

공간 지역성을 기반으로 한 실시간 대규모 평면 물체 인식

이 수 원¹

¹조교수, 경상국립대학교 컴퓨터과학부, 경상국립대학교 기초과학연구소

Real-time Large-scale Planar Object Recognition Based on Spatial Locality

Suwon Lee¹

¹Assistant Professor, School of Computer Science and The Research Institute of Natural Science, Gyeongsang National University

[요 약]

평면 물체 인식은 컴퓨터비전의 중요한 문제 중 하나이다. 특히, 대상 평면 물체의 수가 많을 때인 대규모 평면 물체 인식은 여러 컴퓨터비전 응용에서 중요한 역할을 하는 등 그 가치가 높다. 본 논문에서는 실시간 대규모 평면 물체 인식을 위한 방법을 제안한다. 제안한 방법은 입력 영상을 다수의 지역으로 나누고, 공간 지역성을 기반으로 하여 지역별로 지지 후보를 선택한 후 이를 검증한다. 이 과정에서 평면 물체의 전경만이 인식에 사용되고 불필요한 연산이 최소화되어 빠른 속도로 정확한 인식이 가능하다. 실험을 통해 제안한 방법은 약 7%의 정확도 손실만으로 인식 대상 평면 물체의 수를 2,000개로 확장하여도 실시간 인식이 가능함을 보인다.

[Abstract]

Planar object recognition is one of the important problems of computer vision. In particular, large-scale planar object recognition, when the number of target planar objects is large, has high value, such as playing an important role in various computer vision applications. In this paper, we propose a method for real-time large-scale planar object recognition. The proposed method divides the input image into multiple regions, selects support candidates for each region based on spatial locality, and verifies them. In this process, only the foreground of a planar object is used for recognition and unnecessary calculations are minimized, enabling accurate recognition at high speed. Experimental results show that compared to the conventional method designed for a single planar object, the number of recognition planar objects is increased to 2,000 at the expense of only a 7% reduction in the recognition rate.

색인어 : 평면 물체 인식, 특정 물체 인식, 물체 인식, 공간 지역성, 고속 근사 매칭

Keyword : Planar Object Recognition, Specific Object Recognition, Object Recognition, Spatial Locality, Fast Approximate Matching

<http://dx.doi.org/10.9728/dcs.2021.22.11.1885>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 01 October 2021; **Revised** 29 October 2021

Accepted 29 October 2021

***Corresponding Author; Suwon Lee**

Tel: +82-55-772-1394

E-mail: leesuwon@gnu.ac.kr

I. 서론

평면 물체 인식은 컴퓨터비전의 중요한 문제 중 하나이다. 특히, 평면 물체와 카메라 사이의 3차원 자세를 추정하는 수준의 인식은 시각적 서보잉(visual servoing), 동시적 위치 추정 및 지도 작성(simultaneous localization and mapping, SLAM), 무인 항공기(unmanned aerial vehicle) 제어, 증강현실(augmented reality) 등과 같은 많은 컴퓨터 비전 기반 응용과 관련 분야에서 중요한 역할을 한다.

평면 물체와 카메라 사이의 3차원 자세를 추정하기 위해서는 대상 평면 물체만을 포함하고 있는 등록 이미지와 카메라 영상에서 포착된 평면 물체 사이에서 최소 4쌍의 위치 관련성(correspondence)을 이용해 호모그래피(homography)를 계산해야 한다. 최소 4쌍의 위치 관련성은 두 이미지 사이에서 추출된 키포인트(keypoint) 기반의 시각적 지역특징(local features)을 매칭(matching)하여 생성된다. 매칭 결과는 이상치(outlier)를 포함하기 때문에 매칭 결과의 정제(refinement) 등을 통해 이상치를 최소화한 후, 호모그래피를 계산하고 인식 여부를 판단할 수 있다.

한편, 인식해야 하는 대상 평면 물체의 수가 많아진다면 매칭부터 호모그래피 계산까지의 과정을 평면 물체의 수만큼 반복해야 한다. 이 방법은 대상 평면 물체의 수가 늘어날수록 연산 량 또한 그에 비례하여 증가하기 때문에 실시간성을 요구하는 응용에는 적합하지 않다.

본 논문에서는 인식해야 하는 대상 평면 물체의 수가 많을 때 카메라 영상에서 특정 평면 물체의 3차원 자세를 추정하는 수준의 인식을 대규모 평면 물체 인식이라고 정의하고, 이를 실시간에서 수행하기 위한 방법을 제안한다.

