

혼재된 환경에서의 효율적 로봇 파지를 위한 3차원 물체 인식 알고리즘 개발

Development of an Efficient 3D Object Recognition Algorithm for Robotic Grasping in Cluttered Environments

송 동 운¹ · 이 재 봉² · 이 승 준[†]

Dongwoon Song¹, Jae-Bong Yi², Seung-Joon Yi[†]

Abstract: 3D object detection pipelines often incorporate RGB-based object detection methods such as YOLO, which detects the object classes and bounding boxes from the RGB image. However, in complex environments where objects are heavily cluttered, bounding box approaches may show degraded performance due to the overlapping bounding boxes. Mask based methods such as Mask R-CNN can handle such situation better thanks to their detailed object masks, but they require much longer time for data preparation compared to bounding box-based approaches. In this paper, we present a 3D object recognition pipeline which uses either the YOLO or Mask R-CNN real-time object detection algorithm, K-nearest clustering algorithm, mask reduction algorithm and finally Principal Component Analysis (PCA) algorithm to efficiently detect 3D poses of objects in a complex environment. Furthermore, we also present an improved YOLO based 3D object detection algorithm that uses a prioritized heightmap clustering algorithm to handle overlapping bounding boxes. The suggested algorithms have successfully been used at the Artificial-Intelligence Robot Challenge (ARC) 2021 competition with excellent results.

Keywords: 3D Object Recognition, Cluttered Environment, YOLO, Mask R-CNN

1. 서 론

최근 협동 로봇과 서비스 로봇들의 폭넓은 보급과 기계학습 분야의 급속한 발전으로 인하여^{1,2}, 사전에 지시된 동작만

을 반복적으로 수행하는 것이 아니라 센서를 사용하여 물체를 인식하고 지능적으로 물체를 조작할 수 있도록 하게 하는 연구들이 활발히 진행되고 있다^{3,4}. 이러한 지능적 물체 조작을 위해서는 물체를 인식하고 그 3차원 위치를 추정하는 능력이 필수적으로 요구된다. 이를 위해서는 RGB 이미지로부터 물

Received : May. 24. 2022; Revised : Jun. 26. 2022; Accepted : Jul. 6. 2022

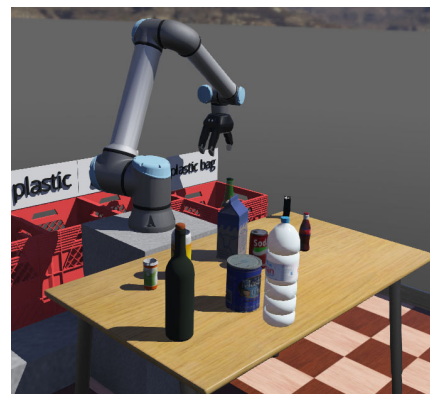
※ This work was partially supported by IITP grant funded by the Korea government (MSIT) (No. 2021-0-01202) and KIAT grant funded by the Korea Government (MOTIE) (P0008473, HRD Program for Industrial Innovation)

This project was funded by Police-Lab 2.0 Program (www.kipot.or.kr) funded by the Ministry of Science and ICT(MSIT, Korea) & Korean National Police Agency (KNPA, Korea) (No. 082021D48000000) and This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2021-0-01202)

1. MS Student, Pusan National University, Busan, Korea (dongwoon@pusan.ac.kr)

2. Ph.D. Student, Pusan National University, Busan, Korea (niteofhunter@pusan.ac.kr)

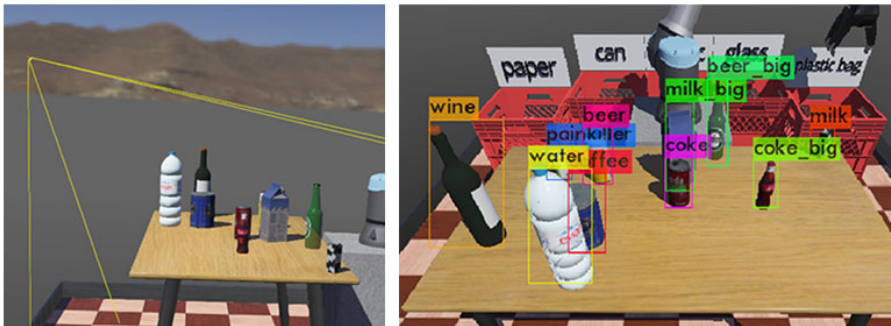
† Associate Professor, Corresponding author: Pusan National University, Busan, Korea (seungjoon.yi@pusan.ac.kr)



