{a}\right| = \left|A_{nm}\right|$ 이다. 즉, 회전 변환에 대해서 Zernike 모멘트의 크기는 불변이다.

그림 (2.19)는 주어진 영상에 대해서 1도 간격으로 회전하여 얻은 영상들 중 몇 개의 샘플을 나타낸다.

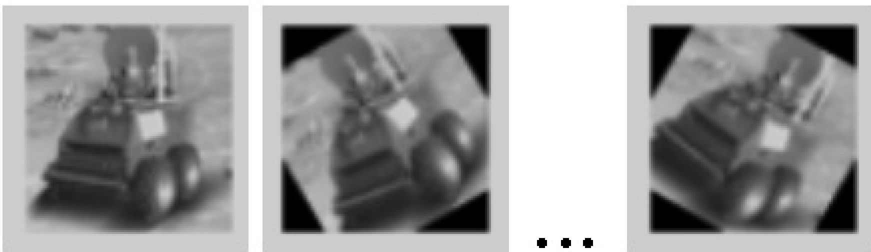


그림 2.19 회전 불변 특성 실험용 영상

그림 (2.20)는 물체를 1도 간격으로 회전해서 1차부터 7차까지의 Zernike 모멘트를 구한 것이다. 회전 변환에 대해 각 차수에 해당하는 모멘트의 크기가 불변 특성을 보인다는 것을 알 수 있다.

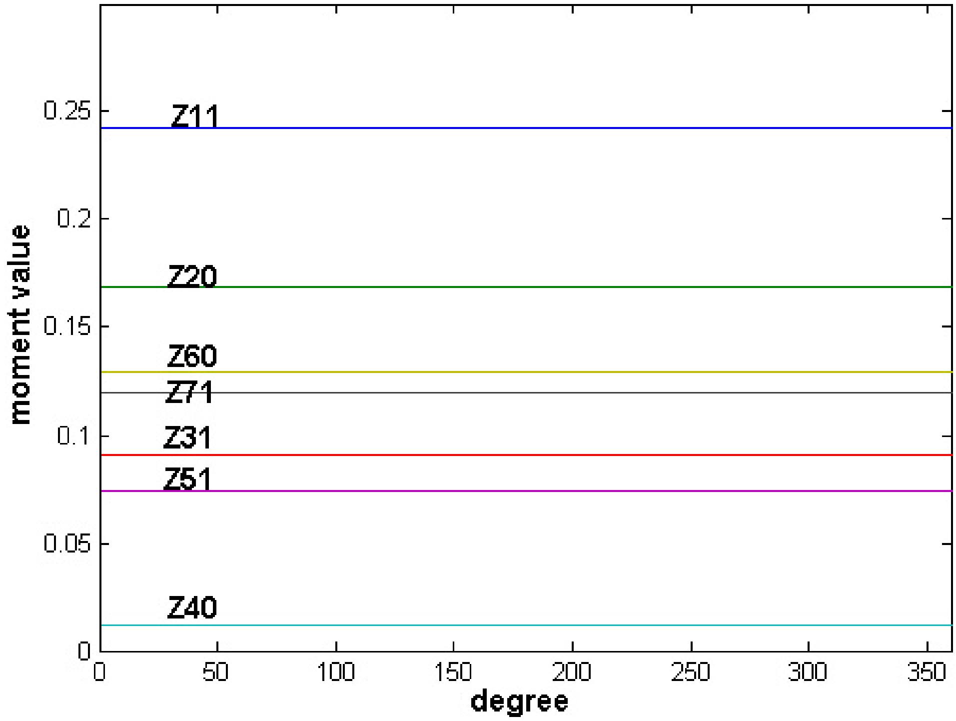


그림 2.20 Zernike 모멘트의 회전 불변 특성 그래프

b) 스케일 불변 특성

지역 특징점에 기반하여 Zernike 모멘트를 계산할 경우, 물체의 크기를 미리 알 수 없다. 또한 Zernike 모멘트의 기하학적인 에러와 수치상 에러가 항상 존재하기 때문에 이러한 점을 고려하여야 한다 [32]. 입력 영상에 대해서 각 코너마다 직경 20 픽셀의 디스크에서 Zernike 모멘트를 추출한다고 할 때, 이러한 문제를 해결하기 위해 모델 영상을 스케일 스페이스로 표현하고 Zernike 모멘트를 각 스케일 스페이스에 대해서 구하였다. 즉, 디스크

직경을 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60 크기로 설정한 뒤 NN(nearest neighbor)기법을 적용하여 크기 20으로 리사이즈(resize)하였다. Bilinear interpolation이나 bicubic interpolation은 분할된 영상의 경계 부근에 왜곡(artifact)이 발생하고 흐려짐(smoothing) 현상이 크게 나타나는 문제가 있다. 이렇게 리사이즈(resize) 하는 것은 스케일 1/3~1배 크기의 물체를 인식할 수 있는 것과 같은 효과를 갖는다. 물체의 모델 영상이 상당히 크기 때문에 오버 샘플링은 배제하였다.

그림 (2.21)은 실제 물체의 모델 영상을 나타낸다. 동그라미 부분은 특징점을 중심으로 하는 지역 영역을 나타낸다.

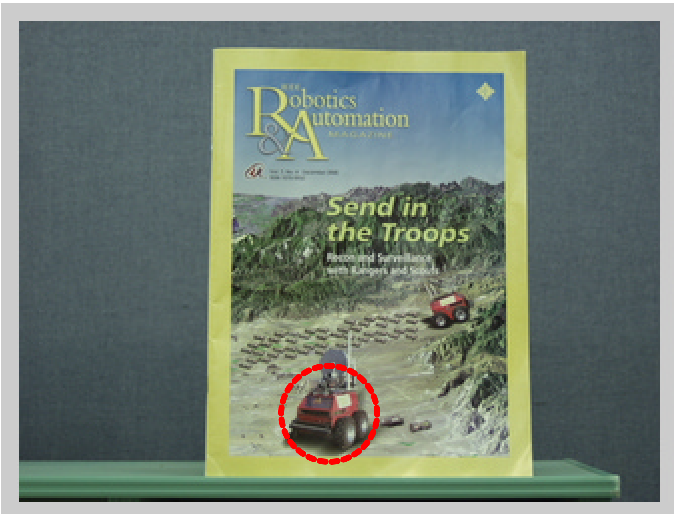


그림 2.21 원래 물체 영상

그림 (2.22)는 위에서 동그라미 한 부분에 대한 스케일 스페이스 영상으로부터 일정한 크기로 리사이즈(resize)한 것이다. 이렇게 함으로써 입력 영상에 다양한 크기의 물체가 들어와도 크기와 무관하게 Zernike 모멘트를 구할 수 있다.

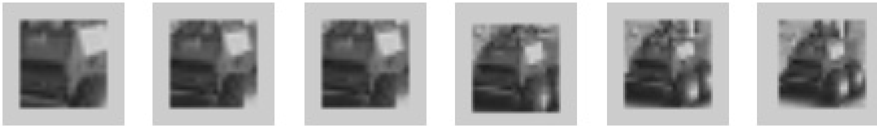
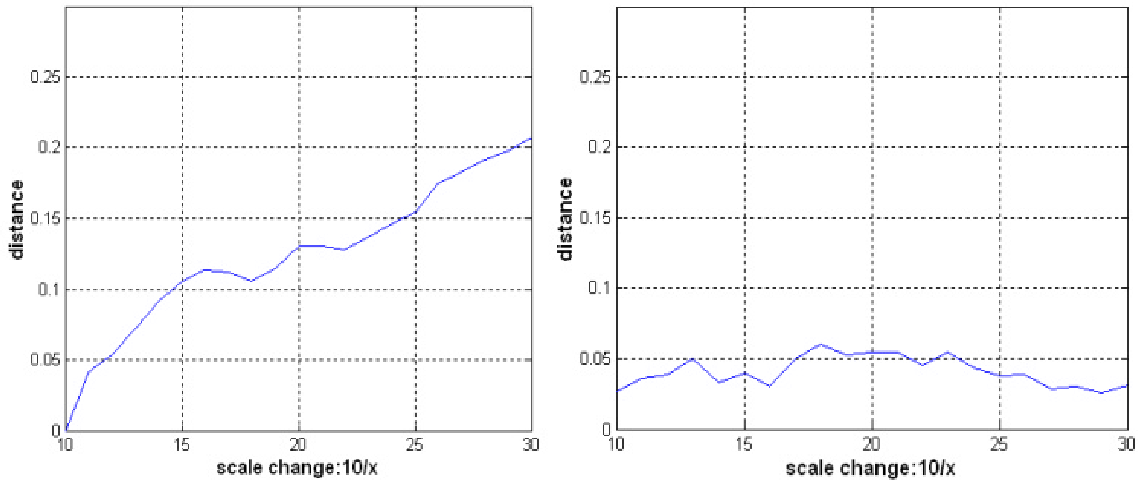


그림 2.22 분할된 지역 영상

그림 (2.23)의 (a)는 입력 물체의 크기를 1.1배 간격으로 작게 하여, 1~1/3스케일의 물체에 대해 스케일 스페이스 방법을 사용하지 않았을 때 모델 Zernike 모멘트와의 유크리드 거리를 나타낸 것이다. 그림 (2.23)의 (b)는 스케일 스페이스에서의 구한 Zernike 모멘트와의 입력 Zernike 모멘트와의 유클리드 거리를 나타낸다. 결과에서 보듯이, 스케일 스페이스를 도입할 경우 다양한 스케일 변화에 대해서 거리 차이가 0.05 내외로 거의 일정한 것을 알 수 있다.



a) 스케일 스페이스 적용하지 않을 때

b)스케일 스페이스 적용했을 때

그림 2.23 스케일 불변 특성

c) 조명 불변 특성

$f(x,y)$가 분할된 영상으로, 원래 물체의 밝기 영상을 나타낸다고 할 때, 스케일 조명 변화가 있을 경우, 새로운 밝기 영상 $f'(x,y)$을 식 (2.15)와 같이 모델링할 수 있다. 이것은 그림 (2.24)와 같은 물리적 현상에 기초한다. 그림 (2.24)에서 첫번째 행은 실제 조명을 변화시켜가면서 촬영한 물체 영상을 나타낸다. 그림 (2.24)의 두번째 행은 인접 영상의 밝기비를 곱하여 밝기가 매우 유사한 결과를 얻은 것을 나타낸다. 따라서, 조명 변화를 식 (2.15)와 같이 스케일 조명 변화로 근사화할 수 있다.

$$f'(x_i, y_i) = a_L f(x_i, y_i) \tag{2.15}$$

여기서 a_L은 지역적인 조명 변화를 의미한다. 지역 영상의 총 매스(mass)-지역 영상의 밝기에 해당-로 Zernike 모멘트를 정규화하면, 즉, Z_{nm}을 Z_{00}으로 나누면

$$\frac{Z(f'(x,y))}{Z_{00}(f')} = \frac{a_L Z(f(x,y))}{a_L Z_{00}(f)} = \frac{Z(f(x,y))}{Z_{00}(f)} \tag{2.16}$$

이다. 여기서 Z_{00}는 영상 $f(x,y)$의 매스(mass), Z는 Zernike 모멘트 연산자를 뜻한다. 따라서 원래 영상을 평균 밝기로 나눠주는 것은 스케일 조명 변화에 대해서 일정하다.

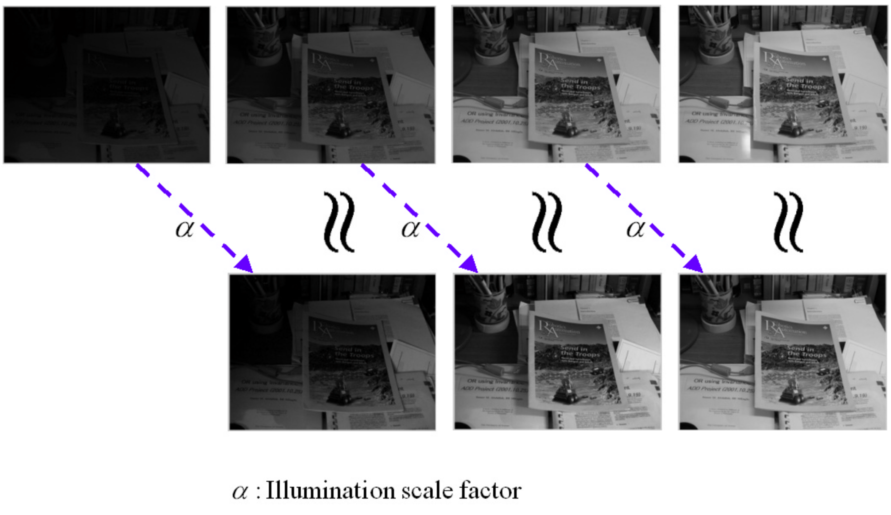
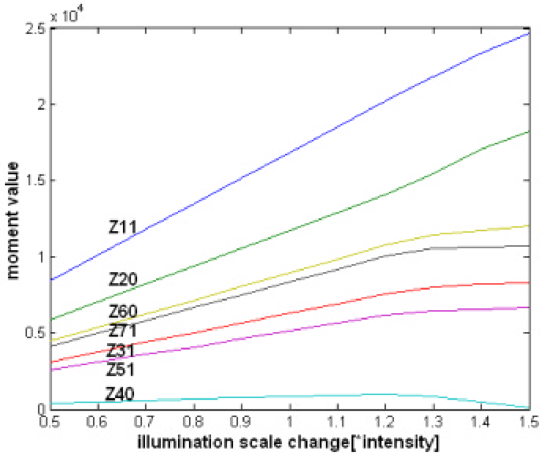


그림 2.24 실제 조명 변화 및 모델링 검증

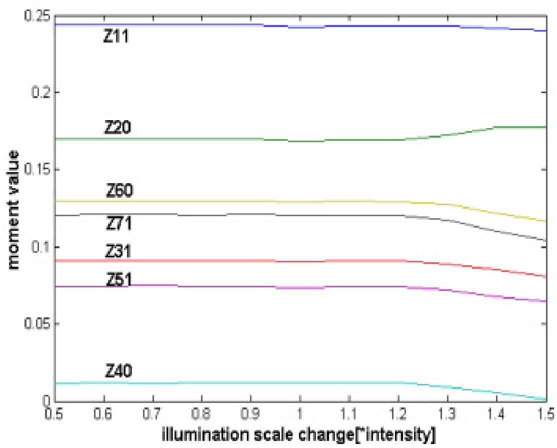
그림 (2.25)는 스케일 조명 변화(0.5~1.5배)를 나타낸 그림이다.

그림 2.25 조명 불변 실험용 영상

위 영상에 대해 원래 Zernike 모멘트를 구하면 그림 (2.26)의 (a)와 같다. 영상의 평균 밝기가 커질수록 이에 비례하여 Zernike 모멘트가 커지는 것을 알 수 있다. 반면, 제안한 방법으로 Zernike 모멘트 크기를 구하면 그림 (2.26)의 (b)와 같다. 이 결과는 영상의 밝기가 0.5~1.5배로 0.1배씩 바뀔 때 각 모멘트를 나타낸 것이다. 밝기 스케일이 0.5~1.3배까지는 거의 일정한 것을 알 수 있다. 스케일이 1.3배보다 커질 경우에는 256 그레이 레벨(gray level)을 벗어나는 픽셀이 많아져 비선형 왜곡이 생겼기 때문이다.



a) 조명불변 특성을 적용하지 않은 경우



b) 조명불변 특성을 적용한 경우

그림 2.26 Zernike 모멘트의 조명 불변 특성 그래프

d) 시각(view angle)에 대한 특성

보완된 Zernike 모멘트 특징량이 시각에 대해서 어떤 특성을 갖는지 파악하는 것은 매우 중요하다. 시각에 대한 특징량의 특성을 파악하기 위해, 그림 (2.27)과 같이 2D 물체와 3D 물체를 0도에서 90도까지 4.5도 간격으로 촬영하였다. 대칭성을 고려해 반대 방향은 생략하였다. 0도 영상을 기준으로 하여, 이 영상에서 추출한 Zernike 모멘트와 회전한 영상으로부터 추출한 Zernike 모멘트의 매칭 비율을 살펴본 것이 그림 (2.28)이다. 매칭은 식 (2.17)과 같이 정의한다.

$$R_M[\%] = \frac{N_A}{N_0} \times 100 \qquad (2.17)$$

여기서 N_0는 0도에서 촬영한 물체의 Zernike 모멘트 개수를 의미하고 N_A는 각도 A에서 촬영한 물체에서 추출한 Zernike 모멘트와 0도에서 촬영한 물체에서 추출한 Zernike 모멘트 사이의 유클리드 거리가 특정 문턱치 이하가 되는 개수를 의미한다. 매칭이 40%이하가 되는 회전각을 보면, 2D 물체는 약 43도, 3D 물체가 약 25도인 것을 알 수 있다.

