SRCNN 을 이용한 HEVC 화면 내 예측 부호화

김남욱
세종대학교
nukim@sju.ac.kr

강정원
전자통신연구원
jungwon@etri.re.kr

이영렬
세종대학교
yllee@sejong.ac.kr
교신저자

HEVC Intra prediction using SRCNN

Nam Uk Kim
Sejong
University

Jung, Won Kang
ETRI

*Yung Lyul Lee
*Sejong
University

요 약

본 논문에서는 최신의 비디오 코덱 표준인 HEVC(High Efficiency Video Coding)의 화면 내 예측 부호화의 성능 향상을 위하여 SRCNN(Super Resolution Convolutional Neural Networks)을 이용하는 방법을 제안한다. SRCNN 은 비교적 최신 기술인 CNN(Convolutional Neural Network)을 사용하여 이미지를 추가적인 데이터 없이 보간 하여 해상도를 증가시키는 기술이다. HEVC 에서는 화면 내 예측의 잔차신호를 부호화 하기 위해 많은 비트를 소모하는데, 본 논문에서는 이 잔차신호들의 해상도를 낮추어 부호화 되는 비트를 줄이며, 복호화기에서 SRCNN 을 이용하여 원래의 해상도로 복원을 수행하여 압축성능을 향상 시키는 방법에 대하여 제안한다. 제안하는 기술은 HM 16.6 에 구현하였으며, CNN 트레이닝에 Caffe 라이브러리를 사용하였다.

1. 서론

최근 신경망 네트워크(Neural Networks)를 활용한 기술이 영상, 의학, 언어, 게임 인공지능 등 다양한 분야에서 활발히 연구되고 있다. 영상관련 분야에서는 다계층 신경망 네트워크 중 하나인 CNN(Convolutional Neural Networks)이 다른 종류의 네트워크 구조보다 높은 성능을 보이고 있으며, 많은 연구가 진행 중 이다. CNN 은 하나 또는 여러 개의 합성곱(Convolution) 계층과 그 위에 올려진 일반적인 신경망 계층들로 구성된다. SRCNN(Super Resolution Convolutional Neural Network)[1][2]은 CNN 구조를 이용하여 bicubic[3] 이나 lanczos[4]와 같이 보간된 픽셀을 생성하여 이미지의 해상도를 증가시키는 기술이다. SRCNN 은 기존의 휴리스틱 보간 알고리즘들과는 많은 차이를 보이며 보다 높은 품질의 보간 이미지를 생성한다. HEVC[5]는 정지 영상을 부호화 하기 위하여 화면 내 예측 부호화를 사용한다. 화면 내 예측 부호화는 영상의 공간적인 중복만을 이용하여 예측 신호를 생성하여 부호화 하는 기술로 대부분의 비트를 예측후에 발생한 잔차신호를 부호화 하기 위하여 사용한다. 본 논문에서는 잔차신호 부호화의 오버헤드를 줄이기 위하여 SRCNN 을 활용하는 방법을 제안한다. 본 논문의 2 장에서는 SRCNN 에 대하여 설명한다. 3 장에서는 트레이닝 방법에 대하여 설명하며, 마지막 4 장에서는 제안하는 방법과 실험 결과를 보이며 본 논문을 마친다.

