행동 복제 강화학습 및 딥러닝 사물 부분 검출 기술에 기반한 사람형 로봇손의 사물 조작

오지헌 [1], 류가현 [1], 박나현 [1], Edwin Valarezo Añazco[1,3], Patricio Rivera Lopez[1], 원다슬 [2], 정진균 [2], 장윤정 [2], 김태성 [1,2]
[1] 경희대학교 전자정보대학 전자정보융합공학과
[2] 경희대학교 전자정보대학 생체의공학과
[3]Faculty of Engineering in Electricity and Computation, FIEC., Escuela Superior Politécnica del Litoral, ESPOL. Guayaquil, Ecuador.
{nhpark, dhwlgjs3, yugacandy, edgivala, patoalejor, dsrina, ynjngchang, tskim}@khu.ac.kr
wjdwlsrbs77@naver.com

Object Part Detection-based Manipulation with an Anthropomorphic Robot Hand Via Human Demonstration Augmented Deep Reinforcement Learning

Ji Heon Oh[1], Ga Hyun Ryu[1], Na Hyeon Park[1], Edwin Valarezo Añazco[1,3], Patricio Rivera Lopez[1], Da Seul Won[2], Jin Gyun Jeong[2], Yun Jung Chang[2], Tae-Seong Kim[1,2]
[1]Dept. of Electronics and Information Convergence Engineering,
[2]Dept. of Biomedical Engineering, College of Electronics and Information,
Kyung Hee University, Republic of Korea
[3]Faculty of Engineering in Electricity and Computation, FIEC., Escuela Superior Politécnica del Litoral, ESPOL. Guayaquil, Ecuador.

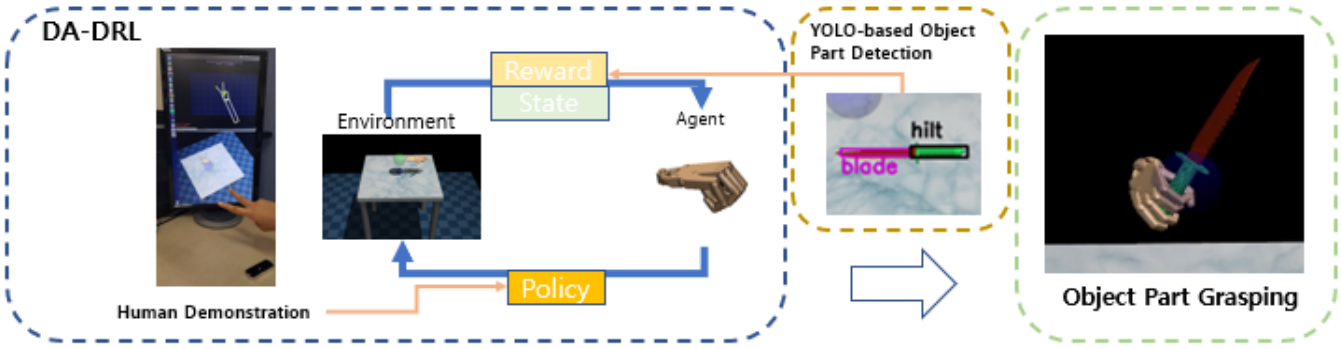
요 약

최근 사람형(Anthropomorphic)로봇손의 사물조작 지능을 개발하기 위하여 행동복제(Behavior Cloning) Deep Reinforcement Learning(DRL) 연구가 진행중이다. 자유도(Degree of Freedom, DOF)가 높은 사람형 로봇손의 학습 문제점을 개선하기 위하여, 행동 복제를 통한 Human Demonstration Augmented(DA)강화 학습을 통하여 사람처럼 사물을 조작하는 지능을 학습시킬 수 있다. 그러나 사물 조작에 있어, 의미 있는 파지를 위해서는 사물의 특정 부위를 인식하고 파지하는 방법이 필수적이다. 본 연구에서는 딥러닝 YOLO기술을 적용하여 사물의 특정 부위를 인식하고, DA-DRL을 적용하여, 사물의 특정 부분을 파지하는 딥러닝 학습 기술을 제안하고, 2종 사물(망치 및 칼)의 손잡이 부분을 인식하고 파지하여 검증한다. 본 연구에서 제안하는 학습방법은 사람과 상호작용하거나 도구를 용도에 맞게 사용해야하는 분야에서 유용할 것이다.

