GIST developed real-time 3D object tracking technology... Enabling remote support for industrial sites using a single camera

- Professor Kwanghee Ko's team from the Department of Mechanical and Robotics Engineering developed a technology that precisely estimates an object's 3D position and rotational direction (6DoF) using only a single RGB camera... This technology enables stable tracking even under changing lighting and background conditions

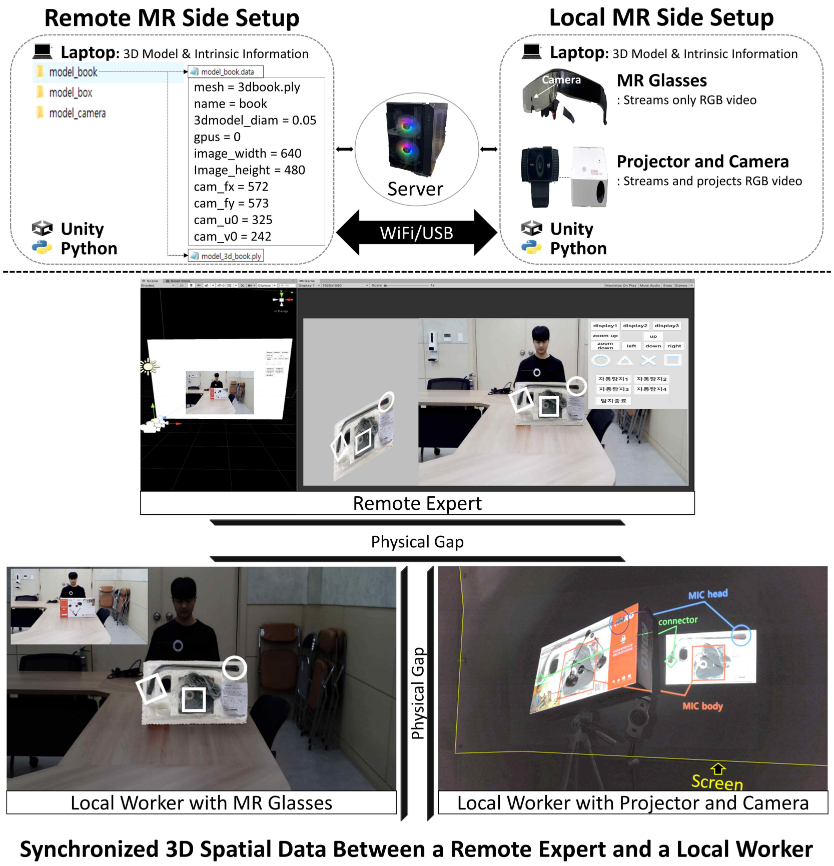
- This smart glasses-camera-projector-linked MR remote collaboration system is expected to be utilized in industrial settings, improving efficiency and accuracy in manufacturing training, remote after-sales service, and other areas... Published in the international journal 《Computers in Industry》

▲ (From left in the top row) Professor Kwanghee Ko (corresponding author, Department of Mechanical and Robotics Engineering, GIST), Inyoung Oh (first author, Department of Mechanical and Robotics Engineering, GIST), Gilsang Jang (second author, Department of Mechanical and Robotics Engineering, GIST), Jinho Song (third author, Department of Artificial Intelligence, Hannam University), Moongu Son (fourth author, Softhills), Daewoon Kim (fifth author, KT), and Junsang Yun (sixth author, Department of Mechanical and Robotics Engineering, GIST)

The Gwangju Institute of Science and Technology (GIST, President Kichul Lim) announced that Kwanghee Ko's team from the Department of Mechanical and Robotics Engineering has developed a technology that accurately identifies the location and tilt of an object in 3D space (six degrees of freedom (6-DoF)*) using a single RGB camera*. Based on this, the team has developed a mixed reality (MR) remote collaboration framework applicable to manufacturing, maintenance, and education.

* RGB Camera: A camera that captures color images using three color channels—red, green, and blue—records the R, G, and B values of each pixel to create a human-visible image. It is used in computer vision, robotics, and augmented reality for object recognition, tracking, and 3D reconstruction.

