

## 신경망 기반 원본영상에서 화면 내 예측 모드로 변환

\*김양우 \*\*이영렬

세종대학교

\*ywkim@sju.ac.kr \*yllee@sejong.ac.kr

### Neural Network based Pixel to Intra Prediction Mode Decision

\*Kim, Yangwoo \*\*Lee, Yung-Lyul

Sejong University

#### 요약

VVC(Versatile Video Codec)의 화면 내 예측은 인코더에서 영상을 적절하게 사각형 블록으로 분할하고, 블록 주변의 먼저 재구성된 참조샘플들을 이용하여 예측블록을 형성한다. 인코더는 화면 내 예측 모드에서 각 PU(Prediction Unit)에 대하여 MIP(Matrix-based weighted Intra Prediction) 적용 여부, MIP에서 matrix의 인덱스, MRL(Multi Reference Line)의 인덱스, DC/Planar/Angular 모드에 대한 최적모드를 고려하여 각 정보를 디코더로 전송하며 각 후보모드들의 압축효율을 비교하는 과정에서 높은 연산량을 요구한다. 본 논문에서는 이러한 모드 결정은 원본영상으로도 대략적인 결정이 가능하다는 전제를 가지고 NN(Neural Network)의 일종인 CNN(Convolutional Neural Network)를 이용하여 복잡한 모드 결정 방법을 생략하는 방법을 제안한다.

#### 1. 서론

VVC는 차세대 비디오 표준으로서 ISO/IEC MPEG(Moving Picture Experts Group)과 ITU-T VCEG(Video Coding Experts Groups)이 JVET(Joint Video Exploration)을 2015년 10월에 결성, 2018년 HEVC(High Efficiency Coding)[1]이후의 비디오 표준화를 목표로 개발되었다.

VVC 인코더에서 영상을 적절하게 QT(Quad Tree)/BT(Binary Tree)/TT(Ternary Tree)로 재귀적으로 분할하고, 분할된 CU(Coding Unit)는 시간적/공간적으로 이미 재구성된 CU의 화소들을 참조샘플로 사용하여 예측블록을 생성하여 원본영상과 예측블록과의 차이를 변환-양자화하여 전송함으로써 압축률을 극대화 한다. 이 일련의 과정 중 공간적으로 재 구성된 주변화소들을 이용하여 현재 CU의 예측블록을 생성하는 기술을 ‘화면 내 예측(Intra Prediction)’이라 한다.

VVC 인코더는 가장 효율적인 예측블록을 도출하기 위하여 MIP, MRL, DC/Planar/Angular 등의 여러 가지 모드를 이용하여 예측블록을 생성하며 각 모드를 선택했을때의 비트율을 고려한 Rd-Cost를 이용하여 최적의 모드를 결정하여 해당 모드 정보를 디코더에 전송한다. 이때 각 모드의 확률적인 빈도수를 고려하여, MIP, Planar, MPM(Most Probability Mode), 나머지 모드 순으로 해당 모드가 아니면 다음 모드 인지 검사하기 때문에 같은 입력영상을 인코딩 하더라도 DC, Planar 등의 모드들은 QP(Quantization Parameter)로 지칭되는 요구압축률이 높으면 비교적 더 많이 선택된다.

VVC의 화면 내 예측 모드 중 MIP는 이미 지정된 matrix들을 이용하여 참조샘플과 matrix간의 행렬 곱으로 예측블록을 생성하며, 몇 번째의 matrix를 사용 할 지는 MIP가 선택 될 경우 추가 비트전송을 통해 결정 할 수 있다. 이때 matrix 종류의 개수는 같이 블록 크기에 따라 달라지며 최대 16개의 인덱스를 가진다.

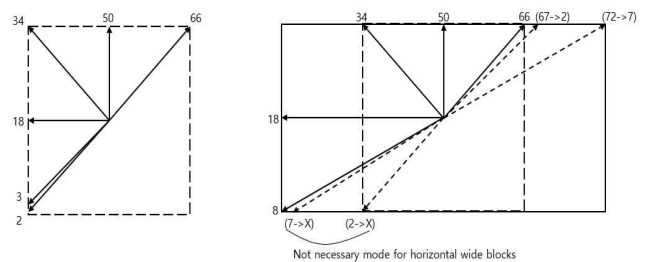


그림 1. 정사각형 블록에서 Angular 모드 범위(왼쪽)와 가로가 세로와 같은 두배인 블록에서의 Angular 모드 범위(오른쪽)