증강현실 등의 실제 응용에서는 평면 물체를 인식한 이후 프레임부터는 해당 평면 물체를 추적(tracking)하는 단계로 전환되어야 하지만 본 논문에서는 대규모 평면 물체를 인식할 때 발생하는 문제점들을 해결하는 데에 초점이 맞춰져 있다. 그 이유는 인식하고자 하는 평면 물체의 수는 인식 단계에만 영향을 미치고, 이후 추적 단계와는 무관하기 때문이다.

본 논문은 다음과 같이 구성된다. 2장에서는 전통적인 평면 물체 인식 방법과 본 논문에서 제안하는 방법에서 사용되는 알고리즘에 대해 살펴본 후, 3장에서는 제안한 방법에 대해 기술한다. 4장에서는 제안한 방법이 대규모 평면 물체의 실시간 인식에 얼마나 적합한지를 실험을 통해 비교하고 분석한다. 마지막으로 5장에서 결론을 맺고 향후 연구 방향을 모색한다.

II. 본론

2-1 전통적인 평면 물체 인식

전통적인 평면 물체 인식은 지역특징 추출, 지역특징의 매칭 및 정제, 호모그래피 계산 및 인식 여부 결정 등의 3단계로 구성된다.

지역특징은 이미지에서 반복성(repeatability)이 높은 지점을 키포인트로 검출한 후, 해당 키포인트의 위치와 크기를 고려하여 생성한 이미지 조각(patch)을 고차원의 벡터로 기술(description)하는 과정을 통해 추출된다. 등록 이미지는 보통 평면 물체의 경계가 주어지기 때문에 이를 이용해 물체 영역에서만 지역특징을 추출할 수 있다. 반면에 카메라 영상은 평면 물체와 배경이 섞여 있기 때문에 전체 영역에서 지역특징을 추출한다.

카메라 영상에서 추출한 지역특징은 등록 이미지에서 추출해놓은 지역특징들 중에서 가장 가까운 거리(distance)의 지역특징과 매칭된다. 매칭 결과 중에는 잘못된 매칭인 이상치(outlier)를 포함하는데 이는 인식을 방해하는 요소로 작용하기 때문에 최근린 거리 비율(nearest neighbor distance ratio, NNDR)과 양방향(bi-directional) 매칭[1,2] 등의 정제과정을 통해 최소화시킨다.

매칭된 지역특징들의 위치 쌍들을 이용하여 등록 이미지와 카메라 영상 사이의 평면 물체의 투시 변환(projective transformation)인 호모그래피를 계산한다. 정제과정을 거친 매칭 결과 중에도 이상치가 포함될 수 있기 때문에 무작위 표본 합의(random sample consensus)[3] 혹은 점진적인 표본 합의(progressive sample consensus, PROSAC)[4]를 이용해 샘플 추출을 달리하면서 호모그래피를 계산하고, 해당 호모그래피에 대한 이상치와 정상치(inlier)를 구분한다. 이 과정에서 계산된 호모그래피들 중에서 가장 많은 정상치(inlier)를 가지는 호모그래피가 최종적으로 계산된다. 정상치의 수 등을 이용해 호모그래피를 평가하여 최종 인식 여부를 결정한다.

전통적인 평면 물체 인식의 성능은 지역특징 추출 알고리즘과 정제 알고리즘, 호모그래피 계산 알고리즘의 조합에 따라 달라진다. 해당 알고리즘들의 조합에 따른 전통적인 평면 물체 인식의 성능을 비교 분석하는 논문[5]에서는 정제 알고리즘과 호모그래피 계산 알고리즘에 힘입어 모든 지역특징에서 준수한 정확도를 보여주고 있으며 특히, NNDR과 PROSAC의 조합이 속도 대비 정확도가 가장 높게 나타났다. 하지만 단일 평면 물체만을 대상으로 했어도 해당 조합은 인식 속도가 평균 100~200ms 정도로 실시간성을 보장하지 못하고 있다. 이렇듯 전통적인 평면 물체 인식 방법은 실시간이나 대규모 물체를 대상으로 할 때에는 적합하지 못하다.