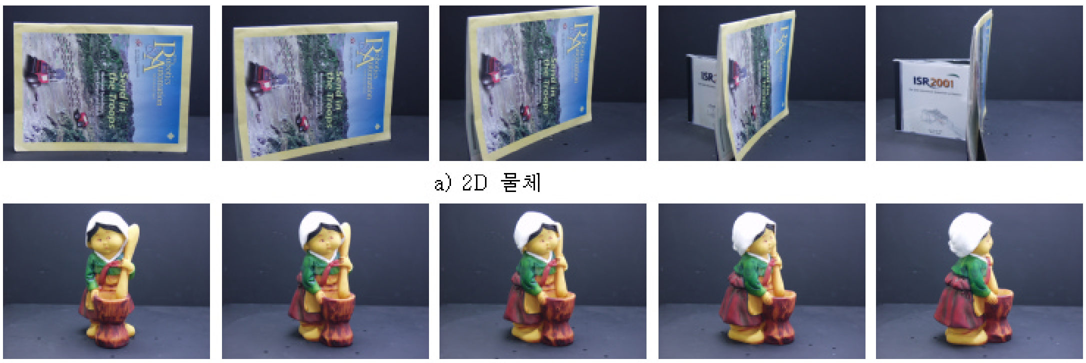


그림 2.27 시각(view angle) 특성 실험용 2D 물체와 3D 물체

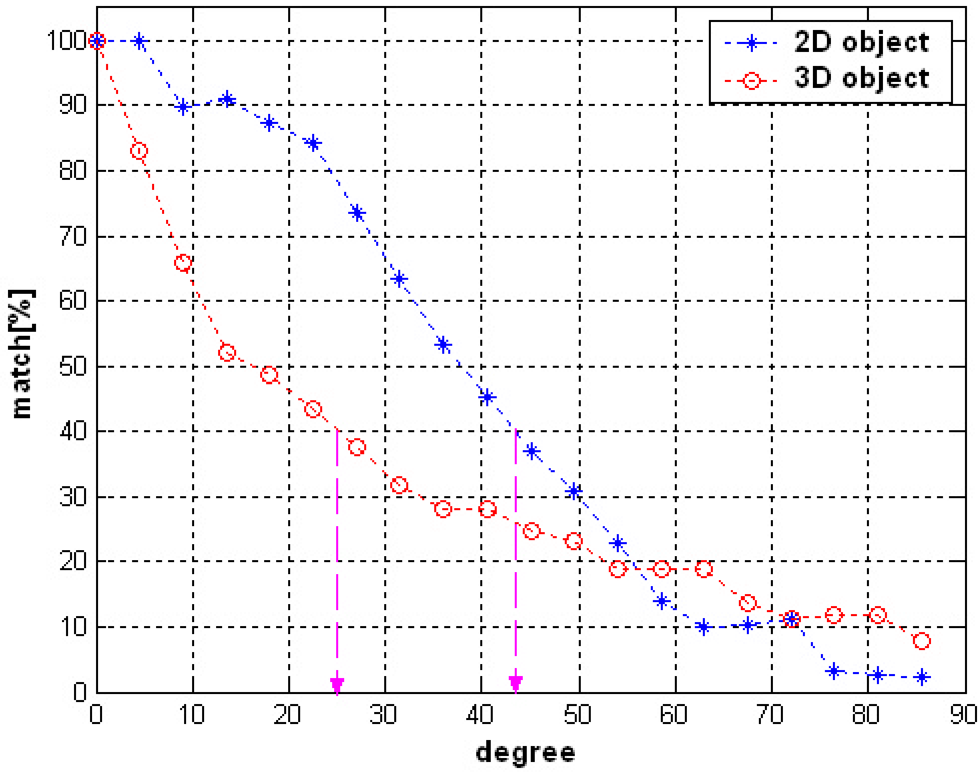


그림 2.28 시각(view angle)에 대한 특징량 매칭

2.5. 특징량 검색

8차 이상의 고차원 벡터 공간에서 특정 벡터와 가까이에 있는 벡터를 빨리 찾는 일은 매우 어렵다. 모든 벡터에 대해 단순한 방법(full search)으로 거리를 계산해서 찾는 것이 가장 정확한 방법이지만, 계산 시간이 벡터의 개수와 차원에 따라 선형적으로 증가하게 된다. 그러나, 가장 가까운 벡터를 근사적으로 구함으로써 탐색 시간을 10배에서 100배 정도 빠르게 할 수 있다 [39].

2.5.1. ANN(approximate nearest neighbor) 검색

가장 가까운 벡터를 찾는 문제는 단순히 전체 검색(brute-force search)을 통해 $O(dn)$ 시간으로 찾을 수 있다. n 은 벡터 개수를 의미한다. 1차원 데이터일 경우 크기 순으로 재배열한 다음 이진 검색(binary search)를 통해 $O(\log n)$ 시간으로 찾을 수 있고, 2차원 데이터일 경우에는 보로노이 다이어그램(Voronoi diagram)을 이용하여 마찬가지로 $O(\log n)$ 시간으로 찾을 수 있다. 문제는 3차원 이상의 벡터이다. 현재까지 3차원 이상의 벡터에 대해서 선형 공간(linear space)과 로그 쿼리 시간(logarithmic query time)을 동시에 만족 시키는 탐색 알고리즘은 없다 [39]. 3차원 이상의 고정된 차원을 갖는 벡터에 대해서 ANN 방법은 효과적인 탐색 방법을 제시한다.

최적 이웃 벡터 문제(Nearest neighbor problem)에서 데이터(S)는 d 차원 공간(R^d) 상에 존재한다. 이러한 데이터는 전처리 되어 구조화된다. 그래서 쿼리(query) 입력 $\mathbf{q}\,(\in R)$가 주어질 경우, 쿼리 $\mathbf{q}$ 와 가장 가까운 k 개의 데이터 값을 효과적으로 찾게 된다. 양의 실수 e 에 대해서, 데이터 벡터와 입력간의 거리가 실제 거리의 $(1+e)$ factor 이내에 있을 때, $\mathbf{q}$ 의 $(1+e)$-ANN을 찾을 수 있다. 두 벡터 간의 거리는 다양한 방법이 있지만 ANN에서는 Minkowski 메트릭을 사용한다. 이 방법은 유클리드 거리(Euclidean distance), 만하탄 거리(Manhattan distance) 그리고 최대 거리(max distance)를 포함한다. 수학적으로 보면, 임의 정수 $m \ge 1$ 에 대해서, $\mathbf{p}=(p_1, p_2, \cdots, p_d)$ 와 $\mathbf{q}=(q_1, q_2, \cdots, q_d)$ 의 거리는 $\sum_{1\le i\le d} |p_i - q_i|^m$ 의 m-th 근으로 정의된다. 가령, L_1, L_2, L_∞ 은 각각 만하탄(Manhattan), 유클리드(Euclidean) 그리고 최대 메트릭(max metric)이라고 부른다.

R^d 공간상에 정의된 데이터 집합 S, query point $\mathbf{q} \in R^d$ 라고 할 때, $e>0$ 에 대해서

$$dist(\mathbf{p},\mathbf{q}) \le (1+e)(\mathbf{p}^*,\mathbf{q}) \quad \mathbf{p}^*: \mathbf{q} \text{ 와 가장 가까운 데이터값} \tag{2.18}$$

을 만족하는 $\mathbf{p}$ 가 찾고자 하는 ANN이다.

ANN은 다음과 같은 성능을 낸다.

R^d 공간상에 n 개의 데이터 집합 S 가 있을 때, 상수 $c_{d,e} \le d[1+6d/e]^d$,

구조크기(structure size) $O(dn)$, 데이터 검색 시간(data construction time) $O(dn\log n)$ 일 때, 어떤 Minkowski 메트릭에 대해서도 다음 사실이 성립한다.

i) 주어진 $e > 0$, $\mathbf{q} \in R^d$ 에 대해서 q 의 ($1+e$)-ANN은 S 공간에서 $O(c_{d,e} \log n)$ 시간으로 찾을 수 있다.

ii) 일반적으로 $e > 0$, $\mathbf{q} \in R^d$, 임의의 k ($1 \le k \le n$), 에 대해서 q 의 k 개 ($1+e$)-ANN은 S 공간에서 $O((c_{d,e} + kd)\log n)$ 시간으로 찾을 수 있다.

ANN 검색 알고리즘은 다음과 같은 두 가지 특성을 갖는다. 하나는 시스템이 데이터를 저장하기 위해 e 에 무관하게 dn 공간만을 필요로 하기 때문에 파라미터 설정 관점에서 최적이다. 다른 하나는, 전처리 과정이 e 와 메트릭(metric)에 무관하기 때문에 데이터 구조가 미리 만들어졌더라도 쿼리(query)에 대한 데이터 값은 임의 e 과 Minkowski 메트릭에 대해서도 구할 수 있다는 점이다.

ANN에서 사용한 데이터 구조는 BBD(balanced box-decomposition) 트리(tree)라고 불리는 계층적 공간 분할을 이용한 것이다. 공간은 반복해서 셀(cell)로 분할되는데 각각은 d 차원의 사각형 모양이다. 트리의 노드는 셀과 연관되어 이 셀 안에 있는 데이터 집합을 함축한다. 각 리프셀(leaf cell)은 셀(cell) 안에 있는 하나의 작은 셀을 말한다. 트리 구조는 $O(n)$ 노드로 구성되고 $O(dn\log n)$ 시간 안에 만들 수 있다. 다음은 ANN 쿼리 검색 과정을 설명한다.

Step1: 주어진 쿼리 q 에 대해서 이 쿼리를 포함하는 리프 셀(leaf cell) 을 찾는다. 이것은 트리 구조에서 단순경사(simple descent) 방법을 이용하여 $O(\log n)$ 시간 안에 찾는다.

Step2: 쿼리에서 거리가 멀어지는 순으로 리프 셀(leaf cell) 에 번호를 매긴다. 이것은 우선 검색(priority search) 이라고도 부른다. 셀(Cell)을 만나게 되면 그 셀에 있는 점과 쿼리 점과의 거리를 계산하고 가장 가까운 점들을 따라간다. 그림 (2.29) 는 그러한 부-분할(sub -division) 과정을 나타낸다.

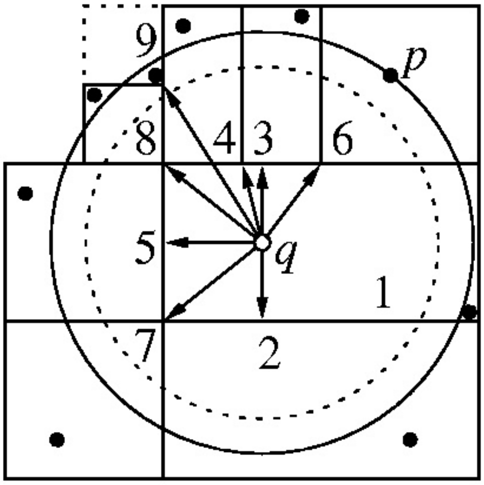


그림 2.29 ANN 검색 과정 [39]

Step3: p 가 찾아낸 가장 가까운 점이라고 할 때, q 에서 현재 리프 셀(leaf cell) 까지의 거리가 $dist(\mathbf{q},\mathbf{p})/(1+e)$ **를 넘는 순간(그림 2.29에서 점선), 탐색 과정은 끝나게 되고 p 가 q 의 ANN이 된다.**

3. 확률적 보팅을 이용한 물체 인식

3.1. 확률적 보팅(probabilistic voting)

입력 영상에 어떤 물체가 있을 때 이 물체가 무엇인지 알기 위해 확률적 물체 인식 기법을 사용한다. 확률적 물체 인식은 특징량의 대응쌍으로 구성되는 가설을 생성하고 각 가설에 합리적인 확률값을 부여하여 확률값이 최대가 되는 물체를 찾는 것이다. 먼저 확률적 모델 생성을 위한 기본 표기법(notation) 및 가설 생성에 대해 알아본다.

입력 영상을 S, 모델 집합을 $M=\{M_i\},\ i=1,\ 2,\ \cdots,N_M$, N_M 은 모델의 개수, 입력 영상 S 의 Zernike 모멘트 집합 $Z=\{\mathbf{Z}_i\},\ i=1,\ 2,\cdots,N_S$, 여기서 $\mathbf{Z}_i$ 는 특징점 C_i 에서 계산된 하나의 Zernike 모멘트를 나타내고, N_S 는 입력 영상 S 의 Zernike 모멘트 개수를 나타낸다. 입력 영상의 i 번째 Zernike 모멘트 $\mathbf{Z}_i$ 에 대응하는 모델 영상의 Zernike 모멘트는 $\hat{Z}_i=\{\hat{\mathbf{Z}}_i^j\}\ \ j=1,\ 2,\cdots,N_C$ 이고, N_C 는 대응 Zernike 모멘트의 개수를 나타낸다.

3.1.1. 확률적 물체 인식 모델 생성

물체 인식은 Bayes 이론에서 위험 함수(risk function)를 최소화함으로써 최적 물체 인식을 수행할 수 있다 [43]. 최적 물체 모델 M_i^* 의 추정에 대한 위험 함수(risk function)는 식 (3.18)과 같이 정의 된다. 증명의 편의상 M_i 는 실수라고 가정한다.

$$R(M_i^*)=\int_{M_i\in M} C(M_i^*, M_i)P(M_i \mid S)dM_i \tag{3.1}$$

여기서 $C(M_i^*,M_i)$ 는 비용 함수(cost function)를 나타내고, $P(M_i\,|\,S)$ 는 사후(posteriori) 확률 분포를 나타낸다. 사후 확률은 선험(a priori) 확률과 가능도(likelihood) 확률을 이용하여 구할 수 있다. 사후 확률은 Bayes 이론에 의해 식 (3.2)와 같이 표현된다.

$$P(M_i | S) = \frac{P(S | M_i)P(M_i)}{P(S)} \tag{3.2}$$

$P(M_i)$ 는 물체 모델 M_i 의 선험 확률을 나타내고, $P(S|M_i)$ 는 고정된 S 에 대해 M_i 의 함수인 가능도(likelihood)를 나타낸다.

비용 함수는 $C(M_i^*, M_i)$ 는 참값이 M_i^* 일 때, 물체 모델 M_i 의 추정 비용을 나타낸다. 비용함수는 이차함수 형태와 식 (3.3)과 같은 $d(0-1)$ 비용함수가 있다.

$$C(M_i^*, M_i) = \begin{cases} 0 & if \quad \left\| M_i^* - M_i \right\| \le d \\ 1 & \quad otherwise \end{cases} \tag{3.3}$$

이차비용함수를 이용한 위험함수는 식 (3.4)과 같다.

$$R(M_i^*) = \int_{M_i \in M} \left\| M_i^* - M_i \right\| P(M_i | S) dM_i \tag{3.4}$$

식 (3.4)를 M_i^*에 대해 미분하고 0으로 두면 식 (3.5)와 같이 된다.

$$M_i^* = \int_{M_i \in M} M_i P(M_i | S) dM_i \tag{3.5}$$

이 결과는 사후 확률의 평균이 된다. 한편, d 비용함수에 대한 위험 함수는 아래 식과 같다.

$$R(M_i^*) = \int_{M_i : \left\| M_i^* - M_i \right\| > d} P(M_i | S) dM_i = 1 - \int_{M_i : \left\| M_i^* - M_i \right\| \le d} P(M_i | S) dM_i \tag{3.6}$$

$d \rightarrow 0$ 일 때, 위험함수는 식 (3.7)과 같이 근사화 된다.

$$R(M_i^*) = 1 - kP(M_i | S) \tag{3.7}$$

k 는 $M_i : \left\| M_i^* - M_i \right\| \le d$ 을 만족시키는 부피를 의미한다. 식 (3.7)을 최소화하는 것은 사후 확률 $P(M_i|S)$ 를 최대화하는 것과 동일하다.

따라서 최적 모델 추정을 위해 위험 함수를 최소화하는 것은 식 (3.8)과 같이 같다.

$$M_i^* = \arg\max_{M_i \in M} P(M_i | S) \tag{3.8}$$

위에서 최적 모델 추정을 위해 M_i 를 실수값으로 일반화하여 유도하였다. 유도 결과는

M_i가 자연수 값을 가지는 경우에도 마찬가지로 성립한다.

따라서 오인식 위험(risk)을 최소화하는 물체 인식은 식 (3.8)과 같이 입력 영상에 대해서 가장 높은 확률값을 갖게 하는 모델을 찾는 것이다.

