

360 도 비디오의 객체 증강 saliency map 생성 방법

심유정, 서지민, 이명진

한국항공대학교

yoojeong2018@kau.kr, cyinxi@naver.com, artistic@kau.ac.kr

Method of creating augmented saliency map for 360-degree video

Yoojeong Shim, Jimin Seo, Myeong-jin Lee

Korea Aerospace University

요 약

360 도 영상은 기존 미디어와 다른 몰입감을 제공하지만 HMD 기반 시청은 멀미, 신체적 불편함 등을 유발할 수 있다. 또한, 시청 디바이스 보급 문제, 네트워크 대역의 문제, 단일 소스 다중 이용의 수요 등으로 일반 디스플레이 기반 서비스 수요도 존재한다. 본 논문에서는 360 도 영상의 일반 디스플레이 서비스를 위한 뷰포트 추출에 필요한 영상 내 객체의 동적 속성을 활용한 시각적 관심 지도 증강 기법과 이를 이용한 서비스 구조를 제시한다.

1. 서론

코로나 이후로 사회적 거리두기, 여행 제한 등에 따른 가상현실을 통한 관광, 공연 관람 등 다양한 실감형 미디어 서비스가 등장하고 있다. 이러한 몰입형 미디어는 기존의 2D 영상에 비해 높은 몰입감이나 현실감을 제공하지만, 시청자들의 360 도 비디오 시청시 HMD 사용으로 인한 멀미, 두통, 신체적 불편함이 유발 가능하다 [1].

360 도 비디오 응용을 위해 영상 시청자의 시청 패턴을 시각적 관심지도의 측정과 추정을 통해 파악하여 콘텐츠 제작, 스트리밍, 영상 압축에 사용하기 위한 연구가 수행되고 있다 [2,3]. 이들 연구들은 360 도 비디오 시청 패턴 데이터를 이용하여 관심지도와 시선이동 모델을 학습시키는 방법들을 사용하기 때문에, 생성된 관심지도의 중심 바이어스 특성과 빠른 동적 속성 객체를 고려하지 못하는 한계가 존재한다. 360 도 비디오에 대해 자동으로 뷰포트를 추출하는 연구들도 수행되어 왔으나, 기존 관심지도나 시청자들의 시청 패턴 데이터만을 사용하기 때문에 생성된 뷰포트가 영상내 동적인 변화를 추종하지 못하고, 다중 객체들이 산재되어 존재하는 영상에 대해 뷰포트 품질이 저하되는 한계가 존재한다 [4,5,6].

본 논문에서는 360 도 비디오로부터 뷰포트 자동 생성에 활용하기 위해 기존 관심지도에 객체의 동적 속성을 증강하는 방법을 제안한다. 360 도 비디오 내 출현 객체들과 해당 객체들의 움직임에 따라 동적 속성을 정의하고, 이를 활용하여 객체 증강 시각적 관심지도를 생성하는 방법을 제안한다.

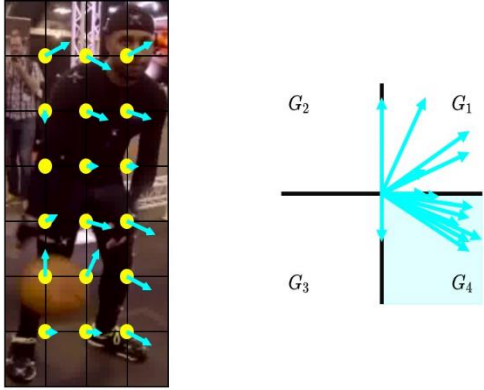
2. 제안하는 객체증강 시각적 관심지도 생성방법

기존 HM position data 를 사용한 시각적 관심지도 생성 또는 추정 은 여러 시청자들의 관심도를 통계적으로 누적하는 방법을 사용하므로 특정 객체에 대한 관심도가 평균적으로 반영되어 있어서, 시각적 관심지도만을 이용한 뷰포트 생성시 주요 객체를 포함하지 못할 수 있다. 또한, HM position data 의 경우에는 data 획득 방식에 따라 central bias 가 존재할 수 있다.

본 논문에서는 객체의 동적 속성을 기반으로 하는 객체 증강 saliency map 을 생성하는 방식을 제안한다. 제안 방법은 영상에서 객체들을 검출하고, 각 객체를 중심으로 한 다음의 Gaussian 커널을 시각적 관심지도에 중첩 시킨다.

$$G_{t,k}(\mathbf{p}) = e^{-\frac{-(\mathbf{p}-\mathbf{p}_{t,k})^2}{2\sigma_{t,k}^2}}, \quad (1)$$

수식(1)에서 $p_{t,k}$ 는 검출 객체의 대표 화소 위치로서, 사람 객체의 경우 얼굴 영역인 바운딩 박스의 6/7 높이 지점이다. $\sigma_{t,k}$ 는 t 번째 프레임의 k 번째 객체 바운딩 박스 너비의 절반이다.



