

## 컨볼루션 신경망을 이용한 다시점 비디오의 중간 시점 양자화 노이즈 제거

함유진, 강제원

이화여자대학교 전자전기공학과

yujinhamee@gmail.com, jewonk@ewha.ac.kr

### Quantization noise removal in an intermediate view of multi-view videos using convolutional neural network

Yu-Jin Ham, Je-Won Kang

Department of Electronic and Electrical Engineering

Ewha W. University

#### 요 약

본 논문에서는 컨볼루션 신경망을 이용하여 다시점 비디오의 중간 시점 양자화 노이즈를 제거하는 방안을 제안한다. 다시점 비디오에서 중간 시점의 화질을 개선하기 위한 방안으로 인접 시점의 정보를 활용하였다. 제안하는 알고리즘을 적용하여 중간 시점에서의 양자화 노이즈를 제거할 수 있으며, 화질 (PSNR, peak-to-noise ratio)를 개선할 수 있다. 인접 시점의 정보를 활용할 경우, 일반적인 양자화 노이즈에 대해서 학습한 결과 대비 성능 향상을 제공한다.

#### 1. 서론

비디오 압축은 저장과 전송을 위해 필수적인 과정이다. 비디오의 각 프레임이 공간적으로 부드럽게 변하는 성질과, 연속된 프레임 간의 시간적인 유사성을 이용하여 데이터 크기를 줄일 수 있다. 반면, 소비자들은 더 높은 화질의 비디오를 요구하고 있으며, 특히 생생한 가상환경 (VR) 장비에 기반한 몰입형 비디오의 수요가 증가하고 있다. 비디오 압축과 인지 화질의 균형을 위해서 몰입형 비디오에 대한 연구와 표준 정의가 활발히 진행되고 있다 [1][2].

몰입형 비디오 서비스는 고품질의 다시점 비디오로 구성되기 때문에, 일반 비디오에 비해 더 많은 비트가 필요하다. 다시점 비디오는 여러 카메라 각도로부터 촬영된 비디오 시퀀스를 합성하여 사용자에게 시점의 자유를 주는 3차원 비디오 기술이다. 다시점 비디오는 시점 최적화를 통해 기본 시점과 추가 시점으로

구분이 되며, 기본 시점은 전체 시점에서 가장 공통의 성분을 갖는 시점으로 화질이 비교적 높다. 본 논문에서 사용하는 데이터셋은 MPEG-I 에서 6 자유도 지원 몰입형 비디오의 공통실험조건을 테스트하기 위한 모델 TMIV[3] 로 복원된 영상이다 [4].

본 논문에서는 컨볼루션 신경망을 기반으로 노이즈를 제거하는 여러 알고리즘 중 높은 성능을 보이고 계산적인 관점에서 효율적인 FastDVDnet (Fast Deep Video Denoising Network)[5] 를 기본 모델로 사용한다. 다시점 비디오에서 인접한 시점은 서로 공간상의 중복이 많이 발생하므로, 이 특성을 활용한다. 또한, 비디오의 시간적 유사성을 학습하는 FastDVDnet[5] 의 네트워크 구조를 참고하여 비디오의 시·공간적인 특성을 모두 고려한다.

본 논문의 구성은 다음과 같다. 2 절에서는 논문에서 제안하는 기법과 FastDVDnet[5] 에 대한 배경에 대해 살펴본 후,

3 절에서는 실험환경과 성능평가를 분석한다. 마지막으로 4 절에서는 본 논문에 대한 결론을 맺고 향후 연구 방향에 대해 소개한다.

