3D 'depth information' found more accurately in videos with AI algorithm!

— Professor Hae-Gon Jeon's team at the AI Graduate School will present at the European Computer Vision Conference (ECCV) in October
— Estimation of depth information of the world's best performance based on focal length... Utilization of applications such as VR, AR, and metaverse

▲ From left, Professor Hae-Gon Jeon and Won Chang-yeon and integrated student Changyeon Won

A research team at GIST (Gwangju Institute of Science and Technology, President Kiseon Kim) has developed a technology that can accurately estimate 'depth information' in an image through an artificial intelligence (AI) algorithm.

If this achievement is applied, more accurate depth information can be obtained from the image through the algorithm, which can improve the quality and development of 3D virtual reality (VR) and augmented reality (AR) images, metaverse, and computer vision applications such as image-based cognitive detection.

Depth information in video is used as a key factor in various fields of application such as 3D space of AR/VR and natural image synthesis. If there is no depth information, there is no information about the 3D space, so an unnatural image is inevitably implemented in the application stage.

Recently, research to obtain image depth information using focus images obtained through the autofocus function of a DSLR camera or smartphone camera is being conducted at global IT companies such as Google or at various universities.

Most of the existing methodologies estimate depth information based on a single image with a narrow depth of field*. Because it is assumed that the auto-focus image is aligned* or the user has to manually align the auto-focus image through additional settings according to the characteristics of the image data, the accuracy of the depth information is limited.

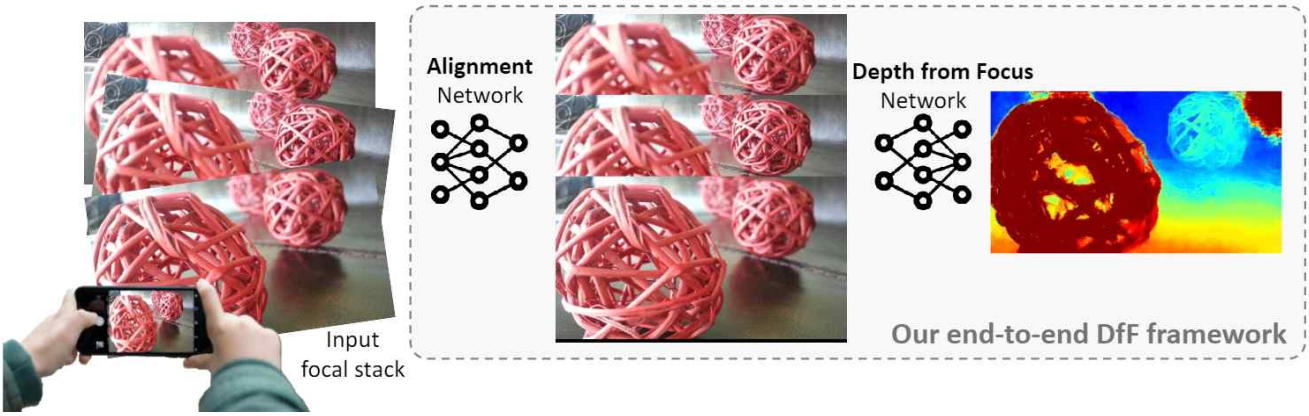
* depth of field: The amount of distance that is in focus before and after the center of the subject. The farther the subject is and the narrower the aperture, the deeper the depth of field.

* alignment: The process of aligning the position of the subject on the image to the position of the other image.

A research team led by Professor Hae-Gon Jeon of GIST AI Graduate School devised an algorithm to more accurately estimate depth information using the entire image used in the autofocus function through AI techniques.

The research team aligns the images obtained through the autofocus function in the metadata output from the camera with an artificial intelligence network. Using this end-to-end method* for estimating depth information was proposed for the first time in the world.

* end-to-end technique: A deep learning technique that learns and infers all processes between the input and output of an algorithm as a single model.

