

계층구조 합성곱 신경망 기반 고해상도 동영상 프레임 고속 보간 방법

안하은^{1,2}, 정진우², 김제우²¹ 광운대학교, ² 전자부품연구원¹mysco226@kw.ac.kr, ²{jw.jeong, jwkim}@keti.re.kr

Hierarchical Convolutional Neural Network based Fast Frame Interpolation for High - Resolution Video

Ha-Eun Ahn^{1,2}, Jinwoo Jeong² and Je Woo Kim²¹Kwangwoon University, ²Korea Electronics Technology Institute

요 약

본 논문에서는 계층구조 합성곱 신경망 기반의 고해상도 동영상 프레임 고속 보간 방법을 제안한다. 기존의 고해상도 동영상 프레임 보간 방법은 시간 해상도와 공간 해상도를 분리하여 보간 하기 때문에, 예측된 보간 프레임이 블러(blur) 열화를 갖는 문제를 보인다. 제안하는 방법에서는 이러한 문제를 해결하기 위하여 계층구조 합성곱 신경망 기반의 보간 방법을 이용한다. 제안하는 계층구조 합성곱 신경망은 우선 저해상도의 광학 흐름 추정지도를 생성하고 이를 고해상도로 복원하여 프레임 보간을 수행한다. 이때, 저해상도 광학 흐름 지도를 추정할 때 사용된 특징 정보들을 활용하여 고품질의 고해상도 광학 흐름 지도를 추정한다. 실험을 통하여 제안하는 방법이 고해상도 프레임을 고속으로 보간하며, 동시에 블러 열화에 대한 성능 향상을 가짐을 보였다.

1. 서론

동영상 프레임 보간은 시간적으로 인접한 프레임들의 중간 프레임을 생성하여, 프레임 율을 향상시킴으로써 동영상의 시간적 해상도를 증가시키는 기술이다. 고전적인 동영상 프레임 보간 방법은 대부분 광학 흐름 추정을 통하여 수행되기 때문에, 프레임 보간 알고리즘의 성능은 정확한 광학 흐름 추정지도를 획득하는 것에 크게 의존한다. 최근에는 합성곱 신경망을 이용하여 보다 정확한 광학 흐름 추정지도를 예측하는 방법이 제안되었다 [1]. Liu 는 비 지도학습을 통하여 광학 흐름 추정지도를 생성하고 이를 통하여 인접한 프레임들을 블렌딩(blending) 하는 방법을 통하여 프레임 보간을 수행하는 방법을 제안하였다 [2]. 기존의 방법들 대비 고품질의 광학 흐름 지도 생성이 가능하지만, 광학 흐름 추정이 어려운 가려진 영역이나 불규칙적인 영상 변화를 가지는 영역에 대하여 블러(blur) 열화를 가지는 문제를 보인다.

이를 해결하고자 시계(視界)지도를 이용하는 방법도 연구되었다 [3]. 이 방법은 가려진 영역에 대하여 기존의 방법 대비 비교적 좋은 성능을 보이지만, 높은 연산복잡도를 가지는 문제가 있다. 이는 4K 프레임과 같은 고해상도 프레임 보간 시 큰 제약사항이 된다. 이러한 문제를 해결하기 위해서, 프레임 보간을 시간 해상도와 공간 해상도를 분리하여 보간 하는 방법도 제안되었다 [4]. 이 방법은 시간 해상도 보간을 저해상도에서 수행하고, 그 결과를 공간 해상도에 대하여 보간을 수행한다. 이를 통하여 고해상도 프레임을 고속으로 보간 가능함을 보였으나, 공간 해상도 보간 시 블러 열화를 가지는 문제를 보인다.

본 논문에서는 이러한 문제점을 해결하기 위하여, 계층구조 합성곱 신경망 기반 고해상도 동영상 프레임 보간 방법을

제안한다. 제안하는 방법에서는 우선 저해상도의 광학 흐름 추정지도를 생성한 뒤 멀티스케일 영상 정보를 이용하여 이를 고해상도로 복원한다. 이를 통하여 고해상도의 광학 흐름 추정지도를 직접 생성하는 방법에 비하여 비적 낮은 연산복잡도를 가지는 합성곱 신경망을 설계 할 수 있다. 또한, 해상도를 점진적으로 확대시키는 형태의 접근 방법을 통하여 고품질의 고해상도 광학 흐름지도를 생성 할 수 있으며, 이는 블러 열화를 제거하는데 도움을 준다.

본 논문의 구성은 다음과 같다. 2 절에서는 제안하는 방법에 대하여 설명한 후, 3 절에서는 실험을 통하여 제안하는 방법의 우수성을 검증한다. 마지막으로 4 절에서는 본 논문에 대한 결론을 맺는다.

2. 제안하는 방법

제안하는 전체 합성곱 신경망은 한 개의 광학 흐름 추정부와 두 개의 광학 흐름 보간부, 총 3 개의 계층구조로 이루어져 있다. 우선 광학 흐름 추정부에서는 시간적으로 인접한 입력 영상들의 광학 흐름 관계를 저해상도에서 추정한다. 광학 흐름 지도 보간부에서는 추정된 저해상도 광학 흐름지도를 고해상도로 복원하는 작업을 수행한다. 첫 번째, 광학 흐름 보간부에서는 FHD 해상도의 광학 흐름 지도 보간을 수행한다. 두 번째 광학 흐름 보간부에서는 이를 입력으로 받아 4K 해상도의 광학 흐름 지도보간을 수행한다. 이때, 저해상도 광학 흐름 지도를 추정할 때 사용된 특징 정보들을 활용하여 고품질의 고해상도 광학 흐름 지도를 추정한다. 최종적으로, 광학 흐름 보간부에서 복원된 광학 흐름지도를 이용하여 프레임 보간을 수행한다.