[Fig. 1] ARC 2021 Simulation environment



[Fig. 2] YOLO based 3D Object Recognition using Toyota HSR Home Service Robot



[Fig. 3] Camera setup and YOLO result of the ARC 2021 Environment

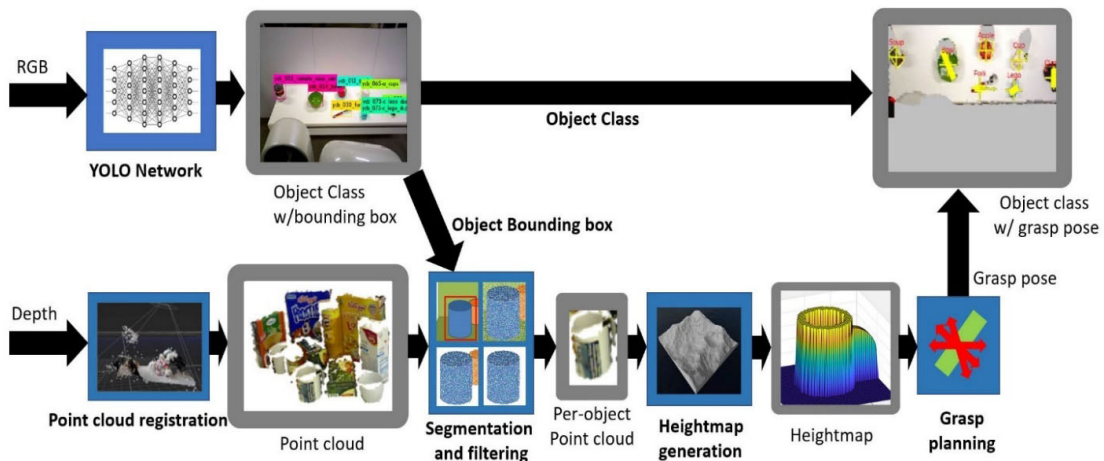
체를 우선 인식하고^[5] 깊이 이미지를 사용하여 인식된 물체들의 3차원 위치 추정을 하는 방법이 일반적이다. 이 경우 물체 인식을 위해서는 데이터 라벨링이 간편하고 다양한 환경에서 우수한 인식 성능을 보이는 YOLO 알고리즘이 폭넓게 쓰이고 있다. 특히 [Fig. 1]과 같이 제한된 시간에 다수의 물체를 학습하여야 하는 로봇 대회 중의 경우 대부분의 출전 팀들이 YOLO 기반 알고리즘을 사용하고 있으며^[6-8], 본 연구팀 또한 [Fig. 2]에서 볼 수 있듯이 YOLO 기반 알고리즘을 사용하여 다수의 로봇 대회에서 우수한 결과를 내어 온 바 있다^[9].

하지만 YOLO 알고리즘은 물체의 바운딩 박스만을 출력하기 때문에 혼재된 상황에서 물체들 간의 구분에 한계가 있다. 2021년 개최된 국제 로봇 컨테스트 인공지능 로봇 챌린지 2021 (ARC 2021) 대회의 경우 [Fig. 3]에서 볼 수 있듯이 이러한 YOLO 알고리즘의 단점들이 크게 부각되는 상황으로, 각각 카메라로 인한 왜곡과 다수의 높이가 높은 물체들의 사용으로 바운딩 박스의 중첩이 매우 많이 일어나 기존 YOLO 기반 3차원 위치 추정 알고리즘의 성능이 크게 저하되게 된다.

본 논문에서는 기존에 사용하였던 YOLO 기반 물체 인식 및 추정 알고리즘을 소개하고, 혼재된 상황에서 물체들의 종류와 3차원 위치를 정확히 추정하기 위하여 픽셀 단위의 Mask

를 출력하는 물체 인식 알고리즘인 Mask R-CNN을 사용한 3차원 물체 인식 알고리즘을 소개한다. 마지막으로 YOLO의 빠른 라벨링 효율을 유지하면서 혼재된 환경에 대응할 수 있도록 하기 위하여, 깊이 이미지 기반 클러스터링과 우선 순위 기반 클러스터 처리 알고리즘을 도입한 개선된 YOLO 기반 물체 인식 및 추정 알고리즘을 제안한다. 개발된 알고리즘들은 시뮬레이션 환경과 실제 로봇을 사용하여 성능이 검증되었고, 실제 ARC 대회에서 사용되어 우수한 결과를 보였다.

본 논문의 구성은 다음과 같다. 2장에서는 기존의 YOLO 기반 3D 물체 인식 및 위치 추정 알고리즘을 소개하고 이를 ARC 2021 환경에 적용 시의 문제점에 관하여 서술한다. 3장에서는 Mask R-CNN을 RGB 인식용으로 사용한 알고리즘을 제시하고, YOLO 기반 알고리즘과 ARC 2021 환경에서 비교한 내용에 대해 서술한다. 4장에서는 혼재된 환경을 위하여 개선한 YOLO 기반 3D 물체 인식 알고리즘을 소개하고, 5장에서는 물체들과 충돌 없이 로봇팔을 사용하여 물체를 파지하게 하는 파지 및 경로 계획 알고리즘에 관해 설명한다. 6장에서는 ARC 2021 환경에서의 알고리즘 성능 비교 결과를 제시하고, 7장에서는 실제 로봇을 사용한 알고리즘 성능 비교 결과를 제시한다. 마지막 8장에서는 결론을 서술한다.



[Fig. 4] YOLO based 3D Object Detection Pipeline

2. YOLO 기반 물체 인식 및 위치 추정

본 연구팀은 YOLO 알고리즘을 이용한 [Fig. 4]와 같은 3차원 물체 인식 및 위치 추정 알고리즘을 개발, 사용하여 왔다. 여기서는 이 알고리즘에 대해 간단히 소개하고, 이를 ARC 환경에 적용시 문제점에 대하여 서술한다.

