

다중 RGB Depth 카메라 및 합성곱 신경망을 이용한 6축 매니퓰레이터의 파지 알고리즘에 관한 연구

감지웅*, 옥승호**

A Study on the Grasping Algorithm of 6-axis Manipulator using Multiple RGB Depth Cameras and Convolutional Neural Networks

Ji-Woong Kam*, Seung-Ho Ok**

이 성과는 2020년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임
(NRF-2018R1C1B5043706)

요 약

본 논문에서는 두 개의 RGB depth 카메라 및 합성곱 신경망(convolutional neural network: CNN)을 이용한 6축 매니퓰레이터의 파지 동작 알고리즘을 제안한다. 제안하는 알고리즘은 작업 공간에 위치한 RGB depth 카메라를 통해 물체의 추론 및 위치를 탐색하고, 매니퓰레이터의 말단 장치(end-effector)에 장착된 RGB depth 카메라를 통해 물체를 파악한다. 이후 물체의 회전 각도와 방향을 계산하여 파지 위치를 탐색 및 집기-놓기(pick-and-place) 동작을 수행한다. 실험 결과 제안하는 알고리즘은 사전에 학습된 사각형의 상자 및 원통형의 물체가 검출되었을 경우 물체의 형상에 따른 파지 동작 및 pick-and-place 동작이 정상적으로 구동되는 것을 확인하였다.

Abstract

In this paper, we propose a gripping algorithm for a 6-axis manipulator using two RGB depth cameras and a convolutional neural network (CNN). The proposed algorithm infers and locates an object through an RGB depth camera located in the working space, and then grasps the object through an RGB depth camera mounted on the manipulator's end-effector. Then, by calculating the rotation angle and direction of the object, a gripping position is searched and a pick-and-place operation is performed. As a result of the experiments, the proposed algorithm showed that the gripping and pick-and-place motion were normally performed according to the shape of the object when a previously learned rectangular box and cylindrical object were detected.

Keywords

visual servoing, CNN, manipulator, pick-and-place, image processing

* 동의대학교 로봇·자동화공학전공(학사)
- ORCID: <https://orcid.org/0000-0001-6726-6207>
** 동의대학교 로봇·자동화공학전공 조교수
(교신저자)
- ORCID: <https://orcid.org/0000-0002-9036-0872>

• Received: Feb. 25, 2021, Revised: Mar. 21, 2021, Accepted: Mar. 24, 2021
• Corresponding Author: Seung-Ho Ok
Dept. of Robot-Automation Engineering, Dong-eui University, Korea,
Tel.: +82-51-890-2265, Email: osh@deu.ac.kr

I. 서 론

비주얼 서보잉(Visual servoing)은 로봇의 작업 공간 또는 매니퓰레이터의 말단 장치(End-effector)에 장착된 비전 센서를 통해 입력된 영상을 처리함으로써 로봇이 지능화된 동작을 수행할 수 있도록 하는 기술이다. 비주얼 서보잉 시스템은 크게 위치기반 비주얼 서보 제어(PBVS, Position-Based Visual Servoing)와 영상기반 비주얼 서보잉 제어(IBVS, Image-Based Visual Servoing)로 구분할 수 있다 [1][2]. 최근에는 딥러닝을 활용한 비정형 환경에서 PBVS 및 IBVS를 융합한 시스템이 연구되고 있다 [3]. 매니퓰레이터의 비주얼 서보잉 시스템은 카메라의 종류 및 설치 구성에 따라 단안 카메라를 이용한 eye-to-hand 및 eye-in-hand[4], 스테레오 카메라 [5]를 이용한 eye-to-hand 및 eye-in-hand, 그리고 이를 융합한 시스템이 있다[6]. Eye-to-hand 방식의 경우 카메라는 로봇의 작업 공간에 설치되며 물체 인식 및 추적 시 카메라의 시야(FOV, Field Of View) 각에 제한을 받는다. Eye-in-hand 방식의 경우 카메라는 로봇의 말단 장치에 설치되며, 매니퓰레이터의 동작 중에도 물체를 인식 및 추적하기 위해서는 높은 프레임 속도(FPS, Frames Per Second)의 영상처리가 필요하다. 이러한 문제를 해결하기 위해 [7]의 연구에서는 eye-to-hand 및 eye-in-hand 방식을 융합하여 작업 성능과 정확성을 개선하였다.

