

컨볼루션 신경망 기반 비디오 디인터레이스 기법

정진우, 안하은, 김제우

전자부품연구원

{jw.jeong, han324, jwkim}@keti.re.kr

Video Deinterlace based on Convolutional Neural Network

Jinwoo Jeong, Ha-Eun Ahn, Je Woo Kim

Korea Electronics Technology Institute

요 약

인터레이스 영상은 지난 수 십 년간 방송 및 비디오 레코딩 등에 광범위하게 사용되고 있으며 디인터레이스의 성능을 향상 시키기 위한 많은 연구가 이루어졌다. 이를 위한 것으로써 본 논문에서는 컨볼루션 신경망을 이용한 비디오 디인터레이스 기법을 제안한다. 제안한 방법은 SKIP 연결을 사용하여 낮은 수준 특징 정보를 뒷 단의 레이어까지 전달함으로써 성능 향상을 달성하였다. 실험 결과는 FFMPEG 에서 제공하는 디인터레이스 기법에 비해 전 영상에 걸쳐 우수한 성능을 제공하며, 특히 복잡한 영상에서 기존 알고리즘 대비 큰 폭의 성능향상을 보인다.

1. 서론

인터레이스(interlace) 영상은 지난 수 십 년간 방송 및 비디오 레코딩 등에 광범위하게 사용되고 있다. 인터레이스 영상은 대역폭을 절약하기 위해 하나의 화면을 두 개의 필드(짝수 라인과 홀수 라인의 화소값)으로 나누어 저장하고 있으며 두 필드사이에는 1/60 초의 시간 간격을 가진다. 데이터의 정보량이 절반으로 줄었기 때문에 단순히 두 필드의 조합으로 하나의 장면을 구성하는 경우 다양한 형태의 화질 열화가 발생할 수 있다.

이를 해결하기 위한 다양한 디인터레이스 기법이 존재하며, 이들은 주로 영상의 시공간의 상관 관계를 이용하여 영상을 복원해 낸다. 이런 방법 이외에 신경망을 이용한 기법들이 연구되었으며 [1], 최근에는 딥러닝을 이용한 디인터레이스 기법도 제안되었다 [2]. 컨볼루션 신경망(Convolutional Neural Network)은 영상 인식, 영상 처리 등 다양한 분야에 사용되고 있으며 기존 기법에 비해 우수한 성능을 제공하고 있다. 따라서 본 논문에서는 성능 향상을 위해 컨볼루션 신경망을 이용한 디인터레이스 기법을 제안한다.

2. 컨볼루션 신경망 기반 디인터레이스

디인터레이스 기법은 인터레이스 영상에서 사라진 짝수 및 홀수 필드의 영상을 복원하여 프로그레시브(progressive) 영상을 생성하는 것이다. t 와 $t+1$ 프레임의 프로그레시브 영상 X_t, X_{t+1} 이 존재한다고 가정하면 t 프레임의 홀수 영상 X_t^{odd} 와 $t+1$ 프레임의 짝수 영상 X_{t+1}^{even} 을 조합하여 인터레이스 영상 $I = \{X_t^{odd}, X_{t+1}^{even}\}$ 를 생성할 수 있다. 프로그레시브 영상을 만들기 위하여는 인터레이스 영상 I 에서 X_t^{even} 과 X_{t+1}^{odd} 을 생성해 내야한다. 따라서 제안한 컨볼루션 신경망 디인터레이스 기법에서 입력은 인터레이스 영상 I 이고 출력은 X_t^{even} 과 X_{t+1}^{odd}

가 된다.

제안한 기법의 신경망 구조는 그림 1 과 같다. 위에서 언급했듯이 입력 영상 I 는 시간 t 에서의 홀수 필드와 $t+1$ 의 짝수 필드로 구성되어 있다. 출력은 t 에서의 짝수 필드와 $t+1$ 에서의 홀수 필드이다. 입력 t 에서의 홀수 필드와 출력 t 에서의 짝수 필드를 조합하여 t 에서의 프로그레시브 영상을 생성한다. 마찬가지로 입력 $t+1$ 에서의 짝수 필드와 출력 $t+1$ 의 홀수 필드를 조합하여 $t+1$ 에서의 프로그레시브 영상을 생성한다.