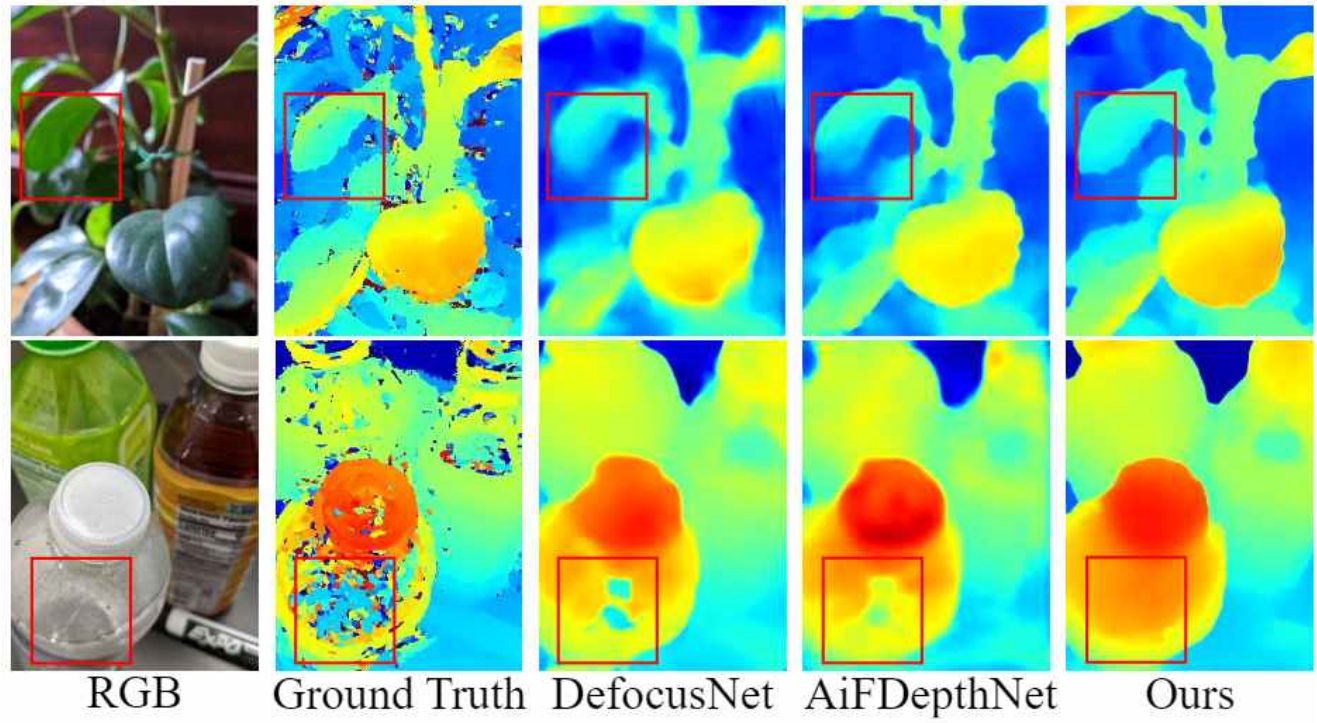


▲ The focal length-based depth information estimation network structure proposed in this study. Unlike previous studies that dealt only with the depth from focus network, the alignment network that aligns the subject positions of the images taken from multiple focal points to the same and accurate depth information estimation to propose an end-to-end structure in which a depth from focus network is combined.

The captured focus images are arranged in the following way. The degree of view of the 1 image, that is, the field of view, is calculated for each focus as metadata, and then the 2 focus images are cut to fit a specific field of view, and the field of view of all the captured images is set equally. 3 By further aligning the focal images with the artificial intelligence network designed by the research team, well-aligned focal images can be finally obtained.

In particular, the depth information measured through this performance ranks first in 17 out of a total of 20 scales measuring error and accuracy in the official benchmark*. It showed consistently excellent performance in various cameras and environmental conditions.

* accredited benchmark: This study is a benchmark for measuring performance on the dataset (DDFF 12 Scene) provided by the research paper (ACCV2018 Conference) of the Technical University of Munich, Germany (https://competitions) .codalab.org/competitions/17807). This benchmark objectively evaluates the performance of the algorithm through 19 scales that quantify the error and accuracy between the actual depth information and the depth information estimated by the algorithm, and one scale that measures the depth information inference speed of the algorithm.



▲ In the case of the far left (RGB), the video screen appears when shooting with a normal camera, and the remaining four images on the right show the depth information as a color video.

Professor Hae-Gon Jeon said, "This study proposes an artificial intelligence model that overcomes the limitations of the existing manual focal length image alignment and depth information estimation algorithms. Using depth information obtained through artificial intelligence networks is expected to contribute to the development of applications such as metaverse, augmented reality, and image-based cognitive detection."

This research was led by AI Graduate School Professor Hae-Gon Jeon (corresponding author) and integrated student Changyeon Won (first author) with support from the Ministry of Science and ICT (Research and Development Zone), Gwangju Metropolitan City, the 'Convergence Culture Virtual Studio for Realization of Artificial Intelligence-based Metaverse' project hosted by GIST, and the 'AI Innovation Hub' project of the Ministry of Science and ICT and will be presented in October at the European Conference on Computer Vision (ECCV), the world's best conference in the field of artificial intelligence and computer vision.

GIST
Since 1993