딥러닝 기반 6D 객체 포즈 개선을 위한 3D 윤곽 부정합 손실

염옥영, 박한훈*

부경대학교

yeyuning12@gmail.com, *hanhoon.park@pknu.ac.kr

A 3D Contour Misalignment Loss for Deep Learning-Based 6D Object Pose Refinement

Yuning Ye, Hanhoon Park*

Pukyong National Univ.

요 약

본 논문은 RGB 입력 영상으로부터 객체의 6D 포즈를 개선(refinement)하기 위한 딥러닝 모델을 학습하기 위해 3D 윤곽 부정합 손실(contour misalignment loss)를 제안한다. LINEMOD 데이터셋을 사용한 실험을 통해 포즈 오차의 크기에 따라 차이는 있으나 제안된 손실 함수를 사용하여 학습된 모델이 효과적으로 객체 포즈 오차를 추정함으로써 반복적으로 객체의 포즈를 개선하여 객체 포즈의 정확도를 향상시킬 수 있음을 확인하였다.

Ⅰ. 서 론

객체의 6D 포즈 추정은 객체가 존재는 월드 좌표계와 카메라 좌표계 사이의 3D 회전 변환과 3D 이동 변환을 찾는 문제이다. RGB 영상만을 사용한 객체 포즈 추정은 3D 정보의 부재 및 다양한 요인으로 인해 부정확한 결과를 초래한다. 그래서, 부정확한 포즈를 개선(refine)하는 것이 중요한데, 최근 딥러닝 기술의 발전으로 인해 깊은 신경망 모델을 활용하여 포즈 오차를 직접적으로 회귀(regress)하는 딥러닝 기반 방법들이 제안되고 있다[1]. 본 논문에서는 딥러닝 기반 6D 객체 포즈 개선을 위한 3D 윤곽 부정합 손실(contour misalignment loss)을 제안하고, 실험을 통해 제안된 손실 함수를 사용하여 학습된 모델이 효과적으로 객체 포즈 오차를 추정함으로써 객체 포즈의 정확도를 향상시킬 수 있음을 확인하였다.

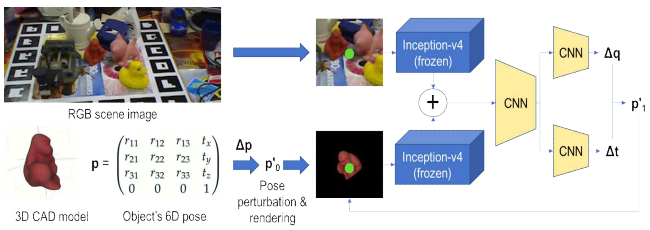

그림 1. 객체 포즈 개선을 위한 딥러닝 모델 구조.

Ⅱ. 실험 및 결과

2.1. 딥러닝 모델 구조 및 학습

그림 1에서 보는 것처럼, 객체의 포즈가 서로 다른 영상을 입력으로 하며, Inception-v4를 백본(backbone)으로 사용하고 샴(Siamese) 구조를 가진 네트워크를 사용하여 두 영상 사이의 포즈 오차를 추정한다. 학습을 위해 객체를 포함하는 RGB 입력 영상, 객체의 6D 포즈, 객체의 3D CAD 모델을 필요로 한다. 주어진 객체의 포즈 $\mathbf{p}$에 임의의 작은 변화(perturbation)를 주어 얻어진 $\mathbf{p}'$를 사용하여 객체의 CAD 모델을 렌더링한 영상을 생성한 후, 입력 영상과 렌더링 영상으로부터 객체 영역의 패치를 잘라서 네트워크의 입력으로 하여 두 패치 사이의 포즈 오차를 추정하도록 네트워크 파라미터를 갱신한다. 이 과정을 다양한 포즈 변화를 주면서 반복한다. 본 논문에서는 포즈 변화의 크기를 회전과 이동에 대해 각각 3 ~ 20도, 0.01 ~ 0.1m로 설정하였다.