본 논문에서는 두 개의 RGB depth 카메라를 사용하여, eye-to-hand에서는 합성곱 신경망(CNN, Convolutional Neural Network)을 통해 물체를 추론하고, eye-in-hand에서는 RGB depth 정보를 기반으로 PBVS 및 IBVS를 융합하여 물체의 회전 정도와 방향을 계산 후 최종적으로 파지 위치를 탐색하는 알고리즘을 제안한다. 본 논문에서는 pick-and-place 동작을 구현하기 위해 기존의 연구내용[8]에서 파지 위치 탐색을 위한 영상처리 기법 및 실험 내용을 추가 확장하였다.

본 논문의 2장에서는 RGB depth 카메라와 비주얼 서보잉 시스템에 관해 설명하며, 3장에서는 제안하는 eye-to-hand 및 eye-in-hand 기반 융합 알고리즘을 설명한다. 이후 4장에서는 본 논문에서 제안하는

다중 RGB depth 카메라 및 딥러닝을 이용한 매니퓰레이터 동작 제어 알고리즘의 성능에 대해 평가하고, 5장에서는 본 논문의 결론과 향후 연구과제에 대해 논한다.

II. 비주얼 서보잉

2.1 비주얼 서보잉 시스템

비주얼 서보잉 방식 중 위치기반 비주얼 서보잉은 로봇의 작업 공간에 설치된 카메라 및 작업 대상 물체의 상대적인 위치 오차를 계산 후 로봇을 제어하는 방법이며, 물체의 정확한 위치 정보를 획득하기 위해 카메라의 교정(Calibration)이 이루어져야 한다. 따라서 카메라의 교정 이후에 오차가 누적되면 정확성이 낮아지게 된다. 이와 반대로 영상기반 비주얼 서보잉 방식은 작업 대상 물체를 카메라 영상 평면 중심에 위치하도록 로봇을 제어하므로, 위치기반 비주얼 서보잉 방법과 달리 상대적으로 교정에 큰 영향을 받지 않는 특징이 있다. 따라서 주로 위치기반 비주얼 서보잉 및 영상 기반 비주얼 서보잉을 융합한 시스템이 사용되고 있다[9][10].

2.2 제안하는 시스템 구성

그림 1은 작업 공간 내 임의의 위치에 놓은 물체를 추론 및 파지하기 위한 시스템 구성을 나타내며, 그림 2는 제안하는 시스템 동작 흐름도를 나타낸다. 그림 1(a)에서 전체 시스템은 6축 수직 다관절 로봇과 이를 제어하기 위한 제어기, 물체를 추론 및 파지를 위한 두 개의 RGB depth 카메라, 그리고 영상처리를 통한 알고리즘 수행 및 로봇 제어기와의 통신을 위한 단일 보드 컴퓨터(SBC, Single Board Computer)로 구성되어 있다.

그림 1(b)의 eye-to-hand에서 검출된 이미지는 SBC로 전송되며, SBC는 물체를 분류하기 위해 CNN을 사용한 추론과 위치 검출을 수행 후 제어기와의 TCP/IP 통신을 통해 매니퓰레이터의 말단 장치를 물체의 상단으로 이동시킨다. 이후 그림 1(c)의 eye-in-hand에서는 물체의 정확한 파지를 위해 그

리퍼(Gripper)의 회전 각도 및 회전 방향을 외곽선 탐색 기법을 통해 계산 후 pick-and-place 동작을 수행한다.

III. 제안하는 알고리즘

3.1 제안하는 eye-to-hand 알고리즘

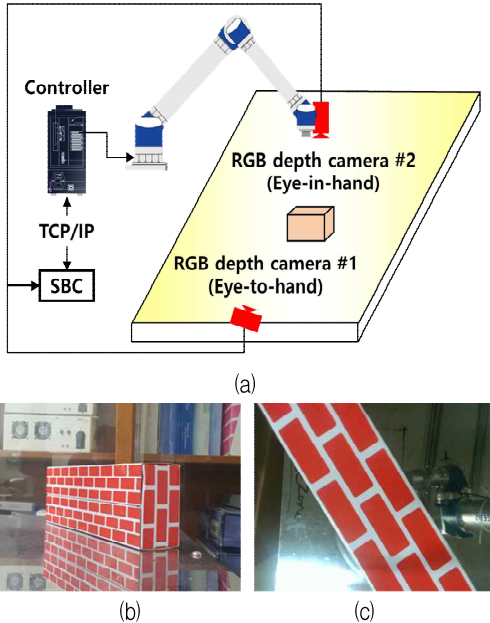


그림 1. 시스템 구성

Fig. 1. System configuration, (a) System configuration, (b) An image from the RGB depth camera #1, (c) An image from the RGB depth camera #2