입력 영상 S 의 Zernike 모멘트와 대응 모델 Zernike 모멘트들로 구성되는 대응쌍을 이용하여 다음과 같은 가설을 만들 수 있다.

$$H_i = \left\{ (\hat{\mathbf{Z}}_1^j, \mathbf{Z}_1) \cap (\hat{\mathbf{Z}}_2^k, \mathbf{Z}_2) \cap \cdots (\hat{\mathbf{Z}}_{N_S}^l, \mathbf{Z}_{N_S}) \right\}, \quad i = 1, \ 2, \cdots, N_H \tag{3.9}$$

여기서 N_H는 입력영상에 대응하는 모델 Zernike 모멘트의 곱집합의 원소 개수와 같고, j, k, l 은 입력 특징량에 대응하는 모델 Zernike 모멘트의 인덱스를 나타낸다. 입력 영상의 전체 가설은 식 (3.10)과 같이 표현된다.

$$H = \left\{ H_1 \cup H_2, \cdots, \cup H_{N_H} \right\} \tag{3.10}$$

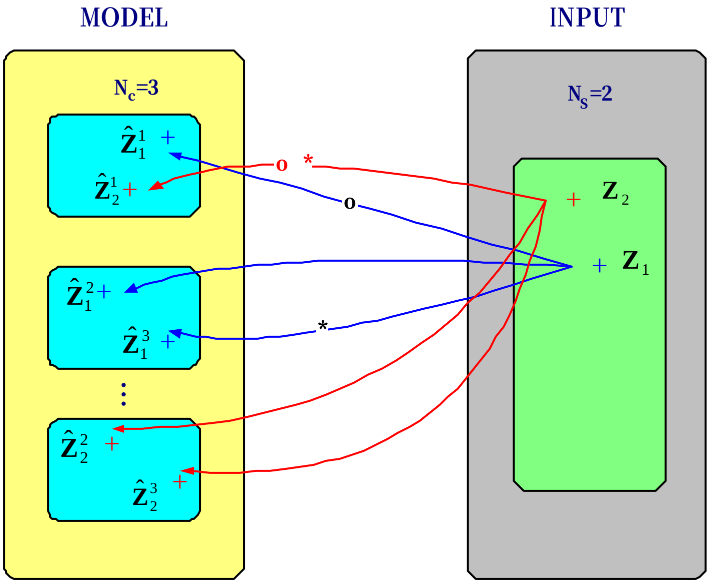


그림 3.1 가설 생성용 대응쌍 예

그림 (3.1)은 입력에서 특징량이 N_S =2이고 각 특징량마다 대응 특징량 N_C =3일 경우 2개의 가설을 나타낸다. 가설1은 'o'로 표시한 부분으로 $H_1 = \left\{ (\hat{\mathbf{Z}}_1^1, \mathbf{Z}_1) \cap (\hat{\mathbf{Z}}_2^1, \mathbf{Z}_2) \right\}$과 같다. 가설2는 '*'로 표시한 부분으로 $H_2 = \left\{ (\hat{\mathbf{Z}}_1^3, \mathbf{Z}_1) \cap (\hat{\mathbf{Z}}_2^1, \mathbf{Z}_2) \right\}$ 와 같다. 위 예에서 가설의 총

개수는 $N_C^{N_S}=3^2=9$와 같다.

한편, 식 (3.9)와 식 (3.10)으로부터 입력 영상 S는 가설 H로 대체할 수 있으므로 $P(M_i|S)$는 $P(M_i|H)$와 동등하다. 또한, 식 (3.8)에 Bayes 이론을 적용시키면 식 (3.11)과 같은 결과를 얻을 수 있다.

$$P(M_i \mid H) = \frac{P(H \mid M_i)P(M_i)}{P(H)} \tag{3.11}$$

모든 물체가 발생할 확률 $P(M_i)$이 동일하고, 각 물체가 서로 독립이라고 가정하면, 식 (3.11)은 다음과 같이 표현할 수 있다.

$$P(M_i \mid H) = \frac{\sum_{h=1}^{N_H} P(H_h \mid M_i)P(M_i)}{\sum_{h=1}^{N_H} P(H_h)} \tag{3.12}$$

분모에서 각 가설에 대한 확률 $P(H_h)$는 전확률 이론(theorem of total probability)에 의해 $P(H_h)=\sum_{i=1}^{N_M} P(H_h \mid M_i)P(M_i)$과 같다.

식 (3.12)에서 $P(M_i)$이 각 물체에 대해서 동일하다고 가정하였고, 분모가 공통되기 때문에 사후 확률값을 결정짓는 중요한 부분은 바로 선험(a priori) 확률 $P(H_h|M_i)$이다. 선험 확률을 구할 때 다음 사항들을 고려한다 (그림 (3.2) 참고).

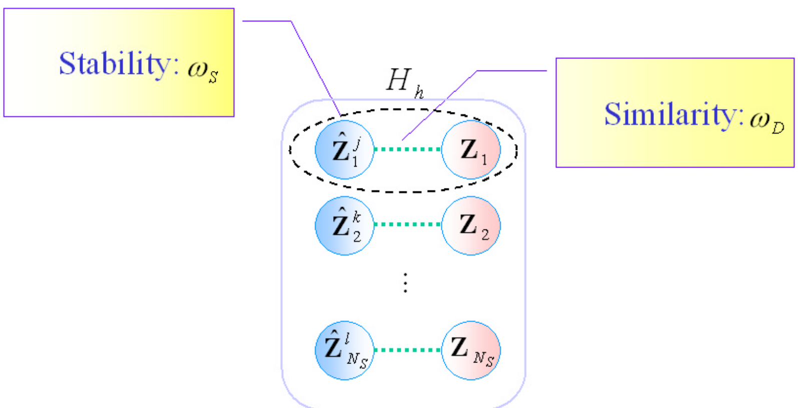


그림 3.2 선험 확률 부여시 고려 사항

1) 대응하는 모델 Zernike 모멘트와의 유사도(w_D)

입력 Zernike 모멘트에 대응하는 모델 Zernike 모멘트와의 유클리드 거리가 가까울수록 모델의 특징량과 입력 특징량이 옳게 매칭될 확률이 높아진다. 따라서, 식 (3.13)과 같이 w_D와 거리(distance)는 반비례 관계에 있다.

$$w_D \propto \frac{1}{\text{distance}} \tag{3.13}$$

2) 대응하는 모델 Zernike 모멘트의 안정도(w_S)

입력 Zernike 모멘트에 대응하는 모델 Zernike 모멘트 안정도를 반영한다. Zernike 모멘트의 안정도는 Zernike 모멘트의 위치에 대한 민감도에 반비례한다. Zernike 모멘트의 민감도는 식 (3.14)와 같이 잡음에 의해 특징점이 a, b, c, d 로 바뀌게 될 때 중심점 e를 기준으로하는 Zernike 모멘트의 표준 편차로 정의 된다.

$$sensitivity = \sqrt{\frac{1}{4}\left[\|\mathbf{Z}_a - \mathbf{Z}_e\|^2 + \|\mathbf{Z}_b - \mathbf{Z}_e\|^2 + \|\mathbf{Z}_c - \mathbf{Z}_e\|^2 + \|\mathbf{Z}_d - \mathbf{Z}_e\|^2\right]} \tag{3.14}$$

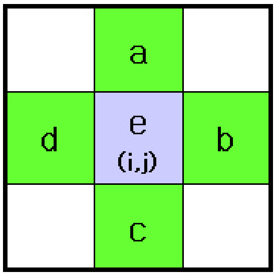


그림 3.3 Zernike 모멘트의 민감도 계산을 위한 특징점 위치

그림 (3.3)에서 (i, j) 는 이상적인 경우의 특징점 위치를 나타낸다. 실제 영상에서는 잡음등의 원인 때문에 특징점의 위치에 오차가 발생할 수 있다. 이러한 위치 오차에 따라 생성되는 특징량인 Zernike 모멘트값도 바뀌게 된다. 민감도가 큰 모델 특징량과 대응할수록 잘못 매칭될 확률이 높아진다. 따라서 이러한 특징점의 위치 오차에 의해 발생되는 Zernike 모멘트의 불확실성을 안정도(stability) 개념을 도입하여 확률적 물체 인식 패러다임에 반영하도록 한다. 식 (3.14)로부터 안정도는 다음 관계에 있다.

$$w_S \propto \frac{1}{sensitivity} \tag{3.15}$$

대응쌍의 유클리드 거리가 가까우면서 모델 특징량의 안정도가 높으면 이 쌍이 제대로 매칭될 확률이 높다. 따라서 선험 확률은 식 (3.16)과 같이 표현할 수 있다.

$$P(H_h \mid M_i) = \prod_{j=1}^{N_S} P((\hat{\mathbf{Z}}_j^k, \mathbf{Z}_j) \mid M_i) \tag{3.16}$$

여기서 $\hat{\mathbf{Z}}_j^k$ 는 *j*번째 입력 특징량에 대응하는 모델 특징량들 중 *k*번째 특징량을 나타낸다. 각 대응쌍에 대한 확률값은 식 (3.17)과 같이 정의한다.

$$P((\hat{\mathbf{Z}}_j^k, \mathbf{Z}_j) \mid M_i) = \begin{cases} \exp\left(-\dfrac{1}{w_S(\hat{\mathbf{Z}}_j^k) w_D(\hat{\mathbf{Z}}_j^k, \mathbf{Z}_j) a}\right) & if \quad \hat{\mathbf{Z}}_j^k \in \hat{Z}(M_i) \\ e & else \end{cases} \tag{3.17}$$

안정도 w_S 와 유사도 w_D 는 식 (3.18), 식 (3.19)와 같이 정의된다.

$$w_S(\hat{\mathbf{Z}}_j^k) = \frac{1}{sensitivity(\hat{\mathbf{Z}}_j^k)} \tag{3.18}$$

$$w_D(\hat{\mathbf{Z}}_j^k, \mathbf{Z}_j) = \frac{1}{\left\|\hat{\mathbf{Z}}_j^k - \mathbf{Z}_j\right\|} \tag{3.19}$$

식 (3.16) 에서 N_S 는 입력 영상의 특징량 개수를 나타내고, 식 (3.17) 에서 a 는 정규화 인자이고, $\hat{Z}(M_i)$ 는 M_i 모델에 해당하는 특징량 집합을 나타낸다. e 은 대응 모델 특징량이 조건을 만족하지 않을 경우 부여하는 값이다. 본 실험에서는 0.2로 설정하였다.

대응쌍 후보는 앞에서 소개한 ANN (approximate nearest neighbor) 검색 알고리즘을 사용하여 찾는다. 각 입력 영상의 Zernike 모멘트에 대해 k-nearest neighbor를 찾아서 가설을 생성하고 각 가설에 대해 위에서 제안한 확률적 보팅 기법을 적용한다.

3.1.2. 단순 보팅(voting)과 제안한 확률적 보팅

단순 보팅(voting)은 각 입력 특징량이 모델 특징량에 해당 되면-이것은 Mahalanobis

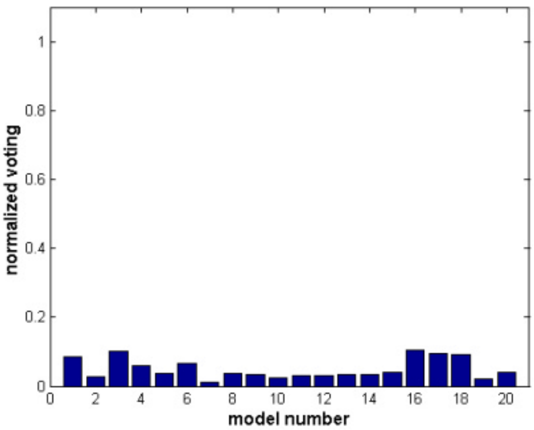
거리나 유클리드 거리로 결정할 수 있다-그 모델에 값을 증가 시키는 방법이다. 수학적으로 표현하면 식 (3.20)과 같다.

$$V(j)=\sum_{i=1}^{N_S}\sum_{k=1}^{N_C} 1 \;\; if \;\; \left\|\hat{\mathbf{Z}}_j^k-\mathbf{Z}_i\right\|\le T \;\;\&\;\; \hat{\mathbf{Z}}_j^k\in\hat{Z}(j) \qquad (3.20)$$

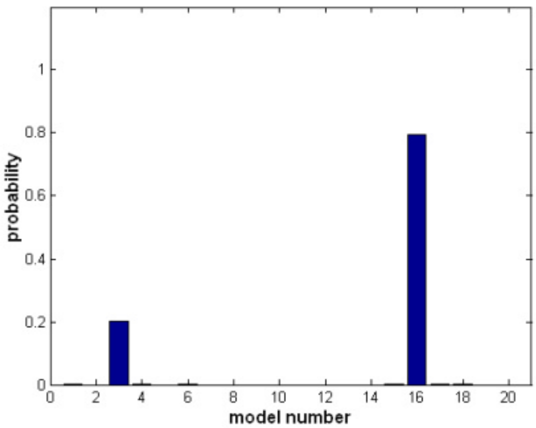
여기서 N_S 는 입력 특징량 개수를 나타내고, N_C 는 대응 모델 특징량 개수를 나타낸다. T 는 유클리드 거리에 대한 문턱치를 의미한다. 이 방법은 매우 빠른 특성이 있지만, 일반 환경에서는 적용하기 힘든 경우가 있다.

(a) 입력 영상

(b) 16번 모델 영상



(c) 단순 보팅



(d) 확률적 보팅

그림 3.4 단순 보팅(voting)과 확률적 보팅 비교

그림 (3.4)는 단순한 보팅 방법을 사용했을 때와 위에서 제안한 확률적 보팅 기법을 적용했을 때 결과를 나타낸다. 복잡한 배경에 16번 물체가 입력으로 들어왔을 때, 확률적 보팅 기법이 단순히 보팅(voting)만 적용했을 때보다 얼마나 모델을 잘 찾는지 보여준다. 이것은 제안한 방법이 매우 강력하다는 것을 의미한다. 좀 더 자세한 것은 4장에서 다루도록 한다.

3.2. 물체 인식 검증

확률적 물체 인식을 통해 입력 영상에 어떤 물체가 있는지 알 수 있다. 그 다음 선택된 모델이 입력 영상에 있는 물체가 맞는지 확인하는 검증 과정이 필요하다. 이것은 모델과 입력 영상과의 대응쌍을 통해서 확인할 수 있다. 물체 인식 검증은 그림 (3.5)처럼 대응쌍 필터링 과정을 거쳐 모델을 입력 영상에 정렬(alignment)하는 전역 매칭을 통해 확인한다. 이 과정은 지역 특징량을 이용하는 방법을 보완하는 과정이면서 결과적으로 입력 영상에 있는 물체의 정확한 위치와 자세도 알 수 있는 특징이 있다.

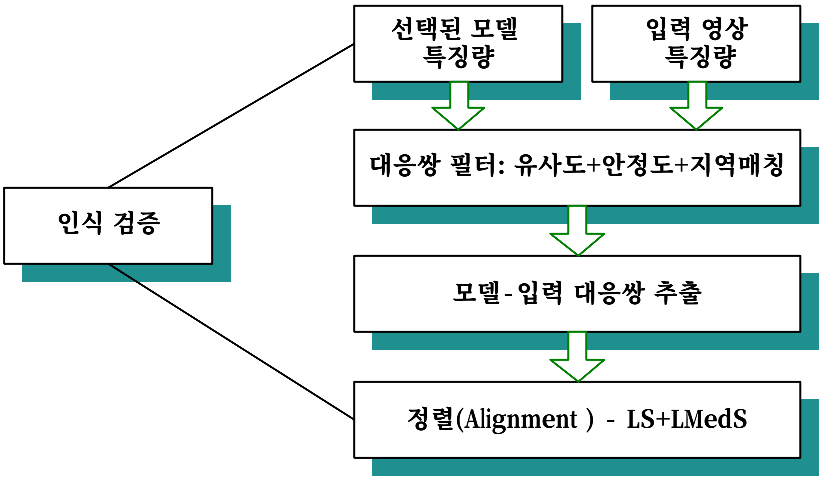
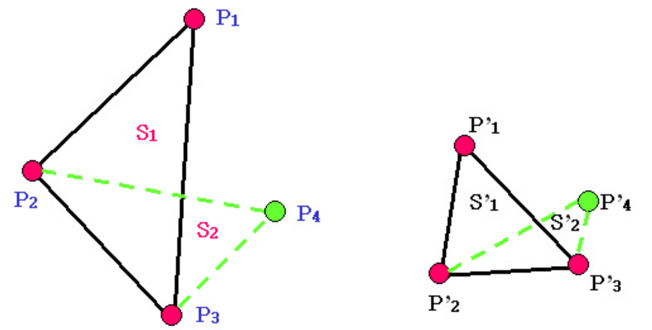


그림 3.5 물체 인식 검증 과정

3.2.1 대응쌍 필터링

확률적 물체 인식을 통해 찾은 모델 영상과 입력 영상과의 안정적인 대응점을 찾기 위해서는 대응쌍 필터링 과정이 필요하다. 대응쌍 검색은 확률적 물체 인식 과정에 사용했던 ANN 기법을 적용한다. 여기서는 각 입력 영상의 Zernike 모멘트에 대응하는 모델 영상의 1 nearest neighbor를 찾은 뒤, 거리 제한(distance constraint), 안정도 제한(stability constraint) 그리고 공간 매칭 제한(local spatial matching constraint)을 두어, 가장 안정적으로 대응쌍을 찾도록 한다. 거리 제한과 안정도 제한은 각 매칭쌍에 문턱값을 주어 1차로 필터링하고 2차는 공간 매칭을 통하여 필터링한다. 공간 매칭은 대응되는 연속 네 쌍의 점을 사용한다. 그림 (3.6)에서 세 점으로 구성되는 삼각형의 면적비(area ratio)를 적용한다. 면적비는 어파인(affine) 변환에 대해서 불변이기 때문에 이 조건을 사용하는 것은 타당하다. 공간 매칭 제한을 이용한 필터링 과정은 다음과 같다. 먼저 모델과 입력 영상 사이의 대응쌍 네 점을 선택한다. 다음, 모델에서 $\frac{\Delta P_2 P_3 P_4}{\Delta P_1 P_2 P_3}$, 입력 영상에 대해서 $\frac{\Delta P'_2\, P'_3\, P'_4}{\Delta P'_1\, P'_2\, P'_3}$ 를 구하여 두 값의 비가 허용값을 넘게 되면 네 번째 쌍을 제거한다. 이 때, 앞의 세 쌍은 매칭 되었다고 가정한다.



