

# 관심영역 기반의 동영상 리타겟팅 사례 연구

배재현<sup>†\*</sup>, 김동휘<sup>†\*</sup>, 이수진<sup>†\*</sup>, 진준규<sup>°</sup>, 배병준<sup>\*</sup>, 조숙희<sup>\*</sup>, 박상효<sup>\*\*°</sup>  
경북대학교, 한국전자통신연구원

\*\*s.park@knu.ac.kr

## A Case Study on ROI-based Video Retargeting

Jaehyun Baet<sup>°</sup>, Donghwi Kim<sup>†°</sup>, Sujin Leet<sup>°</sup>, Jun-gyu Jin<sup>°</sup>,  
Byeongjun Bae<sup>\*</sup>, Sukhee Cho<sup>\*</sup>, Sang-hyo Park<sup>\*\*°</sup>

Kyungpook National University<sup>\*\*°</sup>, ETRI<sup>\*</sup>

### 요약

본 연구는 아날로그 방송 영상인 4:3의 화면비를 가진 영상을 16:9의 화면비로 변환하기 위한 연구의 원형이다. 16:9의 화면비로 변환할 때 영상의 가로길이만 늘려야 하는 경우, 객체의 가로 비율이 극단적으로 늘어날 수 있는 상황이 발생하고, 이를 해결하기 위한 방법으로 일반적인 초해상화 기법을 적용하기에는 한계가 있다. 제안기법에서는 영상의 콘텐츠가 왜곡되는 것을 최소화 하기 위해 ROI 기반의 복잡도 판별을 적용하여 객체의 원 비율을 유지하며 비관심 영역에 대해서만 초해상화 기술 및 보간법을 적용하여 영상의 비율을 변환한다. 실험결과를 통해 변환 전 영상의 객체의 원 비율을 유지할 수 있음을 시각적으로 확인할 수 있었다.

### I. 서론

아날로그 방송의 화면비를 가진 VGA(640 x 480)에서부터 최근 4K-UHD(3840 x 2160)까지 발전하였다. 해상도의 발전이 자연스럽게 화면비의 변화를 일으켰다. VGA의 경우 4:3의 화면비를, 4K-UHD는 16:9의 비율을 각각 가진다. 이로 인해, 과거 아날로그 방송의 영상을 16:9의 화면비를 갖는 디스플레이에서 재생하면 영상의 좌우에 검정색 레터박스가 생기거나, 영상을 화면에 맞추게 되면 영상이 지나치게 왜곡되는 현상이 있다. 이를 해결하기 위해 비디오 리타겟팅(video retargeting)에 관련된 여러 연구가 있었다.

여러 연구들 중 seam carving 기법을 적용하여 비디오 리타겟팅을 수행하는 방법이 있다. Avidan과 Shamir에 의해 제안된 seam carving[1]은 콘텐츠 인지 기반 리사이징(resizing)으로 seam은 이미지 에너지 함수에 의해 정의된 픽셀 단위 이미지로 seam을 제거 또는 삽입하는 것을 반복하여 영상의 해상도를 변환한다. 이는 일반적으로 수행시간이 오래 걸리고, 영상 내 콘텐츠에서 대부분의 영역이 중요한 영역이라고 판단되면 콘텐츠 인지 기반 리사이징이 제대로 수행되지 않을 수 있다. Deep video retargeting[2]는 비디오 리타겟팅을 해결하기 위해 최초로 순환 신경망(recurrent neural network)을 사용하여 리타겟팅이 된 여러 후보군을 생성하여, 영상의 콘텐츠를 가장 잘 보존한 후보를 선택한다. 이때, 일부 배경의 손실을 유발할 가능성이 있다.

본 논문에서는 아날로그 방송의 화면비를 16:9의 화면비로 영상 콘텐츠의 화면비를 변환하는 비디오 리타겟팅 문제에 대한 연구를 진행하였으며,

관심기반영역(region of interests, ROI)에 기반하여 사람이 영상에서 주로 관심을 갖는 영역에 대한 복잡도를 고려하여 최신 영상 초해상화(video super-resolution)기술로 비디오 리타겟팅 문제를 해결하는 원형의 연구를 제시한다. 해당 문제를 최신 YOLO(You Only Look Once)[3]기반의 객체 검출(object detection)과 SOF-VSR[4]모델을 이용한 영상 초해상화의 효용성을 비디오 리타겟팅 문제에서 검증해보았다. 제안기법은 기존 비디오 리타겟팅 방법보다 간단하지만 여전히 개선의 여지가 많다는 것을 보여준다.

### II. 본론

본 논문에서는 세 단계의 과정을 통해 화면비 변환 연구를 진행하였다. 첫째, 딥 러닝 기반의 객체 검출 모델을 통해 영상의 복잡도를 측정한다. 둘째, 오픈소스 소프트웨어를 이용하여 패치를 생성한다. 셋째, 딥 러닝 기반의 영상 초해상화 모델을 이용하여 결과를 생성한다. 이와 같은 일련의 과정을 통해 화면비 변환에 대한 연구를 진행하였으며, 본 연구를 진행하기 위해 REDS[5]데이터셋을 이용하였다. 가로비 변환 연구의 원형으로 640 x 480의 영상을 853 x 480의 해상도로 변환하는 연구를 진행하였다.

