

멀티모달 LLM 기반 Topological Map 작성 방법

Topological Map Building based on Multi-modal LLM

최창조¹ · 원종완¹ · 어규호[†]

Changjo Choi¹, Jong-Wan Won¹, Gyuho Eoh[†]

Abstract: This paper presents a novel approach to building topological maps based on a multi-modal large language model (MLLM). Traditional methods for building topological maps compare the features of the previous and current scenes and separate places by comparing their similarity. However, their reliance on features has made them difficult to use in environments with repetitive patterns. In addition, traditional methods only provide a simple link between places without relative directional information. To overcome these limitations, we propose a method for building a topological map using directional signboards. By using text and directional information from directional signboards, a robot can overcome ambiguity in repetitive patterns and can identify places. Furthermore, the robot can build a topological map without having to visit all the places. To obtain directional signboard data and transform it into a suitable format for further processing, we use a MLLM, which can simultaneously perform directional signboard recognition, text recognition, and orientation recognition within an image. We evaluate the efficiency of MLLM-based topological map building methods in real environments.

Keywords: Large Language Model, Sign Recognition, SLAM, Topological Map

1. 서론

로봇은 위치 파악과 탐색을 수행하기 위해 주변 환경을 표현할 수 있는 지도를 작성해야 한다. 지도 작성 방법은 크게 metric map 작성 방법과 topological map 작성 방법이 있다.

Metric map은 Camera 센서로부터 추출된 feature를 이용하여 2D 또는 3D 지도를 그리거나 LiDAR, RADAR와 같은 거리 센서에서 획득된 정보를 이용하여 단위 격자의 상태를 업데이트하며 지도를 작성한다. Metric map은 환경에 대한 정확한 거리 및 위치 정보를 제공할 수 있다는 장점이 있지만, 많은 메모리 공간이 필요하고 연산 비용이 많이 드는 단점이 있다. 반면,

Topological map은 장소를 나타내는 노드(node), 노드 사이의 관계를 나타내는 엣지(edge)로만 지도를 작성한다. 따라서 공간 복잡도가 낮고, 정확한 위치 파악이 필요하지 않아 metric map보다 필요한 메모리 공간 및 연산 비용이 적다^[1,2]. 전통적인 topological map 작성 방식은 장소 간의 특징을 비교하여 새로운 장소를 인식하고 노드를 생성한다. 그러나 유사 및 반복 패턴이 많은 환경에서는 이전 장면과의 특징 차이가 거의 없기 때문에 노드 생성 시점을 명확히 할 수 없다^[3]. 최근 topological map 작성 방법은 로봇이 이동하며 책상, 의자, 문과 같은 semantic 정보를 인식할 때마다 노드를 생성해 topological map을 작성하는 방식^[4]과 RGB 카메라와 LiDAR 데이터를 결합하여 시각 정보가 풍부하다고 판단되는 지역에서 노드를 생성하여 topological map을 작성하는 방식^[5]이 연구되었다. 그러나 대부분의 topological map은 노드 사이의 상대적인 연결 관계만 나타내기 때문에 노드 간 방향에 대한 정보가 없어 방문하지 않은 장소에 대해서는 지도를 작성할 수 없다. 또한, semantic 정보를 활용한 방식에서는 객체를 정확하게 인식하지 못하면 환경을 부정확하게 표현할 수 있으며, 고해상도의 semantic 정보를 다루기 때문에 복잡한 환경이나 대규모 환경에서는 topological map 구축에 메모

Received : Oct. 2, 2024; Revised : Oct. 28, 2024; Accepted : Oct. 30, 2024

* This work was supported by the GRR program of Gyeonggi province [GRR TUKorea2023-B03, Development of an intelligent inspection system and an autonomous navigation system for the transportation of multi-material parts].

1. Undergraduate Student, Department of Mechatronics Engineering, Tech University of Korea, Gyeonggi, Korea (ckdwh99@gmail.com, jongwon2009@tukorea.ac.kr)

† Assistant Professor, Corresponding author: Department of Mechatronics Engineering, Tech University of Korea, Gyeonggi, Korea (gyuho.eoh@tukorea.ac.kr)

리 사용량과 계산 부담이 크다.

이러한 기존 방식의 문제를 해결하기 위해 본 논문에서는 방향 표지판 인식을 활용한 topological map 작성 방법을 제시한다. 일반적으로 방향 표지판은 서로 구별되는 장소를 가리키고 있으므로, 방향 표지판이 인식되는 시점에 노드를 생성함으로써 노드 생성 시점의 모호성을 줄일 수 있다. 또한, 방향 표지판의 방향 정보를 활용하면 topological map 작성 시 로봇이 방문하지 않았던 장소에 대해서도 노드 생성이 가능하기 때문에 topological map 작성 시간을 단축할 수 있으며, 필요한 메모리 공간 및 계산 비용을 줄일 수 있다. 방향 표지판 인식을 위해서는 그림, 글자와 같은 이중 정보를 동시에 입력하여 원하는 형태의 결과물을 얻을 수 있는 MLLM을 이용한다. 방향 표지판을 MLLM의 입력으로 주고, 우리가 원하는 형태의 topological 정보를 산출할 수 있도록 프롬프트를 작성한다. 이러한 방식으로 이미지 인식, 장소 판단, 위치 관계 산출 등의 복잡한 단계로 이루어지는 topological map 작성을 단순화할 수 있다. LLM은 현재 시점에서 가장 널리 알려지고 사용되어지고 있는 ChatGPT 4o^[6]를 활용하되, Claude^[7], Gemini^[8] 등 다른 LLM에 대해서도 적용 가능한 형태로 제안한다.

본 논문의 기여는 다음과 같다.

1. MLLM 기반 topological map 작성 방법을 제시한다.
2. 방향 표지판 정보를 활용하여 방향 정보를 가진 topological map을 작성하였다.
3. 제안된 MLLM 기반 topological map을 이용하여 localization에 활용할 수 있음을 보인다.
4. 다양한 MLLM을 활용하여 topological map을 작성하는 것을 실제 실험을 통해 검증하였다.

이 논문의 구성은 다음과 같다. 2장에서는 topological map과 MLLM에 관련된 연구를 설명한다. 3장에서는 본 연구의 문제를 정의하고 연구 목표를 제시한다. 4장에서는 본 연구에서 제안하는 방식의 개요를 제공한다. 5장에서는 본 연구에서 제안하는 방식인 MLLM 기반 topological map 작성 방법을 자세히 설명한다. 6장에서는 실제 실험을 통해 LLM의 성능을 비교하고 프롬프트 작성 방법에 따른 성능 변화와 기존 환경과 작성된 topological map의 유사도를 수치화해서 검증하며, 작성된 topological map을 이용해서 localization 한다. 마지막으로 7장은 결론으로 구성되어 있다.

2. 관련 연구

2.1 Topological Map

Topological map은 metric map보다 필요한 메모리 공간 및

연산 비용이 적어 정밀한 공간 측정이 불필요한 상황에서 metric map보다 효율적으로 사용될 수 있다^[9].

초창기의 topological map은 Voronoi diagram에서 임계점과 선을 결정하는 것을 목표로 했으며, 등거리 장애물로부터 노드와 엣지를 추출한다^[10]. 이후 Voronoi diagram을 개선하거나 기하학적 특징을 갖춘 가상 센서 접근 방식을 사용한 방식이 연구되었고^[11,12], 지도학습을 사용한 방식이나 스펙트럼 클러스터링을 사용하는 방식으로 발전했다^[13,14]. 또한, 내비게이션 작업을 위해 프로그래밍한 모바일 로봇이 로봇 주변 영역에서 불변 영역을 감지하고 다각형을 그리는 방식의 topological map 작성 방식^[15]과 센서 입력과 위상 및 지역적 측정 정보를 결합하여 이 정보를 설명할 수 있는 topological map을 작성하는 방법이 연구되었다. 이러한 방식들은 다수의 센서와 사전 작업이 필요해 큰 비용이 드는 문제가 있다. 이후 비용 문제를 해결하기 위해 적은 개수의 센서로 topological map을 작성하는 방법이 등장하였으며, 그 방법 중 단안 휴대용 광각 카메라를 사용하여 topological map을 작성하는 방법이 제안되었다^[16]. 이 방식은 현재 이미지와 이전 이미지의 feature를 비교하여 유사도 함수를 통해 노드 생성을 결정한다. 한편, 지상에서 로봇이 주행하는 방식이 아닌 공중에서 촬영한 영상을 통해 지형 지도를 형성하여 topological map을 작성하는 방식도 제안되었다^[17]. F. Blochliger^[18]은 용도에 맞게 제작된 지도를 로봇에 제공하여 localization 하는 Topomap을 제안하였고, D. S. Chaplot^[19]은 semantic 정보를 학습시키고 목표 물체가 있을 확률이 높다고 판단되는 장소로 주행하며 지도를 작성하고 localization 하는 모델을 제안했다.