제안한 컨볼루션 신경망은 전처리 레이어 (preprocess layer), 공유 레이어 (shared layer), 홀수 필드 레이어 (odd layer), 짝수 필드 레이어 (even layer)로 구성된다. 모든 레이어는 컨볼루션 레이어로 구성되며 SKIP 연결이 존재한다. 전처리 레이어는 5×5 의 필터 크기를 갖는 40 개의 커널로 구성되며 컨볼루션을 수행한 후에 ReLU 레이어가 적용된다. 전처리 레이어에서는 낮은 수준의 특징(feature) 정보를 추출하고 추출된 특징 정보는 SKIP 연결을 통해 공유 레이어와 홀수, 짝수 필드 레이어에 직접 입력으로 전달된다. 전처리 레이어를 제외한 다른 레이어의 필터 크기는 3×3 이며, 공유 레이어 1 과 2 는 40 개의 컨볼루션 커널로 구성되며 컨볼루션을 수행한 후에 ReLU 레이어가 적용된다. 공유 레이어 2 의 입력은 전처리 레이어의 출력과 공유 레이어의 출력의 합이 된다. 이는 낮은 수준의 특징 정보를 네트워크의 뒷부분까지 전달하기 위하여 적용하였다. 전처리 레이어와 공유 레이어는 특징 정보를 추출하는 기능을 담당한다고 볼 수 있으며, 공유 레이어에서의 출력은 홀수 및 짝수 필드 레이어의 입력이 된다. 홀수 및 짝수 레이어는 서로 독립적이다. 홀수 및 짝수 필드

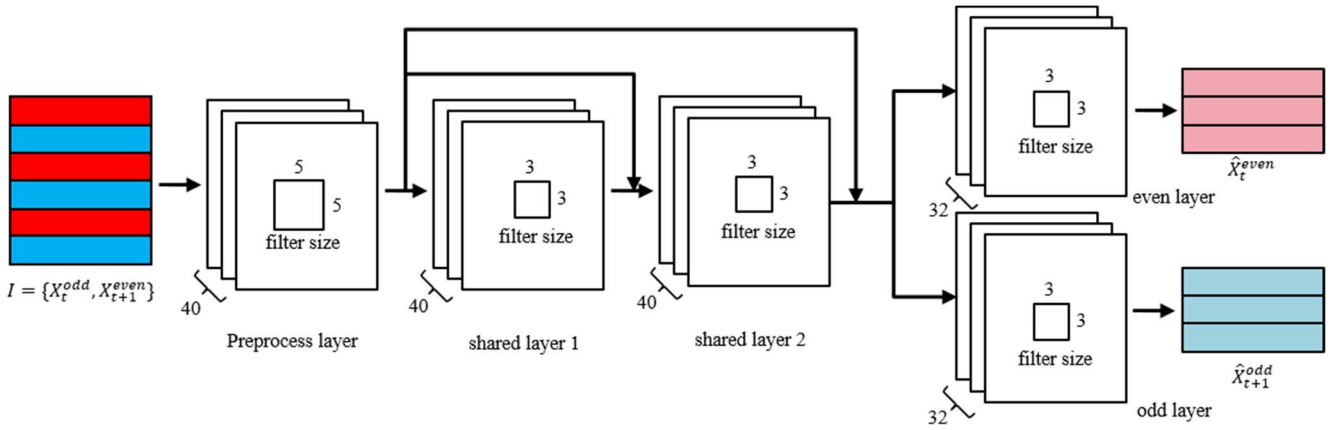


그림 1. 디인터레이스를 위한 제안 컨볼루션 신경망 구조

레이어의 입력은 전처리 레이어의 출력과 공유 레이어의 출력을 더한 것이다. 홀수 및 짝수 필드 레이어는 32 개의 커널을 가지며 컨볼루션 레이어만으로 구성되어 있다. 마지막 레이어는 1 개의 커널을 가지는 컨볼루션 레이어고 수직 방향으로 스트라이드(stride)가 2 이다. 따라서 출력 영상의 크기는 영상의 홀수 또는 짝수 필드의 크기와 일치한다. 마지막 레이어를 제외한 모든 레이어는 스트라이드 가 1 이므로 해상도가 변하지 않는다.

