

A study of virtual human production methods: Focusing on video contents

Kim, Kwang Jib

Assistant Professor, School of Media, Seoul Institute of the Arts, Korea
Kangjib@seoularts.ac.kr

Abstract

Interest in virtual humans continues to increase due to the development of generative AI, extended reality, computer graphics technology, and the spread of a converged metaverse that goes beyond the boundaries between reality and virtuality.

Despite the negative public opinion that virtual humans were just temporary form of entertainment event in the early days of their emergence, the reason they are showing continuous growth is due to the unique characteristics of virtual humans and the expansion of diverse usage from technological advancements. The production of video content using virtual humans is becoming vigorously active, but currently there is limitation and no exact process for the technology to apply virtual humans to video content for it to be produced accordingly to the characteristics or situations of virtual humans.

In this study, we investigated the characteristics of virtual human production technology methods & processes, and identifying the impact of each production technology on the production environment through examples of virtual human content applied to domestic and international video contents.

In conclusion, by proposing an appropriate production method for each content, we hope to develop and assist production practitioners so they can effectively use virtual humans in video content production.

Keywords: *Virtual Human, Digital Double, Deepfake, Game Engine, Generative AI*

1. Introduction

The ongoing development of generative AI, augmented reality, computer graphics technology, and the widespread adoption of the converged metaverse, transcending the boundaries between reality and virtuality, have led to a sustained increase in interest in virtual humans. As the demand grows for online experiences resembling offline ones, there is an evolution in metaverse spaces, allowing individuals to create avatars that represent them or engage in real-time communication with virtual humans resembling real people, crossing the boundaries between reality and virtual worlds. Virtual humans are establishing themselves not only as one-time events but gradually as sustainable content, surpassing initial negative perceptions of being one-time events.

Despite initial negative opinions about virtual humans being perceived as one-time events, they are showing continued growth due to the unique characteristics of virtual humans and the expanding scope of their utilization resulting from technological advancements [1]. Virtual humans offer advantages such as unlimited constraints in time and space, adaptable to various fields, and effective reduction of diverse risks associated with real individuals. Consequently, they have gained sustainability endorsements from various industries and businesses, engaging in activities such as official modeling, appearing in TV advertisements, and expanding their scope beyond the entertainment sector into e-commerce, beauty, health, and various other fields.

While the use of virtual humans in video content creation is becoming prevalent, there is currently no precise process for applying virtual humans to video content, leading to frequent challenges in high costs or inadequate quality. In this study, the aim is to investigate the characteristics of virtual human production techniques and processes, examining the impact of these production techniques on the environment through case studies of virtual human content applied in domestic and international video content. In conclusion, proposing suitable production methods for each type of content is intended to assist production practitioners in effectively incorporating virtual humans into video content production.

2. Background Theory

2.1 Definition of Virtual Human

The term "virtual human" is used diversely depending on the industry utilizing it, with variations such as digital human, cyber human, artificial person, and more. Recent distinctions are also made based on content platforms, referring to them as virtual influencers or virtual YouTubers. Among these, "virtual human" and "digital human" are commonly used, and some experts differentiate them based on the implementation of bidirectional interaction using artificial intelligence technology. However, according to the research on the semantic shift of virtual human conducted by Hwang Seo-ee and Lee Myung-cheon, following the occurrence of COVID-19, artificial intelligence technology is included within the meaning of virtual human as a presence resembling real humans, serving roles like advertising models, influencers, YouTubers, etc.

From a content perspective, William R. Swartout (2006), proposing a training system utilizing virtual humans, defines them as "software-created objects that appear, behave, and interact like humans but exist in a virtual environment [2]." Researchers Nadia Magnenat-Thalmann & Daniel Thalmann (2005), focusing on realistic virtual human creation, define virtual humans as "computer graphics simulations of real people," emphasizing the need for realistic external modeling, smooth and flexible motion, and high-level realistic behavior for creating lifelike virtual humans [3].