2.2 Multi-modal Large Language Model

MLLM의 등장으로 언어 모델이 해결할 수 있는 작업의 범위가 확장되었고, 로봇에 MLLM을 적용시키는 방식이 연구되고 있다. K. Rana^[20]은 MLLM의 이미지 인식을 통해 장면 내의 물체를 구분하였으며, Z. Long^[21]은 MLLM에 로봇 비전 작업을 요청하여 객체 감지 및 식별에 활용했다. H. Li^[22]은 MLLM을 이용해 현재 이미지와 이전 이미지를 구별하여, 반복 패턴이 있는 환경에서 지도를 작성하였다. S. Grag^[23]은 MLLM을 통해 이미지에 대한 정보를 얻고, 그 정보를 통해 지도를 작성하였다. 그리고 MLLM을 통해 장면에서 텍스트를 감지하고 랜덤마크를 나타내는지 판단하여 SLAM에 활용한 연구^[24]와 실내 환경에서 장면 이해 및 장소 인식을 위해 텍스트 의미 파악에 MLLM을 사용하여 topological map을 작성하고 localization 하는 방법이 연구되었다^[25]. 또한 최근 연구에서는 자연어 경로 설명을 LLM에 입력하고, 경로 설명에서 장소와 동작을 추출해 노드와 엣지를 만들어 topological map을 작성하였다^[26]. 그러

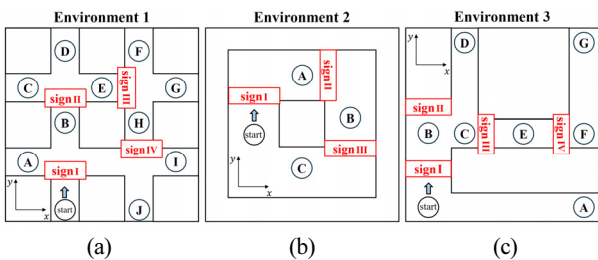
나 해당 방법에서는 경로 설명을 일정한 구조로 작성해야 하는 한계와 LLM이 자연어 설명을 통해 경로의 공간 구조를 완벽하게 이해하지 못한다는 단점이 있다.

본 논문에서는 MLLM의 이미지 인식을 통해 방향 표지판과 방향 표지판의 정보를 검출하고, 노드 생성 시점을 명확하게 하여 기존 방식들의 문제점을 극복한다.

3. 문제 정의

본 연구의 문제는 로봇이 주행하며 얻은 이미지에서 방향 표지판 정보를 추출하고, 추출된 정보를 통해 topological map을 작성하는 것이다. 방향 표지판이 있는 환경의 예는 [Fig. 1]에 나타내었다. 다양한 환경에서 실험하기 위해 다음과 같은 3가지 형태로 구성하였다. [Fig. 1(a)]는 일반적인 맨해튼(Manhattan) 구조^[27]인 격자형 구조이며 최단 거리 이동으로 환경 전체를 표시할 수 있는 환경을 나타내었다. [Fig. 1(b)]는 정사각형 구조로 환경 자체는 대칭이지만, 방향 표지판 인식으로 장소가 구별될 수 있음을 보여준다. [Fig. 1(c)]는 비대칭 구조로 주어진 환경이 균일하지 않더라도 topological map을 그릴 수 있는 환경인지 검증한다. 방향 표지판의 정보를 알고 있다면 해당 환경에서 장소 사이의 방향 관계를 파악할 수 있고, 방향 관계를 통해 topological map을 작성할 수 있다. 방향 표지판의 정보를 얻기 위해서는 방향 정보에 해당하는 화살표를 읽을 수 있어야 한다.

본 연구에서는 다음과 같이 3가지의 가정을 한다. 먼저, 방향 표지판은 로봇의 정면에서 인식된다는 것이다. 일반적으로 카메라는 로봇의 전방에 부착되어 있으며, 이 카메라를 통해 인식하기 때문에 방향 표지판은 정면에서 인식된다. 따라서, 로봇이 방향 표지판의 측면을 관측하여 생길 변수를 고려하지 않는다. 둘째는 로봇이 주행하는 환경은 방향 표지판을 통해 묘사할 수 있는 환경이어야 한다. 즉, 전체 환경내의 방향 표지판을 모두 모았을 때, 지도를 그리기 위한 충분한 정보가 포함되어 있어야 한다. 본 논문에서 제안하는 방식은 방향 표지판의 정보를 MLLM을 통해 추출하여 topological map을 생성한다. 마지막으로 맨해튼 세계 가정(Manhattan-world assumption)을 한다.

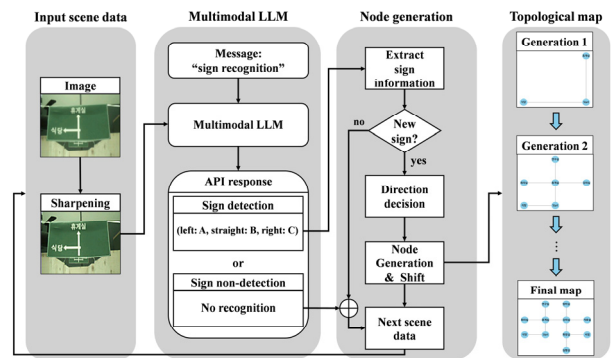


[Fig. 1] Examples of environments with signs: (a) lattice structure (b) square structure and (c) asymmetric structure

방향 표지판의 방향 정보로 정면, 좌측, 우측을 사용한다. 따라서 대각선 방향의 경우 LLM의 방향 표지판 인식에서 방향 인식의 모호성이 생기기 때문에 해당 내용은 고려하지 않는다.

4. 개요

이 장에서는 본 논문에서 제안하는 방식인 MLLM 기반 topological map 작성 방법의 개요에 대해 [Fig. 2]를 통해 설명한다. 해당 방식은 Input scene data, Multimodal LLM, Node generation, Topological map으로 구성되어 있다. 먼저, Input scene data에서는 로봇이 주행하며 카메라를 통해 얻은 이미지가 입력되며, 글자를 선명하게 하기 위해 sharpening을 한다. Sharpening 된 이미지는 방향 표지판 인식 요청 메세지와 함께 MLLM에 입력되며 방향 표지판에서 인식된 정보 또는 인식되지 않았음을(“no recognition”) 알려준다. 방향 표지판이 인식되지 않았을 경우에는 다음 scene data를 입력받으며, 방향 표지판이 인식되었을 경우에는 Node generation 단계로 넘어간다. Node generation 단계에서는 인식된 방향 표지판 정보를 가져와 이전 응답과 비교하여 관측되지 않았던 방향 표지판인지 판단한다. 기존에 관측되었던 방향 표지판이라면 다음 scene data 입력을 받는다. 기존에 관측되지 않은 방향 표지판이 인식되었다면 방향 지정을 한다. 방향 지정은 로봇의 wheel odometry를 통해 구해진 로봇의 헤딩 값(yaw)으로 결정한다. 방향이 결정되었으면 인식된 방향 표지판 정보에 따라 노드를 생성하고 해당 방향으로 이동한다. 노드를 생성하고 나면 topological map을 출력한다. 관측되지 않았던 방향 표지판이 인식되면 Generation 1이 생성되며, 관측되지 않았던 새로운 방향 표지판이 인식될 때마다 topological map을 갱신한다. 더 이상의 입력 이미지가 없으면 final map을 출력한다.



