“저기 소파 옆 테이블 위에 있는 빨간 책 찾아줘”
GIST, 문장으로 설명한 물체를 3D 공간에서 찾아내는 AI 로봇 기술 개발

- AI융합학과 김의환 교수팀, 문장 전체 맥락과 주변 사물 간 위치 관계 함께 이해하는 AI 로봇 내비게이션 기술 ‘Context-Nav’ 개발
- 추가 학습 없이 기존 대비 2.3배 높은 정확도 달성... 서비스 로봇 활용 가능성 제시
- AI 분야 국제학술대회 ‘CVPR 2026’ 발표 예정

▲ (왼쪽부터) AI융합학과 김의환 교수, 장원식 석박통합과정생

광주과학기술원(GIST·지스트, 총장 임기철)은 AI융합학과 김의환 교수 연구팀이 **사람이 문장으로 설명한 물체를 3D 공간 속에서 이해하고 정확히 찾아내는 ‘인공지능(AI) 로봇 내비게이션 기술**(Context-Nav)’**를 개발**했다고 밝혔다.

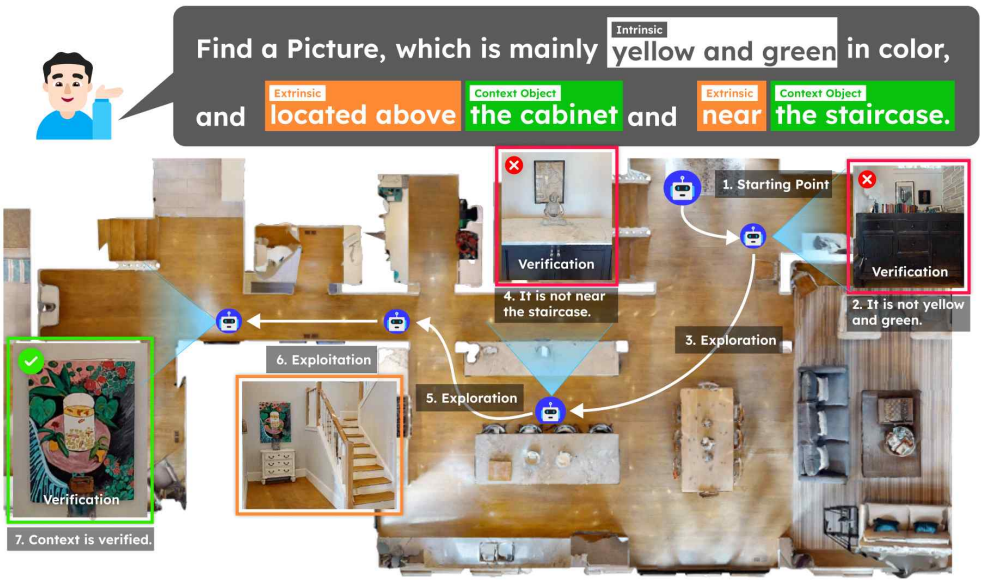
이 기술은 물체의 색과 모양 같은 물리적 특징을 포함해 **다른 사물과의 상대적인 위치까지 함께 분석해 다양한 서비스 로봇 분야로의 확장 가능성**을 보여준다.

실내 환경에서 청소·배달·안내 등 **다양한 작업을 수행하는 자율 로봇**(서비스 로봇)**이 사람의 언어 지시를 이해하고 정확히 수행하기 위해서는 주변 사물과 위치 관계를 종합적으로 파악해야 한다.**

기존 연구에서는 **로봇이 행동 전략을 스스로 배우도록 시도와 실패를 반복하며 최적의 행동을 찾는 ‘강화학습**(Reinforcement Learning)**’ 방식을 주로 사용했지만, 방대한 데이터와 높은 학습 비용이 필요**했다.

