

패치 기반 생성형 모델을 적용한 메타버스 이미지 리타겟팅 연구

김건아, 문채원, 김은지, 김찬호[°], 정설영*

경북대학교 컴퓨터학부

kga0416@knu.ac.kr m6011703@gmail.com now0104@knu.ac.kr coolhol123@knu.ac.kr

snowflower@knu.ac.kr

Metaverse image retargeting study with a patch-based generative model

Geonah Kim, Chaewon Moon, Eunji Kim, Chanhoo Kim[°], Seolyoung Jeong*

School of Computer Science and Engineering, Kyungpook National University

요 약

메타버스 산업의 성장과 함께 메타버스 기술의 구현에 대한 중요성이 부각되고 있으나, 화면을 최적화하는 리타겟팅 기술에 대한 연구가 부족하다. 이 연구는 메타버스에서의 화면 리타겟팅 필요성을 언급하며, 패치 기반 이미지 생성 모델인 GPNN을 활용하여 메타버스 데이터의 리타겟팅을 수행하는 프레임워크를 실험적으로 적용했다. 이후 시각적 자연스러움을 평가하는 비교 분석을 실시하였으며, 그 결과 일관성 측면에서는 비교적 우수하나 완전성 측면에서는 개선이 필요함을 확인할 수 있었다. 실험결과를 통해 패치 기반 모델의 한계를 파악할 수 있었고, 향후 메타버스 이미지 기반의 리타겟팅을 수행하기 위해서는 객체에 대한 이해를 가진 모델 연구의 필요성을 파악하였다.

1. 서 론

메타버스 산업의 발전으로 메타버스의 활용 및 적용에 대한 연구가 활발히 이루어지고 있으나, 메타버스를 실현하는 기술 연구는 부족한 상태이다.[1] 특히, 메타버스는 3차원(3D) 공간을 기반으로 하여, 새로운 3D 데이터를 생성하는 데 비용과 시간이 많이 소요되는 한편, 기존의 3D 데이터를 활용하여 원하는 비율의 데이터를 생성할 수 있는 리타겟팅 기술에 대한 연구가 필요하다.

화면 리타겟팅은 메타버스 환경을 위한 중요한 기술 중 하나로, 다양한 화면 크기 및 비율에 맞게 콘텐츠를 최적화하는 것을 말한다. 컴퓨터 및 가상현실(VR) 기기와 같이 다양한 플랫폼에서 메타버스를 활용하는 사용자들에게 품질 높은 경험을 제공하기 위해 화면 리타겟팅 기술은 필수적이다.

본 연구는 메타버스 환경에서의 리타겟팅을 위해 패치 기반 생성형 모델(GPNN)을 실험적으로 활용한다. GPNN은 다른 이미지 생성 모델에 비해 시간이 적게 걸리고 자연 이미지를 생성하는 데 우수한 성능을 보이는 알고리즘이다. 따라서 GPNN을 활용하여 메타버스 데이터의 리타겟팅 수행 능력을 탐구하고자 한다.

이번 연구에서는 MPEG MIV 데이터를 이미지로 변환하여 활용하며, 입력 데이터에 대한 결과로 높이는 유지하고 너비를 2배로 조정하는 방식으로 진행한다. 이를 통해 메타버스 데이터의 리타겟팅을 실험적으로 보여준다.

2. 관련 연구

InGAN[2]은 한 장의 입력 이미지만으로 이미지의 내부 패치 분포를 학습하여 출력 이미지로 매칭하는 패치 기반 이미지 리타겟팅 모델이다. 모델 아키텍처는 여러 데이터 타입을 가진 이미지를 다양한 사이즈와 모양, 중첩비율 가진 이미지로 리타겟팅하여 일관성 있고 완전한 결과를 도출한다. 그러나 입력 이미지 외의 추가적인 정보가 없기에 문맥적으로 의미가 맞지 않은 부자연스러운 결과를 도출한다는 한계가 존재한다.

패치 기반 모델인 GPNN[3]은 GAN과 유사한 생성 기능을 최근접 이웃 패치 기반 방법에 주입한 모델로, 빠른 생성 속도와 시각적으로 완성도 높은 이미지를 출력한다는 특징을 가진다. 모델의 구조는 멀티 스케일링 구조를 사용함으로써 이미지의 다양한 특징을 포착하고, 각 스케일마다 최근접 이웃 패치 방법을 통해 입력 이미지와 출력 이미지 간 내부 확률의 유사성을 강화한다. GPNN은 이러한 특성을 유지함과 동시에 파라미터 조절을 통해 리타겟팅, 이미지 합성, 이미지 편집 등의 응용에 적용할 수 있다. 본 연구에서는 GPNN의 리타겟팅 응용에 초점을 두고, 메타버스 콘텐츠에 적용한 결과를 살펴본다.

