

휘도 및 색차 성분에 따른 딥러닝 기반 SHVC 계층간 참조 픽처 생성 방법의 부호화 성능 분석

이우주, 이민훈, 황기수, 성준영, *오승준

광운대학교 전자공학과

{krosea, minhun, kisu031, jysung13mmlab, sjoh}@kw.ac.kr

Encoding Performance Analysis of Deep Learning based SHVC Inter-Layer Reference Picture Generation Method by Luma and Chroma Component

Wooju Lee, Minhun Lee, Gisu Hwang, Junyoung Sung, *Seoungjun Oh
Kwangwoon University

요 약

본 논문에서는 휘도 및 색차 성분에 따른 SHVC 계층간 참조 픽처 생성 방법의 부호화 성능을 분석한다. SHVC 상위 계층에서는 하위 계층의 픽처를 DCT-IF 기반 업샘플링하여 사용한다. 상위 계층의 부호화 성능을 높이기 위해 딥러닝 기반 필터링을 이용하여 휘도, 색차 성분의 고주파 신호 복원이 부호화 성능에 미치는 영향을 분석한다. 기존 Y 성분에만 VDSR 네트워크를 이용하여 필터링을 적용하였을 때보다 색차 성분까지 필터링을 진행할 경우 최대 2.18%, 평균 1.5% 감소된 결과를 보였다.

1. 서론

SHVC(Scalable High Efficiency Video Coding)는 실시간 비디오 전송 및 이종망(Heterogeneous) 서비스에 적합한 공간적, 시간적, 화질적 스케일러빌리티(Scalability)를 제공하는 확장 표준이다. SHVC 는 하나의 비트스트림을 통하여 다양한 해상도, 프레임률, 화질 등을 지원하도록 계층적 방식으로 부호화를 진행하고 사용자의 환경에 따라 비트스트림의 일부를 이용하여 복호화 가능한 방식이다. SHVC 의 계층 간 참조 방법은 그림 1 과 같이 시간적으로 동일한 위치에 있는 하위 계층의 픽처(Picture)를 DCT-IF(Discrete Cosine Transform-based Interpolation Filter) 기반 업샘플링(Upsampling) 방법을 사용하여 상위 계층의 참조 픽처 리스트에 추가하여 상위 계층 부호화에서 화면 간 예측과 같이 사용할 수 있도록 지원한다[1]. 하지만 DCT-IF 기반 업샘플링 방법은 고주파 성분을 온전히 재구성하기 어렵기 때문에 예측 신호로 선택되는 빈도수가 적다. 예측 신호로 선택되는 빈도수를 높이기 위한 방법으로 SC(Sparse-Coding)방식의 사전 학습(Dictionary Learning)을 이용한 SISR (Single Image Super Resolution) 접근법[2] 과 딥러닝을 이용한 SISR 방법[3]이 있다. 두 방법 모두 휘도 성분에 대해서만 SISR 을 적용하여 계층간 참조 픽처를 생성하였다. 비디오 코딩의 특성상 예측 신호를 원본에 가깝게 만들수록 압축 효율이 증가하기 때문에 색차 성분에서도 적용할 경우 추가적인 압축 효율을 기대할 수 있다.

따라서 본 논문에서는 기존의 방법[3]을 바탕으로 Y 성분뿐만 아니라 Cb, Cr 성분에 대해서도 SISR 을 적용하여 상위 계층의 픽처를 생성하고 생성된 픽처를 통해 부호화 된 압축 성능을 분석한다.

본 논문의 구성은 다음과 같다. 2 장에서는 적용한 방법에 대해 설명하고, 3 장에서는 적용한 방법의 성능을 분석하고 4 장에서 결론을 맺는다.