1. 서론

현재 제조업에서 사용되는 로봇손들은 그리퍼(Gripper)형태로 집게 모양의 손 혹은 공기를 빨아들이는 석션(Suction)형태를 띄고 있다[1]. 이와 같은 로봇손들은 조립라인 같은 익숙한 환경에서 반복적인 작업을 하는데 뛰어나다. 그러나 파지하는 사물이 제한된다는 점과 조작 응용 능력에서 떨어지는 모습을 보여준다. 로봇손의 활용분야가 확대됨에 따라 사람형(Anthropomorphic) 로봇손이 각광받고 있다. 사람형 로봇손에 대한 연구는 지난 수십년간 로봇손의 정교함을 높이는 하드웨어와 로봇의 움직임을 제어하는 소프트웨어 두가지 방향으로 진행되었다[1][2]. 그러나 사람형 로봇은 다관절에서 오는 움직임의 자유도(Degree of Freedom)가 많아서 학습이 어렵다.

이를 개선하기 위해 사람 데모를 사용하여 행동복제를 통한 Deep Reinforcement Learning(DRL)연구가 진행되었다. 사람 데모(human demonstration)를 통해 행동을 복제하여 로봇에 학습시키는 연구[4]가 소개되었고, 이에 기반한, Demonstration Augmented Policy Gradient(DAPG)정책 최적화 방법이 개발되었다[5].



(그림 1) 제안하는 딥러닝 YOLO 를 이용한 사물 부위 검출 및 사람형 로봇손의 사물 특정부위 파지 강화 학습 시스템

또한 사람의 사물 조작은 사물의 부위를 인식하여 이루어지는 반면[6,7], 대부분의 로봇손 사물 조작 연구는 사물의 부위를 이해하지 못하고, 중심부를 파지 하는 것을 목표로 하고 있다. 즉 사물 부분의 의미 정보를 추출하지 못하면, 의도된 용도의 사물 파지가 이루어지지 못한다[8]. 최근 연구에서 의미 있는 사물 파지를 위해서는 사물의 특정 부분을 인식하여 파지하는, 즉 사물의 용도를 이해하는 방식의 연구가 진행되고 있다[3,4].

본 논문에서는 의미 있는 사물의 파지를 위해, 딥러닝 YOLO[9]를 통해 사물의 부위를 인식하고, DAPG DRL 을 통하여 사물의 특정 부위를 조작하는 학습 시스템을 제안하고, 2 종의 사물에 대하여 특정 부위를 인식하고 파지하여 검증한다.

2. 방법

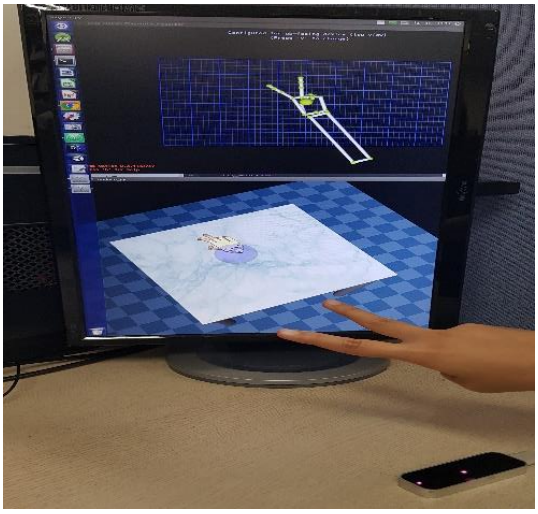
본 연구의 사물 부분 인식 및 사물 조작 강화학습 지능 시스템을 그림 1 에 도시하였다. 주어진 사물의 RGB 영상에서 사물의 부위를 검출하고 DAPG 를 통하여 사물의 특정 부위를 파지하고 조작하는 강화학습을 수행한다.

A. 시뮬레이션 환경

사람형 로봇손과 사물은 가상 시뮬레이션 환경인 Multi-Joint dynamics with Contact(Mujoco)[10]를 통해 구현했다. 구현한 Adroit 로봇손은[11] 30 개의 자유도를 가지고 있다. Adroit 로봇손은 손 과 팔의 회전과 관련된 8 개의 자유도를 가지고, 엄지와 새끼손가락에 각각 5 개의 자유도를 가지며 나머지 손가락은 각각 4 개의 자유도를 가지고 있다. 본 연구에서는 이 로봇손을 학습시켜 칼과 망치에 대한 사물 부분 조작 작업을 확인한다.

B. 사람 데모 생성

사람의 손동작 인식에 Leap Motion LM-010 카메라[12]를 사용하였다. 그림 2 는 Leap Motion 으로 인식한 손동작을 Mujoco 내 Adroid 로봇 손에 손동작을 매핑한 결과를 보여준다.



(그림 2) Leap Motion 센서를 이용한 손동작 인식 및 Mujoco Adroit 로봇손의 행동 복제 환경.

