

# 강화학습 기반 3D 객체복원 데이터 획득 시뮬레이션 설계

진영훈\*  
백석대학교 첨단IT학부 교수

## Designing a Reinforcement Learning-Based 3D Object Reconstruction Data Acquisition Simulation

Young-Hoon Jin\*  
Professor, Division of Advanced IT, BaekSeok University

**요약** 물체나 공간을 디지털화하는 기술인 3D 복원은 주로 포인트 클라우드 데이터를 활용한다. 본 논문은 강화학습을 활용하여 주어진 환경에서 포인트 클라우드의 획득을 목표로 한다. 이를 위해 시뮬레이션 환경은 유니티를 이용하여 구성하고, 강화학습은 유니티 패키지인 ML-Agents를 활용한다. 포인트 클라우드 획득 과정은 먼저 목표를 설정하고, 목표 주변을 순회할 수 있는 경로를 계산한다. 순회 경로는 일정 비율로 분할하여 각 스텝마다 보상한다. 이때 에이전트의 경로 이탈을 방지하기 위해 보상을 증가시킨다. 에이전트가 순회하는 동안 목표를 응시할 때마다 보상을 부여하여 각 순회 스텝에서 포인트 클라우드의 획득 시점을 학습하도록 한다. 실험결과, 순회 경로가 가변적이지만 상대적으로 정확한 포인트 클라우드를 획득할 수 있었다.

**주제어** : 유니티, ML-Agents, 포인트 클라우드, 강화학습, 봇, 자율주행

**Abstract** The technology of 3D reconstruction, primarily relying on point cloud data, is essential for digitizing objects or spaces. This paper aims to utilize reinforcement learning to achieve the acquisition of point clouds in a given environment. To accomplish this, a simulation environment is constructed using Unity, and reinforcement learning is implemented using the Unity package known as ML-Agents. The process of point cloud acquisition involves initially setting a goal and calculating a traversable path around the goal. The traversal path is segmented at regular intervals, with rewards assigned at each step. To prevent the agent from deviating from the path, rewards are increased. Additionally, rewards are granted each time the agent fixates on the goal during traversal, facilitating the learning of optimal points for point cloud acquisition at each traversal step. Experimental results demonstrate that despite the variability in traversal paths, the approach enables the acquisition of relatively accurate point clouds.

**Key Words** : Unity3D; ML-Agents; Point Cloud; Reinforcement Learning, Bot, Autonomous Driving

## 1. 서론

3차원 공간 및 객체의 복원은 현실 세계의 물체와 환경을 컴퓨터 비전 및 그래픽스 기술을 사용하여 3차원

모델로 변환하는 과정이다. 이러한 기술은 컴퓨터 비전, 로봇 공학, 게임 개발, 문화유산 보존, 의료 영상 처리 및 기타 다양한 분야에서 활용되고 있으며 메타버스 및 디지털 트윈 기술의 발전으로 더욱 가속화되고 있다[1,2].

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2022R1G1A1012974).

\*교신저자 : 진영훈(devjay@bu.ac.kr)

접수일 2023년 10월 18일 수정일 2023년 11월 21일 심사완료일 2023년 11월 23일

로봇 공학 분야에서는 주변 환경을 분석하고 3D 모델로 재구성하기 위해 센서, 포인트 클라우드 등을 통해 다양한 자동화 작업 및 로봇 응용 프로그램을 연구한다.

게임 분야에서는 게임 환경 및 캐릭터를 현실적으로 시각화하기 위해 사용된다. 이는 게임의 시각적 품질과 현실감을 크게 향상시키며, 플레이어에게 몰입감 있는 경험을 시간과 비용을 절약하면서 빠르게 구축할 수 있도록 한다.

문화유산 보존 분야에서 3D 복원 기술은 역사적 건물, 조각, 유물, 미술품 및 유산의 보존, 연구 및 전시를 위해 사용된다. 이 기술은 특히 실제 피질에 손상을 입히지 않으면서도 유산의 중요한 측면을 보존할 수 있는 장점이 있다.