2-2 지역특징과 이진특징, 그리고 ORB

평면 물체 인식의 첫 번째 단계인 지역특징 추출은 키포인트 검출과 기술로 구성된다. 가장 대표적인 알고리즘으로는 SIFT(Scale Invariant Feature Transform)[2]가 있다. SIFT는 크기 공간(scale space)에서 지역 극값(local extrema)을 키포인트로 검출한다. 검출된 키포인트는 위치와 크기, 방향 정보를 가지며, 이를 이용해 잘라낸 이미지 조각은 균일하게 정규화(normalization)된다. 이미지 조각을 여러 개의 격자로 나누고, 격자 단위로 계산한 경사도

(gradient)의 분포(histogram)를 하나의 벡터로 연결하여 지역특징을 추출한다. SIFT는 현재까지 가장 성공한 지역특징으로 인정받고 있지만 연산량이 크다는 단점이 있다. SIFT의 연산을 근사하여 연산속도를 대폭 향상시킨 SURF[6] 등이 제안되었지만 이 또한 실시간 시스템에는 적합하지 않다.

지역특징 추출의 높은 연산량을 개선하기 위한 방법으로 이진특징(binary features)이 연구되어 왔다. 이진특징은 단순히 두 픽셀 쌍들의 밝기 값을 비교하여 키포인트를 검출하고 기술한다. FAST(Features from Accelerated Segment Test)[7]는 특정 반지름 상에 놓인 이웃 픽셀보다 어둡거나 밝은 지점을 키포인트로 검출한다. 키포인트가 아닌 지점은 대부분 한두 번의 비교에서 걸러지기 때문에 매우 빠르게 전체에 대해 연산이 가능하다. BRIEF(Binary Robust Independent Elementary Feature)[8]는 주어진 이미지 조각에 대해 픽셀 쌍들의 밝기 값 비교를 여러 번 수행하여 0과 1로만 이루어진 벡터로 기술하는 이진기술자이다.

ORB(Oriented FAST and Rotated BRIEF)[9]는 FAST 키포인트에 방향 속성을 부여하고, 방향 속성에 대해 정규화한 이미지 조각을 BRIEF 기술자로 기술하여 추출한 이진특징이다. ORB는 SIFT나 SURF에 비해 시점 변화 등에 대한 강인함(robustness)이 다소 떨어지지만 최대 10배 이상의 빠른 연산량을 자랑한다. 뿐만 아니라 ORB는 최종 특징 벡터가 0과 1로만 이루어진 비트 열(string)이기 때문에 해밍 거리(Hamming distance)를 이용해 고속으로 매칭이 가능하다는 추가 장점을 지닌다.

2-3 고속 근사 매칭과 LSH

인식 대상의 수가 증가하면 자연스럽게 지역특징의 매칭 대상의 수도 증가한다. 매칭 대상의 수에 비례하여 매칭을 위한 연산량 또한 증가하기 때문에 대규모 평면 물체를 실시간으로 인식하기 위해서는 전수 조사를 통한 매칭은 현실적인 방법이 아니다.

매칭 대상이 많을 때 연산량 문제를 해결하는 방법은 고속 근사 최근린 검색(fast approximate nearest neighbor search) 기법을 이용한 고속 근사 매칭을 통해 약간의 정확도를 희생하여 연산량 이득을 크게 취하는 것이다. 고속 근사 최근린 검색은 가장 유사한 점을 찾는 것이 보장되지는 않지만 매우 빠른 속도로 최근린에 가까운 점을 찾아준다. 얼마나 빠르게 얼마나 높은 확률로 최근린을 찾는지가 관심사이다. 핵심 아이디어는 탐색해야 하는 범위, 즉 검색 공간의 축소에 있으며, 이는 최근린은 축소된 공간에 존재할 확률이 높다는 점에 기초한다. 대표적인 방법으로는 LSH(Locality Sensitive Hashing)[10,11]로 대표되는 해싱이 있다.

해싱은 여러 개의 이진사상(binary embedding) 함수를 정의하여 검색 공간을 축소한다. 하나의 이진사상 함수는 고차원의 벡터를 1비트로 압축하고, 여러 개의 이진사상에 의해 압축된 결과들이 모여 해시 테이블(hash table)을 구성한다.

