사람의 얼굴을 이용한 동물 3D 모델링 생성 서비스

이종혁[01], 소병욱[02], 조현태[1], 이준수[2], 백지웅[1], 최명걸[2*]
[1]가톨릭대학교 컴퓨터정보공학부
[2]가톨릭대학교 미디어기술콘텐츠학과
{ljh20011, thquddnr123, hyuntae9912, land8746, jw6133, }@catholic.ac.kr

Service for Generating Animal 3D Objects Using Human Faces

Jong-Hyuk Lee[01], SO BYUNGUK[02], JO HYUNTAE[1], LEE JUNSOO[2], BACK JIWOONG[1], 최명걸[2*]
[1]School of Computer Science and Information Engineering, The Catholic University of Korea
[2]Dept. of Media Technology and Media Contents, The Catholic University of Korea

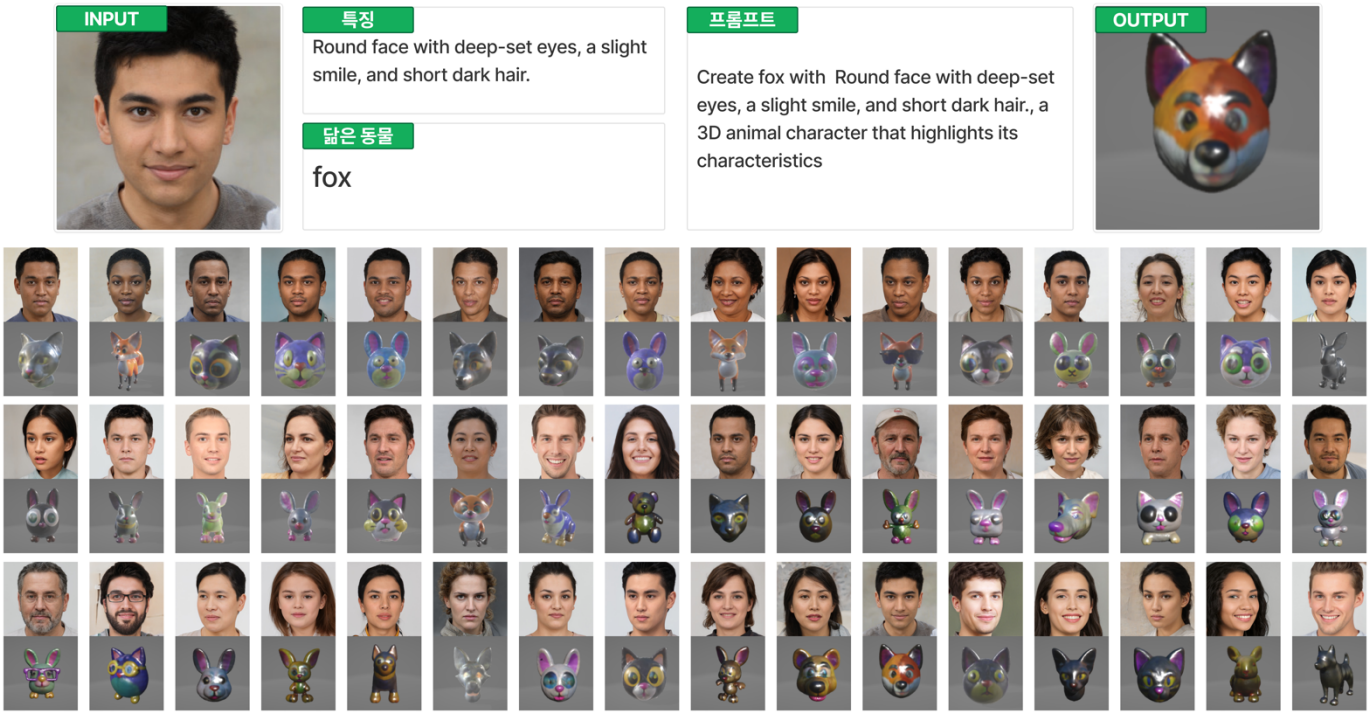

그림 1 : 사람의 얼굴 이미지를 사용하여 생성된 3D 모델링

요약

최근 일어난 생성형 AI의 발전은 컴퓨터 그래픽스 분야에서, 텍스트만으로도 3D 모델링을 생성할 수 있게 되었다. 본 연구에서는 사람의 얼굴 이미지로부터 추출한 특징들을 프롬프트로 사용하여 3D 동물 모델링을 생성하는 서비스를 제안한다. 유사한 동물을 추론하기만 하는 기존의 서비스와 달리, 인물의 특징과 닮은 동물의 정보를 포함한 프롬프트를 생성하여 보다 개성 있는 3D 동물 모델링을 생성한다. 이를 위해 ResNet34 모델을 사용하여 유사한 동물을 추론하고, ChatGPT API를 통해 인물의 특징을 추출하였다. 이렇게 추출된 정보들은 하나의 프롬프트로 연결되어, meshy AI의 API를 통해 3D 모델링으로 변환되었다. 실험 결과, 전문가 평가를 통해 생성된 3D 모델의 질적 우수성을 확인하였다. 본 연구는 얼굴 이미지 한 장만으로도 특징이 잘 드러나는 3D 동물 모델링을 제공하며, 향후 다양한 동물에 대한 실험을 통해 사용자 경험을 극대화할 계획이다.

* 구두(포스터) 발표논문
* 학부생 주저자 논문임.
* 본 연구는 xxx 지원으로 수행되었음

1. 서론

