



# A Study on the Realization of Virtual Simulation Face Based on Artificial Intelligence

Zheng-Dong Hou<sup>1</sup>, Ki-Hong Kim<sup>2\*</sup>, Gao-He Zhang<sup>3</sup>, and Peng-Hui Li<sup>4</sup>

<sup>1</sup>Department of Visual Contents, Dongseo University, Busan 47011, Republic of Korea

<sup>2</sup>Department of Visual Animation, Dongseo University, Busan 47011, Republic of Korea

<sup>3</sup>Department of Visual Contents, Dongseo University, Busan 47011, Republic of Korea

<sup>4</sup>Department of Visual Contents, Dongseo University, Busan 47011, Republic of Korea

## Abstract

In recent years, as computer-generated imagery has been applied to more industries, realistic facial animation is one of the important research topics. The current solution for realistic facial animation is to create realistic rendered 3D characters, but the 3D characters created by traditional methods are always different from the actual characters and require high cost in terms of staff and time. Deepfake technology can achieve the effect of realistic faces and replicate facial animation. The facial details and animations are automatically done by the computer after the AI model is trained, and the AI model can be reused, thus reducing the human and time costs of realistic face animation. In addition, this study summarizes the way human face information is captured and proposes a new workflow for video to image conversion and demonstrates that the new work scheme can obtain higher quality images and exchange effects by evaluating the quality of No Reference Image Quality Assessment.

**Index Terms:** Artificial Intelligence, Deepfake, Facial animation, Animation

## I. INTRODUCTION

People rely on facial expressions to indicate emotions and intentions. Due to the sensitivity of humans to subtle facial movements, many details of features such as muscle movements, wrinkles and skin composition must be considered when creating realistic facial expressions using computer technology, making it difficult to achieve realistic facial animations from computer graphics models. Realistic facial animation production consists of facial modelling and animation data acquisition techniques. Facial modelling was earlier used to create models and mapping using 3D software, then the face models were given to skeletons for control and then animated, to obtain realistic face models, researchers worked with laser scanning or image scanning [1-3]. Face animation

data acquisition techniques include speech-driven techniques, image-based techniques, and data capture techniques [4,5]. Nguyen, Tan-Nhu, et al. (2020) proposed a computer vision system for data acquisition using a non-contact Kinect sensor for real-time tracking of rigid head and non-rigid face imitation movements while designing subject-specific Texture generation subsystem to enhance the realism of generative models with texture information. A head animation subsystem with a graphical user interface was also developed. Pan, Ye, et al. (2022) proposed a real-time motion capture system named MienCap by combining traditional blend shape animation techniques with machine learning models. It drives character expressions in a geometrically consistent and perceptually efficient way, a system that could potentially find its way into VR filmmaking and animation pipe-

Received 20 October 2022, Revised 03 January 2023, Accepted 06 January 2023

\*Corresponding Author Ki-Hong Kim (E-mail: khkim@g.dongseo.ac.kr)

Department of Visual Animation, Dongseo University, Busan 47011, Republic of Korea

Open Access <https://doi.org/10.56977/jicce.2023.21.2.152>

print ISSN: 2234-8255 online ISSN: 2234-8883

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © The Korea Institute of Information and Communication Engineering

lines. Ye, Yuping, Zhan Song, and Juan Zhao (2022) developed a facial acquisition system based on an infrared structured light sensor to obtain high-fidelity and accurate facial expression models. Accurate and dense point clouds, then morphing template models into captured facial expressions, textured real-time 3D meshes using high-resolution images captured by three color cameras. Gu K, Zhou Y, Huang T (2020) propose a Landmark-driven network to generate realistic speaking facial animations, where more facial details are created, preserved, and transferred from multiple source images rather than a single source image. The acquisition subnetwork learns to carefully warp and merge facial regions directly from five source images with unique landmarks, while the learning pipeline renders facial organs from the training facial space to compensate. K Vougioukas, S Petridis, M Pantic (2019) present a system for generating talking head videos, which achieves this by using a temporal GAN with 2 discriminators that capture different aspects of the video to produce facial animations. In recent years, artificial intelligence techniques have been widely used for their ability to solve complex problems. Due to this ability, researchers have turned to machine learning to improve the quality and feasibility of facial animation [S.W. Bailey, Prashanth Chandran, 2020] [Thanh Thi Nguye, 2019] [T. Karras, 2017]. With a sufficiently large dataset, AI can learn how to produce facial animations for a variety of humans. The increased speed, quality, and usability of the AI model learning results make it an excellent solution to some of the major problems of traditional approaches to facial animation. Since the research on AI in facial animation is relatively new and information on its use cases is limited, Deepfake is one of the AI techniques and this study will focus on Deepfake technique, detailing the important parts of the realistic facial animation solution as well as the improved parts of the workflow, and finally the quality evaluation of the produced

