International Journal of Advanced Culture Technology Vol.9 No.4 385-391 (2021)
DOI https://doi.org/10.17703/IJACT.2021.9.4.385

IJACT 21-12-45

Hair Segmentation using Optimized Fully Connected Network and 3D Hair Style

[1]Junghyun Kim, [2]Yunhwan Lee, [3]Seongah Chin*

[1,2]Department of Media Software, Sungkyul University, XICOM LAB
*[3]Prof. Department of Media Software, Sungkyul University, XICOM LAB **
*{wjdgusrla125, unan98}@naver.com, solideo@sungkyul.ac.kr**

Abstract

3D modeling of the human body is an integral part of computer graphics. Among them, several studies have been conducted on hair modeling, but there are generally few studies that effectively implement hair and face modeling simultaneously. This study has the originality of providing users with customized face modeling and hair modeling that is different from previous studies. For realistic hair styling, We design and realize hair segmentation using FCN, and we select the most appropriate model through comparing PSPNet, DeepLab V3+, and MobileNet. In this study, we use the open dataset named Figaro1k. Through the analysis of iteration and epoch parameters, we reach the optimized values of them. In addition, we experiment external parameters about the location of the camera, the color of the lighting, and the presence or absence of accessories. And the environmental analysis factors of the avatar maker were set and solutions to problems derived during the analysis process were presented.

Keywords: *FCN, Avatar Maker, Hair Modeling, Face Modeling*

1. INTRODUCTION

Recently, as the technology of 3D modeling of the human body has been developed, the market size of the field using 3D modeling of the human body is expanding. The size of the game industry including the metaverse, which is the main field of human modeling is 17.93 trillion won as of December 2020, which is an increase of 9.2% compared to the previous year and is expected to increase by about 7.4% to 18.268.3 billion won in 2021. As such, 3D modeling of the human body is used in various fields such as games, movies, medicine, and more recently, VR/AR contents [1, 2] such as metaverse. As interest in 3D modeling of the human body increases, the importance of 3D hair styling is expected to increase. As a related study of 3D hair styling [3], an automatic 3D hair model restoration method using a deep neural network was developed. High-quality hair modeling can be implemented using deep neural networks [4]. Another study [5] describes a 3D hairstyle simulator using augmented reality. A 3D hairstyle simulator was developed based on NUI (Natural User Interface) and EHCI (Extended Human Computer Interaction). There are also several studies related to segmentation [6]. In this paper, to improve this problem, we propose a realistic hair styling technique that combines search-based automatic 3D hair modeling [7] and Avatar Maker [8] through FCN (Fully Connected Network). Executes partial hair region zoning using FCN specialized for regulating input portrait photos. It

Manuscript received: November 26, 2021 / revised: November 30, 2021 / accepted: December 7, 2021
Corresponding Author: solideo@sungkyul.ac.kr
Tel: +82-31-467-8198, Fax: +82-31-467-8067
Professor, Dept. of Media Software, Sungkyul Univ., Korea

also provides users with faster and more accurate face modeling using the avatar maker.

In this paper, hair zoning using FCN and face modeling using an avatar maker are described. As such, it is intended to contribute to the development of 3D hair styling technology by providing realistic hair and face modeling to users.

2. SYSTEM PROCESS

The system process in this study consists of the sequence shown in Figure 1. In the hair modeling selection step, hair segmentation of the input image is performed using MobileNet, a FCN learning module. MobileNet proceeded with a module that optimized the epoch and iteration values. And the comparison procedure is carried out for a mapping of hair modeling from the hair modeling sets, which is similar to the hair segmentation image. In addition, the face modeling configuration step outputs a picture by optimizing the position of the camera. It also sets the color of the lighting and the presence or absence of accessories. Then, input the selected image of Avatar Maker and generate the face modeling. Finally, it provides users with a combination of hair modeling and face modeling.