(a) 모델 영상의 지역 특징점 (b) 입력 영상의 지역 특징점

그림 3.6 공간 매칭 제한

그림 (3.7)은 이러한 과정을 통해 얻은 결과를 나타낸다. 대응점을 선으로 연결하였다. 대부분 대응쌍의 위치가 일치하는 것을 확인할 수 있다.

(a) 모델 영상 (b) 입력 영상

그림 3.7 모델 영상과 입력 영상의 대응쌍 필터링 결과

3.2.2 정렬 (Alignment)

대응쌍으로부터 모델이 어디에 어떤 자세로 있는지 알기 위해 호모그래피(homography)를 이용한다 [40]. 호모그래피는 동차(homogeneous) 좌표계에서 2장의 2D 영상을 맵핑 해주는 행렬이다. 모델에 있는 영상과 입력에 들어온 영상과는 어파인 변환 관계에 있다고 가정을 한다. 동차 좌표계에서 모델점 $(x,y)^T$ 는 $(x \;\; y \;\; 1)^T$ 로 표현하고, 어파인 변환(affine transformation)된 영상상의 점 (x',y')는 $(x', \;\; y', \;\; 1)^T$ 로 표현한다. 이 때 어파인 변환식은 식 (3.21)과 같다.

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} m_1 & m_2 & t_x \\ m_3 & m_4 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{3.21}$$

어파인 변환은 식 (3.21)에서 보는 것처럼 6파라미터로 표현할 수 있다. 모델의 이동은 (t_x, t_y) 에 반영되어 있고, 모델의 어파인 회전(affine rotation), 스케일(scale), 확장(stretch)는 m_i 에 반영된다. 실제로 본 논문에서 모델은 심도(depth)가 그렇게 크지 않은 3차원 물체를 60도 간격으로 모델링하기 때문에 어파인 변환 가정은 타당하다.

호모그래피을 구하기 위해 LS (least square)와 LMedS (least median of squares)를 결합 시키는 방법을 제안한다. 이 방법은 50% 이상점(outlier)에 강인한 특성이 있다.

a) 초기값 선택 (H_0)-LS

LMedS를 이용하여 호모그래피를 최적화하기 위해서는 초기값을 잘 선정하여야 한다. 본 논문에서는 대응쌍을 랜덤하게 샘플링하여(최소 4쌍) LS(least square) 방법으로 호모그래피를 구한다. 이상점(outlier) 개수가 최소가 되게 하는 호모그래피를 초기 호모그래피로 설정한다.

선택한 대응쌍이 n 개이고, 선택한 n 개의 모델점을 $\mathbf{X}_m$, 입력 영상의 점 $\mathbf{X}_s$ 라고 할 때, 대응쌍은 호모그래피에 의해 식 (3.22)와 같은 관계를 갖는다.

$$\mathbf{X}_s = \mathbf{H}_0 \mathbf{X}_m \tag{3.22}$$

위 식의 해는 양변에 $\mathbf{X}_m{}^T$ 를 곱하여 구한다.

$$\begin{aligned} \mathbf{X}_s \mathbf{X}_m{}^T &= \mathbf{H}_0 \mathbf{X}_m \mathbf{X}_m{}^T \\ \mathbf{H}_0 &= \mathbf{X}_s \mathbf{X}_m{}^T (\mathbf{X}_m \mathbf{X}_m{}^T)^{-1} \end{aligned} \tag{3.23}$$

여기서 문제는 $\mathbf{H}_0$ 는 랜덤하게 선택한 여러 개의 대응쌍으로부터 구하고 이 $\mathbf{H}_0$ 로부터 이상점(outlier)이 최소가 되게 하는 $\mathbf{H}_0$ 를 선택한다. 반복하는 횟수는 식 (3.24)에 기초한다 [41].

$$N = \frac{\ln(1-C)}{\ln(C-(1-e)^S)} \tag{3.24}$$

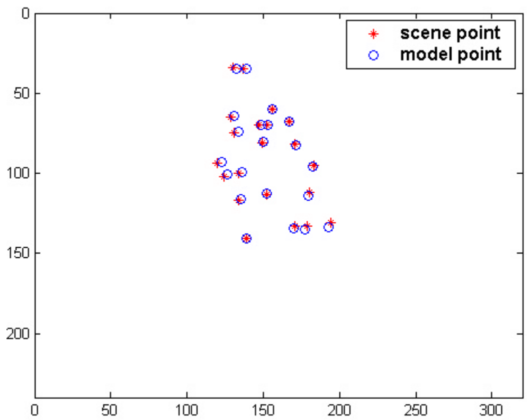
식 (3.24)에서 N 은 반복횟수, C 는 신뢰도(confidence), e 은 이상점(outlier) 비율, S 는 사용된 대응쌍의 개수를 의미한다. 본 논문에서 사용된 값은 $C = 0.95$, $e = 0.4$, $S = 3$ 일

때 $N = 9.7$ 로, 올림하여 10회 반복하였다.

참고로 그림 (3.7) 의 모델 영상 (a)와 입력 영상 (b)에 대해 구한 결과 이상점(outlier)이 최소가 되는 $\mathbf{H}_0$ 는 식 (3.25)와 같이 나왔다.

$$\mathbf{H}_0 = \begin{bmatrix} -0.0065 & -0.4635 & 157.8246 \\ 0.4462 & -0.0253 & 110.5371 \\ 0 & 0 & 1 \end{bmatrix} \tag{3.25}$$

식 (3.25)로 구한 초기 호모그래피($\mathbf{H}_0$)로 모델 영상을 입력 영상에 정렬한 결과는 그림 (3.8)과 같다.



(a) 변환된 모델 특징점과 입력 영상 특징점

(b) 변환된 모델 영상의 에지

그림 3.8 초기 호모그래피로 정렬한 결과

b) 최적 변환-LMedS

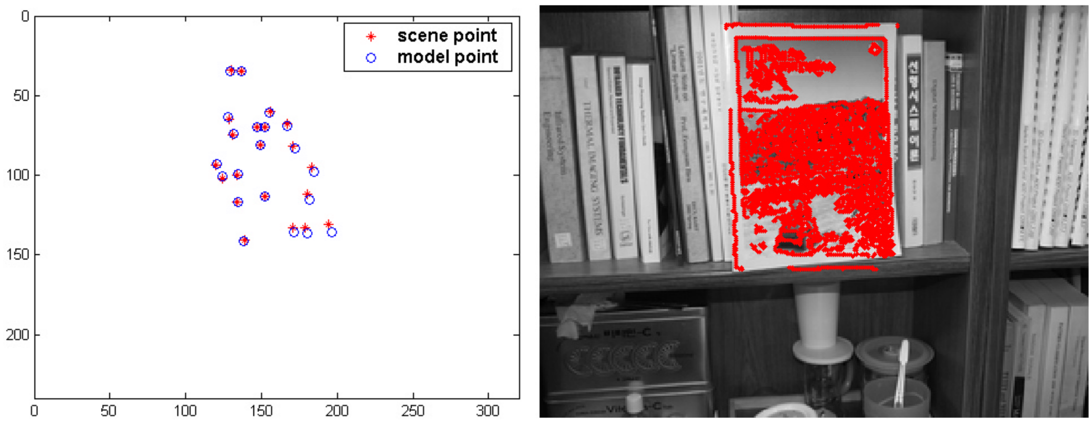
앞 절에서 LS 방법으로 구한 $\mathbf{H}_0$ 를 초기값으로 하여 LMedS (least median of squares)를 Nelder-Mead 방법 [42]으로 호모그래피를 최적화 (refine) 시킨다. (식 (3.26) 참조).

$$\min_{i=1,2,\cdots N} med \left\{ (\mathbf{HX}_{Mi} - \mathbf{X}_{Si})^2 \right\} \tag{3.26}$$

Nelder-Mead 방법은 미분을 이용하지 않고 simplex를 이용하여 다변수를 최적화 시키는 알고리즘이다. 또한 LMedS는 50% 이상점(outlier)에 대해 강인하게 호모그래피를 찾을 수 있다. 이러한 방법을 이용하여 위에서 구한 호모그래피를 최적화(refine)한 결과는 식 (3.27)과 같다.

$$\mathbf{H} = \begin{bmatrix} -0.0069 & -0.4635 & 157.6104 \\ 0.4462 & -0.0253 & 110.5371 \\ 0 & 0 & 0.9968 \end{bmatrix} \tag{3.27}$$

이 결과는 초기값과 약간 다르게 수치가 조정된 것을 알 수 있다. 식 (3.27)을 이용하여 모델을 입력 영상에 정렬(alignment)시킨 결과는 그림 (3.9)와 같다. 그림 (3.8)의 (a)와 그림 (3.9)의 (a)를 비교해 보면, 모델의 특징점이 입력 영상의 특징점에 맞게 약간 조절된 것을 알 수 있다. 또한 최종 이상점(outlier)이 0.0%를 나타내 성공적으로 인식이 검증되는 것을 알 수 있다. 만약 이상점(outlier)이 50.0%를 넘으면 모델 검증이 실패한 것으로 판단한다.



(a) 변환된 모델 특징점과 입력 영상 특징점 (b) 변환된 모델 영상의 에지

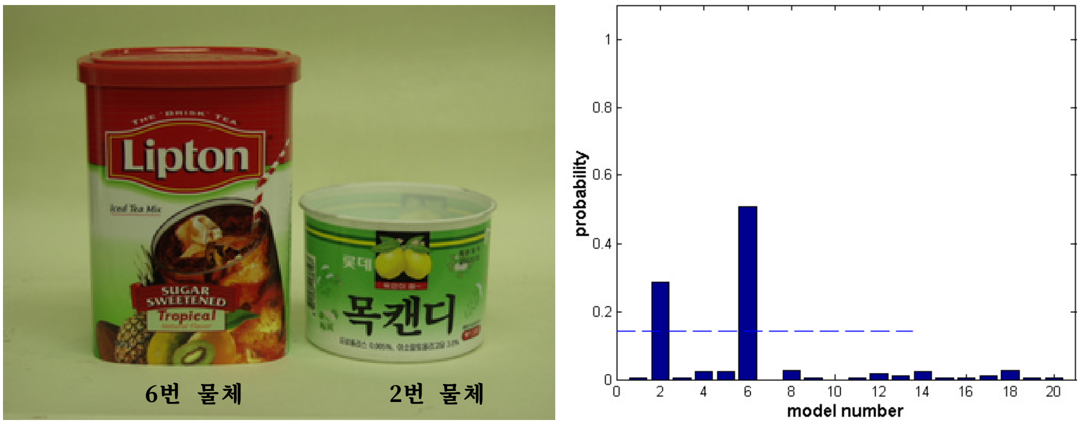
그림 3.9 최적 호모그래피로 정렬한 결과

위와 같은 모델 검증 과정을 통해 확률적으로 물체를 인식한 것이 맞은지 틀린 지 알

수 있고, 맞다면 찾은 물체가 어디에 어떤 자세로 있는지 자세(pose)까지 구할 수 있는 장점이 있다.

3.3. 다중 물체 인식

입력 영상에 물체 여러 개가 있는 경우가 있다. 이 때, 물체 인식 기법은 기존 인식 프레임을 유지한다. 다만, 그림 (3.10)과 같이 확률적 보팅을 하기 전에 단순 보팅을 통해 각 모델의 분포를 미리 계산한다 (prior 모델 생성). 각 모델의 존재 확률이 0.15를 넘는 모델의 개수를 체크하고 이 개수만큼 인식을 수행한다. 단, 인식 과정을 반복할 때 앞 단계에서 인식한 물체는 제외하고 차례대로 인식을 수행한다.



(a) 입력 영상　　　　(b) 단순 보팅 결과

그림 3.10 각 물체의 존재 확률 (prior 모델)

3.4. DB에 없는 물체 처리

입력 영상에 DB에 없는 물체가 들어올 수 있다. 이 물체의 특징점 분포가 DB의 특징치와 유사한 경우 확률적 보팅 과정을 통해 오인식을 하게 된다. 그러나, 물체 인식 검증 과정을 통해 대응쌍이 모두 필터링 되거나 이상점(outlier)이 50%를 넘게 되면 입력 영상에 대한 최종 인식을 실패하므로, 결국 reject하게 된다.

4. 물체 인식 실험

4.1. 특징량 차수 실험

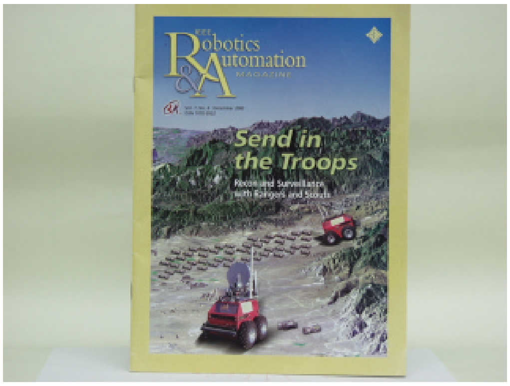
Zernike 모멘트의 차수를 정하는 것은 매우 중요하다. 차원을 정하기에 앞서 변별인자(discriminative factor) 개념을 제안하다. 물체 인식 관점에서 볼 때, 물체를 변별하는 능력은 매우 중요하다. 따라서 변별인자를 다음과 같이 정의한다.

$$DF = \frac{N_C}{N_D} \tag{4.1}$$

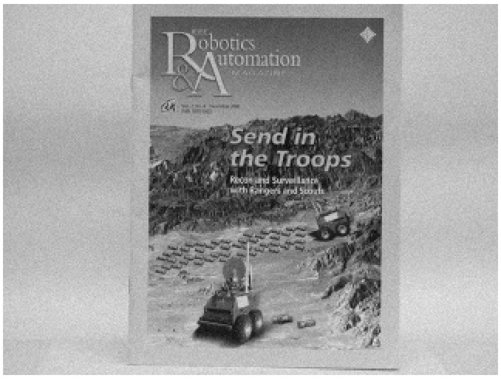
여기서 N_C 는 입력 영상에 있는 물체로부터 추출한 특징량과 이 물체에 해당하는 모델로부터 추출한 특징량과의 대응쌍 개수를 나타내고, N_D 는 다른 모델과의 대응쌍 개수를 나타낸다. 따라서, 변별인자(DF) 값이 높을수록 물체 변별력, 즉, 인식력이 높아진다고 볼 수 있다.

최적 차수를 정하기 위해서 Zernike 모멘트의 차수 5, 10, 15, 20, 25, 30에 대해서 동일 물체간, 다른 물체간 대응하는 수를 센다. 단, 실험 조건을 동일하게 하기 위해 잡음이 없는 입력 영상에 대해 동일 물체 사이의 대응점 수가 100개가 되게 유클리드 거리의 문턱값을 조절한다. 그런 다음, 잡음이 없는 환경, 가우시안 잡음이 있는 환경 하에서 각각 변별인자(DF)를 계산한다. 또한 각 차수별 계산 시간도 고려한다. 그림 (4.1)은 이러한 실험을 위한 영상을 나타낸다. 그림 (4.1)의 (a)는 잡음이 섞여 있지 않은 입력 영상을 나타내고, (b)는 표준편차 5의 가우시안 잡음이 첨가된 영상을 나타낸다. (c)는 입력 영상과 맞는 모델 영상을 나타내고 (d)는 입력 영상과 다른 모델을 나타낸다. 다음 그림 (4.2)는 Zernike 모멘트의 최적 차수를 정하기 위해 실험한 결과 그래프이다. 잡음이 없는 이상적인 경우만 볼 경우 차수가 높을수록 DF 값이 증가한다. 하지만, 이에 따른 처리 시간은 선형적으로 증가한다. 입력 영상에 표준 편차가 5인 가우시안 잡음이 첨가될 경우,

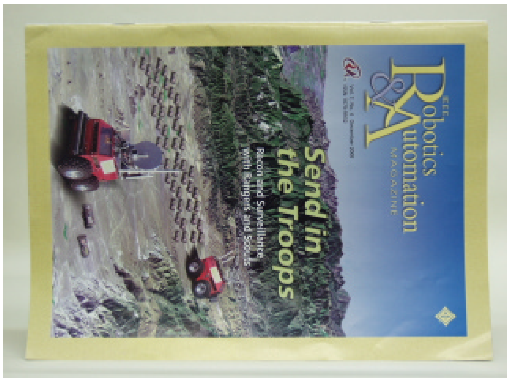
DF 값이 증가하다가 20차일 때 최고값을 갖고 다음 차수부터 감소하는 경향을 나타낸다. 따라서, 처리 시간, 잡음 등의 인자를 고려할 경우, Zernike 모멘트의 최적 차수는 20차라고 결정할 수 있다.



