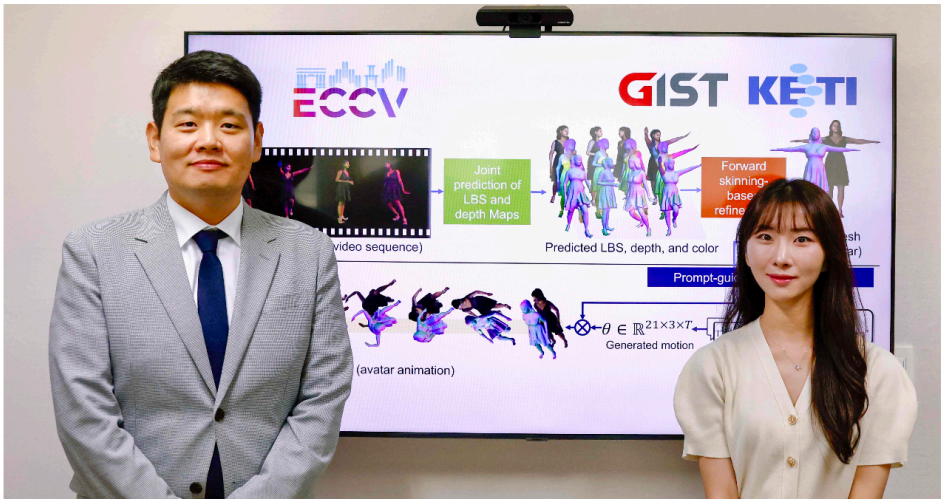
GIST develops AI technology to create 3D avatars that move using only 2D flat images

- AI Graduate School Professor Hae-Gon Jeon's team develops an AI algorithm that implements an image of a single person making a simple movement into a 3D avatar with natural movements without a separate, complex process
- Compared to the current technique that requires the touch of a graphic expert, it greatly reduces time and cost... "Expected to be used in cultural industries such as action movies and games that require complex movements that are difficult for humans to perform"
- Scheduled to be presented on October 2nd at 'ECCV 2024', one of the world's top 3 computer vision conferences held in Milan, Italy



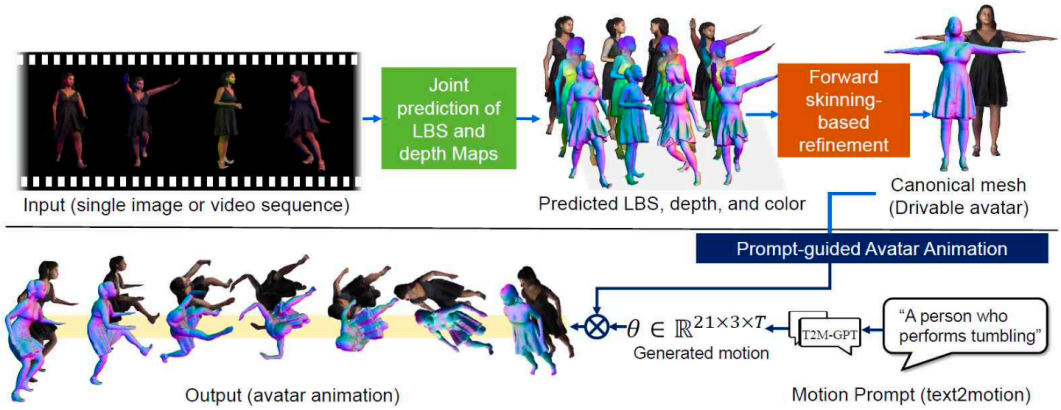
▲ (From left) AI Graduate School Professor Hae-Gon Jeon and Jisu Shin, a combined master's and doctoral program student

An artificial intelligence (AI) technology has been developed that allows anyone to create a high-quality 3D avatar that can perform any action using only a 2D image or video. It is expected to be used to create natural movements like real people in videos.

The Gwangju Institute of Science and Technology (GIST, President Kichul Lim) announced that Professor Hae-Gon Jeon's research team from the AI Graduate School has developed a technology to create high-quality 3D characters capable of various movements from 2D images.

