

VVC 의 다중 변환 선택을 고려한 신경망 기반 화면내 예측

박도현, 문기화, *임성창, 김재곤
 한국항공대학교, *한국전자통신연구원
 {dhpark, ghmoon}@kau.kr, *sclim@etri.re.kr, jgkim@kau.ac.kr

Neural Network-Based Intra Prediction Considering Multiple Transform Selection in Versatile Video Coding

Dohyeon Park, Gihwa Moon, *Sung-Chang Lim, and Jae-Gon Kim
 Korea Aerospace University, *ETRI

요 약

최근 VVC(Versatile Video Coding) 표준 완료 이후 JVET(Joint Video Experts Team)에서는 NNVC(Neural Network-based Video Coding) EE(Exploration Experiment)를 통하여 화면내 예측을 포함한 신경망 기반의 부호화 기술들을 탐색하고 검증하고 있다. 본 논문에서는 VVC 에 채택되어 있는 다중 변환 선택(MTS: Multiple Transform Selection)에 따라서 적절한 예측 블록을 선택할 수 있는 TDIP(Transform-Dependent Intra Prediction) 모델을 제안한다. 실험결과 제안기법은 VVC 의 AI(All Intra) 부호화 환경에서 VTM(VVC Test Model) 대비 Y, U, V 에 각각 0.87%, 0.87%, 0.99%의 BD-rate 절감의 비디오 부호화 성능 향상을 보였다.

1. 서론

최근 방송, 모바일, 스트리밍 기반의 다양한 미디어 서비스의 발전과 함께 UHD(Ultra High Definition), AR(Augmented Reality), VR(Virtual Reality)와 같은 초고용량의 동영상에 대한 수요가 급속도로 증가하고 있다. 이에 따라, MPEG(Moving Picture Experts Group)과 VCEG(Video Coding Experts Group)은 공동으로 JVET(Joint Video Experts Team)을 구성하여 HEVC(High Efficiency Video Coding) 대비 월등히 개선된 압축 성능을 갖는 VVC(Versatile Video Coding) 표준화를 2020 년 7 월 완료하였다[1].

JVET 은 잔차신호(Residual Signal)에 DCT(Discrete Cosine Transform), DST(Discrete Sine Transform) 기반의 다양한 변환 커널을 선택하여 사용하는 다중 변환 선택(MTS: Multi Transform Selection)과 같은 새로운 변환 기술과 MIP(Matrix-weighted Intra Prediction), LFNST(Low Frequency Non-

Separable Transform)과 같이 학습된 신경망 모델을 이용한 부호화 기술들을 VVC 비디오 코덱의 부호화 성능 향상을 위해 채택하였다. 또한, VVC 표준 완료 이후 신경망 기반의 비디오 부호화 기술에 대한 NNVC(Neural Network-based Video Coding) EE(Exploration Experiment)를 두고 관련 기술을 활발하게 탐색하고 검증하고 있다. 특히, 영상의 공간적 특성 및 참조샘플의 부족으로 인해 DC, Planar 및 방향성 모드와 같은 일반적인 화면내 예측의 부호화 성능 향상은 한계에 다다랐으며, 이를 극복할 수 있는 딥러닝 기반의 화면내 예측 기술의 중요성이 부각되고 있다. 그러나 종래의 신경망 기반의 화면내 예측 모델 및 그 모델을 학습함에 있어서는 비디오 코덱의 변환 기반 부호화를 반영하지 않거나 반영을 하더라도 하다마드(Hadamard) 변환 또는 DCT2 변환만을 반영하기 때문에 VVC 에 채택되어 있는 MTS 에 적합하지 않다.

따라서, 본 논문에서는 NNVC EE 에서 검증된 SAIP(Single Additional Intra Prediction)[2, 3]의 신경망 모델을 기반으로

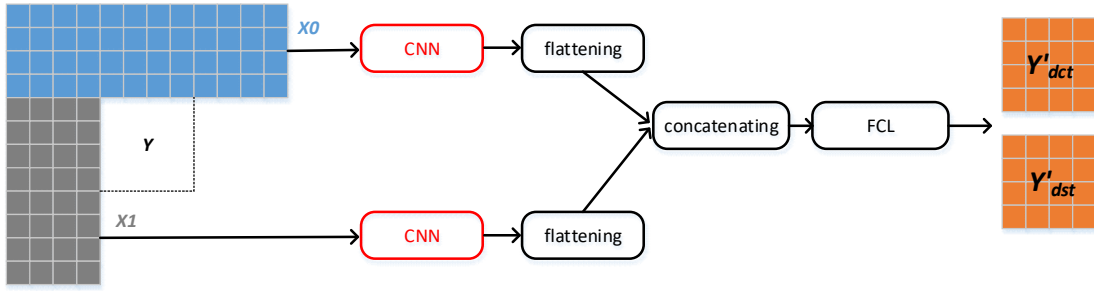


그림 1. MTS 를 위한 신경망 기반 화면내 예측 모델 구조

VVC MTS 를 고려한 신경망 기반 화면내 예측 기법을 제안한다.