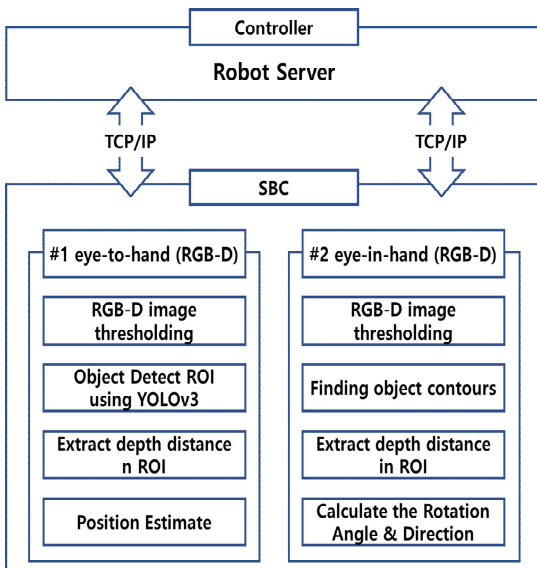


그림 2. 시스템 동작 흐름도
Fig. 2. System operation flowchart

본 논문에서 사용된 eye-to-hand 카메라는 그림 1(a) 및 그림 3과 같이 로봇의 작업 공간에 고정되며, 매니플레이터의 역기구학을 위한 월드 좌표계의 기준점이 된다. Eye-to-hand 카메라의 경우 그림 4와 같이 RGB depth 카메라에서 입력된 영상으로부터 작업 대상 물체를 인식 후 관심 영역(ROI, Region Of Interest)을 추출한다.

이를 위해 그림 4(a)에서 파지가 불가능한 영역의 물체 검출을 사전에 방지하기 위해 깊이 맵(Depth map)을 사용한 문턱 값(Thresholding) 처리를 통해 4(b)와 같은 영상을 추출한다. 배경이 제거된 그림 4(b)에서 사전에 학습된 물체를 추론하기 위해 합성곱 신경망 YOLOv3 모델을 사용하여 그림 4(c)와 같이 학습된 물체의 관심 영역을 추출한다. 그림 4(c)까지의 과정을 통해 획득한 물체의 3차원 거리 정보를 획득하기 위해 그림 4(c)에서 검출된 물체의 좌표를 그림 4(d)의 깊이 영상에 적용하여 그림 4(e)와 같이 물체의 ROI 및 거리 정보를 획득한다.

Eye-to-hand 카메라에서는 그림 4(a)의 영상에서 식 (1)을 적용하여 그림 4(b)와 같이 물체 검출 영역을 제한함으로써 매니플레이터가 이동 불가능한 영역으로의 이동을 방지한다. 여기서 D 는 물체와 RGB depth 카메라 사이의 거릿값이며 $MinD$ 는 RGB depth 카메라의 최소 검출 거리, $MaxD$ 는 최대 검출 거리를 나타낸다.

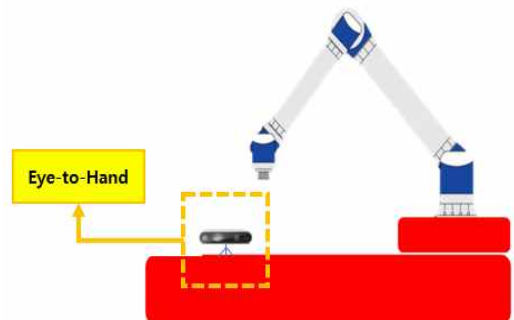


그림 3. Eye-to-hand 카메라 위치
Fig. 3. Position of eye-to-hand camera

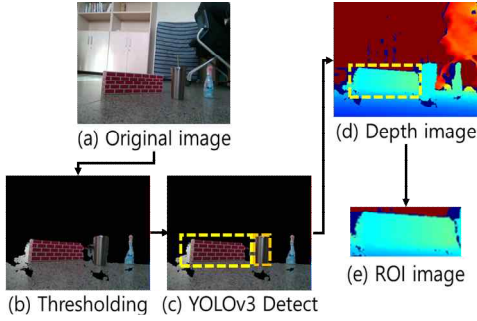


그림 4. Eye-to-hand ROI 검출 과정 예시
 Fig. 4. Example of eye-to-hand ROI detection process