학습을 위하여 데이터는 $I, X_t^{even}, X_{t+1}^{odd}$ 의 트리플렛 (triplet) 형태로 구성하였다. 기준 영상 (ground truth)을 생성하기 위하여 프로그래시브 영상의 X_t, X_{t+1} 영상에서 입력 영상 I 와 출력 영상 $X_t^{even}, X_{t+1}^{odd}$ 로 생성하였다. 학습을 위하여 영상을 64x64 의 크기를 갖는 패치로 분할하였다. 학습 영상으로는 유튜브에서 다운로드 받은 영상 [3] 및 MPEG 의 테스트 영상을 사용하였으며 [4], 약 21 만개의 패치로 학습을 수행하였다.

손실 (loss) 함수는 다음과 같이 예측 영상과 원영상의 차이의 제곱인 L2 norm 손실 함수와 smooth 손실 함수를 사용하였다. λ_r 과 λ_s 는 각각 1 과 2×10^{-8} 로 설정하였다.

$$l_T = \lambda_r l_r + \lambda_s l_s \quad (1)$$

$$l_r = \frac{1}{N} (\| \hat{X}_t^{even} - X_t^{even} \|_2^2 + \| \hat{X}_{t+1}^{odd} - X_{t+1}^{odd} \|_2^2) \quad (2)$$

$$l_s = \frac{1}{N} (\| \nabla \hat{X}_t \|_1 + \| \nabla \hat{X}_{t+1} \|_1) \quad (3)$$

제안한 방법은 Pytorch 로 구현하여 테스트하였다. 최적화 기법으로는 Adam 최적화기를 사용하였으며 기본 설정값을 사용하였다. 학습율은 0.001 로 설정하였으며 100 EPOCH 만큼의 학습을 수행하였다.

3. 실험결과

제안한 디인터레이스 기법의 성능을 객관적으로 평가하기 위하여 우리는 PNSR (Peak signal to noise ration)을 사용하였다. 비교 알고리즘으로는 FFMPEG 에서 제공하는 디인터레이스 기법들인 NNEDI (neural network edge directed interpolation), YADIF (yet another deinterlacing filter),

BWDIF (Bob Weaver Deinterlacing Filter) 를 사용하였다 [6]. 이들 알고리즘으로는 우수한 성능을 제공하며 광범위하게 사용되고 있다. 테스트 시퀀스로는 MPEG 시퀀스 중의 학습에 사용되지 않은 영상의 일부를 사용하였다. 제안한 알고리즘은 밝기 신호에 대하여만 적용하였다.

표 1 은 제안 알고리즘과 비교 방법의 PSNR 결과를 보여준다. 제안 방법은 NNEDI, YADIF, BWDIF 대비 평균 1.71, 2.54, 1.26 dB 의 화질 향상을 보인다. 그리고 모든 영상에 대하여 기존 알고리즘 대비 우수한 결과를 보임을 알 수 있다. 특히 PSNR 이 35dB 이하로 낮은 영상에서는 1dB 에서 4dB 이상의 높은 화질을 제공하고 이는 제안 알고리즘이 복잡한 영역에서도 잘 동작하고 있음을 보여준다. 그림 2 는 주관적 화질 결과를 보여준다. 제안 알고리즘이 계단 현상 없이 더욱 선명한 화질을 보여줌을 알 수 있다. 제안 알고리즘은 1920x1080 해상도 영상 1 프레임을 처리하는 데 Titan Xp 에서 0.06 초의 시간이 소요된다.