(a) 입력 영상(잡음 없음)



(b) 입력 영상(잡음 있음)



(c) 맞는 모델(물체 1)

(d) 다른 모델(물체 17)

그림 4.1 Zernike 모멘트 차수 실험용 영상들

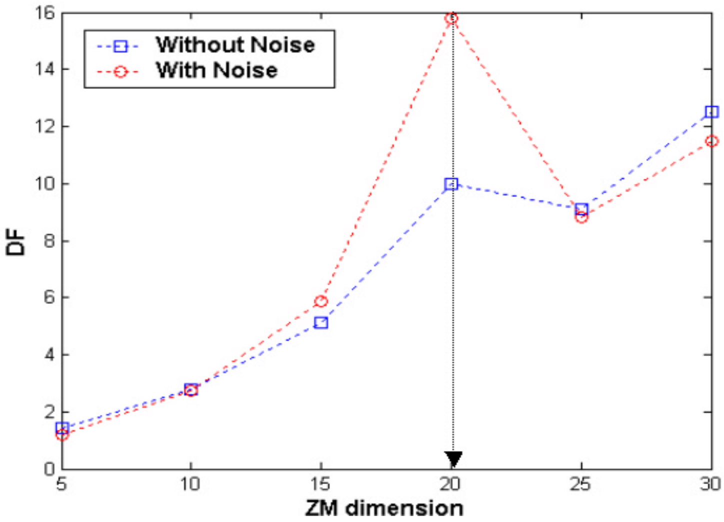


그림 4.2 Zernike 모멘트의 차수에 따른 DF 분포

표 4.1 Zernike 모멘트의 최적 차수을 구하기 위한 실험 결과

항목 \ 차수		5	10	15	20	25	30
문턱값		0.12	0.20	0.21	0.22	0.21	0.19
잡음 없음	입력-옳은 모델쌍	100	100	100	100	100	100
	입력-다른 모델쌍	69	36	20	10	11	8
DF		1.45	2.77	5.00	10.00	9.09	12.50
처리 시간[초]		2.04	3.67	5.37	7.46	9.63	11.86
잡음 있음	입력-옳은 모델쌍	89	93	94	95	88	92
	입력-다른 모델쌍	76	34	16	6	10	8
DF		1.17	2.74	5.86	15.80	8.80	11.50

표 (4.1)은 차수 실험을 위해 사용된 유클리드 거리의 문턱치 및 DF, 처리시간을 나타낸다.

4.2. 특징점 영역 크기 실험

제안한 인식 시스템은 지역 영상 정보를 사용하기 때문에 지역 영상의 크기를 정하는 것 또한 매우 중요하다. 단순히 작게만 또 크게만 할 수 없다. 본 실험에서는 앞 실험에서처럼 변별인자(DF)를 이용하여 지역 영상의 영역 크기를 정하기로 한다.

실험용 입력 영상은 그림 (4.1)의 (a), 맞는 모델 영상은 (c), 다른 모델 영상은 (d)로 한다. 지역 영상의 크기는 5×5, 11×11, 15×15,, 21×21, 25×25, 31×31 크기로 변화 시키며 변별인자(DF)를 계산한다. 실험 조건에서 Zernike 모멘트 차수는 20으로 고정 시키고, 입력 영상과 맞는 모델과의 대응쌍 수가 100개가 되도록 문턱값을 정한다 (표 4.2 참고). 그런 다음, 입력 영상과 다른 모델과의 대응쌍 수를 구해서 변별인자(DF)를 계산한다.

그림 (4.3)은 이러한 실험을 통해 얻은 결과 그래프이다. 결과를 보면, 변별인자(DF) 값이 지역 영상 영역의 크기가 증가함에 따라 감소하다가 (21×21) 크기에서 최고값을 나타내고 다음부터는 급격히 감소하는 것을 알 수 있다. 영역의 크기가 (21×21)보다 클 때 변별인자(DF) 값이 급격히 떨어지는 것은 지역 영역이 커짐에 따라 물체의 3차원 투영 효과가 커져 어파인(affine) 가정을 벗어나기 때문이다.

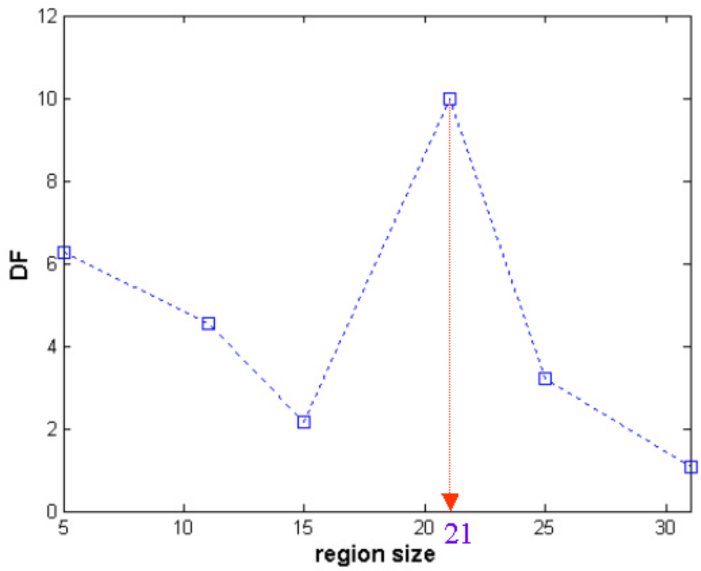


그림 4.3 Zernike 모멘트의 지역 영역 크기에 따른 DF 분포

표 4.2 지역 영상 영역의 최적 크기를 위한 실험 결과

윈도우크기 / 항 목	5×5	11×11	15×15	21×21	25×25	31×31
문턱값	0.016	0.06	0.23	0.22	0.26	0.34
입력-옳은 모델쌍	100	100	100	100	100	100
입력-다른 모델쌍	16	22	46	10	31	94
DF	6.25	4.54	2.17	10.00	3.22	1.06
처리 시간[초]	4.71	5.40	6.04	7.32	9.96	12.46

4.3. 단순 리사이징(resizing)과 스케일 스페이스

3장에서 특징점 추출기가 스케일에 불변 특성을 갖지 못하기 때문에 스케일 스페이스를 사용하여 보완한다고 하였다. 본 절에서는 실제로 스케일 스페이스가 변별인자(DF)에 어떤 영상을 주는지 간단히 실험해 본다.

실험 환경은 Zernike 모멘트의 차수가 20, 지역 영상 영역의 크기가 (21×21)로 하였다. 앞에서와 마찬가지로 입력 영상의 물체와 맞는 모델과의 대응쌍 수가 100이 되도록 각각 문턱값을 설정하고 두 가지 경우, 즉 단순 리사이징(resizing)하는 경우와 스케일 스페이스 이론을 이용하는 경우에 대해서 변별인자를 구하여 비교하였다.

표 4.3 단순 리사이징과 스케일 스페이스 방법 비교

방 법 / 항 목	단순 리사이징	스케일 스페이스
문턱값	0.23	0.22
입력-옳은 모델쌍	100	100
입력-다른 모델쌍	14	10
DF	7.14	10.00

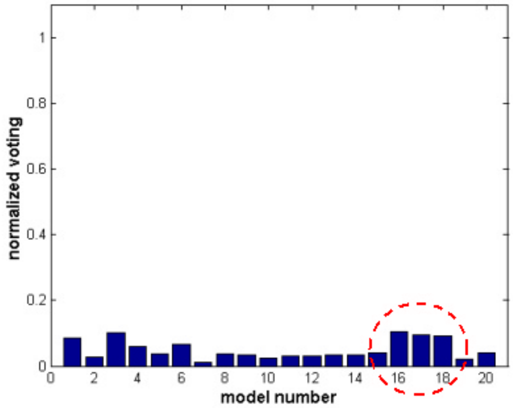
표 (4.3)에서 보면, 단순 리사이징(resizing)할 때보다 스케일 스페이스 이론을 도입할 경우 변별인자(DF) 값이 더 큰 것을 알 수 있다. 따라서 스케일 스페이스를 사용하는 것은 물체 인식률을 높이는 중요한 요소가 된다.

4.4. 단순 보팅(voting) 과 확률적 보팅 실험

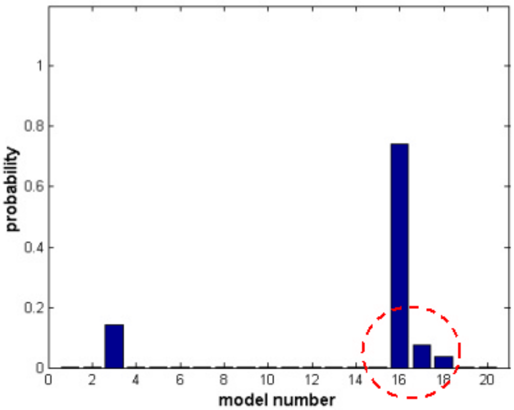
제안한 확률적 물체 인식이 얼마나 강력한지 보여주기 위해 일반적으로 사용하는 단순 보팅 방법과 비교해 본다. 단순 보팅은 3장에서 잠깐 언급하였듯이 각 입력 영상의

특징량에 대해서 가장 가까운 모델의 특징량을 찾고 문턱값을 만족하는 특징량의 모델 숫자를 센다. 모든 입력 영상의 특징량에 대해서 거리를 계산하여 보팅(voting)이 가장 많이 된 모델을 찾게 된다. 반면, 제안한 확률적 보팅 방법은 입력 영상의 특징량에 가장 가까운 후보 k 개를 선택하여 가설을 생성하고 각 가설에 대해 안정도, 유사도 개념을 반영하여 이에 따라 확률값을 계산한다. 각 입력 영상의 특징량마다 k 개의 가장 가까운 후보가 선택되므로 가설을 생성할 때 가우시안 분포를 갖도록 후보를 각각 선택한다. 이렇게 하였을 때 몇 가지 실험 예는 다음 그림 (4.4), (4.5)와 같다.

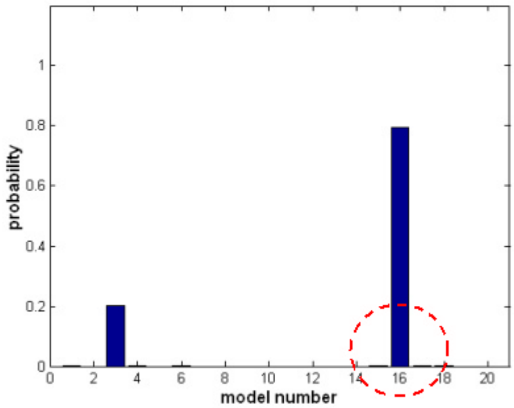
(a) 입력 영상: 모델 16



(b) 단순 보팅



(c)확률적 보팅(유사도)

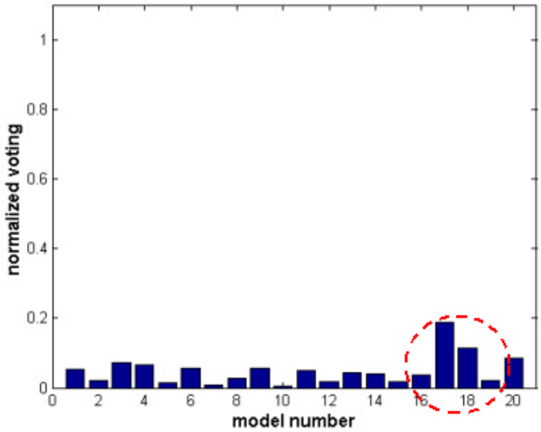


(d) 확률적 보팅(유사도+안정도)

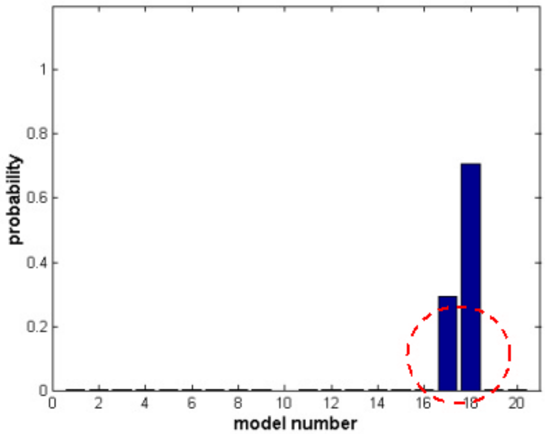
그림 4.4 단순 보팅과 확률적 물체 인식 비교 1

그림 (4.4)는 단순 보팅 방법과 확률적 물체 인식 기법이 모두 성공적으로 강아지 인형(16번 물체)을 인식하는 것을 보여준다. 그러나 단순 보팅의 방법은 찾고자 하는 물체의 보팅값이 다른 물체의 보팅값과 매우 유사하여 인식에 실패할 확률이 매우 높다. 유사도만을 이용한 확률적 보팅 기법을 적용한 결과 (c)는 매우 높은 확률로 물체 16을 인식한다. 유사도와 안정도를 동시에 고려하여 확률적 보팅을 할 경우, 그림 (4.4)의 (d)와 같이 매우 안정적으로 물체 인식을 한다.

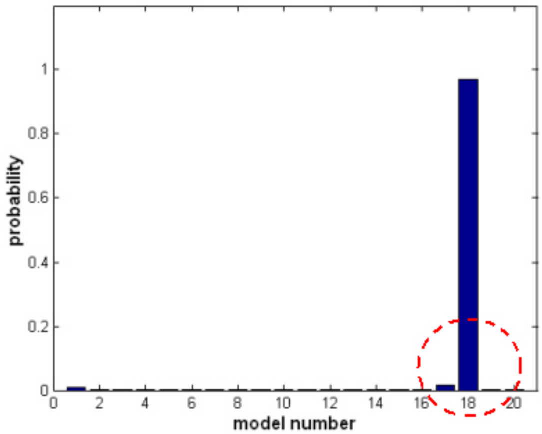
a) 입력 영상:모델 18



(b) 단순 보팅



(c) 확률적 보팅(유사도)



(d) 확률적 보팅(유사도+안정도)

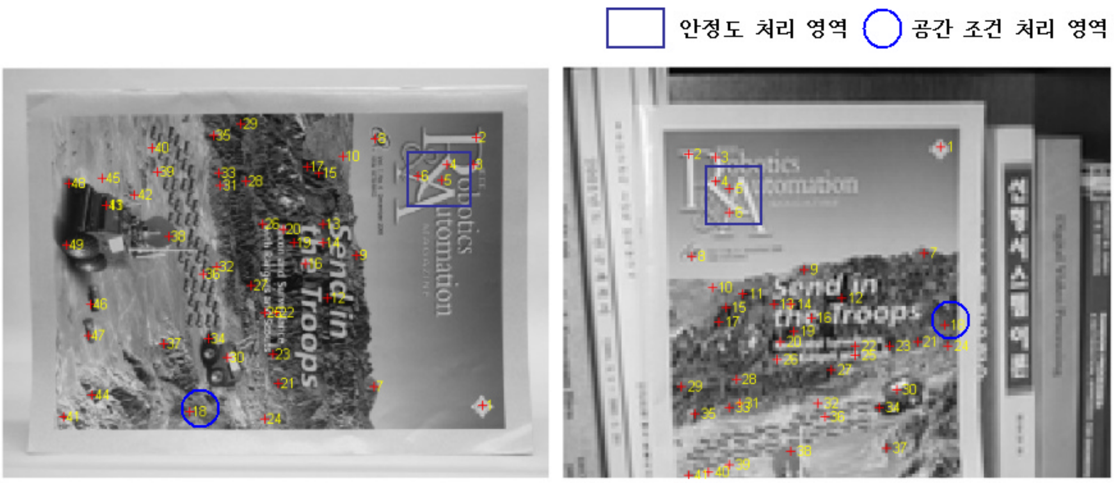
그림 4.5 단순 보팅과 확률적 물체 인식 비교 2

그림 (4.5)는 그림 (4.4)와 유사하다. 그러나 이 경우는 단순 보팅의 방법이 인식에 실패하지만, 제안한 확률적 물체 인식 기법을 적용할 경우에는 올바르게 입력 영상에 있는 박스 물체(18번 물체)를 인식하는 것을 보여준다. 그림 (4.5)의 (c)와 (d)를 통해 유사도와 안정도를 함께 고려하는 확률적 보팅 방법이 안정적으로 물체 인식한다는 것을 보여준다.

그림 (4.4), (4.5)로부터 제안한 확률적 물체 인식 기법이 단순 보팅의 문제점을 잘 해결한다는 것을 확인할 수 있다.