2. SRCNN

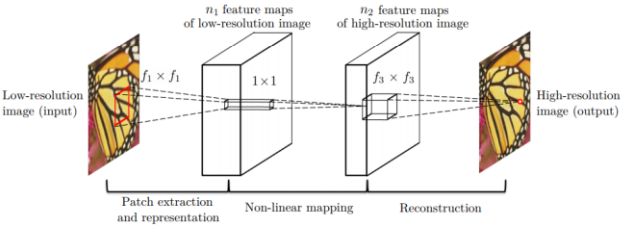

그림 1. SRCNN 모델

SRCNN 은 그림 1 과 같이 단순한 3 콘볼루션 계층을 갖는다. 각각의 콘볼루션 계층은 ReLU 를 Activation function 으로 사용한다. $f_1 \times f_1$ 크기의 입력은 bicubic interpolation 을 통해 미리 확대를 하기 때문에 디콘볼루션[6]이나 언풀링 과 같은 레이어를 통해 중간에 맵의 크기를 확대할 필요가 없다. Loss function 으로는 일반적으로 사용하는 MSE(Mean Square Error)를 사용한다. SRCNN 은 비교적 작은 크기의 레이어와 적은 파라메터들을 사용하기 때문에 한번의 앞 전파(forward propagation)를 수행할 때 신경망 알고리즘 치고는 적은 약 50 만번 혹은 300 만번의 곱셈 연산과 가감연산만을 수행한다(입력 단위의 크기에 따라

다름).

3. HEVC 를 위한 SRCNN 트레이닝

기존 SRCNN 은 21x21 크기의 입력만 사용을 하며, 3 배의 확대만을 지원하기 때문에 네트워크 모델을 약간 수정을 할 필요가 있다. HEVC 에서는 모든 블록이 2 의 지수승의 크기를 갖기 때문에 SRCNN 의 입력 또한 2 의 지수 승으로 맞추어야 하며, 확대의 배율도 2 배 혹은 4 배 중 하나로 정하여 SRCNN 을 새로이 트레이닝을 수행하여야 한다.

HEVC 에 적용하기 위한 새로운 SRCNN 모델은 입력은 항상 8x8 크기로 고정하였으며, 출력도 8x8 크기로 설정하였다. 16x16 이나 32x32 과 같은 더 큰 크기의 입력을 줄 경우에는 네트워크가 잘 수렴하지 않았다. 이미지의 확대 배율도 4 배로 할 경우 잘 트레이닝이 되지 않아 2 배로 설정하였다. 트레이닝에 사용된 이미지는 본 저자가 카메라로 직접 촬영한 무손실 사진을 Full-HD (1920x1080)해상도급으로 다운스케일링 하여 사용하였다.

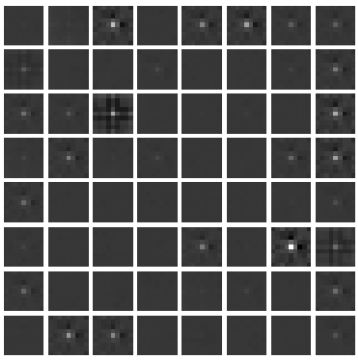

그림 2. 첫번째 커널

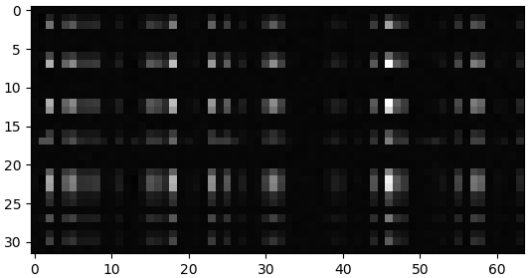

그림 3. 두번째 커널

그림 2~4 는 8x8 샘플들로 4 천만번 트레이닝 후 얻어낸 콘볼루션 계층의 커널들이다. 이후 HEVC 에서 적용한 SRCNN 은 해당 모델을 사용하였다.