치료 영상 처리 분야에서 3D 복원 기술은 환자의 해부학적 구조를 정확하게 복원하고 시각화하는데 사용되며, 환자의 진단, 치료 및 관리에 필수적이고 더 나은 의료 결과를 제공한다. 이와 같이 다양한 영역에서 활용되는 3차원 복원 기술은 일반적으로 <Table 1>과 같은 절차로 진행된다.

<Table 1> 3D Reconstruction Process

Process	Content
Data Collection	Measuring and collecting data of real-world objects or scenes
Point Cloud Generation	Processing the information obtained during the data collection phase
Data Refinement and Filtering	Refining and filtering noise, outliers, and empty spaces
Point Cloud Registration	Integrating data acquired from various perspectives
Point Cloud Processing	Extracting surfaces, mapping textures, and adjusting color and lighting conditions
3D Model Generation	Processing the data to create a polygon mesh model
3D Model Post-Processing	Adding texture, color, and lighting information

<Table 1>의 절차 중에서 데이터의 수집은 스캐닝 장비를 이용하여 사람에 의해 이루어지는데 이때 많은 시간과 비용이 발생한다. 또한, 접근하기 어려운 환경은 데이터 수집이 어려워 드론을 활용하는 경우가 많다. 이러한 문제를 해결하기 위해 본 논문은 강화학습을 기반으로 데이터 수집의 자동화를 제안한다.

강화학습 기반 데이터 수집은 목표점의 이동과 포인트 클라우드 획득의 2단계로 이루어진다. 이렇게 수집된 데이터는 포인트 클라우드의 생성 및 정합, 처리에 활용된다.

## 2. 관련연구

3차원 복원 기술은 컴퓨터 비전 영역에서 많은 연구가 이루어졌으며[3,4,5], 특히 최근에는 딥러닝 및 강화학습 기술의 발전으로 다양한 연구가 진행되고 있다[6].

Shixiang Gu는 로봇 조작을 위해 딥 강화학습에서 비동기적 오프-폴리시 업데이트를 통해 학습 과정을 효율적으로 가속화하였다. 이는 현재 학습 중인 정책을 방해하지 않고 데이터를 수집하여 대규모 환경에서 복잡한 에이전트를 비동기적으로 학습시켜 정책 업데이트의 효율을 높이는 방식이다[7].

Lillicrap, Timothy P.는 Deep Deterministic Policy Gradients (DDPG)라는 알고리즘을 제안하여 연속적인 액션 공간에서의 강화학습 문제를 효과적으로 해결하였다. 경험 재사용을 통해 데이터 효율성을 높여 학습 안정성을 향상시킨다[8].

Mirowski, Piotr는 복잡한 환경에서 로봇이 학습하고 이동하는 방법을 딥러닝과 강화학습을 기반으로 구성하였다[9]. 네비게이션 문제는 항상 에이전트가 센서에 의존할 경우 주변 환경의 빠른 변화에 대처하기 어려운 점을 엔드투엔드(end-to-end) 학습 프레임워크를 구성하였다. 학습 프레임워크는 먼저, A3C(Asynchronous Advantage Actor-Critic) 알고리즘을 통해 누적 보상을 최대화하고, RGB 관측값에서 심도맵(Depth Map)을 추론하는 손실값을 최소화한다. 끝으로 에이전트가 자신의 이동 경로에서 이전에 지나간 위치를 다시 인식하는 루프 클로저(Loop Closure)를 감지하도록 훈련을 수행하여 네비게이션 문제를 해결하려 하였다.

Tatarchenko는 단일 이미지(Single View)를 기반으로 3D 객체복원 네트워크가 어떻게 학습하는지 연구하였다[10].