## 2. 본론

### 2.1 데이터셋

실험을 위해서 사용된 데이터는 MIV 모드로 복호화된 다시점 비디오 데이터셋으로, 총 7 개의 시퀀스로 구성된다 [6]. 시퀀스별 자세한 정보는 <표 1> 과 같다. 본 실험에서는 컴퓨터 그래픽이 아닌 자연적인 콘텐츠의 양자화 노이즈 제거를 위해서, 테스트 환경에서 시퀀스 E 와 L 을 사용하였다. 시점 별로 영상의 화질에 차이가 있기 때문에 가장 PSNR 이 낮은 비디오의 양자화 노이즈를 제거하는 것을 목표로 한다. 즉, 가장 PSNR 이 낮은 시점의 비디오를 테스트 데이터로 사용하고, 나머지 시점의 비디오를 학습 데이터로 사용하였다. 시퀀스 E 는 총 13 개의 시점으로 구성되어 있으며, 평균 화질 (PSNR)은 27.6645(dB) 이고 시점 4 의 화질이 25.4607(dB)로 가장 낮다. 시퀀스 L 은 총 12 개의 시점으로 구성되어 있으며, 평균 화질은 32.7969 이고 시점 6 의 화질이 27.8804(dB)로 가장 낮다.

표 1 다시점 비디오 데이터셋의 특성

Seq.	Name	Format	# of View	Resolution
A	Classroom Video	ERP	15	4096 × 2048
B	Technicolor Museum	ERP	23	2048 × 2048
C	Technicolor Hijack	Semi-ERP	10	4096 × 4096
D	Technicolor Painter	Perspective	16	2048 × 1088
E	Intel Frog	Perspective	13	1920 × 1080
J	Orange Kitchen	Perspective	25	1920 × 1080
L	Poznan Fencing	Perspective	10	1920 × 1080

### 2.2 FastDVDNet

Fastdvdnet[5] 은 컨볼루션 신경망 네트워크를 기반으로 비디오에서 노이즈를 제거하는 알고리즘이다. 타겟 프레임의 노이즈를 제거하기 위해 연속된 5 개의 프레임과, 표준편차가  $\sigma$ 인 노이즈맵을 입력으로 사용한다. 본 실험에서는 양자화 파라미터 37 로 복호화된 영상의 양자화 노이즈를 제거하는 것이 목표이기 때문에 복호화된 5 개의 연속된 프레임을 사용했지만 입력 가우시안 노이즈는 무시할 만한 수준으로 하여 입력으로 취한다. 일반적으로 노이즈맵은 가우시안인 경우 표준편차로 일정하며, 포아송인 경우 영상의 강도에 따라 달라진다.

비디오의 시간적으로 중복되는 특성을 반영하기 위해서 모션 추정과 보완 과정을 거치지 않고, 멀티-스케일 U-Net 과 Cascading 구조를 통해서 내재적으로 모션 정보를 반영한다. 첫 번째 디노이징 블록에서 U-Net 의 인코더단에 3 장의 프레임과 노이즈맵이 입력으로 들어가고 각 블록마다 1 장의 디노이징된 프레임이 출력되며, 모든 디노이징 블록은 동일한 가중치를 가진다. 2 번째 디노이징 블록에서는 1 번째 디노이징 블록에서 출력된 3 개의 프레임을 사용하여 가운데 프레임의 노이즈를 제거한다.

FastDVDnet의 모델 구조는 <그림 1>과 같고, 각 Denoising 블록은 변형된 U-Net [7]구조로 구성된다. FastDVDnet 은 별도의 모션 블록이 없기 때문에 End-to-end 학습이 가능하다.

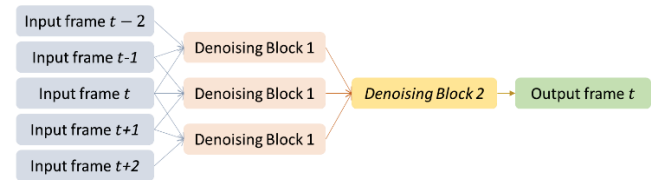


그림 1 FastDVDnet 아키텍처

## 3. 실험 결과

### 3.1 학습 및 테스트 환경

실험에서 사용한 비디오는 MPEG-I 몰입형 비디오의 공통실험조건에 따라 양자화 파라미터 (QP) 37 로 부-복호화된 영상이다. 시점 별로 영상의 화질에 차이가 있기 때문에 가장 PSNR 이 낮은 비디오의 양자화 노이즈를 제거하는 것을 타겟으로 한다. 즉, 가장 PSNR 이 낮은 시점의 비디오를 테스트 데이터로 사용하고, 나머지 시점의 비디오를 학습 데이터로 사용하였다. 학습 데이터와 테스트 데이터의 크기는 모두 1920 × 1080 으로 일정하고, 학습 시에는 랜덤한 영역에서 256 × 256 패치를 잘라서 사용한다.