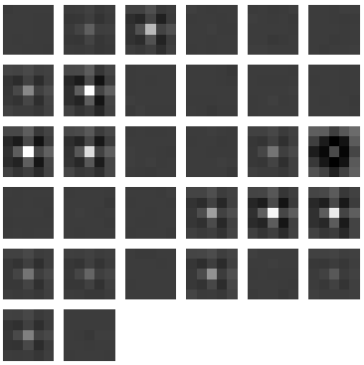

그림 4. 세번째 커널

4. 제안하는 방법 및 실험 결과

제안하는 방법은 다음과 같다. HEVC 의 화면 내 예측 과정 중 예측 후 잔차신호를 변환하기전에 추가적인 플래그에 따라 잔차신호의 해상도를 1/4 로 줄이거나 줄이지 않는다. 만약 해상도를 줄이지 않는다면 기존 HEVC 와 동일한 과정을 수행하게 된다. 만약 해상도가 1/4 로 줄게 된다면 1/4 크기의 잔차신호를 그대로 변환 양자화를 수행한 후 엔트로피 부호화 복호화 그리고 역양자화 역변환을 통해 1/4 크기의 복원 신호를 생성하게 된다. 복원 신호는 bicubic interpolation 을 통해 원래크기로 복원되며, 원본과의 오차를 줄이기 위해 SRCNN 을 통해 복원 신호의 화질을 개선한다.

제안하는 방법을 HEVC 에 구현하기 위해서는 컨벌루션 계층과 ReLU(Rectified Linear Unit)가 정수연산으로 구현 되어야 한다. 현실적으로 이러한 구현을 C++ 코드로 직접 구현하기 위해서는 오랜 시간이 소요되기 때문에 본 논문에서는 우선 직접적으로 구현하지 않고, 코스트 계산을 통해 간접적으로 제안하는 방법의 성능을 예측하였다.

구체적인 성능 계산 과정은 다음과 같다. 우선 HEVC 부호화기로 영상을 부호화 하고 각각의 TB(Transform Block)으로부터 3 가지 정보인 복원 블록, 1/4 크기의 복원블록 그리고 비트량 차이를 추출하였다. 복원 블록은 TB 의 양자화 계수를 가지고 복원한 블록이다. 1/4 크기의 복원 블록은 TB 의 잔차신호를 1/4 로 다운샘플링 후 변환 양자화한 계수로 복원한 블록이다. 마지막 비트량 차이는 TB 의 양자화 계수를 부호화하기 위한 비트량과 1/4 로 다운샘플링 후 변환 양자화한 계수를 부호화하기 위한 비트량과의 차이이다. 모든 TB 에 대해 이 3 가지 데이터를 수집한 후 식 (1)과 같이 비용 차이를 계산하였다.

$$\Delta Cost = \lambda * \Delta BitCost + \Delta Distortion \quad (1)$$

ΔBitCost 위 HEVC 부호화기로부터 얻어낸 비트량 차이에 해당한다. ΔDistortion HEVC 로 복원한 블록과 원본 영상과의 SAD(Sum of Absolute Difference)와 1/4 로 다운샘플링된 복원블록을 SRCNN 을 이용해 보간한 블록과 원본 영상과의 SAD 를 뺀 수치이다. λ는 QP(Quantization Parameter)에 따라 바뀌는 값으로 1 비트 코스트의 가치를 대략 λ 만큼의 Distortion 의 가치와 동일하게 보고 비용을 계산하기

위해 사용된다. ΔCost가 음수이고 더 작은 값일 수록 SRCNN 이 해당 블록을 코딩하는데 유리하다고 예측할 수 있다. 아래의 표는 1920x1080 크기의 Cactus 시퀀스 영상을 사용하여 SRCNN 과 성능을 비교한 결과를 나타낸다.