facial animations to demonstrate the solution’s Feasibility. Fig. 1 is an overview of the proposed scheme in this thesis. First, the same face model as the model is created by photogrammetry, and then a simulated avatar is created using a plug-in of the game engine, and Audio2Face provides the face animation data for the avatar. Then a digital camera is used to capture the model’s face to get a video file, and Adobe Media Encoder is used to decompose the video into pictures delivered to the AI model for training, and finally a hyper-realistic face animation is obtained.

## II. DEEPPAKE PREPARATION WORK

### A. Introduction of Decoder and Encoder

Deepfake refers to techniques for specific types of synthetic media in which a person in an image or video swaps faces with another person. Common underlying mechanisms for deepfake are deep learning models, such as autoencoders and generative adversarial networks (Gans), which have been widely used in computer vision. Deepfake methods typically require large amounts of image and video data to train models to create photo-realistic images and videos [15-17]. Because of the simplicity of Deepfake applications can be used by both professionals and users with low computer skills [18].

In Deepfake technique work is done by computer to learn two data sets to generate AI model by which face interchange can be achieved, auto encoder size, encoder size and decoder size are important components of AI model, auto encoder size is the middle layer of AI model and affects the number of AI complementary images generated. The decoder divides the extracted face image matrix style to be learned by the model, the larger the size the more the num-

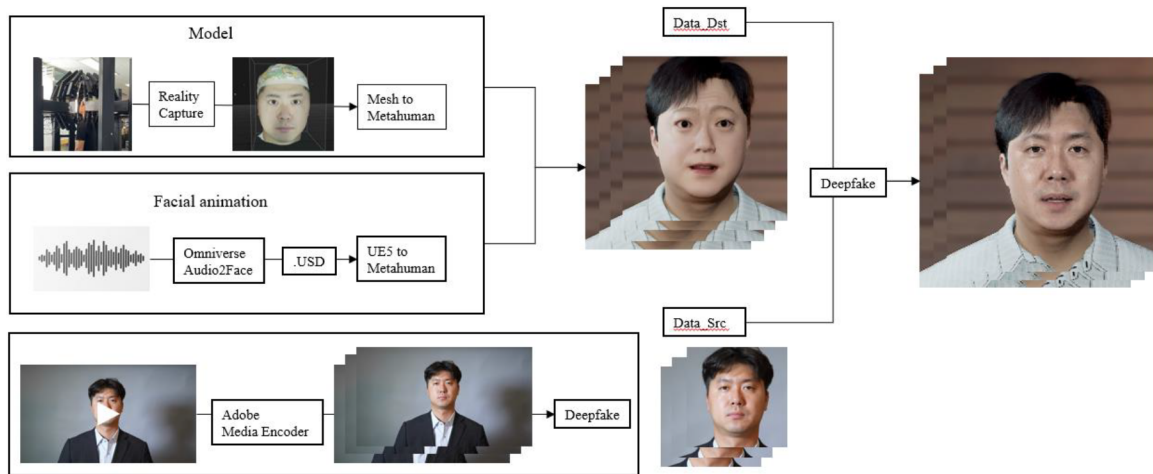
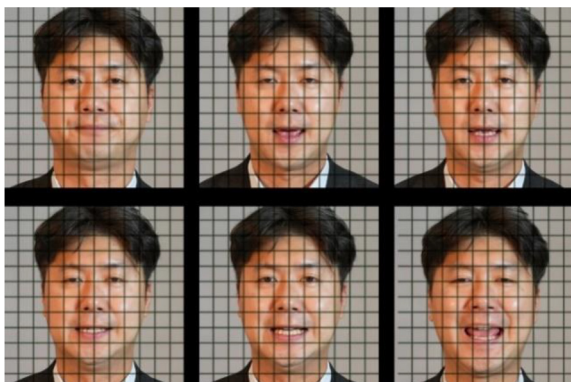


Fig. 1. Create super-realistic face animation scheme illustration.