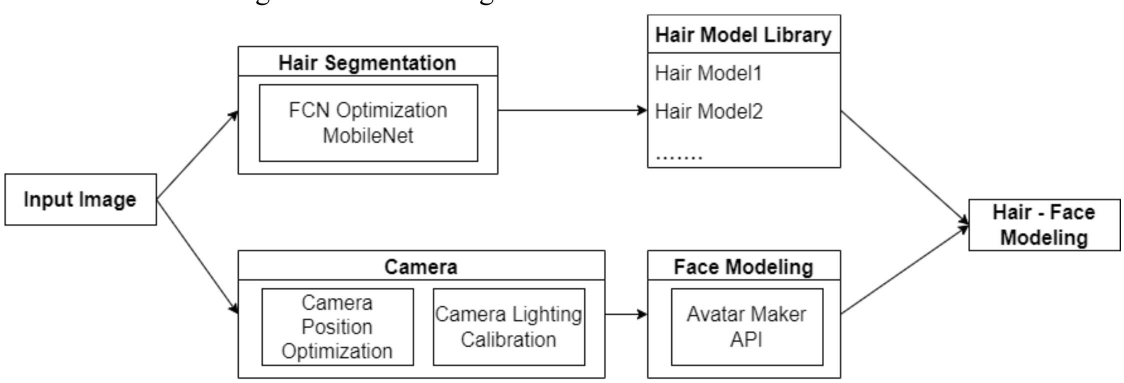


Figure 1. System process

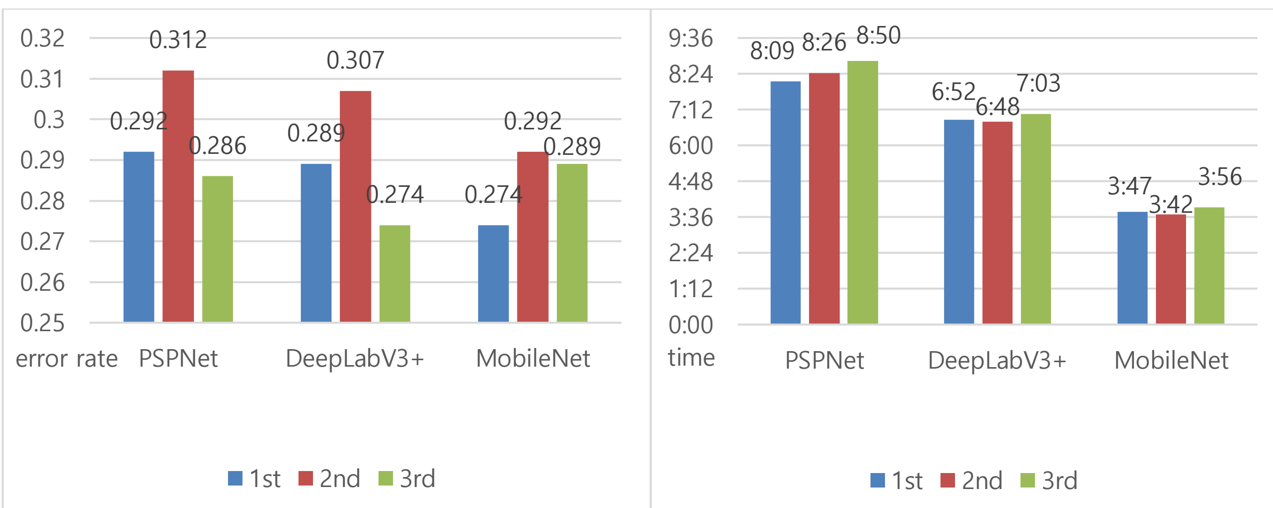
3. HAIR SEGMENTAITION USING FCN

FCN is used to extract the hair region from the image of a person. FCN is a network that stores location information by changing the fully connected layer part of the upsampling part of the existing CNN [9] to a 1 x 1 convolution layer [10]. PSPNet, DeepLab V3+, and MobileNet are selected to compare the error rate and learning speed according to the type of FCN model. The dataset used is Figaro1k [11], and as shown in Figure 2, 840 training pictures and 210 testing pictures are provided.

(a) Straight Hair (b) Short Hair (c) Men's Short Hair (d) Cully Hair

Figure 2. Figaro1k dataset training images

The difference between PSPNet (Pyramid Scene Parsing Network) and general FCN is the Pyramid Pooling Module [12]. The existing FCN progressively interpolates the encoded information through the convolutional layer. The interpolated information goes through the convolutional layer again and becomes denser. It is a structure in which the Pyramid Pooling Layer is added in the middle [13]. DeepLabV3+'s spatial pyramid pooling (SPP) and encoder-decoder structures are used in semantic segmentation. SPP allows encoding of multi-scale contextual information, and the encoder-decoder structure allows for finer capture of object boundaries. It is DeepLabV3+ that combines the advantages of both methods and adds a simple but effective decoder to the previous version, DeepLabV3 [14]. MobileNet is a focused on model of weight reduction. MobileNet v1 reduced the total number of weights in the vgg model by replacing the convolutional layer with depth wise separable convolution and 1x1 conv [15]. As shown in Figure 3, the error rate and learning time were measured by learning using three modules. For the accuracy of the experiment, it was repeated a total of 3 times. The epoch value was set to 5 times and the iteration value was set to 10 times.