[Fig. 2] Overview of topological map building based on multimodal LLM: this figure consists of Input scene data, Multimodal LLM, Node generation, and Topological map

5. MLLM 기반 Topological Map 작성

5.1 MLLM을 활용한 방향 표지판 인식

Topological map을 그릴 때 사용하는 정보는 로봇이 주행하면서 얻은 장면 이미지에 대한 MLLM의 응답과 그 장면에서 로봇의 odometry를 통해 구한 헤딩 값이다. 로봇이 주행하며 얻은 장면 이미지의 방향 표지판과 글자를 MLLM에서 더 정확하게 인식할 수 있도록 sharpening을 통해 이미지의 선명도를 높인다. 본 연구에서는 글자 인식을 위해 에지 강화의 효과가 강한 unsharp mask^[28]를 사용하였다. 방향 표지판을 인식한 MLLM은 방향 표지판 인식 정보에 대한 응답을 준다. 이 응답을 이용해 topological map을 생성하기 위해서는 인식된 방향과 장소에 대한 정보를 담아 일정한 구조로 출력되도록 프롬프트를 작성해야 한다. 해당 프롬프트 작성 구조는 [Fig. 3]과 같고, 방향 표지판 인식 지정, 출력 형태 지정, 예시 입력, 방향 표지판 글자 표시의 총 4개의 과정으로 구성했다. 첫째, 방향 표지판 인식 지정은 글자 인식이 가능한 방향 표지판을 인식하도록 하여 부정확한 노드 생성을 줄인다. 둘째, 출력 형태 지정은 좌측, 정면, 우측에 대한 장소 정보를 담아 동일한 구조로 출력되어 topological map 생성 알고리즘에서 활용될 수 있도록 한다. 셋째, 예시 입력은 예시를 통해 MLLM이 출력해야 하는 형태를 지정한다. 마지막으로 방향 표지판의 언어가 어떤 언어로 되어 있는지 지정하여 MLLM에서 이미지를 분석하는 시간을 줄이고, 인식 정확도를 높일 수 있다.

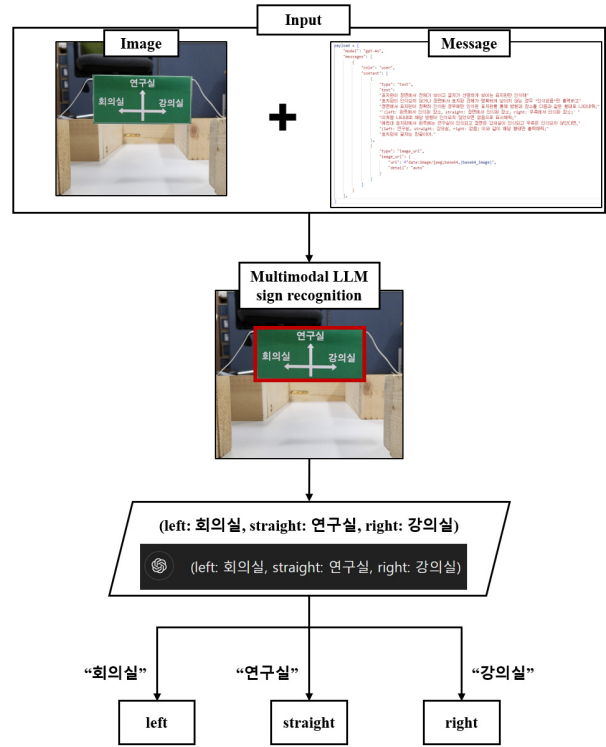
작성된 프롬프트와 장면의 이미지를 MLLM에 입력하여 방향과 장소에 대한 정보를 일정한 구조의 응답으로 받는다. 받은 응답은 각 방향에 저장한다. 이 동작 예시와 응답의 형태는 [Fig. 4]에 나타내었다. 방향 표지판이 인식되면 입력되는 이미지마다 API 응답의 방향에 대한 장소를 순차적으로 저장한다. 즉, {left, straight, right}에 인식된 장소가 각각 저장된다.

```

"model": "gpt-4o",
"messages": [
  {
    "role": "user",
    "content": [
      {
        "type": "text",
        "text": "표지판이 장면에서 전체가 보이고 글자가 선명하게 보이는 표지판만 인식해" → Sign recognition
      },
      {
        "type": "text",
        "text": "표지판이 인식되지 않거나 장면에서 표지판 전체가 명확하게 않은 경우 '인식없음'만 출력하고"
      },
      {
        "type": "text",
        "text": "장면에서 표지판이 정확히 인식된 경우에만 인식된 표지판의 방향과 장소를 다음과 같은 형태로 나타내줘;"
      },
      {
        "type": "text",
        "text": "'(left: 왼쪽에서 인식된 장소, straight: 장면에서 인식된 장소, right: 오른쪽에서 인식된 장소)'"
      },
      {
        "type": "text",
        "text": "'이처럼 나타내며 해당 방향이 인식되지 않았으면 빈문으로 표시해줘;"
      }
    ]
  },
  {
    "role": "assistant",
    "content": "왼편에 표지판에서 왼쪽에는 연구실이 인식되고 정면은 강의실이 인식되고 우측은 인식되지 않았다면,"
  },
  {
    "role": "assistant",
    "content": "'(left: 연구실, straight: 강의실, right: 없음) 이와 같이 해당 형태만 출력해줘;"
  },
  {
    "role": "text",
    "text": "표지판의 글자는 한글이야." → Sign language
  }
]
},
{
  "type": "image_url",
  "image_url": {
    "url": "#data:image/jpeg;base64,(base64_image)",
    "detail": "auto"
  }
}
]

```

[Fig. 3] Prompt writing structure: 1) sign recognition, 2) output format, 3) examples, and 4) sign language



[Fig. 4] Example of saving a location in a direction: when images and messages are entered into the MLLM, the MLLM provides recognition information about the direction and location and saves the output for each direction

5.2 Topological Map 작성

MLLM의 응답을 이용하여 topological map을 작성하는 방법 [Fig. 5]의 의사코드를 통해 제시한다. 로봇이 주행하며 얻은 장면 이미지에 Unsharp Mask를 적용한 이미지 $I_0, I_1, \dots, I_{n-1}, I_n$ 와 로봇의 헤딩 값에 해당하는 $\psi_0, \psi_1, \dots, \psi_{n-1}, \psi_n$ 이 입력된다(line 2-3). 입력된 이미지는 MLLM API에 전달되고 (line 6) API에서는 이미지 정보에 대한 응답 R_{cur} 를 준다(line 7). 여기서 이전 응답은 R_{pre} , 현재 응답은 R_{cur} 로 나타내며 이전 응답에서 로봇의 헤딩 값은 ψ_{pre} , 현재 응답에서의 로봇의 헤딩 값은 ψ_{cur} 로 나타내었다. 첫 번째 응답에서는 R_{pre} 값과 ψ_{pre} 은 비워져 있으므로 첫 번째 응답이 R_{pre} 에 저장되고, 첫 번째 응답에서의 로봇의 헤딩 값이 ψ_{pre} 에 저장된다(line 8-11). 이후 현재 API 응답과 이전 API 응답을 비교하여 이전에 관측되지 않았던 새로운 방향 표지판인지 판단한다(line 12). 이전에 관측되지 않은 방향 표지판인지 판단하는 로직은 이전 방향 표지판 인식과 현재 방향 표지판 인식을 비교하여 각 방향에 저장된 장소가 모두 바뀌었을 경우로 판단한다. 이전에 관측되지 않은 방향 표지판이라고 판단하게 되면 방향 결정 단계로 넘어간다(line 14). 방향 결정 단계에서는 관측되지 않았던 방향 표

Algorithm 1 Topological Map generation

```

1 Initialize  $R_{pre}$  to None and  $\psi_{pre}$  to None
2 Input image  $I_0, I_1, \dots, I_{n-1}, I_n$ 
3 Input yaw  $\psi_0, \psi_1, \dots, \psi_{n-1}, \psi_n$ 
4 for  $k$  from 0 to  $n$  do
5   Input  $I_k, \psi_k$ 
6   Send  $I_k$  to API
7   Receive  $R_{cur}$  from API
8   if  $R_{pre}$  is None then
9     update  $R_{pre}$  to  $R_{cur}$  and  $\psi_{pre}$  to  $\psi_k$ 
10    continue
11  end if
12  Compare  $R_{pre}$  and  $R_{cur}$ 
13  if detect new sign then
14    direction decision  $D_{dec} = \psi_{cur} - \psi_{pre}$ 
15    if  $-\frac{\pi}{2} - \alpha \leq D_{dec} \leq -\frac{\pi}{2} + \alpha$  then
16      shifting right and generating Node
17    else if  $\frac{\pi}{2} - \alpha \leq D_{dec} \leq \frac{\pi}{2} + \alpha$  then
18      shifting left and generating Node
19    else if  $-\alpha < D_{dec} < \alpha$  then
20      shifting straight and generating Node
21    else
22      continue
23    end if
24    update  $R_{pre}$  to  $R_{cur}$  and  $\psi_{pre}$  to  $\psi_k$ 
25  else
26    continue
27  end if
28 end for