제안하는 방법에서는 계층구조 합성곱 신경망의 훈련을 위하여 YCbCr 색상 공간을 가지는 영상 패치(patch)를 생성한다. 한 개의 훈련 샘플(sample)은 T-1, T, T+1 의 시간정보를 가지는 YCbCr420 형태의 영상 패치들을 포함한다. 훈련 샘플의 Y 영상은 512x512 크기를 가지며 CbCr 영상은 256x256 크기를 가진다. 훈련을 위하여 사용한 손실함수 l 은 식 (1) 을 통하여 계산된다.

$$l = w_s l_s + w_w l_w + w_c l_c + w_e l_e \quad (1)$$

여기서 l_s 와 l_w 는 평활도와 워핑(warping) 손실함수이다. 평활도 손실함수는 추정된 광학 흐름지도가 지역적으로 닮아있을 경우 감소하고 그 반대의 경우 증가하도록 설계한다. 워핑 손실함수는 광학 흐름지도의 정확도를 향상시키도록 설계되며, 입력 프레임의 워핑 결과가 실제 정답과 상이할 때 증가한다. l_c 는 영상의 화소 손실함수로서, 광학 흐름지도를 통하여 생성한 프레임과 정답 프레임의 화소 차이에 의해 결정된다. l_e 는 에지(edge) 손실 함수로서 보간된 프레임과 정답 프레임의 에지영상의 차이로 계산된다. 에지 손실 함수 l_e 은 식 (2) 와 같이 계산된다.

$$l_e = \| \mathcal{E}(I') - \mathcal{E}(I_{gt}) \|_2 \quad (2)$$

여기서 \mathcal{E} 는 입력 영상에서 에지 지도를 추출한다. 본 논문에서는 holistically nested edge detection [5] 를 이용하여 에지 지도를 생성한다. I' 는 예측된 보간 프레임이며, I_{gt} 는 생성된 보간 프레임 정답 프레임을 의미한다. 식(1)의 w 는 각 손실함수에 적용하는 가중치로 본 논문에서는 $w_s=1$, $w_w=1$, $w_c=1$, $w_e=1$ 로 설정한다.

3. 실험 결과

제안하는 방법의 우수성을 보이기 위하여 기존의 프레임 보간 방법과 PSNR 비교 측정 실험을 수행한다. 성능 비교를 위하여, 본 논문에서는 고해상도 동영상 프레임 복원을 별도의 메모리 해제 없이 수행할 수 있는 [4]을 선택한다. 실험에 사용한 고해상도 동영상은 Ultra Video [6]의 Beauty 와 ShakeNDry 영상으로, 광학 추정에 있어 블러(blur) 현상을 야기하는 큰 움직임과 불규칙적인 영상 변화를 다량 포함한다. 또한, 추가적으로 고해상도 프레임 보간 시간 측정 실험을 수행한다. 표 1 은 제안하는 방법과 기존 방법들의 PSNR 측정 결과를 비교하여 보여준다.

표 1. 기존 방법과의 PSNR 비교

	Beauty	ShakeNDry
[4]	30.38	33.80
Ours	30.52	34.15

표 2 는 제안하는 방법과 기존의 방법들의 4K 영상에 대한 프레임 보간 시간을 비교하여 보여준다. 보간 시간 측정 실험을 위하여 1 대의 Nvidia Titan Xp GPU 장비를 이용하여, 연속된

1,000 개의 프레임을 보간 한 후, 전체 소모 시간의 평균치를 계산한다. 제안하는 방법은 4K 의 고해상도 영상에서 기존의 방법대비 약 1.37 배의 보간 속도 향상을 보인다.

표 2. 4K 영상에 대한 프레임 보간 시간 비교

	Running time(ms)
[4]	620ms
Ours	450ms

4. 결론

본 논문에서는 계층구조 합성곱 신경망 기반 고해상도 동영상 프레임 고속 보간 방법을 제안하였다. 제안하는 방법에서는 우선 저해상도의 광학 흐름지도를 추정하고 이를 광학 흐름 지도 보간부에서 고해상도로 복원한다. 이때, 각 계층 신경망에서 생성되는 특징 지도를 다음 신경망으로 전달하여 활용하였다. 이를 통하여 고해상도 동영상 프레임을 고속으로 수행할 수 있음을 보였다. 또한 실험을 통하여 제안하는 방법이 블러(blur) 열화에 대한 성능 향상을 가짐을 보였다.

ACKNOWLEDGMENT

이 논문은 2018 년도 과학기술정보통신부(정보통신·방송연구개발사업)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임.(2018-0-00837, 영상콘텐츠 초고속/초고화질 변환 기술 개발).

참고 문헌

- [1] Dosovitskiy, Alexey, et al. "Flownet: Learning optical flow with convolutional networks." Proceedings of the IEEE international conference on computer vision. 2015.
- [2] Liu, Ziwei, et al. "Video frame synthesis using deep voxel flow." Proceedings of the IEEE International Conference on Computer Vision. 2017.
- [3] Jiang, Huaizu, et al. "Super slomo: High quality estimation of multiple intermediate frames for video interpolation." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.
- [4] Ahn, Ha-Eun, Jinwoo Jeong, and Je Woo Kim. "A Fast 4K Video Frame Interpolation using a Hybrid Task -Based Convolutional Neural Network." Symmetry 11.5 (2019): 619.
- [5] Xie, Saining, and Zhuowen Tu. "Holistically -nested edge detection." Proceedings of the IEEE international conference on computer vision. 2015.
- [6] Le Feuvre, Jean, et al. "Ultra high definition HEVC DASH data set." Proceedings of the 5th ACM Multimedia Systems Conference. ACM, 2014.