4.4.1. 대응쌍 필터링 유무 실험

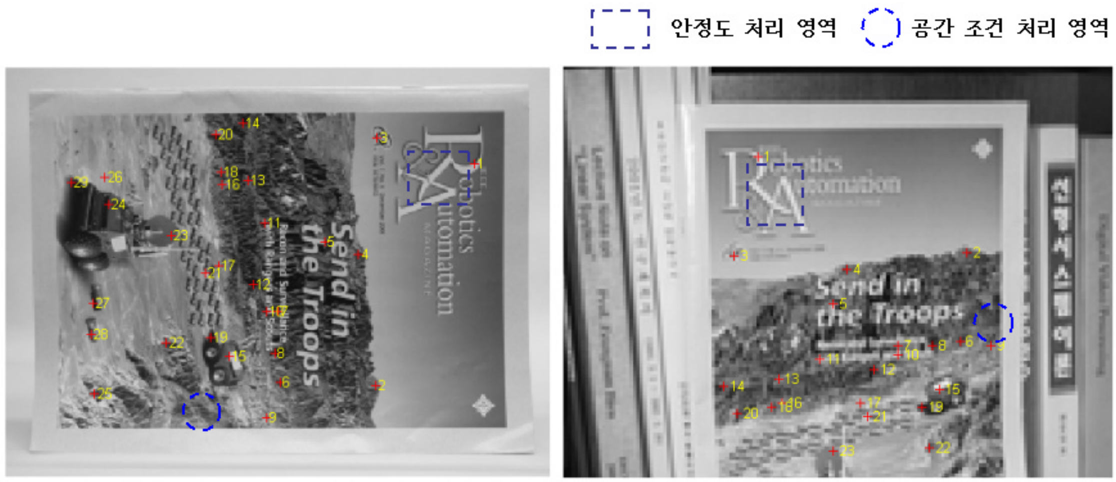
찾은 물체가 맞는 물체인지 검증(verification)을 하기 위해 대응쌍을 이용한 정렬(alignment) 과정이 필요하다. 입력 영상과 찾은 모델과의 거리가 가장 가까운 특징점을 대응쌍으로 했을 때 잘못 대응되는 경우(outlier)가 있다. 그러나 확률적 물체 인식을 할 때, 사용한 안정도 개념을 적용하고, 면적비를 이용한 공간 매칭 조건(local spatial constraint)을 적용하는 필터링 과정을 더 거칠 경우, 이상점(outlier)이 대부분 제거된다. 그림 (4.6)은 필터링 적용전의 모델과 확대한 입력 영상을 나타낸다.



a) 모델 영상　　　　b) 확대한 입력 영상

그림 4.6 필터링 적용 전

그림 (4.7)은 제안한 필터링의 기능을 잘 보여준다. 네모 박스는 안정도 필터링에 의해 제거되는 것을 나타낸다. 제거된 특징점은 특징점 오차에 대해 Zernike 모멘트 값이 크게 변하는 점이다. 또한 동그라미는 공간적으로 잘못 대응된 경우인데, 지역 공간 조건을 적용하여 필터링한 결과 잘 제거되는 것을 알 수 있다. 만약, 이렇게 필터링 과정에서 대응점이 3개 이하가 될 경우에는 제안한 필터링 기법을 적용하지 않는다. 제거 되지 않은 이상점(outlier)은 강인한 매칭 기법으로 어느 정도 해결되기 때문이다.

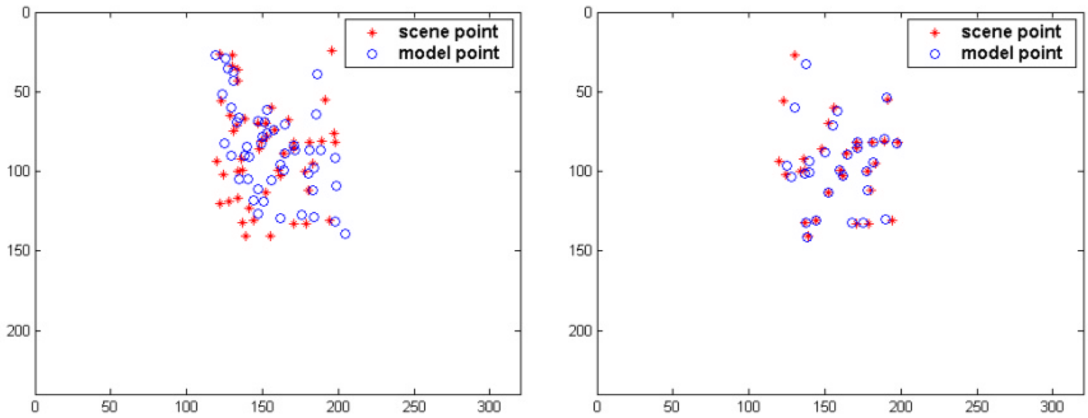


a) 모델 영상　　　b) 확대한 입력 영상

그림 4.7 제안한 필터링 적용 후

그림 (4.8)은 모델 특징점에 호모그래피를 적용한 것이다. (a)에서는 단순하게 거리 정보만을 이용하여 대응쌍을 찾고 LS+LMedS 기법을 이용하여 호모그래피(homography)를 찾았다. 잘못 정렬되는 경우가 10번에 1번 정도 발생한다. 이것은 강인한 호모그래피 계산 기법을 사용했기 때문이다. 그 중 (a)는 잘못 정렬된 결과를 보여주는데 이것은 대응쌍을 랜덤하게 추출 할 때 이상점(outlier)이 포함되었기 때문이다. 이렇게 잘못 정렬되는 문제를 해결하기 위해 본 논문에서 제안한 필터링 기법을 적용하였다. 초기 대응쌍에 그림 (4.8)의 (b)와 같이 필터링하고 이 대응쌍에 대해

LS+LMedS 기법으로 호모그래피를 구하여 대응쌍에 적용한 결과이다.



(a) 거리 조건만을 이용한 경우(outlier:30.6%)　　(b) 거리+안정도+지역공간조건을 이용한 경우(outlier:0%)

그림 4.8 필터링 유무에 따라 모델 특징점에 호모그래피를 적용한 결과

모델의 에지 영상에 대해서도 각각 호모그래피를 적용한 것이 그림 (4.9)이다. 필터링을 적용한 경우에 정렬이 잘 되는 것을 확인할 수 있다. 따라서 본 논문에서 제안한 대응쌍 필터링 과정이 매우 효과적으로 작동한다는 것을 확인할 수 있다.

(a) 거리 조건만을 이용한 경우(outlier:30.6%)　　(b) 거리+안정도+지역공간조건을 이용한 경우(outlier:0%)

그림 4.9 필터링 유무에 따라 모델 에지에 호모그래피를 적용한 결과

4.5. 다양한 환경에 대한 물체 인식 실험

앞 절에서는 제안한 인식기의 최적 셋팅을 위한 기본 실험을 하여 Zernike 모멘트의 차수, 지역 영역의 크기를 결정하였다. 또한 스케일 스페이스, 확률적 물체 인식, 대응점 필터링 방법이 인식 성능을 향상 시킨다는 것을 확인했다.

그림 4.10 실험에 사용된 물체 모델

본 절에서는 제안한 인식 시스템이 다양한 입력 영상에 대해서 올바로 인식하는지 그 강인성을 검증하기로 한다. 단, 본 실험에서는 앞 절 실험을 통해서 검증 받은 시스템 변수를 사용하기로 한다. 즉, Zernike 모멘트의 차수는 20, 지역 영상 영역의 크기는 21×21, 스케일 스페이스 사용, 확률적 물체 인식 기법 적용, 마지막으로 대응점 필터링 기법을 사용한다. 또한 실험을 위한 물체로는 주변에서 쉽게 구할 수 있는 20개를

선정하여 실험한다. 실험에 사용된 물체는 그림 (4.10)과 같다.

크기, 회전, 시각, 조명, 배경 잡음, 가려짐, 다중 물체 실험을 할 때, 인식 실험은 각 물체 당 1회로 하였다.

4.5.1. 크기(size) 실험

물체의 크기가 다양하게 변할 때(1~2.5배 축소) 각 물체의 인식 성공 여부를 확인한다. 성공 여부는 두 단계로 확인한다. 첫번째 단계는 확률적 물체 인식의 결과가 맞게 나오는지, 두 번째 단계는 물체 검증에서 모델이 입력 물체에 50%이하의 이상점(outlier)으로 정렬하는지 검사한다.

(a) scale=1/1.5 (b)scale=1/2 (c) scale=1/2.5

그림 4.11 크기 실험용 샘플 영상들

그림 (4.11)은 스케일 실험을 위한 샘플 영상을 나타낸다. 입력 영상에서 변하는 것은 물체의 크기만으로 한정한다.

그림 (4.12)는 샘플 영상에 대한 인식 결과 얻은 물체의 위치와 포즈를 나타낸 것이다. 결과에서 보듯이 다양한 크기의 물체에 대해서 정확하게 인식하는 것을 알 수 있다. 표 (4.4)는 20개의 물체에 대해 3가지 다른 스케일에 대해서 인식 능력을 검사해본 결과이다. 실험은 인식 부분과 검증 부분으로 두 가지 경우에 대해 살펴 보았다. 표 (4.5)는 표 (4.4)를 세부화한 것으로 입력 물체를 제대로 인식하는지 여부와 모델을 입력 영상에 있는 물체에 제대로 정렬(align)하는지 % 이상점(outlier)으로 확인한 세부 실험 결과이다.

이상점(outlier)이 50%를 넘는 경우는 정렬 실패로 간주하였다. 크기 실험 결과 1.5배, 2배, 2.5배 축소된 물체 20개에 대해서 두 가지 인식 실험이 모두 성공하였다.

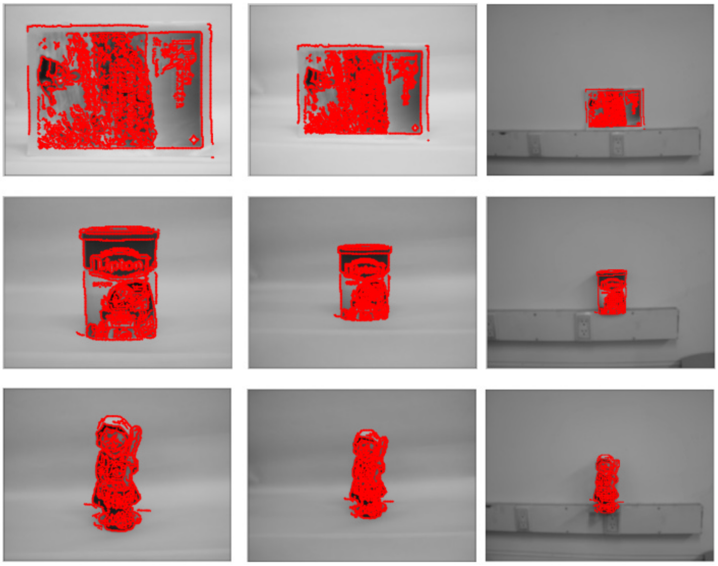


그림 4.12 크기 변화된 영상에 대한 물체 인식 결과 영상

표 4.4 크기에 따른 인식 결과

실험 \ 스케일	1/1.5	1/2	1/2.5
인 식[성공/횟수]	20/20	20/20	20/20
검 증[성공/횟수]	20/20	20/20	20/20

표 4.5 크기에 따른 각 물체의 세부 인식 실험

[S1: 인식, S2: %이상점, S: success, F: fail]

모델 #	Scale 1.5		Scale 2		Scale 2.5	
	S1	S2	S1	S2	S1	S2
1	S	2	S	0	S	1
2	S	0	S	0	S	0
3	S	0	S	0	S	0
4	S	0	S	0	S	0
5	S	0	S	0	S	0
6	S	0	S	0	S	0
7	S	0	S	0	S	0
8	S	0	S	0	S	0
9	S	0	S	0	S	0
10	S	0	S	0	S	0
11	S	0	S	2	S	0
12	S	3	S	0	S	0
13	S	0	S	0	S	0
14	S	2	S	0	S	5
15	S	3	S	0	S	0
16	S	2	S	2	S	0
17	S	2	S	2	S	0
18	S	0	S	2	S	2
19	S	0	S	0	S	0
20	S	0	S	0	S	0

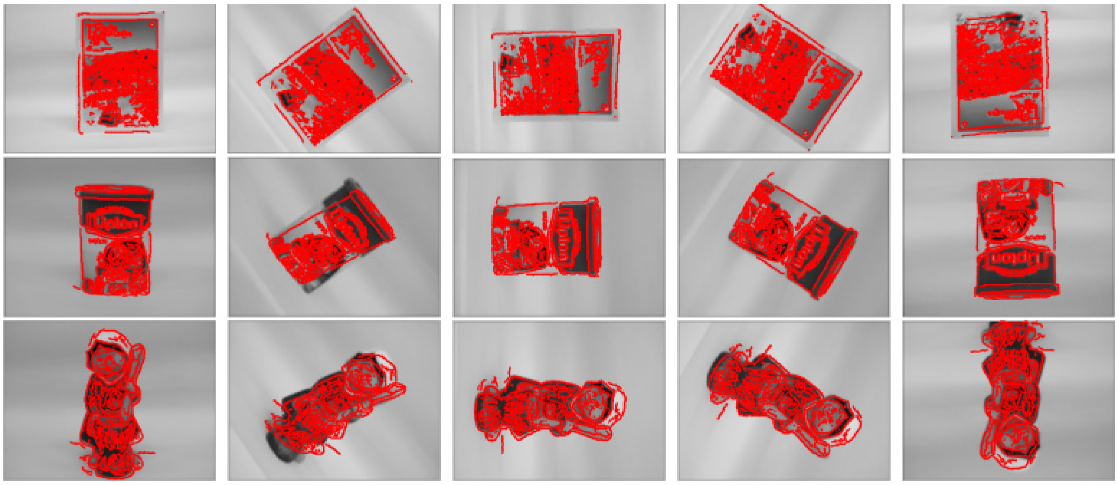
4.5.2. 회전(planar rotation) 실험

물체가 평면 회전만 할 때 각 물체의 인식 능력을 검사한다. 물체 크기, 보는 각도, 조명 등 다른 조건은 모두 동일하게 하였다. 입력 영상은 0°, 45°, 90°, 135°, 180° 회전한 실제 물체에 대해서 인식 검사를 하였다. 그림 (4.13)은 몇 가지 물체에 대한 실제 영상을 나타낸다.

(a) 0° 회전 (b) 45° 회전 (c) 90° 회전 (d) 135° 회전 (e) 180° 회전

그림 4.13 평면 회전 실험용 샘플 영상들

그림 (4.14)는 샘플 물체에 대한 인식 결과를 각각 나타낸다. 다양한 평면 회전에 대해서 각 물체의 위치와 포즈를 잘 찾는다. 표 (4.6)는 물체가 평면 회전 할 때, 각 물체에 대한 성공 횟수를 나타낸다. 표 (4.7)은 표 (4.6)의 세부 실험 결과로 인식 능력을 두 단계로 살펴 본 것이다. S1은 물체의 확률적 인식 성공 여부를 살펴보고, S2는 찾은 모델이 입력 영상에 있는 물체에 잘 정렬하는지 % 이상점(outlier)으로 확인한다.



(a) 0° 회전　(b) 45° 회전　(c) 90° 회전　(d) 135° 회전　(e) 180° 회전

그림 4.14 평면 회전된 물체에 대한 인식 결과 영상

표 4.6 회전에 따른 인식 결과

평면회전 / 실험	0°	45°	90°	135°	180°
인 식[성공/횟수]	20/20	20/20	20/20	19/20	20/20
검 증[성공/횟수]	20/20	19/20	20/20	18/20	19/20

표 4.7 회전에 따른 각 물체의 세부 인식 실험

[S1: 인식, S2: %이상점, S: success, F: fail]

모델	0°		45°		90°		135°		180°	
	S1	S2	S1	S2	S1	S2	S1	S2	S1	S2
1	S	4	S	11	S	0	S	0	S	0
2	S	0	S	40	S	8	S	0	S	0
3	S	3	S	0	S	8	S	0	S	4
4	S	0	S	0	S	21	S	20	S	4
5	S	0	S	0	S	0	S	0	S	0
6	S	0	S	0	S	26	S	0	S	0
7	S	0	S	F	S	0	S	25	S	0
8	S	2	S	12	S	27	S	13	S	2
9	S	17	S	44	S	11	S	23	S	28
10	S	0	S	25	S	40	S	F	S	F ①
11	S	2	S	0	S	4	S	0	S	8
12	S	0	S	25	S	2	S	40	S	37
13	S	0	S	13	S	8	S	17	S	8
14	S	8	S	0	S	8	S	F	S	7
15	S	0	S	40	S	20	F	F ②	S	0
16	S	24	S	43	S	42	S	F	S	40
17	S	25	S	40	S	4	S	0	S	33
18	S	5	S	15	S	4	S	45	S	5
19	S	0	S	40	S	7	S	0	S	0
20	S	2	S	0	S	9	S	20	S	9

결과를 보면 모든 물체가 5가지 회전에 대해서 15번 물체(②)를 제외하고 성공적으로 물체를 인식하는 것을 알 수 있다. 15번 물체를 인식하고 정렬하는데 모두 실패를 하였는데 이것은 그림 (4.15)의 (a)에서 보는 것처럼 물체 표면에 스페큘라(specular) 성분이 많아 각도에 따라 외관이 많이 바뀌어서 다른 물체로 인식하였기 때문이다. 10번 물체(①)가 135도, 180도 회전한 경우에 올바로 정렬하지 못하였다. 물체 인식은 제대로 하였는데 정렬을 제대에 실패한 것은 10번 물체가 대칭성을 갖기 때문이다. 특징량 값은 유사하지만, 대응점이 다른 곳에 분포하여 정렬이 제대로 되지 못하였다. 또한 45도 회전한 물체에 대해서 이상치(outlier)가 다소 높은데 이것은 Harris 코너 추출기가 45도에서 반복성이 85%로 떨어지는데 원인이 있다.

