

Bi-Segmentation 을 활용한 비디오 아웃페인팅 아티팩트 개선 연구

오명훈, 이성배, 김규현*
경희대학교

mhun9801@khu.ac.kr, rhee@khu.ac.kr, kyuheonkim@khu.ac.kr*

Study on the Improvement of Video Outpainting Artifacts using Bi-Segmentation

Myeonghun OH, Seongbae Rhee, Kyuheon Kim*
Kyung Hee University

요 약

최근 다양한 화면비의 디스플레이가 등장하며, 기존 (16:9)의 화면비로 제작되었던 비디오를 다양한 화면비의 디스플레이에 맞추는 리타겟팅 기술이 활발히 연구되고 있다. 이러한 비디오 리타겟팅 기술 중 하나로 아웃페인팅 기술이 있으며, 아웃페인팅 기술은 원본 영상의 양측에 빈 영역을 외삽하고 이후 전경 부분이 배경으로 채워진 배경 프레임에서 픽셀을 참조하는 방식으로 외삽 영역을 채우게 된다. 배경 프레임은 원본 영상에서 전경을 대상으로 세그멘테이션을 적용해 마스킹 한 후, 마스킹된 부분을 인페인팅하게 되는데, 이 과정에서 마스킹의 정확도에 따라 배경 프레임에 전경의 픽셀이 남게 된다. 이러한 전경에서 비롯된 픽셀은 이후 아웃페인팅된 외삽 영역에서 아티팩트라고 불리는 비정상적인 잡음이나 형태를 유발한다. 따라서 본 연구는 외삽된 영역에 정상적인 픽셀이 참조될 수 있도록 Bi-Segmentation 을 활용해 전경이 배경으로 채워지는 과정을 도우며, 결론적으로 이를 통해 아티팩트를 개선하는 연구를 진행한다.

1. 서론

최근 폴더블, 롤러블등의 다양한 화면비를 가지는 디스플레이가 등장하며 기존 (16:9) 화면비로 제작되었던 비디오들의 재생에 있어서 디스플레이 전체 영역을 사용하는 것이 아닌, 일부분만을 사용하는 문제점들이 발생되었다. 따라서 이를 해결하기 위해 비디오에 대해 디스플레이의 화면비와 해상도에 맞추어 적절히 변환할 수 있는 영상 리타겟팅 기술의 필요성이 대두되었다. 그러나 리타겟팅의 초기 기술은 보간법을 통한 영상 리사이징에 의존하고 있으며, 이로 인해 이미지 내부 객체들의 비율이 변경되어 이미지 정보가 훼손돼 이용자들의 몰입도를 저하시키는 문제를 유발하였다.

따라서 단순한 리사이징이 아닌 비디오 프레임의 시간적 유사성을 이용하는 리타겟팅 기법인 아웃페인팅[1] 기법이 등장하였으며, 해당 기법은 우선 원본 이미지에서 전경을 세그멘테이션을[2] 적용해 마스킹 하여 빈 공간으로 만든 후, 광학흐름[3]을 이용해 다른 프레임의 픽셀로부터 채운 배경 이미지를 완성하고, 완성된 배경 이미지의 픽셀을 광학흐름을 이용하여 원본 이미지의 양측에 시간적 문맥에 맞게 픽셀을 덧붙이는 방식으로 진행되며, 광학흐름에서 찾지 못한 픽셀은 생성형 모델을 사용하여 생성하게 된다.

이로 인해 마스킹의 정확도에 따라 전경의 픽셀들이 배경 이미지에 남아있을 가능성이 존재하였고, 전경 픽셀이 남게 된다면 이러한 픽셀들은 다른 프레임의 빈 공간에 참조 픽셀로 활용되어 마스킹의 정확도에 따라 아티팩트라[1]고 불리는 비정상적 잡음이나 형태가 발생하는

것을 확인하였다. 이러한 마스킹은 현재와 미래 프레임을 고려하는 단방향 세그멘테이션을 통해 수행되었는데, 그러나 비디오에서 현재 프레임은 과거, 미래 프레임과도 연관이 있으므로 양방향 마스킹 사용 시 아티팩트 개선 유무에 관한 실험이 필요하기에, 이에 본 논문에서는 양방향 세그멘테이션 기법을 적용해 마스킹 성능을 확인하고 아티팩트 개선 유무를 확인하고자 한다.

2. 본론

본 논문에서 비디오 리타겟팅을 위해 활용하는 아웃페인팅 기법은 그림 1 과 같이 광학흐름, 세그멘테이션, 픽셀 참조를 통한 인페인팅과 아웃페인팅으로 구성된다.

