

# 메타버스를 위한 실감형 미디어 기반 데이터셋의 리타겟팅 적용 연구

김동휘, 김아로, 이동훈, 김찬호, 김은지, 문채원, 김건아, 박상효\*  
경북대학교 IT 대학 컴퓨터학부

## A Study on Retargeting Via Realistic Media-based Datasets for Metaverse

Dong-hwi Kim, Aro Kim, Dong-hun Lee, Chanhoo Kim, Eunji Kim, Chaewon Moon, Geonah Kim,  
Sang-hyo Park\*

A School of Computer Science and Engineering

Kyungpook National University

E-mail : dhwi@knu.ac.kr, \*s.park@knu.ac.kr

### Abstract

In this paper, we propose an experimental application of retargeting method to immersive media-based datasets for metaverse content. The past retargeting techniques such as resizing, cropping, and seam carving may distort the main objects. Therefore, our study employs a specialized retargeting technique that detects and preserves the visual quality of main objects. As a result, we compared the original frame and retargeted frame with a perceptual quality-based comparative analysis. Finally, we found that the relative location of coordination offset could affect the main object-oriented retargeting framework. We present the results of applying metaverse media content to the retargeting algorithm by identifying complementary points through the experimental results.

### I. 서론

메타버스 콘텐츠 및 실감형 미디어 기반 서비스 공급이 증가함으로써, 사용자들은 다양한 화면 비율을 가진 플랫폼 및 디바이스에서 실감형 콘텐츠를 접할 수 있는 기회가 늘고 있다. 이에 따라 영상 미디어의 화면 비율 조정이 점차 중요한 과제로써 부각된다. 동시에, 다방향의 실감형 미디어 기반 서비스를 효과적으로 제

공하기 위한 연구가 요구된다. 그 중 서비스 제공자의 입장에서는 해상도와 같은 기술적 요소를 고려해야 하기에, 본 연구에서는 리타겟팅을 실험적으로 메타버스 콘텐츠에 적용해 봄으로써 사용자의 재생기기에 적합한 종횡비의 해상도로 변환 및 최적화를 이루어 낼 수 있도록 연구한다. 리타겟팅 기반의 실험적 연구 결과를 통해 다시점 뷰 포인트 데이터셋에 대해 리타겟팅 적용 연구 결과를 살펴본다.

본 연구를 진행하기 위해서는 기존의 가로, 세로로만 표현 가능한 2D 기반의 데이터셋과는 다른 메타버스 콘텐츠를 이용하는 사용자를 위해 특수한 데이터셋을 사용하여야 한다. 예를 들면, MPEG MIV [7] 와 같은 6 degrees of freedom (DoF) 기반의 다각도, 다방면의 뷰 포지션을 가지는 캡처 시스템 기반의 데이터셋이나, RGB+D 포인트 클라우드 등과 같은 특수한 데이터셋 등이 있다. 본 연구에서는 다시점 뷰 포인트를 가진 데이터셋을 사용하여 리타겟팅을 진행하였다. 목표 종횡비를 고려함과 동시에, 중요 객체에 대한 정보를 활용하여야 인지적 결과를 향상시킬 수 있다.