(a) Error Rate (b)Training Time

Figure 3. Error rate and training time with different FCN module

(a) Epoch (b) Iteration

Figure 4. Error rate with different epoch and iteration value

As the result of the experiment, the change in the error rate according to the model was confirmed to be within 0.05, and it was confirmed that the speed of the lightweight model, MobileNet, was the fastest in the learning time. Therefore, as a result of comparing the learning time and error rate, MobileNet was selected for learning. In order to measure the effect of changes in Epoch and Iteration values on the error rate, an experiment was conducted with the selected module, MobileNet. The epoch value is the total number of sets that have been trained. Iteration determines how many training sessions per set. Epoch values were set 5 times, 10 times, 15 times, and 20 times, and the iteration value was performed 10 times and 20 times. For accuracy, it was learned a total of 3 times. As shown in Figure 4, the error rate decreased as the epoch value increased. Also, there was no change in the error rate even when the iteration value was increased. However, as the number of times increases, the learning time is proportional, so when the learning time and the error rate are considered, the epoch value is set to 10 times and the iteration value is set to 10 times to construct the final learning model.

4. FACE MODELING USING AVATAR MAKER

Avatar Maker is a Unity editor extension that creates 3D avatars from portraits. You can use and create existing photos using your web camera in the Unity editor, and instantly create photo-like avatars. Before using the avatar maker, it was confirmed that the modeling changes depending on the location of the web camera and the presence or absence of accessories. In addition, lighting is a variable that determines the quality of the avatar. Therefore, the most efficient and accurate camera position, lighting, and presence or absence of accessories were set as parameters and analyzed. Before recognizing a person using a web camera, the position of the camera was set to 30 cm, 50 cm, and 100 cm, and the skin color of the neck and face was compared. Looking at the results in Figure 5, it can be observed that the skin color of the front and side parts of the face is most similar at 100 cm. The change of skin color according to the color of lighting was modeled using images illuminated with yellow and white lights.

(a) 30cm (b) 50cm (c) 100cm

Figure 5. Modeling differences by distance

As shown in Figure 6, in the image using white light, it can be seen that the overall skin tone is white, and in the image using yellow light, the skin tone is determined to be yellow. Therefore, realistic modeling can be obtained only when the lighting color is set according to the skin color of the person. We use HSV color model [16]. The accessories worn on the face were modeled with the most common image wearing glasses.

(a) White Light (b) White light with hair model (c) Yellow Light (d) Yellow light with hair model