Eye-to-hand 카메라에서는 매니플레이터의 말단 장치를 물체의 상단부로 이동시키기 위한 x, y축 좌표 벡터를 계산한다. 이를 위해 그림 4(c) ~ 4(e) 과정을 통해 검출된 물체의 ROI 및 거릿값을 사용하며, 식 (2)에서는 y축 좌표 벡터 그리고 식 (3)에서는 x축 좌표 벡터를 획득한다. 여기서 식 (2)의 stereo_Dscale의 경우 RGB depth 카메라의 깊이 맵을 거릿값으로 변환하기 위한 상숫값이며, 식 (3)의 $x_1, y_1, distance_{y_1}$ 의 경우 그림 4(a) 원본 영상 중심의 x, y 좌표와 거리 정보, 그리고 $x_2, y_2, distance_{y_2}$ 의 경우 그림 4(e)에서 CNN을 통해 획득한 ROI 중점의 x, y 좌표와 거리 정보이다.

$$Binarization = \begin{cases} 1 & MinD \leq D < MaxD \\ 0 & Otherwise \end{cases} \quad (1)$$

$$distance_y = (stereo_Dscale \times ROI_pixel) \quad (2)$$

$$distance_x = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (distance_{y_1} - distance_{y_2})^2} \quad (3)$$

3.2 제안하는 eye-in-hand 알고리즘

Eye-in-hand 카메라는 그림 5와 같이 매니플레이터의 그리퍼에 고정되며, 카메라에 인식된 물체를 파지하기 위한 회전 각도를 계산한다. 그림 6 및 7은 eye-in-hand 카메라로부터 입력된 영상 및 깊이 맵을 사용하여 물체의 윤곽선(Contour)을 검출 후 회전 각을 계산하고, 그리퍼를 회전한 결과를 나타

낸다. 이 과정에서 RGB depth 카메라를 통해 획득한 depth image를 그림 6(b)와 같이 깊이 값을 RGB 값으로 표현 후 빛에 의한 간섭 완화 및 바닥과 물체의 분리를 위해 RGB를 HSV로 변환한다. 그리고 Hue 영역과 Value 영역의 정보를 기준으로 문턱값 처리를 통해 그림 6(c)와 같이 이진화를 수행한다.

이진화된 이미지를 기반으로 물체의 회전 각도 및 방향을 계산하기 위한 외곽선을 그림 6(d)와 같이 탐색하고 검출된 물체의 각 꼭짓점 좌표를 기반으로 그리퍼 회전 각도를 식 (4)를 통해 계산한다. 식 (4)에서 box의 가로 index의 경우 그림 6(d)에서 검출된 box의 최소경계 사각형의 꼭짓점 좌표이며, 꼭짓점의 x축 좌표값은 가로 index의 0번에 y축 좌표값은 가로 index의 1번에 할당된다. 식 (4)에서 box의 세로 index의 경우 가로 index의 1번에 할당된 y축 좌표값이 가장 큰 값이 세로 index의 0번에 할당되며 그림 6(d)의 최소경계 사각형 꼭짓점에서 반시계방향 순서대로 할당된다.

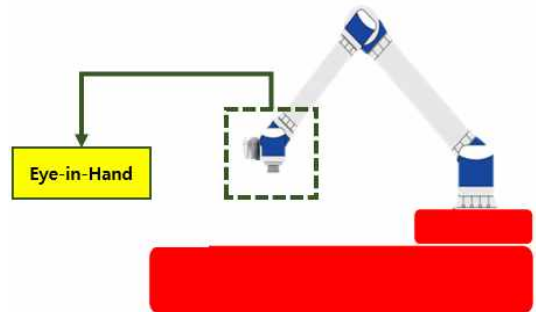


그림 5. Eye-in-hand 카메라 위치
 Fig. 5. Position of eye-in-hand camera

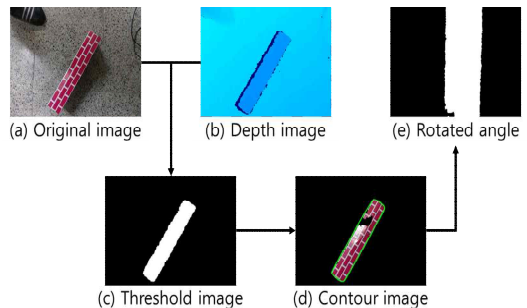


그림 6. Eye-in-hand 각도 검출 과정 예시
 Fig. 6. Example of eye-in-hand angle detection process

