VVC 행렬가중 화면내 예측(MIP) 학습기법 분석

박도현, *권형진, *정세윤, 김재곤
한국항공대학교, *한국전자통신연구원
dhpark@kau.kr, {kwonjin, jsy}@etri.re.kr, jgkim@kau.ac.kr

Analysis of Training Method for Matrix Weighted Intra Prediction (MIP) in VVC

Dohyeon Park, *Hyoungjin Kwon, *Seyoon Jeong, and Jae-Gon Kim
Korea Aerospace University, *ETRI

요 약

최근 VVC(Versatile Video Coding) 표준 완료 이후 JVET(Joint Video Experts Team)은 인공신경망 기반의 비디오 부호화를 위한 AhG(Ad-hoc Group) 구성하고 인공지능을 이용한 비디오 압축 기술들을 검증하고 있으며, MPEG(Moving Picture Experts Group)에서는 DNNVC(Deep Neural Network based Video Coding) 활동을 통해 딥러닝 기반의 차세대 비디오 부호화 표준 기술을 탐색하고 있다. 본 논문은 VVC 에 채택된 신경망 기반의 기술인 MIP(Matrix Weighted Intra Prediction)를 참조하여, MIP 모델의 학습에서 손실함수가 예측 성능에 미치는 영향을 분석한다. 즉, 예측의 왜곡(MSE)만을 고려한 경우와 예측오차의 부호화 비용도 함께 반영한 손실함수를 비교한다. 실험을 위해 HEVC(High Efficiency Video Coding) 화면내 예측 대비 평균적인 PSNR 향상 정도를 나타내는 성능 지표($\Delta PSNR$)를 정의한다. 실험결과 예측오차의 부호화 특성을 반영하는 손실함수를 이용한 학습이 MSE 만 고려한 학습 대비 $\Delta PSNR$ 기준 평균 0.4dB 향상됨을 보였다.

1. 서론

MPEG(Moving Picture Experts Group)과 VCEG(Video Coding Experts Group)은 공동으로 JVET(Joint Video Experts Team)을 출범하여 다양한 비디오 서비스에 적합하며 HEVC(High Efficiency Video Coding) 대비 월등히 개선된 압축 성능을 갖는 VVC(Versatile Video Coding) 표준화를 2020 년 7 월 완료하였다[1, 2].

JVET 은 비디오 코덱의 성능의 향상을 위해 MIP(Matrix Weighted Intra Prediction) 및 LFNST(Low Frequency Non-Separable Transform)와 같은 학습된 신경망 모델을 이용한 부호화 기술들을 채택하였다. 또한 VVC 를 확장할 수 있는 신경망 기반의 비디오 부호화 기술의 잠재성을 확인하기 위한 AhG11(Ad-hoc Group11)을 두고 관련 기술을 연구하고 발전시키고 있다[3]. 또한, MPEG 에서는 딥러닝 기반 비디오 코딩 기술들을 탐색 및 개발하는 DNNVC(Deep Neural Network based Video Coding) 활동을 통해 딥러닝 기반 비디오 코덱의 차세대 표준 잠재성을 확인하고 있다[4]. 특히, 영상의 공간적 특성 및 참조샘플의 부족으로 인해 일반적인 화면내 예측은 성능 향상의 한계에 다다랐으며, 이를 극복할 수 있는 딥러닝 기반의 화면내 예측 기술의 중요성이 부각되고 있다.