또한 **'의자','컵'처럼 물체의 짧은 속성적 정보만 활용해 사람이 길게 설명하면서 제공하는 ▲주변 사물과의 위치 ▲상대적 방향·배치 ▲상황적 단서 등 문장 속 맥락을 충분히 반영하지 못하는 한계가 있다.**

특히 **공간 관계**(왼쪽·오른쪽·앞·뒤)**는 관찰자의 시점이나 위치에 따라 달라지기 때문에 기존 방법에서는 로봇이 실제 목표 물체가 아닌 잘못된 후보를 목표로 오인할 가능성이 높다.**



▲ 긴 문장 설명을 활용한 3차원 문맥 기반 로봇 탐색 과정. 로봇은 긴 자연어 설명 속에서 색·모양 같은 물체의 고유 특징뿐 아니라 다른 사물과의 상대적 위치 정보를 함께 활용해 탐색한다.

연구팀은 이러한 문제를 해결하기 위해, **사람이 제공하는 긴 문장 설명 전체를 로봇의 탐색 과정에 활용**해 **목표 물체의 특징과 주변 사물 간 3차원 공간 관계를 함께 이해하도록 하는 방법을 제시**했다.

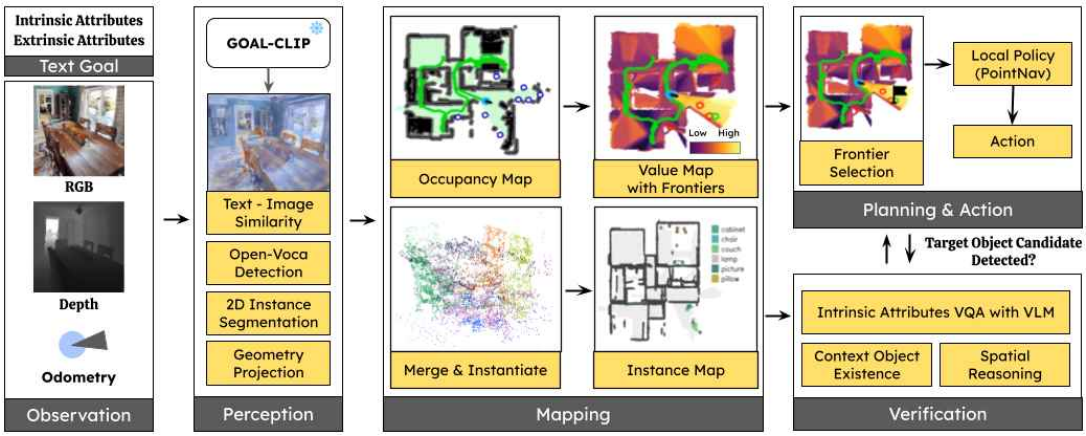
예를 들어 사람이 "거실 소파 옆 테이블 위에 있는 빨간 책을 찾아줘"라고 설명하면, **로봇은 이 문장을 단순한 물체 정보가 아니라 3차원 공간 속 위치 정보로 해석**한다.

먼저 **RGB 카메라*와 깊이 센서를 이용해 주변 환경을 인식**한 뒤, **설명과 일치할 가능성이 높은 영역을 실시간으로 확인**한다. 그리고 **후보 공간이 목표와 얼마나 잘 맞는지 적합도를 계산해 '가치지도**(Value Map)**'에 점수로 기록**한다.

이후 **가장 높은 점수가 매겨진 곳을 중심으로 효율적인 탐색 경로를 결정**한다. 후보 물체를 발견하면 이미지와 텍스트 정보를 함께 이해하는 비전언어모델(Vision Language Model)*을 활용해 **속성을 확인하며 3차원 공간 추론을 통해 주변 물체와의 위치 관계**(위·아래·왼쪽·오른쪽·앞·뒤)**를 정밀하게 검증**한다.