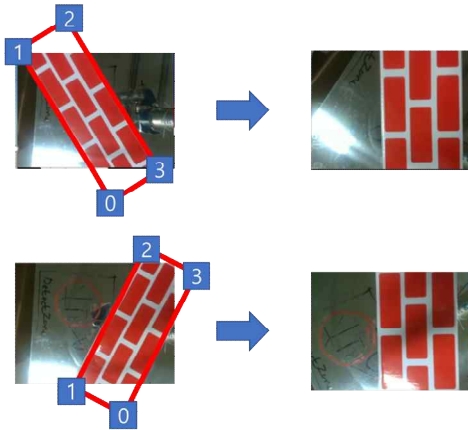


그림 7. 회전각 및 방향 적용 결과
Fig. 7. Rotation angle and direction application result

즉, 그림 7에서 $box[i][j]$ 가 $i=0, j=0$ 인 경우 y 축 좌푯값이 가장 큰 위치의 x 축 좌푯값이며 $i=3, j=1$ 경우 $i=0$ 의 위치에서 반 시계 방향으로 3번째 순서인 꼭짓점의 y 축의 좌푯값이다. 최종적으로 획득한 회전 각도와 방향을 로봇 제어기로 전송함으로써 매니플레이터는 pick-and-place 동작을 수행한다.

$$\theta_1 = \tan^{-1} \left(\frac{\begin{bmatrix} 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 3 \\ 1 \end{bmatrix}}{\begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 3 \\ 0 \end{bmatrix}} \right) \quad (4)$$

IV. 매니플레이터 동작 제어 실험 결과

4.1 성능 평가 환경 및 방법

본 논문에서 제안하는 시스템은 영상처리를 위해 OpenCV와 Intel의 RealSense 라이브러린 librealsense를 사용하였으며 매니플레이터는 그림 8과 같이 $\pm 0.02mm$ 의 반복 정밀도를 가진 ZEUS사의 ZERO(ZRA-0515P)를 사용하였다. 실험에서는 각 물체에 대해 25번의 pick-and-place 실험을 진행하였으며 표 1은 실험 환경에 사용된 SBC의 H/W 및 S/W 사양을 나타낸다.

물체를 추론하여 ROI 기법을 적용하기 위한 네트워크로는 YOLOv3를 사용하였으며, 학습 모델의 경우 darknet-53 그리고 액티베이션 함수로 leaky ReLu를 사용하였다. 그림 9는 YOLOv3 학습 결과로 검출된 영상을 나타내며 본 실험에서는 평행이동

그리퍼에서 파지할 수 있는 박스와 컵으로 실험 대상을 한정하여 학습 및 실험을 진행하였다. 학습데이터의 경우 박스와 컵 두 개의 클래스를 각 클래스당 원본 이미지 100장에서 밝기, 노이즈, 회전, 색상 변화를 주는 어그멘테이션(Augmentation)을 수행하였으며 학습 및 검증을 위해 데이터셋을 9:1 비율로 나눠 진행했다.

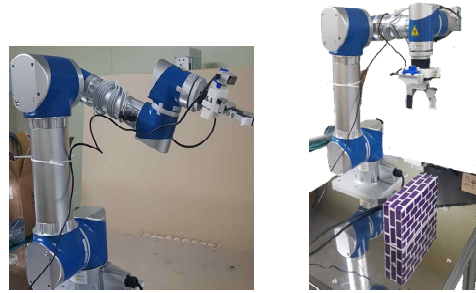


그림 8. 실험 환경
Fig. 8. Experimental environment

표 1. 실험 환경
Table 1. Experimental environment

H/W	CPU	8-Core ARM v8.2 64-Bit CPU, 8 MB L2 + 4 MB L3
	GPU	512-Core Volta GPU with tensor cores
	Memory	32 GB 256-Bit LPDDR4x
	#1, 2 CAM	Intel Realsense D435i
	Manipulator	6-DOF ZEUS-ZERO
S/W	CNN model	YOLOv3