Kevin Rioussms 분류 성능과 탐사 품질 모두에 대해 보상을 받는 새로운 RL 프레임워크를 제안하였다[11].

Waleed는 LiDAR를 이용하여 포인트 클라우드에서 실시간으로 3D 객체 위치, 방향 및 경계 검출연구를 수행하였다[12].

Julio는 강화학습 기반 Active SLAM을 통해 로봇이 환경을 모르는 상태에서 동시에 위치 추정과 지도 작성을 수행하는 연구를 진행하였다[13].

위와 같이 강화학습은 로봇제어, 자율주행, 객체 인식 및 복원 등 다양한 영역에 활용된다. 본 연구는 강화학습을 기반으로 물리 세계의 객체에서 포인트 클라우드 획득을 할 수 있는 시뮬레이션 설계를 제안한다.

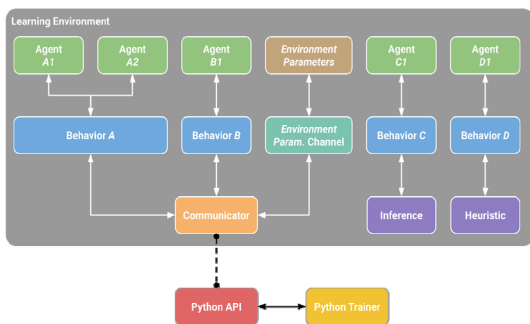
### 3. 시뮬레이션 설계

일반적으로 강화학습은 모델 유무에 따라 분류한다. 모델이 없는 강화학습(Model Free RL)은 시스템에 적용되는 모델 없이 환경과 상호작용을 통해 데이터를 획득하고 정책을 획득할 수 있지만, 단순한 환경에서도 모델 기반 강화학습에 비해 상대적으로 많은 시도횟수가 필요하다. 반면 모델 기반 강화학습(Model Based RL)은 데이터 효율성, 안정성 및 효과적인 정책 개선 등의 장점이 있지만, 모델 구성을 위한 상대적 기회비용이 필요하다 [14]. 따라서 3장에서는 비 모델 강화학습을 위한 환경구성 및 보상체계를 논한다.

#### 3.1 ML-Agents

ML-Agents는 Unity 엔진에서 강화학습을 수행할 수 있는 프레임워크로, Unity 엔진에서 환경을 구성하여 에이전트를 훈련할 수 있다. 이를 이용하면 게임, 가상 훈련 환경, 로봇제어, 자율주행 등 다양한 환경에 대한 시뮬레이션이 가능하다.

[Fig. 1]에서 에이전트는 병렬 구성이 가능하고, 같거나 다른 행동을 할 수도 있다. 이를 통해 더 빠르게 학습을 할 수 있으며 환경(State)과 상호작용(State)을 통한 측정값을 ML Core(Python Trainer)에 전달한다. 훈련을 통한 추론을 기반으로 행동(Action)하고 상호작용하는 과정을 통해 강화학습이 수행된다.

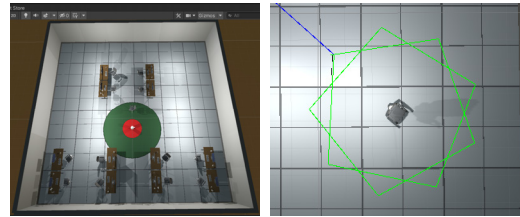


[Fig. 1] The Operation of ML-Agents[15]

#### 3.2 Unity 환경구성

환경은 [Fig. 2]와 같이 구성하여 에이전트가 일정한 간격을 두고 [Fig. 3]과 같이 목표점을 360도 순회하며 포인트 클라우드를 획득할 수 있도록 한다.

포인트클라우드의 획득은 물리세계의 라이다(Lidar) 센서를 모방할 수 있는 레이캐스트(raycast)를 이용한다. 레이캐스트는 레이(가상의 선)를 쏘 안에서 방사하여 다른 객체와의 상호작용을 감지하는 방식으로, 3D 공간에서 객체와의 충돌 및 특정 지점에 대한 정보를 얻기 위해 사용된다.

