

합성곱 신경망 기반의 화면 내 예측 모드 결정

정지연, 이영렬
세종대학교

jyung@sju.ac.kr, yllee@sejong.ac.kr

Decision of the Intra Prediction mode using Convolutional Neural Network

JiYeon Jung, Yung-Lyul Lee
Sejong University

요 약

본 논문에서는 최신 비디오 압축 표준인 HEVC 를 기반으로, 영상의 화면 내 예측 모드를 결정하기 위하여 기존의 RDO 프로세스 대신 합성곱 신경망을 접목시켜 예측 모드를 결정하는 연구를 진행하였다. 네트워크 학습에는 실제 화면 내 예측 모드를 결정하는 데 사용되는 데이터들이 사용되었다. 실험 결과 본 제안 기법은 HM 16.20 대비 0.05%의 부호화 성능 향상을 보였다.

1. 서론

HEVC(High Efficiency Video Coding) [1]는 ITU-T 의 VCEG(Video Coding Experts Group)과 ISO/IEC 의 MPEG(Moving Picture Expert Group)의 비디오 전문가들에 의해 진행된 비디오 압축 기술 표준이다. 2010 년 4 월 독일 드레스덴 논의를 시작으로 표준화가 진행되었고, 2013 년도에 표준화가 완료되었다.

HEVC 는 압축 효율 향상을 위해 CU(Coding Unit), PU(Prediction Unit), TU(Transform Unit) [2]을 포함하는 계층적 블록 구조를 갖는다. 이 구조에 따라 예측, 변환, 양자화, 엔트로피 부호화 순서로 압축 과정이 이루어진다. 예측은 이미 복원된 주변 블록으로부터 현재 블록을 예측하는 과정으로 화면 내 예측, 화면 간 예측 두 가지 방법이 존재한다. 두 가지 예측 방법 중 하나를 선택해 예측 신호를 생성한 후 원본 신호로부터 빼 주어 잔차 신호를 생성한다. 이후 변환을 통해 주파수 영역에서 압축을 수행하며 변환 과정에서 출력되는 변환 계수들은 양자화되어 엔트로피 부호화를 통해 비트스트림으로 변환된다. 해당 비트스트림은 복호화기에 전송한다.

화면 내 예측 시, 블록 간의 공간적 상관관계 및 블록의 방향적 특성을 사용하여 예측 블록이 생성된다. HEVC 에서는 그림 1 [1]과 같이 Planar 모드, DC 모드 등을 포함한 총 35 개의 예측 모드를 사용한다. 각 블록의 특성과 Bit 및 Distortion 간의 trade-off 에 따라 적절한 예측 모드가 결정이 되며 이 모드를 통해 예측 블록을 생성하게 된다. 이때 예측 모드를 결정하기 위해 Rough-Mode Decision (RMD) 및 Rate-Distortion Optimization (RDO) 등의 과정이 수행된다. RMD 과정에서는 총 35 개의 모드에 대하여 Hadamard Transform 을 적용하여 N 개의 후보 모드를 추출해낸 뒤, RDO 과정을 통해 해당 N 개의 모드에 대한 RD cost 를 구하여 가장 작은 cost 값을 가진 모드를 선택하게 된다. 이때 N 의 값은 작은 블록 (8x8 보다 작은 블록)은 8 개, 큰 블록 (16x16 보다 큰 블록)은 3 개로 구성된다.

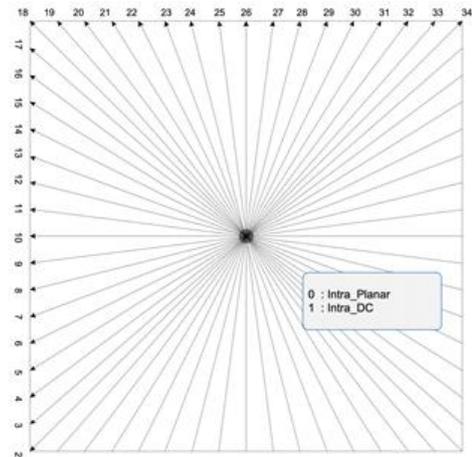


그림 1. HEVC에서의 화면 내 예측 모드 [1]

본 논문에서는 최근 활발히 연구되고 있는 딥러닝을 이용하여 화면 내 예측 모드를 결정하는 방법을 제안한다. 네트워크 학습에 사용되는 데이터는 실제 화면 내 예측 모드를 결정하는 데 사용되는 데이터로 구성되어 있으며, 네트워크는 합성곱 신경망 네트워크 구조를 사용하였다.