C. Object Part Detection via YOLO

사물의 부분 인식을 위해 Darknet 기반인 YOLOv3[13]를 사용하였다. YOLO 는 Unifiied Detection 방법으로 24 개의 convolution 레이어와 1 개의 fully connected 레이어로 구성되었다. 사물을 구분하는 confidence 를 파악하기 위해 S*S 그리드로 이미지를 나눈 후 각 셀마다 B 개의 bounding box 를 만들어 어떤 사물이 있는지 확률적으로 나타낸다. Bounding box 는 x, y, width, height, confidence 로 구성되며 x, y 를 제외한 3 가지 변수를 조절하며 특정 클래스의 bounding box 를 만드는 것이 최종 목적이다. Bounding box 와 사물이 얼마나 겹쳐 있는지를 Intersection over union(IOU)로 계산한다. IOU 와 사물의 확률 분포를 통해 confidence 는 식(1)로 계산한다. 최종적인 클래

스 확인을 위해 조건부 확률과 식(2)를 통해 클래스를 판별한다.

$$confidence = probability(object) * IOU(truth, predict) \quad (1)$$

$$classclassificationconfidencescore = \mathrm{P}\left(\frac{\mathrm{class}}{\mathrm{object}}\right) * \mathrm{p(object)} * \mathrm{IOU(truth, predict)} \quad (2)$$

본 연구에서는 2종의 사물로 망치와 칼을 선정하여 사물 모델을 생성했다. 망치는 헤드와 손잡이로 구분, 칼은 blade와 손잡이로 구분하여 사물 당 310장의 학습데이터를 생성하여 YOLO 학습을 진행했다.

그림 3은 칼과 망치의 부위 인식 결과이다. 그림 3-(a)은 hammer_handle과 hammer_head 부위의 인식 결과이고, 그림 3-(b)의 blade 및 hilt 부위의 인식 결과이다.

(a) (b)

(그림 3) YOLO를 통해 (a) 망치와 (b)칼 부위 검출 결과

D. DAPG

DRL의 정책 최적화 방법인 Natural Policy Gradient(NPG)는 기존 Policy Gradient(PG)의 non-covariant한 문제를 해결하기 위해 Natural gradient를 추가해 steepest descent direction으로 학습이 진행되도록 설계되었다[14]. 하지만 NPG만 사용하는 경우에는 많은 변수로 인하여, 장시간 학습시간이 소요되고, 사람과 유사한 동작으로 사물을 조작하지 못한다.

이런 문제를 해결하기 위해 NPG에 사람 데모를 추가한 DAPG 방법이 개발되었다[5]. (3)을 통하여, 정책 최적화가 이루어지고, 식 (4)를 통하여 사람 데모를 정책에 복제한다. 행동 복제과정에서 사람 데모의 정보를 통해 사물로 접근하는 경로와 손 모양 값들로 가이드 정책을 만든다.

$$g_{aug} = \sum_{(s,a)\in\rho_\pi} \nabla_\theta ln\pi_\theta(a|s)A^\pi(s,a) + \sum_{(s,a)\in\rho_D} \nabla_\theta ln\pi_\theta(a|s)w(s,a) \quad (3)$$

$$w(s,a) = \lambda_0\lambda_1^k \max_{(s',a')\in\rho_\pi} A^\pi(s',a')\forall(s,a)\rho_D \quad (4)$$

g_{aug}는 augmented gradient이다. $\sum_{(s,a)\in\rho_\pi} \nabla_\theta ln\pi_\theta(a|s)A^\pi(s,a)$는 평균 보상의 gradient를 나타낸다. π_θ는 정책을 나타내며, $A^\pi(s,a)$는 s에 따른 리워드 함수이다.

w(s,a)는 가중치 함수이다. K는 iteration 횟수이며 $\lambda_0 = 1.0, \lambda_1 = 0.95$이다. K가 커짐에 따라서 사람 데모에 대한 가중치가 작아진다.

보상의 정도를 통해 정책을 갱신하는 과정에서는 YOLO를 통해 얻은 파지 부분을 잡았을 때 높은 보상을 얻도록 해 로봇 손이 도구 사용법에 맞는 부분을 잡도록 학습하였다.

3. 결과

A. 망치 조작

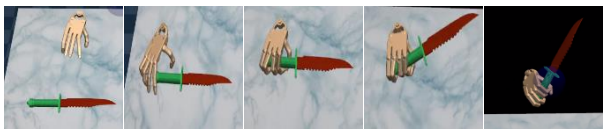
그림 5에서 손이 망치의 손잡이로 이동하여 파지 한 후, 조작하는 모습을 확인할 수 있다.

(그림 5) DAPG를 통해 학습한 로봇손이 시간에 따라 망치를 파지하는 과정

그림 7은 보상을 도시한 그래프로 초록색은 망치의 보상을 나타낸다. 망치의 보상을 보면 500번 반복한 지점부터 최종 결과와 비슷한 정도의 보상을 얻었다. 최종 최적화된 정책으로 94%의 조작 성공률을 획득하였다.