Research by Kwak Bo-eun and Heo Jeong-yun (2021), focusing on digital human utilization services, defines digital human as a "3D virtual being with the appearance and speech of a real person [4]." Another classification by Seo Young-ho, Oh Moon-seok, and Han Gyu-hun (2021) categorizes digital humans as "3D humanoid models created to imitate the features and appearance of real people for the purpose of replacing human roles [5]."

While it's challenging to confine digital humans and virtual humans to a single definition, considering their diverse representations as virtual influencers, virtual models, etc., across various digital content fields, this study collectively refers to them as "virtual humans." Additionally, for the focus on video content creation using virtual humans, this study defines virtual humans as "virtual objects expressed with a level of realistic representation similar to real people, utilizing artificial intelligence and computer graphics technology."

2.2 Emergence and Market Trends of Virtual Humans

Virtual humans have undergone a significant evolution alongside advancements in computer graphics technology. In its early stages, virtual humans represented 3D models with human-like appearances, such as personas and avatars. However, less sophisticated than the expanded meanings they carry today. In 1998, South Korea saw the debut of the first virtual humans, the cyber singers 'Adam' and 'Lusia,' engaging in various activities like album releases, music video production, and advertisements. However, creating realistic virtual humans faced challenges due to the limitations of distinguishing them from real individuals in terms of appearance and movement, high production costs, and lengthy production times. Despite a brief period of popularity, these early virtual humans disappeared from the market.

Following Adam, who first virtual human in South Korea, his debut, about a decade of advancements in computer graphics technology, driven by hardware and software innovations, enabled the creation of results almost indistinguishable from reality. Alongside technological progress, virtual humans gradually achieved realistic graphics and natural movements. Recent developments in technologies such as artificial intelligence, deepfakes, 3D scanning, and game engines have directly influenced the emergence of diverse virtual humans. Realistic graphics, including details like skin texture, fur, and pores, are now achievable, and virtual idols' debut music videos have garnered millions of views. Virtual humans have also appeared as show hosts in live commerce and home shopping, successfully selling luxury bags. Additionally, there is a growing trend of industrial attempts, such as AI anchors, AI bankers, and AI professors, aiming to perform various roles on behalf of busy individuals. Globally, well-known virtual humans are actively engaged in various activities. [Figure 1] illustrates the appearance of the first domestic virtual humans and international human-like virtual humans.



Figure 1. Virtual humans

In particular, the rapid shift to a contactless society following the acceleration of the post-COVID-19 digital transformation has intensified societal interest in virtual humans, who are free from constraints of time and space in their activities. Building on this, the advantages of securing a stable corporate image without temporal and spatial limitations have been emphasized, leading to rapid growth in the entertainment industry, including broadcasting and advertising. It is anticipated that in the future, virtual human utilization will expand beyond entertainment into various sectors such as education, commerce, virtual assistant services, counseling, and therapy. According to Grand View Research, a global research and consulting company, the global virtual human market was valued at \$14.34 billion in 2022, with an expected compound annual growth rate (CAGR) of 42.6% from 2023 to reach \$270.61 billion by 2030. The global virtual influencer marketing market is projected to grow from \$3 billion in 2021 to \$14.8 billion in 2026. In the domestic market, the digital human market is forecasted to grow at an annual rate of 46.40%, reaching approximately KRW 3.89 trillion by 2026 from KRW 575.3 billion in 2021. The domestic virtual influencer market is expected to grow at an annual rate of 34.55%, reaching around KRW 3.768 trillion by 2026 from KRW 919 billion in 2021. Virtual human

technology is currently being utilized in various fields and is expected to expand its applications across even more sectors as it continues to advance.