**Fig. 2.** Decoder and encoder subdivision images.

ber of divided squares, while the faster the AI model learns. Fig. 2 shows the autoencoder and encoder working process of Deepfake, where the decoder converts the data squares in the model into images [19,20].

## B. Reduce Loss During Image Conversion

Deepfake’s encoder is divided into images in the form of square matrix, so the clarity of AI model learning picture material is crucial, Deepfake data collection method generally by digital camera shooting video, and then the video file into pictures, with the development of film and television technology, there are many ways to convert video to pictures, but due to the different compression methods, different conversion methods The quality of the pictures obtained is also different due to the different compression methods, the reason for this difference lies in the different picture compression techniques. For digital images, each pixel is used as a sampling point and has a corresponding sampling value. The finer the image segmentation, the more the number of pixels, the more the sampling points, the higher the image clarity; conversely, the fewer the number of pixels, the lower the image clarity. Since the human eye has different subjective sensitivity to brightness and chromaticity, it is difficult to distinguish the difference in quality between pictures with the naked eye, so the source video with the same conversion method is different, and it needs to be calculated by a function to carry out the work of evaluation of the clarity of the non-reference picture. The general principle of BRISQUE algorithm is to extract the mean subtracted contrast normalized (MSCN) coefficients from the image, and to fit the MSCN Tenengrad function is a gradient-based function. In image processing, it is generally considered that well-focused images have sharper edges and therefore have larger gradient function values. The Lapras algorithm is sensitive and can obtain fast results in images of different sizes [21-24].

Adobe Media Encoder is a video and audio encoding application that can be used for different applications and audiences in various distribution formats to Audio and video files are encoded in a variety of distribution formats. Adobe Media Encoder combines the numerous settings provided by the major audio and video formats and includes preset settings specifically designed to export files compatible with specific delivery media [25]. Deepfake technology uses FFmpeg for image conversion, which is a set of formats that can be used to record, convert digital audio and video, and can convert them to streams as an open-source computer program. It is licensed under the LGPL or GPL. It provides a complete solution for recording, converting, and streaming audio and video. It contains the very advanced audio/video codec library libavcodec. To ensure high portability and codec quality, much of the code in libavcodec was developed from scratch [26,27]. To ensure the objectivity of the data, this study will use these two methods to convert the video files to pictures, and then work on the No Reference Image Quality Assessment by three algorithms, BRISQUE, Tenengrad, and Laplacian, which are converted to Python code in Spyder (Anaconda3) is implemented, the following table shows the specific code.

**Table 1.** Implementation code

```
def BRISQUE (img):
    model_path="brisque_model_live.yml"
    range_path="brisque_range_live.yml"
    result_static=cv2.quality.QualityBRISQUE_compute(img,model_path,
range_path)
    score=np.mean([i for i in result_static if (i != 0 and not np.isinf(i))])
    score=0 if np.isnan(score) else score
return score
def Tenengrad (img):
    img_gray=cv2.cvtColor(img,cv2.COLOR_RGB2GRAY)
    fx=np.power(img_gray[:,1:-1]*2 - img_gray[:,:-2]-img_gray[:,2:],2)
    fy=np.power(img_gray[1:-1,:]*2 - img_gray[:,:-2,:]-img_gray[:,2:],2)
return np.sqrt(fx.mean()+fy.mean())
def Laplacian (img):
    img_gray=cv2.cvtColor(img,cv2.COLOR_RGB2GRAY)
    s=img_gray[1:-1,:-2]+img_gray[1:-1,2:]+img_gray[:,:-2,1:-1]+img_gray
[2:,1:-1]-4*img_gray[1:-1,1:-1]
    return np.abs(s).mean()
```