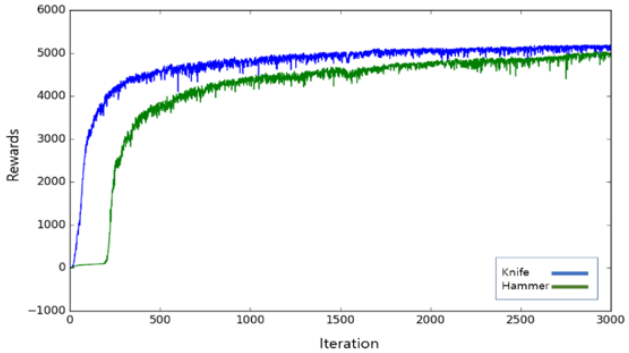
B. 칼 조작

그림 6에서 손이 칼의 손잡이로 이동하여 파지한 후, 조작하는 모습을 확인할 수 있다.



(그림 6) DAPG를 통해 학습한 로봇손이 시간순서에 따라 칼을 조작하는 과정

그림 7에서 파란색이 칼의 조작 보상을 나타낸다. 칼 조작은 250번 반복 후부터 최종 결과와 비슷한 보상을 얻은 것을 확인할 수 있다. 최종 최적화된 정책으로 100%의 조작 성공률을 획득하였다.



(그림 7) 망치와 칼 조작 보상 그래프

4. 결론

본 연구에서는 딥러닝 YOLO를 이용하여 칼과 망치의 부위를 인식하고, 사람형 로봇손을 DAPG DRL을 통해 학습시켜, 도구의 용도에 맞게 파지하고 조작하는 지능을 개발하였다. 제시한 학습법은 로봇 손이 사람이 사용하는 도구를 써야하는 분야에서 용도에 맞게 도구를 사용하는 로봇에 이용될 수 있다.

ACKNOWLEDGEMENT

이 논문은 2019년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(2019R1A2C1003713). 이 논문은 2020 년도 정부(미래창조과학부)의 재원으로 한국연구재단 -현장맞춤형 이공계 인재양성 지원사업의 지원을 받아 수행된 연구임(No. 2017H1D8A1031522). 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 디지털콘텐츠원천기술개발사업의 연구결과로 수행되었음 (IITP-2017-0-00655).

참고문헌

[1] Richard Hodson, “A gripping problem,” Nature, Vol557, pp.23-25, 2018.

[2] Aude Billard, Danica Kragie, “Trends and challenges in robot manipulation,” Science, Vol 364, Issue 6446, 2019.

[3] Nikolaus Vahrenkamp, Leonard Westkamp, Natsuki Tamanobe, Eren E.Aksoy, and Tamim Asfour, “Part-based Grasp Planning for Familiar Objects,” IEEE-RAS 16th International Conference on Humanoid Robots(Humanoids) Cancun, Mexico, pp.919-925, 2016.

[4] T.Osa, Jan Peters and G.Neumann, “Hierarchical reinforcement learning of multiple grasping strategies with human instructions,” Advanced Robotics, Vol32 ,No18 , pp.955-968,2018.

[5] Aravind Rajeswaran, Vikash Kumar, Abhishek Gupta, Giulia Vezzani, John Schulman, Emanuel Todorov, Sergey Levine. “Learning Complex Dexterous Manipulation with Deep Reinforcement Learning and Demonstrations,” Robotics: Science and Systems (RSS), 2018, arXiv:1709.10087v2.

[6] I.Biederman, “Recognition-by-components: a theory of human image understanding,” Psychological review, Vol94, No 2, pp.115-147, 1987

[7] D.D.Hoffman and W.A. Richards, “Parts of recognition,” Cognition, Vol 18, No 1, pp.65-94,1984

[8] S.H.Creem and D.R.Proffitt, “Grasping objects by their handles: a necessary interaction between cognition and action.,” Journal of Experimental Psychology: Human Perception and Performance, Vol 27, No 1 , pp.218 -228, 2001

[9] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” 2015, arXiv 1506 02640.

[10] E. Todorov, T. Erez and Y. Tassa, "MuJoCo: A physics engine for model-based control," 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, Vilamoura , 2012, pp.5026-5033.

[11] V. Kumar, Z. Xu and E. Todorov, "Fast, strong and compliant pneumatic actuation for dexterous tendon-driven hands," IEEE International Conference on Robotics and Automation, Karlsruhe, 2013, pp.1512-1519.

[12] https://www.ultraleap.com/product/leap-motion-controller/

[13] Joseph Redmon ,Ali Farhadi , “YOLOv3: An Incremental Improvement,” 2018

[14] S.Kakade, “A natural policy gradient,” Proceedings of the 14th International Conference on Neural information Processing Systems: Natural and Synthetic , Vancouver, 2001, pp.1531-1538