This is an innovative method that uses artificial intelligence (AI) technology to easily create a 3D human model using only 2D images without the need for multiple viewpoint cameras or expensive equipment.

In addition, it can significantly reduce costs and resources because it does not require separate filming equipment and studio facilities, and can be used to create realistic characters like real people, especially in various cultural industries such as games, movies, AR, and VR.



▲ Summary and application of this study. Each 3D model is implemented using an image of a single person, and the information of the models is integrated to create a realistic avatar that can move.

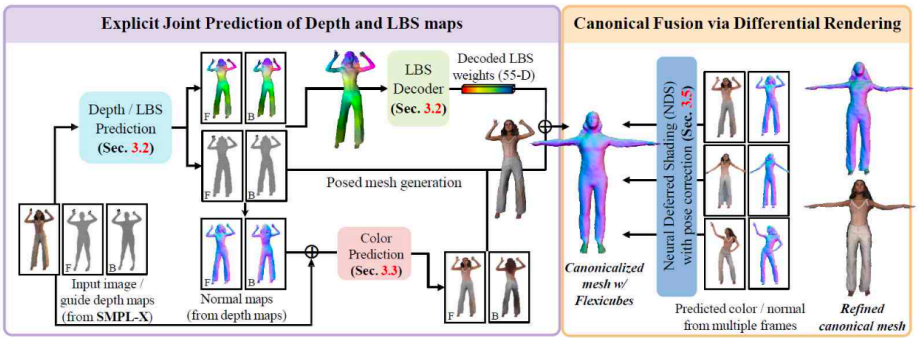
Existing methodologies for generating 3D characters from a single image are limited to static avatars that simply create shapes in 3D space and do not allow free movement.

To make these static avatars actually move, a process called 'rigging' is essential, which involves setting up a virtual bone and joints for the character so that each part can move naturally.

This process requires a very sophisticated task, and skilled graphic designers must manually specify each and every joint and range of motion to ensure that the character's movements are realistic.

Therefore, considerable expertise and resources are still required to create movable characters, and research is currently being conducted into automated rigging technologies and more efficient animation techniques.

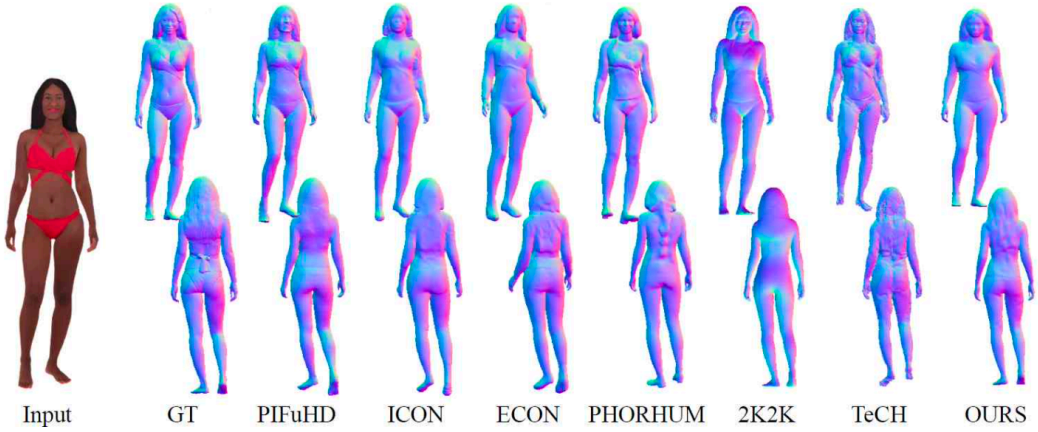
The core of this study is to simultaneously predict depth information and a skinning map (linear blend skinning, LBS map)*, then generate an initial dynamic avatar based on this, and integrate and improve multiple image information.



▲ Avatar generation pipeline proposed in this study. It is divided into a part that creates a 3D mesh by extracting a depth map, skinning map, and color map from an image, and a part that integrates each implemented mesh and creates a final avatar by improving the difference between the normal map and color map.