(a) motion vectors (b) motion vector classification

그림 1 대표 포인트의 모션 벡터와 방위에 따른 그룹화 예시

객체의 동적 속성을 이용하여 시각적 관심지도를 증강시키기 위해 객체의 활동성과 변위를 사용한다. 변위 추정엔 Lukas Kanade optical flow 알고리즘을 사용하여 그림 1과 같이 객체의 18 개 대표 화소의 optical flow 를 추정하였다. 움직임 벡터의 방향은 4개의 bin 으로 분류되며, 주 방향 bin 의 motion vector 크기 평균이 최종 객체 변위이다. 4 개 방향 bin 인 G_d 는 방향의 분류를 위해 정의되었으며 기저 벡터는 다음과 같이 정의된다.

$$n_d = \left[\cos \frac{(2d-1)\pi}{4}, \sin \frac{(2d-1)\pi}{4} \right], d = 1,2,3,4 \quad (2)$$

여기서 d 번째 방향 bin 은 수식 (3)과 같이 정의한다.

$$G_d = \{ p \mid \operatorname{argmax}_{m=1,2,3,4} v(p) \cdot n_m = d, \forall p \in S \} \quad (3)$$

객체 O 의 방향 인덱스는 다음과 같다.

$$d_{max}(O) = \operatorname{argmax}_d |G_d| \quad (4)$$

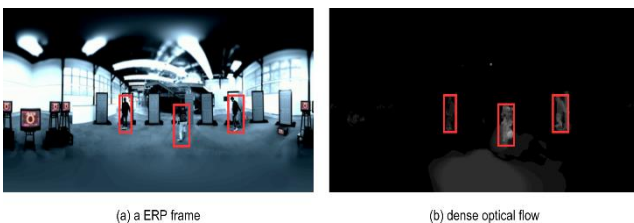
각 객체의 변위는 다음과 같다.

$$d(o_{t,k}) = \frac{1}{|G_{d_{max}}|} \sum_{p \in G_{d_{max}}} \|v(p_{t,k})\| \quad (5)$$

객체의 변위에 따른 가중치 식은 다음과 같다.

$$w_d(o_{t,k}) = (1 - \exp(-\frac{d(o_{t,k})}{d_{th}})) \quad (6)$$

d_{th} 는 변위 크기 임계 값이다.



(a) ERP frame (b) dense optical flow

그림 2 인접 프레임의 dense optical flow 예시

객체의 활동성은 Gunnar Farneback 이 제안한 2 차원 밀집 optical flow 방식을 사용해 전체 픽셀의 optical flow 를 계산하여 크기를 평균한다.

그림 2 은 테스트 시퀀스의 활동성 크기 분포를 보여준다. 해당 영상의 중앙 객체는 제자리에서 회전하며, 양 옆 객체는 약간의 움직임만 존재하기 때문에 그림과 같이 중앙 부근이 강조된 결과가 나타난다. 이는 영상의 실 시청자들이 위치이동을 하는 객체 뿐만 아니라 활동성을 보이는 객체에 관심을 가질 가능성이 존재하므로 이를 반영하기 위함이다. 객체의 활동성은 객체 바운딩 박스 내부 밀집 optical flow 의 변화량을 통해 계산하며 계산 수식은 다음과 같다.

$$a(O) = \frac{1}{A(R)} \sum_{p \in R} \|v(p)\| \quad (7)$$

R 은 객체 O 의 영역을 나타내며 A(·)은 영역의 면적을 나타낸다. 객체의 활동성 가중치는 수식(8)과 같다.

$$w_a(o_{t,k}) = 1 - e^{-\frac{a(o_{t,k})}{a_{th}}} \quad (8)$$

객체의 동적 속성이 증강된 시각적 관심지도는 시각적 관심지도 $S_f^{GT}(p)$ 와 증강계수 α 를 이용하여 다음과 같이 계산한다.

$$S_t(p) = (1 - \alpha)S_f^{GT}(p) + \alpha \sum_{k=1}^{N_t} \frac{w_d(o_{t,k}) + w_a(o_{t,k})}{2} G_{t,k}(p) \quad (9)$$

3. 실험결과

실험을 위해 사용한 제안 방법의 객체 검출기는 YOLO(You Only Look Once) v2 로 VOC 2007 과 VOC 2012 데이터셋을 사용하여 학습된 모델을 사용하였다. 테스트 시퀀스는 PVS-HM 데이터 셋에서 동적인 사람 객체를 포함하는 360 도 영상으로 영상 및 머리 움직임 센서 자료를 수집해 사용하였다[6]. 시퀀스들은 최소 29.97 Hz, 최대 30.00Hz 의 fps 를 가진다.