2.2. 손실 함수

2.1 절에 설명한 학습을 위해서는 손실 함수가 정의되어야 하며, 학습은 손실 함수를 최소화하면서 네트워크 파라미터를 갱신하는 과정이다. 본 논문에서는 손실 함수를 식 (1)과 같이 객체의 윤곽점들의 3차원 좌표 사이의 거리로 정의된다.

$$L(\Delta q, \Delta t) = \sum_{v' \in b(V_{p'})} d_{\min}\left(\Delta q v' \Delta q^{-1} + \Delta t,\ V_p\right). \quad (1)$$

여기서, $\Delta\mathbf{q}$는 회전 변환을 나타내는 쿼터니언(quaternion) 벡터이고, $\Delta\mathbf{t}$는 이동 변환 벡터이다. $\mathbf{V}_p$와 $\mathbf{V}_{p'}$는 객체의 3D CAD 모델의 점들을 각각 $\mathbf{p}$와 $\mathbf{p}'$를 사용하여 변환된 점들의 집합이다. $b(\mathbf{V})$는 집합 $\mathbf{V}$ 내의 점 중에서 가장자리(윤곽)에 있는 점들로 이루어진 부분 집합을 얻는 함수이고, $d_{\min}(v, \mathbf{V})$은 점 v로부터 집합 $\mathbf{V}$ 내의 가장 가까운 점 사이의 거리를 계산하는 함수이다.

2.3. 실험 환경

딥러닝 모델의 구현, 학습, 검증을 위해 TensorFlow를 사용하였다. 학습을 위해 배치 크기는 16, 학습률은 1e-4로 설정하고, Adam 옵티마이저를 사용하였다. 학습된 모델의 검증은 학습 시와 유사한 형태로 이루어진다. 객체의 6D 포즈 $\mathbf{p}$를 알고 있는 입력 영상에 대해 임의로 작은 포즈 변화를 준 $\mathbf{p}'_0$를 사용하여 객체의 CAD 모델을 렌더링한 영상을 생성한 후, 입력 영상과 렌더링 영상으로부터 객체 영역의 패치를 잘라서 학습된 모델

의 입력으로 사용하여 두 영상 사이의 포즈 오차 $\Delta\mathbf{p}$를 추정한다. 추정된 포즈 오차를 사용하여 $\mathbf{p}'_0$를 개선하여 $\mathbf{p}'_1$을 얻는다. 위의 과정을 N번 반복해서 $\mathbf{p}$와 $\mathbf{p}'_N$ 사이의 차이를 구하였다.

2.4. 데이터셋

6D 객체 포즈 추정을 위한 벤치마크 데이터셋인 LINEMOD를 사용하였다[2]. LINEMOD는 15개의 객체에 대한 18,273장의 RGB 영상으로 이루어지며, 각 영상에서의 객체의 포즈 정보 및 3D CAD 모델이 함께 제공된다. 데이터셋은 복잡한 장면, 텍스처가 없는 객체, 조명 조건의 변화 등 객체 포즈 추정에서의 도전적인 환경을 포함하고 있다.

2.5. 실험 결과

그림 2는 포즈 변화의 크기에 따른 포즈 개선 결과를 보여준다. 포즈 변화의 크기에 상관없이 포즈를 반복적으로 개선함으로써 개선된 포즈를 사용하여 렌더링된 객체와 입력 영상 내의 객체 영역이 거의 일치하였다. 이는 포즈가 올바른 방향으로 개선되었음을 의미한다.

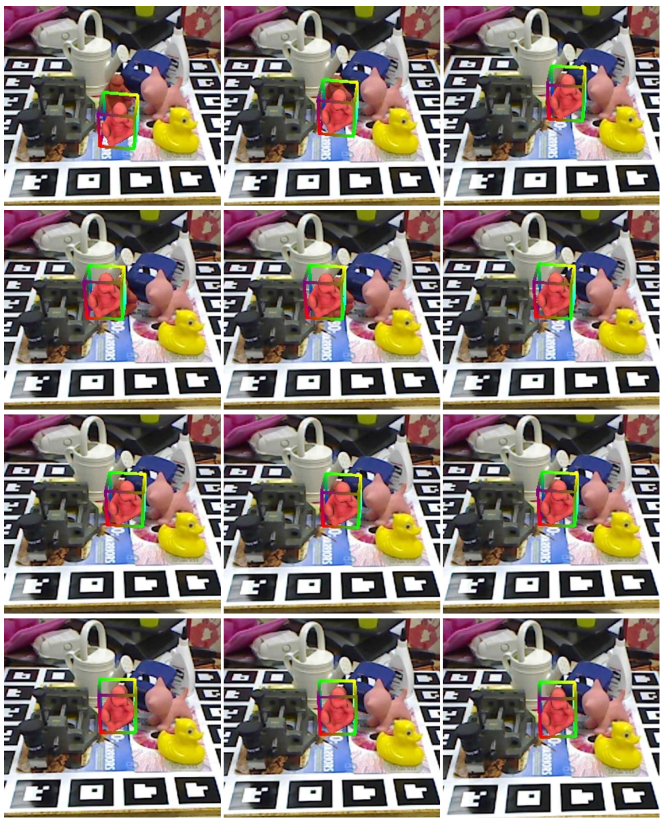

그림 2. 크기가 서로 다른 포즈 변화를 가지는 입력 영상에 대한 포즈 개선 결과. 왼쪽 열: 포즈 개선 전, 중간 열: 포즈 개선 진행 중, 오른쪽 열: 포즈 개선 후.