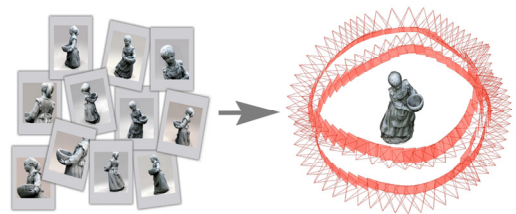


[Fig. 2] Virtual environment and object traversal path

#### 3.3 강화학습

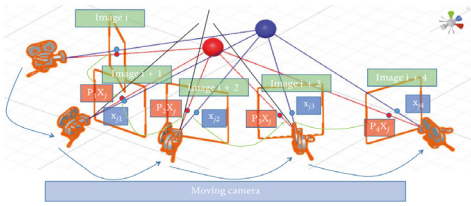
강화학습(Reinforcement Learning)은 에이전트(Agent)가 환경과 상호작용하며 특정 작업을 수행하여 학습하는 방식으로 어떤 행동에 대한 최대의 보상을 통해 시행착오를 극복하는 알고리즘이다. 강화학습기반 객체복원을 위해서는 객체를 구성하는 포인트 데이터셋을 획득해야 한다.

객체복원을 위해서는 목표를 360도 순회하며 데이터를 획득하는 것이 중요하다. [Fig. 3]은 SfM(Structure from motion)알고리즘을 이용한 객체복원과정이다. SfM알고리즘은 여러 개의 2D 이미지에서 3D 장면의 구조와 카메라의 동작을 추정하는 것을 목표로 한다 [16,17].



[Fig. 3] The Object Reconstruction Process Using SfM Algorithm[16]

SfM과 같은 과정으로 강화학습 기반 객체복원을 구성하면 [Fig. 4]와 같이 객체를 순회하며 특징이 잘 표현되는 시야각에서 데이터획득이 가능하도록 <Table 2>와 같이 설계되어야 한다.



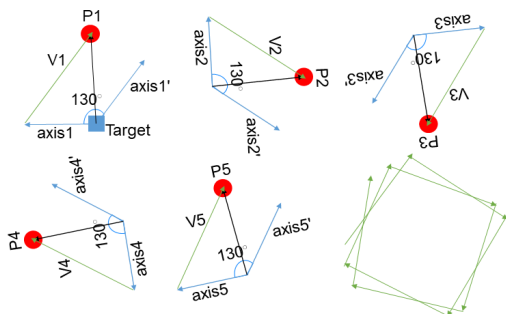
[Fig. 4] Reconstruction Procedure Through SfM Feature Correspondences[17]

<Table 2> Reconstruction Procedure Using Reinforcement Learning in 3D

Process	Content
Objective Verification	Target Detection Using Raycasting
Traversal Path Confirmation	Calculation of a 360-Degree Traversable Path
Acquisition of Point Set	Acquisition of Point Cloud
Reward System	Construction of Rewards for Traversal/Data Acquisition

목표 확인절차는 360도 방향으로 레이를 방사하여 벽 이외의 객체를 확인하는 과정이다. 목표가 감지되면 좌표계(기준 축)를 구성하고, 목표를 중심으로 오각형과 비슷한 형태의 순회 경로 확인절차를 [Fig. 5]와 같이 반복적으로 계산한다. 이때  $V_1$ 의 길이는 단위벡터의 4배를 한다. 최종 순회 경로는  $V_1 + V_2 + \dots + V_n$ 을 통해 완성되며, 수식 (1)을 통해 구성할 수 있다.