(a) 135도 회전한 15번 물체의 입력 영상

(b) 135도 회전한 10번 물체의 입력 영상

그림 4.15 인식/정렬에 실패한 입력 영상들

4.5.3. 시각(view angle) 실험

제안한 인식 시스템의 모델은 물체의 정면을 기본으로 한다. 3차원 물체를 모든 방향에서 인식하기 위한 모델 영상을 생성하기 위해서는 제안한 인식 시스템이 물체 정면을 모델로 갖고 있을 때, 입력으로 들어오는 물체의 인식 가능한 시각(view angle) 범위를 알아야 한다. 가령, 인식 가능한 시각(view angle 범위가 정면을 기준으로 -30°~+30° 라면

3차원 물체에 대해서 60° 각도 간격으로 본 6장의 영상만 있으면 된다. 이번 실험은 정면에 대해 인식 가능한 시각(view angle)을 알아보는 것이다. 그림 (4.16)은 시각 실험용 샘플 영상을 나타낸다. 실험 영상은 정면을 기준으로 10도 간격씩 오른쪽으로 회전하여 촬영하였다. 반대 방향은 대칭성을 고려해 생략하였다.

(a) 0° view (b) 10° view (c) 20° view (d) 30° view

그림 4.16 시각(view angle) 실험용 샘플 영상

그림 (4.17)는 물체를 바라보는 각도에 따라 인식한 결과를 보여준다. 물체가 많이 회전할 경우 정렬이 좀 어긋나는 것을 알 수 있다. 표 (4.8)은 정면 뷰에서 0°, 10°, 20°, 30° 씩 시점(view angle)이 바뀌었을 때 인식 결과를 나타낸다. 물체 인식은 모든 물체에 대해서 성공적이다. 표 (4.9)는 각 물체별, 각도별 인식 결과를 나타낸다. 정렬에 실패한 것은 회전각에 따라 스페큘러 때문에 매칭이 안되었기 때문이다.

(a) 0° view (b) 10° view (c) 20° view (d) 30° view

그림 4.17 시각(view angle)에 따른 물체 인식 결과 영상

표 4.8 시각에 따른 인식 결과

실험 \ 시각	0°	10°	20°	30°
인 식[성공/횟수]	20/20	20/20	20/20	20/20
검 증[성공/횟수]	20/20	18/20	19/20	18/20

표 4.9 시각(view angle)에 따른 각 물체의 세부 인식 실험

[S1: 인식, S2: %이상점, S: success, F: fail]

모델	0°		10°		20°		30°	
	S1	S2	S1	S2	S1	S2	S1	S2
1	S	3	S	5	S	1	S	4
2	S	0	S	0	S	20	S	0
3	S	0	S	2	S	0	S	30
4	S	11	S	33	S	20	S	0
5	S	0	S	5	S	0	S	0
6	S	4	S	5	S	8	S	20
7	S	0	S	0	S	0	S	0
8	S	37	S	2	S	5	S	21
9	S	19	S	F	S	40	S	F
10	S	0	S	33	S	40	S	40
11	S	4	S	4	S	0	S	4
12	S	3	S	17	S	11	S	17
13	S	6	S	10	S	20	S	32
14	S	7	S	10	S	4	S	23
15	S	33	S	F	S	F	S	F
16	S	12	S	12	S	40	S	25
17	S	37	S	39	S	19	S	5
18	S	4	S	5	S	5	S	16
19	S	0	S	0	S	25	S	0
20	S	1	S	0	S	3	S	4

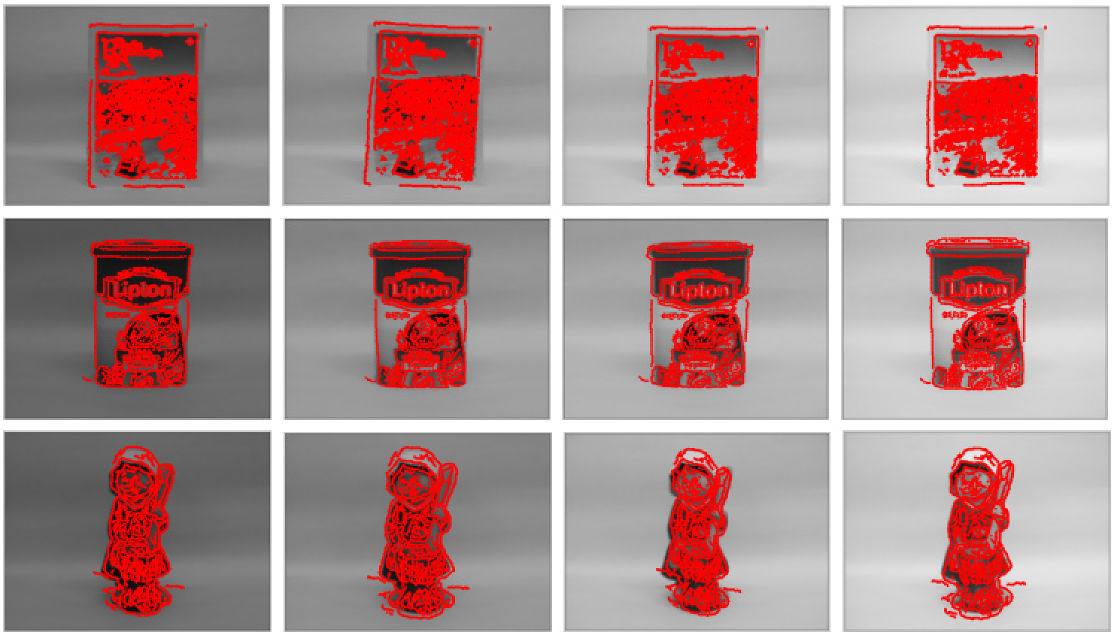
4.5.4. 조명(illumination) 실험

물체 인식기가 조명에 무관하게 물체 인식을 할 수 있는 능력 또한 중요하다. 본 실험에서는 그레이(gray) 영상을 사용하기 때문에 조명의 밝기만 고려한다. 그림 (4.18)은 영상의 평균 밝기가 100, 130, 160, 190인 경우의 샘플 영상들을 보여준다.

(a) 평균 밝기: 100 (b) 평균 밝기: 130 (c) 평균 밝기: 160 (d) 평균 밝기: 190

그림 4.18 조명 실험용 샘플 영상

그림 (4.19)는 다양한 조명 하에서 물체의 인식 결과를 보여준다. 물체의 평균 밝기가 100부터 190까지 30씩 변할 때 각 물체를 매우 정확하게 인식하는 것을 보여준다. 표 (4.10)은 네 가지 조명에 대해서 각 물체에 대한 인식 실험을 한 결과이다. 물체 인식 실험에서 모든 물체가 성공적이다. 표 (4.11)은 세부 실험 결과이다. 정렬 실험에서 15번 물체만 실패하였다. 실험 결과를 종합하면 상당히 넓은 조명 밝기에 대해서 제안한 물체 인식 시스템이 매우 강인하게 동작하는 것을 알 수 있다.



(a) 평균 밝기: 100 (b) 평균 밝기: 130 (c) 평균 밝기: 160 (d) 평균 밝기: 190

그림 4.19 변하는 조명 하에서 물체 인식 결과 영상

표 4.10 조명에 따른 인식 결과

평균밝기 / 실험	100	130	160	190
인 식[성공/횟수]	20/20	20/20	20/20	20/20
검 증[성공/횟수]	19/20	20/20	20/20	20/20

표 4. 11 조명에 따른 각 물체의 세부 인식 실험

[S1: 인식, S2: %이상점, S: success, F: fail]

모델	100		130		160		190	
	S1	S2	S1	S2	S1	S2	S1	S2
1	S	3	S	20	S	4	S	16
2	S	6	S	0	S	13	S	15
3	S	14	S	0	S	2	S	2
4	S	0	S	18	S	14	S	9
5	S	0	S	2	S	2	S	2
6	S	0	S	0	S	3	S	2
7	S	0	S	0	S	0	S	0
8	S	0	S	3	S	2	S	2
9	S	20	S	5	S	8	S	13
10	S	0	S	43	S	0	S	35
11	S	0	S	9	S	6	S	2
12	S	0	S	0	S	8	S	8
13	S	0	S	0	S	0	S	0
14	S	14	S	16	S	23	S	15
15	S	F	S	14	S	0	S	17
16	S	42	S	31	S	6	S	4
17	S	7	S	30	S	6	S	19
18	S	4	S	4	S	3	S	1
19	S	0	S	0	S	0	S	0
20	S	2	S	2	S	0	S	4

4.5.5. 배경 잡음(background clutter) 실험

일반적으로 물체가 잘 분할되어 있는 경우가 드물다. 이번 실험은 물체 뒤에 다양한 잡음(background c lutter)이 있을 때 제안한 인식기의 특성을 확인하도록 한다. 그림 (4.20)은 단순한 배경 잡음과 복잡한 배경 잡음에 대한 샘플 영상을 나타낸다.

(a) 단순한 배경

(b) 복잡한 배경

그림 4.20 배경 잡음 실험용 샘플 영상

그림 (4.21)은 배경 잡음이 있을 때 물체를 인식한 결과를 나타낸다. 배경이 단순하든 복잡하든 물체의 위치와 포즈를 잘 인식하는 것을 알 수 있다.

(a) 단순한 배경

(b) 복잡한 배경

그림 4.21 배경 잡음이 있을 때 물체 인식 결과 영상

표 (4.12)는 각 물체에 대해서 단순한 배경일 때와 복잡한 배경일 때, 각각 실험한 결과를 나타낸다. 표 (4.13)은 세부 인식 실험으로 단순한 배경 및 복잡한 배경에서 물체 인식이 모두 성공적이었다. 그러나 복잡한 배경에서는 물체 10을 인식하고 정렬하는데 실패했다. 원인은 물체의 특징점보다 배경의 특징점이 더 많았기 때문이다.

표 4.12 배경 잡음에 대한 인식 결과

배경 / 실험	단순한 배경	복잡한 배경
인 식[성공/횟수]	20/20	19/20
검 증[성공/횟수]	20/20	19/20

표 4.13 배경 잡음이 있을 때 각 물체에 대한 세부 인식 실험
[S1: 인식, S2: %이상점, S: success, F: fail]

모 델	단순한 배경		복잡한 배경	
	S1	S2	S1	S2
1	S	0	S	3
2	S	0	S	0
3	S	0	S	0
4	S	10	S	0
5	S	2	S	0
6	S	2	S	3
7	S	14	S	17
8	S	0	S	4
9	S	0	S	40
10	S	31	F	F
11	S	8	S	0
12	S	17	S	17
13	S	0	S	17
14	S	0	S	32
15	S	38	S	25
16	S	44	S	0
17	S	7	S	6
18	S	4	S	1
19	S	0	S	0
20	S	0	S	0

4.5.6. 가려짐(occlusion) 실험

물체가 다른 물체에 의해 가려질 수 있다. 이렇게 가려짐(occlusion)이 있는 경우에 제안한 시스템이 얼마나 강인하게 동작하는지 실험해본다. 실험 영상은 그림 (4.22)와 같이 가려짐 실험을 하기 위해 25%, 50% 가려짐을 가정하였다.

그림 4.22 가려짐 실험용 샘플 영상

그림 (4.23)은 25%, 50% 가려짐이 있는 영상에 대해 제안한 인식 시스템을 적용 시킨 샘플 결과들이다. 반 정도가 가려졌는데도 물체의 위치와 자세를 잘 찾는 것을 알 수 있다.

그림 4.23 가려짐 영상에 대한 인식 결과 영상

표 (4.14)은 물체가 25%, 50%정도 다른 물체에 의해서 가려졌을 때 각 물체에 대한 인식 실험 결과이다. 표 (4.15)는 세부 인식 결과로 20개의 가려짐이 있는 입력 영상에 대해서 물체 인식 및 모델 정렬 모두 성공하는 것을 알 수 있다.

표 4.14 가려짐에 대한 인식 결과

실험 \ 가려짐	25% 가려짐	50% 가려짐
인 식[성공/횟수]	20/20	20/20
검 증[성공/횟수]	20/20	20/20

표 4.15 가려짐(occlusion)이 있을 때 각 물체에 대한 세부 인식 실험

[S1: 인식, S2: %이상점, S: success, F: fail]

모델	25% 가려짐		50% 가려짐	
	S1	S2	S1	S2
1	S	0	S	3
2	S	0	S	0
3	S	0	S	5
4	S	9	S	6
5	S	0	S	11
6	S	0	S	0
7	S	0	S	3
8	S	0	S	3
9	S	0	S	9
10	S	8	S	0
11	S	0	S	31
12	S	0	S	29
13	S	0	S	0
14	S	2	S	7
15	S	40	S	45
16	S	8	S	14
17	S	0	S	11
18	S	0	S	3
19	S	0	S	0
20	S	7	S	0

4.5.7. 다중 물체 실험

그림 (4.24)는 실험용 다중 물체를 나타내고, 그림 (4.25)는 이들 입력 영상에 대한 인식 결과를 나타낸다. 입력의 다중 물체가 합성일 때, 2개로 구성됐을 때, 3개로 구성됐을 때 모두 성공적으로 인식하는 것을 보여준다.

그림 4.24 다중 물체 인식 실험용 입력 영상

그림 4.25 다중 물체 인식 결과

4.6. 일반 환경에서의 실험

4.1절, 4.2절, 4.3절에서는 제안한 물체 인식 시스템의 내부 특성을 파악하고자 변별인자(discriminative factor) 개념을 도입하여 최적의 실험 파라미터를 찾았다. 4.5절에서는 다양한 실험 즉, 물체가 크기, 평면 회전, 시각, 조명, 배경잡음, 가려짐, 다중 물체 등에 의해 변화가 독립적으로 있을 때 제안한 인식기가 얼마나 강인하게 동작하는지 살펴보았다. 본 절에서는 이러한 실험을 바탕으로 물체가 다양한 환경에 임의로 놓여 있을 때 인식 실험을 수행한다.

실험은 세 경우에 대해서 수행하였다. 첫째는 시점 변화(크기, 회전, 시각)와 배경 잡음이 동시에 있는 경우이고 둘째는 첫째 실험 환경에 조명 변화와 가려짐 현상이 복합적으로 있는 경우이고, 셋째는 수 많은 물체 영상에 대해 통계적 인식 특성을 살펴보는 경우이다.

4.6.1. 일반 환경 실험 I

첫째 실험은 물체가 배경 잡음이 있는 곳에 임의의 자세로 있는 경우에 대해서 인식 실험을 하였다. 조명은 모델을 촬영할 때와 거의 같은 밝기로 하고 가려짐 현상은 전혀 없게 하였다.

그림 (4.26) 은 시점 변화 즉, 물체가 크기, 회전, 시각(view angle) 변화가 있고 동시에 배경 잡음이 있는 실험용 입력 영상을 나타낸다. 물체가 임의의 배경에 놓이기 때문에 배경잡음이 존재하고, 임의 위치에 놓여있기 때문에 물체에 크기, 회전, 시각 변화가 동시에 생기게 된다.

그림 4.26 시점+배경 잡음이 있는 입력 영상

그림 (4.27)은 시점과 배경 잡음이 있을 때 최종적으로 물체 인식한 결과를 나타낸다.

그림 4.27 시점+배경 잡음이 있을 때 물체 인식 결과 영상

임의로 놓여 있는 각 물체에 대해 물체를 인식하고 위치와 자세까지 잘 인식한다. 다른 물체에 대해서도 물체 인식 20/20, 물체 검증 20/20의 성공률을 보였다.

4.6.2. 일반 환경 실험 II

일반 환경 실험 II는 물체에 시각 변화, 배경 잡음, 조명 변화, 가려짐 현상이 복합적으로 있는 가장 일반적인 환경에서의 실험이다. 그림 (4.28)은 이러한 환경에서 촬영한 샘플 입력 영상을 나타낸다. 물체가 임의의 배경에 임의의 자세로 있고 조명이 어둡고 가려짐 현상이 동시에 나타나 있다.

그림 4.28 시점+배경 잡음+조명+가려짐이 있는 입력 영상

그림 4.29 시점+배경 잡음+조명+가려짐이 있을 때 물체 인식 결과 영상

그림 (4.29)는 물체에 크기, 회전, 시각 변화가 있고, 배경이 복잡한 곳에서 부분 가려짐이 있는 어두운 곳에서의 물체 인식 결과를 나타낸다. 열악한 환경에서 모든 물체에 대해서 실험한 결과 15/20의 인식 성공률을 보였고, 14/20의 정렬 성공률을 보였다. 나머지 6개를 인식하는데 실패한 것은 여러 가지 환경이 복합적으로 나타나 특징점 추출 및 대응이 불완전 해졌기 때문이다. 참고로 모델 DB에 없는 물체를 입력으로 할 경우, 90%이상 제거하는 것을 확인하였다. 이것은 모델 DB에 없는 물체가 들어올 경우, 대응쌍이 필터링 과정에서 대부분 제거되기 때문이다.

4.6.3. 일반 환경 실험 III

표 (4.16)은 물체가 임의의 환경에서 임의 자세로 있을 때, 각 샘플 물체에 대해서 100회 내외로 반복 실험한 결과이다. 6번 물체는 78.81%의 높은 인식률을 보였지만 나머지 8, 11번 물체는 50% 내외의 인식률을 보였다.