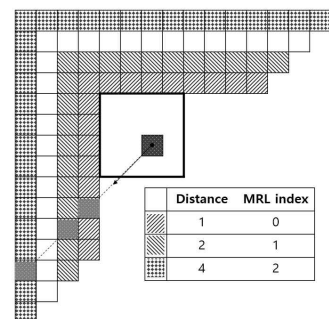


그림 2. MRL에서 참조샘플 거리 별 MRL index의 예시

#### 2. 신경망 기반 원본영상에서 화면 내 예측 모드로 변환

##### 2.1. VVC의 화면 내 예측

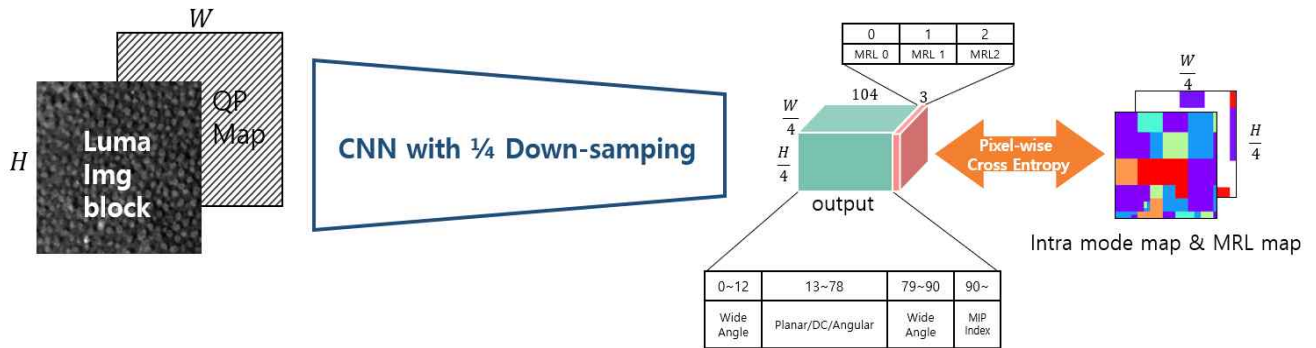


그림 3. 신경망 네트워크의 입력과 출력

전통적인 화면 내 예측 기법인 DC, Planar, Angular 모드를 이용한 예측블록 생성은 VVC에서도 유효하며, 총 67개의 모드를 가진다. VVC에서는 가로와 세로의 크기가 다른 화면 내 예측 블록이 존재하며, 이러한 블록의 Angular 모드에 대하여 WAIP(Wide-Angle Intra Prediction)를 사용한다. 그림1과 같이 블록의 짧은변쪽의 Angular 모드의 일부를 긴변의 모드로 간주하고 사용할 수 있다. VVC의 Main Profile상에서 긴 변이 짧은 변보다 8배를 초과하여 길어질 수 없기 때문에, DC/Planar/Angular 모드는 [-12, 78] 범위의 모드를 가진다.

MRL은 그림 2와 같이 가장 인접한(1-distance) 참조샘플이 아닌 2-distance, 4-distance의 참조샘플을 사용하는 방법이다. DC/Planar/Angular 모드들은 별도의 비트전송을 통해 몇 번째 참조샘플을 사용할지 결정한다.

## 2.2. 신경망 구조 및 입출력

신경망 네트워크 구조에서 이미지 분류에 해당하는 문제는 CNN으로 이미지의 특징을 추출한 후에 FCN(Fully Connected Network)를 통한 정답과의 Cross-Entropy를 손실함수로 사용하여 해결한다. 하지만 이러한 블록 단위 접근방식[2][3]은 하나의 블록의 모드를 판단하는데 네트워크를 한번 통과하여야 하며, 블록의 크기마다 네트워크에 의존성이 생기는 문제점이 있다.

본 논문에서는 입력으로 현재 코딩하려는 원본영상 전체를 네트워크 입력으로 하고, VVC의 휘도영상 화면 내 예측에서 4픽셀 단위의 블록마다 예측모드는 반드시 같으므로 2번의 Down-Sampling을 통하여 네트워크의 복잡도를 감소시켰다.

네트워크 출력은 입력 영상보다 가로, 세로의 크기가 1/4만큼 작은 107개의 채널이며, 104개의 채널은 화면 내 예측의 모드 결정을 위하여 사용되고, 3개의 채널은 MRL 인덱스를 결정하기 위해 사용된다. 104개의 채널과 3개의 채널은 각각의 Cross-Entropy 손실함수에 의해서 학습되며 두 개의 손실은 0.75과 0.25의 가중치를 가진다.

## 3. 실험결과

본 논문의 방법으로 VVC 인코더에서 4K영상 1장을 화면 내 예측 모드를 도출하기 위해서 RTX2080ti GPU 1개를 사용하여 평균 15.6초가 소요되었으며 전체 예측블록들에 대한 모드 예측 정확도는 45.6%이다..

## 4. 결론

본 논문은 VVC 화면 내 예측 모드를 신경망 구조로 예측하기 위하여 원본영상만을 통한 화면 내 예측 모드 결정을 제안한다. 해당 방법은 기존의 블록단위 신경망 예측 방법에 비하여 GPU 통신 횟수를 대폭 줄일 수 있다.

## 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] Rui Song, Dong Liu, Houqiang Li, and Feng Wu, "Neural network-based arithmetic coding of intra prediction modes in HEVC," in VCIP, 2017, pp. 1-4.
- [3] H. Ting, H. Fang and J. Wang, "Complexity Reduction on HEVC Intra Mode Decision with modified LeNet-5," 2019 IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS), Hsinchu, Taiwan, 2019, pp. 20-24, doi: 10.1109/AICAS.2019.8771586.