

“사람에겐 당연한 ‘손잡이·버튼의 역할’, AI도 이해한다” GIST, 물체의 세부 부품과 기능까지 이해하는 로봇 AI 기술 개발

- AI학과 김의환 교수팀, 물체·부품의 위치부터 기능·가능한 행동까지 연결하는 ‘통합 3차원 장면 그래프’ 개발... 부품 검색 성능 83.6%로 기존 최고 성능보다 31.2%p 향상
- 이동형 로봇 ‘Stretch 3’에 적용해 주변 물체 파악부터 탐색·이동까지 구현
- 국제학술대회 ‘ECCV 2026’ 발표



▲ (왼쪽부터) GIST AI학과 김의환 교수, 김이름 석·박사통합과정생

광주과학기술원(GIST·지스트, 총장 임기철)은 **AI학과 김의환 교수 연구팀이 로봇이 주변의 물체뿐 아니라 손잡이·버튼 같은 세부 부품과 그 기능까지 파악할 수 있는 인공지능(AI) 기술을 개발했다고 밝혔다.**

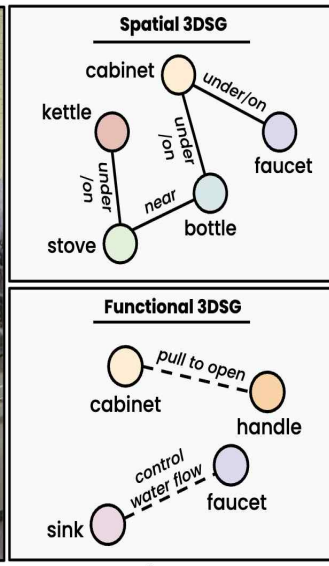
예를 들어 로봇이 전자레인지의 위치를 아는 것만으로는 음식을 데울 수 없다. 문을 열기 위해 손잡이를 찾아야 하고, 작동을 위해 버튼의 위치와 기능도 알아야 한다. 사람에게는 당연한 정보지만, 로봇이 실제 환경에서 사람의 지시를 수행하려면 이처럼 **물체와 그 주변의 세부적인 구성 요소, 기능을 함께 이해하는 것이 중요하다.**

연구팀은 이러한 문제를 해결하기 위해 **물체와 부품의 위치뿐 아니라 각각의 기능과 가능한 행동까지 하나로 연결해 보여주는 ‘통합 3차원 장면 그래프(Unified 3DSG, 이하 통합 3DSG)’를 제시하고, 실제 카메라 영상으로부터 이를 자동으로 구축하는 AI 기술 ‘OP3DSG(Open-Vocabulary Part-Aware 3D Scene Graph Generation)’을 개발했다.**

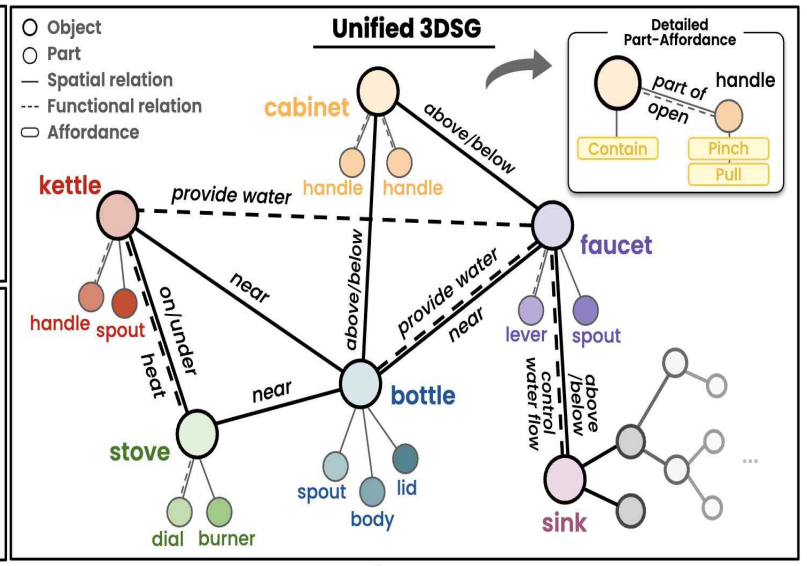
‘**3차원 장면 그래프(3D Scene Graph, 3DSG)**’는 로봇이 주변 환경을 이해하고 이동하거나 작업 순서를 계획할 수 있도록 공간 속 물체와 물체 사이의 관계를 연결해 표현하는 일종의 ‘**로봇용 3차원 지식지도**’다.