Deepfake’s video to picture command provides two kinds of formats: PNG and JPG, so two sets of picture clips named DF\_PNG and DF\_JPG were obtained by Deepfake’s command, and the video was converted to PNG and JPG using Adobe Media Encoder. The total duration of this experiment is 36 s, the resolution is 3840\*2160 px, and the frame rate is 60 F/s. The DF\_PNG image group is decomposed into 2190 images, each image resolution is 3840\*2160 px, and the bit depth is 24. The average image size in the DF\_JPG image group is 680KB, and the average image size in the DF\_PNG

image group is 8 MB. using Adobe Media Encoder to decompose the same video, the final AME\_PNG image group decomposes 2910 images, with a resolution of 3840\*3160 px image and a bit depth of 24. The average image size in the AME\_JPG image group was 3 MB, and the average image size in the AME\_PNG image group was 8.5 MB. 30 images were randomly selected from each image group for the evaluation of the clarity of the non-reference images.

The results in Fig. 3 show the results of four groups of picture clarity measurements. The smaller the value of BRISQUE represents the better picture clarity, and the larger the value of Tenengrad and Laplacian, the better the picture clarity. Among the results of the four groups of pictures, the best picture clarity is found in the case of converting the video to JPG format by Adobe Media Encoder, so this study proposes to use Adobe Media Encoder to produce the picture material used for AI learning in Deepfake.

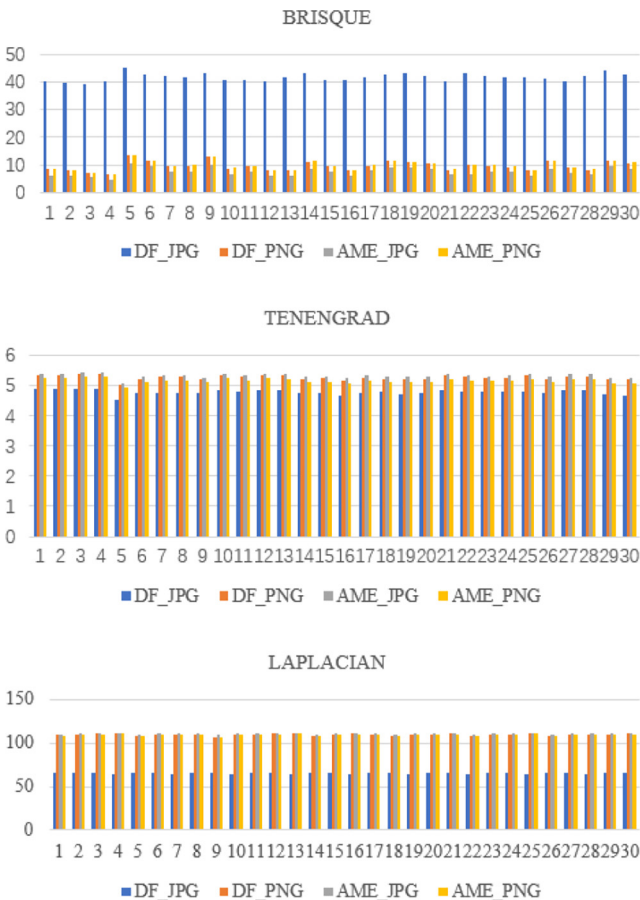


Fig. 3. BRISQUE, Tenengrad, and Laplacian evaluation Data.

### III. AI MODEL TRAINING

#### A. Pre-training

To reuse AI models generated by Deepfake, they must be pre-trained because Deepfake maps the style effect of the previous training. This study uses DeepfakeLab for AI model training. The AI model file size does not change with the number of materials or training time, and pre-training can save the mapping effect after training as a starting point, so the AI model after pre-training is more efficient for reuse. The AI model generated by Deepfake includes 6 files, among which the replacement of \_SAEHD\_data.dst file can complete the reuse of pre-training, which can reduce a great deal of training time to imaging. Because the pre-training only learns on a set of picture material, only mask training is set to True. if the face angle in Src is missing will lead to longer training time, so the AI model needs to add various face angles during the pre-training, Fig. 4 shows the pre-training process of the AI model used in this study, you can see that the simulation of the character expression has been basically completed and The Loss value is the evaluation value of the face exchange result, and a smaller value means a better face exchange effect, so it is judged that the AI model can be used for formal training.