Figure 6. Face Modeling differences by Light Colors

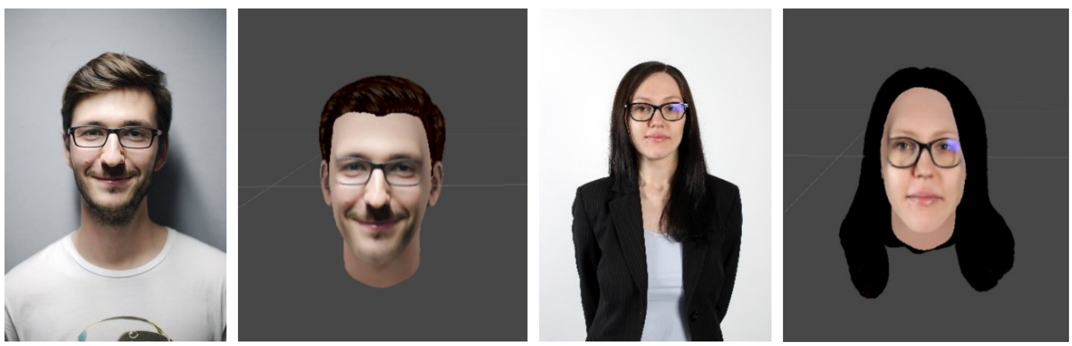
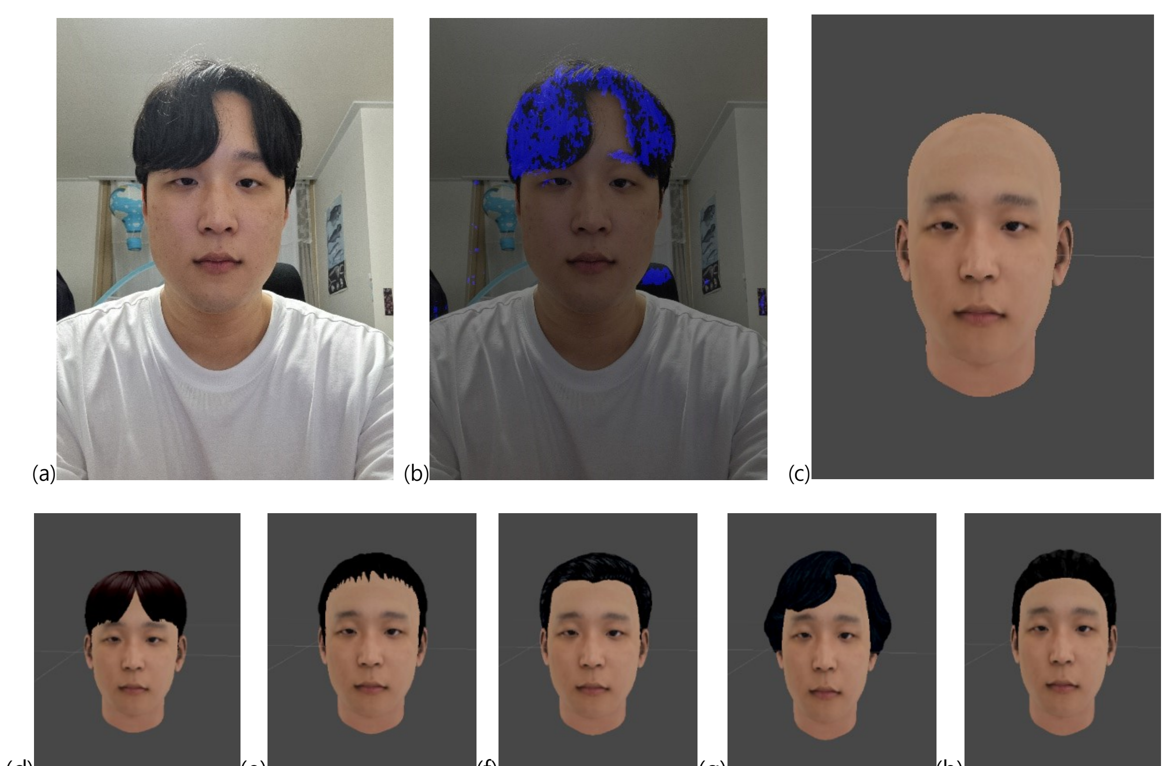


Figure 7. Face Models including glasses



(a) Input photo, (b) Hair segmentation, (c) Face Model (d),(e),(f),(g), and (h) Face models with hair models

Figure 8. Hair Segmentation and Hair-Face Modeling Pipeline

Figure 7 shows the face modeling method of the avatar maker by cutting a person's face and attaching it to the front of the head created in advance. If modeling is carried out while wearing glasses, modeling of glasses is not implemented, and if a person with a small head or oversized glasses is worn, it can be observed that the glasses are distorted. The solution to the problems revealed through the experiments is to set the distance of the camera to 80cm to 100cm, to model after wearing glasses, to remove them, and to set the color of the lighting to be the most similar to the skin color. Based on these analysis results, Figure 8 shows the results generated by semi-automatically comparing the prepared hair dataset with the hair zoning part after realizing the face modeling using the avatar maker after performing hair zoning using FCN. Among them, select the hair modeling result most similar to the original photo.