```

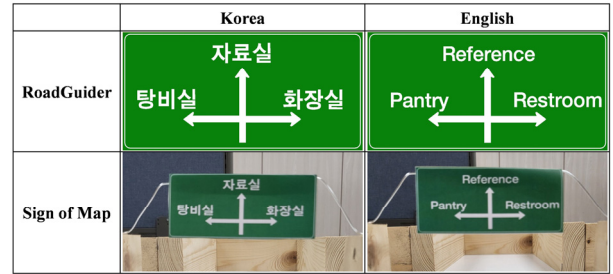
[Fig. 5] Topological Map generation pseudo code: the pseudo code represents the algorithm from scene input to node generation in topological map generation

지판이 인식된 시점의 로봇의 헤딩 값과 이전 노드를 생성했을 때의 로봇의 헤딩 값의 차이를 이용하여 방향을 결정한다. D_{dec} 는 방향을 결정하는 값을 나타내며 현재 응답에서의 로봇의 헤딩 값과 이전 응답에서의 로봇의 헤딩 값의 차이인 $\psi_{cur} - \psi_{pre}$ 이다. 로봇의 헤딩 값은 초기 지점에서 z축을 기준으로 얼마나 회전했는지 알 수 있기 때문에 방향 표지판을 새로 인식한 시점과 이전 방향 표지판을 인식했던 시점의 로봇의 헤딩 값의 차이를 이용해서 로봇의 방향을 알 수 있다. α 는 로봇의 방향을 보정해주는 상수이다. 로봇의 주행은 비선형성을 가지며, wheel odometry가 differential drive kinematics를 통해 계산되기 때문에 실제 환경에서는 오차누적이 발생한다. 따라서 오차보정 값인 α 를 설정해주었다. 본 연구에서는 $\pi/4$ 로 설정하였다. 방향 결정 값인 D_{dec} 의 값에 따라 right, straight, left 중 방향을 결정한다. 방향이 결정되면 해당 방향으로 이동하고 이동한 방향을 정면으로 하여 인식된 방향 표지판의 정보를 통해 노드를 생성한다(line 15-24). 이는 모든 이미지 입력에 대한 동작이 마칠 때까지 반복한다(line 4-28).

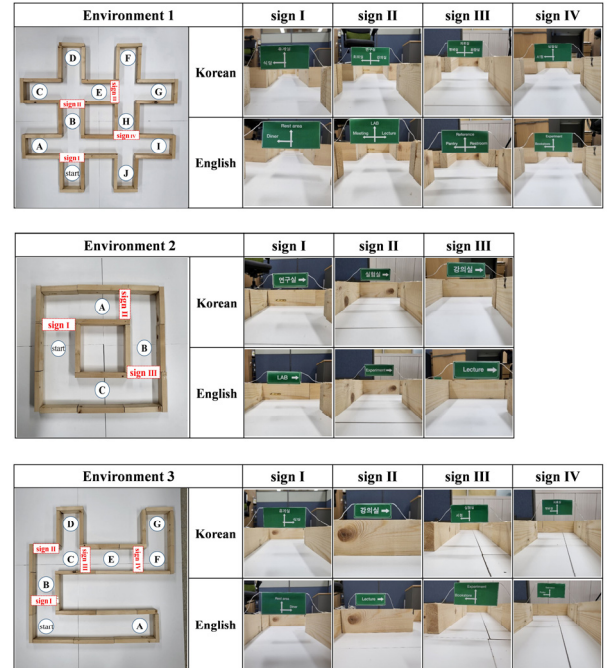
6. 실험

6.1 실험 환경

먼저, 로봇은 E-puck2^[29]를 사용하였고, [Fig. 1]에서 제시한 3가지 환경을 구성하여 실험을 진행하였다. 카메라는 로봇의 부착되어 있는 카메라를 사용하였다. 카메라 성능은 최대 이미지 전송 속도는 30 fps, 해상도는 640×480 이다. 방향 표지판은 도로표지 안내 시스템 RoadGuider^[30]로 제작해 부착했다. [Fig. 6]에는 제작한 방향 표지판의 예시가 있다. RoadGuider를 통해 제작할 수 있는 방향 표지판은 실제 도로교통 방향 표지판과 동일하게 제작할 수 있으며 초록색 배경에 흰색 글씨



[Fig. 6] Example of a sign: the top two signs are signs created using RoadGuider^[36], while the bottom two are signs attached to the actual experimental environment



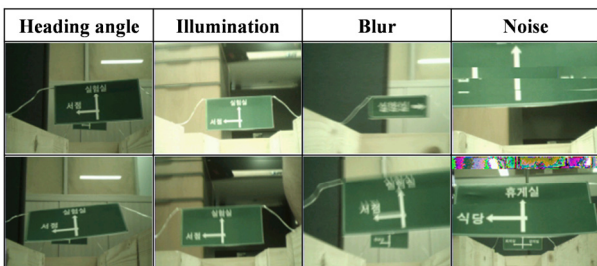
[Fig. 7] Experimental environments: three environments were created with the structure of [Fig. 1], and Korean sign environment and an English sign environment were configured

로 쓰여진다. 방향은 좌측, 정면, 우측을 그릴 수 있다. [Fig. 7]은 실제 구성된 환경이다. 총 3개의 환경을 로봇이 주행할 수 있도록 목재 재료를 가공하여 벽을 만들어 구성했다. 각 환경마다 방향 표지판의 위치를 sign I에서 sign IV로 표기하였고, 방향 표지판의 정보에 해당하는 장소를 알파벳 A에서 G로 표현하였다. [Fig. 7]에서 왼쪽은 구성된 환경의 top view를 기록하였으며, 오른쪽은 각 sign 위치에 부착한 방향 표지판을 기록하였다.

6.2 MLLM별 인식 성능 비교

이 절에서는 각 MLLM별 인식 성능을 비교한다. 성능 비교에 사용되는 MLLM은 ChatGPT- 4o, ChatGPT-4, Claude 3.5 Sonnet, Gemini로 총 네 개의 MLLM을 비교한다. 비교 방법은 동일한 내용으로 작성된 프롬프트와 로봇이 환경을 주행하며 얻은 이미지 중 다양한 각도, 조명 변화, 블러, 노이즈가 있는 이미지를 영어와 한국어로 각각 30장씩 MLLM에 입력한다. 입력 이미지의 예시는 [Fig. 8]에 기록하였다. 각 MLLM의 성능 측정을 위해 인식 방해 요소가 있는 이미지들을 사용하여 인식 정확도를 평가한다. 이 입력에 대한 MLLM의 출력과 입력 이미지를 비교하여 인식 정확도를 기록한다. 인식은 방향 표지판에 나타난 방향과 그 방향에 나타난 장소가 일치하도록 인식하였으면 success로 기록하였고, 방향 표지판이 있는 이미지지만 LLM에서 인식하지 못했을 경우에는 failure, 방향 표지판을 인식하여 응답을 주었지만, 실제 방향 표지판의 방향 또는 장소와 일치하지 않을 경우에는 false positive로 기록하였다. 인식률은 인식된 횟수를 총 입력 횟수로 나누어서 계산하였다. MLLM별 프롬프트의 예시는 [Fig. 9]에 기록하였다. 그리고 해당 실험 결과는 [Table 1]에 나타내었다.

해당 실험의 결과를 통해 한국어와 영어 모두 ChatGPT-4o의 인식률이 가장 높은 것을 확인할 수 있다. 다른 MLLM들도 영어에 대한 인식 성능 지표는 높았으나, 방향에 대한 정보나 한국어 인식 성능은 ChatGPT-4o보다 낮은 성능을 보였다.