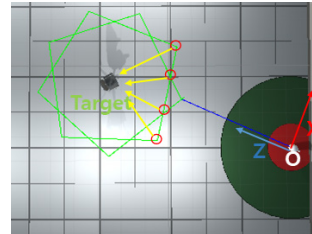
$$\begin{aligned}
 axis1 &= -\overrightarrow{Target}_{raycast} \\
 axis1' &= Q_{(130, Y)} \times axis1 \\
 V &= axis1' \times 4 \\
 axis_{pre} &= P_{pre} - Target \\
 Y &= (0, 1, 0)
 \end{aligned}
 \tag{1}$$



[Fig. 5] Traversal Path Computation

포인트 셋 획득 절차는 순회 경로를 따라 이동하는 과정에서 목표점을 바라보며 포인트 클라우드를 획득하는 절차이다. 마지막으로 보상체계 절차는 순회 경로 및 포인트 셋 획득 절차를 [Fig. 6]과 같이 학습하기 위한 절차이다.

먼저 순회 경로에서 보상체계는 [Fig. 6]과 같이 각 경로를 4 분할하여 순차적으로 도달할 시 보상의 강도를 강하게 주는 방법을 사용한다. 이를 통해 경로 이탈을 막고 더 빠른 학습을 할 수 있다.



[Fig. 6] Reinforcement Learning Reward System and Coordinate System Configuration

포인트 셋 획득을 위한 보상체계는 [Fig. 6]과 같이 현재 에이전트가 바라보는 방향과 위치에서 목표점의 방향의 내적값이 일정 값 이상이면 보상하는 방식이다. 이러한 보상체계를 통해 데이터 자율 획득이 가능하게 된다.

모든 절차가 완료되면 강화학습이 가능하지만, 좌표체계를 구성이 필요하다. 유니티 환경은 에이전트, 목표점 등 모든 객체에 대한 위치, 회전에 대한 정보를 확인할 수 있지만, 실세계는 그렇지 못하다. 본 논문은 실제 빛에 탐재를 위한 시뮬레이션이므로 빛을 중심으로 하는 좌표체계를 [Fig. 6]과 같이 구성해야 한다.

먼저, <Table 2>의 목표 확인절차에서 빛이 목표를 감지할 때 빛을 원점으로 목표 방향의 벡터를 월드좌표계의 앞방향(Z축)으로 구성한다. Y축은 지면에 수직인 벡터이고, X축은 Z축을 Y축 방향을 중심으로 90도 회전하는 사원수(Quaternion)를 구성하여 계산하면 된다. 식 (2)은 위 과정을 도식화한다.

$$\begin{aligned}
 O &= Bot_{position}, Z = \overrightarrow{Target}_{raycast} \\
 Y &= (0, 1, 0), X = Q_{(90, Y)} \times Z
 \end{aligned}
 \tag{2}$$

이렇게 구성된 환경을 통해 자율주행하며 포인트 클라우드 데이터를 획득하는 빛 시뮬레이션을 구성할 수 있다.

### 4. 실험

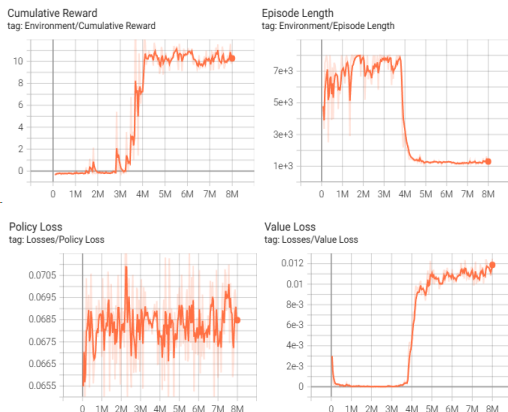
시뮬레이션을 위한 시스템은 Core(TM) i9-12900K 3.19 GHz와 32GB 메모리, ML-Agents 0.30, PyTorch 2.0 및 Cuda 117 버전을 사용한다. 실험은 <Table 3>의 의사 코드(pseudo-code)를 따라 구성되며 PPO (Proximal Policy Optimization) 알고리즘을 사용한다.