* 6-DoF (6 Degrees of Freedom): Refers to the six independent degrees of freedom an object can have in 3D space, divided into position translation and rotation. Position translation refers to the degrees of freedom to move left and right, up and down, and forward and backward along the three axes X, Y, and Z, while rotation refers to the degrees of freedom to rotate around the three axes Roll, Pitch, and Yaw. In other words, 6-DoF is a concept that can simultaneously define where an object is located in space and in which direction it rotates.



▲ Overview of the MR Remote Collaboration Framework (System Architecture). It provides a visual overview of 3D spatial data synchronization, annotation transfer, and device-specific rendering flow between remote experts and field workers. Alignment of the object coordinate system is key.

To support remote work on-site, it's crucial that the virtual markers or guide lines sent by the expert accurately align with the actual object surface. To achieve this alignment, existing MR remote collaboration systems* primarily used markers attached to objects or depth sensors to secure the object's coordinate system.

However, this approach suffers from the hassle of fixing and reinstalling markers, and depth sensors have limitations in practical industrial settings due to lighting and background variations and equipment costs.

In particular, while RGB camera-based estimation technology has advanced, it has been difficult to ensure both stability and real-time performance in real-world environments where lighting, background changes, and occlusion are repetitive.

* MR Remote Collaboration Framework: This system enables real-time collaboration by allowing remote experts and field workers to share the same object coordinate system, accurately aligning and projecting annotations, instructions, and work procedures onto the object's surface.

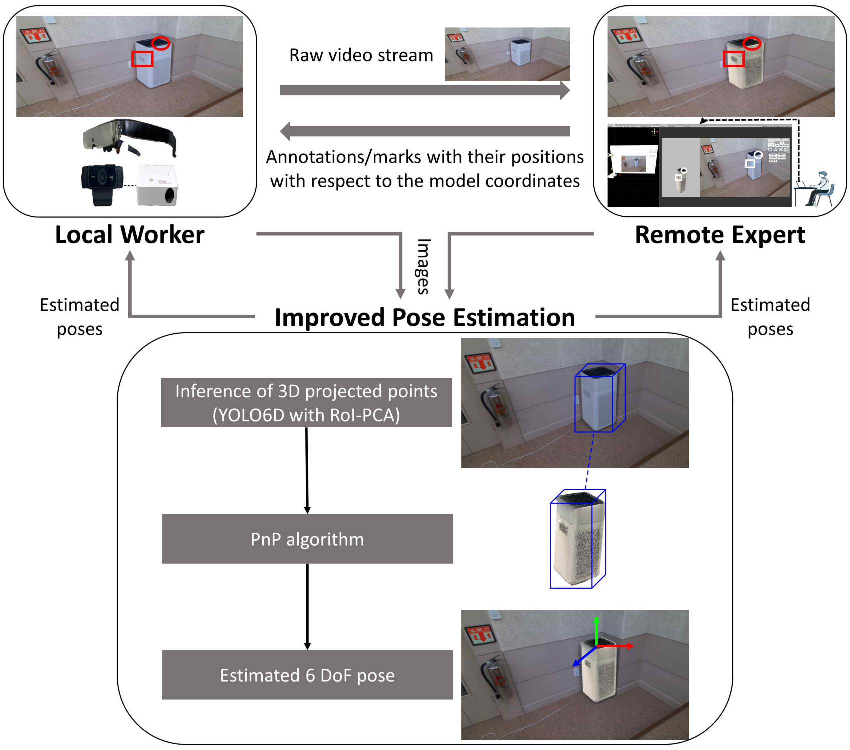
To address these limitations, the research team applied a new approach that enables stable object tracking despite lighting or background changes. This approach adds a color processing technique called "RoI-PCA" to the existing YOLO6D*-based key point estimation process.

RoI-PCA selectively transforms the color of only the object's region, rather than the entire image, thereby maintaining the object's unique color while reducing errors caused by lighting and background variations, thereby improving accuracy.

This allows for stable identification of an object's position and rotational direction using a single RGB camera.

