

흡착 그리퍼 강체 파지법을 위한 인공지능 퓨전 네트워크 개발

김찬혁, 양경옥, 최준원*

한양대학교, *서울대학교

{chkim, koyang}@spa.hanyang.ac.kr, *junwchoi@snu.ac.kr

Development of an Artificial Intelligence Fusion Network for Suction Gripper Rigid-Body Grasp Synthesis

Chanhyeok Kim, Kyoung Ok Yang, Jun Won Choi*

Hanyang Univ., *Seoul Univ.

요약

최근 스마트 물류 환경에서 매니퓰레이터 (Manipulator)를 활용한 픽앤플레이스 (Pick and Place)에 대한 수요가 증가하고 있다. 그 중 협소한 공간에서 물체들을 잡기 위하여 흡착 그리퍼 (Suction Gripper)의 파지점 생성 (Grasp Synthesis)에 대한 연구들이 이뤄지고 있다. 본 논문에서는 흡착 그리퍼 (Suction Gripper)의 강체 파지점 생성을 위한 인공지능 퓨전 네트워크를 제안한다. 제안한 네트워크는 Gray 이미지와 Depth 이미지를 이용한 파지점 생성 방식을 사용한다. 각 Gray 이미지와 Depth 이미지는 독립적인 특징 추출 과정을 거친 후 퓨전 모듈을 사용하여 특징값을 효과적으로 융합하였다. 제안한 네트워크의 성능 평가를 위해 MuJoCo에서 YCB object를 사용해 이미지 데이터를 수집했으며, 제안한 네트워크의 성능은 기존보다 2.7% 개선된 99.02%의 mAP를 보였다.

I. 서론

매니퓰레이터 (Manipulator)란 인간의 팔과 유사한 동작을 수행할 수 있는 다관절 로봇을 말한다. 스마트 팩토리, 스마트 물류와 같은 산업에서 매니퓰레이터는 간단한 조립 업무나, 종단간 생성품 이동 작업을 수행함으로써 인간의 노동을 대체하여 작업 생산성을 높여 주었다. 물리적인 힘을 많이 요구하는 업무이거나 위험한 환경에서도 작업을 수행할 수 있기 때문에 작업 생산성 향상과 운영 비용 절감 측면에서 유용하게 사용되고 있다. 매니퓰레이터를 이용해 물체를 픽앤플레이스 (Pick and Place)하기 위해서는 파지점 생성(Grasp Synthesis)에 대한 기술이 필수적이다[1]. 특히, 물류 환경에서는 다양한 물체에 대한 픽앤플레이스가 요구되며 생산성과 직결된 문제이기 때문에 강인한 파지점을 생성할 수 있는 기술이 다른 매니퓰레이터 환경에 비해 중요하다.

물류 환경에서 매니퓰레이터로 물체를 픽앤플레이스하기 위해 인공지능을 활용한 연구가 진행되고 있다. 아마존에서는 자동 파지 시스템을 위한 Benchmark 데이터셋인 ARMBench[3]을 만들어 제공하고, 2015년부터 Amazon Picking Challenge를 개최하는 등 물체의 파지에 대한 연구를 활성화하는데 힘쓰고 있다. 또한, 버클리 대학교에서 DexNet[2]을 개발하여 파지점 생성이 수학 모델 기반 방식뿐 아니라 데이터 기반 방식의 연구도 수행 되었다. 물체를 안정적으로 파지하여 중력 및 이동에 의한 관성에 안정적으로 저항할 수 있는 상태를 힘폐쇄 (Force Closure)라고 한다. DexNet은 지점의 힘폐쇄 확률을 예측하는 방식으로 파지점을 생성한다. 물체를 파지하기 위한 매니퓰레이터의 그리퍼로는 크게 집게 그리퍼 (Parallel-jaw Gripper)와 흡착 그리퍼 (Suction Gripper)가 있다. 집게 그리퍼는 두 개의 손가락을 이용해 물체를 잡는 방식의 엔드 이펙터 (End-Effector)로 물체의 종류에 구애받지 않고 물체를 집을 수 있는 장점이 있다. 하지만, 물체가 집게보다 큰 경우이거나, 벽이나 다른 물체로 손가락이 들어갈 공간이 협소한 경우에 사용하기 어렵다. 흡착 그리퍼는

진공의 압력을 이용해 물체를 잡는 방식을 사용한다. 집게 그리퍼에 비해 접촉점을 최소화할 수 있다는 장점을 가지고 있어서 다양한 물건이 좁은 공간에 들어가는 경우에 용이하게 사용되어진다.