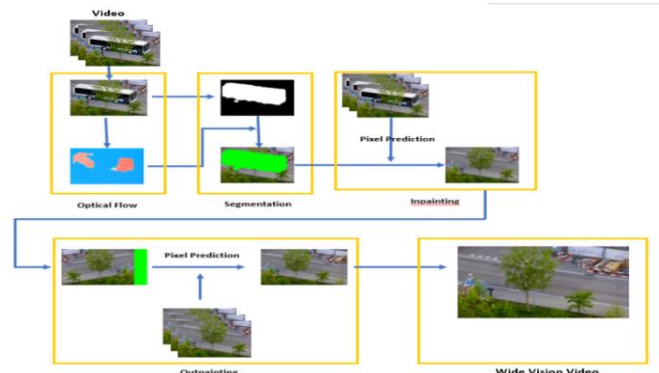


그림 1. 아웃페인팅 파이프라인

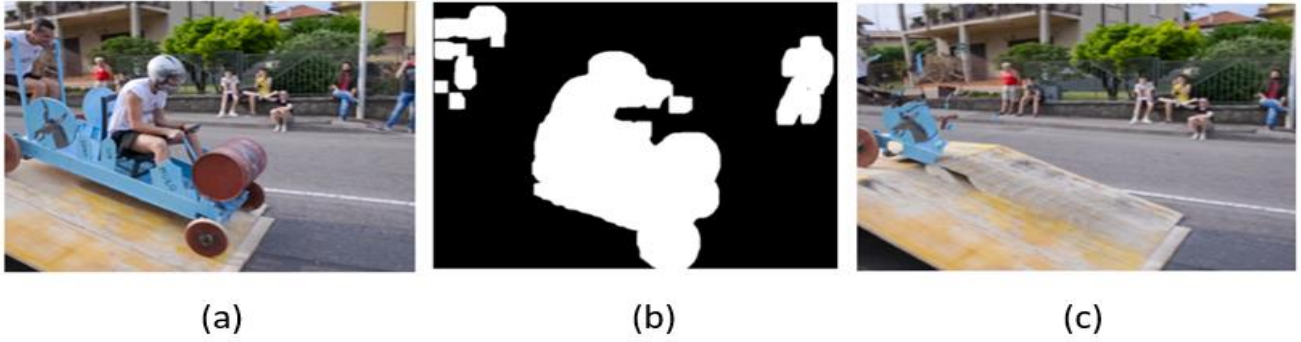


그림 2. 단방향 마스킹과 아티팩트

앞서 설명한 바와 같이, 해당 기법은 원본 이미지에 대하여 광학흐름 이미지, 전경에 대한 마스킹 이미지를 생성한다. 이때 생성되는 마스킹은 원본 이미지에서 빈 공간으로 표기되며 앞서 생성한 광학흐름을 이미지를 이용해 다른 프레임에서 픽셀을 가져오는 방식으로 빈 공간을 예측한다. 생성된 이미지는 배경 프레임이라 불리며, 배경 프레임에 빈 공간을 외삽 후 다시 이 빈 공간을 광학흐름을 이용한 픽셀 예측을 통하여 다른 프레임에서 픽셀을 가져오게 되며 이후 예측하지 못한 픽셀은 생성형 모델을 사용하여 생성하게 되며 아웃페인팅된 외삽 영역은 원본 이미지 우측에 병합된다. 이후 배경 이미지를 반전하여 다시 빈 공간을 외삽 후 픽셀 예측을 수행하여 아웃페인팅을 진행하고 아웃페인팅된 외삽 영역은 반전되어 원본 이미지 좌측에 병합된다.

그림 2의 (b)는 (a)의 원본 이미지에 단방향 세그멘테이션 적용시 생성되는 마스크이며 이로 인해 (c)와 같은 아티팩트가 발생한 것을 확인할 수 있다. (b)와 같은 마스크는 현재 프레임과 그 다음 프레임간 상관관계를 측정하는 Co-Attention Mechanism을 이용하여 운동성이 큰 전경에 마스킹이 적용되며, 현재 프레임과 그 다음 프레임을 이용하는 순방향이 적용된다.

3. 실험결과

본 논문에서는 비디오의 현재 프레임은 과거, 미래와 연관이 있다는 점을 고려할 때, 단방향 세그멘테이션이 아닌 양방향 세그멘테이션을 통해 아티팩트 개선 유무를 확인하고자 한다. 기존 세그멘테이션 기법[2]이 현재와 미래 프레임을 이용하므로, 양방향 세그멘테이션을 위해 광학흐름을 이용한 세그멘테이션 모델[4]에 현재와 과거 프레임의 광학흐름을 이용하여 마스킹을 진행하였다.

순방향과 역방향 정보가 모두 들어있는 마스킹을 생성하기 위하여 실험은 현재, 미래 프레임을 이용한 마스킹과 현재, 과거 프레임을 이용한 마스킹을 서로 더하는 방식으로 이루어졌다. 제안 방법으로 재생성한 세그멘테이션은 그림 3과 같으며, 해당 마스킹 영역으로 생성한 실험 결과는 그림 4와 같다.



그림 3. 양방향 마스킹



그림 4. 제안한 방법의 실험 결과

4. 결론

현재, 미래 프레임을 이용하는 순방향 마스킹과 현재, 과거 프레임을 이용하는 역방향 마스킹을 합성할 시 그림 2의 (b)와 같이 단방향에선 검출되지 않은 전경이 검출되는 것을 실험을 통해 확인하였다.

마스킹을 합성하는 것은 마스킹이 가지는 오류 또한 같이 합성되게 되지만, 적절한 임계값 설정을 통해 오류를 줄일 수 있었으며 양방향 마스킹을 통해 아티팩트가 개선된 것을 확인할 수 있었다.

ACKNOWLEDGMENT

"This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2024-2021-0-02046) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation)"

참고 문헌

- [1] Dehan, Loïc, et al. "Complete and temporally consistent video outpainting." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.
- [2] Lu, Xiankai, et al. "See more, know more: Unsupervised video object segmentation with co-attention siamese networks." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019.
- [3] Teed, Zachary, and Jia Deng. "Raft: Recurrent all-pairs field transforms for optical flow." *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*. Springer International Publishing, 2020.
- [4] Pei, Gensheng, et al. "Co-attention Propagation Network for Zero-Shot Video Object Segmentation." *arXiv preprint arXiv:2304.03910* (2023).