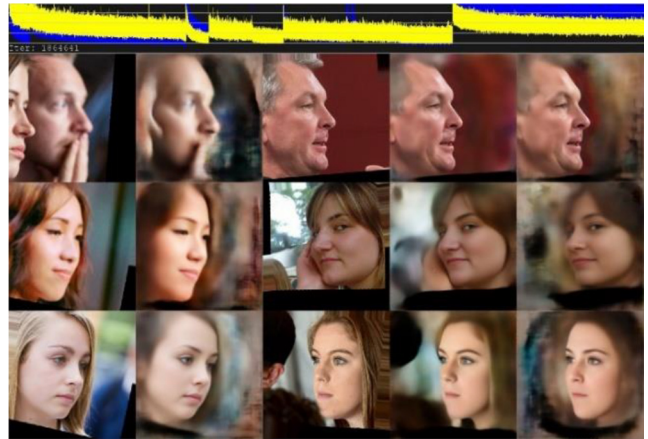


Fig. 4. AI model pre-training process.

#### B. Formal Training

Two sets of data (Src & Dst) are required for formal training of the AI model. Dst material used in this thesis is MetaHuman animation material, MetaHuman character model is made by real people through photo-scanned face data, and face ani-

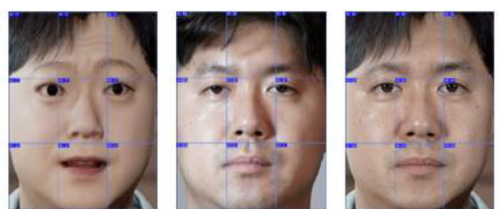


mation is made by Omniverse Audio2Face, Fig. 5. Src material is made by digital camera shooting 13 videos. Adobe Media Encoder output a total of 33,554 valid Jpg images.

When training, we need to turn on masked\_training and random\_warp, the loss value will drop very fast at the beginning of training, and the image will appear quickly. When most of the outline of the fifth column in the preview image is basically accurate, you need to set lr\_dropout to True to continue training, when the loss value drops very slowly or there are signs of rebound one after another. At this time, check the eye direction in the training results, in order to ensure the clarity of the eye direction and the side profile of the face need to set eyes\_mouth\_prio and uniform\_yaw to True, set lr\_dropout and random\_warp and random\_flip to False. the final stage to set the value of Gan, GAN The value of Gan is the speed of AI model learning, if set too large will lead to AI model error, so the value range is better to start with 0.0001, and the setting of GAN will increase the GPU occupancy.

#### IV. ANALYSIS OF EXPERIMENTAL RESULTS

In this study, the synthesized face after training is divided into 9 parts (Fig. 5) equal to the original face, and SSIM evaluation is performed for the real face and Deepfake face from the whole and each part, respectively, Structural Similarity Index Measure (SSIM) is a metric to measure the degree of similarity between two digital images. This SSIM evaluation was performed in the software Matlab, and the SSIM index was 0.6398 for the overall two face images and 0.3326, 0.7299, 0.4737, 0.694, 0.7127, 0.8046, 0.6696, 0.7664, and 0.5956 for each part. observing the images and SSIM values reveal that Deepfake simulates and replicates the real face details and has good effect on the processing of light and shadow (Table 2).



3D face model      real face      Deepfake results

**Fig. 5.** Image cut-out display.

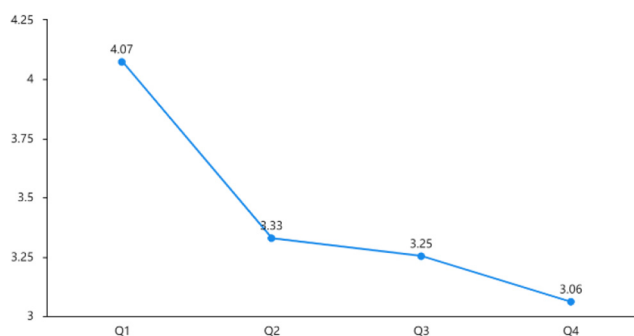
**Table 2.** SSIM evaluation results

SSIM Evaluation Values		
0.333	0.730	0.474
0.694	0.713	0.805
0.670	0.766	0.596
SSIM Overall Evaluation: 0.6398		

Although it was objectively verified that the AI model generated realistic face animations, a test of the Valley of Terror effect was needed to confirm whether viewers perceived the animation as realistic. Regarding the measurement of the questionnaire this study was based on some of the questions proposed by Bartneck, Kulic, and Croft (2009) [28]. The scale itself was set up for robots, so this study first screened through the respondents to set up suitable questions, and after screening a total of four topics were identified, and the topics were set up as shown in Table 3. this survey was conducted at Dongseo University in Busan, Korea, starting on October 15, 2022, and ending on October 18, 2022, the survey was conducted offline, allowing respondents to watch a produced video and record their feelings based on the questionnaire questions. 67 people participated, with the participants consisting mainly of professors and students.