본 연구에서 지향하려는 메타버스 콘텐츠 강화의 경우

## II. 관련 연구

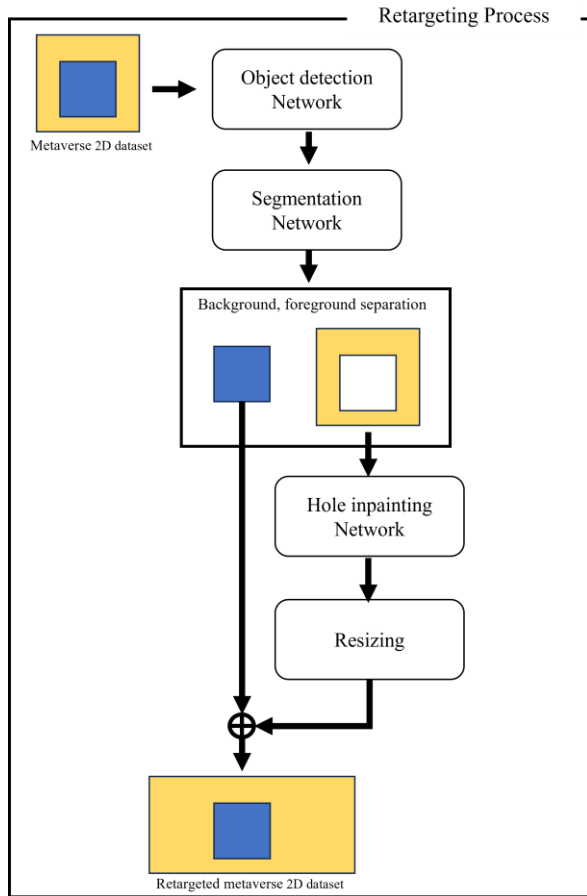


그림 1 본 연구에 사용된 메타버스 콘텐츠 데이터셋 기반 리타겟팅 프레임워크

시퀀스 명	원본 해상도	리타겟팅 후 해상도
Hi jack	4096×2048	2560×1080
Cyberpunk	2048×2048	2560×1080
Carpark	1920×1088	1451×1088
Street	1920×1088	3414×1088
Group	1920×1080	3414×1088
Dancing	1920×1080	2560×1080

표 1 본 연구에 사용된 MPEG MIV 데이터셋 시퀀스 정보

서비스 내 가상 캐릭터, 배경 등과 같은 인지적 화질 요소들을 최대한 보존하기 위해, 중심 객체의 왜곡을 최소한으로 줄이는 비디오 데이터셋 기반의 리타겟팅 선행 연구 [2, 3]를 통해 메타버스 콘텐츠 기반의 데이터셋 리타겟팅 선행 연구를 조사하였으며, 본 연구에서는 제안된 프레임 워크 [3]를 사용하였다.

리타겟팅 방법의 경우, 고전적인 처리기법인 리사이징, 잘라내기 등의 방법, 에너지 그래디언트를 활용한 특징점 기반의 영상 조절 방법, 그리고 딥러닝 기반의 방법들을 예로 들 수 있다. 전통적인 잘라내기, 리사이징 등의 리타겟팅 기법들을 살펴보면 잘라내기의 경우, 영상 및 콘텐츠 내에 존재하는 일부 배경, 자막 등의 영상 정보의 유실을 유발할 수 있다는 것을 확인할 수 있다. 또한 리사이징의 경우, 영상 내의 중요 객체의 원본 비율이 변하게 되기 때문에, 시각적으로 왜곡된 결과를 보여지게 되므로, 리타겟팅 이후 몰입도가 떨어질 수 있다. 따라서, 객체 정보의 유실율을 줄이고, 몰입도를 향상시키기 위한 방법으로 패치 (patch), 심 (seam) 등의 도구 기반의 연구가 고안되었다. Seam carving [7]은 이미지 내의 그래디언트 기반의 에너지 함수를 통해 심을 정의하고, 위에서 아래 혹은 왼쪽에서 오른쪽으로 방향성을 정하여 연결된 픽셀 경로를 의미한다. 심을 제거하거나 삽입함으로써 이미지 전체의 종횡비를 변경할 수 있으나, 원본 영상 내의 콘텐츠를 인식하여 변환을 진행하지 않기 때문에, 객체를 온전히 보존하기 힘들며, 일부 왜곡이 발생한다는 점이 있다.