첫 번째 단계인 동영상 복잡도 측정의 경우, YOLOv5를 이용하였고, 객체의 위치, 개수 등의 기준을 세워 복잡도를 파악하여 ROI 기반의 초해상화를 사용하여 화면비 변환, 혹은 단순히 리사이징을 이용하여 변환을 수행할 것인지에 대한 기준을 마련하였다. 복잡도를 판단하는 기준은 다음과 같다. 먼저, 영상의 프레임을 1:1:1 면적 비율로 T1, T2, T3 총 세 구역으로 나눈 후, T2에 객체의 개수가 3개 이상이거나 T2에서 객체가 차지하는



원본



보간법



제안기법

그림 1. 보간법과 제안기법의 비교

비율이 90% 이상인 경우, 복잡한 영상으로 판단하여 ROI 기반의 초해상화를 수행한다. 이 때, T2 구역 내의 객체 수를 구할 때, 다른 구역에도 객체가 걸쳐져 있을 경우 T2 구역 내의 객체가 차지하는 비율이 50% 이상일 경우에 복잡도를 판단하는 전체 객체 수로 포함한다. 여기서 T2에 대한 복잡도만 판단하는 이유는 사람이 영상에서 가장 중요하다고 생각하는 영역이 영상의 가운데 영역일 것이라고 판단하였다. 복잡하다고 판단되지 않을 경우 단순히 보간법을 사용한 가로 리사이징을 수행한다.

영상의 복잡도를 측정 후, 이를 기반으로 각각의 패치를 만든다. 이는 FFmpeg 프로그램을 이용하여 패치로 분할하여 생성하였다. 본 연구의 경우, 가로 및 세로의 크기가 동일한 3개의 패치를 이용하여 연구를 진행하였다.

세 번째 단계에서 사용한 영상 초해상화 모델로는 SOF-VSR을 사용하였다. SOF-VSR은 3개의 저해상도 영상의 정보들을 활용하여 레퍼런스 영상, 즉 가운데 영상 1개를 초해상화하는 딥 러닝 기반의 모델이다.

딥 러닝 기반의 모델은 보통 디컨볼루션 층이나 파이토치의 PixelShuffle 층을 통해 저해상도의 특징 맵이나 영상을 고해상도로 변환한다. 하지만 앞서 언급한 층들은 영상의 가로와 세로 모두에서 초해상화를 진행하기 때문에 화면의 가로 비율만 변환하는 본 과제를 직접적으로 해결하기 쉽지 않다. 따라서, 영상의 초해상화를 진행한 후, 이를 수학적 계산을 사용한 보간법을 통해 리사이징을 수행한다.

화면비 변환 방법은 다음과 같다. 복잡하다고 판단된 영상의 경우, 영상을 세 등분의 패치로 나누어진 T1 영역과 T3 영역에 대해서만 초해상화를 진행한다. 이는 중요하다고 판단된 T2의 객체가 왜곡되는 것을 방지하기 위함이다. T1과 T3에 대해서 SOF-VSR을 통해 초해상화를 진행한 후, T2 패치의 해상도와 동일하도록 보간법을 사용하여 해상도를 맞춰준다. 마지막으로 T1, T2, 그리고 T3를 합쳐서 변환이 완료된 영상을 얻는다. 그림 1은 4:3 화면비를 갖는 영상을 각각 보간법과 제안기법을 사용해서 얻은 16:9로의 화면비 변환의 결과를 같은 영역에 대해 자른 예시이다. 제안기법이 T2 영역에서 원본 객체의 크기를 그대로 유지하기 때문에 객체가 가로로 과도하게 늘어나는 왜곡이 덜한 것을 시각적으로 확인할 수 있다.

### III. 결론

본 논문에서는 4:3 화면비의 영상을 객체 검출 및 영상 초해상화와 보간법을 통해 16:9의 화면비를 갖도록 변환하는 방법의 원형을 제안한다. 제안기법을 통해 객체의 비율에 대한 왜곡이 감소함을 보여주었지만, 계단

현상 등의 문제점에 대한 개선의 여지가 있다는 것을 확인할 수 있었으며 이를 보완하는 연구를 지속하여 수행할 예정이다.

### ACKNOWLEDGMENT

본 논문(저서)은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No.2021-0-00087-002, SD/HD급 저화질 미디어의 고품질 변환 기술 개발, 50%)과 교육부 및 한국연구재단의 4단계 BK21 사업(경북대학교 컴퓨터학부 지능융합 소프트웨어 교육연구단)으로 지원된 연구임(4199990214394, 50%).

† : 모두 연구에 동등하게 기여하였음.

### 참고 문헌

- [1] Avidan and Shamir, "Seam Carving for Content-Aware Image Resizing," ACM Transactions on Graphics, Vol. 26, No. 3, pp.267-276, 2007
- [2] Cho, Donghyeon, et al. "Video Retargeting: Trade-off between Content Preservation and Spatio-temporal Consistency." Proceedings of the 27th ACM International Conference on Multimedia. 2019.
- [3] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
- [4] Wang, Longguang, et al. "Deep video super-resolution using HR optical flow estimation." IEEE Transactions on Image Processing 29 (2020): 4323-4336.
- [5] Nah, Seungjun, et al. "Ntire 2019 challenge on video deblurring and super-resolution: Dataset and study." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2019.