**Table 3.** The questionnaires defined used to assess the impressions of the robot facial animation

Please rate your impression of the Facial animation on these scales:						
Fake	1	2	3	4	5	Natural
Machinelike	1	2	3	4	5	Humanlike
Artificial Moving	1	2	3	4	5	Lifelike Moving
Rigidly	1	2	3	4	5	Elegantly



**Fig. 6.** Average Comparison.

After averaging the questionnaire data (Fig. 6), the pair-produced video scores higher in the case of comparison with the real person, and the motion perception score of the animation is also higher than the middle value, thus, the feasibility of the solution proposed in this study can be demonstrated. To summarize respondents' feedback outside of the questionnaire questions, the animation of the eyes and the authenticity of the hair can affect the overall judgment.

#### V. CONCLUSION

The application of artificial intelligence technology in the film and television industry is becoming more and more

mature, with the development of the Internet, the variety of film and television content has become rich, movies, TV series, etc. using artificial intelligence technology to change the age of actors and even let the deceased actors appear in the virtual world, the development of film and television technology has made the production of virtual simulators become simple, especially the facial animation of simulators can already be made in various ways through facial capture technology, etc. However, both facial capture technology and traditional facial animation production technology require a lot of labor and time cost. The solution proposed in this thesis, using Deepfake to train AI models to learn real human faces and expression movements, and then replace the faces of CG characters, because AI models have the feature of reuse, and there are also skin and other facial details information are almost perfectly replicated, which can greatly save the time cost and labor cost of facial animation production, which is a new attempt in CG film and television industry. From the experimental results and the results of the subjectivity survey, the solution proposed in this study has good results in terms of real human production, however, the fact that the face has only simple animation and hair rendering quality leads to a small difference in the mean score of the face animation evaluation, although it is higher than the middle value. Future studies on eye animation and hair realism will be attempted.

The material obtained from the Internet in this study was used for study and research purposes only, and all data that might violate personal interests were removed after the experiment. Consent for the use of face information was obtained from the individual.