This technology, which can generate a 3D character model with only a few pose images of a character, first predicts the character's front and back depth information and skinning map through the first neural network from each image. In order to simply express the high-dimensional skinning map, a second neural network model consisting of an autoencoder and decoder is used to compress the complex skinning information, and then decompress it to implement accurate movements.

* skinning map (linear blend skinning map, LBS map): The skinning map used in this data refers to a two-dimensional map that contains information on the movement method of each joint of a three-dimensional person for each pixel of the person image.



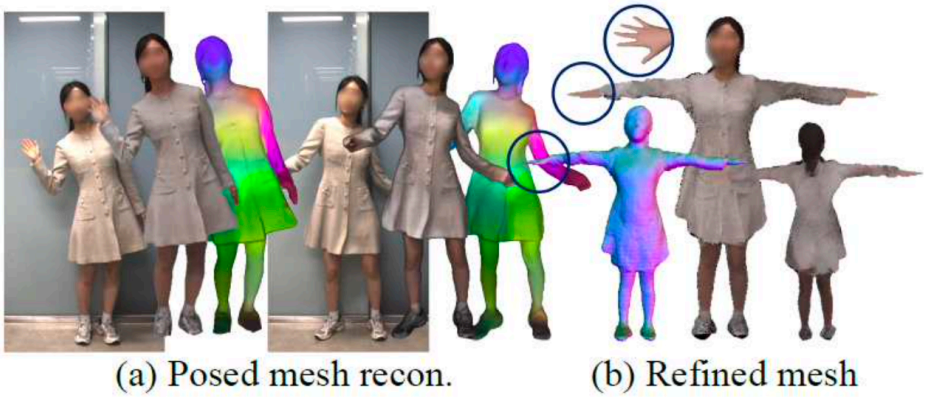
▲ Comparison of front and back normal maps between the existing results that performed 3D implementation based on the input image and the results implemented using the method proposed by our research team (Ours).

Next, based on the predicted depth information and skinning map, the initial 3D avatar, or canonical mesh*, is generated. Previously, it was necessary to calculate skinning values for the model implemented using a character template model and generate a canonical mesh, but with the technology developed by the research team, a canonical mesh can be generated without a separate post-processing process.

* canonical mesh: The basic pose of a 3D human model. It is used as a starting point for various movements, and is usually defined as a pose of standing still with arms outstretched.

Next, the initial avatar is refined and textured using a rendering technique based on forward skinning proposed by the research team.

This technique can improve the accuracy of the avatar by integrating information from multiple images and reducing the color and location differences based on a two-dimensional image. In addition, it has the advantage of reducing the pose difference between the joints of the generated avatar and the pose of the avatar generated through the first neural network, so that the pose error is not large. Through this method, a three-dimensional avatar capable of natural movement can be created.



▲ An example of an avatar generated by the algorithm by photographing an actual person wearing a loose-fitting one-piece dress.

In addition, this technology has the advantage of not being limited by the number of input images and can obtain realistic results even with a single image. In particular, it shows the possibility that avatars can implement various movements

when combined with the technology of generating movement from text based on a super-large language model that has recently attracted attention.

▲ An example of an application where an avatar created using the algorithm of this study dances when the text ‘light dancing motion’ is provided.

Professor Hae-Gon Jeon said, "Through this research result, existing graphic designers can greatly reduce the time and cost required to move 3D objects, and ordinary people can easily create and use 3D avatars with only 2D images. It is expected that in the future, it will be used as a realistic avatar in various cultural industries such as action movies or games that require complex movements that are difficult for humans to perform."

This study, led by Jisu Shin, a combined master's and doctoral student in the AI Graduate School's Professor Hae-Gon Jeon Lab, and jointly conducted by GIST and the Korea Electromagnetic Engineering Institute (KETI), was supported by the Development of a User-Participatory Metaverse Performance Solution Based on Neural Human Modeling (Ministry of Trade, Industry and Energy, Korea Institute for Advancement of Technology), an international R&D collaboration project, and the Development of Object Media Processing Technology for Multi-Source Images (Ministry of Science and ICT, National IT Industry Promotion Agency). It is scheduled to be presented at the 'ECCV (European Conference on Computer Vision),' one of the world's top three academic conferences in the field of computer vision, on October 2, 2024.

GIST
Since 1993