[Fig. 8] Input data for the MLLM recognition performance comparison experiment



ChatGPT-4o

Response

이미지를 분석한 결과, 다음과 같은 표지판이 나타났습니다.

- 왼쪽: 회의실
- 정면: 연구실
- 오른쪽: 강의실

따라서 다음과 같이 방향과 장소를 나타낼 수 있습니다.
(left: 회의실, straight: 연구실, right: 강의실)

ChatGPT-4

Response

The sign in the image you uploaded displays directions in Korean. Here's the breakdown of the locations and their respective direction as indicated on the sign

(left: 회의실, straight: 연구실, right: 강의실)

Claude 3.5 Sonnet

Response

이미지에 표지판이 인식되었습니다. 표지판의 내용을 분석한 결과는 다음과 같습니다.
(left: 회의실, straight: 없음, right: 화장실)

Gemini

Response

The image shows a rectangular sign with three Korean words written on it:

- left: Research Room
- front: Meeting Room
- right: Lecture Room

Sign description:
Based on the image, the sign can be described as follows:
• left: Research Room
• front: Meeting Room
• right: Lecture Room

[Fig. 9] Example prompts for each LLM

6.3 프롬프트별 인식 성능 비교

이 절에서는 ChatGPT- 4o의 프롬프트 내용에 따른 방향 표지판 인식 성능을 비교한다. 그리고 구현한 알고리즘을 통

[Table 1] Recognition rate by LLM

Model	Recognition	Korean	English
		Rate (count/total)	
ChatGPT-4o ^[6]	Success	73.3% (22/30)	86.7% (26/30)
	Failure	6.7% (2/30)	6.7% (2/30)
	False positive	20% (6/30)	6.7% (2/30)
ChatGPT-4 ^[6]	Success	10% (3/30)	80% (24/30)
	Failure	36.7% (11/30)	6.7% (2/30)
	False positive	53.3% (16/30)	13.3% (4/30)
Claude 3.5 Sonnet ^[7]	Success	23.3% (7/30)	26.7% (8/30)
	Failure	13.3% (4/30)	0% (0/30)
	False positive	63.3% (19/30)	73.3% (22/30)
Gemini ^[8]	Success	6.7% (2/30)	16.7% (5/30)
	Failure	16.7% (5/30)	23.3% (7/30)
	False positive	76.7% (23/30)	60% (18/30)

해 작성된 topological map이 실제 환경과 얼마나 유사한지 유사도 측정을 통해 확인한다. 프롬프트 내용은 [Table 2]에 나와 있다.

프롬프트는 다음 세 가지로 구성했다. 첫째는 방향 표지판에 인식되는 언어 지정, 둘째는 방향 표지판을 인식하는 조건 지정, 셋째는 방향 표지판에 인식되는 언어와 방향 표지판을 인식하는 조건 지정한다. 총 3가지의 프롬프트 형식으로 인식 정확도를 비교한다. 해당 실험의 결과는 [Table 3]에 기록하였다. [Table 3]에서 방향 표지판의 글자를 지정하고 인식되는 조건을 지정한 프롬프트의 응답 인식 성능이 가장 좋은 것을 확인할 수 있다.

해당 프롬프트별 인식 성능 비교 실험에서 작성된 topological map과 기존 환경이 얼마나 유사한지 측정한다. 본 논문에서는 옛지와 노드의 유사도를 확인할 수 있는 Jaccard Similarity^[31]를 이용해 유사도 측정을 하였다.

Jaccard Similarity 수식은 다음과 같다.

$$\text{Node Similarity: } J(N_1, N_2) = \frac{|N_1 \cap N_2|}{|N_1 \cup N_2|} \quad (1)$$

$$\text{Edge Similarity: } J(E_1, E_2) = \frac{|E_1 \cap E_2|}{|E_1 \cup E_2|} \quad (2)$$

$$\text{Average Similarity} = \frac{J(N_1, N_2) + J(E_1, E_2)}{2} \quad (3)$$

[Table 2] Prompt input type: type 1 specifies the recognition language, type 2 specifies the recognition condition, and type 3 specifies the recognition language and recognition condition

	Prompt
Specifies the recognition language (type 1)	표지판을 인식해 표지판이 인식되지 않거나 정면에서 표지판 전체가 명확하게 보이지 않는 경우 ‘인식없음’만 출력하고 정면에서 표지판이 정확히 인식된 경우에만 인식된 표지판을 통해 방향과 장소를 다음과 같은 형태로 나타내줘 (left: 왼쪽에서 인식된 장소, straight: 정면에서 인식된 장소, right: 우측에서 인식된 장소) 예컨대 표지판에서 왼쪽에는 연구실이 인식되고 정면은 강의실이 인식되고 우측은 인식되지 않았다면, (left: 연구실, straight: 강의실, right: 없음) 이와 같이 해당 형태만 출력해줘 표지판의 글자는 한글이야.
Specifies the recognition condition (type 2)	표지판이 정면에서 전체가 보이고 글자가 선명하게 보이는 표지판만 인식해 표지판이 인식되지 않거나 정면에서 표지판 전체가 명확하게 보이지 않는 경우 ‘인식없음’만 출력하고 정면에서 표지판이 정확히 인식된 경우에만 인식된 표지판을 통해 방향과 장소를 다음과 같은 형태로 나타내줘 (left: 왼쪽에서 인식된 장소, straight: 정면에서 인식된 장소, right: 우측에서 인식된 장소) 예컨대 표지판에서 왼쪽에는 연구실이 인식되고 정면은 강의실이 인식되고 우측은 인식되지 않았다면, (left: 연구실, straight: 강의실, right: 없음) 이와 같이 해당 형태만 출력해줘
Specifies the recognition language and recognition condition (type 3)	표지판이 정면에서 전체가 보이고 글자가 선명하게 보이는 표지판만 인식해 표지판이 인식되지 않거나 정면에서 표지판 전체가 명확하게 보이지 않는 경우 ‘인식없음’만 출력하고 정면에서 표지판이 정확히 인식된 경우에만 인식된 표지판을 통해 방향과 장소를 다음과 같은 형태로 나타내줘 (left: 왼쪽에서 인식된 장소, straight: 정면에서 인식된 장소, right: 우측에서 인식된 장소) 예컨대 표지판에서 왼쪽에는 연구실이 인식되고 정면은 강의실이 인식되고 우측은 인식되지 않았다면, (left: 연구실, straight: 강의실, right: 없음) 이와 같이 해당 형태만 출력해줘 표지판의 글자는 한글이야.

[Table 3] Prompt input comparison experiment result

Prompt type	Environment	Korean			English		
		Success rate (count/total)	Failure (count/total)	False positive (count/total)	Success rate (count/total)	Failure (count/total)	False positive (count/total)
1	Env 1	82.9% (34/41)	4.9% (2/41)	12.2% (5/41)	82.9% (34/41)	0% (0/41)	17.1% (7/41)
	Env 2	100% (36/36)	0% (0/36)	0% (0/36)	100% (39/39)	0% (0/39)	0% (0/39)
	Env 3	70.6% (24/34)	5.9% (2/34)	23.5% (8/34)	78.4% (29/37)	5.4% (2/37)	16.2% (6/37)
2	Env 1	80.5% (33/41)	4.9% (2/41)	14.6% (6/41)	82.9% (34/41)	7.3% (3/41)	9.8% (4/41)
	Env 2	83.3% (30/36)	13.9% (5/36)	2.8% (1/36)	92.3% (36/39)	7.7% (3/39)	0% (0/39)
	Env 3	76.5% (26/34)	5.9% (2/34)	17.6% (6/34)	78.4% (29/37)	8.1% (3/37)	13.5% (5/37)
3	Env 1	90.2% (37/41)	0% (0/41)	9.8% (4/41)	90.2% (37/41)	9.8% (4/41)	0% (0/41)
	Env 2	94.4% (34/36)	0% (0/36)	5.6% (2/36)	100% (39/39)	0% (0/39)	0% (0/39)
	Env 3	91.2% (31/34)	0% (0/34)	8.8% (3/34)	100% (37/37)	0% (0/37)	0% (0/37)