3. Virtual Human Production Technology

3.1 Classification of Production Technology

Various technologies are used in the production of virtual humans. Photorealistic virtual humans, based on technologies such as 3D scanning, deep fake, and Generative Adversarial Network (GAN), are being created to the extent that it is difficult to distinguish them from real photos. Differences in the technology used, production methods, and the form of the output result in variations in the expression and scope of activities of virtual humans. It is necessary to apply and utilize virtual humans in content in various ways according to the characteristics of each content.

Based on various prior research and practical experiences [6, 7], virtual humans can be broadly classified into two main categories depending on the production method: Full 3D method and Image Synthesis method. Further subcategories include Digital Double method, Engine-Based method, Deepfake method, and Generative AI method. Table 1 classifies the types of virtual human production methods.

Table 1. classifies the types of virtual human production methods

| Production Method | | |
|-------------------|-----------------|---|
| Digital Double | Method | Utilizes Full 3D technology to create faces. |
| | Strength | Graphic fidelity is high enough for use in Hollywood movies, and there are no issues with angles. |
| | Weakness | Demands high production costs and an extended production timeline. |
| Engine Based | Method | Involves the use of real-time rendering game engines capable of real-time motion and facial expression capture when using motion capture devices. |
| | Strength | Offers real-time movement and facial expression capabilities. |
| | Weakness | Graphics quality achieved is lower compared to digital doubles. |
| Deepfake | Method | Integrates 2D images based on virtual faces onto real bodies captured during filming. |
| | Strength | Economical with a fast workflow, making it the most widely used method. |
| | Weakness | Limited by low resolution and constraints in depicting facial angles, restricting movement expression. |
| Generative AI | Method | Builds extensive image datasets to generate results through text input. |
| | Strength | Allows quick production of results without engaging in graphic work. |
| | Weakness | Faces challenges in creating consistent images and has limitations in implementing video content. |

3.2 Process of Virtual Human Production

3.2.1 Full 3D Method

The Digital Double method involves creating 3D models to implement real individuals or virtual characters. Primarily used in the fields of film and game production, it aims to accurately reproduce the physical features, appearance, and movements of the original subject in 3D. Its high quality has led to its use in Hollywood films for quite some time. [Figure 2] illustrates the process of creating virtual humans using the Digital Double method, utilizing a photogrammetry system.



Photogrammetry system



Virtual human based digital double

Figure 2. Photogrammetry system and virtual human using digital double method

The Digital Double method used in virtual human production involves creating the entire body in 3D for use. However, it is more common to digitally create only the face and then synthesize it with the actual person's body. Due to the necessity of accurate 3D modeling, tasks such as 3D scanning of individuals, high-quality texture work, shading, and more are required. This process incurs high costs and demands a significant production time. [Figure 3] depicts a virtual human created using the Digital Double method.



ROZY



ARA

Figure 3. Virtual human with digital double method

The Engine-Based method, similar to the Digital Double method, is based on game engines such as Unreal and Unity. The key difference lies in not synthesizing the face with the actual body; instead, the entire body is typically created in 3D for implementation. This method allows for the creation of virtual humans with real-time interactions, thanks to features like real-time rendering and physics simulation. Additionally, it enables real-time expression of movements and various effects in the surrounding background through motion capture

and facial capture. Although it offers the advantage of quickly verifying results in real-time, it faces challenges in achieving the same level of realism as the Digital Double method due to limitations in real-time rendering and hair simulation. However, the Engine-Based method allows for the creation of new faces, utilizes 3D data, and offers relatively natural direction from various angles. This makes it suitable for various applications, such as live shopping, real-time YouTube and broadcast appearances. [Figure 4] showcases a virtual human created using the Engine-Based method.



