GIST develops AI that detects changes in objects in videos even when filming time and location are different

- Research team led by Professor Ue-Hwan Kim of the Department of AI Convergence develops 'VSCDNet,' an AI model that precisely detects changes in actual objects by comparing past and present images of the same space, much like finding differences within the same picture

- Performance verified through the construction of a dataset of over 1.1 million frames and actual robot experiments

- Expected to be utilized in indoor patrol robots and smart security surveillance fields... Scheduled to be presented at the international conference 'ICML 2026'

▲ (From left) Professor Ue-Hwan Kim of the Department of AI Convergence, and integrated master's and doctoral student Jiae Yoon

A research team led by Professor Ue-Hwan Kim of the Department of AI Convergence at the Gwangju Institute of Science and Technology (GIST, President Kichul Lim) announced the development of 'VSCDNet,' an artificial intelligence (AI) model that automatically detects changes in actual objects by comparing videos of the same space taken at different times and along different paths.

Unlike existing methods that compare individual photos, this model is characterized by its ability to analyze the entire video to identify changes such as the appearance, disappearance, and movement of objects.

> ** VSCD (video-based scene change detection): Refers to the problem of detecting changes in objects by comparing videos of the same space taken at different times and along different paths. VSCDNet is an AI model proposed by the research team to solve the VSCD problem.*

For autonomous robots to navigate indoor spaces autonomously for extended periods, they must be able to continuously perceive changes in their surrounding environment.

However, existing change detection technologies primarily rely on comparing images taken at similar locations and viewpoints, which limits their ability to accurately detect changes in videos taken along different paths.

To overcome these limitations, the research team focused on the overall flow of the video instead of comparing individual images.

▲ ***Description of the video-based scene change detection (VSCD) problem. The reference video and the query video are footage of the same indoor space captured at different times and locations. Unlike existing change detection datasets, VSCD requires consideration of the correspondence relationships across the entire video rather than simple frame comparisons, as there are significant differences in camera movement paths and viewpoints.***

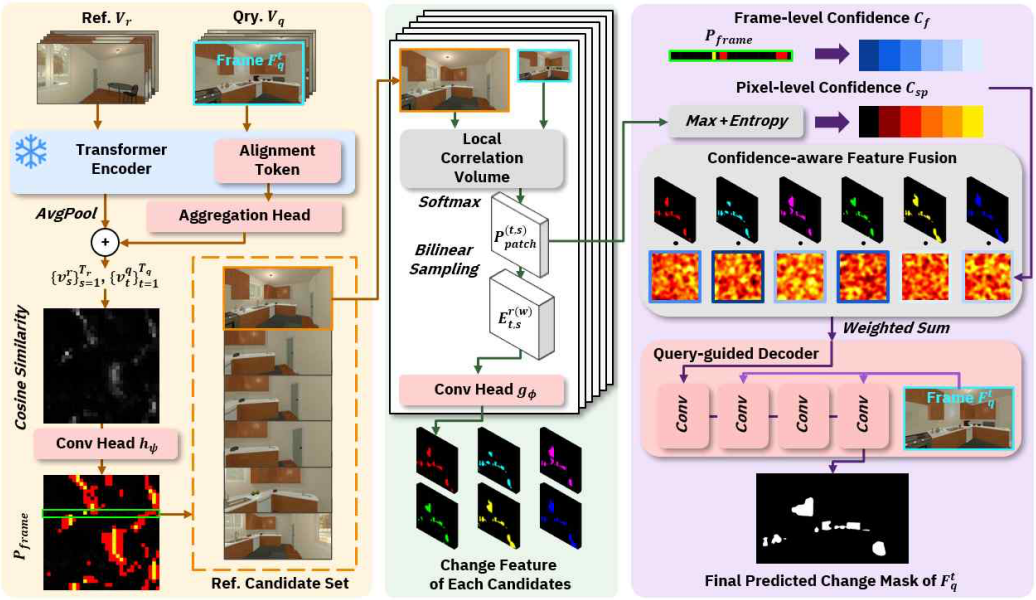
The developed AI model, 'VSCDNet,' compares the reference video captured in the past with the current video of the same space to identify corresponding scenes within the entire video and precisely detect only the areas where actual object changes have occurred.

Based on this, it generates a 'change mask'* that visually marks the areas where changes have occurred, presenting the final areas of change. In other words, it can automatically detect actual object changes, such as a laptop disappearing or an object's position shifting.

** change mask: A type of map that represents pixels where changes have occurred as '1' and areas without changes as '0'.*

To systematically verify the performance of this model, the research team directly constructed a large-scale dataset containing virtual space and actual indoor environment data. The dataset consisted of a total of 1,090 images (approximately 1.13 million or more frames).

