RGBd 카메라 기반 Symmetric Object의 6D 자세 추정

6D Pose Estimation using RGBd Camera for Symmetric Objects

이정우[1] · 김송우[1] · 박찬진[1] · 현승민[1] · 정호진[2] · 윤한얼[3]
Jungwoo Lee, Songwoo Kim, Chanjin Park, Seungmin Hyun, Ho-Jin Jung and Han Ul Yoon

[1]연세대학교 컴퓨터정보통신공학부
E-mail: {jw.lee, swkim, cjpark04, smhyun}@yonsei.ac.kr
[2]연세대학교 일반대학원 전산학과
E-mail: hojinj@yonsei.ac.kr
[3]연세대학교 소프트웨어학부
E-mail: huyoon@yonsei.ac.kr

요 약

본 논문은 목표 객체의 6D 자세를 구하는 신경망 구조에 대해 논한다. 과실 수확에서 줄기와 연결된 꼭지를 정확히 끊어 과실을 수확하기 위해서는 과실의 6D 자세를 파악하는 것이 중요하다. 본 연구에서는 6D 자세 추정이 어려운 symmetric object로서 사과를 목표 객체로 고려하며, FFB6D 신경망을 적용하여 6D 자세 추정을 시도한다. 현실 환경과 유사한 Isaac Sim 가상 환경으로부터 카메라 객체를 통해, 학습을 위한 데이터 세트를 구성하기 위해 과실을 촬영한 이미지와 6D 자세 정답 데이터를 획득한다. 최종적으로, 이 방법론을 통해 symmetric object인 사과의 6D 자세를 강인하게 추정할 수 있는지 검증해본다.

키워드 : 6D 자세 추정, RGBd 카메라, PoseCNN, DenseFusion, PVN3D, FFB6D

1. 서 론

객체의 6D 자세 추정은 자율 주행이나 로봇의 파지 작업 수행을 위한 필수적인 요소이다. 카메라 센서를 통한 목표 객체의 6D 자세 추정은 다른 객체에 의해 목표 객체가 가리워지는 경우, 센서 노이즈, 조도와 같은 객체가 놓인 환경에 민감하게 반응한다. Yu et al.은 RGB 이미지를 사용하여 6D 자세를 추정하는 PoseCNN을 제시하고, 앞서 언급된 환경 조건들을 포함한 YCB-Video dataset을 이용해 성능을 검증하였다[1][2]. 이 성능검증에서 PoseCNN은 여러 객체가 겹친 상황이나 다양한 조도의 환경에도 기존 방법들보다 높은 성능을 보였다. 그러나 조도가 낮거나 객체의 텍스처 정보가 부족한 경우 낮은 정확도를 보였다. 이러한 문제점을 보완하기 위해 Wang et al.은 DenseFusion 네트워크를 제안하여 RGB와 depth 이미지를 이어붙이는 fusion 연산으로 6D 자세 추정의 성능을 PoseCNN에 비해 향상시켰다[3]. 이후 He et al.은 PVN3D 네트워크를 제안하여 fusion을 통해 등장하는 feature에 3D keypoint를 적용하였고, 이는 DenseFusion보다도 더 나은 성능을 보였다[4]. 또한, PVN3D를 개선한 FFB6D 네트워크를 제안하였으며 이는 RGB 이미지와 depth 이미지로부터 획득한 feature로 bidirectional fusion module을 사용한 것이 특징이다[5].

감사의 글 : 이 성과는 정부(과학기술정보통신부, 농림축산식품부)의 재원으로 한국연구재단(No. 2021R1F1A1063339, 시각 정보가 제한된 비정형 환경에서의 로봇 매니퓰레이터 물체 파지를 위한 메타-러닝 기반 Peep-and-Pick 알고리즘) 및 농림식품기술기획평가원(No. 122032-03-1-HD020, 노지분야 스마트농업기술단기 고도화사업)의 지원을 받아 수행되었습니다. 연구비 지원에 감사드립니다.