딥러닝 기반의 리타겟팅의 경우 패치 기반이나, 객체 보존성을 강화한 리타겟팅의 연구 사례가 있다. Internal GAN (InGAN) [4] 등의 리타겟팅 기법의 연구가 패치 기반의 리타겟팅 기법의 예이며, 순환 신경망 기반의 deep video retargeting [5] 연구 또한, 딥러닝 기반의 리타겟팅 연구의 예이다. InGAN [4]의 경우, 입력된 이미지에 대해 내부 패치 분포를 학습하여 다양한 크기의 리타겟팅 결과를 나타낸다. 목표의 해상도 달성을 위해 멀티 스케일 분별망을 사용하여 패치 분포를 사용하여 전통적인 리타겟팅 방법에 비해서는 나은 성능의 결과를 보여주지만, 입력 이미지 외의 추가적인 정보가 없기에, 중심 객체의 왜곡은 피할 수 없는 한계가 존재한다. Deep video retargeting [5] 모델은 여러 후보군을 생성하여 보다 우수한 품질의 결과를 생성할 수 있지만, 일부 배경의 손실을 유발할 가능성이 있다는 한계가 존재한다. 따라서, 본 연구는 정보의 손실률과 왜곡을 최대

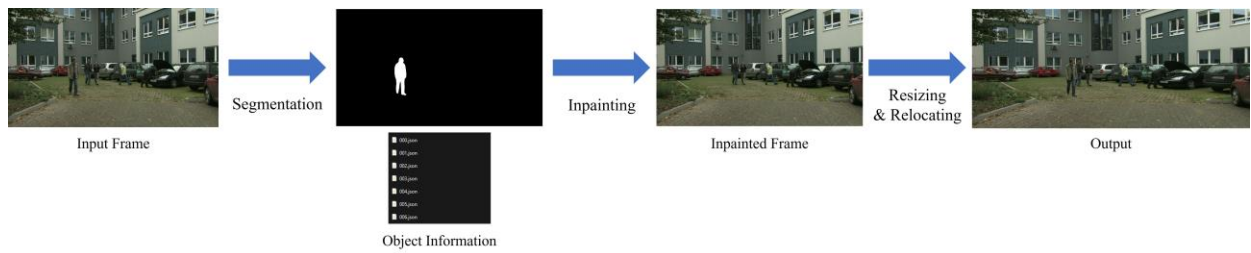


그림 2 MPEG MIV 데이터셋의 리타겟팅 결과



그림 3 MPEG MIV - Group, Carpark, Dancing 데이터셋의 리타겟팅 결과

한 줄이기 위해 여러 단계의 전처리 과정을 통해 연구를 진행하였다.

### III. 메타버스 콘텐츠의 리타겟팅 적용 연구

본 연구에서는 중심 객체 기반의 리타겟팅 모델인 J. G. Jin et al [3] 모델을 기반으로 사람 객체를 중심 객체로 실험적으로 선정하여 리타겟팅 실험을 진행하였다. 기존 다 시점 기반의 데이터셋의 경우, 모델 연산량 증감 및 기존 모델과의 효율성 증대를 위해 뷰 포인트를 고정된 상태로 변환하여 실험을 진행하였다. 그림 1은 본 연구에 사용된 메타버스 콘텐츠 데이터셋 기반의 리타겟팅 프레임워크를 나타낸다. 그림 2는 제안된 프레

임워크에 대해 MPEG MIV 데이터셋에 대해 리타겟팅을 적용해 본 결과를 나타낸다. 전체 과정을 자세히 설명하면 첫째, 여러 깊이(depth)와 텍스처(texture)를 가지고 있는 MIV 데이터셋을 2D 기반의 영상 데이터로 변환한다. 그 후 객체 탐지 및 분할 모델로 중요 객체를 탐지하여 해당 객체들을 분리한다. 그 후 분리된 부분을 채워 넣는 인페인팅 기법을 적용한다. 그 후 배경이 크기를 조정하는 리사이징 과정을 거치고 이 전 단계에서 분리한 객체들을 리사이징된 프레임에 재배치하여 해상도 변환 리타겟팅을 수행한다. 객체 탐지 및 분할의 경우 Mask2Former [8] 모델을 사용하였고, 인페인팅 모델은 E2FGVI [9]를 사용하여 리타겟팅 프레임을 생성하였다. 실험적으로 16:9 영상을 21:9로 변환하여 영상의 프레임을 비교하여 성능을 비교하였다. 일부 Hijack 데

이더넷 등과 같은 8K 이상의 초고해상도 영상의 경우, 다운샘플링 된 데이터를 통해 나머지 영상의 리타겟팅 해상도와 비슷하게 맞추었다. 그림 3 은 MPEG MIV 데이터셋 내에 있는 Group, Carpark, Dancing 데이터셋에 적용하여 일부 뷰 포인트 프레임의 결과를 나타낸다.