Plave



MAVE:

Figure 4. Virtual human with game engine method

To create a Digital Double and a game engine-based virtual human, initially, a 3D model must be produced using either 3D scanning technology or sculpting. Sculpting involves directly carving the model, showcasing a high level of precision but requiring significant costs and time. In recent times, the use of 3D scanning has facilitated a simpler and faster production process. However, it is dependent on hardware, and variations in results may occur based on hardware performance. After the model is created, rigging for natural movement of gestures and expressions, as well as texturing for material representation, are essential tasks. Animation work is required to express movements necessary for video content application, achieved through frame-by-frame animation or using motion capture equipment. Finally, the video is extracted through rendering and composited into video content. [Figure 5] illustrates the process of creating a Full 3D-based virtual human.

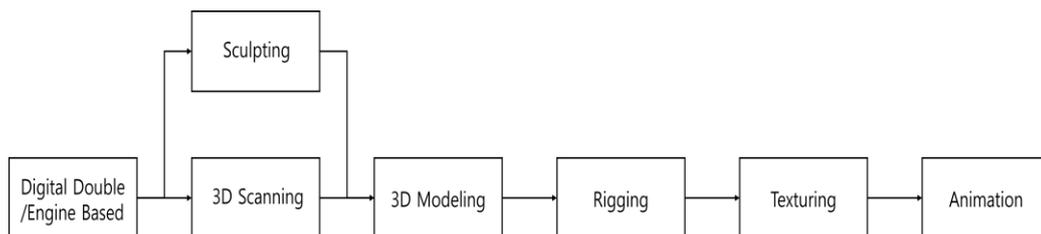


Figure 5. Full 3D based virtual human production process

3.2.2 Image Synthesis Methods (Deepfake, Generative AI)

The Deepfake method differs from the previous two approaches by using real-life models. It represents the actual person's body as is and uses artificial intelligence to train the face of a virtual character, which is then synthesized onto the face of the existing model. Since only the face is synthesized as a virtual character, it

allows for the most realistic representation. After an initial learning period, additional content can be rapidly produced at a low cost, enabling the creation of a large amount of content in a short time. However, there are drawbacks, such as difficulties in obtaining successful results when the implementation angle is limited or when facial features are obscured during movements. Therefore, Deepfake methods are primarily designed for frontal facial production and are limited in usage, mainly for social media content like Instagram, Reels, and Kiosk. Most virtual humans based on Deepfake appearing in dramas or variety shows exhibit a fixed frontal appearance, indicating a limited application. Some employ additional synthesis processes to create artistic works. Many virtual human companies in South Korea utilize this method. [Figure 6] illustrates a virtual human created using the Deepfake method.



Figure 6. Virtual human with deepfake method

Generative AI is a technology based on deep learning models that learn from various types of data, such as text, voice, and images, to generate natural content. Specifically, it learns statistical patterns from given data, allowing it to create new content. Using this technology, attempts to create virtual humans have increased, particularly by collecting real-life images of individuals and training AI models to generate results resembling human appearances. Unlike the conventional Deepfake method that synthesizes virtual faces onto real images, Generative AI creates entirely new images. With a substantial amount of data, it can efficiently generate a large number of images. While it excels in producing photo results, there are still technical limitations for video implementation, as challenges persist in achieving consistency in output and accurately capturing the creator's intent. Ongoing research explores various applications of Generative AI for creating consistently unified videos, but commercialization is still challenging.



Figure 7. Virtual human with generated AI method

The virtual human production process using image synthesis methods differs from Full 3D in that it requires an initial phase of collecting/generating training data. Typically, 2D images representing the appearance of virtual humans are collected and used as training data. In cases where data for non-existent virtual characters is needed, 3D modeling may be used to create models, and images are then extracted for use. Once the training data collection is complete, for Deepfake synthesis, real-life models are filmed, and the synthesized video is completed through a face swap process based on the collected training data. For Generative AI, synthesis is possible without a separate filming of real-life models, but there may be challenges in achieving desired poses. [Figure 7] illustrates Generative AI example.