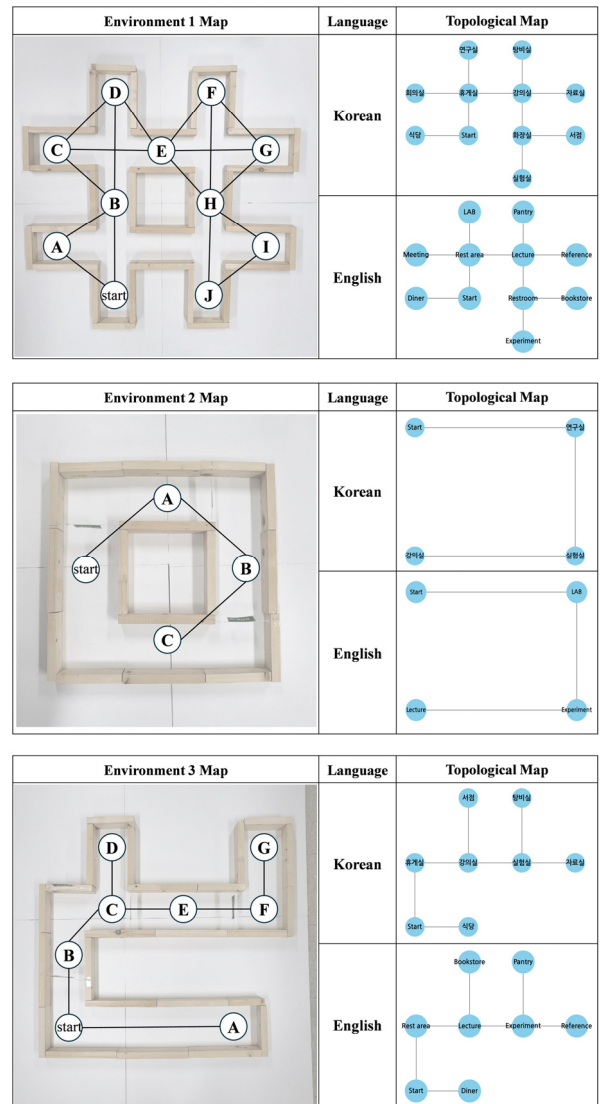
[Table 4] Jaccard Similarity by prompt contents

Prompt type	Environment	Similarity (Korean)			Similarity (English)		
		Node	Edge	Average	Node	Edge	Average
1	Env1	0.58	0.32	0.45	1	0.56	0.78
	Env2	1	1	1	1	1	1
	Env3	0.67	0.3	0.48	0.78	0.17	0.47
2	Env1	0.92	0.56	0.74	1	0.56	0.78
	Env2	0.50	0.33	0.42	1	1	1
	Env3	0.88	0.44	0.66	0.80	0.60	0.70
3	Env1	1	0.56	0.78	1	0.56	0.78
	Env2	1	1	1	1	1	1
	Env3	1	1	1	1	1	1

식 (1)과 식 (2)는 각각 노드, 엣지의 유사도를 나타낸다. 이 노드 유사도와 엣지 유사도의 평균을 나타내는 식 (3)을 통해 실제 환경과 그려진 topological map의 유사도를 평가한다. 프롬프트 작성 내용별 Similarity가 계산된 결과는 [Table 4]에 기록되어 있다. 각 환경별로 유사도의 평균이 가장 높은 값은 bold로 표현하였다.

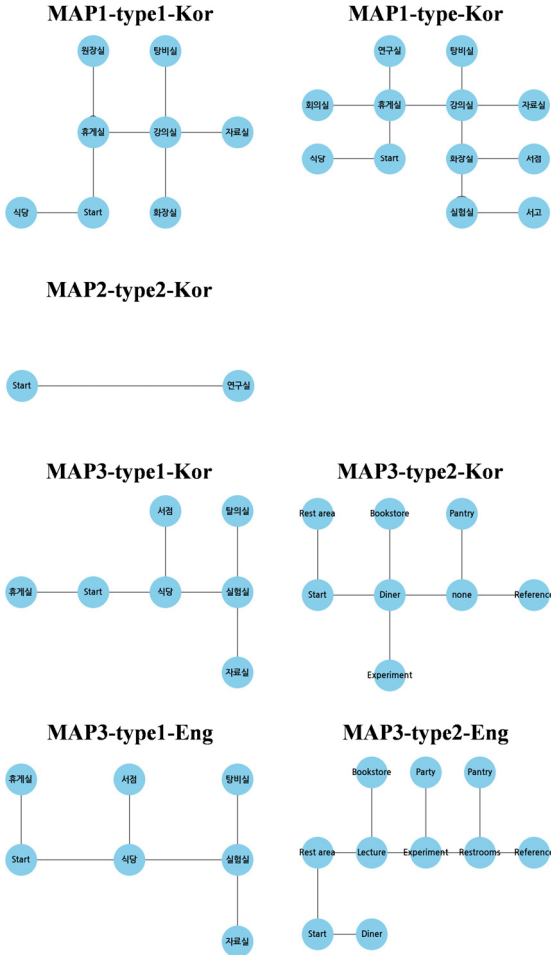
[Fig. 10]에는 구성된 환경과 본 연구에서 제안하는 방식을 통해 작성된 topological map을 기록했다. 좌측은 주어진 환경에서 적용 가능한 엣지를 연결한 이미지이며, 우측은 구현한 알고리즘을 통해 작성된 topological map 중 유사도가 가장 높은 topological map이다. 작성된 topological map이 실제 환경의 노드 사이의 관계와 상대적인 위치 정보를 담고 있는 것을 확인할 수 있다.

[Fig. 11]은 해당 실험에서 방향 표지판을 인식하지 못하거나 방향 표지판의 정보를 잘못 인식하는 상황이 지속적으로 발생하게 될 경우 생성되는 부정확한 topological map이다. ‘환경-프롬프트 타입-언어’로 부정확하게 그려진 topological map을



[Fig. 10] Real environment and generated topological map: the left side shows the actual configured environment, and the right side shows the topological map drawn for each language

나타내었다. 실제 환경과 비교하여 보면 노드 명이 다르거나 엣지가 부정확하게 그려진 것을 확인할 수 있다.



[Fig. 11] Incorrectly created Topological Map: the map result drawn in the form of ‘Map drawn according to environment-Prompt type-Language’ was recorded

6.4 동일 데이터 및 프롬프트에 대한 반복 실험

이 절에서는 topological map 작성 결과가 일관성 있게 나타나는지 확인한다. 실험 환경에서 수집한 동일 데이터와 동일 프롬프트를 입력으로 하여 topological map 작성을 3회 반복 시행하여 각 시행별로 인식률을 측정한다. 해당 실험 결과는 [Table 5]에 기록되어 있다. Environment1, Environment2, Environment3에 대해 측정하고 방향 표시판의 방향과 장소에 대해 올바르게 식별했다면 success, 방향 표시판이 있는 이미지지만 방향 표시판을 인식하지 못했다면 failure, 방향 표시판을 인식했으나 방향이나 장소 정보가 일치하지 않을 때는 false positive로 기록하였다. 실험 결과는 각 trial 별로 인식 정확도가 거의 동일하게 나타났다. 따라서 MLLM의 응답이 일관성 있게 나타남을 입증한다.

6.5 Localization 결과

작성된 topological map의 활용성을 확인하기 위해 localization 실험을 진행하였다. 로봇이 주행하며 얻은 장면의 이미지를 MLLM을 통해 인식했다. 인식된 정보가 topological map 작성에서 노드를 생성한 시점의 응답과 일치하면 해당 노드로 localization 하며 로봇의 odometry 값을 통해 주행한다. 해당 실험 결과는 [Fig. 12]에 나타내었다. 각 환경에 대해 localization 실험을 진행하였으며, 로봇의 실제 주행의 top view와 localization 결과를 나타낸다. Topological map은 노드와 노드 사이의 관계 즉, 위상학적 관계를 나타내기 때문에 실제 위치에 적합하게 localization 되고 있음을 확인할 수 있었다. [Table 6]는 localization의 success rate이다. Topological map을 통해 localization을 하는 로봇의 위치와 top view의 로봇의 위치를 비교하여 올바른 노드에 위치하거나 노드와 노드 사이에 해당하는 엣지에 일치하게 위치하였으면 success로 기록하였다.