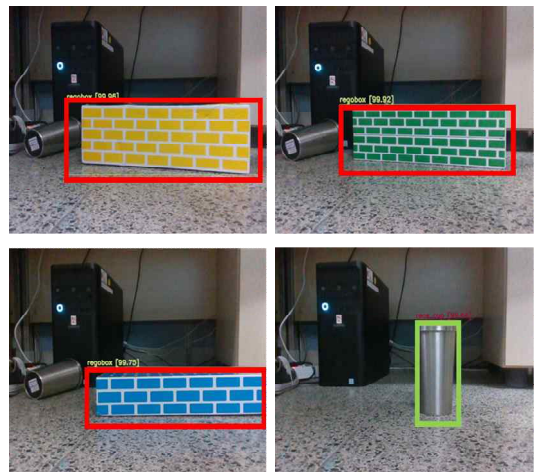


그림 9. YOLOv3 검출 결과
Fig. 9. Results of YOLOv3 detection

학습 결과 2,700회 반복 학습에서 평균손실 0.057%을 나타냈으며 객체 검출 정확도의 경우 박스에서 99.8% 컵에서 99.64%의 결과를 확인하였다.

Eye-to-hand 카메라에서는 CNN을 사용하여 물체를 추론 후 위치기반 비주얼 서보잉을 통해 물체의 위치를 계산하였다. 그리고 eye-in-hand 카메라에서는 영상기반 비주얼 서보잉을 통해 물체의 회전 각을 계산 후 파지를 수행하였다.

4.2 실험 결과

표 2는 박스 및 컵에 대한 pick-and-place 실험 결과를 나타내며, 제안하는 eye-to-hand 및 eye-in-hand 알고리즘의 각 동작 성공률 및 전체 pick-and-place 동작 성공률을 나타낸다. 실험 결과 로봇의 작업 환경 내에 물체가 놓일 경우 학습된 물체에 따라 pick-and-place 동작이 정상 수행되는 것을 확인하였다. 즉, 박스가 검출될 경우 박스의 회전 각을 계산 후 그리퍼를 회전하고, 컵이 검출되면 회전 동작은 수행하지 않고 pick-and-place 동작을 정상 수행하는 것을 확인하였다.

표 2. Pick-and-place 실험 결과
Table 2. Pick-and-place experimental results

Object Step	Rotation angle of the object (°)					
	0			30		
	Box experiment success rate(%)			Cup experiment success rate(%)		
eye-to-hand	100	96	96	100	100	100
eye-in-hand	100	100	100	100	100	100
Total	100	96	96	100	100	100

실험 중 eye-to-hand 알고리즘에서 발생한 한 번의 파지 실패의 경우 실험에서 사용된 평행이동 그리퍼의 폭이 최대 7.5cm에서 최소 5.5cm로 이동 반경이 2cm로 한정되어 가로 폭이 7cm인 박스를 파지하기 위해 박스 상단부 중점으로부터 그리퍼의 최대 폭 사이의 오차범위 2.5mm 이내의 거릿값 측정이 요구된다. 하지만 박스가 30°이상 회전된 상태에서 RGB depth의 거릿값 측정 시 빛에 의한 표면 반사 및 노이즈로 인해 박스의 상단부의 중점에서

2.5mm 이상 벗어나는 거릿값이 측정된 것으로 분석된다. 이를 해결하기 위해 향후 최대 폭 및 이동 반경이 더 넓은 그리퍼로 교체 또는 빛에 의한 외란을 제거하기 위한 알고리즘의 수정 보완으로 성능 개선이 가능할 것으로 사료된다.

V. 결론 및 향후 과제

본 논문에서는 eye-to-hand 카메라에서 합성곱 신경망을 사용하여 물체를 추론 후 깊이 맵을 통해 물체의 위치를 획득하고, eye-in-hand 카메라에서 RGB 영상 및 깊이 맵을 사용하여 물체의 윤곽선 및 회전 각을 계산 후 최종적인 파지 위치를 탐색하는 비주얼 서보잉 알고리즘을 제안하였다. 제안하는 알고리즘은 물체를 인식하기 위한 딥러닝 네트워크로 YOLOv3를 사용함으로써 물체의 모양에 따라 서로 다른 파지 동작 수행이 가능하다. 즉, eye-in-hand에서는 물체의 모양에 따라 이를 파지하기 위한 회전 각도를 계산하고 이를 통해 6축 매니퓰레이터의 말단 장치를 제어함으로써 물체의 모양에 따른 파지 및 pick-and-place 동작을 수행한다.