The advent of this technology marks a revolutionary shift, simplifying all the manual processes involved in face synthesis for traditional Visual Effects (VFX). It enables obtaining synthesized videos at a low cost and in a short time. However, in terms of the quality of the final video, limitations such as facial occlusion errors, challenges in representing gaze processing for teeth and pupils, make it unsuitable for direct use in movies or dramas. Additional post-processing is essential to enhance the quality of the final video. [Figure 8] illustrates the virtual human production process using image synthesis methods.



Figure 8. Image based virtual human production process

3. Analysis of Utilization Cases in Video Contents

3.1 Utilization Cases of Virtual Humans

3.1.1 Digital Double (Digital Double)

Full 3D-based Digital Double, capable of achieving high-quality results, has been prominently used in numerous Hollywood films. An example of production utilizing Digital Double is the 2021 film "Free Guy." The movie explores the story of NPCs (Non-Player Characters) gaining consciousness in the metaverse game 'Free City.' Antwan, the villain, creates a final weapon called 'Dude,' resembling the protagonist Guy (Ryan Reynolds). Digital Double technology was employed by first creating a digital replica of the protagonist's face through 3D scanning. The body was portrayed by a separate actor, and the face was replaced during post-production using a face swap technique. [Figure 9] and [Figure 10] illustrates the application of Digital Double in the movie.



Figure 9. Application examples the digital double in 'Free Guy'

The introduction of highly realistic virtual humans, offering a level of realism comparable to real individuals, has brought fresh enjoyment to movie audiences. Despite Dude's brief appearance in the film, it was utilized in promotional content, such as Ryan Reynolds' personal social media channel featuring content named "Next Level Reynolds." However, it's notable that scenes were directed to hide the face in all instances except for interview scenes in the videos posted on social media, serving as a method to minimize the costly and time-consuming aspects of Digital Double work.



Figure 10. Application examples the digital double for Ryan Reynolds

Additionally, the 2021 film "Gemini Man" depicted a story where retired special forces operative Henry Brogan (Will Smith) confronts his clone sent to assassinate him, showcasing a notable example of using Digital Double to realistically portray Brogan's clone. [Figure 11] illustrates digital double case in Gemini Man.



Figure 11. Application examples the digital double in Gemini Man

In South Korea, virtual humans are not only used as part of film technology but also created for advertisements and TV appearances, generating a new content market. An exemplary case is the virtual human "Rozy" created by Locus X. Rozy, a Full 3D virtual human model, gained high recognition for its realistic

facial features and movements, showcased in domestic advertisements, music videos, and dramas. [Figure 12] shows scenes from the Tving original drama "Head of Internal Medicine," where Rozy appears as a nurse, and the music video "Fly so Higher."



MV - Fly so higher



TV DRAMA – Dr Park

Figure 12. Application examples the digital double In 'Rozy'

3.1.2 Engine-Based (Game Engine Based)

Unlike real-time 2D virtual humans using conventional deepfake optimized for frontal speech without additional video synthesis, engine-based virtual humans allow real-time interactions and implementation of various angles, clothing, hair, and movements. An exemplary case is the virtual human idol group "MAVE," produced by MetaVerse Entertainment, a subsidiary of Netmarble. MAVE debuted on MBC's "Show! Music Core," gaining attention for its debut stage. The music video for the pre-released title track "Pandora" quickly achieved over a million views within a day. MAVE stands out due to its high-quality, realistic 3D models, unlike traditional virtual humans that synthesize with real individuals. Utilizing game engine technology, MAVE was produced using Epic Games' Unreal Engine, specifically the "MetaHuman Creator" tool, streamlining the character creation pipeline. With about 800 automated facial expressions and shared animations, MAVE achieved a high level of efficiency in production. [Figure 13] showcases scenes from MAVE's music video.