3. 실험 진행 및 결과

3.1. 실험 진행

GPNN은 패치 기반 생성형 모델들 중 높은 시각적 완성도와 빠른 생성 속도를 보유하고 있다. 이러한

특성을 고려하여, 본 연구에서는 GPNN 에 메타버스 콘텐츠를 기반으로 한 MIV 데이터 셋의 일부 프레임을 활용하여 리타겟팅 실험을 수행하였다(표 1). 입력 프레임의 해상도는 800 x N 으로 설정하였다. 입력 이미지의 높이를 유지하면서 너비를 2 배로 늘리는 리타겟팅을 수행하였다. 그림 1 은 본 연구에서 활용된 패치 기반 생성형 모델을 이용한 메타버스 콘텐츠 리타겟팅 프레임워크를 나타낸다.

실험 모델의 하이퍼 파라미터 값을 설정하기 위해 패치의 크기와 피라미드의 높이, 다운 스케일링 비율을 조절하여 총 4 가지의 경우로 실험을 진행하였다. Street 데이터를 활용한 실험 결과(그림 2), 모델에 입력하는 하이퍼 파라미터의 값에 따라 결과가 변하는 것을 확인하였다.

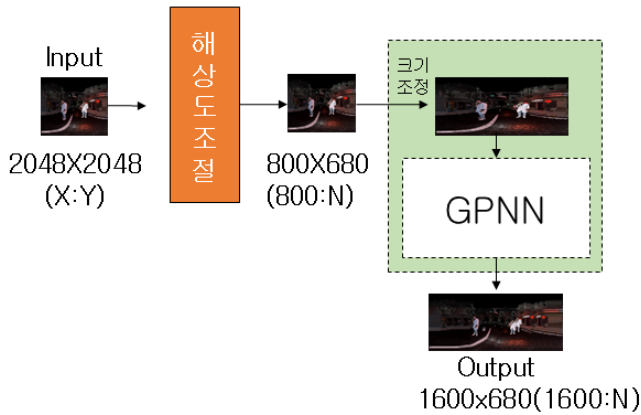


그림 1 메타버스 콘텐츠 리타겟팅 프레임워크

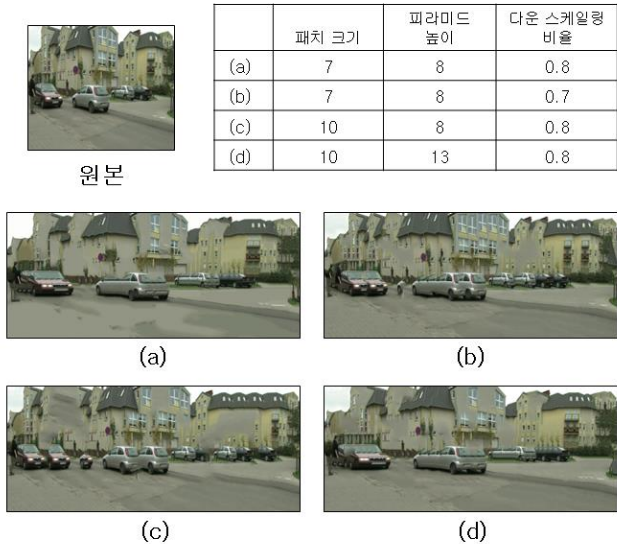


그림 2 하나의 데이터에 다양한 하이퍼 파라미터로 실험한 결과

그림 2 의 (b)와 (c)는 자동차 객체가 연속되지 않고 중복적으로 추가된 것을 볼 수 있으나, (a)와 (d)의 경우의 자동차 객체는 비교적 연속성을 가지고 생성되었다. 추가적으로 (a)의 배경에서 창문의 대칭성이 (d)에 비해 높은 것으로 나타났다. 이에, (a)는 일관성을 유지하며 생성된 것으로 확인할 수 있다. 따라서, 본 연구에서는 모델의 파라미터로, 패치 크기

7, 피라미드 높이 8, 다운 스케일링 비율 0.8 로 설정하여 진행하였다.