[Table 5] Experimental results using identical prompts

trial	environment	Korean			English		
		success rate (count/total)	failure rate (count/total)	false positive (count/total)	success rate (count/total)	failure rate (count/total)	false positive (count/total)
1	Env 1	90.2% (37/41)	0% (0/41)	9.8% (4/41)	90.2% (37/41)	9.8% (4/41)	0% (0/41)
	Env 2	94.4% (34/36)	0% (0/36)	5.6% (2/36)	100% (39/39)	0% (0/39)	0% (0/39)
	Env 3	91.2% (31/34)	0% (0/34)	8.8% (3/34)	100% (37/37)	0% (0/37)	0% (0/37)
2	Env 1	85.4% (35/41)	0% (0/41)	14.6% (6/41)	87.8% (36/41)	0% (0/41)	12.2% (5/41)
	Env 2	88.9% (32/36)	11.1% (4/36)	0% (0/36)	97.4% (38/39)	2.6% (1/39)	0% (0/39)
	Env 3	79.4% (27/34)	0% (0/34)	20.6% (7/34)	89.2% (33/37)	2.7% (1/37)	8.1% (3/37)
3	Env 1	87.8% (36/41)	0% (0/41)	12.2% (5/41)	90.2% (37/41)	9.8% (4/41)	0% (0/41)
	Env 2	88.9% (32/36)	11.1% (4/36)	0% (0/36)	97.4% (38/39)	2.6% (1/39)	0% (0/39)
	Env 3	82.4% (28/34)	8.8% (3/34)	8.8% (3/34)	94.6% (35/37)	0% (0/37)	5.4% (2/37)

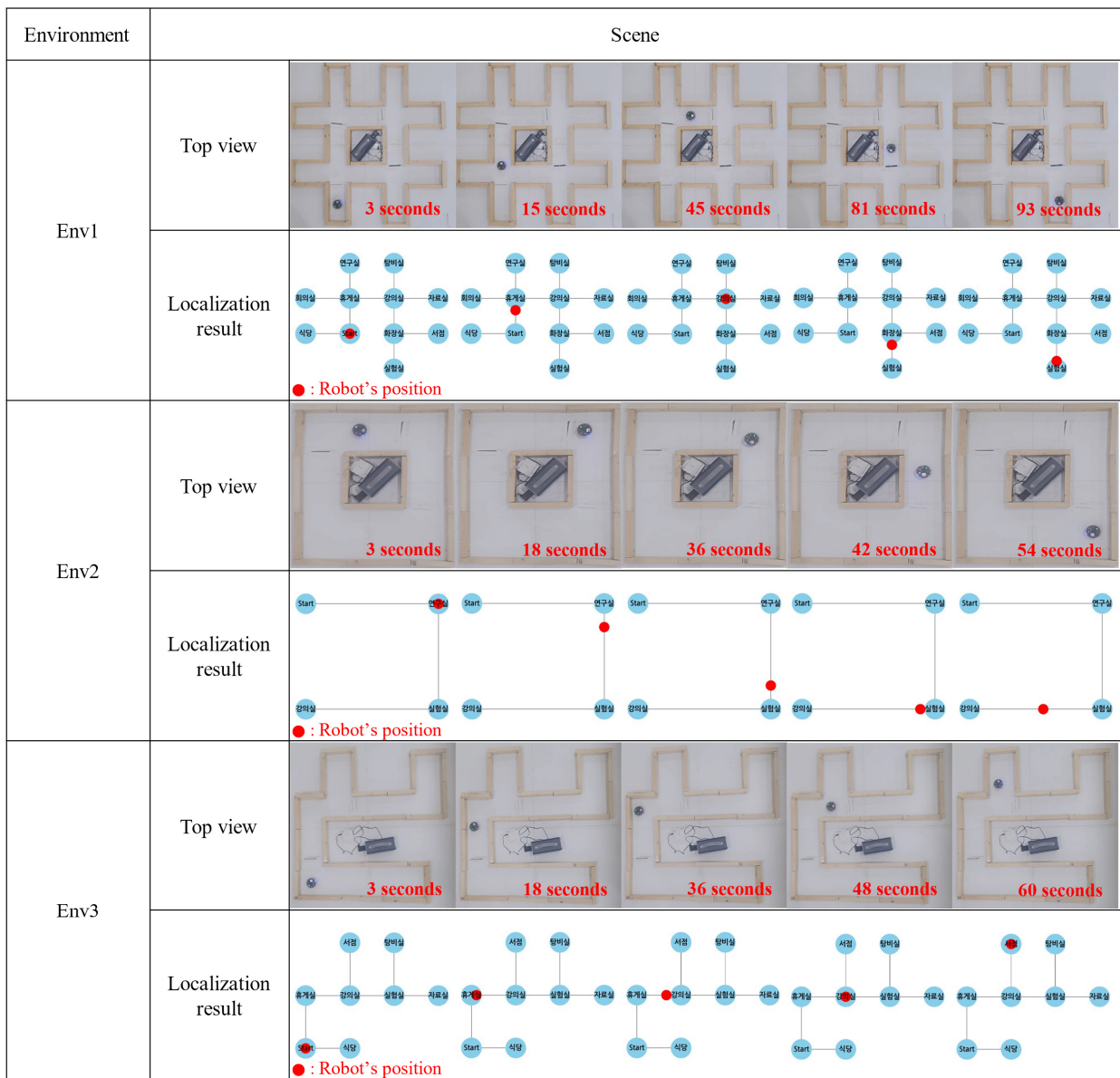
[Table 6] Localization success rate

	Success rate (success/total)
Env1	87.2% (34/39)
Env2	89.5% (17/19)
Env3	86.2% (25/29)

7. 결 론

본 연구는 방향 표지판 정보를 이용해 topological map을 작성하는 방법을 제안한다. 방향 표지판의 인식은 MLLM을 사용

하였고 MLLM 종류별 비교를 통해 ChatGPT-4o의 방향 표지판 인식 성능이 가장 뛰어난 것을 확인하였다. 또한, 실험을 통해 실제 환경과 그려진 topological map의 유사도를 확인하였다. 본 연구에서 제안하는 방법은 장소 간의 위상학적 관계를 파악하여 방향 정보를 포함한 topological map을 작성할 수 있다. 또한, 방향 표지판이 인식될 때 노드를 생성함으로써 topological map에서의 장소 인식 시점을 명확하게 할 수 있다. 향후 연구에서는 방향 표지판을 측면에서도 인식하여 topological map을 작성할 수 있는 알고리즘을 구현할 계획이다. 또한, 이 방식을 SLAM에 적용하여 실시간으로 지도를 작성하면서 위치 파악



[Fig. 12] Results of localization experiments: localization was performed using the generated topological map, and images of the robot's movement were captured from a top view, displaying the corresponding times for each environment. The robot's position at each time is indicated with a red point as the localization result

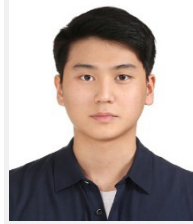
이 가능한 방향 표지판 인식을 활용한 Topological SLAM을 연구할 계획이다.