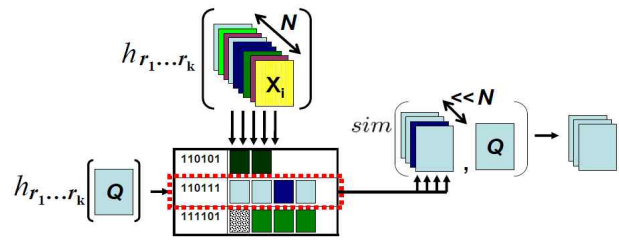


그림 1. LSH를 이용한 고속 근사 최근린 검색의 개요

Fig. 1. Overview of fast approximate nearest neighbor search using LSH

질의(query) 벡터에 대해 동일한 압축과정을 거치고, 압축 결과가 같은 벡터들에 대해서만 검색을 수행한다. 해싱은 이진사상 함수를 어떻게 정의하느냐에 따라 성능이 크게 좌우되며, 다루고자 하는 벡터의 도메인에 적합한 이진사상 함수의 정의를 필요로 한다. 그림 1[12]은 LSH를 이용한 고속 근사 최근린 검색의 개요를 보여준다.

III. 제안한 방법

3-1 오프라인 학습

인식 대상 평면 물체를 포함하고 있는 다수의 등록 이미지에서 각각 ORB 특징을 추출한다. 다수의 등록 이미지는 대상 평면 물체만을 포함하거나 대상 평면 물체의 경계가 주어지기 때문에 이를 이용해 대상 평면 물체 영역에서만 ORB 특징을 추출할 수 있다. 그림 2(a)는 등록 이미지의 대상 평면 물체에서 추출된 ORB 특징의 FAST 키포인트를 보여준다. 인식 단계에서 고속 근사 매칭을 위해 다수의 등록 이미지에서 추출된 모든 ORB 특징들을 모아서 LSH를 학습한다. LSH를 구성하는 파라미터는 이진사상 함수의 개수와 해시 테이블의 개수이다. 특징 벡터가 0과 1로만 이루어진 ORB 특징의 이진사상 함수는 특정 위치의 비트 값을 취하는 것으로 단순화된다. 학습된 LSH는 새로운 ORB 특징을 질의로 입력받아 학습 단계에서 추출된 ORB 특징 중에서 고속 근사 최근린을 검색하여 반환한다. 이때, 반환된 ORB 특징은 어떤 이미지에서 뽑힌 몇 번째 특징이라는 두 개의 인덱스를 가진다. 학습은 오프라인에서 한번만 수행된다.

3-2 실시간 인식

실시간으로 입력되는 카메라 영상은 대상 평면 물체의 위치에 대한 정보가 주어지지 않기 때문에 전체 영역에서 FAST 키포인트를 검출한다. 그림 2(b)에서 보듯이 대상 평면 물체 이외의 부분인 배경에서도 다수의 키포인트가 검출된다. 이렇게 검출된 키포인트들을 구분 없이 사용한다면 인식 성능의 저하로 이어질 가능성이 크다.

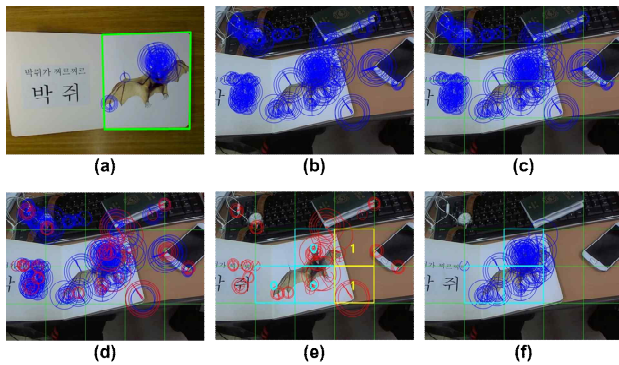


그림 2. 제안한 방법의 전체 과정
 Fig. 2. The whole process of the proposed method

이러한 문제를 해결하기 위해 공간 지역성(spatial locality)을 도입한다. 공간 지역성의 가정은 이미지에 투사된 물체는 공간적으로 이어져 있다는 것이다. 같은 대상 평면 물체의 영역에서 검출된 키포인트들은 서로 가까운 위치에 존재하며 키포인트의 위치로 서로 간의 인접성을 판단할 수 있다. 공간 지역성을 위해 그림 2(c)와 같이 영상을 여러 개의 격자로 이루어진 지역으로 분할하고, 지역 별로 균등하게 분포되도록 FAST 키포인트를 선택한다. 그림 2(d)는 선택된 FAST 키포인트를 붉은 색으로 시각화하고 있다. 선택된 FAST 키포인트에서만 BRIEF 기술자를 이용해 ORB 특징을 추출한다. 선택되지 않은 FAST 키포인트는 ORB 특징으로 추출되지 않을 수도 있기 때문에 연산 속도에 있어서도 장점을 갖게 된다.