본 논문은 VVC 에 채택되어 있는 MIP 의 구조를 참조하여

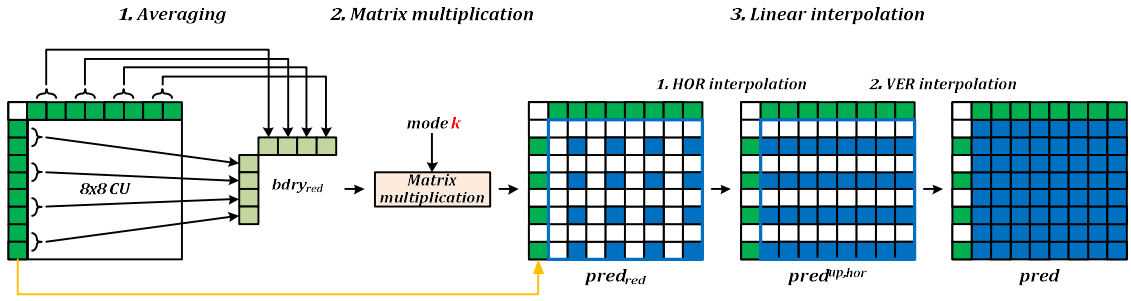

그림 1. VVC MIP 프로세스(8x8 블록)

MIP 모델 학습에 있어 손실함수(loss function)가 비디오 부호화 성능에 미치는 영향을 분석한다.

2. VVC MIP 및 MIP 참조모델 학습

(1) VVC MIP

그림 1 은 8x8 블록에 대한 VVC MIP 의 프로세스로 참조샘플 감소, MIP 예측 및 예측샘플 보간의 순서로 크게 3 단계로 구성된다. 참조샘플 감소 및 예측샘플 보간은 MIP 의 복잡도 감소를 위한 단계이며 실제 예측샘플은 MIP 예측단계에서 생성된다. MIP 예측단계에는 감소된 참조샘플과 학습을 통하여 기정의된 행렬과의 곱셈 연산을 통해 예측샘플을 생성한다. MIP 예측 모드의 수는 블록 크기에 따라 다르게 정의되어 있으며, 4x4 코딩 블록에 대해서는 30 개, 4x4 보다 크며 8x8 이하의 코딩 블록에 대해서는 16 개, 8x8 보다 큰 블록에 대해서는 6 개의 MIP 예측 모드가 정의되어 있다.

(2) MIP 참조모델 학습

MIP 모델을 학습하기 위해 VVC MIP 모델을 참조하여 MIP 참조모델을 구현하였다. MIP 참조모델의 각 모드는 그림 2 와 같이 예측하는 블록의 주변 참조샘플 벡터를 단일 전연결계층에 입력하여 예측블록을 생성하며 이는 식 1 과 같이 표현할 수 있다.

$$p_k = A_k r + b_k \tag{1}$$

r은 예측하고자 하는 블록의 참조샘플이며 k는 모드를 의미한다. A_k 및 b_k는 k 모드에서의 선형변환 행렬과 오프셋이며 p_k는 k 모드의 예측샘플이다. 모드의 수는 VVC MIP 와 동일한 수를 사용한다. 예를 들어, 8x8 블록에 대한 MIP 참조모델은 16 개의 단일 연결계층 모델로 구성된다.

본 논문에서는 MIP 모델을 학습할 때 손실함수의 영향을 알아보기 위해 두가지의 손실함수를 정의하며 모델을 학습한다. 첫번째 손실함수 L_{mse} 는 비디오 부호화 왜곡에 일반적으로 사용되는 MSE(Mean Squared Error) 기반으로 다음과 같이 정의한다.

$$\check{k} = \underset{k}{\operatorname{argmin}} \frac{1}{n}\sum_i (o_i - p_{k,i})^2 \tag{2}$$

$$L_{mse} = \frac{1}{n}\sum_i (o_i - p_{\check{k},i})^2 \tag{3}$$

o는 예측하고자 하는 블록에 해당하는 원본샘플이며 n은 예측샘플의 수이다(예를 들어, 8x8 블록의 n은 64). $\check{k}$는 최적의 비용을 갖는 모드를 의미하며 L_{mse} 는 $\check{k}$ 모드에 대한 MSE 이다. 두번째 손실함수 L_{codec}는 비디오 부호화의 변환 및 양자화를 반영하는 손실함수이며 다음과 같이 정의한다.

