

화면 간 예측에서 인코딩 정보를 고려한 딥러닝 기반 인루프 필터

김양우, 이영렬

세종대학교

ywkim@sju.ac.kr ylee@sejong.ac.kr

Considering Encoding Information for CNN based In-loop Filter in Inter Video Coding

Kim Yang-Woo, Lee Yung-Lyul
Sejong University

요약

VVC (Versatile Video Coding)는 HEVC이후 차세대 표준 비디오 코딩으로 JVET(Joint Video Exploration)에 의해 2018년 표준화를 시작하였다. VVC에는 복원픽처의 변환-양자화에 의해 발생한 블로어, 블로킹, 링잉 아티팩트를 감소시키기 위하여 deblocking filter (DF), sample adaptive offset (SAO), adaptive loop filter(ALF)와 같은 모듈을 사용한다. 한편 CNN (Convolutional Neural Network)은 최근 이미지와 비디오 복원에 높은 성능을 보이고 있다. VVC에서 픽처는 CTU (Coding Tree Unit)으로 분할되고 각 CTU는 다시 CU (Coding Unit)으로 분할된다. 그리고 인코딩을 위한 중요한 정보들이 Picture, CTU, CU단위로 디코더에 전송된다. 이 논문에서는 화면 간 예측으로 인코딩 된 픽처에서 블록과 픽처정보를 이용한 딥러닝 기반의 인루프 필터 모델을 제안한다. 제안하는 모델은 화면 간 예측에서 QP, 4x4 블록단위의 모션벡터, 참조블록과의 시간적거리, CU의 깊이를 모델에 추가적인 정보로 이용한다.

1. 서론

Versatile video Codec (VVC)는 ITU-T VCEG와 ISO/IEC MPEG에 의해 결성된 Joint Video Exploration team (JVET)에 의해 표준화 중인 High Efficiency Video Codec (HEVC) 이후의 차세대 비디오 코덱이다. VVC는 UHD 환경에서 HEVC보다 40%이상 higher Bjøntegaard delta bit-rate (BD-rate) 성능을 보인다. VVC는 블록 단위 코딩, 화면 내 예측, 화면 간 예측, 변환 및 양자화, 인루프 필터, 엔트로피 코딩등의 전통적인 모듈들로 이루어져 있다.

VVC는 영상을 고정된 정사각형 크기의 Coding Tree Unit (CTU)로 분할하고, 각 CTU는 Quad Tree, Binary Tree, Ternary Tree (QTBT)를 통해 최적의 Coding Unit (CU)로 분할된다. 각 CU는 복호화를 위한 중요한 정보를 담은 단위이다. 한편 화면 간 예측은 먼저 코딩된 영상을 참조하여 현재 영상을 예측하는 예측모드이다. 화면 간 예측에서 각각의 CU는 현재블록이 참조블록과 얼마나 시공간적으로 차이가 발생했는지 알 수 있고, 이를 이용하여 예측블록을 생성한다. 예측블록과 원본영상의 차이인 잔차신호를 변환-양자화를 통해 손실압축하고 엔트로피 코딩으로 이진화하여 디코더에 전송한다. 이때 변환-양자화 과정 또는 압축률을 위하여 잔차신호를 아예 전송하지 않는 결정으로 인해 원본영상과 복원영상과의 에러가 발생한다. 이러한 Deblocking, Ringing, Blurring등의 에러를 제거하고 객관적-주관적 화질 향상을 위해, 전통적인 비디오 코덱은 인-루프 필터를 사용한다. VVC의 인-루프 필터는 Deblocking Filter, Sample Adaptive offset, Adaptive Loop Filter가 순차적으로 적용된다.

최근 Convolutional Neural Network (CNN) 기반의 모델들은 이미지 디-노이징, 인페인팅, 슈퍼-레졸루션등에 높은 성능을 보인다. 비

디오 코덱에 CNN 기반 인-루프 필터를 추가하거나, 인-루프 필터 전체를 CNN 기반 모델로 대체하는 연구들이 기존의 전통적인 방법보다 높은 부호화 성능을 보이고 있다.