지역별로 선택된 FAST 키포인트에서 추출된 ORB 특징을 3-1절에서 학습한 LSH에 질의로 입력한다. LSH는 질의로 입력된 ORB 특징과 3-1절에서 추출된 ORB 특징들과의 고속 근사 매칭을 수행한다. 매칭된 특징이 속하는 이미지의 인덱스를 가지고 지역별로 투표(voting)를 수행한다.

지역별로 가장 많은 표를 받은 이미지를 지역별 지지 후보로 결정하고, 같은 후보를 지지하면서 인접한 지역을 연결하여 국소 연결 지역(locally connected regions)을 생성한다. 그림 2(e)에서 보듯이 국소 연결 지역은 해당 후보에 대한 개략적 위치정보를 가진다. 그림 2(e)와 같이 여러 개의 국소 연결 지역이 생성된다면 이후 카메라 영상에 포착된 다수의 평면 물체를 동시에 인식하는 것도 가능하다.

국소 연결 지역별로 지지하는 이미지를 검증하기 위해 국소 연결 지역에 속한 모든 FAST 키포인트로부터 BRIEF 기술자를 이용해 ORB 특징을 추출한다. 이때, 이전 과정에서 선택되었던 FAST 키포인트는 이미 ORB 특징으로 추출되었기 때문에 생략된다. 그림 2(f)는 하나의 국소 연결 지역에서 최종적으로 추출된 ORB 특징의 FAST 키포인트를 보여준다. 해당 국소 연결 지역에 속하지 않는 배경에서 검출된 FAST 키포인트는 ORB 특징으로 추출되지도 않고, 최종 검증을 위한 매칭에도 쓰이지 않기 때문에 정확도와 연산 속도 모든 면에서 효과를 볼 수 있다.

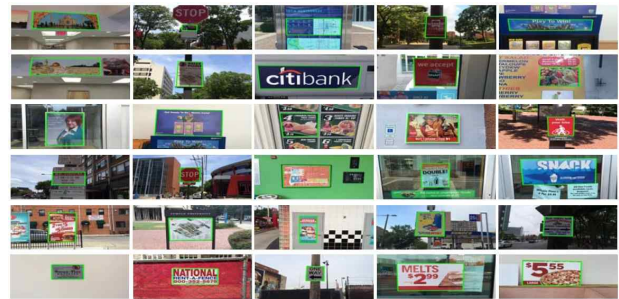


그림 3. 데이터베이스에 포함된 30개의 평면 물체
 Fig. 3. 30 planar objects in the database



그림 4. UKBench 데이터베이스의 견본 이미지들
 Fig. 4. Sample images from UKBench database

국소 연결 지역에서 추출된 ORB 특징을 이용해 해당 국소 연결 지역이 지지하는 후보의 등록 이미지에서 추출된 ORB 특징과의 매칭을 수행한다. 이후 매칭 결과의 정제, 호모그래피 계산 및 인식 여부 결정은 2-1절에서 설명한 전통적인 평면 물체 인식 방법과 동일하다. 실제 구현에서는 속도 대비 정확도가 가장 높은 NNDR과 PROSAC을 이용하였다.

IV. 실험 및 비교 분석

4-1 실험 환경 구축

실험을 위해 평면 물체 추적 문제의 벤치마킹 목적으로 제작된 POT(Planar Object Tracking)[13] 데이터베이스를 사용한다. POT 데이터베이스는 그림 3과 같이 30개의 평면 물체를 대상으로 7가지의 서로 다른 변화요소를 고려하여 촬영하여 총 210개의 영상으로 구성된다. 하나의 영상은 501개의 프레임에 포함하는데 첫 번째 프레임은 등록 이미지로 사용하고 나머지 500개의 프레임은 실시간 인식을 위한 이미지로 사용한다. 영상의 모든 프레임에는 인식의 성공여부를 판단하기 위한 평면 물체의 네 모서리 좌표가 주석으로 달려있다.

POT 데이터베이스는 특정 영상에서 인식 대상의 평면 물체가 하나라고 가정하고 제작되었기 때문에 본 연구에서 제안한 대규모 평면 물체의 실시간 인식 방법의 성능을 평가하기에는 적합하지 않다. 추가적인 이미지를 도입하기 위해 이미지 검색의 성능 평가를 목적으로 제작된 UKBench(Univ. of Kentucky Recognition Benchmark)[14] 데이터베이스를 추가로 사용하였다.