*** RGB 카메라:** 빨강(Red), 초록(Green), 파랑(Blue) 빛의 강도를 각각 기록해 사람의 눈처럼 색상을 인식하는 카메라이다. 촬영한 영상에서 사물의 색과 모양을 구분해 로봇이 무엇을 보고 있는지 이해하는 데 활용된다. 기존 로봇 학습 방식에서는 이런 영상 정보만으로 데이터를 수집했다.

*** 비전언어모델(Vision Language Model):** 이미지(시각 정보)와 텍스트(언어 정보)를 동시에 처리하고 그 관계를 이해하도록 학습된 인공지능 모델을 의미한다.



▲ AI 로봇 내비게이션 기술(Context-Nav) 구조도. RGB 카메라, 깊이 센서, 위치 정보, 목표 설명을 활용해 주변 환경을 3차원 지도로 만들어 로봇이 목표 물체를 탐색하는 과정을 보여준다. 물체가 발견되면 AI가 색·모양과 위치를 확인해 목표인지 판단하며, 맞으면 정지하고 아니면 탐색을 계속한다.

연구팀의 기술은 **로봇이 사물과의 관계, 색, 모양 등 세부적인 속성이 포함된 긴 문장의 이해도를 평가하는 '목표 찾기 시험**(CoIN-Bench)*' **에서도 높은 성과를 나타냈다.**

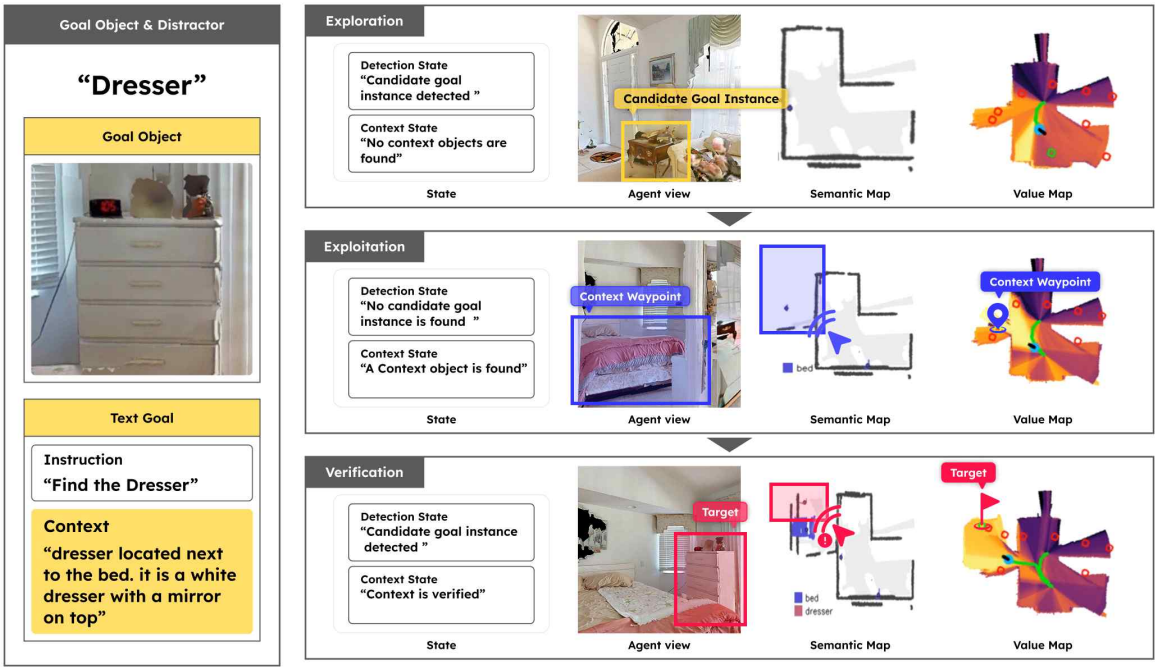
로봇이 목표를 찾는 최적 행동을 반복적인 시도와 실패를 통해 스스로 배우게 되는 기존의 일반적인 강화학습과 달리 **연구팀의 기술은 추가 학습 없이도 20.3%의 성공률을 기록해, 기존 강화학습 기반 방법(8.9%) 대비 약 2.3배 높은 성능을 달성**했다.

