

**“AI understands the functions of handles and buttons just like humans do”
GIST develops robot AI technology capable of understanding an object’s
detailed parts and functions**

- *A research team led by Professor Ue-Hwan Kim (Department of AI) has developed an “Integrated 3D Scene Graph” that links objects and parts by connecting their locations to their functions and potential actions, achieving a part retrieval performance of 83.6%, a 31.2 percentage-point improvement over the previous state-of-the-art*
- *The technology was implemented on the mobile robot “Stretch 3,” enabling capabilities ranging from identifying surrounding objects to navigation and movement*
- *The findings were presented at the international conference ECCV 2026*



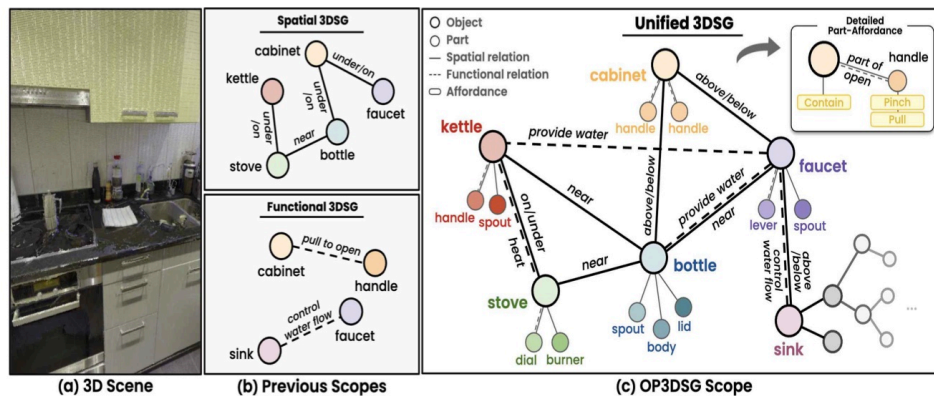
▲ (From left) Professor Ue-Hwan Kim and Yirum Kim (combined master’s/doctoral student) from the GIST Department of AI

The Gwangju Institute of Science and Technology (GIST; President Kichul Lim) Department of AI announced that a research team led by Professor Ue-Hwan Kim has developed an AI technology enabling robots to recognize not only surrounding objects but also specific components—such as handles and buttons—and their associated functions.

For instance, a robot cannot heat food simply by knowing the location of a microwave; it must locate the handle to open the door and identify the position and function of the buttons to operate it. While such information is intuitive to humans, a robot must understand these objects, their detailed components, and their functions in unison to successfully carry out human instructions in real-world environments.

To address this challenge, the research team introduced the "Unified 3D Scene Graph (Unified 3DSG)," a framework that integrates the locations of objects and components with their respective functions and potential actions. They also developed "OP3DSG" (Open-Vocabulary Part-Aware 3D Scene Graph Generation), an AI technology capable of automatically constructing this graph from real-world camera footage.

A "3D Scene Graph (3DSG)" serves as a "3D knowledge map for robots," representing the relationships between objects in a space to help robots understand their surroundings, navigate, and plan task sequences.



▲ *A comparison between conventional 3D scene graphs and the "Integrated 3DSG." Conventional methods treated spatial information (indicating where an object is located) and functional information (indicating which part of an object performs what function) separately. The "Integrated 3DSG" proposed by the research team represents objects, their constituent parts (such as handles and buttons), their locations and functions, and even possible actions are all connected within a single graph.*

However, conventional 3DSGs focused primarily on representing information at the object level, such as "chair" or "microwave", limiting their ability to specifically represent the smaller parts that make up an object or their functions.

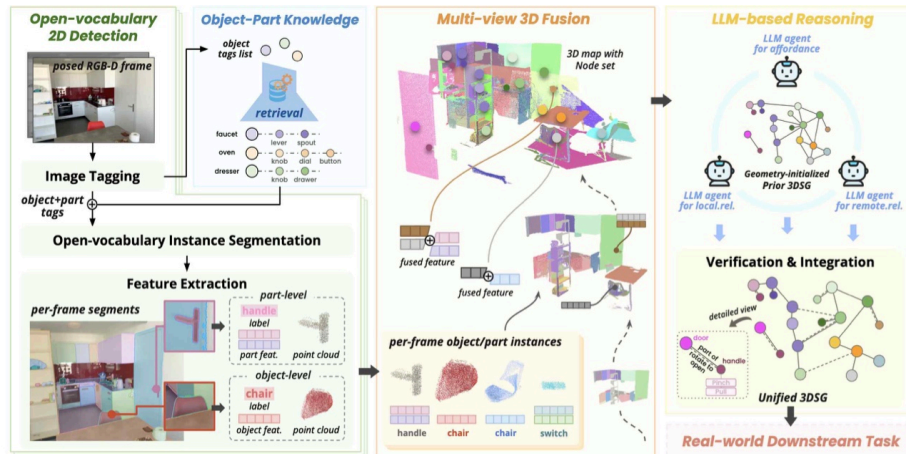
For instance, if a person is told to "find the chair leg," they can easily identify the specific part; however, when an AI attempts to locate it in a video, it might only detect the chair as a whole or fail to correctly link the same part when viewed from different angles. This is because small parts are diminutive in size and their appearance changes significantly depending on the camera angle.

OP3DSG, the method proposed by the research team, is an AI technology that automatically constructs an Integrated 3DSG from actual camera footage.