본 논문에서는 흡착 그리퍼의 파지점 생성을 위한 인공지능 네트워크를 제안하고자 한다. 기존 연구한 인공지능 네트워크인 DexNet은 흡착 그리퍼의 파지점을 생성하기 위해 Gray 이미지와 Depth 이미지를 병합하여 결과를 예측하는 구조를 가진다. 하지만 DexNet에서는 Multimodal 데이터에 대하여 적절한 퓨전 방식이 적용되지 않아서 Multimodal의 데이터 값을 효과적으로 적용하지 못했다는 단점이 있었다. 따라서 본 논문에서는 DexNet과 동일하게 Gray 이미지와 Depth 이미지 데이터를 사용하여 흡착 파지점을 생성하지만 Depth 이미지 데이터의 특징점 추출 레이어를 강화하고, Multimodal 특징값을 합치는 과정을 새롭게 제안하여 성능을 향상시키고자 하였다. 결과적으로 본 논문에서 제안한 네트워크와 기존 DexNet을 비교 실험했을 때 mAP가 96.30%에서 99.02%로 약 2.7% 성능이 향상되는 결과가 나타났다.

II. 본론

본 논문에서는 흡착 그리퍼의 파지점 생성을 위한 새로운 네트워크를 제안하고자 한다. 본 논문에서 제안하는 네트워크는 Gray 이미지와 Depth 이미지를 입력으로 받고 서로 다른 Multimodal 데이터에 대한 결과를 퓨전하여 최종 파지점을 예측하는 모델을 개발했다. 흡착 그리퍼를 이용한 파지점은 표현 방식에 따라 Yaw 값을 제외한 5 자유도 파지 방식과 조감도 (Bird-Eye View)에서 파지를 예측하는 탑 다운 (Top-Down) 방식이 있다. 이 중 우리는 탑다운 방식을 사용하여 파지점 생성을 진행하였다. 본 논문에서는 그림 1과 같은 DexNet-F를 제안한다. DexNet-F는 Gray 이미지와 Depth 이미지를 입력으로 받아서 특징값을 추출하기 위한 독립적인 추출 과정을 거친 후 통합 결과값을 도출하도록 설계되었다. Depth

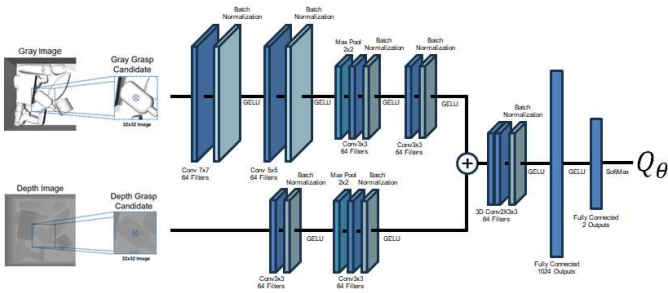


그림 1. 제안하는 DexNet-F 네트워크의 아키텍처

이미지에서 특징 값을 추출할 때 Depth가 가지는 공간적인(Spatial) 정보를 충분히 획득하기 위하여 컨볼루션 레이어 (Convolution layer) 구조를 가지도록 하였다. Depth 이미지에서 추출되는 특징값을 학습하는 과정에서 일반화를 확보하기 위하여 배치 정규화 (batch normalization)를 추가하였다.

Gray 이미지에서는 기본적으로 DexNet과 동일한 추출 과정을 거치도록 설계하였다. 기존과의 다른 점은 각 레이어가 GELU를 활성화 함수로 사용한다는 것에 있다. 음의 결과 값도 사용함으로써 결과값의 범위를 다양하게 사용할 수 있도록 하였다. 이러한 과정을 거친 특징값들을 각 Multimodal 특징값을 채널로 쌓아서 3D 컨볼루션을 수행하여 퓨전을 진행하도록 하였다. 마지막으로 완전 연결 레이어를 통과해 힘폐쇄(force closure) 확률을 예측하였다. 네트워크가 예측한 힘폐쇄 확률이 임계치 이상인 후보군들 중에서 최대값을 가지는 지점을 파지점으로 선정하도록 설계하였다.

III. 실험

본 논문에서 제시한 DexNet-F의 성능을 테스트하기 위해서 Accuracy, Recall, Precision, F1 Score 그리고 mAP 메트릭을 가지고 성능을 평가하였다. 실험에서 이미지 데이터는 YCB Object[4]를 사용했고 시뮬레이션 결과는 MuJoCo에서 수집하였다. 인공지능 네트워크는 Intel i7-12700F, Nvidia GeForce Titan X 12G 환경에서 학습하였고, 하이퍼파라미터로는 배치 사이즈 32, 학습률 0.0001, 20 epoch으로 설정하였다.