5. CONCLUSION & DISCUSSION

As the human body 3D modeling related field has developed and the market size has grown, the importance of 3D hair styling has also increased. Existing hair modeling has a problem that the face and hair cannot be modeled together, and the variety of hair is insufficient. In this study, we investigated the error rate by Epoch and Iteraion values compared with the model implementation process of hair region using FCN to implement realistic hair styling. And we used the avatar maker to create the face model which could be the most efficient and accurate environment for face modeling. By analyzing the principles of PSPNet, Deeplab V3+, and MobileNet, and comparing the error rate and learning time for each model, the best model was selected and trained. When using the avatar maker, the location of the camera, the color of the lighting, and the presence or absence of accessories were set as environmental analysis factors. In addition, as a solution to the problems derived by analyzing the elements, the method of setting the distance of the camera to 80~100cm, matching the color of the lighting with the skin color, and detaching the accessories were presented. In future research, the goal is to search for the most similar hair model in the database using aspect ratio and HoG and to improve the real-time rendering color change and FCN speed by measuring the image color. Providing users with realistic hair styling is expected to not only increase the realism of various contents, but also promote the development and commercialization of hair styling technology in the field of 3D modeling of the human body.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (No. 2021R1F1A104540111).

REFERENCES

[1] Jahanzeb Hafeez, Seunghyun Lee, Soonchul Kwon, and Alaric Hamacher, “Image Based 3D Reconstruction of Texture-less Objects for VR Contents,” *The International Journal of Advanced Smart Convergence*, 6(1), *9-17*, 2017. DOI: 10.7236/IJASC.2017.6.1.9

[2] Y. S. Kang, J. Y. Kang, H. Y. Yoon, J. J. Hwang, Y. H. Chang, and C. B. Ko, “A Study on the AR Game Analysis and Business Model,” *The Journal of the Convergence on Culture Technology*, 2(4), *49-54*, 2016. DOI: 10.17703/JCCT.2016.2.4.49

[3] M. Chai, T. Shao, H. Wu, Y. Weng, and K. Zhou, “Autohair: Fully automatic hair modeling from a single image,” *ACM Transactions on Graphics*, 35(4), *1-12*, 2016. DOI: 10.1145/2897824.2925961

[4] S. Lee, “Deep structured learning: architectures and applications,” *International Journal of Advanced Culture Technology*, 6(4), *262-265*, 2018. DOI: 10.17703//IJACT2018.6.4.262

[5] Sung-Ho Kim, “Development of the 3D Hair Style Simulator using Augmented Reality,” *Journal of*

Digital Convergence, 13(1), *249-255*, 2015. DOI: 10.14400/JDC.2015.13.1.249

[6] Y. Huang and J. Y. Song, "Recognition model of road signs using image segmentation algorithm," *The Journal of The Institute of Internet, Broadcasting and Communication*, 13(2), *233-237*, 2013. DOI: 10.7236/JIIBC.2013.13.2.233

[7] J. W. Lee and I. K. Park, "Retrieval-Based Hair Model Augmentation for 3D Face Modeling," *Journal of KIISE*, 46(5), *405-412*, 2019. DOI: 10.5626/JOK.2019.46.5.405

[8] Unity Asset Store. *https://assetstore.unity.com/packages/tools/modeling/avatar-maker-pro-3d-avatar-from-a-single-selfie-134800*

[9] H. Kim, and Y. Chung, "Effect of Input Data Video Interval and Input Data Image Similarity on Learning Accuracy in 3D-CNN," *International Journal of Internet, Broadcasting and Communication*, 13(2), *208-217*, 2021. DOI: 10.7236/IJIBC.2021.13.2.208

[10] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, *3431-3440*, 2015. DOI: 10.1109/tpami.2016.2572683

[11] Share Your Project. *http://projects.i-ctm.eu/it/progetto/figaro-1k*

[12] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid Scene Parsing Network," *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, *2881-2890*, 2017. DOI: 10.1109/CVPR.2017.660

[13] B. Yu, L. Yang and F. Chen, "Semantic Segmentation for High Spatial Resolution Remote Sensing Images Based on Convolution Neural Network and Pyramid Pooling Module," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(9), *3252-3261*, 2018. DOI: 10.1109/JSTARS.2018.2860989

[14] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff and H. Adam, "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation," *Proceedings of the European Conference on Computer Vision (ECCV)*, *801-818*, 2018. DOI: 10.1007/978-3-030-01234-2_49

[15] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, *4510-4520*, 2018. DOI: 10.1109/CVPR.2018.00474

[16] H. S. Hwang, J. W. Jung, Y. H. Kim and Y. S. Choe, "A Study on Improving Speed of Interesting Region Detection Based on Fully Convolutional Network," *Proc. of the Conference of The Korea Society of Broad Engineers*, *322-325*, 2018.