2.1 2D Object Detection

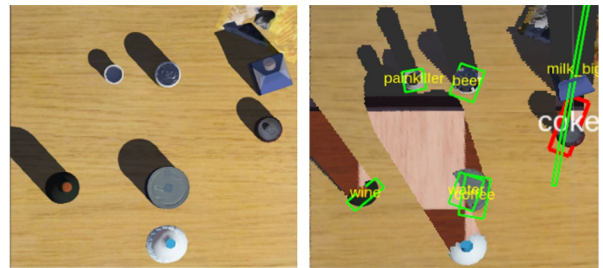
첫 단계로 YOLO 알고리즘을 사용하여 RGB 이미지로부터 물체들의 종류와 바운딩 박스를 얻어낸다. 물체 인식 성능을 향상시키기 위하여 초기 데이터를 사용한 학습 후 인식이 떨어지는 위치들을 찾아 수동으로 추가 데이터를 생성한 후 학습을 반복하는 방법을 사용하였다.

2.2 Point Cloud Segmentation and filtering

물체들의 종류와 바운딩 박스를 얻어낸 후에는 대응하는 깊이 이미지를 각 물체별 바운딩 박스로 자른 뒤 3차원에 투영하여 물체별 포인트 클라우드를 생성한다. 여기에는 주위의 벽이나 근처의 다른 물체들이 포함되어 있을 수 있기 때문에 RANSAC 알고리즘으로 얻은 평면 정보를 사용하여 벽에 해당하는 점들을 제거하고, K-Nearest Neighbor (KNN) clustering을 적용하여 [10,11] 바운딩 박스의 중심에 가장 가까운 클러스터만을 남긴다.

2.3 3D Pose estimation

물체를 파지하기 위해서는 물체의 위치와 자세를 추정해야 한다. 필터링된 물체별 포인트 클라우드에 주성분 분석(PCA)를 적용하여 해당 물체의 장축과 단축을 구하여 물체의 위치와 2.5D 자세를 추정하게 된다.



[Fig. 5] Performance issue of basic YOLO based algorithm in cluttered ARC environment

2.4 Grasp planning

복잡한 형태의 물체를 정확하게 파지하기 위하여, 물체별 포인트 클라우드로부터 물체의 2D 높이맵을 생성 후 다양한 파지 위치와 각도를 탐색하여 충돌 없이 물체를 파지할 수 있는 최적의 파지 방법을 찾는다[12].

2.5 로봇 대회에서의 YOLO 기반 인식 알고리즘 활용

본 연구팀은 이 알고리즘을 사용하여 대규모 국제 서비스 로봇 대회인 RoboCup@Home과 World Robot Summit에 출전하여 2021년 양 대회 모두 우승하는 등 매우 우수한 성과를 내어 왔다. 또한 2018년의 RoboCup@Home과 World Robot Summit 대회 우승팀인 일본의 Hibikino-Musashi@Home^[13]과 2019년 RoboCup@Home 우승팀인 네덜란드의 Tech United Eindhoven팀^[14]도 YOLO 알고리즘 기반의 3차원 물체 인식 및 위치 추정 알고리즘을 사용하였다.

2.6 ARC 2021 환경에 알고리즘 적용시의 문제점

화각이 좁은 카메라를 사용하여 작은 물체를 인식하는 기준의 로봇 대회들과 달리 ARC 환경에서는 높은 물체들과 광각카

메라의 영향으로 [Fig. 3]에서 볼 수 있듯이 바운딩 박스의 중첩이 다수 발생하고, 그 결과 기존 알고리즘을 적용시 [Fig. 5]와 같이 인접한 물체들간의 간섭이 발생하여 위치 추정 성능이 크게 떨어지는 문제가 발생하였다.

3. Mask R-CNN 기반 물체 인식 및 위치 추정

혼재된 환경에서 물체들의 바운딩 박스가 중첩되는 YOLO 알고리즘의 단점을 해결하기 위해, [Fig. 6]와 같이 직사각형의 바운딩 박스가 아닌 픽셀 단위의 Mask 형식으로 각 물체의 정보를 얻을 수 있는 Mask R-CNN 알고리즘을 사용하여 물체들이 중첩되고 기울어진 상황에서도 정확한 위치 추정값을 얻어내도록 하였다.

3.1 Mask R-CNN을 활용한 물체 인식

[Fig. 4]에 제시된 파이프라인에서 RGB 인식 부분에서 YOLO 대신 Mask R-CNN을 대체하여 사용한다. 사각형의 바운딩 박스를 출력하는 YOLO와 달리 Mask R-CNN은 이미지의 모든 픽셀에 대하여 클래스 값을 출력하기 때문에, YOLO 기반 알고리즘에서 사용한 벽 정보 필터링 과정 없이 물체별 포인트 클라우드를 바로 생성할 수 있다.

3.2 중첩된 물체 경계면의 마스크 축소

Mask R-CNN은 물체의 외형에 매우 근접한 마스크 값을 출력하지만, 물체가 이미지 상에서 중첩된 경우 마스크의 경계와 실

제 물체의 경계가 수 픽셀 정도의 오차가 발생할 수 있다. 이러한 경우 마스크를 사용하여 포인트 클라우드를 생성시 값이 픽셀이 잘못 분류되어 위치 추정에 오류가 생길 수 있다. 이 문제를 해결하기 위하여, 인접한 픽셀이 서로 다른 마스크를 갖는 경우 해당 부분의 마스크들을 [Fig. 7]과 같이 축소시켜 인접부분의 오인식을 방지하였다. 결과적으로 물체들이 기울어지고 중첩된 상황에서도 각 물체들의 정확한 종류와 위치를 얻을 수 있었다.