## REFERENCES

- [ 1 ] L. Dzelkalēja, J. K. Knēts, N. Rozenovskis, and A. Silītis, "Mobile apps for 3D face scanning," *Proceedings of SAI Intelligent Systems Conference*, pp. 34-50, 2022. DOI: 10.1007/978-3-030-82196-8\_4.
- [ 2 ] P. Amornvit and S. Sanohkan, "The accuracy of digital face scans obtained from 3D scanners: an in vitro study," *International Journal of Environmental Research and Public Health*, vol. 16, no. 24, p. 5061, Dec. 2019. DOI: 10.3390/ijerph16245061.
- [ 3 ] Z. Wang, "Robust three-dimensional face reconstruction by one-shot structured light line pattern," *Optics and Lasers in Engineering*, vol. 124, p. 105798, Jan. 2020. DOI: 10.1016/j.optlaseng.2019.105798.
- [ 4 ] A. Richard, C. Lea, S. Ma, J. Gall, F. de la Torre, and Y. Sheikh, "Audio-and gaze-driven facial animation of codec avatars," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. Waikoloa, USA, Jan. 2021. DOI: 10.1109/wacv48630.2021.00009.
- [ 5 ] V. Barrielle, and N. Stoiber, "Realtime performance-driven physical simulation for facial animation," in *Computer Graphics Forum*, vol. 38, no. 1, Feb. 2019. pp. 151-166. DOI: 10.1111/cgf.13450.
- [ 6 ] T-N. Nguyen, S. Dakpé, M-C. Ho Ba Tho, and T.-T. Dao, "Real-time computer vision system for tracking simultaneously subject-specific rigid head and non-rigid facial mimic movements using a contactless sensor and system of systems approach," *Computer Methods and Programs in Biomedicine*, vol. 191, p. 105410, Jul. 2020. DOI: 10.1016/j.cmpb.2020.105410.
- [ 7 ] Y. Pan, R. Zhang, J. Wang, N. Chen, Y. Qiu, Y. Ding, and K. Mitchell, "MienCap: Performance-based facial animation with live mood dynamics," in *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, Christchurch, New Zealand, pp. 654-655, 2022. DOI: 10.1109/VRW55335.2022.00178.
- [ 8 ] Y. Ye, S. Zhan, and Z. Juan, "High-fidelity 3D real-time facial animation using infrared structured light sensing system," *Computers & Graphics*, vol. 104, pp. 46-58, May. 2022. DOI: 10.1016/j.cag.2022.03.007.
- [ 9 ] K. Gu, Y. Zhou, and T. Huang, "Flnet: Landmark driven fetching and learning network for faithful talking facial animation synthesis," *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 07, pp. 10861-10868, 2020. DOI: 10.1609/aaai.v34i07.6717.
- [ 10 ] K. Vougioukas, S. Petridis, and M. Pantic, "End-to-end speech-driven realistic facial animation with temporal GANs," *CVPR Workshops*, pp. 37-40, 2019.
- [ 11 ] S. W. Bailey, D. Omens, P. Dillorenzo, and J. F. O'Brien, "Fast and deep facial deformations," *ACM Transactions on Graphics*, vol. 39, no. 4, Aug. 2020. DOI: 10.1145/3386569.3392397.
- [ 12 ] P. Chandran, D. Bradley, M. Gross, and T. Beeler, "Semantic deep face models," in *2020 International Conference on 3D Vision (3DV)*, Fukuoka, Japan, pp. 345-354, 2020. DOI: 10.1109/3DV50981.2020.00044.
- [ 13 ] T. Karras, T. Aila, S. Laine, A. Herva, and J. Lehtinen, "Audio-driven facial animation by joint end-to-end learning of pose and emotion," *ACM Transactions on Graphics*, vol. 36, no. 4, pp. 1-12, Jul. 2017. DOI: 10.1145/3072959.3073658.
- [ 14 ] T. T. Nguyen, C. M. Nguyen, T. D. Nguyen, T. Duc, S. Nahavandi, "Deep learning for deepfakes creation and detection," *arXiv preprint arXiv:1909.11573*, vol. 1, no. 2, p. 2, Sep. 2019.
- [ 15 ] A. Tewari, M. Zollhoefer, F. Bernard, P. Garrido, H. Kim, P. Perez, and C. Theobalt, "High-fidelity monocular face reconstruction based on an unsupervised model-based face autoencoder," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 357-370, Feb. 2020. DOI: 10.1109/TPAMI.2018.2876842.
- [ 16 ] J. Lin, Y. Li, and G. Yang, "FPGAN: Face deidentification method with generative adversarial networks for social robots," *Neural Networks*, vol. 133, pp. 132-147, Jan. 2021. DOI: 10.1016/j.neunet.2020.09.001.
- [ 17 ] M-Y. Liu, X. Huang, J. Yu, T-C. Wang, and A. Mallya, "Generative adversarial networks for image and video synthesis: Algorithms and applications," *Proceedings of the IEEE*, vol. 109, no. 5, pp. 839-862, May 2021. DOI: 10.1109/JPROC.2021.3049196.
- [ 18 ] T. T. Nguyen, Q. V. H. Nguyen, D. T. Nguyen, D. T. Nguyen, T. Huynh-The, S. Nahavandi, T. T. Nguyen, Q-V. Pham, and C. M. Nguyen, "Deep learning for deepfakes creation and detection: A survey," *Computer Vision and Image Understanding*, vol. 223, p. 103525, Oct. 2022. DOI: 10.1016/j.cviu.2022.103525.
- [ 19 ] I. Perov, D. Gao, N. Chervoniy, K. Liu, S. Marangonda, C. Umé, Mr. Dpfks, C. S. Facenheim, L. RP, J. Jiang, S. Zhang, P. Wu, B. Zhou, and W. Zhang, "DeepFaceLab: Integrated, flexible and extensible face-swapping framework," *arXiv preprint arXiv:2005.05535*, May 2020. DOI: 10.48550/arXiv.2005.05535.
- [ 20 ] F. Jia and S. Yang, "Video face swap with DeepFaceLab," in *International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2021)*, Harbin, China, vol. 12168, pp. 326-332, Mar. 2022. DOI: 10.1117/12.2631297.
- [ 21 ] L. Her and X. Yang, "Research of image sharpness assessment