2. 화면 간 예측 모드 영상을 위한 CNN기반 인-루프 필터

2.1 학습데이터 구축

딥러닝 모델 학습을 위하여 80여개의 비디오 시퀀스를 VVC Test Model (VTM) 8.0를 사용하고, Random Access (RA)로 Quantization Parameter (QP) 22,27,32,37 설정 하에 각각 복-부호화 한다. 딥러닝 모델에 복원픽셀 이외의 추가적인 정보를 부여하기 위하여 QP, 4x4 블록단위의 모션벡터, 참조 블록과 현재 블록간의 시간적거리, CU의 깊이를 픽셀과 같은 크기의 맵으로 만들어서 모델의 중간 입력으로 사용한다. 비디오 코덱에서 일반적으로 사용하는 YUV4:2:0 포맷에 맞추어 휘도성분과 색차성분을 분리하여 입력한다.

2.2 CNN 모델 구조

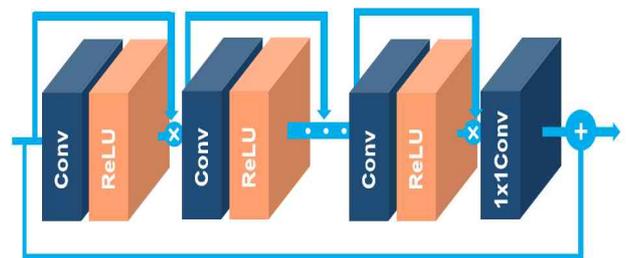


그림 1. Residual Dense Block (RDB) 구조 제안하는 CNN 모델 구조는 인-루프 필터를 거치지 않은 복원영상

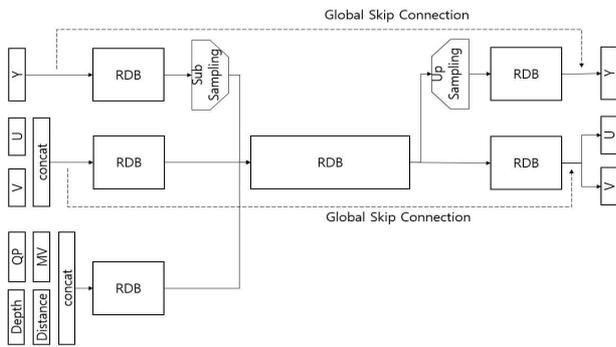


그림 2. CNN 모델 구조

YUV(4:2:0)의 휘도성분과 색차성분을 각각 Residual Dense Block (RDB)로 특징맵을 추출한다. 그리고 QP, 모션벡터, 참조픽처와 시간적 거리, CU의 깊이를 이차원 맵으로 변환하고 RDB를 이용하여 특징맵을 추출한다. 휘도성분의 특징맵을 서브-샘플링 하여 색차성분의 특징맵과 크기를 같게 맞추고 특징맵을 채널 단위로 합쳐 RDB의 입력으로 사용한다. 이후 휘도성분 특징맵과 색차성분 특징맵으로 다시 분할 하여 각각을 RDB를 이용하여 각각 휘도성분 잔차신호와 색차성분 잔차신호로 복원한다. 각 잔차신호를 입력신호와 픽셀단위로 더하여 복원신호를 구성한다.

3. 실험 결과

VVC의 Common Test Condition(CTC)의 1GOP 분량의 Random access 설정에서 2.35%의 성능향상을 보인다.

3. 결론

제안하는 인코딩 정보를 이용한 화면 간 예측에서 딥러닝 기반 인-루프필터를 이용하여 VVC 8.0 대기 2.35%의 부호화 성능향상을 보인다.

감사의 글

이 논문의 일부는 2020년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. 20200348)

참고문헌

- [1] B. Bross, W.-J. Han, G. J. Sullivan, J.-R. Ohm, T. Wiegand, "High efficiency video coding (HEVC) text specification draft 7", document JCTVC-I1003, Jul. 2012 K. D. Hong and K. J. Lim, "A study on image understanding," IEEE Trans. Image Processing, vol. 3, no. 2, pp. 1-10, 2007.
- [2] B. Bross, W.-J. Han, G. J. Sullivan, J.-R. Ohm, T. Wiegand, "High efficiency video coding (HEVC) text specification draft 7", document JCTVC-I1003, Jul. 2012 K. D. Hong and K. J. Lim, "A study on image understanding," IEEE Trans. Image Processing, vol. 3, no. 2, pp. 1-10, 2007.
- [3] Zhang, K., Zuo, W., Chen, Y., et al.: 'Beyond a Gaussian denoiser: residual learning of deep Cnn for image denoising', IEEE Trans. Image Process., 2017, 26, (7), pp. 3142 - 3155.