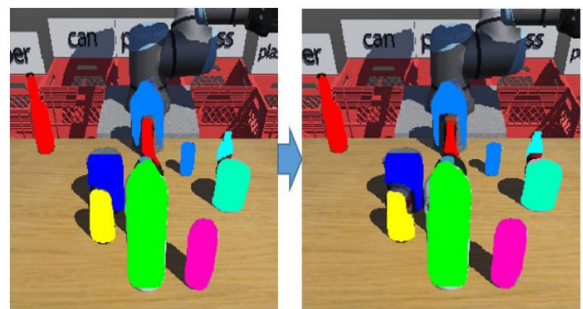
3.3 YOLO 알고리즘과의 비교

기존 YOLO 기반 알고리즘의 경우 로봇 대회 환경에서는 우수한 성능을 보여왔다. 하지만, [Fig. 3]과 같이 광각 카메라를 사용하고 물체의 높이가 높은 ARC 2021 환경의 경우 물체가 RGB 화면에서 기울어지고 겹치게 되어 결과적으로 출력되는 물체들의 바운딩 박스의 중첩이 다수 발생하였다. 이 경우 물체의 추정 위치가 [Fig. 5]에서 볼 수 있듯이 주위 물체들의 영향을 받아 오인식되는 오류가 다수 발생하였다. 반면 Mask R-CNN을 사용한 방법의 경우 [Fig. 6]에서 볼 수 있듯 각 물체의 마스크를 픽셀 별로 중첩 없이 얻을 수 있다. 따라서 혼재된 환경에서도 [Fig. 8]처럼 각 물체들의 3차원 위치를 오류 없이 정확하게 추정할 수 있었다.

단 Mask R-CNN은 학습 데이터 라벨링시 바운딩 박스 정보만 입력하면 되는 YOLO와 달리 물체 외부의 정확한 형태 정보를 라벨링하여야 하기 때문에, Mask R-CNN 기반 알고리즘은 YOLO 기반 알고리즘에 비해 학습 데이터 생성에 훨씬 오랜 시



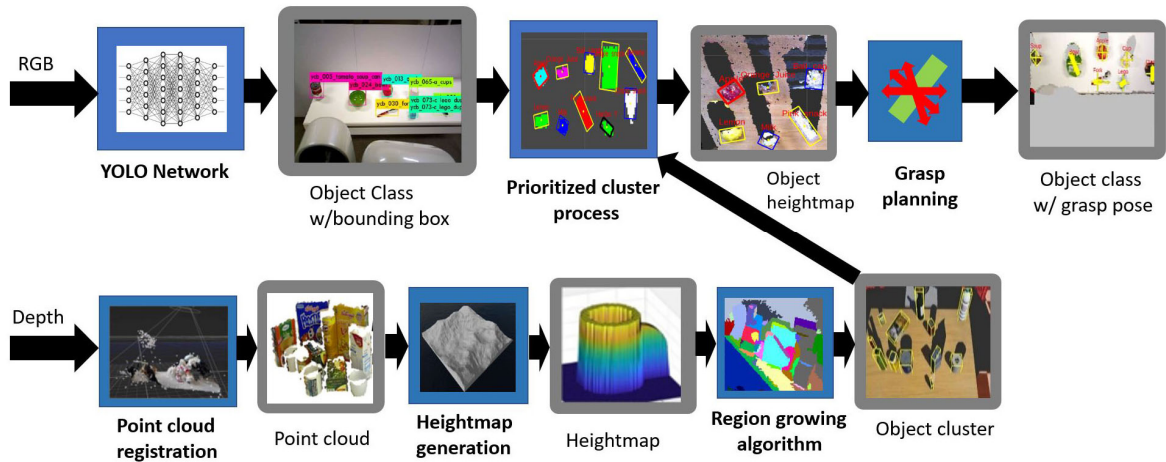
[Fig. 6] Comparison of YOLO and Mask R-CNN results



[Fig. 7] Mask shrinking process



[Fig. 8] Result of the Mask R-CNN based algorithm in cluttered ARC environment



[Fig. 9] Improved YOLO based 3D object detection algorithm

간이 걸리는 모습을 보여 단시간에 다수의 물체에 대한 빠른 학습이 요구되는 로봇 대회와 같은 환경에서는 불리한 모습을 보였다.

4. 개선된 YOLO 기반 물체 인식 및 위치 추정

3장에서 제안한 Mask R-CNN 기반 물체 인식 및 위치 추정 알고리즘은 혼재된 환경에서 발생하는 기존 YOLO 기반 3D 물체 인식 알고리즘의 단점을 대부분 극복 가능하였으나, 트레이닝 데이터를 제작하는 시간이 매우 오래 걸려 단시간에 다수의 물체를 학습하여야 하는 로봇 대회 환경과 같은 경우 적용이 어려운 문제가 있었다.

이러한 두 알고리즘의 단점들을 극복하기 위하여, [Fig. 9]와 같이 데이터 제작 효율성이 높은 YOLO를 RGB 인식 알고리즘으로 사용하면서 추가적으로 3D 위치 추정 알고리즘을 개선하여 혼재된 환경에서 오류 없이 물체 위치를 추정할 수 있도록 하였다.