표 1 데이터 셋 명세

시퀀스 명	원본 해상도	GPNN 입력 해상도	리타겟팅 해상도
Cyberpunk	2048x2048	800x680	1600x680
Carpark	1920x1088	800x450	1600x450
Street	1920x1088	800x450	1600x450
Dancing	1920x1080	800x450	1600x450

3.2. 실험 결과

원본 이미지와 리타겟팅 이후 이미지를 비교하여 시각적 자연스러움을 평가한 결과, 여러 문제점을 확인할 수 있었다. 그림 3 과 그림 4 의 리타겟팅 결과를 보면 각각의 객체들의 크기와 비율이 비교적 잘 유지되었다. 이를 통해 일관성을 갖추었다고 판단할 수 있다. 그러나 완전성의 측면에서 살펴보면 이미지가 확장되면서 하나의 객체가 여러 개로 분할되어 객체의 형태가 완전하게 보존되지 않았다.

또한, 그림 4(b)의 왼쪽 하단을 보면, 넓은 부분에 직선으로 이어져 있는 화단 부분이 리타겟팅 이후 네 부분으로 끊어진 것을 볼 수 있다. 이는 GPNN 이 이미지를 생성할 때 전체 흐름을 반영하지 않고, 패치 사이즈를 바탕으로 초근접한 이웃 패치만을 학습하여 새로운 이미지를 생성함으로써 발생하는 문제로 추론된다.

결과적으로, 현재 모델은 메타버스 데이터에 대하여 일관성 측면에서는 어느 정도 성능을 보이지만, 완전성과 전체적인 흐름을 반영하는 능력에서 개선이 필요한 것으로 나타났다. 이는 패치 기반 생성형 모델의 구조적인 문제로 비롯되며, 모델을 유지하여 성능을 개선하는 것에는 어려움이 있음을 파악하였다.

따라서, 주변 문맥을 고려하여 더 넓은 영역의 정보를 학습할 수 있는 새로운 모델 아키텍처를 고안할 필요가 있다. 뿐만 아니라, 메타버스 이미지의 특성을 반영한 객체 감지 및 분할과 같은 컴퓨터 비전 기술을 함께 적용한다면 현재 모델의 한계점을 극복할 수 있을 것으로 기대된다.



(a) GPNN 입력 이미지



(b) 리타겟팅 결과 이미지

그림 3 Dancing 리타겟팅



(a) GPNN 입력 이미지



(b) 리타겟팅 결과 이미지

그림 4 Carpark 리타겟팅

4. 결론

본 논문에서는 메타버스 데이터를 리타겟팅 하기 위해 패치 기반 이미지 생성형 모델인 GPNN 을 실험적으로 적용하였다. 그 결과 메타버스 객체들이 끊어지거나 중복적으로 생성되어 완전성이 보장되지 않는 문제가 있었다. 이는 패치 기반 모델이 근접 부분만을 학습하여 이미지를 생성하기 때문에 발생한 특성으로 파악된다. 특히, 사람과 같은 주요 객체가 중복적으로 생성된 경우 시각적으로 더 어색하게 인식될 수 있다.

결론적으로 대부분의 메타버스 콘텐츠는 복잡한 배경과 명확한 주요 객체를 가지고 있어서 기존의 패치 기반 생성 모델을 사용하는 것이 적합하지 않음을 알 수 있다. 따라서 주변 문맥을 고려하는 더 넓은 영역의 정보를 학습할 수 있는 새로운 모델 아키텍처의 개발을 고려하여야 한다. 또한, 메타버스 이미지에서 주요 객체의 중복 생성 및 이미지 완전성 부족 문제를 개선하기 위해서는 객체 감지 및 분할에 특화된 딥러닝 모델이나 컴퓨터 비전 기술을 통합하는 것이 유용할 것으로 판단된다. 이러한 발전된 모델들은 메타버스 데이터의 특성에 더 적합한 이미지 생성 및 리타겟팅을 가능케 할 것으로 기대되므로 이에 따른 후속 연구가 필요하다.

5. 참고 문헌

- [1] SHIM, Mi-Na. Analysis of Research Trends on the Domestic Metaverse using Topic Modeling: Focusing on Research from 2007 to 2022, 디지털콘텐츠학회논문지 (J. DCS), 2457-2468. 2022. 23.12.
- [2] SHOCHER, Assaf, et al. Ingan: Capturing and retargeting the "dna" of a natural image. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. p. 4492-4501. 2019.
- [3] GRANOT, Niv, et al. Drop the gan: In defense of patches nearest neighbors as single image generative models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. p. 13460-13469. 2022.

[4] SIMAKOV, Denis, et al. Summarizing visual data using bidirectional similarity. In: 2008 IEEE conference on computer vision and pattern recognition. IEEE, p. 1-8. 2008.

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학사업의 연구결과로 수행되었음" (2021-0-01082)