텍스트로부터 이미지를 생성하는 모델인 Stable Diffusion[1]과 같은 Diffusion 모델의 성공으로 인해 텍스트-이미지로 구성된 엄청난 양의 학습 데이터가 확보되면서, 텍스트를 사용하여 고품질의 2D 이미지 생성이 가능하게 되었다. 이는 2D 이미지를 기반으로 3D 모델링을 생성하는 과정에서 요구되는 비용을 감소시켰고, DreamFusion[2]과 같이 텍스트만으로도 고품질의 3D 모델링을 생성하는 모델의 발전에 기여하였다. 동시에

LLaVA[3], CLIP[4]와 같이 이미지와 텍스트를 한번에 처리할 수 있는 VLM (Vision-Language Model)의 등장은 하나의 모델을 사용하여 이미지에 대한 정보를 처리할 수 있게 되었다.
우리는 이러한 기술들을 활용하여 사람의 얼굴 이미지로부터 추출한 특징을 사용하여 3D 동물 모델링을 생성하는 서비스를 제안한다. 기존에 분류 모델을 사용하여 유사한 동물을 추론하는 서비스를 제공한 경우[1]는 있었지만, 추론 결과로 나온 동물에 대한 정보만을 사용자에게 전달하였다. 본 연구에서는 이에서 더 나아가, 단순히 동물을 추론하고 결과를 반환하는 것이 아니라 분류 모델이 반환한 인물과 닮은 동물에 VLM이 반환하는 이미지 속 인물의 특징을 프롬프트에 첨가하여 3D 모델링을 생성하는 방법을 통해 보다 현실감 있고 개성 있는 동물 캐릭터를 만들어내는 서비스를 제안한다.

2. 방법론

사람의 사진을 통해 유사한 동물을 추론하기 위해 ResNet34[5]를 사용하였다. 동물의 이미지로 학습된 ResNet 모델은 사용자가 입력한 사람 이미지에 대해 가장 확률적으로 유사한 동물을 텍스트의 형태로 반환한다. 동시에, ChatGPT API[2]를 이용하여 사람의 이미지에 대한 특징을 추출한다. 이렇게 만들어진 동물 키워드와 인물의 특징을 사용하여 프롬프트를 제작한 뒤, 제작된 프롬프트를 사용하여 3D 모델링을 생성하였다. 3D 모델링 생성에는 Meshy AI[3]의 API를 활용하였으며, 이는 고품질의 3D 모델을 효과적으로 생성하는 데 사용되었다.
모든 과정을 원활하게 연결하기 위해 Spring Boot와 Flask를 사용하여 시스템을 구축하였다. Spring Boot는 백 엔드 서버를 구축하고 프론트 엔드와의 연동에 사용되었으며, Flask는 모델들의 API를 담당하였다. 이 시스템을 통해 사용자는 웹사이트에 이미지를 업로드하면 3D 모델링이 생성되고, 결과를 확인할 수 있다.