References

- [1] S. Thrun, "Learning metric-topological maps for indoor mobile robot navigation," *Artificial intelligence*, vol. 99, no. 1, pp. 21-71, Feb., 1998, DOI: 10.1016/S0004-3702(97)00078-7.
- [2] B. Kaleci, Ç. M. Senler, O. Parlaktuna, and U. Gürel, "Constructing Topological Map from Metric Map Using Spectral Clustering," *2015 IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI)*, Vietri sul Mare, Italy, pp. 139-145, 2015, DOI: 10.1109/ICTAI.2015.33.
- [3] A. Yang, Y. Luo, L. Chen, and Y. Xu, "Survey of 3D map in SLAM: Localization and navigation," *Advanced Computational Methods in Life System Modeling and Simulation*, pp. 410-420, Aug., 2017, DOI: 10.1007/978-981-10-6370-1_41.
- [4] Y. C. N. Sousa and H. F. Bassani, "Topological Semantic Mapping by Consolidation of Deep Visual Features," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 4110-4117, Apr., 2022, DOI: 10.1109/LRA.2022.3149572.
- [5] K. Song, W. Liu, G. Chen, X. Xu, and Z. Xiong, "FHT-Map: Feature-Based Hybrid Topological Map for Relocalization and Path Planning," *IEEE Robotics and Automation Letters*, vol. 9, no. 6, pp. 5401-5408, Jun., 2024, DOI: 10.1109/LRA.2024.3392493.
- [6] OpenAI, *ChatGPT*, [Online], <https://chat.openai.com>, Accessed: Jul. 3, 2024.
- [7] Anthropic, *Claude*, [Online], <https://www.anthropic.com>, Accessed: Jul. 4, 2024.
- [8] Google DeepMind, *Gemini*, [Online], <https://www.deepmind.com>, Accessed: Jul. 5, 2024.
- [9] S. Thrun and A. Bücken, "Integrating Grid-based and Topological Maps for Mobile Robot Navigation," *AAAI/IAAI*, vol. 13, pp. 944-951, Aug., 1996, [Online], <https://cdn.aaai.org/AAAI/1996/AAAI96-140.pdf>.
- [10] H. Choset and J. Burdick, "Sensor Based Motion Planning: The Hierarchical Generalized Voronoi Graph," Ph.D. dissertation, California Institute of Technology, Pasadena, CA, USA, 1996, DOI: 10.7907/49ee-a204.
- [11] E. G. Tsardoulis, A. T. Serafi, M. N. Papazoglou, and L. Petrou, "Construction of Minimized Topological Graphs on Occupancy Grid Maps Based on GVD and Sensor Coverage Information," *Journal of Intelligent and Robotic Systems*, vol. 75, pp. 457-474, Feb., 2014, DOI: 10.1007/s10846-013-9995-3.
- [12] P. Buschka and A. Saffiotti, "A virtual sensor for room detection," *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Lausanne, Switzerland, pp. 637-642, 2002, DOI: 10.1109/IRDS.2002.1041463.
- [13] O. M. Mozos and W. Burgard, "Supervised Learning of Topological Maps using Semantic Information Extracted from Range Data," *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Beijing, China, pp. 2772-2777, 2006, DOI: 10.1109/IROS.2006.282058.
- [14] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and computing*, vol. 17, pp. 395-416, Dec., 2007, DOI: 10.1007/s11222-007-9033-z.
- [15] F. de la Paz López and J. R. Álvarez-Sánchez, "Topological maps for robot's navigation: a conceptual approach," *Bio-Inspired Applications of Connectionism*, pp. 459-467, Jun., 2001, DOI: 10.1007/3-540-45723-2_55.
- [16] A. Angeli, S. Doncieux, J.-A. Meyer, and D. Filliat, "Incremental vision-based topological SLAM," *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Nice, France, pp. 1031-1036, 2008, DOI: 10.1109/IROS.2008.4650675.
- [17] Z. Li, J. D. Wegner, and A. Lucchi, "Topological Map Extraction From Overhead Images," *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Republic of Korea, pp. 1715-1724, 2019, DOI: 10.1109/ICCV.2019.00180.
- [18] F. Blochliger, M. Fehr, M. Dymczyk, T. Schneider, and R. Siegwart, "Topomap: Topological Mapping and Navigation Based on Visual SLAM Maps," *2018 IEEE International Conference on Robotics and Automation (ICRA)*, Brisbane, QLD, Australia, pp. 3818-3825, 2018, DOI: 10.1109/ICRA.2018.8460641.
- [19] D. S. Chaplot, R. Salakhutdinov, A. Gupta, and S. Gupta, "Neural Topological SLAM for Visual Navigation," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, pp. 12872-12881, 2020, DOI: 10.1109/CVPR42600.2020.01289.
- [20] K. Rana, J. Haviland, S. Garg, J. Abou-Chakra, I. Reid, and N. Suenderhauf, "Sayplan: Grounding large language models using 3d scene graphs for scalable task planning," *Proceedings of the 7th Conference on Robot Learning*, PMLR, vol. 229, pp. 23-72, 2023, [Online], <https://proceedings.mlr.press/v229/rana23a.html>.
- [21] Z. Long, G. Killick, R. McCreadie, and G. Aragon-Camarasa, "RoboLLM: Robotic Vision Tasks Grounded on Multi-modal Large Language Models," *2024 IEEE International Conference on Robotics and Automation (ICRA)*, Yokohama, Japan, pp. 12428-12435, 2024, DOI: 10.1109/ICRA57147.2024.10610797.
- [22] H. Li, S. Yu, S. Zhang, and G. Tan, "Resolving Loop Closure Confusion in Repetitive Environments for Visual SLAM through AI Foundation Models Assistance," *2024 IEEE International Conference on Robotics and Automation (ICRA)*, Yokohama, Japan, pp. 6657-6663, 2024, DOI: 10.1109/ICRA57147.2024.10610083.
- [23] S. Garg, K. Rana, M. Hosseinzadeh, L. Mares, N. Sünderhauf, F. Dayoub, and I. Reid, "Robohop: Segment-based topological map representation for open-world visual navigation," *arXiv:2405.05792*, 2024, DOI: 10.48550/arXiv.2405.05792.
- [24] W. Zhang, Y. Guo, L. Niu, P. Li, C. Zhang, Z. Wan, J. Yan, F. U. D. Farrukh, and D. Zhang, "Lp-slam: Language-perceptive rgb-d slam system based on large language model," *arXiv:2303.10089*, 2023, DOI: 10.48550/arXiv.2303.10089.
- [25] C. Kassab, M. Mattamala, L. Zhang, and M. Fallon, "Language-

EXtended Indoor SLAM (LEXIS): A Versatile System for Real-time Visual Scene Understanding,” *2024 IEEE International Conference on Robotics and Automation (ICRA)*, Yokohama, Japan, pp. 15988-15994, 2024, DOI: 10.1109/ICRA57147.2024.10610341.

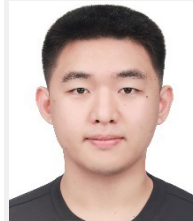
- [26] H. Deguchi, K. Shibata, and S. Taguchi, “Language to Map: Topological map generation from natural language path instructions,” *2024 IEEE International Conference on Robotics and Automation (ICRA)*, Yokohama, Japan, pp. 9556-9562, 2024, DOI: 10.1109/ICRA57147.2024.10611377.
- [27] J. M. Coughlan, and A. L. Yuille, “The manhattan world assumption: Regularities in scene statistics which enable bayesian inference,” *Advances in Neural Information Processing Systems 13(NIPS 2000)*, Denver, CO, USA, pp. 830-836, 2000, [Online], https://proceedings.neurips.cc/paper_files/paper/2000/file/90e1357833654983612fb05e3ec9148c-Paper.pdf.
- [28] A. Polesel, G. Ramponi, and V. J. Mathews, “Image enhancement via adaptive unsharp masking,” *IEEE Transactions on Image Processing*, vol. 9, no. 3, pp. 505-510, Mar., 2000, DOI: 10.1109/83.826787.
- [29] F. Mondada, M. Bonani, X. Raemy, J. Pugh, C. Cianci, A. Klaptocz, S. Magnenat, J.-C. Zufferey, D. Floreano, and A. Martinoli, “The e-puck, a robot designed for education in engineering,” *the 9th conference on autonomous robot systems and competitions*, Castelo Branco, Portugal, pp. 59-65, 2009, [Online], <https://infoscience.epfl.ch/entities/publication/d8bfd306-faf0-4dd3-9bf8-c00e455ade97>.
- [30] MOLIT/Road Sign Management Center, *RoadGuider*, [Online], https://www.roadsign.go.kr:8444/main_form.asp, Accessed: Jun. 20, 2024.
- [31] L. F. Costa, “Further generalizations of the Jaccard index,” *arXiv:2110.09619*, 2021, DOI: 10.48550/arXiv.2110.09619.



최창조

2019~현재 한국공학대학교 메카트로닉스 공학과(학사)

관심분야: Visual Recognition, Visual SLAM, Deep Learning



원종완

2020~현재 한국공학대학교 메카트로닉스 공학과(학사)

관심분야: MAPF, SLAM, 제어이론



어규호

2009 서울대학교 전기공학부(공학사)
 2011 서울대학교 전기공학부(공학석사)
 2016 서울대학교 전기공학부(공학박사)
 2016~2020 LG전자 로봇선행연구소 책임 연구원

2021 충북대학교 산업인공지능연구센터 초빙교수
 2022~현재 한국공학대학교 메카트로닉스공학부 교수

관심분야: 다개체 로봇 제어 및 협업, Fusion SLAM, 강화학습