4.1 깊이 이미지 기반 물체 클러스터링

[Fig. 4]에 제시된 기존 3D 물체 인식 파이프라인의 경우 RGB 이미지로부터 각 물체를 우선 인식 후 그 바운딩 박스 정보를 기반으로 깊이 이미지를 분할하여 물체별 포인트클라우드를 생성하게 된다. 따라서 RGB 물체 인식 알고리즘의 결과

에 따라 물체별 포인트 클라우드에 여러 물체가 들어갈 수 있게 되어, 물체가 중첩된 경우 오인식이 발생할 수 있게 된다. 이를 방지하기 위하여 [Algorithm 1]과 같이 우선 깊이 이미지만을 사용하여 포인트 클라우드를 만들고, 이를 바탕으로 2D 높이맵(Height map)을 생성한다. 생성된 2D 높이맵으로부터 각 물체들을 구분하기 위하여 영역 확대 알고리즘(Region growing algorithm)^[15,16]을 사용하여 높이맵을 이어진 구획들로 분할하여 물체별 클러스터를 생성한다. 이후 각 클러스터에 식 (1), (2)와 같이 주성분 분석(PCA)^[17]를 적용하여 각 클러스터의 위치 및 방향을 추정한다.

$$\hat{X}_k = X - \sum_{s=1}^{k-1} X w_{(s)} w_{(s)}^T \quad (1)$$

$$w_k = \underset{\|w\|=1}{\operatorname{arg\,max}} \left\{ \|\hat{X}_k w\|^2 \right\} = \underset{\|w\|=1}{\operatorname{arg\,max}} \left\{ \frac{w^T \hat{X}_k^T \hat{X}_k w}{w^T w} \right\} \quad (2)$$

4.2 우선 순위 기반 높이맵 처리

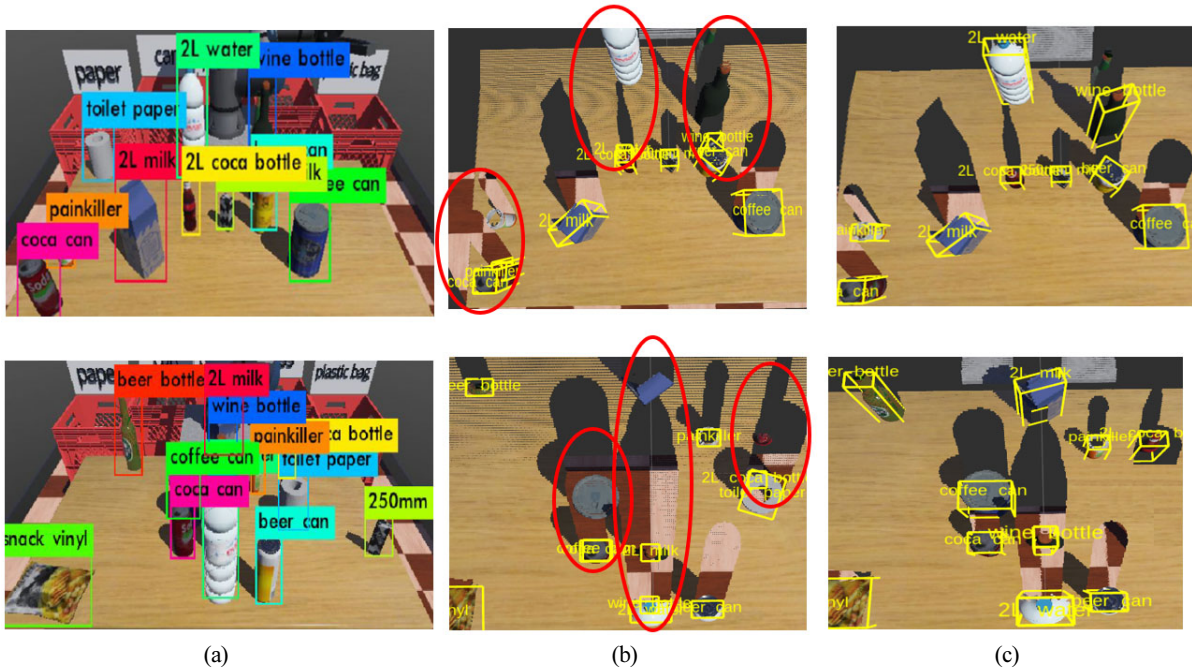
깊이 이미지만을 사용하여 높이맵을 생성하고 물체별 클러스터로 분할한 후, YOLO 정보를 사용하여 각 클러스터에 해당하는 물체를 결정한다. 바운딩 박스가 중첩되는 경우를 해결하기 위하여, 현재 미처리된 물체들의 바운딩 박스들 중

[Algorithm 1] Depth image-based object clustering

- 1 3D_point_cloud ← ProjectTo3D(depth image)
- 2 2D_height_map ← GenHeightMap(3D_point_cloud)
- 3 while any nonzero cell in 2D_height_map do
- 4 start_cell ← closest 2D cell in 2D_height_map
- 5 cluster[i] ← Region_growing(2D_height_map, start_cell)
- 6 cluster_stats[i] ← PCA(cluster[i])
- 7 delete_cells(2D_height_map, cluster[i])