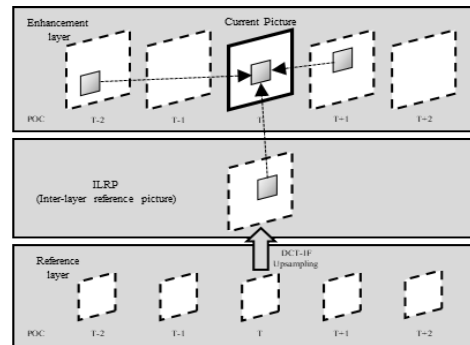


그림 1. SHVC 의 계층 간 참조 방법의 예

2. 딥러닝 기반 계층간 참조 픽처 생성 방법

2.1 계층간 참조 픽처 생성 방법

그림 2 는 계층 간 참조 픽처 생성을 위해 적용한 방법이다 [3]. 위 방법은 다양한 배율을 고려하기 위하여 SISR 에서 다양한 배율에 강인한 VDSR 모델을 사용하고 다양한 부호화 손실을 고려하기 위하여 하위계층의 복원된 픽처와 원본 영상을 이용하여 VDSR 네트워크를 학습하여 사용한다. 위 방법은 VDSR 모델의 파라미터 필터 크기 및 개수, 계층의 깊이는 [4] 에서 사용한 값을 차용하였다.

본 논문에서는 휘도 및 색차 성분에 따른 계층간 참조 픽처 생성 방법의 부호화 성능을 분석한다. 기존 방법은 휘도 성분에 대해 학습과 필터링을 제안하였다. 본 논문에서는 휘도 성분 뿐만 아니라 색차 성분에 대해서도 학습을 진행하고 계층간 참조 픽처를 생성하여 부호화 성능을 분석한다.

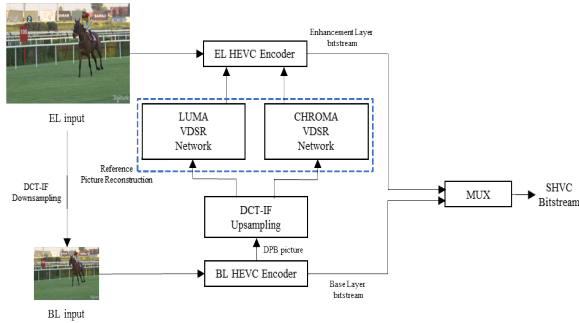


그림 2. 적용한 방법의 부호화기 블록도

2.2 학습 방법

기존 VDSR 에서는 배율이 다른 영상에 대하여 학습을 진행하였다. 하지만 본 논문에서는 부호화 손실에 강인한 VDSR 네트워크를 학습시키기 위하여 QP 값에 대하여 학습을 진행하였다. 학습에 사용된 QP 값은 {22, 27, 32, 37}이며, Ultra-Video Group 의 120fps FHD 테스트 시퀀스를 SHM CTC(Common Test Conditions) 임의의 접근 환경에서 부호화하여 사용했다. 상위 계층의 원본 영상과 하위계층의 복원된 영상을 60 프레임마다 한 장씩 추출하여 학습 데이터로 사용하였다. 기존 계층간 업샘플링 방법과 성능 비교를 위하여 Y 성분과 Cb, Cr 성분을 각각 학습을 진행하였다. 학습은 L2 손실 함수(Loss function)를 사용하여 원본 영상과 QP 값에 따라 부호화된 후 DCT-IF 기반 업샘플링 된 픽처를 사용하였다.

3. 실험 결과 및 분석

본 논문에서 제안하는 부호화기의 성능을 평가하기 위해서 MPEG 표준 테스트 시퀀스 표준 영상을 SHM CTC Spatial 2x 환경, RA(Random Access), AL(All Intra), LP(Low delay P)에 대하여 SHM12.0 을 이용하여 부호화를 진행하였다. 본 논문에서 제안하는 방법의 부호화 효율 평가를 위한 지표로 BD-BR(Bjontegaard-Delta Bitrate) 을 사용하여 표준 SHM 대비 상위 계층의 비트율을 비교하였다. 표 1 은 SHM 대비 비트율을 정리한 결과이다. 좌측부터 첫번째는 사전 학습 방법을 적용한 결과이다. 두번째는 Y 성분에 대해서만 VDSR 을 적용한 기존 방법이다. 세번째 방법은 하나의 네트워크를 Y, Cb, Cr 모든 성분에 대하여 학습한 방법이다. 네번째 방법은 Y 성분을 학습시킨 네트워크와 Cb, Cr 두가지 성분을 학습시킨 네트워크 2 개의 네트워크를 사용한 결과이다. 다섯번째 방법은 Y, Cb,

Cr 세가지 성분을 각각의 네트워크를 학습시켜 3 개의 네트워크를 사용한 결과이다. Y 성분에 대해서만 VDSR 필터링을 진행한 경우와 비교하여 색차 성분까지 필터링을 진행할 경우 최대 2.18%, 평균 1.5% 감소된 결과가 측정되었다. 색차 성분의 경우 Cb, Cr 성분에 대하여 각각 학습을 진행한 경우와 Cb, Cr 성분을 동시에 학습시킨 네트워크는 유사한 비트감소율이 측정되었다.