FFB6D를 활용하여 객체의 6D 자세 추정을 하기 위해서는, 이미지와 정답 레이블로 구성된 데이터 세트가 필요하다. 본 논문에서는 먼저, FFB6D의 학습을 위한 데이터 세트를 Isaac Sim 기반 가상 환경에서 생성하는 방법론에 대해 논한다. 본 연구의 궁극적인 목표는, 가상 환경의 객체가 아닌 현실 세계 객체의 6D 자세 추정이므로, 텍스처나 물리 엔진과 같은 컴포넌트들의 reality gap이 적은 Isaac Sim 환경에서 데이터세트 생성을 진행한다. 데이터 세트는 매 프레임마다 객체가 다르게 표현되는 비디오 영상으로부터 생성한다. 이후, 생성된 데이터세트를 FFB6D에 적용하여, 가상 환경 객체의 6D 자세 추정에 대한 성능을 검증한다.

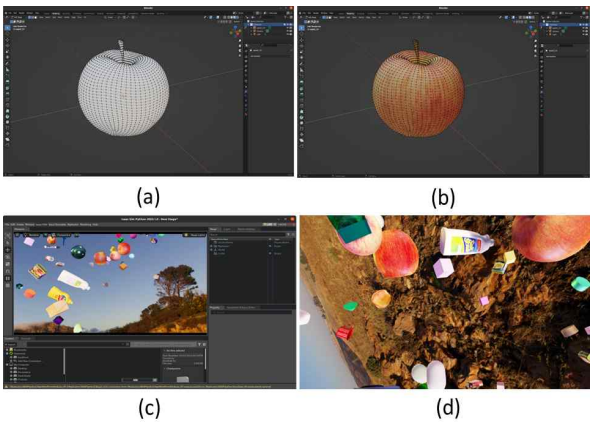

그림 1. (a) 3D 모델링된 사과 객체, (b) 텍스처를 입힌 사과 객체, (c) Isaac Sim에서 생성한 클러터 장면 (d) 생성된 장면을 Isaac Sim 카메라 객체로 촬영한 이미지

2. 제안하는 방법론

2.1 학습을 위한 Isaac Sim 환경에서의 데이터 생성

그림 1은 학습 및 테스트 데이터 획득을 위한 사과의 3D 모델링과 이미지를 생성하는 과정을 보여준다. 그림 1(a)는 Blender(v3.6, Blender Foundation, Amsterdam, the Netherlands)로 모델링한 사과 객체의 3D 모델을 보여주고, 그림 1(b)는 모델링된 매쉬에 텍스쳐를 적용한 모습이다. 이때, 꼭지 모양, 특정 위치의 반점, 줄무늬와 같은 특징에 따라 5가지 다른 3D 사과 모델을 생성한다. 그림 1(c)는 사과 객체들과 다른 객체들로 이루어진 클러터(clutter)로 구성된 Isaac Sim 장면(stage)를 보여준다. 장면들을 생성할 때, variety를 증가시켜 domain randomization 효과를 주기 위해, 다양한 조명과 배경의 조건을 적용한다. 마지막으로 그림 1(d)는 생성된 장면을 해당 장면 내부에 배치된 Intel realsense d435i Isaac Sim 카메라 객체 모델로 촬영한 이미지이다.

위 방법을 통해 획득한 데이터는 {RGB 이미지, depth 이미지, 배경이 마스크된 정답(label) 이미지, 객체들의 6D 자세 값}으로 구성된다. 데이터 요소들의 각 차원은 다음과 같다: RGB 이미지(1280×720), depth 이미지(1280×720), 배경이 마스크된 정답(label) 이미지(1280×720), 객체들의 6D 자세 값(5×(3×4)=클래스 개수×(마지막 행을 제외한 homogeneous transformation matrix)). 이를 FFB6D를 통한 목표 객체의 6D 자세 추정을 위한 학습 데이터 세트로 이용한다.