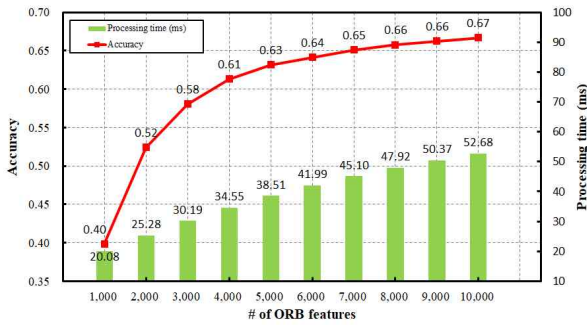


그림 5. ORB 특징 개수에 따른 베이스라인 방법의 단일 평면 물체 인식의 정확도 및 연산 시간

Fig. 5. Accuracy and processing time of single planar object recognition of baseline method according to the number of ORB features

UKBench 데이터베이스는 2,550개의 물체를 서로 다른 4개의 시점으로 촬영하여 총 10,200장의 이미지로 구성되어 있다(그림 4). 인식 대상이 대규모로 많은 상황을 고려하기 위해 POT 데이터베이스의 등록 이미지들과 함께 UKBench 데이터베이스의 이미지들을 같이 학습한다. 실시간 인식 단계에서 UKBench 데이터베이스의 이미지들은 인식을 방해하는 오답 역할을 하게 된다. 즉, 국소 연결 지역에서 UKBench 데이터베이스의 이미지를 지지하면 인식에 실패한 것이다. UKBench 데이터베이스의 하나의 물체를 다른 시점으로 촬영한 여러 개의 이미지가 포함되면 물체의 중복이 발생한 것이기 때문에 물체의 중복 없이 하나의 물체 당 하나의 이미지를 이용하여 총 2,550개의 오답 이미지를 준비하였다. 오프라인 학습 단계에서 학습에 참여하는 오답 이미지의 수가 실시간 인식 단계에서 인식하고자 하는 평면 물체의 수를 의미하며, 이를 늘려가면서 확장성(scalability)을 테스트한다. 모든 실험은 i7-6700(3.40GHz) CPU와 16.0GB RAM이 장착된 PC에서 수행되었다.

4-2 베이스라인: 단일 평면 물체의 실시간 인식

제안한 방법이 대규모 평면 물체 인식에 얼마나 적합한지를 비교하기 위해 베이스라인으로 ORB와 NNDR, PROSAC을 이용하여 2-1절에서 설명한 전통적인 방법의 단일 평면 물체 인식 시스템을 구현하였다. ORB는 추출할 최대 특징의 개수를 파라미터로 조절할 수 있는데 이는 평면 물체 인식의 정확도와 연산 속도 사이에 크게 영향을 미친다. 추출된 특징이 많을수록 정확도는 올라가지만 그만큼 연산 속도도 느려진다. 실시간성이 보장될 때의 단일 평면 물체 인식 시스템의 정확도가 어느 정도인지를 파악하기 위해 카메라 영상에서 추출되는 ORB 특징의 최대 개수를 조절하면서 정확도와 연산 속도를 측정하였다. 등록 이미지에서는 동일하게 최대 1,000개의 ORB 특징을 추출하였다.

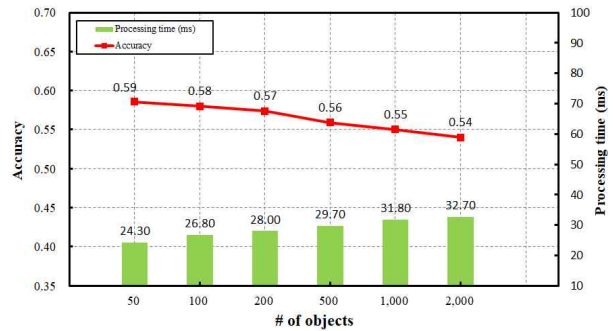


그림 6. 인식 물체의 수에 따른 제안한 방법의 대규모 평면 물체 인식의 정확도 및 연산 시간

Fig. 6. Accuracy and processing time of large-scale planar object recognition of proposed method according to the number of objects