제안하는 알고리즘은 6축 수직 다관절 로봇 및 평행이동 그리퍼를 사용하여 실험을 진행하였으며 그리퍼에서 파지할 수 있는 박스와 컵으로 한정하여 학습 및 실험을 진행하였다. 실험 결과, 학습된 물체가 검출되었을 때 pick-and-place 동작이 정상 수행되는 것을 확인하였으며, 향후 평행 이동 그리퍼 이외에 다양한 그리퍼 및 이를 통해 파지 가능한 대상 물체의 종류를 확대하여 pick-and-place 동작의 구현 및 모바일 로봇에 적용을 위한 연구를 진행할 계획이다.

References

[1] B. Espiau, F. Chaumette, and P. Rives, "A new approach to visual servoing in robotics", IEEE Transactions on Robotics and Automation, Vol. 8, No. 3, pp. 313-326, Jun. 1992.
[2] S. Hutchinson, G. Hager, and P. Corke, "A tutorial on visual servo control", IEEE Transactions of

- Robotics and Automation, Vol. 12, No. 5, pp. 651-670, Oct. 1996. <https://doi.org/10.1109/70.538972>
- [3] J. Liu and Y. Li, "An Image Based Visual Servo Approach with Deep Learning for Robotic Manipulation", The 6th Int. Workshop on Advanced Computational Intelligence and Intelligent Informatics, Chengdu, China, pp. 61-66, Nov. 2019.
- [4] J. S. Yoon, C. H. Kwak, Y. H. Kim, W. H. Kim, and S. H. Ok, "Implementation of a Table Tennis Robot Using Monocular Camera-based 3D Object Recognition Method and Path Planning for 6-Axis Manipulator", The Journal of Korean Institute of Information Technology, Vol. 16, No. 8, pp. 35-44, Aug. 2018. <https://doi.org/10.14801/jkiit.2018.16.8.35>
- [5] I. Jee, "A Study on the Generation and Processing of Depth Map for Multi-resolution Image Using Belief Propagation Algorithm", The Journal of Institute of Internet, Broadcasting and Communication, Vol. 15, No. 6, pp. 201-208, Dec. 2015. <https://doi.org/10.7236/JIIBC.2015.15.6.201>
- [6] D. Kragic and H. Christensen, "A Framework for Visual Servoing", International Conference on Computer Vision Systems, ICVS, Graz, Austria, Vol. 2626, pp. 345-354, Apr. 2003.
- [7] C. Yang, S. Amarjyoti, X. Wang, Z. Li, H. Ma and C. Y. Su, "Visual servoing control of Baxter robot arms with obstacle avoidance using kinematic redundancy", International Conference on Intelligent Robotics and Applications, Portsmouth, UK, Vol. 9244, pp. 568-580, Aug. 2015.
- [8] J. W. Kam and S. H. Ok, "A Study on Manipulator Motion Control Algorithm Using Multiple RGB depth Cameras and Deep-learning", Proceedings of KIIT Conference, Cheongju, Korea, pp. 522-524, Oct. 2020.
- [9] D. H. Kim and M. J. Chung, "A Switched Visual Servoing Technique Robust to Camera Calibration Errors for Reaching the Desired Location Following a Straight Line in 3-D Space", The Journal of Korea Robotics Society, Vol. 1, No. 2, pp. 125-134, Jan. 2006.
- [10] N. R. Gans and S. A. Hutchinson, "Stable Visual Servoing Through Hybrid Switched-System Control", IEEE Transactions on Robotics, Vol. 23, No. 3, pp. 530-540, Jun. 2007. <https://doi.org/10.1109/TRO.2007.895067>

저자소개

감 지 웅 (Ji-Woong Kam)



2021년 3월 현재 :
동의대학교(학사)
로봇·자동화공학전공
관심분야 : 로봇비전, 비주얼
서보잉, 임베디드 제어

옥 승 호 (Seung-Ho Ok)



2008년 2월 : 경북대학교
전자공학과(공학석사)
2014년 2월 : 경북대학교
전자공학부(공학박사)
2014년 3월 ~ 2017년 8월 :
삼성전자 책임연구원
2017년 9월 ~ 현재 : 동의대학교

로봇·자동화공학전공 조교수
관심분야 : 로봇비전, SoC, VLSI