### 3.2 성능평가

시퀀스 E 와 L 에 대해서 테스트를 한 결과, 시퀀스 별 특성에 의해 성능에 차이를 보였다 [6].

시퀀스 E 의 경우, 인접 시점끼리 영상의 변화가 적고 프레임 간 움직임이 상대적으로 적어 공간적, 시간적으로 화질 개선이 어려운 반면, 시퀀스 L 은 프레임 간 변화가 크고 사물이 빠르게 움직여 시-공간적인 정보를 활용하여 노이즈를 제거하는 것이 성능 향상에 도움이 되었다. 표 2 는 시퀀스 L 과 E 의 잡음 제거 후 향상된 성능을 보인다. 각각 0.31 dB 및 0.05 dB 의 PSNR 향상이 있었다.

또한, 인접 시점으로부터 학습한 모델과 같은 양자화 파라미터로 복호화한 비디오로부터 학습한 모델의 성능 비교를 위한 실험을 구성하였다. 일반적인 데이터는 YouTube 에서 다운 받은 데이터로 장면 변화 (Scene change) 를 준 데이터셋으로 학습하였다. 표 3 은 그 결과로 인접 시점에서 학습하고 다른 시점에서 테스트한 결과 0.12 dB 의 성능 향상이 있었다.

**표 2 FastDVDnet을 이용해 중간 시점의 노이즈를 제거한 결과 (평균값)**

	PSNR output	PSNR input	Gain
Seq L	28.1930	27.8804	0.3126
Seq E	27.7127	27.6645	0.0482

**표 3 인접 시점 데이터와 일반적인 데이터로 학습했을 때의 성능 비교**

	PSNR output	PSNR input	Gain
인접시점	28.1930	27.8804	0.3126
일반	28.0797	27.8804	0.1993

#### 4. 결론 및 향후 연구방향

본 논문에서는 컨볼루션 신경망을 활용한 딥러닝을 기반으로 몰입형 비디오의 양자화 노이즈를 제거하는 실험 및 분석을 진행하였다. 복호화가 된 비디오 프레임들로부터 노이즈를 제거하는 알고리즘으로 기존의 시간적으로 중복이 많은 비디오의 특성을 활용하는 FastDVDnet[5] 을 활용하였다. 향후 양자화 노이즈에 적합한 노이즈맵을 예측하여 양자화 노이즈를 제거할 수 있는 알고리즘에 대한 연구를 진행할 예정이다.

#### Acknowledgement

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터지원사업의 연구결과로 수행되었음 (IITP-2020-0-01460)

#### 참고문헌

- [1]“ICT R&D 기술로드맵 2023 - 방송·콘텐츠”, 정보통신기획평가원, 2018. 12.
- [2]“ICT 표준화전략맵 Ver.2019, 실감방송·미디어”, 한국정보통신기술협회, 2018. 10.
- [3] “Test Model 3 for Immersive Video”, ISO/IEC JTC 1/SC 29/WG 11 N18795 (October 2019)
- [4] 박현수, 박상효, 강제원. (2019). 6 자유도 전방위 몰입형 비디오의 압축 코덱 개발 및 성능 분석. 방송공학회논문지, 24(6), 1035-1052.
- [5] Tassano, Matias, Julie Delon, and Thomas Veit. "FastDVDnet: Towards Real-Time Deep Video Denoising Without Flow Estimation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.
- [6] 여은, 강제원 "컨볼루션 신경망을 이용한 6 자유도 비디오의 프레임 화질 개선," 전자공학회 하계학술대회 2020
- [7] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015.