It begins by identifying key objects in the video and then searches for the detailed parts expected to be associated with those objects. For example, upon detecting an oven, the system adds the oven's constituent parts, such as buttons, the door, and the handle, to its search targets.

Next, it integrates video data captured from multiple angles into a 3D spatial framework. To distinguish whether small parts appearing in different frames are indeed the same component, the system compares not only their locations and semantic meanings but also their color information, thereby linking data regarding the identical part.

After establishing basic relationships based on the location and distance data of objects and parts, the system utilizes a Large Language Model (LLM)* to review and refine the functional relationships and potential actions associated with those parts and objects.



▲ *The operational process of the OP3DSG technology, which automatically constructs an "Integrated 3DSG." It identifies objects and their components, such as handles and buttons, from camera footage and integrates information captured from multiple angles into a 3D space. It then establishes relationships between the objects and components based on data like actual location and distance. Finally, it utilizes a Large Language Model (LLM) to analyze and refine functional connections and potential actions, thereby constructing the "Integrated 3DSG."*

For instance, beyond the spatial relationship of "microwave-handle," it links the functional relationship, such as recognizing that the handle is "used to open the door", and uses this to represent the actions a robot can perform.

* large language model (LLM): An AI model trained on vast amounts of text that understands and generates sentences, and infers knowledge and relationships.

The research team verified the performance of OP3DSG using a newly developed evaluation framework called "UniGraph3D."

The results showed that the correct component was included among the top three candidates identified by the AI in 83.6% of cases. This represents a 31.2 percentage point improvement over the highest-performing existing method, which stood at 52.4%.

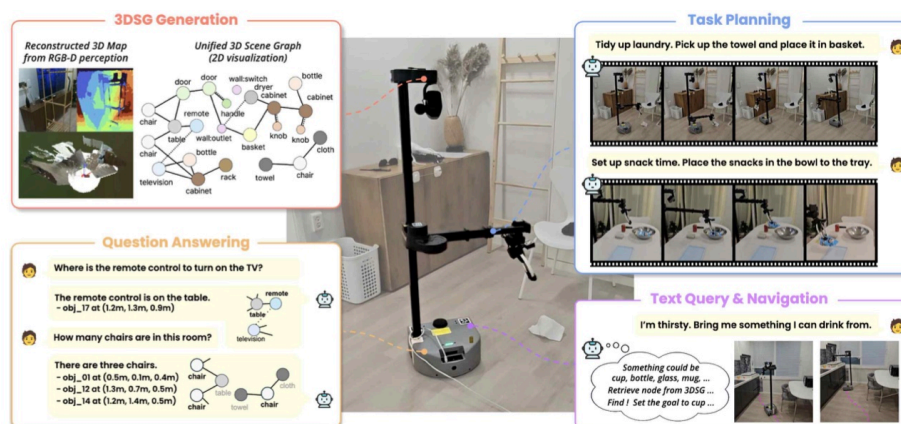
Performance in identifying spatial relationships between objects and components was 7.9 percentage points higher than the previous best, while performance in identifying functional relationships improved by 9.2 percentage points.

The research team also applied the developed technology to the mobile robot "Stretch 3" to confirm its practical applicability in real-world environments.

For example, the robot was able to count nearby chairs and locate the remote control needed to turn on the TV, then navigate to it.

Furthermore, when given an instruction like "tidy up the laundry," the robot could plan a sequence of actions, such as locating towels and placing them in a basket. It could also handle instructions that did not explicitly name specific objects, such as "I'm thirsty; bring me something to drink", by identifying the necessary object based on functional relationships and setting its location as the navigation target.

This demonstrates the potential for robots to translate human instructions into specific tasks based on the functional relationships between objects and components in the environment, without requiring humans to specify the exact names or locations of objects.



▲ *Practical application of the "Integrated 3DSG" (built using OP3DSG technology) in a robot. The mobile robot "Stretch 3" navigates an indoor space, using the "Integrated 3DSG" to answer questions about its surroundings and break down human instructions into specific action sequences. It can also locate relevant objects and set their positions as navigation targets based on descriptions of required functions or situations, even without being told the object's name directly.*

Professor Ue-Hwan Kim stated, "The core of this research lies in integrating the information necessary for a robot to understand not only 'what is where' but also 'which components are available and how they can be used.'" He added, "We expect this to serve as a foundational technology enabling service robots to comprehend complex real-world environments and translate natural language instructions from humans into concrete task plans and movements."

This research, supervised by Professor Ue-Hwan Kim of the GIST Department of AI and led by Yirum Kim (a student in the combined master's and doctoral program) as the first author, was supported by the "AI Bots Collaboration Platform and Self-Organizing AI Technology Development" project (Ministry of Science and ICT/Institute for Information & Communication Technology Planning & Evaluation), the "Early-Career Researcher Support" project (Ministry of Science and ICT/National Research Foundation of Korea), and the "Global TOP Strategic Research Group" project (Ministry of Science and ICT/National Research Council of Science & Technology).

The findings — OP3DSG: Open-Vocabulary Part-Aware 3D Scene Graph Generation for Real-World Environments — were presented at the European Conference on Computer Vision (ECCV) 2026, held in Malmö, Sweden, from September 8 to 12, 2026. ECCV is a premier international conference in the fields of computer vision and machine learning, organized by the European Computer Vision Association (ECVA).

Meanwhile, GIST announced that it is accepting inquiries from companies and organizations interested in the technology commercialization and industrial application of this research through the Technology Commercialization Center at the GIST Institute of Entrepreneurship and Innovation (hgmoon@gist.ac.kr).