(a) 3D Scene



(b) Previous Scopes



(c) OP3DSG Scope

▲ 기존 3차원 장면 그래프와 '통합 3DSG' 비교. 기존 방식은 '어떤 물체가 어디에 있는지'를 나타내는 공간 정보와 '물체의 어느 부품이 어떤 기능을 하는지'를 나타내는 기능 정보를 각각 다뤘다. 연구팀이 제시한 '통합 3DSG'는 물체와 손잡이·버튼 같은 부품, 이들의 위치·기능과 가능한 행동까지 하나의 그래프로 연결해 표현한다.

하지만 기존 3DSG는 주로 '의자', '전자레인지'와 같은 물체 단위의 정보를 표현하는 데 초점이 맞춰져 있어, 물체를 구성하는 작은 부품이나 그 기능까지 구체적으로 표현하는 데 한계가 있었다.

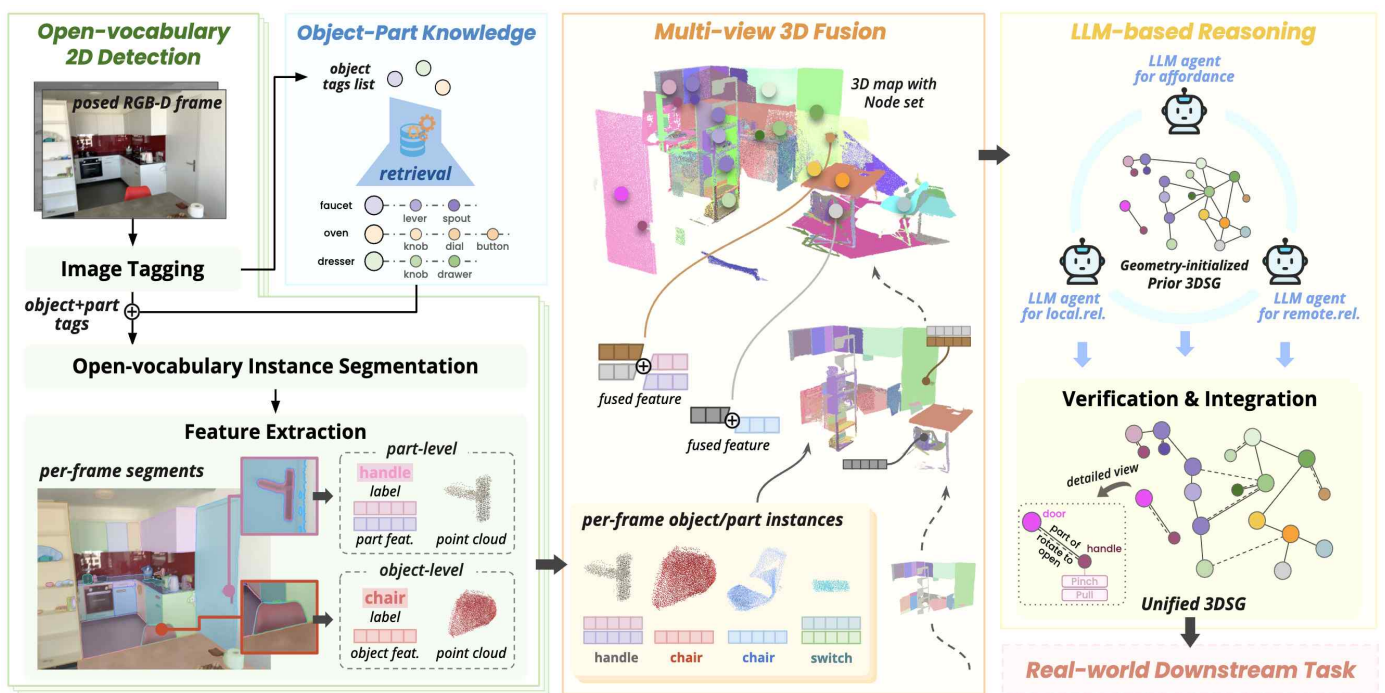
가령 사람에게 "의자 다리를 찾아라"라고 하면 의자의 어느 부분을 말하는지 쉽게 알 수 있지만, AI가 영상에서 이를 찾을 때는 의자 전체를 찾아내는 데 그치거나 서로 다른 각도에서 촬영된 같은 부품을 제대로 연결하지 못할 수 있다. 작은 부품은 크기가 작고 촬영 각도에 따라 모습이 크게 달라지기 때문이다.

연구팀이 제안한 OP3DSG는 실제 카메라 영상으로부터 통합 3DSG를 자동으로 구축하는 AI 기술이다.