2. MTS 를 위한 신경망 기반 화면내 예측

본 논문에서 제안하는 다중 선택 변환에 적합한 신경망 기반 화면내 예측 모델은 그림 1과 같이 구성된다. MTS를 위한 신경망 기반 화면내 예측(TDIP: Transform Dependent Intra Prediction) 모델은 현재 예측하고자 하는 부호화 블록 주변의 참조 블록이 입력되어 CNN(Convolutional Neural Network) 및 FCN(Fully Connected Network)를 통해 DST 기반의 예측 블록과 DCT 기반의 예측 블록을 동시에 출력한다. TDIP 모델의 파라미터들은 블록 크기 마다 정의되고 학습되며 각 TDIP 모델을 학습함에 있어 다음과 같은 손실함수를 이용한다.

$$L = \sum |T_{DCT2} \cdot R_{dct} \cdot T'_{DCT2}| + \sum |T_{DST7} \cdot R_{dst} \cdot T'_{DST7}| \quad (1)$$

R_{dct} 및 R_{dst} 는 원본 신호(Y)에 DCT 기반 예측신호(Y'_{dct}) 및 DST 기반 예측신호(Y'_{dst})를 각각 뺀 잔차신호이며, 손실함수 L 은 각 잔차신호에 해당하는 변환을 수행하여 얻어진 변환계수의 합으로 정의된다. TDIP 모델을 VVC 에 통합하기 위해 부호화 블록 단위로 신경망 기반 화면내 예측의 수행여부를 의미하는 지시자(flag)를 추가하였으며 DCT 기반의 예측을 사용할 지 또는 DST 기반의 예측을 사용할 지에 대한 여부는 추가적인 지시자 없이 기존의 MTS 색인(index) 정보에 따라서 결정된다.

3. 실험결과

제안된 화면내 예측 모델의 성능을 평가하기 위하여 제안기법 TDIP 와 제안기법과 동일한 학습환경에서 학습된 SAIP 모델을 각각 VTM(VVC Test Model)-11 위에 추가적인 화면내 예측 모드로 구현하였다. 실험은 JVET CTC(Common Test Condition)에 정의된 시퀀스에 대해 AI(All Intra) 부호화 환경에서 실험을 진행하였으며 MTS에 많은 영향을 미칠 수 있는 LFNST 를 배제하였다. 실험결과 SAIP 모델은 VTM-11 대비 Y, Cb, Cr 에 각각 0.72%, 0.48%, 0.79%의 BD-rate 절감을 보였으며 TDIP 모델은 Y, U, V 에 각각 0.87%, 0.87%, 0.99%의 BD-rate 이득을 보였다. 또한, SAIP 대비 TDIP 는 Y, U, V 에 각각 0.15%, 0.39%, 0.2%의 BD-rate 절감을 보였다.

표 1. VTM-11 대비 SAIP 실험결과

Class	Y	U	V
Class A1	-1.57%	-0.69%	-1.14%
Class A2	-0.53%	-0.15%	-0.57%
Class B	-0.57%	-0.43%	-0.47%
Class C	-0.45%	0.06%	0.16%
Class E	-0.64%	-1.41%	-2.50%
Overall	-0.72%	-0.48%	-0.79%

표 2. VTM-11 대비 제안기법(TDIP) 실험결과

Class	Y	U	V
Class A1	-1.70%	-1.93%	-1.10%
Class A2	-0.70%	-0.18%	-0.65%
Class B	-0.73%	-0.86%	-0.83%
Class C	-0.51%	-0.34%	-0.21%
Class E	-0.89%	-1.22%	-2.53%
Overall	-0.87%	-0.87%	-0.99%

3. 결론

본 논문에서는 NNVC EE 에서 검증된 SAIP 모델을 이용하여 VVC 의 다중 변환 선택에 따른 신경망 기반의 화면내 예측 기법을 제안하였다. 제안기법을 VTM-11 구현하여 실험 및 검증을 진행하였다. 실험결과 휘도성분에 대해 VTM-11 대비 0.87% 부호화 성능 향상을 확인하였으며 SAIP 대비 0.15%의 부호화 성능 향상을 확인하였다.

Acknowledgement

본 논문은 2022 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No. 2017-0-00072, 초실감 테라미디어를 위한 AV 부호화 및 LF 미디어 원천기술 개발)

참고 문헌(References)

- [1] *Versatile Video Coding*, Standard ISO/IEC 23090-3, ISO/IEC JTC1/SC29, Jul. 2020
- [2] T. Duman, F. Galpin, P. Bordes, and F. Léanneec, "AHG11: neural network-based intra prediction with transform selection in VVC," *JVET of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29*, document JVET-T0073, Oct. 2020
- [3] 문기화, 박도현, 김재곤, "문맥적응적 화면내 예측 모델 학습 및 부호화 성능분석", 방송공학회논문지, 제 27 권 제 3 호, 2022 년 5 월.