E(ΔCost)	MSE1	MSE2	E(SAD)1	E(SAD)2
−239.17	48.16	11.71	112.81	110.48

MSE1 은 SRCNN 복원신호의 MSE, MSE2 는 기존 HEVC 복원신호의 MSE 를 의미한다 (SAD 도 동일한 순서이다). 예상외로 ΔCost 측면에서는 SRCNN 이 많은 경우 좋은 결과를 보였다. ΔCost의 평균이 음수이며 작은 값이기 때문에 HEVC 화면내 예측 부호화의 RDO 과정에서는 대부분 SRCNN 을 사용하여 복원된 블록이 선택될 것이다. 언뜻 보기엔 SRCNN 이 높은 성능을 나타낼 것으로 예상될 수 있었지만 MSE 가 예상외로 SRCNN 으로 복원된 신호가 HEVC 로 복원된 신호보다 높은 수치를 보였기 때문에 결국 PSNR(Peak Signal to Noise Ratio)이 많이 감소할 것으로 예상된다. CU(Coding Unit)혹은 TB(Transform Block) 당 플래그를 전송하여 SRCNN 사용여부를 결정할 경우 SRCNN 이 잘 동작하는 경우도 있기 때문에 약간의 BD-rate 향상을 기대할 수 있을 것으로 보인다.

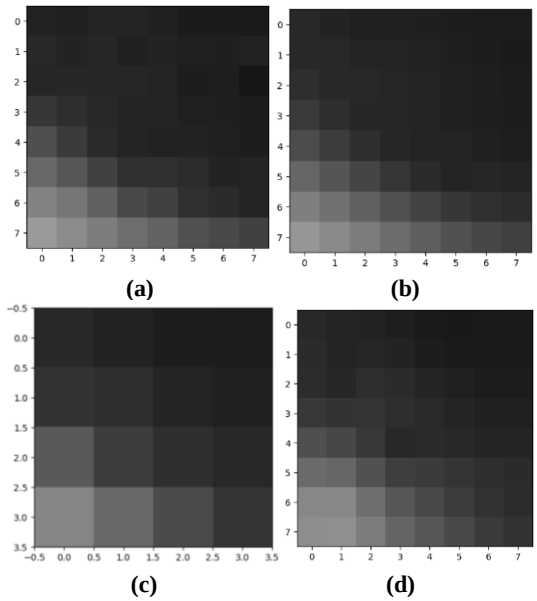

그림 5. Cactus 영상의 일부
(a)원본블록, (b)HEVC 복원블록,
(c)HEVC 복원블록을 1/4 로 다운샘플링한 블록
(d) SRCNN 복원블록

차후 연구에서는 단순히 SRCNN 을 HEVC 에 바로 적용하는 것은 비효율 적이기 때문에, HEVC 에 맞는 신경망 모델을 수립하는 방법에 대해 연구 주제로 가질 예정이다. 신경망의 입력으로 화면 내 예측신호와 다운 샘플링된 잔차신호를 입력받고 신경망을 거쳐 원래 해상도의 잔차신호를 예측하는 방법으로 구현한다면, 압축률을 많이 올릴 수 있을 것으로 기대된다.

감사의 글

이 논문은 일부 2017 년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No. 2016-0-00572, 초고실감 미디어 서비스 실현을 위해 HEVC/3DA 대비 2 배 압축을 제공하는 5 세대 비디오/오디오 표준 핵심 기술 개발 및 표준화)

참고 문헌

[1] Chao Dong, Chen Change Loy, Kaiming He, Xiaoou Tang. Learning a Deep Convolutional Network for Image Super-Resolution, in Proceedings of European Conference on Computer Vision (ECCV), 2014

[2] Chao Dong, Chen Change Loy, Kaiming He, Xiaoou Tang. Image Super-Resolution Using Deep Convolutional Networks, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), Preprint, 2015

[3] R. Keys (1981). "Cubic convolution interpolation for digital image processing". IEEE Transactions on Acoustics, Speech, and Signal Processing. 29 (6): 1153-1160

[4] Ken Turkowski and Steve Gabriel (1990). "Filters for Common Resampling Tasks". In Andrew S. Glassner. Graphics Gems I. Academic Press. pp. 147-165.

[5] G. J. Sullivan; J.-R. Ohm; W.-J. Han; T. Wiegand (December 2012). "Overview of the High Efficiency Video Coding (HEVC) Standard". IEEE Transactions on Circuits and Systems for Video Technology. IEEE. 22 (12)

[6] O'Haver T. "Intro to Signal Processing Deconvolution". University of Maryland at College Park