먼저 영상에서 주요 물체를 찾은 뒤, 해당 물체에 포함될 것으로 예상되는 세부 부품을 함께 탐색한다. 예를 들어 오븐을 발견하면 버튼, 문, 손잡이 등 오븐을 구성하는 부품을 찾아야 할 대상으로 추가해 탐색하는 방식이다.

이후 여러 각도에서 촬영된 영상 정보를 3차원 공간으로 통합한다. 특히 작은 부품이 서로 다른 영상에서 동일한 부품인지 구분하기 위해 부품의 위치와 의미뿐 아니라 색상 정보도 함께 비교해 같은 부품에 대한 정보를 연결한다.

이렇게 물체와 부품의 위치 및 거리 정보를 바탕으로 기본적인 관계를 구성한 뒤에는 대규모 언어 모델(LLM)*을 활용해 부품과 물체 사이의 기능적 관계와 가능한 행동을 검토·보완한다.



▲ ‘통합 3DSG’를 자동으로 구축하는 OP3DSG 기술의 작동 과정. 카메라 영상에서 물체와 손잡이·버튼 같은 부품을 찾아 여러 방향에서 촬영한 정보를 3차원 공간에 합친다. 이어 실제 위치와 거리 등의 정보를 바탕으로 물체와 부품의 관계를 구성하고, 대규모언어모델(LLM)을 활용해 기능상 어떻게 연결되는지와 어떤 행동이 가능한지를 검토·보완해 ‘통합 3DSG’를 구축한다.

예를 들어 ‘전자레인지-손잡이’라는 공간적 관계에 더해 손잡이가 ‘문을 여는 데 사용된다’는 기능적 관계를 연결하고, 이를 바탕으로 로봇이 수행할 수 있는 행동까지 표현할 수 있도록 하는 것이다.

* **대규모언어모델(LLM):** 방대한 텍스트를 학습해 문장을 이해하고 생성하며, 지식과 관계를 추론하는 인공지능 모델이다.

연구팀은 새롭게 개발한 ‘유니그래프3D(UniGraph3D)’ 평가 체계를 활용해 OP3DSG의 성능을 검증했다.

그 결과, AI가 찾아낸 상위 3개 후보 안에 정답 부품이 포함되는 비율은 83.6%로 나타났다. 기존 비교 방법 가운데 가장 높은 성능인 52.4%보다 31.2%포인트 향상된 수치다.

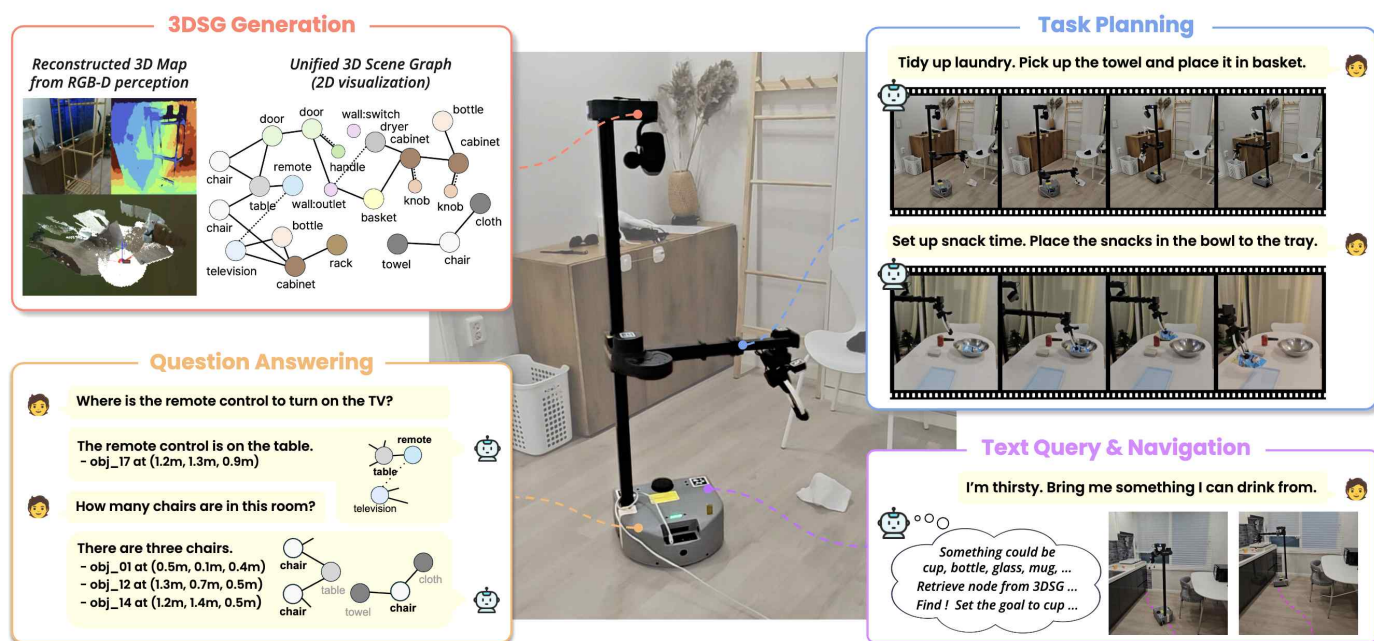
물체와 부품 사이의 공간적 관계를 파악하는 성능도 기존 최고 성능보다 7.9%포인트 높았으며, 기능적 관계를 파악하는 성능 역시 9.2%포인트 향상됐다.