표 4.16 통계적 실험 결과

Model #	Iteration	Success	Recognition rate[%]
Model 6	118	93	78.81
Model 8	125	55	44.00
Model 11	95	53	58.24

8번 물체와 과 11번 물체의 인식률이 낮은 것은 그림 (4.30)과 같이 촬영한 물체 영상의 1/3정도가 너무 어둡거나 배경이 물체보다 훨씬 복잡한 경우이거나 카메라 초점이 맞지 않은 경우, 스펙큘러 반사가 심한 영상들이었기 때문이다.

(a) 너무 어두운 경우 (b) 배경잡음이 심한 경우 (c) 초점이 틀린 경우 (d) 스페귤러가 있는 경우

그림 4.30 물체 인식에 실패한 경우의 입력 영상들

5. 결론 및 향후 과제

본 논문에서는 일반 환경에 있는 복잡한 물체를 인식하기 위한 새로운 인식 패러다임을 제안하였다. 기존 물체 인식에 대한 많은 자료 조사를 통해 물체 인식은 특징량 매칭(feature matching)으로 귀결된다고 결론 내리고, 이를 바탕으로 새로운 형태의 물체 인식 시스템을 제안하였다. 제안한 시스템은 크게 두 부분으로 구성된다. 첫번째 부분은 특징량을 추출하는 부분이고 두 번째 부분은 특징량을 매칭하는 부분이다. 첫번째 부분에서 특징량을 추출할 때 다음과 같은 문제에 대해서 해결하고자 하였다. 일반 환경에 있는 물체는 보는 위치에 따라서 물체의 모양이 바뀌게 되고, 조명이 변함에 따라서 물체의 외관도 달라진다. 또한 물체 뒤에 여러 가지 배경 잡음(background clutter)도 있을 수 있고 다른 물체에 의해 가려짐(occlusion) 현상이 발생하기도 한다. 먼저, 물체를 보는 위치에 따라서 물체의 모양이 바뀌는 문제는 어파인 변환에 대해서 불변 특성을 갖는 Zernike 모멘트를 사용하였다. 단, Zernike 모멘트가 스케일에는 불변 특성을 갖지 못하기 때문에 스케일 스페이스 이론을 도입하여 물체의 스케일 문제를 해결하였다. 또한 조명이 변하는 문제는 Zernike 모멘트의 (0,0) 모멘트를 이용하여 해결하였다. 조명을 스케일 변환으로 모델링할 경우, 영상을 평균 밝기로 나눠주게 되면 조명에 불변인 영상이 나오게 된다. 마찬가지로 (0,0) 모멘트는 영상의 평균 밝기에 상응하므로, 원래 Zernike 모멘트를 (0,0) Zernike 모멘트로 나누게 되면, 스케일 조명에 불변인 특징량이 나오게 된다. 또한, 물체의 배경 잡음, 가려짐 현상은 영상의 전역 특성보다는 지역 특성에 착안하여 이러한 문제를 해결하였다. 결국, 물체 인식을 위한 특징량으로 지역 영상의 Zernike 모멘트로 하였다. 한편, 제안한 시스템의 두 번째 부분은 인식 즉, 특징량 매칭 부분이다. 특징량이 고차원의 벡터로 표현되기 때문에 최적의 대응쌍을 얻기 위해 기존의 단순 보팅(voting) 방법과 이 방법의 단점을 해결하기 위한 확률적 매칭 방법론을 제안하였다. 후보 모델을 선정할 때 단순 보팅 방법이 물체를 인식할 때 원하지 않는 물체를 찾는 경우가 있다. 이를 해결하기 위해 Zernike 모멘트의 안정도, 특징량 사이의 유사도를 확률적 물체 인식 기법에 반영하는 방

법을 제안하여 그 타당성을 검증하였다. 또한 모델과 입력 영상의 대응쌍을 구할 때, 단순하게 거리 정보만 쓸 경우, 정렬이 부정확 해지기 때문에 본 논문에서는 안정도, 유사도, 그리고 네 점을 이용하는 지역 공간 매칭 기법을 적용하여 대응쌍을 필터링하였다. 마지막으로 물체 인식이 제대로 되었는지 검증하기 위해 모델 특징량과 입력 물체의 특징량 간의 대응 관계를 나타내주는 호모그래피를 사용하였다. 호모그래피를 구하기 위해 기존의 방법들을 결합시켰다. 이상점(outlier)에 약한 LS(least square) 방법과 50% 이상점에 강한 LMedS(least median of squares) 방법을 결합 시켰다. 랜덤하게 대응쌍을 선택하여 LS 방법으로 호모그래피를 구하여 이상점(outlier)이 최소가 되게 하는 호모그래피를 선택한 후, 이 값을 초기값으로 하여 LMedS 방법으로 최적의 호모그래피를 구하였다.

20개의 복잡한 일반 물체에 대해 크기 실험, 평면 회전 실험, 시각(view angle) 실험, 조명 실험, 배경 잡음 실험, 가려짐 실험, 다중 물체 인식 실험 마지막으로 일반적인 환경에서의 실험을 통해 제안한 인식 시스템의 성능을 검증하였다.

제안한 시스템이 지역 특징점을 중심으로 한 모델 영상 기반의 인식 기법이기 때문에 몇 가지 해결할 과제가 있다.

첫째, 제일 큰 문제는 시각(view angle) 문제이다. 시각 실험에서 -30° ~ +30° 회전한 물체에 대해서 최소 80%의 인식률을 보인다는 점이다. 완전 3차원 물체를 해결하기 위해서는 최소 6 방향에 대해서 물체의 모델을 만들 필요가 있다는 점이다. 따라서 모델을 위한 저장 공간이 늘어나는 문제가 있다. 이런 문제를 해결하기 위해 좀 더 넓은 시각에서 불변 특성을 갖는 특징량 개발이 필요하다.

둘째, 물체가 평면 회전할 때 모델 영상에 대해 45° 회전할 경우, 다른 각도에 비해 인식률이 10%정도 떨어지는 문제가 있다. Harris 코너 검출기가 가로, 세로 두 방향으로만 코너를 검출하기 때문에 대각선 방향에 대해 반복성(repeatability)이 떨어진다. 따라서 전 방향에 대해서 불변 특성을 갖는 특징점 검출기 개발이 필요하다.

셋째, 본 논문에서는 물체의 밝기 분포만을 이용하여 확률적 물체 인식을 수행하였지만 앞으로 칼라 정보나, 모양 정보, 혹은 다른 센서(가령, 적외선) 정보와의 융합을 통해 물체 인식 시스템의 강인성을 높일 필요가 있다.

6. 참고 문헌

[1] S. Ullman, “High-level Vision: Object Recognition and Visual Cognition”, The MIT Press, 1997.

[2] C. Francis Olson, “Fast Object Recognition by Selectively Examining Hypothesis”, Doctor’s Thesis, University of California, 1994.

[3] A. R. Pope, “Learning to Recognize Objects in Images: Acquiring and Using Probabilistic Models of Appearance”, Doctor’s Thesis, The University of British Columbia, 1995.

[4] K. S. Roh, I. S. Kweon, "2-D Occluded and Curved Object Recognition using Invariant Descriptor and Projective Refinement”, IEICE Transactions on Information and Systems, Vol. E81-D, No. 5, 1998

[5] M. R. Stevens, J. R. Beveridge, “Precise Matching of 3-D Target Models to Multisensor Data”, CSU, Jan. 20, 1997.

[6] D. Keren, M. Osadchy, C. Gotsman, “Antifaces: A Novel, Fast Method for Image Detection”, IEEE Tras. On Pattern Recognition and Machine Intelligence, VOL. 23, NO.7, pp.747-761, 2001.

[7] A. R. Pope D. G. Lowe, “Probabilistic models of appearance for 3-D object recognition”, International Journal of Computer Vision, 40, 2, pp. 149-167, 2000.

[8] S. Z. Li, J. Hornegger, ``A Two-Stage Probabilistic Approach For Object Recognition''. In Proceedings of the European Conference on Computer Vision. University of Freiburg, Germany. 2-6 June, 1998.

[9] A. T. S. Au, P. W. M. Tsang, “Affine Invariant Recognition of 2D Occluded Objects using Geometric Hashing and Distance Transformation”, TENCON '96. Proceedings., IEEE TENCON. Digital Signal Processing Applications , Volume: 1 , 1996.

[10] G R. Ji, G Y. Wang, Z. Houkes, B. Zheng, Y. P. Han,” A New Method for Fast Computation of Moments based on 8-Neighbor Chain Code applied to 2-D Object Recognition”, Intelligent Processing Systems, ICIPS '97. 1997 IEEE International Conference on , Volume: 2 , 1997.

[11] S. Lin, S. W. Lee, ”Using Chromaticity Distributions and Eigenspace Analysis for Pose-, Illumination-, and Specularity-Invariant Recognition of 3D Objects”, Computer Vision and Pattern Recognition, Proceedings., IEEE Computer Society Conference on , Page(s): 426 –431, 1997.

[12] F. Jurie, “Solution of the Simultaneous Pose and Correspondence Problem Using Gaussian Error Model”, COMPUTER VISION AND IMAGE UNDERSTANDING, vol. 73 No3, pp. 357-373, 1999.

[13] D. W. Jacobs, R. Basri, ”3-D to 2-D Recognition with Regions”, Computer Vision and Pattern Recognition, Proceedings., IEEE Computer Society Conference on , Page(s): 547 –553, 1997.

[14] A. E. Johnson, M. Hebert, , ”Efficient Multiple Model Recognition in Cluttered 3-D Scenes”, Computer Vision and Pattern Recognition, Proceedings. IEEE Computer Society Conference on , Page(s): 671 –677, 1998.

[15] B. Cui, W. von Seelen, , ”Object Recognition of Robot based on Hidden Markov Models “, Signal Processing Proceedings, WCCC-ICSP. 5th International Conference on , Volume: 2 , Page(s): 851 –854, 2000.

[16] D. Roobaert, P. Nillius, J. O. Eklundh , “Comparison of Learning Approaches to Appearance-based 3D ObjectRecognition with and without Cluttered Background”, ACCV2000, Taipei, Taiwan, January 8-11, 2000.

[17] H. Zha, H. Nanamegi, T. Nagata, “Recognizing 3-D objects by using a Hopfield-style Optimization Algorithm for Matching Patch-based Descriptions”, Pattern Recognition 31 (6), pp. 727-741, 1998.

[18] Y. H. Gu, T. Tjahjadi, “Coarse-to-fine Planar Object Identification using Invariant Curve Features and B-spline Modeling”, Pattern Recognition 33 (9), pp. 1411-1422, 2000.

[19] K. C. Lo, S. K. W. Kwok, “Recognition of 3D Planar Objects in Canonical Frames”, Pattern Recognition Letters 22(6/7): 715-723, 2001.

[20] H. Murase, S. K. Nayar, “Visual Learning and Recognition of 3D Object from A ppearance”, International Journal of Computer Vision 14,5-24, 1995.

[21] C. Schmid, A. Zesserman, R. Mohr, “Integrating Geometric and Photometric Information for

Image Retrieval", In International Workshop on Shape, Contour and Grouping in Computer Vision, 1998.

[22] D. Lowe, "Object Recognition from Local Scale-Invariant Features", Proc. International Conference on Computer Vision, IEEE Press, 1999.

[23] K. Ohba, K. Ikeuchi, "Detectability, Uniqueness, and Reliability of Eigen-Windows for Robust Recognition of Partially Occluded Objects", IEEE Pattern Analysis and Machine Intelligence 19, no.9, 1043-1048.

[24] D. Jugessur, "Robust Object Recognition using Local Appearance based Methods", Master Thesis, McGill Univ., 2000.

[25] A.Treisman, "Features and Objects in Visual Processing", Scientific American, November 1986.

[26] C. Harris, M. Stehens. "A Combined Corner and Edge Detector", In Alvey Vision Conference, pages 147-151, 1988.

[27] S. C. Bae, "COP: A New Method for Extracting Edges and Corners", Mater's thesis, KAIST, 1997.

[28] C. Tomasi, T. Kanade. "Detection and Tracking of Point Features", Technical report, CMU, 1991.

[29] S. M. Smith and J. M. Brady, "SUSAN – A New Approach to Low Level Image Processing", Int. Journal of Computer Vision, 23(1):45-78, May 1997.

[30] C. Schmid, R. Mohr, C. Bauckhage, "Comparing and Evaluating Interest Points", In ICCV, pp. 230-235, 1998.

[31] T. Lindeberg, "Scale-space thery: A Basic Tool for Analysing Structures at Different Scales", Journal of Applied Statistics, 20, 2, pp. 224-270, 1994.

[32] C. H. The, R. T. Chin, "On Image Analysis by the Methods of Moments", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 10, no. 4, 1988.

[33] M. Hu, "Visual Pattern Recognition by Moment Invariants", IEEE Trans. on Inf. Theory, 8:179-834, 1962.

[34] T. H. Reiss, "The Revised Fundamental Theorem of Moments Invariants", IEEE Trans. on

Pattern Analysis and Machine Intelligence,13:830-834, 1991.

[35] S. X. Liao, M. Pawlak, “On the Accuracy of Zernike Moments for Image Analysis”, IEEE Tras. On PAMI, vol. 20, no.12, 1998.

[36] E. Bigorgne, C. Achard, J. Deavars, “An Invariant Local Vector for Content –based Image Retrieval. Pattern Recognition”, Proceedings. 15th International Conference on , Volume: 1 , 2000.

[37] R. Mariani, “Local Invariants and Local Constraints for Face Recognition, Pattern Recognition”, Proceedings. 15th International Conference on , Volume: 2 , 2000.

[38] S. M. Abdallah, B. MengSc, “Object Recognition via Invariance”, Doctor’s Thesis, The Univ. of Sydney, 2000.

[39] S. Arya, D. S. Netanyahu, R. Silverman, “An Optimal Algorithm for Approximate Nearest Neighbor Searching in Fixed Dimensions”, J. ACM45(6):891-923, Nov 1998.

[40] R. Hartley, “Multiple View Geometry in Computer Vision”, Cambridge Univ. Press, 2000.

[41] O. L. Rosin, “Robust Pose Estimation”, IEEE Trans. On System, Man, and Cybernetics, Vol. 29, No. 2, 1999.

[42] J. A. Nelder, R. Mead, “A Simplex Method for Function Minimization”, Computer J., 1965.

[43] S. Z. Li, “Markov Random Field Modeling in Computer Vision”, Springer-Verlag, 2nd Edition, 2001.

감사의 글

먼저 논문이 나오기까지 아낌없는 관심으로 논문 지도를 해 주신 권인소 교수님께 머리 숙여 감사 드립니다. 또한 논문 심사에 소중한 시간을 내셔서 좋은 조언을 해주신 김성대 교수님과 정명진 교수님께도 감사의 말씀을 드립니다.

또한 석사과정 동안 함께 한 연구실 선배님들, 후배님들께도 감사 드립니다. 연구실의 제일 맏형으로서 여러 가지 도움을 주신 왕헌이 형, 연구실 생활을 시작하면서 좋은 조언을 해 주신 기정이 형께 먼저 감사 드립니다. 연구실 동기로 같이 수업 들으며 함께 할 수 있었던 준혁이, 연구실이 잘 돌아갈 수 있도록 힘써 주신 승훈이 형, 호원이 형께도 감사 드립니다. 저의 사회성을 함양 시키기 위해 힘써 주신 국진이 형, 연구실 자산 관리 요령을 친절하게 가르쳐 주고 석사 논문 심사 연습 때 꼼꼼히 체크해 주신 준식이 형, 지금은 졸업했지만, 옆에 있어 항상 든든했던 승현이 형에게도 감사 드립니다. 그리고 1년 동안 연구실 식품을 관리해오느라 고생한 성림이형, 연구실 웹지기로 컴퓨터 관련 일을 도맡아해 온 영배, 그리고 저 대신 자산 관리를 잘 해 준 욱이에게도 고맙습니다.

힘들 때 같이 술도 마시며 고민을 나눈 룸메이트 웅희와 일홍에게도 고맙습니다. 그리고 알게 모르게 힘이 되어준 고대 선배님들께도 고맙습니다.

그리고 저에게 매우 헌신적인 사랑과 아낌없는 보살핌으로 든든한 힘이 되어준 선이에게 다시 한번 더 감사의 말을 전하고 싶습니다.

마지막으로 저를 항상 격려해 주신 아버지, 세상 그 누구보다도 훌륭하신 어머니, 큰 누나, 큰 형님, 시골 형님, 작은 형님, 작은 누나께 감사 드립니다.

이 력 서

성　　명 : **김 성 호** (**金 聖 浩**)

생년월일 : 1977**년** 7**월** 27**일**

출 생 지 : **경상북도 문경시**

본　　적 : **경상북도 문경시 산양면 형천**1**리** 96

학 력

1996.3 – 2000.2 **고려대학교 전기전자전파 공학부 학사과정** (B.S.)

2000.3 – 2002.2 **한국과학기술원 전자전산학과**

전기및전자공학전공 석사과정 (M.S.)