특히 **사람의 설명을 3D 공간에서 해석하고 문장 전체의 맥락을 반영해 가능성이 높은 위치부터 이동·탐색한 뒤, 주변 사물과의 위치 관계를 검증하는 전략이 로봇 행동의 정확도를 크게 높이는 데 효과적임을 입증**했다.

나아가 **긴 문장 설명 전체를 탐색 과정에 충분히 반영할수록 불필요한 이동이 줄어들고, 여러 시점에서 3차원으로 위치 관계를 확인하는 과정을 통해 목표를 잘못 인식하는 경우도 감소하는 것으로 확인**됐다.

*** 목표 찾기 시험(CoIN-Bench):** 로봇이 긴 자연어 설명을 바탕으로 목표 객체를 찾아가는 능력을 평가하는 공개 테스트. 사물과의 관계, 색·모양 등 세부 속성이 포함된 환경에서 성공률과 이동 효율을 측정하며, ICCV 2025 등 학술 대회에서 공개된 평가 세트를 기반으로 세계 연구자들이

동일한 조건에서 성능을 비교할 수 있는 신뢰할 수 있는 기준이다.



▲ 단계별 문맥 기반 탐색 과정 예시. '침대 옆, 거울이 놓인 흰색 서랍장'을 찾는 상황에서 로봇이 초기 후보를 잠시 보류하고 문장 속 문맥과 일치하는 방으로 이동한 뒤 목표 물체를 최종적으로 검증하는 과정을 보여준다.

김의환 교수는 "이번 연구는 **로봇이 물체 자체의 특징만 보는 수준을 넘어, 주변 맥락과 3차원 공간 관계까지 함께 이해하도록 정밀한 기술을 제시한 것**"이며, "**특정 과제에 맞춘 별도의 학습이나 조정 없이 새로운 공간이나 처음 보는 물체에 대한 설명에도 바로 적용이 가능해 향후 실내 서비스 로봇과 지능형 로봇 시스템의 실제 활용 가능성을 높이는 핵심 기반 기술이 될 것**"이라고 밝혔다.

AI융합학과 김의환 교수가 지도하고 장원식 석박통합과정생이 수행한 이번 연구는 **과학기술정보통신부·한국연구재단 우수신진연구자지원사업, 정보통신기획평가원**(IITP) **문제가설과 자가지도 기반의 자기주도 시각지능 기술 개발 사업, 국가과학기술연구회**(NST) **글로벌 TOP 전략연구단 사업의 지원**을 받았다.

연구 결과는 **국제 학술 서버 'arXiv'에 2026년 3월 18일 사전 공개**됐으며, **AI 분야의 권위 있는 국제학술대회 《Computer Vision and Pattern Recognition Conference**(CVPR 2026)**》에서 발표될 예정**이다.

CVPR 2026은 오는 6월 3일부터 7일까지 미국 콜로라도주 덴버(Denver)**에서 개최**된다.

한편 GIST는 이번 연구 성과가 **학술적 의의와 함께 산업적 응용 가능성까지 고려한 것으로, 기술이전 관련 협의는 기술사업화실(hgmoon@gist.ac.kr)을 통해 진행**할 수 있다고 밝혔다.

논문의 주요 정보

1. 논문명, 저자정보

- 학회명 : IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2026
  ※ 한국정보과학회 및 BK21+ 기준 최우수 학술대회
- 논문명 : Context-Nav: Context-Driven Exploration and Viewpoint-Aware 3D Spatial Reasoning for Instance Navigation
- 저자 정보 : 장원식(제1저자, GIST AI융합학과 석박통합과정), 김의환(교신저자, GIST AI융합학과 부교수)