<Table 3> ML-Agents pseudo code

Init	agent forward : $\vec{A}_{fw}$ , agent to target : $\vec{V}_{tg}$ , agent : $A$ , step position : $\oplus$ , distance between $A$ and $\oplus$ : $D$
1	360° loop ray each 1°
2	observation $\{\vec{A}_{fw}, \vec{V}_{tg}, D, \vec{A}_{fw} \cdot \vec{V}_{tg}\}$
3	if distance $\oplus, A < \theta$ then reward, step++ if distance $\oplus, A < \phi$ and $\vec{A}_{fw} \cdot \vec{V}_{tg} > \tau$ then reward
4	repeat 2 ~ 3 lines

강화학습 8백만 번 시도 후 [Fig. 7]의 누적 보상(Cumulative Reward), 에피소드 길이(Episode Length), 정책손실(Policy Loss), 가치손실(Value Loss) 결과를 보면 3백만 번 시도 후 누적 보상이 증가하며, 에피소드 길이는 누적 보상이 극에 달하는 4백만 번 즈음부터 급격히 하락한다. 이러한 결과는 정상적인 학습으로 평가할 수 있다.

정책손실을 보면 1 이하에서 진동하고 있는데, 이는 훈련 중에 보이는 현상이며 더 나은 정책을 학습하는 중에는 값이 낮아진다.

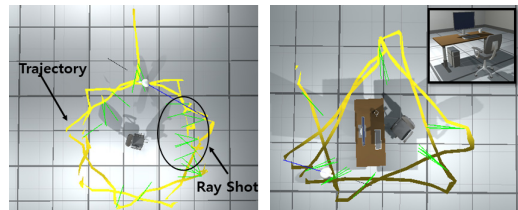


[Fig. 7] Results Reinforcement Learning

가치손실은 에이전트가 학습하는 동안 증가하고 보상이 안정화되면 감소하며, 보상이 증가함에 따라 증가하고 보상이 안정되면 감소하는 것이 일반적이다. 따라서 그래프를 확인하면 정상적으로 학습했음을 확인할 수 있다.

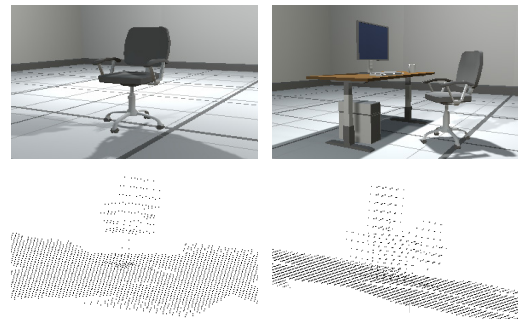
학습결과를 토대로 단일/다중 객체에 대한 포인트 데이터 획득수행은 [Fig. 8]과 같다.

[Fig. 8]에서 단일 객체 탐색은 제시된 오각형 모양의 순회를 거치며 레이를 잘 발사한 모습을 보인다. 반면 다중 객체 탐색은 찌그러진 오각형 형태의 모양을 보이며 레이 발사 역시 특정 위치에 편중된 모습을 보인다. 이는 객체가 커짐에 따라 학습한 반경을 초과해서 발생하는 현상으로 보인다.



[Fig. 8] Results with single/multi Object

순회를 통한 포인트 클라우드 획득결과는 [Fig. 9]와 같다.



[Fig. 9] Point Cloud Acquisition Results

### 5. 결론

본 논문에서 제안하는 강화학습 기반 포인트 클라우드 추출은 [Fig. 9]와 같이 가능함을 보였다. 비록 이동 경로와 레이 발사 위치가 가변적임에도 불구하고 제안하는 방법으로 학습하면 비교적 정확한 데이터 추출이 가능하다. 추후 포인트 클라우드의 밀도를 높이고, 정교한 움직임

임이 가능하도록 학습방법을 개선하여 바퀴 구동형 붓에 탑재하는 연구를 수행할 예정이다.