2.2 6D 자세 추정을 위한 FFB6D 신경망 구조

그림 2는 FFB6D(입력: RGBd 이미지, 출력: 6D 자세 추정값)의 파이프라인을 보여준다. FFB6D는 DenseFusion이 적용된 PVN3D를 백본 아키텍쳐로 한 구조이다[4][5]. FFB6D는 RGB 이미지로부터 CNN 연산을 통해 이미지 pixel 정보를, depth 이미지로부터 point cloud network(PCN) 레이어를 통과시켜 point 정보를 획득한다. 이후, RGB 이미지로부터 획득한 정보와 depth 이미지로부터 획득한 정보를 공유하기 위해 bidirectional fusion module로 point-to-pixel과 pixel-to-point 연산을 수행한다. 또한, 객체의 keypoint를 획득하기 위해 scale invariant feature transform-farthest point sampling(SIFT-FPS) 사용하였다. SIFT-FPS는 2D에서 scale invariant한 pixel들을 먼저 구하고, 3D로 lifting을 거쳐 FPS를 적용한다. 이로 인해, 모델이 object 구별에 더 효과적인 keypoint들을 선택할 수 있게 된다. 결과적으로 FFB6D는 YCB-Video 데이터 세트를 통한 검증에서 DenseFusion과 PVN3D보다 더 높은 정확도를 보였다[5].

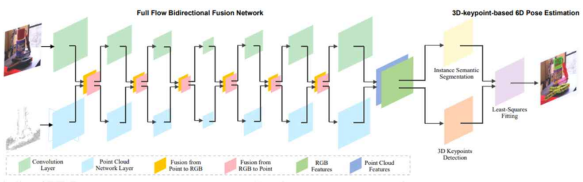

그림 2. FFB6D(입력: RGBd 이미지, 출력: 6D 자세 추정값)의 파이프라인[5]

3. 시뮬레이션 및 결과

그림 3은 테스트에 사용한 원본 이미지와 FFB6D 모델을 통해 획득한 6D 자세 추정에 대한 결과 이미지를

그림 3. 테스트 이미지(좌), FFB6D 모델로 추정한 결과 이미지(우)

보여준다. 그림 3(좌)에 보이듯, 테스트 이미지는 랜덤한 조도와 배경을 가지며, 랜덤한 자세의 다양한 클러터와 5개의 사과를 포함한다. 또한, 자세 추정 난이도를 올리기 위해 다른 객체가 사과를 가리고 있는 경우가 포함되었다. 그림 3(우)는 학습된 모델에 의해 추정된 6D 자세를 point mesh로 보여준다. 총 5개의 사과에 대해 객체에 대한 인식과 동시에 각 객체의 6D 자세를 성공적으로 추정하고 있는 것을 목표 객체들 위에 overlay된 point mesh들을 통해 확인할 수 있다.

4. 결론 및 향후 연구

본 논문에서는 FFB6D를 통해 symmetric object인 사과에 대해 6D 자세 추정을 진행하였으며, 이때 가상 환경인 Isaac Sim을 이용하여 데이터 세트를 생성하고 학습에 이용하였다. 테스트 결과, 목표 객체인 사과에 대해 허용 오차 이내로 6D 자세를 추정하는 모습을 확인할 수 있었다.

향후 이 연구를 로봇 매니퓰레이터를 이용한 정밀한 과실 수확의 핵심 요소 기술로 사용하기 위해, 현실에서 로봇 매니퓰레이터의 손목에 장착된 RGBd 카메라를 통해 실제 사과의 6D 자세 추정을 시도하고, 성능을 검증해볼 계획이다.

참 고 문 헌

[1] X. Yu et al., “Posecnn: a convolutional neural network for 6d object pose estimation in cluttered scenes,” *arXiv preprint* arXiv:1711.00199, 2017.

[2] B. Calli et al., “The ycb object and model set: towards common benchmarks for manipulation research,” in *Proceedings of the International Conference on Advanced Robotics (ICAR),* Washington, USA, May 26-30, 2015, pp. 510-517.

[3] C. Wang et al., “Densefusion: 6d object pose estimation by iterative dense fusion,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR),* Long Beach, CA, USA, Jun. 18-20, 2019, pp. 3343-3352

[4] Y. He et al., “Pvn3d: a deep point-wise 3d keypoints voting network for 6dof pose estimation,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR),* Seattle, Washington, Jun. 16-18, 2020, pp. 11632-11641.

[5] Y. He et al., “Ffb6d: a full flow bidirectional fusion network for 6d pose estimation,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* online, June. 19-25, 2021, pp. 3003-3013.