Figure 13. Music video for game-engine based virtual human idol 'MAVE'

In the field of broadcast video content, examples such as "Avatar Singer," aired by MBN, and "AvaDream," broadcasted on TV Chosun in 2022 [Figure 14], represent prominent instances of real-time game engine applications. These programs showcase participants presenting their avatar virtual humans on stage, engaging in real-time motion capture behind the scenes for performances, forming a competitive music survival program. Utilizing game engines and establishing a live link system, they synchronize the movements of real individuals

with virtual humans in real-time. Additionally, to broadcast the real-time appearance of virtual humans to the audience, a Virtual Production (VP) system was concurrently incorporated.



Figure 14. MBN 'Avatar Singer', TVCHOSUN 'Ava-Dream'

Despite each program incurring production costs exceeding 1 billion KRW, the fusion of emerging technologies such as virtual humans and the metaverse initially captured significant attention. However, the perceived drawbacks include lower-than-expected quality and limitations in 3D graphics and simulation, leaving room for improvement.

3.1.3 Deepfake

Moving on to Deepfake technology, it is widely adopted due to its economic viability and high quality. However, being 2D-based, it faces limitations in metaverse platform utilization and challenges in diverse perspectives. Noteworthy deepfake applications involve de-aging scenarios, exemplified by the 2022 Adidas advertisement "THE IMPOSSIBLE RONDO" [Figure 15]. This ad portrays the past and present appearances of football player Messi, who participated in five World Cups. Realistic expression involved collecting and learning image data, including Messi's hairstyle and uniform, over 16 years, selecting a model with a physique similar to Messi's, and utilizing deepfake technology to synthesize the face.



Figure 15. Application example of 'The Impossible Rondo' using Deepfake method

In a domestic context, KB Life Insurance released a 2022 ad campaign [Figure 16], leveraging deepfake technology to reproduce actress Yoon Yeo-jung's appearance in her 20s. Aligning with the service's concept of providing comprehensive coverage throughout an individual's life, the ad showcases both past and future appearances using deepfake technology.



Figure 16. Application example of 'KB advertise' using Deepfake method

A representative case of deepfake usage in entertainment content is the 2021 TVing program "I Want to be a Star in the Virtual World" [Figure 17]. This program employed deepfake technology to create virtual humans for six participants, aiming to become virtual stars in a metaverse space. Real-time voice modulation was added for identity concealment, enhancing viewer engagement. Although positioned as a metaverse program, challenges such as restricting participants' movements and difficulties in eye movement and gaze expression highlighted issues that future developments in deepfake technology need to address.



Figure 17. Application example of TVing using Deepfake method

Moving on to Generative AI applications, the utilization of virtual humans is a captivating field. By using Generative AI, it is possible to create virtual characters capable of natural interaction and conversation, finding applications in various fields such as entertainment, education, counseling, and marketing. For example, Generative AI-driven avatars could interact with users, providing counseling or educational programs. Additionally, virtual humans could be integrated into marketing campaigns to enhance product promotion and customer service. Despite the rapid content creation capabilities of Generative AI, its integration with other technologies, such as voice AI services like ChatGPT, is seen as a key strategy for increased versatility [Figure 18].



Figure 18. Generated AI service 'virtual announcer'

4. DISCUSSION

In this study, we investigated the classification and characteristics of virtual human production technologies. Through an analysis of domestic and international video content cases, we examined the impact of each technology on the production environment. The virtual human production technologies were categorized as Digital Double, Game Engine, Deepfake, and Generative AI. We analyzed cases utilizing these technologies.

The Full 3D-based Digital Double method can produce high-quality results with unrestricted perspectives for expressive freedom. However, it was noted for its high production costs and time-consuming nature. The Game Engine method involves creating a full body using real-time physics engines like Unreal. While it allows real-time movement and facial expression, there are limitations in achieving realistic graphic representations. Deepfake employs Deep Learning to synthesize a virtual face onto a 2D image-based real model. Although cost-effective and widely used due to its fast workflow, it has limitations such as lower facial expression resolution and restricted movement angles. Generative AI, exemplified by Stable Diffusion, constructs a massive image dataset to generate images based on text input. While advantageous for rapid results, it faces challenges in producing consistent images and has limitations in video implementation.