[Algorithm 2] Priority-based height map processing

- 1 yolo_bbox, yolo_class ← YOLO(rgb_image)
- 2 while any bbox in yolo_bbox do
- 3 bbox_idx ← find_lowest_bbox_idx(yolo_bbox)
- 4 cluster_idx = match_cluster(clusters, yolo_bbox[bbox_idx])
- 5 cluster_class[cluster_idx] ← yolo_class[bbox_idx]
- 6 remove_bbox(yolo_bbox, bbox_idx)
- 7 remove_cluster(clusters, cluster_idx)



[Fig. 10] Comparison of previous YOLO based algorithm and improved YOLO based algorithm in ARC environment

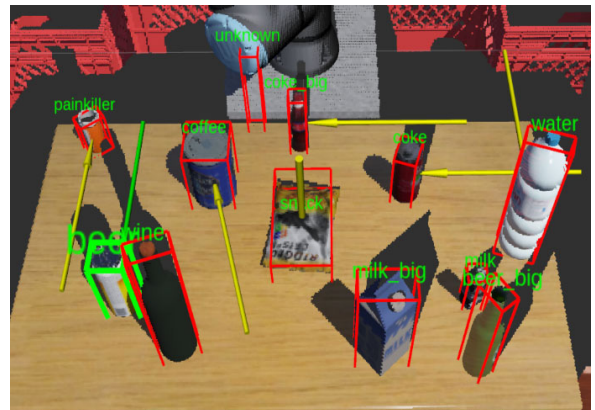
메라로부터 가장 가까운 바운딩 박스를 우선적으로 뽑고, 그 바운딩 박스 내부의 점군과 가장 유사도가 높은 클러스터를 찾아 해당 물체의 클래스를 부여한다. 클래스가 부여된 후에는 중첩을 막기 위하여 해당 물체에 해당하는 높이맵 클러스터를 삭제한다. 이 과정을 모든 물체들의 클래스를 얻어낼 때까지 반복하고, 남은 클러스터는 미확인으로 분류한다. 이 과정은 [Algorithm 2]에 정리되어 있다.

4.3 YOLO 기반 알고리즘의 개선 결과

YOLO 기반 3D 비전 알고리즘의 단점을 극복하기 위해 깊이 기반 물체 클러스터링과 우선 순위 기반 높이맵 처리를 도입한 결과, [Fig. 10]와 같이 물체들이 중첩되거나 광각 왜곡으로 바운딩 박스가 크게 출력되어 기존 알고리즘이 물체들의 위치를 오인식하는 경우에도 물체들의 올바른 위치를 오류 없이 강인하게 인식하는 모습을 보였다.

5. 파지 및 팔 경로 계획

ARC 2021 대회 물체들은 대부분 물병과 같이 수직으로 긴 형태이기 때문에, 수직이 아닌 수평 방향으로 파지를 할 것이 요구된다. 이 경우 로봇 팔의 손목 부분이 타 물체와 충돌할 수 있고, 로봇팔의 작업 공간 제한으로 인하여 물체에 접근할 수 있는 방약에 제약이 존재하기 때문에, 로봇 팔의 작업 공간 하에서 충돌을 회피하며 물체를 파지할 수 있는 파지 및 팔 경로



[Fig. 11] Grasp planning for ARC 2021 environment

계획이 요구된다.

이를 위하여 각 물체별로 일정 각도로 파지 방향을 샘플링한 후, 주위 물체와의 안전 거리, 로봇의 작업 공간 한계와 카메라에서 가까운 정도를 기준으로 각 파지 정책을 평가 후 그 중 가장 우수한 계획을 선택하여 파지를 실행한다^[18]. ARC 2021 환경에서 파지 계획을 수행한 예는 [Fig. 11]에 제시되어 있으며, 가능한 파지 계획들은 노란색 화살표, 가장 우수한 파지 계획이 초록색 화살표로 표시되어 있다.

6. ARC 2021 환경에서의 실험 결과

개발한 알고리즘들 중 개선된 YOLO 기반 3D 물체 인식 알

[Table 1] Specifications of PCs used

	CPU	GPU
Learning	Intel i9-9900K	RTX 3090
Test 1	AMD 5800X	RTX 3080
Test 2	Intel i9-9900KF	RTX 2080 Ti
Test 3	AMD 2600X	GTX 1660 super
Competition	Intel i7-8750H	RTX 2080

고리즘과 Mask R-CNN 알고리즘을 각각 사용하여 ARC 2021 대회에 2개의 팀으로 참가하였다. YOLO 학습을 위해서 총 720장의 라벨링 된 이미지를 만들어 사용하였고, 총 라벨링 시간에는 약 360분이 소요되었다. Mask R-CNN 학습을 위해서는 시간 관계상 총 327장의 라벨링 된 이미지를 만들어 사용하였고, 라벨링에는 약 1000분이 소요되었다. 모델의 학습은 [Table 1]와 같이 3090 GPU가 장착된 학습용 PC를 사용하여 두 알고리즘 모두 약 12시간 진행하였고, 대회에서는 i7-8750H CPU와 2080 GPU가 장착된 동일한 노트북을 사용하였다. RGB와 깊이 데이터가 주어졌을 경우, YOLO 기반 알고리즘은 대회용 노트북에서 평균 40ms, Mask R-CNN 기반 알고리즘은 평균 250ms의 시간이 소요되었다.