본 논문의 구성은 다음과 같다. 2 절에서는 제안하는 기법에 대해 살펴본 후, 3 절에서는 실험 결과 및 분석, 마지막으로 4 절에서는 본 논문에 대한 결론을 맺는다.

2. 제안하는 방법

기존의 HEVC 에서의 화면 내 예측 후보 모드는 블록 크기에 따라 달라진다. 블록의 크기가 작을 수록 복잡한 텍스처를 가진 경우가 많기 때문에 8 개의 후보 모드를

사용하게 되고, 블록의 크기가 클수록 텍스처가 단순한 특성이 있기 때문에 8 개보다 더 적은 3 개의 후보 모드를 사용하게 된다. 후보 모드에 대한 RD cost 를 구하는 RDO 과정에서의 부호화 시간 소요가 큰 비중을 차지하는데, 큰 블록의 경우 비교적 결정되는 예측 모드의 분포가 넓지 않다는 특성을 가지고 있다. 즉, 큰 블록에 대해서는 기존의 복잡한 RDO 과정을 수행하는 것이 비효율적일 수 있다. 따라서 본 논문에서는 큰 블록에 대하여 기존의 RDO 과정이 아닌 네트워크 학습을 통해 빠르게 예측 모드를 결정하는 방법을 제안한다.

2.1. 네트워크 구조

입력 데이터는 기존의 HEVC 에서의 예측 모드를 결정하기 위해 수행되는 RMD 과정에서 생성되는 데이터를 사용한다. RMD 과정을 통해 큰 블록의 경우 총 3 개의 후보 모드와 3 개의 Most Probable Mode(MPM)에 대한 distortion 값과 해당 모드를 복호화기에 전송하기 위해 필요한 bit 값 등이 계산되는데, 이 데이터를 입력 데이터로 사용하였다. 이때, 각 모드에 따른 데이터의 특성이 다르기 때문에 병렬적으로 컨볼루션을 수행한다. 네트워크는 그림 2 와 같이 총 3 개의 컨볼루션 층과 2 개의 완전연결 층을 사용하였으며, 각각의 컨볼루션 층은 8 개, 16 개, 16 개의 특징 맵으로 구성되어 있다. 두 번째 컨볼루션 후, 모든 데이터를 연결한 후에 세 번째 컨볼루션이 진행된다. 세 번째 컨볼루션은 커널의 크기가 1x1 으로, 연산량과 복잡도를 줄이기 위해 사용되는 컨볼루션 층이다. 컨볼루션 후 총 두 번의 완전연결 층을 통해 RMD 후보 모드 개수와 동일한 6 개의 클래스로 분류된다.

2.2. 학습 방법

학습에 사용된 영상은 기존의 HEVC 의 공통 실험 조건 [3]에 사용되는 영상 데이터 중 B, C, D, E, F 클래스와 임의의 이미지 400 여장을 사용하였다. 이때, 학습에 사용된 총 데이터의 개수는 300,000 개, 검증 및 테스트에 사용된 데이터의 개수는 37,500 개이다. 네트워크는 기존의 HEVC 를 통해 결정된 실제 예측 모드와 네트워크의 학습을 통해 결정된 예측 모드 값을 비교하여 그 확률 값을 손실 비용으로 사용하는 cross-entropy 함수를 사용하였다. 손실 함수는 식

(1)과 같다.

$$\text{Loss} = - \sum_{i=0}^5 y_i' \cdot \log(y_i) \quad \text{-- 식 (1)}$$

y_i' 은 실제 예측 모드이며, y 는 네트워크를 통해 분류된 각 클래스의 softmax 값이다.

초기 학습률은 10^{-3} 로 설정하였으며, 그 뒤의 학습률 조정은 Adam Optimizer 를 사용하여 학습하였다. 또한 120 의 배치 사이즈와 총 20 번의 epoch 을 통해 학습이 진행되었다. 네트워크 학습을 위해 PyTorch API 를 사용하였다.

3. 실험 결과 및 분석

본 논문의 실험에 사용된 참조 소프트웨어는 HM 16.20 버전이며, JCT-VC CfP(Call for Proposal) 단계에서 결정된 공통 실험 조건에 맞추어 진행하였다. 실험은 All Intra (AI) configurations 에서 진행하였고, A, B, C, D, E, F 클래스들을 사용하여 총 24 개의 시퀀스들을 사용하였으며, 각 시퀀스마다 22, 27, 32, 37 의 QP 로 설정하여 실험 결과 및 분석을 진행하였다. 각각의 실험 결과를 BD-rate 로 표현하였다.