$$C = T \cdot (O - P_k) \tag{4}$$

$$\check{k} = \underset{k}{\operatorname{argmin}} \sum_i \operatorname{abs}(c_{k,i} + (\alpha g(c_{k,i}) + \beta)) \tag{5}$$

$$L_{codec} = \sum_i \operatorname{abs}(c_{\check{k},i} + (\alpha g(c_{\check{k},i}) + \beta)) \tag{6}$$

T는 2D DCT-2 변환 행렬이고 c 는 변환 계수이다. g는 로지스틱 함수를 의미하며($g(x) = 1/(1+e^{-x})$), L_{codec} 은 $\check{k}$ 모드에 대한 예측오차의 부호화 특성이 반영된 손실이다.

MIP 참조모델 학습은 8x8 블록에 대한 16 개의 모델에 대해서 진행한다. 학습에 사용된 데이터셋은 JVET 테스트 시퀀스와 COCO2017 데이터셋을 이용하여 구성하고, Adam 최적화 방법과 학습률(learning rate) 감소 기법을 이용하여 모델을 최적화 한다[5].

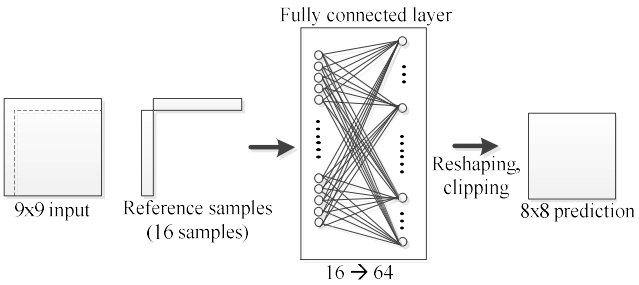

그림 2. MIP 참조모델(8x8 블록)

3. 실험결과

손실함수에 따른 MIP 참조모델 성능을 평가하기 위해 JVET CTC(Common Test Condition)의 A, B, C 클래스 시퀀스를

사용하였으며 각 클래스는 UHD, HD, SD 해상도를 갖는다. 실험은 VVC MIP 와 MIP 참조모델의 예측 성능을 HEVC 예측 성능과 PSNR 측면에서 비교하였다(p_{gain}). 실험 모델에 입력되는 참조샘플은 해당하는 HEVC 복원샘플을 이용하였다. 실험에 필요한 HEVC 의 예측블록 및 복원블록은 HM16.16 에서 QP22 로 압축하는 과정에서 구하였다. 해당 영상에 대한 실험을 진행하기 위해 각 시퀀스 별로 3 만개의 테스트 데이터를 추출하였으며, $\Delta PSNR$ 이라는 성능지표를 식 (8)과 같이 정의하여 각 시퀀스에 대한 성능을 제시하였다.

$$p_{gain} = \begin{cases} p_{mip} - p_{hevc}, & if\ p_{mip} > p_{hevc} \\ 0, & otherwise \end{cases} \quad (7)$$

$$\Delta PSNR = \frac{\sum_{i=0}^{n} p_{gain}^{i}}{n} \quad (8)$$

표 1. 시퀀스 별 VVC MIP 및 손실함수에 따른 MIP 참조모델 $\Delta PSNR$ 성능

		VVC MIP	MIP 참조모델 (L_{mse})	MIP 참조모델 (L_{codec})
Class	**Seq. name**	**$\Delta PSNR$**	**$\Delta PSNR$**	**$\Delta PSNR$**
A (3840x2160)	Campfire	0.89	0.80	1.17
	CatRobot1	1.63	1.49	1.96
	DaylightRoad2	1.43	1.16	1.58
	FoodMarket4	2.58	2.21	2.85
	ParkRunning3	3.05	2.36	3.14
	Tango2	1.53	1.45	1.79
	Average	**1.85**	**1.58**	**2.08**
B (1920x1080)	BQTerrace	0.79	0.60	0.89
	BasketballDrive	0.94	0.65	0.96
	Cactus	1.13	0.90	1.18
	MarketPlace	1.52	1.50	1.82
	RitualDance	1.61	1.30	2.05
	Average	**1.20**	**0.99**	**1.38**
C (832x480)	BQMall	1.11	0.91	1.28
	BasketballDrill	0.97	0.64	0.95
	PartyScene	1.09	0.88	1.19
	RaceHorsesC	1.53	1.13	1.55
	Average	**1.17**	**0.89**	**1.24**