Experimental results showed that 'VSCDNet' recorded the best performance compared to existing change detection techniques in both virtual space datasets and real indoor environment datasets. It maintained stable detection performance even under various conditions where image length, quality, and the number of changed objects varied.



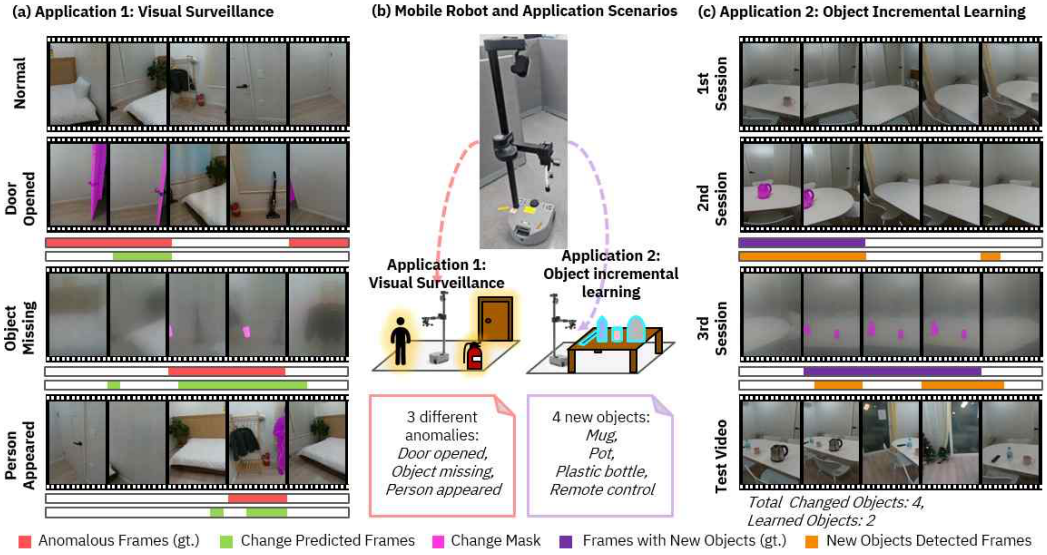
▲ ***VSCDNet model structure. VSCDNet identifies candidate reference frames related to the query frame, corrects time-space differences through patch-level correspondence, and predicts the final change mask by performing confidence-based fusion.***

Furthermore, in experiments using actual mobile robots, the system automatically detected situations where doors opened or objects disappeared within footage captured while the robot moved along different paths. It also verified the ability to remember and learn newly appearing objects.

This research is significant in that it has enabled the stable detection of actual object changes even in footage of the same space captured at different times and along different paths.

In particular, it demonstrated consistent performance under various environmental conditions, enhancing its applicability to real-world environments. The feasibility of monitoring environmental changes and learning new objects was also verified through actual robot experiments.

It is expected to be utilized in various fields in the future, including indoor patrol robots, smart security surveillance, facility management, and IoT-based smart indoor systems.



▲ ***Example of VSCD application in a real robot environment. By comparing footage captured while a mobile robot repeatedly visits the same space, it detects changes such as doors opening or objects newly appearing or disappearing. The detected change regions can be utilized for visual surveillance and object incremental learning.***

Professor Ue-Hwan Kim stated, “VSCDNet is an AI model that goes beyond simply recognizing the current scene to independently identify what has changed compared to the past.” He added, “Since it can compare videos captured from different paths without the need for separate location information or spatial maps, we expect it to be utilized in various fields such as indoor patrol robots, smart security surveillance, and facility management.”

This research, supervised by Professor Ue-Hwan Kim of the Department of AI Convergence and conducted by integrated master’s and doctoral student Jiae Yoon, was supported by the Ministry of Science and ICT and the National Research Foundation of Korea’s Excellent Young Researcher Support Program, the GIST-InnoCORE project, the Korea Institute of Information & Communication Technology

Planning & Evaluation's Self-Directed Visual Intelligence Technology Development Project, and the National Science and Technology Council's Global TOP Strategic Research Group Support Project.

The research results — VSCD: Video-based Scene Change Detection in Unaligned Scenes — are scheduled to be presented at 'ICML 2026 (International Conference on Machine Learning),' the world's premier conference in the fields of AI and machine learning, which will be held at the COEX Convention Center in Seoul from July 6 to 11. Prior to this, the paper was published on the preprint site arXiv on May 20.

Meanwhile, GIST stated that this research achievement takes into account both its academic significance and potential for industrial application, and that discussions regarding technology transfer can be conducted through the Technology Commercialization Center (hgmoon@gist.ac.kr).