표 1. 제안 알고리즘과 비교 방법의 Y PSNR 비교

	NNEDI	YADIF	BWDIF	제안 방법
Crowdrun	33.32	30.84	33.33	34.61
Parkjoy	30.04	27.52	29.89	31.46
kimono	42.56	41.45	42.49	43.38
Parkscene	37.16	35.61	38.23	39.77
Cactus	36.85	36.73	37.47	38.61
Basketballdrive	35.78	35.54	35.92	37.15
BQTerrace	33.71	33.21	33.73	35.82
Sunflower	45.73	43.95	45.38	46.04
Toys_and_calender	39.05	38.19	38.53	39.36
City_corr	32.49	31.46	33.74	34.71
foreman	31.27	34.30	34.24	35.92
평균	36.18	35.35	36.63	37.89

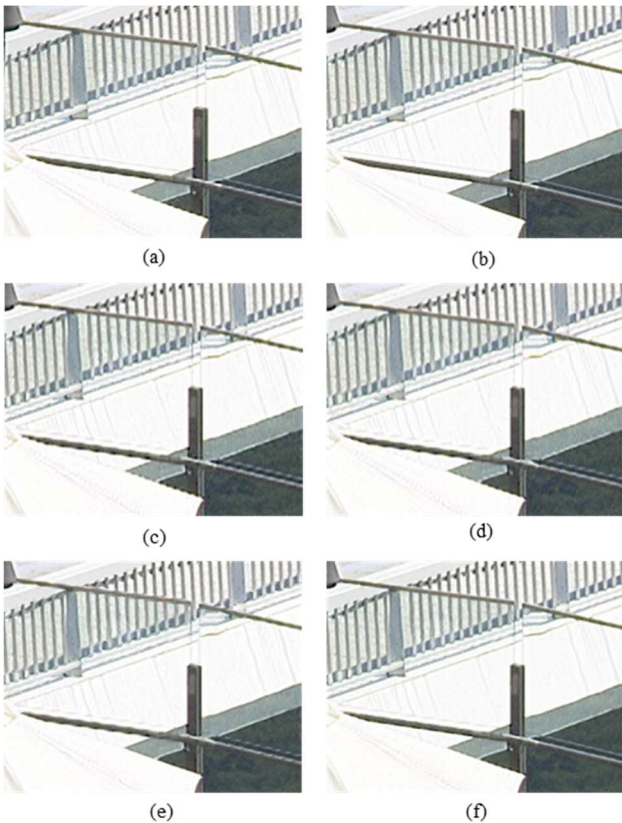


그림 2. 제안 알고리즘과 비교 방법의 주관적 화질 비교, (a) 인터레이스 영상, (b) 원영상, (c) NNEDI, (d) YADIF, (e) BWDIF, (f) 제안 방법

4. 결론

본 논문에서는 컨볼루션 신경망을 이용한 비디오 디인터레이스 기법을 제안하였다. 제안 방법은 낮은 수준의

특징 정보와 높은 수준의 특징 정보를 더함으로써 성능 향상을 제공하였다. 전 영상에 걸쳐 기존 기법 대비 우수한 성능을 제공하며, 특히 복잡한 영상에서 기존 알고리즘 대비 큰 폭의 성능향상을 보였다. 또한 Titan Xp 에서 FHD 크기의 영상을 초당 30 프레임씩 생성함으로써 고속으로 디인터레이스를 수행할 수 있다.

감사의 글

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

5. 참고문헌

- [1] Xianglin Wang and Yeong Taeg Kim, "An edge direction based neural network interpolator for video deinterlacing," International Conference on Neural Networks and Signal Processing, 2003. Proceedings of the 2003, Nanjing, 2003, pp. 1225-1228 Vol.2
- [2] Zhu, Haichao, Xueting Liu, Xiangyu Mao and Tien-Tsin Wong. "Real-time Deep Video Deinterlacing." CoRR abs/1708.00187 (2017)
- [3] Youtube, <https://youtube.com/>
- [4] F. Bossen, Common test conditions and software reference configurations, JCTVC-L1100, January, 2013
- [5] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization, " Proc. Int. Conf. Learning Representations. (ICLR), San Diego, USA, May 2015
- [6] FFMPEG, <https://ffmpeg.org/>