알고리즘의 성능 테스트를 위하여 ARC 2021 환경에서 서로 성능이 다른 3대의 테스트용 PC를 사용하여 총 200가지의 랜덤 시드를 사용하여 12개의 물체를 랜덤 배치한 뒤 인식율을 테스트한 결과는 [Table 2]와 같다. 기존의 YOLO 기반 알고리즘의 경우 바운딩 박스 중첩으로 인하여 인식율이 58.31%에 머물렀으나, Mask R-CNN 기반 알고리즘과 개선된 YOLO 알고리즘은 동일 조건에서 97% 이상의 우수한 인식율을 보였다. 특히 개선된 YOLO 알고리즘은 Mask R-CNN 보다 오히려 0.2% 정도 우수한 성능을 보였는데, 6배에 달하는 라벨링 속도 차이로 인하여 2배 많은 트레이닝 데이터를 준비할 수 있었기 때문이라고 생각된다.

인식한 물체에 대한 파지는 시뮬레이션 환경 특성상 100%에 가까운 성공율을 보였고, 실제 ARC 2021 대회에서도 참가한 양 팀 모두 주어진 물체들을 전혀 오류 없이 완벽하게 인식, 파지 및 분류하는 우수한 성과를 보였다.



[Fig. 12] Test setup using HSR home service robot

[Table 3] Result in real environment

	Recognition Rate (%)		
	6 items	9 items	12 items
YOLO	97.34	94.94	78.35
Improved-YOLO	100	98.98	97.76

7. 실제 환경에서의 실험 결과

실제 환경에서의 알고리즘의 성능 검증을 위해 RoboCup@Home 2022 국제대회의 Grocery Task 환경에서 기존의 YOLO 기반 3차원 물체인식 알고리즘과 개선된 YOLO 기반 3차원 물체인식 알고리즘의 성능을 비교하였다. 실험 환경은 [Fig. 12]와 같이 가로 120 cm, 세로 70 cm, 높이 72 cm의 테이블에 각각 6개, 9개, 12개의 서로 다른 물체들을 랜덤하게 배치한 뒤, RoboCup@Home에 사용되는 Toyota HSR 서비스 로봇을 사용하여 RGBD 이미지를 취득하여 물체를 인식하였다.

3가지 상황에서의 실험 결과는 [Table 3]에 정리하였다. 카메라의 화각이 ARC 환경대비 좁고 물체의 높이가 낮기 때문에 물체의 수가 적을 경우 기존 YOLO 기반 알고리즘도 ARC 환경보다는 상대적으로 우수한 결과를 내었으나, 물체 수가 증가하여 중첩이 발생하는 12개 케이스의 경우 기존 YOLO 알고리즘은 78.35%의 인식율을 보인 반면 개선된 YOLO 알고리즘은 97.76%의 높은 인식율을 유지하는 결과를 보였다. 물체 12개에서 양 알고리즘의 인식 결과의 예는 [Fig. 13]에 제시되어 있다.

[Table 2] Performance comparison of three algorithms in the ARC 2021 environment

	No. of Training images	Labeling Time (min)	Training Time (min)	Training Time/Image (sec)	Recognition Rate (%)	Processing Time (ms)	Grasping Success Rate (%)
YOLO	720	360	720	30	> 72	40	
Improved-YOLO	720	360	720	30	> 95	40	
Mask R-CNN	327	1,000	720	183.4	> 95	250	



[Fig. 13] Comparison of previous YOLO based algorithm and improved YOLO based algorithm in real environment

8. 결 론

본 연구는 ARC 2021 대회 환경과 같이 물체들이 중첩되어 있고 원근 왜곡이 존재하는 혼재된 환경에서 기존 바운딩 박스 기반 3D 물체 인식 알고리즘의 성능 개선을 다루고 있다. 바운딩 박스의 중첩으로 인한 성능 저하 문제는 Mask R-CNN 과 같은 마스크 기반 알고리즘을 적용하여 해결이 가능하나, Mask R-CNN의 경우 YOLO 기반 알고리즘에 비하여 인식 시간과 라벨링 작업에 소요되는 시간이 길어 다량의 데이터를 빠르게 학습하기에는 적합하지 않다.

이를 해결하기 위해 본 논문에서 제안한 개선된 YOLO 기반 3D 물체 인식 알고리즘은 시뮬레이션 환경을 사용한 ARC 2021 국제 로봇 대회에서 사용되어 모든 물체를 오류 없이 완벽하게 인식하는 우수한 성과를 보였고, Toyota HSR 서비스 로봇을 사용한 실제 환경에서도 혼재된 상황에서의 인식율을 크게 향상시키는 것이 검증되었다.

제안된 알고리즘은 기존 YOLO 기반 알고리즘의 빠른 데이터 준비 시간과 연산 시간의 장점을 그대로 가지고 있으면서 혼재된 환경에서도 Mask R-CNN 기반 알고리즘과 동급의 우수한 성능을 보여 앞으로 다양한 분야에서 폭넓게 이용될 수 있을 것으로 기대된다.