연구팀은 개발한 기술을 이동형 로봇 ‘스트레치 3(Stretch 3)’에 적용해 실제 환경에서의 활용 가능성도 확인했다.

예를 들어 주변에 있는 의자의 개수를 파악하거나, TV를 켜기 위해 필요한 리모컨의 위치를 찾아 이동하는 작업이 가능했다.

또한 “빨래를 정리해 줘”라는 지시를 받으면 수건을 찾아 바구니에 넣는 작업 순서를 계획할 수 있었으며, “목이 마르다. 마실 때 쓸 수 있는 것을 가져다 달라”와 같이 구체적인 물체 이름을 직접 제시하지 않은 지시에도 기능적 관계를 바탕으로 필요한 물체를 찾아 해당 위치를 이동 목표로 설정할 수 있었다.

이는 사람이 로봇에게 물체의 정확한 이름이나 위치를 일일이 알려주지 않아도, 주변 환경에 있는 물체와 부품의 기능적 관계를 바탕으로 사람의 지시를 구체적인 작업으로 연결할 가능성을 보여준다.



▲ OP3DSG 기술로 구축한 ‘통합 3DSG’의 실제 로봇 활용. 이동형 로봇 ‘Stretch 3’가 실내를 탐색하며 구축한 ‘통합 3DSG’를 활용해 주변 환경에 관한 질문에 답하고, 사람의 지시를 구체적인 작업 순서로 나눠 계획한다. 또한 물체 이름을 직접 말하지 않고 필요한 기능이나 상황을 설명해도 관련 물체를 찾아 그 위치를 이동 목표로 정할 수 있다.

김의환 교수는 “이번 연구의 핵심은 로봇이 ‘무엇이 어디에 있는지’뿐 아니라 ‘어떤 부품을 사용할 수 있고, 어떻게 사용할 수 있는지’까지 이해하는 데 필요한 정보를 하나로 연결한 것”이라며 “앞으로 서비스 로봇이 복잡한 실제 환경을 이해하고, 사람의 자연어 지시를 구체적인 작업 계획과 이동으로 연결하는 데 기반 기술로 활용될 수 있을 것으로 기대한다”고 말했다.

GIST AI학과 김의환 교수가 지도하고 김이름 석·박사통합과정생이 제1저자로 수행한 이번 연구는 과학기술정보통신부·정보통신기획평가원 AI Bots 협업 플랫폼 및 자기조직 인공지능 기술개발 사업, 과학기술정보통신부·한국연구재단 신진연구지원사업, 과학기술정보통신부·국가과학기술연구회 글로벌 TOP 전략연구단 지원사업의 지원을 받았다.

연구 결과는 2026년 9월 8일부터 12일까지 스웨덴 말뚤에서 열린 국제학술대회 'European Conference on Computer Vision(ECCV) 2026'에서 발표됐다. ECCV는 유럽컴퓨터비전협회(ECVA)가 주관하는 컴퓨터비전·머신러닝 분야의 주요 국제학술대회다.

한편 GIST는 이번 연구 성과의 기술사업화 및 산업적 활용에 관심 있는 기업·기관의 문의를 창업혁신진흥원 기술사업화센터(hgmoon@gist.ac.kr)를 통해 받고 있다고 밝혔다.

논문 정보

1. 논문명, 저자정보

- 학회명 : European Conference on Computer Vision (ECCV) 2026
- 논문명 : OP3DSG: Open-Vocabulary Part-Aware 3D Scene Graph Generation for Real-World Environments
- 저자 정보 : 김이룸(제1저자, GIST AI학과 석·박사통합과정), 김의환(교신저자, GIST AI학과 부교수)