그림 5는 베이스라인 방법으로 단일 평면 물체를 인식할 때 카메라 영상에서 추출되는 ORB 특징의 최대 개수를 1,000개에서 10,000개로 증가하면서 측정한 정확도와 연산 시간을 보여준다. 특징의 수가 증가하면서 정확도와 연산 시간이 같이 증가함을 볼 수 있다. 증가 폭 또한 점점 완만해지는데 특징의 수가 1,000개 증가할 때마다 증가하는 비율은 정확도가 더 급격하게 떨어지는 것을 볼 수 있다. 특징의 수가 증가할수록 시간대비 효율성이 떨어진다고 볼 수 있다. 단순히 정확도를 연산 시간으로 나눈 값을 시간대비 효율성으로 정의한다면 최대 특징의 수가 2,000개일 때 가장 높게 나타났다. 마지막으로 실시간 시스템의 기준을 초당 30프레임으로 정의한다면 실시간성이 보장될 때의 정확도는 최대 특징의 수가 3,000개일 때의 정확도인 0.58이었다.

4-3 제안한 방법: 대규모 평면 물체의 실시간 인식

3장에서 설명한 제안한 방법으로 대규모 평면 물체 인식 시스템을 구현하였다. 확장성을 함께 테스트하기 위해 POT 데이터베이스의 영상과 오답 이미지의 수를 달리하면서 테스트 세트를 구성하였다. 오답 이미지의 수는 최소 50개에서 최대 2,000개 사이의 여섯 구간을 사용하였다. 즉, 210개의 영상 × 6구간으로 총 1,260개의 테스트 세트를 구성하였다.

오프라인 학습 단계에서는 210개의 등록 이미지와 2,000개의 오답 이미지에서 각각 최대 1,000개의 ORB 특징을 추출하였다. 테스트 세트에 따라 등록 이미지 하나와 오답 이미지들을 다르게 구성하여 각각 LSH를 학습하였다. LSH는 20개의 이진사상함수를 가지는 3개의 해시 테이블로 구성하였다. 실시간 인식 단계에서는 카메라 영상에서 최대 1,000개의 FAST 키포인트 추출하였다. 공간 지역성을 이용하기 위해 4×5의 격자로 이루어진 지역으로 분할하고, 지역 별로 20개씩 FAST 키포인트를 선택하였다. LSH의 결과로 생성한 국소 연결 지역에서 추출한 ORB 특징을 이용해 최종 인식을 수행하였다.

그림 6은 제안한 방법으로 다수의 평면 물체를 인식할 때 인식하고자 하는 평면 물체의 수를 달리하면서 측정한 정확도와 연산 시간을 보여준다. 인식 물체의 수가 증가하면서 정확도와 연산 시간이 같이 증가하지만 최대 2,000개의 평면 물체에 대해서도 초당 30프레임 이상으로 실시간 인식이 가능함을 볼 수 있다. 실시간 시스템의 기준을 초당 30프레임으로 정의했을 때, 베이스라인 방법에 비해 약 7%의 정확도 손실만으로 인식 대상의 수를 한 개에서 2,000개로 확장하여 실시간 인식이 가능하였다. 인식 대상의 수가 50개 일 때는 오히려 높은 정확도를 보였는데 이는 제안한 방법의 공간 지역성으로 인한 효과로 해석된다. 제안한 방법은 평면 물체가 있을 확률이 높은 지역을 찾은 후에 해당 지역에서만 추출된 특징으로 매칭을 하기 때문에 배경에서 추출된 특징으로 인해 발생하는 이상치 매칭이 획기적으로 줄어들기 때문이다.

V. 결 론

본 논문은 인식 대상의 수가 많은 상황에서 평면 물체를 실시간으로 인식하기 위한 방법을 제안하였다. 제안한 방법의 핵심은 카메라 영상을 여러 개의 격자로 이루어진 지역으로 분할하고, 지역별로 지지하는 물체를 선택한 후, 같은 물체를 지지하면서 인접한 지역에서만 추출한 지역특징만으로 해당 물체를 검증하는 것이다. 이 과정에서 공간 지역성을 정의하고 ORB와 LSH를 이용하였다.

비교를 위해 전통적인 방법의 단일 평면 물체 인식 시스템을 구현하고, 제안한 방법으로 대규모 평면 물체 인식 시스템을 구현하였다. 인식 대상이 대규모인 상황을 시뮬레이션하기 위해 각 영상의 등록 이미지와 함께 방해 이미지들을 함께 학습하였다. 실험 결과, 제안한 방법이 전통적인 방법의 단일 평면 물체 인식 시스템에 비해 약 7%의 정확도 손실만으로 인식 대상의 수를 2,000개로 확장하여 실시간 인식이 가능함을 보였다.