* YOLO6D: An indirect pose estimation network that predicts the eight corners and centroid of a 3D object from a 2D image and then calculates the 6-DoF pose using the Perspective-n-Point (PnP) algorithm.

* RoI-PCA (Region of Interest Principal Component Analysis): This data augmentation technique applies PCA-based color space perturbation only to the object's region of interest (RoI) rather than the entire image. This maintains the object's unique color characteristics while minimizing the influence of background and lighting changes, thereby improving the stability of keypoint prediction.



▲ Information flow diagram and 6-DoF estimation pipeline. Single RGB input → YOLO6D with RoI-PCA-based keypoint estimation → PnP-based 6-DoF pose estimation process shows the local/remote bidirectional information flow.

The research team integrated the developed pose estimation technology into an MR remote collaboration system for practical use in industrial settings.

Images captured by a field worker's smart glasses or camera-projector equipment are transmitted in real time to an expert. The expert's annotations and instructions are precisely aligned to the analyzed object coordinate system and projected back onto the field equipment.

When using smart glasses, a simple eye correction UI* was implemented to compensate for minor alignment errors that can occur due to parallax between users. For the camera-projector system, a calibration technique (Gray code pattern-based calibration*) utilizing continuously changing patterns ensures stable projection of virtual information even on complex surfaces.

* Eye-Alignment User Interface (EUI): This interface compensates for subtle visual errors caused by parallax when using smart glasses in a Mixed Reality (MR) environment. By simply adjusting the user's eye position or gaze reference, real-time corrections can be made to ensure accurate alignment of virtual information with the object's surface.

* Gray Code Pattern-Based Calibration: This method, used in projector-camera systems to accurately position and align an object's surface, sequentially projects a gray code pattern with a single bit change and captures it with a camera to calculate the corresponding coordinates for each pixel. This technology enables stable alignment and 3D coordinate extraction even under lighting conditions or on complex surfaces.

The research team verified the technology's performance using the LINEMOD benchmark*, a representative test dataset, and proprietary 3D data (RGB-D camera capture) containing both color and distance information.

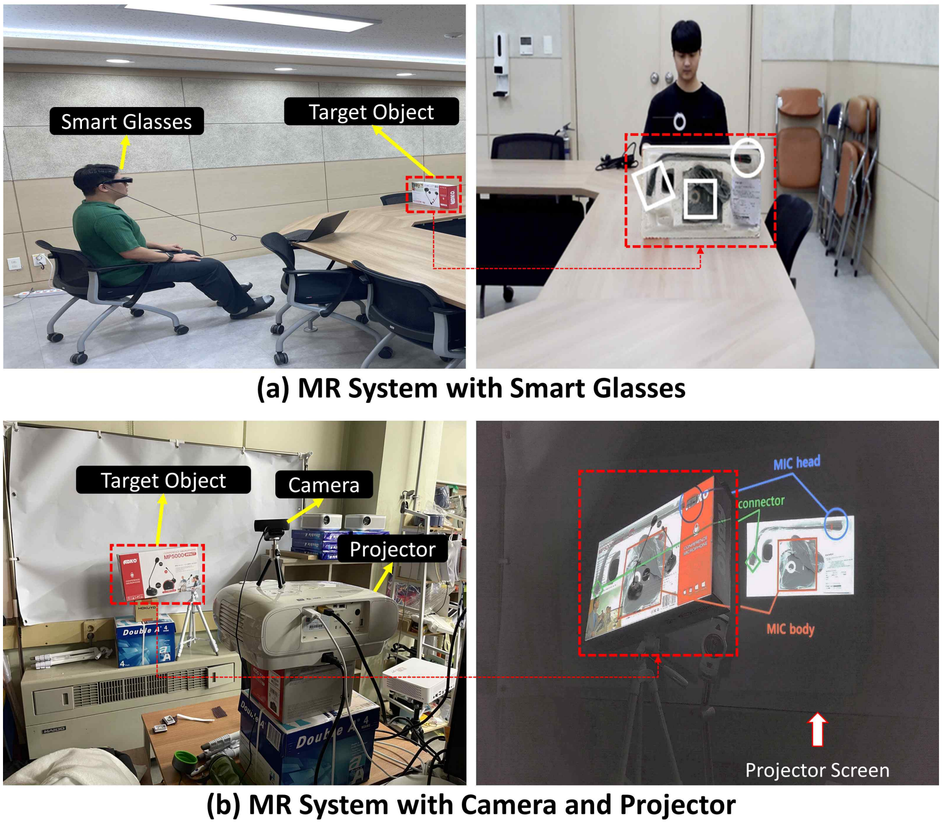
The model applying the new technique (RoI-PCA) demonstrated stable performance improvements over existing models in all criteria assessing the accuracy of object position and rotational orientation (2D reprojection error, ADD, and industrial pose estimation metrics such as the 5 cm·5° standard).