표 1. DexNet과 DexNet-F의 성능표(%)

Network	Accuracy	Recall	Precision	F1 Score	mAP
DexNet[2]	94.53	94.53	94.90	94.62	96.32
DexNet-F (Ours)	98.70 (+4.2)	98.70 (+4.2)	98.69 (+3.7)	98.70 (+4.1)	99.02 (+2.7)

DexNet과 제안한 네트워크인 DexNet-F의 성능을 비교하여 표 1로 정리하였다. 실험에서 DexNet-F는 DexNet에 비하여 정확도 (accuracy), F1 score 및 mAP 등 모든 성능지표가 개선되는 것을 볼 수 있다. 특히, 정확도와 Recall에서는 4.2%가 상승되는 것을 알 수 있다. 또한 mAP는 2.7% 상승하여 99.02%의 높은 성능을 보여주었다.

그림 2에서는 힘폐쇄 확률을 히트맵으로 시각화 하였다. YCB Object는 실생활에서 흔하게 볼 수 있는 물체들로 구성된 데이터 셋이다. YCB Object를 이용해 치약, 과자, 드릴, 통조림 등 다양한 물체들이 상자에 들어있는 상황을 MuJoCo 시뮬레이션을 통하여 구성하고 데이터를 수집하였다. 각 네트워크를 통해 수집한 데이터에서 힘폐쇄 확률을 예측하고 이를 히트맵으로 시각화하였다. 히트맵에서 힘폐쇄 확률이 높을수록 노란색을 갖게 된다. 중앙 그림처럼 DexNet은 높이가 낮고 평평한 물체와 바닥을 명확하게 구분하지 못하는 모습을 보였다 (빨간 점선). 또한, 물체의 경

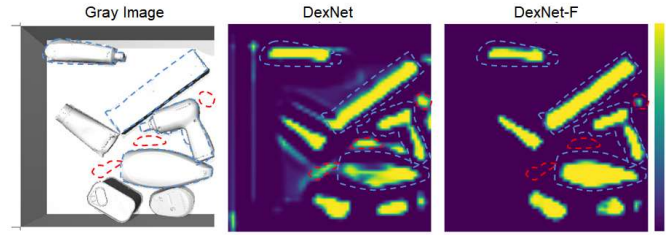


그림 2. 이미지 데이터 (좌)와 그에 따른 모델별 힘폐쇄 확률 (중앙, 우)

계면 (파란 점선)에서는 확률 값이 임계값 보다 높게 나타나는 모습을 보여주면서 일부 불안정한 파지점을 생성하는 모습을 보여줬다. 반면 오른쪽 그림처럼 논문에서 제안한 DexNet-F는 물체의 중심 또는 흡착 그리퍼가 파지하기 쉬운 지점의 확률을 높게 예측하고, 바닥 및 물체의 테두리와 각진 부분의 확률을 낮게 예측하는 결과를 보여주었다. 제안한 네트워크가 파지점 생성에서 에러를 줄이는 것을 확인할 수 있었다.

IV. 결론

본 논문에서는 흡착 그리퍼의 강체 파지점 생성을 위한 인공지능 네트워크인 DexNet-F를 제안했다. DexNet-F는 Gray 이미지와 Depth 이미지의 파라미터를 합치는 병합 모듈을 추가하여 성능을 향상시켰다. DexNet-F는 원본 이미지 (Raw Data)에서 파지점 후보군 (Grasp Candidate)을 바로 생성할 수 있기 때문에 실시간으로 데이터가 처리되어야 하는 환경에서 효율적으로 활용할 수 있는 장점을 가진다. 실험에서는 기존의 간결한 구조를 유지하면서도 보다 강인한 파지점을 생성하는 것을 보였다. 본 논문에서 사용한 흡착 그리퍼는 물체를 파지하는 측면에서 강력하지만 표면이 거칠거나 직물처럼 강체가 아닌 물체는 집을 수 없다는 한계를 갖는다. 향후 흡착 그리퍼 뿐 아니라 평행 그리퍼에서도 사용할 수 있는 네트워크를 개발하여 실생활에 활용도를 높이고자 한다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No.2020R1A2C2012146)

참고 문헌

- [1] Mousavian, Arsalan, Clemens Eppner, and Dieter Fox. "6-dof graspnet: Variational grasp generation for object manipulation." Proceedings of the IEEE/CVF international conference on computer vision. 2019.
- [2] Mahler, Jeffrey, et al. "Learning ambidextrous robot grasping policies." Science Robotics 4.26 (2019): eaau4984.
- [3] Mitash, Chaitanya, et al. "ARMBench: An object-centric benchmark dataset for robotic manipulation." 2023 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2023.
- [4] Calli, Berk, et al. "The ycb object and model set: Towards common benchmarks for manipulation research." 2015 international conference on advanced robotics (ICAR). IEEE, 2015.