표 1 은 본 논문에서 제안하는 화면 내 예측 모드 결정 방법을 적용한 결과이다. 표 1 과 2 에서 HM 16.20 대비 16x16 크기의 블록은 0.0%, 32x32 크기의 블록은 0.1%의 부호화 성능 향상을 보였다.

표 1. 제안된 알고리즘의 성능분석 (16x16)

Class	BD-rate (16x16)		
	Y	U	V
A	0.5%	0.9%	1.0%
B	0.3%	0.6%	0.6%
C	-0.2%	0.3%	0.7%
D	-0.3%	0.2%	0.6%
E	0.6%	1.6%	2.3%
F	-0.7%	0.0%	0.7%
Overall	0.0%	0.6%	0.7%

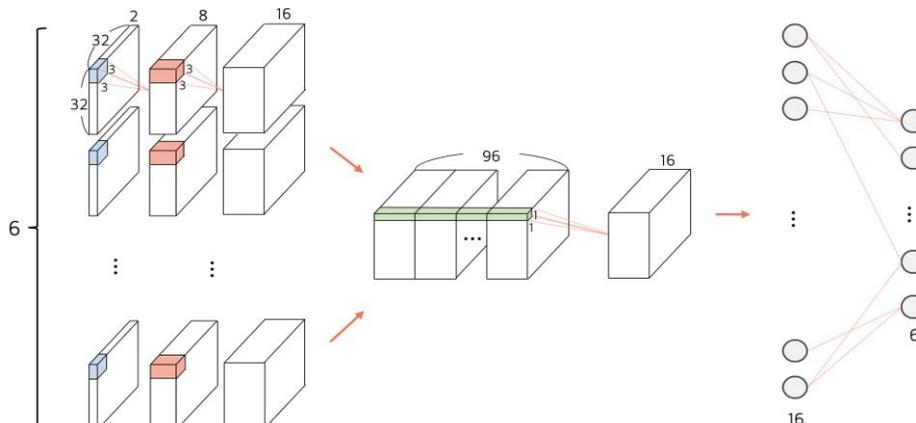


그림 2. 제안하는 네트워크 구조

표 2. 제안된 알고리즘의 성능분석 (32x32)

Class	BD-rate (32x32)		
	Y	U	V
A	0.1%	0.1%	0.2%
B	0.4%	0.7%	0.5%
C	-0.5%	-0.3%	-0.5%
D	-0.5%	-0.6%	-0.4%
E	-0.1%	0.4%	0.5%
F	0.0%	0.0%	0.3%
Overall	-0.1%	0.0%	0.1%

4. 결론

본 논문에서는 최신 비디오 압축 표준인 HEVC 를 기반으로, 영상의 부호화 과정에서 수행되는 화면 내 예측 모드를 결정하기 위하여 기존의 방법 대신 합성곱 신경망을 접목시켜 예측 모드를 결정하는 연구를 진행하였다. 네트워크 학습에는 실제 화면 내 예측 모드를 결정하는 데 사용되는 데이터들이 사용되었으며, 학습에 사용된 데이터들은 기존의 HEVC 공통 실험 조건에 사용되는 영상들을 사용하여 학습하였다. 실험 결과 본 제안 기법은 0.05%의 부호화 성능 향상을 보였다. 본 논문의 제안 방법을 토대로 블록의 특성 및 블록 간의 상관관계 등을 고려하여 부호화 성능 향상 및 화질 개선 등의 추가적인 방법에 대한 연구가 필요하다고 보여진다.

5. 감사의 글

이 논문의 일부는 2019 년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. 20190173)

6. 참고문헌

- [1] B. Bross, W.-J. Han, J. -R. Ohm, G. J. Sullivan, Y.-K. Wang, and T. Wiegand, "High Efficiency Video Coding (HEVC) Text Specification Draft 10 (for FDIS & Consent)", JCT-VC document JCTVC-L1003, vol. 1, 12thmeeting, Geneva, Switzerland, 2013.
- [2] G.J. Sullivan and R.L. Baker, "Efficient Quadtree Coding of Images and Video", IEEE Trans. Image Process, Vol. 3, no.3, pp. 327-331, May 1994.
- [3] JCT-VC document, JCTVC-L1100, Common HM Test conditions and software reference configurations, Geneva, Switzerland, Jan. 2013.