4. 결론

본 논문에서는 휘도 및 색차 성분에 따른 계층간 참조 픽처 생성 방법의 부호화 성능을 분석한다. YCbCr 의 Y 성분뿐만 아니라 Cb, Cr 성분에 대해서도 SISR 방법의 VDSR 네트워크를 이용하여 필터링을 진행하고 고주파 성분이 복원된 픽처를 상위 계층의 참조 픽처로 사용하는 방법이다. 제안하는 방법의 성능의 평가를 위하여 기존 방법과 제안하는 방법을 비트율 측면에서 비교하였다. 결과적으로 기존 Y 성분에만 VDSR 네트워크를 이용하여 필터링을 적용하였을 때보다 색차 성분까지 필터링을 진행할 경우 최대 2.18%, 평균 1.5%감소된 결과를 보인다. 그러나 계산 복잡도 측면에서 개선해야할 여지가 존재한다.

감사의 글

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터지원사업의 연구결과로 수행되었음" (IITP-2019-2016-0-00288)

참고 문헌

- [1] Haechul Choi, Kyung Il Lee, Jung Won Kang, Seong-Jun Bae, and Jeong-Ju Yoo, "Overview and Performance Analysis of the Emerging Scalable Video Coding", Journal of Broadcast Engineering Vol.12, no. 6, pp. 542-554, November, 2007.
- [2] Schneider, Jens, Johannes Sauer, and Mathias Wien. "Dictionary learning based high frequency inter-layer prediction for scalable HEVC", Proceeding of IEEE Visual Communications and Image Processing (VCIP). IEEE, St. Petersburg, USA, pp.1-4, December, 2017.
- [3] 이우주, 이종석, 심동규, 오승준, "CNN 기반 SHVC 계층간 예측을 위한 업샘플링 방법", Workshop on Image Processing and Image Understanding(IPIU), Jeju, Korea, February, 2019.
- [4] J. Kim, J. K. Lee and K. M. Lee, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, USA, pp. 1646-1654, June, 2016.

표 1. 적용한 방법의 SHM 대비 테스트 시퀀스별 비트율 비교

Method	Dictionary Learning [2]			Y VDSR [4]			Y & Cb & Cr VDSR			Y / Cb & Cr VDSR			Y / Cb / Cr VDSR		
	AI(%)	RA(%)	LP(%)	AI(%)	RA(%)	LP(%)	AI(%)	RA(%)	LP(%)	AI(%)	RA(%)	LP(%)	AI(%)	RA(%)	LP(%)
BQTerrace	-2.25	-0.89	-0.33	-15.39	-7.62	-5.56	-14.67	-7.04	-5.02	-17.56	-8.21	-6.02	-17.57	-8.23	-6.04
BasketballDrive	-1.28	-1.37	-1.19	-8.75	-6.91	-5.39	-7.98	-5.95	-4.86	-11.09	-7.52	-5.87	-11.10	-7.54	-5.88
Cactus	-1.88	-1.04	-0.59	-11.55	-7.14	-4.41	-11.02	-6.54	-4.05	-14.08	-7.87	-4.91	-14.12	-7.68	-4.92
Kimono	-0.58	-0.59	-0.35	-4.12	-3.32	-3.55	-4.02	-3.11	-3.24	-4.63	-4.55	-3.79	-4.67	-4.53	-3.79
ParkScene	-1.26	-1.05	-0.52	-6.07	-3.87	-2.72	-5.98	-3.44	-2.98	-6.2	-6.71	-3.03	-6.35	-6.73	-3.05
AVG	-1.45	-0.99	-0.60	-9.18	-5.63	-4.56	-8.34	-5.22	-4.03	-10.72	-6.97	-4.72	-10.76	-6.94	-4.73