실험결과, 그림 3 에서 보이는 것과 같이 MSE 기반의 손실함수를 이용하여 학습된 MIP 참조모델 보다 비디오 부호화의 특성이 반영된 손실함수를 이용하여 학습된 MIP 참조모델이 $\Delta PSNR$ 기준으로 대략 0.4dB 향상된 성능을 보였다.

4. 결론

본 논문에서는 VVC 의 MIP 를 참조하여 MIP 모델을 학습함에 있어 손실함수에 따른 영향을 분석하였다. MSE 손실함수와 비디오 부호화의 특성이 반영된 손실함수를 이용하여 학습된 두 MIP 참조모델을 성능을 비교하였다. MIP 예측블록의 평균적인 PSNR 증가 정도를 표현하는 $\Delta PSNR$ 을 정의하여 성능지표로 사용하였으며, 실험결과 비디오 부호화의 특성을 반영한 손실함수로 학습된 MIP 참조모델의 성능이 MSE 로 학습된 모델 대비 평균 0.4dB 향상됨을 확인하였다. 또한, 실제 비디오 코덱에 MIP 참조모델을 통합하여 제안 모델의 부호화 성능을 BD-rate 성능으로 확인하는 추가적인 연구가 필요하다.

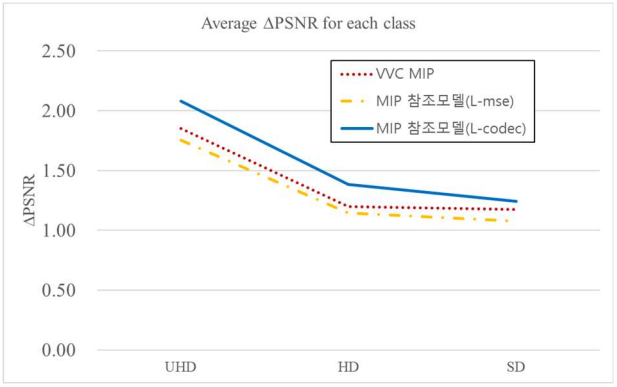

그림 3. 클래스 별 VVC MIP 및 손실함수에 따른 MIP 참조모델 $\Delta PSNR$ 성능

Acknowledgement

이 논문은 2020 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No. 2017-0-00072, 초실감 테라미디어를 위한 AV 부호화 및 LF 미디어 원천 기술 개발)

참 고 문 헌(References)

[1] High Efficiency Video Coding, Version 1, Rec. ITU-T H.265, ISO/IEC 23008-2, Jan. 2013.

[2] Versatile Video Coding, ISO/IEC FDIS 23090-3, Jul. 2020.

[3] S. Liu, E. Alshina, J. Pfaff, M. Wien, P. Wu and Y. Ye, "JVET AHG report: Neural-network-based video coding," Joint Video Experts Team of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29, JVET-T0011, Oct. 2020.

[4] "Use cases and requirements for Deep Neural Networks based Video Coding," ISO/IEC JTC 1/SC 29/WG 2, N22, Oct. 2020.

[5] P. Helle, J. Pfaff, M. Schäfer, R. Rischke, H.Schwarz, D. Marpe, and T. Wiegand, "Intra Picture Prediction for Video Coding with Neural Networks," In Proc. Data Compression Conference 2019, IEEE, Mar. 2019.