In experiments using small boxes and air purifier buttons, similar to real-world industrial environments, the field demonstration was successfully conducted, confirming that the virtual markers were accurately aligned without blurring even under lighting changes, background complexity, and surrounding objects.

* LINEMOD Benchmark: A widely used dataset and evaluation standard for evaluating 3D object recognition and 6-DoF pose estimation performance. It includes RGB-D images and 3D CAD models captured under various lighting conditions, backgrounds, and viewpoints, and is used by researchers to compare the accuracy and robustness of object detection, keypoint extraction, and pose estimation algorithms.

* RoI-PCA: This color space augmentation method applies PCA chromatic variation only to the region of interest (RoI) of the target object, rather than the entire image. This method maintains consistent object color characteristics across frames while minimizing the influence of background and lighting.

* 2D reprojection error is an accuracy metric based on the pixel-by-pixel difference between the predicted 6-DoF pose and the actual position (GT) of 3D model points projected onto the image plane. ADD (Average Distance of Model Points) measures 3D registration error by calculating the average Euclidean distance between model surface points and their actual positions based on the predicted pose. Furthermore, the 5 cm·5° criterion is used as a high-precision evaluation criterion, deeming a translation error of 5 cm or less and a rotation error of 5° or less as a correct answer.



▲ Multi-Device MR Remote Collaboration Demo: Smart Glasses and Camera-Projector-Based Implementation. (a) MR System with Smart Glasses: An example of precisely aligned circular and rectangular guide markers on a target object using the 6-DoF pose estimated from a single RGB input while wearing smart glasses. The enlarged inset shows the annotations precisely superimposed on the object's coordinate system within the user's field of view. (b) MR System with Camera and Projector: An external camera estimates the 6-DoF of the target object, and a calibrated projector correctly projects the same annotations onto the actual surface (projector screen). Camera-projector alignment is maintained using Gray code-based calibration, demonstrating stable remote command visualization even against complex backgrounds.

Professor Kwanghee Ko stated, "This technology eliminates the need for markers or depth sensors, significantly reducing field application costs and maintenance burdens." He added, "We anticipate that it will contribute to reducing work errors and shortening processing times in various fields, such as training courses (manufacturing line onboarding) that help new workers quickly learn equipment and processes to operate safely and accurately, remote after-sales service, quality inspection, and practical training."

The research team plans to expand this technology to include pose estimation technology, which can accurately estimate object position and orientation even in industrial settings with different environments, such as lighting and background, by combining semi-supervised learning, which provides correct information only to a portion of the data and allows the remaining data to be learned autonomously, with federated learning, which allows data from multiple work sites to be learned simultaneously without direct sharing.

This research, supervised by Professor Kwanghee Ko and conducted by doctoral student Inyoung Oh of the Department of Mechanical and Robotics Engineering at GIST, was supported by the Industrial Innovation Talent Growth Support Program of the Ministry of Science and ICT, the Institute of Information and Communications Technology Planning and Evaluation (IITP), the Ministry of Trade, Industry and Energy, and the Korea Institute for Advancement of Technology (KIAT). The results — A mixed reality-based remote collaboration framework using improved pose estimation — were published online in the international academic journal 《Computers in Industry》 on November 20, 2025.

Meanwhile, GIST stated that this research achievement considered both academic significance and industrial applicability, and that technology transfer inquiries can be made through the Technology Commercialization Center (hgmoon@gist.ac.kr).

GIST
Since 1993