The investigation of content utilizing these technologies revealed that some limitations not only lower video quality but also lead to viewer discomfort. To activate and expand the use of virtual humans, various attempts to overcome these limitations are necessary. Currently, most content production relies on a single virtual human production technology, but employing different technologies for each screen through content analysis could be a solution. For scenes requiring wide-angle views and emotional expressions, the Digital Double method could be used, while Deepfake could be applied to scenes focusing on front shots. Recognizing the limitations and utilizing the strengths of each production method can secure content quality and save both costs and time. However, achieving this requires a high understanding of these technologies during production. [Figure 19] shows that propose process.

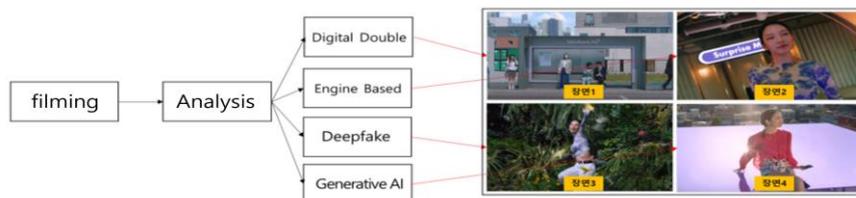


Figure 19. Virtual human production process through screen analysis

5. Conclusion

As interest in virtual humans capable of human-like expressions increases, discussions on legal and regulatory aspects beyond efforts for video content application will be necessary. Issues such as copyright, ownership, regulation of illegal activities such as harassment in metaverse spaces, ethical concerns, and personal information protection require consideration. Appropriate ethical guidelines and legal measures are crucial to protect user privacy and rights and regulate the use of such technologies. However, excessive regulation could hinder the development of emerging industries, so a cautious approach to regulatory and institutional measures is advised.

Expectations are high for the expansion of diverse video content through the technological advancement of virtual humans.

Acknowledgement

This work was supported by Seoul Institute of the Arts in 2024.

References

- [1] Kim, Kwang-Jib. "Evolution of the Real and Virtual Worlds Through Meta-universe Cases." *Broadcasting and Media*, vol. 26, no. 3, 2021.
- [2] Swartout, W., Gratch, J. Hill, Jr, Randall, Marsella, S., Rickel, Jeff, Traum, David. "Toward Virtual Humans". *AI Magazine*. 27, 2006. pp.96-108.
- [3] Magnenat-Thalmann, N., Thalmann, D, "Virtual humans: thirty years of research, what next?", *Visual Comput*, 21, 2005. pp.997–1015.
<<https://doi.org/10.1007/s00371-005-0363-6>>
- [4] Kwak, Bo-Eun, & Heo, Jung-Yoon. "Analysis of the Impact of Non-verbal Expression Styles on the Acceptance of Digital Human Service - Focused on Hotel Services." *Proceedings of the Korean Society of Design Conference*, 2021.
- [5] Seo, Young-Ho, Oh, Moon-Seok, & Han, Gyu-Hoon. "Current and Future of Digital Humans." *Broadcasting and Media*, vol. 26, no. 4, 2021.
- [6] Lee, Da-Young, & Han, Jung-Yeop. "A Study on the Utilization Cases and Characteristics of Visual Media Based on Digital Human Production Technology - Focused on Real-time / Non-real-time Videos." *Journal of the Korean Society of Spatial Design*, vol. 17, no. 7, 2022.
- [7] Lee, Yeon-Kyung, Kim, Hyun-Ji, & Park, Geon-Woo. "Presentation and Status Study on Types of Virtual Singers Using Metaverse." *Journal of the Korean Entertainment Industry Association*, vol. 16, no. 6, 2022.
<<https://doi.org/10.21184/jkeia.2022.8.16.6.107>>