algorithm for autofocus,” in *2019 IEEE 4th International Conference on Image, Vision, and Computing (ICIVC)*, IEEE, Xiamen, China, pp. 93-98, 2019. DOI: 10.1109/ICIVC47709.2019.8980980.

- [22] M. K. Rohil, N. Gupta, and P. Yadav, “An improved model for no-reference image quality assessment and a no-reference video quality assessment model based on frame analysis,” *Signal, Image and Video Processing*, vol. 14, No. 1, pp. 205-213, Feb. 2020. DOI: 10.1007/s11760-019-01543-z.
- [23] X. Zhou, J. Zhang, M. Li, X. Su, F. Chen, “Thermal infrared spectrometer on-orbit defocus assessment based on blind image blur kernel estimation,” *Infrared Physics & Technology*, vol. 130 p. 104538, May 2022. DOI: 10.1016/j.infrared.2022.104538.
- [24] J. Rajevenceltha and V. H. Gaidhane, “An efficient approach for no-reference image quality assessment based on statistical texture and structural features,” *Engineering Science and Technology, an International Journal*, vol. 30, p. 101039, Jun. 2022. DOI: 10.1016/j.jestch.2021.07.002.
- [25] J. Harder, “What other programs that are part of Adobe Creative Cloud can I use to display my graphics or multimedia online?,” in *Graphics and Multimedia for the Web with Adobe Creative Cloud*, Apress, Berkeley, CA, pp. 993-1000, Nov. 2018. DOI: 10.1007/978-1-4842-3823-3\_40.
- [26] X. Wu, P. Qu, S. Wang, L. Xie, J. Dong, “Extend the FFmpeg framework to analyze media content,” *arXiv preprint arXiv:2103.03539*, Mar. 2021. DOI: 10.48550/arXiv.2103.03539.
- [27] M. Gupta, S. Shah, and S. Salmani, “Improving whatsapp Video Statures using FFMPEG and Software based encoding,” in *2021 International Conference on Communication information and Computing Technology (ICCICT)*, IEEE, Mumbai, India, pp. 1-6, 2021. DOI: 10.1109/ICCICT50803.2021.9510129.
- [28] C. Bartneck, D. Kulić, E. Croft, et al., “Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots,” *International Journal of Social Robotics*, vol. 1, no. 1, pp. 71-81, Jan. 2009. DOI: 10.1007/s12369-008-0001-3.



### ZhengDong Hou

received a bachelor's degree in advertising from Qingdao Agricultural University, China (2016). Received Master's degree in Dongseo University from the Department of Visual Contents, Korea (2019). Currently a PhD student at the Department of Visual Contents, Dongseo University (2022). Research interests include virtual characters, 3D reconstruction, and artificial intelligence learning.



### KiHong Kim

2006: Alexander the Great R&D Supervisor in the animated feature film. 2007: San Antonio R&D Supervisor in the animated feature film. 2008: Carol R&D Supervisor in the animated feature film. 2010: 7 Ride Films Executive Production. 2010: Parada of PotteryIs TV series R&D Supervisor. 2010~Present: Professor of Visual Animation at Dongseo University/ Director of Software Convergence Center. Areas of interest: animation content, 3D CG, visual artificial intelligence, motion data, photo surveying and 3D implementation.



### GaoHe Zhang

born in Shandong Province, China in 1996, received a Bachelor of Arts degree from Zhongnan University of Economics and Law in 2019 and a Bachelor of Arts degree from Dongseo University in South Korea. In 2022, he obtained a master's degree in engineering from Dongseo University, and in 2022, he began to engage in artificial intelligence research at Dongseo University in South Korea.



### PengHui Li

rachelor of Arts, China-Korea Institute of New Media, Zhongnan University of Economics and Law. Studying for a master's degree, Department of Visual Contents, Dongseo University.