## IV. 연구 결과

원본의 프레임과 리타겟팅 후 각각의 프레임을 시각적 평가에 기반하여 개선점 파악을 진행하였다. 그림 2, 그림 3 의 리타겟팅 결과를 통하여 개선점이 필요한 결과를 파악하였다. 연구 결과, 객체의 크기가 상대적으로 작거나 객체들이 많고 서로 겹칠 때, 객체 인식 성능이 저조하였다. 객체 중첩 프레임의 경우, 정확한 객체의 분할 영역 파악이 어렵기 때문에, 객체 분할 과정에서 다소 개선이 이루어져야 할 것으로 파악된다.

결과적으로, 본 연구에 사용된 리타겟팅 프레임워크는 전반적인 객체 감지가 잘 된 경우와, 제한적인 움직임 상황에서 효과적임을 알 수 있었다. 이는 적용한 프레임워크가 메타버스 콘텐츠 기반의 데이터셋에 있어서 적용시에 일부 상황에서 효과적일 수 있음을 확인할 수 있었지만, 일부는 개선이 필요한 부분이 있음을 확인할 수 있었다.

## V. 결론

본 논문에서는 메타버스 및 실감형 미디어 기반 플랫폼 기기에서의 적응적 해상도 조절을 위해 MPEG MIV 데이터셋에 객체 중심의 리타겟팅 기법을 적용한 연구결과를 제시한다. 더욱이, 메타버스와 실감형 미디어 분야에서의 콘텐츠 제공과 시각 경험 최적화를 위해 리타겟팅 및 관련 기술의 발전이 중요하며, 미래에는 양질의 이용자 중심의 서비스를 제공하기 위해 제안된 프레임워크의 개선이 필요함을 알 수 있었다. 객체 분할 모델에서의 중요 객체의 분할 오프셋을 더해줌으로써 개선의 여지가 있음을 확인할 수 있었다.

## 참고문헌

- [1] K. Lai et al, "Unsupervised Feature Learning for 3D Scene Labeling" Proceeding of the IEEE International Conference on Robotics and Automation (ICRA), 2014.
- [2] J. H. Bae et al. "A Case Study on ROI-based Video Retargeting", Proceeding of the 3rd Korea Artificial Intelligence Conference. pp. 76-77, 2022.
- [3] J. G. Jin et al. "Object-Ratio-Preserving Video Retargeting Framework based on Segmentation and Inpainting." Proceeding of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 497-503, 2023.
- [4] S. Assaf, et al. "Ingan: Capturing and retargeting the "dna" of a natural image." Proceeding of the IEEE/CVF International Conference on Computer Vision, pp. 4492-4501, 2019.
- [5] D. Cho et al. "Video Retargeting: Trade-off between Content Preservation and Spatio-temporal Consistency." Proceedings of the 27th ACM International Conference on Multimedia. 2019.
- [6] Avidan, S., & Shamir, A. "Seam carving for content-aware image resizing." In Seminal Graphics Papers: Pushing the Boundaries, Volume 2 (pp. 609-617). 2023
- [7] ISO/IEC JTC 1/SC 29/WG 04 N 0307, "Common test conditions for MPEG immersive video". February 2023.
- [8] Cheng, Bowen, et al. "Masked-attention mask transformer for universal image segmentation." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022.
- [9] Li, Zhen, et al. "Towards an end-to-end framework for flow-guided video inpainting." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022.