본 논문에서 제안한 방법은 대규모 평면 물체를 인식하고 추적해야하는 모바일 증강현실의 응용이나 실시간 SLAM 시스템 등을 개발하는 데 있어서 중요한 역할을 할 것으로 기대된다. 실험을 통해 제안한 방법이 2,000개의 평면 물체를 대상으로 실시간 인식이 가능함을 보였지만 인식 대상의 수가 무한히 많아지는 등의 확장성 문제를 근본적으로 해결한 것은 아니다. 이를 위해 인식을 더 세분화하거나 고속 근사 매칭의 효율성을 극대화하는 연구 등이 필요해 보인다.

감사의 글

본 연구는 2021년도 중소벤처기업부의 기술개발사업[과제번호: S3071851] 지원에 의한 연구입니다.

참고문헌

- [1] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE transactions on pattern analysis and machine intelligence*, Vol. 27, No. 10, pp. 1615-1630, Oct 2005. <https://doi.org/10.1109/TPAMI.2005.188>
- [2] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 91-110, November 2004. <https://doi.org/10.1023/B:VISI.0000029664.99615.94>
- [3] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, Vol. 24, No. 6, pp. 381-395, June, 1981. <https://doi.org/10.1145/358669.358692>
- [4] O. Chum and J. Matas, "Matching with PROSAC-Progressive Sample Consensus," in *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, San Diego, CA, pp. 220-226, 2005. <https://doi.org/10.1109/CVPR.2005.221>
- [5] S. Lee, "Planar Object Recognition using Keypoint-based Visual Local Features: A Comparative Analysis," *Journal of Digital Contents Society*, Vol. 21, No. 9, pp. 1685-1690, 2020. <https://doi.org/10.9728/dcs.2020.21.9.1685>
- [6] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, Vol. 110, No. 3, pp. 346-359, June 2008. <https://doi.org/10.1016/j.cviu.2007.09.014>
- [7] E. Rosten and T. Drummond, "Machine Learning for High-Speed Corner Detection," in *Proceeding of European Conference on Computer Vision*, Graz, Austria, pp. 430-443, 2006. https://doi.org/10.1007/11744023_34
- [8] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary Robust Independent Elementary Features," in *Proceeding of European Conference on Computer Vision*, Crete, Greece, pp. 778-792, 2010. https://doi.org/10.1007/978-3-642-15561-1_56
- [9] E. Rublee, V. Rabaud, K. Konolige, and G.R. Bradski, "ORB: an Efficient Alternative to SIFT or SURF," in *Proceeding of International Conference on Computer Vision*, Barcelona, Spain, pp. 2564-2571, 2011. <https://doi.org/10.1109/ICCV.2011.6126544>
- [10] P. Indyk and R. Motwani, "Approximate nearest neighbors: Towards removing the curse of dimensionality," in *Proceeding of Symposium on Theory of Computing*, Dallas, Texa, pp. 604-613, 1998. <https://doi.org/10.1145/276698.276876>

- [11] M. Charikar, "Similarity estimation techniques from rounding algorithms," in Proceeding of Symposium on Theory of Computing, Montreal, Canada, pp. 380-388, 2002. <https://doi.org/10.1145/509907.509965>
- [12] B. Kulis, and K. Grauman, "Kernelized locality-sensitive hashing for scalable image search," in Proceeding of International Conference on Computer Vision, Kyoto, Japan, pp. 2130-2137, 2009. <https://doi.org/10.1109/ICCV.2009.5459466>
- [13] P. Liang, Y. Wu, H. Lu, L. Wang, and C. Liao, "Planar object tracking in the wild: A benchmark," in Proceeding of IEEE International Conference on Robotics and Automation, Brisbane, Australia, pp. 651-658, 2018. <https://doi.org/10.1109/ICRA.2018.8461037>
- [14] D. Nister and H. Stewenius, "Scalable Recognition with a Vocabulary Tree," in Proceeding of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, NY, pp. 2161-2168, 2006. <https://doi.org/10.1109/CVPR.2006.264>



이수원(Suwon Lee)

2012년 : 한국과학기술원 (공학석사)

2017년 : 한국과학기술원 (공학박사)

2018년~현재: 경상국립대학교 컴퓨터과학부 조교수

※ 관심분야 : 증강현실(Augmented Reality), 컴퓨터비전(Computer Vision) 등