포즈 개선 정도를 정량적으로 분석하기 위해 반복 횟수에 따른 회전과 이동 변환의 오차를 계산하였다. 표 1과 2는 포즈 개선을 10번 반복했을 때의 개선량을 보여준다. 회전 및 이동 변환의 오차가 클수록 개선량은 증가했다. 그러나, 포즈 개선을 반복하더라도 오차가 계속해서 줄지는 않았는데, 이는 그림 3을 통해 확인할 수 있다. 즉, 반복 횟수가 일정 이상이 되면 오차는 오히려 증가할 수 있으며, 포즈 오차가 클 경우 포즈 개선을 통해 줄일 수 있는 오차에 한계가 있음을 알 수 있다. 결과적으로, 반복 횟수는 실험적으로 적절히 설정할 필요가 있다.

표 1. 이동 변환의 변화(perturbation) 크기에 따른 개선 결과

이동 변환의 변화 크기	개선량 (평균)
<= 0.07m	0.02013m
> 0.07m	0.03458m

표 2. 회전 변환의 변화(perturbation) 크기에 따른 개선 결과

회전 변환의 변화 크기	개선량 (평균)
<= 10도	1.53305도
> 10도	1.87388도

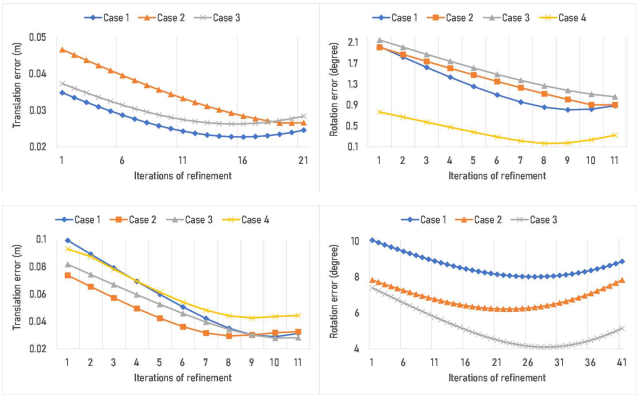

그림 3. 반복 횟수에 따른 포즈 개선 결과. 위 행: 포즈 변화의 크기가 작을 때, 아래 행: 포즈 변화의 크기가 클 때.

Ⅲ. 결론

본 논문에서는 객체의 3D CAD 모델과 대략적인 객체의 6D 포즈가 주어진 RGB 입력 영상에 대해 객체의 정확한 포즈를 얻기 위해, 주어진 객체의 포즈를 이용하여 객체를 렌더링한 영상과 입력 영상으로부터 주어진 포즈와 ground-truth 포즈 사이의 포즈 오차를 추정하도록 딥러닝 모델을 학습하였다. 학습을 위한 손실 함수로 3D 윤곽점 사이의 거리를 사용하는 것을 제안하였으며, 실험을 통해 학습된 딥러닝 모델을 사용하여 주어진 객체의 포즈를 반복적으로 개선할 수 있음을 확인하였다.

그러나, 객체의 포즈 오차가 클 경우 제안된 방법을 사용하더라도 일정한 크기 이하로 감소하지 않았으며, 반복 횟수를 증가시킬 경우 오차가 오히려 커질 수도 있는 문제가 있었다. 향후 이러한 문제를 해결하기 위해 딥러닝 모델의 구조나 손실 함수 등을 변경하고 성능을 분석하는 실험을 수행하고자 한다.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) Grant by the Korean Government through the MSIT under Grant 2021R1F1A1045749.

참 고 문 헌

[1] F. Manhardt, W. Kehl, N. Navab, and F. Tombari, “Deep model-based 6D pose refinement in RGB,” Proc. of ECCV, pp. 833-849, 2018.

[2] S. Hinterstoisser, C. Cagniart, S. Ilic, P. Sturm, N. Navab, P. Fua, and V. Lepetit, “Gradient response maps for real-time detection of textureless objects,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 5, pp. 876-888, 2012.