References

- [1] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *arXiv:2004.10934*, 2020, DOI: 10.48550/arXiv.2004.10934.
- [2] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017, DOI: 10.1109/ICCV.2017.322.
- [3] Y. Yoo, C.-Y. Lee, and B.-T. Zhang, "Multimodal Anomaly Detection based on Deep Auto-Encoder for Object Slip Perception of Mobile Manipulation Robots," *2021 IEEE International Conference on Robotics and Automation (ICRA)*, Xi'an, China, 2021, DOI: 10.1109/ICRA48506.2021.9561586.
- [4] W. Kehl, F. Manhardt, F. Tombari, S. Ilic, and N. Navab, "SSD-6D: Making RGB-Based 3D Detection and 6D Pose Estimation Great Again," *arXiv:1711.10006*, 2017, DOI: 10.48550/arXiv.1711.10006.
- [5] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common Objects in Context," *European Conference on Computer Vision*, pp. 740-755, 2014, DOI: 10.1007/978-3-319-10602-1_48.
- [6] D. Müller, N. Y. Wettengel, and D. Paulus, "homer@UniKoblenz: Winning Team of the RoboCup Virtual @Home Open Platform League 2021," *RoboCup 2021: Robot World Cup XXIV*, 2022, DOI: 10.1007/978-3-030-98682-7_23.
- [7] M. F. B. van der Burgh, J. J. M. Lunenburg, R. P. W. Appeldoorn, L. L. A. M. van Beek, J. Geijsberts, L. G. L. Janssen, P. van Dooren, H. W. A. M. van Rooy, A. Aggarwal, S. Aleksandrov, K. Dang, A. T. Hofkamp, D. van Dinther, and M. J. G. van de Molengraft, "Tech United Eindhoven @Home2019 Champions Paper," *RoboCup 2019: Robot World Cup XXIII*, 2019, DOI: 10.1007/978-3-030-35699-6_43.
- [8] S. G. McGill, S.-J. Yi, H. Yi, M. S. Ahn, S. Cho, K. Liu, D. Sun, B. Lee, H. Jeong, J. Huh, D. Hong, and D. D. Lee, "Team THOR's Entry in the DARPA Robotics Challenge Finals 2015," *Journal of Field Robotics*, vol. 34, no. 4, 2017, DOI: 10.1002/rob.21672.
- [9] J.-B. Yi, T. Kang, D. Song, and S.-J. Yi, "Unified software platform for intelligent home service robots," *Applied Sciences*, vol. 10, no. 17, 2020, DOI: 10.3390/app10175874.
- [10] S. Zakharov, I. Shugurov, and S. Ilic, "DPOD: 6D Pose Object Detector and Refiner," *Computer Vision and Pattern Recognition*, 2019, DOI: 10.48550/arXiv.1902.11020.
- [11] H. Sun, Z. Meng, X. Du, and M. H. Ang, "A 3D Convolutional Neural Network Towards Real-Time Amodal 3D Object Detection," *2018 IEEE/RISJ International Conference on Intelligent Robots and Systems (IROS)*, Madrid, Spain, 2018, DOI: 10.1109/IROS.2018.8593837.
- [12] B. Tekin, S. N. Sudipta, and P. Fua, "Real-time seamless single shot 6d object pose prediction," *Computer Vision and Pattern Recognition*, 2018, DOI: 10.48550/arXiv.1711.08848.
- [13] Y. Tanaka, Y. Ishida, Y. Abe, T. Ono, K. Kabashima, T. Sakata, M. Fukuyado, F. Muto, T. Yoshii, K. Kanamaru, D. Kamimura,

- K. Nakamura, Y. Nishimura, T. Morie, and H. Tamukoh, "Hibikino-Musashi@Home 2019 Team Description Paper," *arXiv:2006.01233*, 2019, DOI: 10.48550/arXiv.2006.01233.
- [14] 2019 - TDP - Tech United Eindhoven - RoboCup@Home - Domestic Standard Platform, [Online], <https://tdp.robocup.org/tdp/2019-tdp-tech-united-eindhoven-robocuphome-domestic-standard-platform/>, Accessed: 03 16, 2022.
- [15] R. Adams and L. Bischof, "Seeded region growing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 6, pp. 641-647, June, 1994, DOI: 10.1109/34.295913.
- [16] J. Fan, D. K. Y. Yau, A. K. Elmagarmid, and W. G. Aref, "Automatic image segmentation by integrating color-edge extraction and seeded region growing," *IEEE Transactions on Image Processing*, vol. 10, no. 10, pp. 1454-1466, Oct., 2001, DOI: 10.1109/83.951532.
- [17] H. Abdi and L. J. Williams, "Principal component analysis," *WIREs Computational Statistics*, vol. 2, no. 4, pp. 433-459, 2010, DOI: 10.1002/wics.101..
- [18] T. Kang, J.-B. Yi, D. Song, and S.-J. Yi, "High-speed autonomous robotic assembly using in-hand manipulation and re-grasping," *Applied Sciences*, vol. 11, no. 1, 2021, DOI: 10.3390/app11010037.



송 동 운

2020 부산대학교 지질환경과학과(학사)
2020~현재 부산대학교 전기전자공학과
(석사과정)

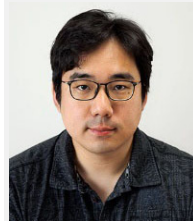
관심분야: 로봇 비전, 매니플레이터



이 재 봉

2020 부산대학교 전기컴퓨터공학부(학사)
2020~현재 부산대학교 전기컴퓨터공학부
(석박사통합과정)

관심분야: 지능로봇, 인공지능



이 승 준

2000 서울대학교 전기공학부(학사)
2002 서울대학교 전기컴퓨터공학부(석사)
2013 서울대학교 전기컴퓨터공학부(박사)
2017~현재 부산대학교 부교수

관심분야: 지능로봇, 인공지능