3. 실험 결과

실험에 사용한 ResNet34는 Colab환경에서 16GB NVIDIA Tesla T4를 통하여 학습하였고, 데이터셋은 저작권이 없는 동물 사진 600장을 사용하였다. 동물은 개, 고양이, 사슴, 곰, 토끼, 여우의 6종류로 구분하였다. ChatGPT API의 경우 이미지와 텍스트를 모두 처리할 수 있어야 하므로 GPT-4o를 사용하였다.
그림 1을 통해 사람의 이미지에 대해 유사한 동물과 특징을 포함하고 있는 프롬프트를 통해 3D 모델링을 생성할 수 있음을 확인할 수 있다. 우리의 결과에 대해 사용자 평가를 수행하지 않았지만, 전문가를 대상으로 질적 평가를 수행하였다. 질적 평가를 위해 다양한 실험 결과를 생성할 필요가 있었고, 이때 사용된 사람의 얼굴 이미지는 생성형 AI를 이용하여 만들어진 사람의 얼굴 이미지[4]를 사용하였다. 최종적인 전문가 평가 결과 질적으로 우수하다는 평가를 받았다.

4. 결론

본 연구에서는 사람의 얼굴 이미지로부터 특징과 동물을 추출하여 3D 모델링을 만드는 서비스를 제안한다. 우리의 서비스는 얼굴 이미지 한장만으로 특징이 잘 드러나는 3D 동물 모델링을 제공한다. 또한, 추가적인 개발을 통해 사진 한 장으로 3D 모델링을 생성하는 과정을 자동화할 수 있는 가능성도 존재한다. 그러나 전문가의 평가 결과 유용성에 대해 긍정적인 평가를 받았음에도 불구하고, 동물 외형의 스타일이 정형화 되어 있고, 표정이 풍부하지 못하다는 피드백을 받았다. 또한, 소수의 전문가에 의한 질적인 평가만 수행하였다는 한계점 역시 존재한다.
향후 연구에서는 이러한 한계점을 해결하기 위해, 동물의 표정 생성에 특화된 인공신경망을 개발하여 현재 3D 모델링을 생성하는데 사용하고 있는 Meshy AI의 API를 대체할 계획이다. 그리고 다수를 대상으로 체계적인 사용자 평가를 진행하여 결과를 수치화 하고, 성능을 개선해 나갈 계획이다. 마지막으로는 다양한 동물에 대해 실험을 진행하여 사용자 경험을 극대화할 수 있는 방법을 지속적으로 탐구할 것이다.

참조

[1] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, **"High-resolution image synthesis with latent diffusion models,"** in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 10 684–10 695
[2] B. Poole, A. Jain, J. T. Barron, and B. Mildenhall, **"Dreamfusion: Text-to-3d using 2d diffusion,"** *arXiv preprint arXiv:2209.14988,* 2022
[3] H. Liu, C. Li, Q. Wu, and Y. J. Lee, **"Visual instruction tuning,"** Advances in neural information processing systems, vol. 36, 2024
[4] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, et al., **"Learning transferable visual models from natural language supervision,"** in International conference on machine learn- ing. PMLR, 2021, pp. 8748–8763.
[5] K. He, X. Zhang, S. Ren, and J. Sun, **"Deep residual learning for image recognition,"** in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770– 778.

[1] https://www.youtube.com/watch?v=_DmM8EO3mb8
[2] https://chatgpt.com/
[3] https://www.meshy.ai/
[4] https://generated.photos/