그림 3 은 실험영상의 GT, 제안 방식을 사용해 추출해 생성한 saliency map 을 나타낸 그림이다. 상단의 GT saliency map 은 중앙 부근만 높은 시각적 관심도를 보이지만, 제안 방법으로 생성한 객체 증강 saliency map 은 영상 내 전체 객체를 고려하므로 검출 객체 중심으로 시각적 관심도가 형성된 것을 확인할 수 있다. 중앙의 테스트 시퀀스는 영상 적도 중앙부근에 넓게 시각적 관심도가 분포하며 객체가 존재하지 않는 영상의 중앙에도 시각적 관심도가 분포한다. 이는 central bias 현상에 의해 발생하며, 제안 방식으로 생성한 객체 증강 saliency map 은 영상 내 객체 위주로 시각적 관심도를 형성한 것을 확인할 수 있다. 마지막으로, 하단 영상의 GT saliency map 은 중앙 부근에 높은 시각적 관심도를 보이지만, 제안 방식의 결과는 객체의 이동 위치에 중앙보다 높은 시각적 관심도를 보이는 것을 확인할 수 있다.

현재 실험에 사용한 가중치인 α 는 0.3 으로 이는 적용 대상 시스템이나 응용에 따라 효과적으로 설정할 필요가 있다.

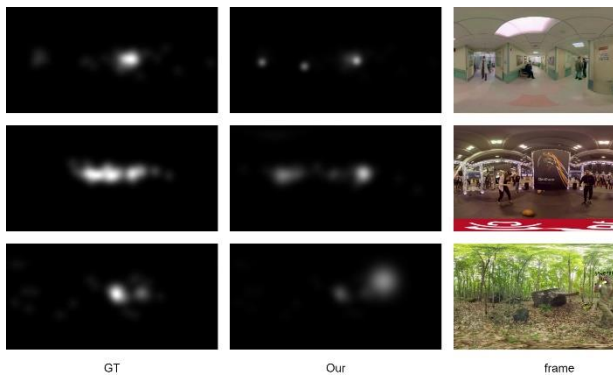


그림 3 GT /제안 saliency map 과 해당 프레임 이미지

4. 결론

본 논문에서는 360 도 영상에 대해 객체 증강 saliency map 의 생성 방식을 제안하였다. 제안 방식은 영상 내 객체의 동적 속성을 흥미 요소라 가정하고 객체의 동적 속성이 변위, 활동성에 대한 정의 및 각 객체의 동적 속성에 따른 증강 방식을 제시하였다. 이는 향후 동적 속성뿐만 아니라 영상 내 객체 혹은 이벤트의 상대적인 관심도에 따른 증강을 통해 응용할 수 있다.

Acknowledgement

이 논문은 2021 년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(NRF-2018R1D1A1B07050603, 360 도 비디오의 시각적 관심도 추정 및 활용 방안 연구, 50%)과 2021 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(2016-0-00144, 시청자 이동형 자유시점 360VR 실감미디어 제공을 위한 시스템 설계 및 기반기술 연구, 50%)

참고문헌

- [1] Lee, B.-R., et al. "표준 소개-멀미 저감을 위한 머리 장착형 영상 장치 기반 가상현실 콘텐츠 제작 지침." TTA Journal (2018): 88-94.
- [2] M. V. Broeck, F. Kawsar, J. Schöning, It's All Around You: Exploring 360° Video Viewing Experiences on Mobile Devices, ACM Multimedia 2017: 762-768
- [3] 이명진, 강두식, 박건우. "360 도 파노라마 비디오의 시각적 관심도 연구 동향." 한국통신학회지 35.9 (2018): 25-31.
- [4] Y.-C. Su, D. Jayaraman, and K. Grauman, "Pano2vid: Automatic cinematography for watching 360 degree videos," ACCV 2016.

[5] H. Hu, Y. Lin, M. Liu, H. Cheng, Y. Chang, and M. Sun, "Deep 360 pilot: Learning a deep agent for piloting through 360ports videos," CVPR, July 2017, pp. 1396-

[6] M. Xu, et al., "Predicting head movement in panoramic video: A deep reinforcement learning approach," IEEE Trans Pattern Analysis and Machine Intelligence, 2018.