## REFERENCES

- [1] To, Alex, et al. "Drone-based AI and 3D reconstruction for digital twin augmentation." International Conference on Human-Computer Interaction. Cham: Springer International Publishing, pp.511-529, 2021.
- [2] Navarro, Francisco, et al. "Integrating 3D reconstruction and virtual reality: A new approach for immersive teleoperation." ROBOT 2017: Third Iberian Robotics Conference: Springer International Publishing, Vol.2, pp.606-616, 2018.
- [3] Ham, Hanry, Julian Wesley, and Hendra Hendra. "Computer vision based 3D reconstruction: A review." International Journal of Electrical and Computer Engineering Vol.9, No.4, pp.2394-2402, 2019.
- [4] Han, Xian-Feng, Hamid Laga, and Mohammed Bennamoun. "Image-based 3D object reconstruction: State-of-the-art and trends in the deep learning era." IEEE transactions on pattern analysis and machine intelligence Vol.43, No.5, pp.1578-1604, 2019.
- [5] Guo, Yulan, et al. "Deep learning for 3d point clouds: A survey." IEEE transactions on pattern analysis and machine intelligence Vol.43, No.12, pp.4338-4364, 2020.
- [6] Samavati, Taha, and Mohsen Soryani. "Deep learning-based 3D reconstruction: A survey." Artificial Intelligence Review pp.1-45, 2023.
- [7] Gu, Shixiang, et al. "Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates." 2017 IEEE international conference on robotics and automation (ICRA). IEEE, pp.3389-3396, 2017.
- [8] Lillicrap, Timothy P., et al. "Continuous control with deep reinforcement learning." arXiv preprint arXiv:1509.02971, 2015.
- [9] Mirowski, Piotr, et al. "Learning to navigate in complex environments." arXiv preprint arXiv:1611.03673, 2016.
- [10] Tatarchenko, Maxim, et al. "What do single-view 3d reconstruction networks learn?." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp.3405-3414, 2019.
- [11] Riou, Kevin, Kevin Subrin, and Patrick Le Callet. "Reinforcement Learning Based Point-Cloud Acquisition and Recognition Using Exploration-Classification Reward Combination." 2022 IEEE International Conference on Multimedia and Expo (ICME). IEEE, pp.1-6, 2022.
- [12] Ali, Waleed, et al. "Yolo3d: End-to-end real-time 3d oriented object bounding box detection from lidar point cloud." Proceedings of the European conference on computer vision (ECCV) workshops. pp.716-728, 2018.
- [13] Placed, Julio A., and José A. Castellanos. "A deep reinforcement learning approach for active SLAM." Applied Sciences Vol.10, No.23, p.8386, 2020.
- [14] Swazinna, Phillip, et al. "Comparing model-free and model-based algorithms for offline reinforcement learning." IFAC-PapersOnLine Vol.55, No.15, pp.19-26, 2022.
- [15] The Operation of ML-Agents Graphics [Internet]. <https://unity-technologies.github.io/ml-agents/ML-Agents-Overview/>.
- [16] Bianco, Simone, Gianluigi Ciocca, and Davide Marelli. "Evaluating the performance of structure from motion pipelines." Journal of Imaging Vol.4, No.8, p.98, 2018.
- [17] Jin, Young-Hoon, Kwang-Woo Ko, and Won-Hyung Lee. "An indoor location-based positioning system using stereo vision with the drone camera." Mobile Information Systems, Vol.2018, 2018.

진 영 훈(Young-Hoon Jin)

[중심회원]



- 2019년 8월 : 중앙대학교 (영상학박사)
- 2021년 3월 ~ 현재 : 백석대학교 첨단IT 조교수

<관심분야>

XR, RL, IoT, Network, Reverse Engineering