레터논문 (Letter Paper)
방송공학회논문지 제23권 제5호, 2018년 9월 (JBE Vol. 23, No. 5, September 2018)
https://doi.org/10.5909/JBE.2018.23.5.718
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

비디오 압축을 위한 딥러닝 기반 화면 간 예측 부호화 기법

이 정 경[a)], 김 나 영[a)], 강 제 원[a)‡]

Deep Learning based Inter Prediction Technique for Video Coding

Jeongkyung Lee[a)], Nayoung Kim[a)], and Je-Won Kang[a)‡]

요 약

최근 차세대 국제 비디오 압축 표준 제정에 딥러닝을 이용하여 비디오 부호화 효율을 향상시키기 위한 다양한 시도가 이루어지고 있다. 본 논문에서는 참조프레임 리스트에 포함된 복원 프레임을 이용하여 현재 프레임의 가상 참조프레임을 딥러닝으로 생성하여 화면 간 예측 부호화에 이용하는 알고리즘을 제안한다. 실험에 따르면 제안 알고리즘은 HEVC 참조 소프트웨어 대비 Random Access 실험 환경에서 평균 1.9%의 BD-rate 감소 효율을 제공한다.

Abstract

This paper presents an inter-prediction technique using deep learning, where a virtual reference frame of the current frame is synthesized by using the reconstructed frames to improve coding efficiency. Experimental results demonstrate that the proposed algorithm provides 1.9% BD-rate reduction on average as compared to HEVC reference software in the Random Access condition.

Keyword : HEVC, deep learning, virtual reference frame

Ⅰ. 서 론

최근 MPEG (Moving Picture Experts Group) 및 VCEG (Video Coding Experts Group)에서는 High Efficiency Video Coding (HEVC) 대비 2배 부호화 성능을 제공하기 위한 차세대 비디오 압축 표준 제정을 목표로 Versatile Video Coding (VVC) 표준화를 시작하였다. 특히 VVC call-for-proposal (cfp) 응답에는 딥러닝 기반의 부호화 기술들이 다수 포함되었다[1]. 최근의 딥러닝 기반의 압축 기술 연구는 크게 두 가지 방향으로 요약이 가능하다. 첫째, 오토엔코더 기반의 코덱을 이용하여 입력 영상을 인코더 신경망의 종단에서 은닉벡터로 변환하고 엔트로피 부호화를 수행하는 종단 간 압축 방식이다. 현 시점에서 오토엔코

a) 이화여자대학교 엘텍공과대학 전자전기공학과 (Department of Electronic and Electrical Engineering, Ewha W. University)
‡ Corresponding Author : 강제원(Je-Won Kang)
E-mail: jewonk@ewha.ac.kr
Tel: +82-2-3277-2347
ORCID:http://orcid.org/0000-0002-1637-9479
※ This work was supported by IITP grant funded by the Korea government(MSIP)(2017-0-00072)
· Manuscript received August 7, 2018; Revised September 10, 2018; Accepted September 10, 2018.

Copyright ⓒ 2016 Korean Institute of Broadcast and Media Engineers. All rights reserved.
"This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (http://creativecommons.org/licenses/by-nc-nd/3.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered."

더 기반 코덱은 정지영상 압축에서 HEVC 화면 내 압축 성능에 근접하는 연구결과가 보고되고 있다[2]. 둘째, 딥러닝 기반 영상처리 기법을 기존 비디오 부호화 요소 기술에 적용하는 방식이다. 대표적인 응용 예로 콘볼루셔널 뉴럴 네트워크 (CNN) 기반 초해상도 기법 및 디노이징 기법을 인루프필터에 이용하여 복원 프레임의 화질을 개선하고 압축 효율을 증대시키는 알고리즘이 제안되었다[3,4].

딥러닝을 이용하여 영상의 공간해상도를 향상시키고 부호화에 응용하는 다양한 연구가 진행되어 왔음에도 불구하고 비디오 프레임의 시간해상도를 향상하여 부호화 효율을 증대시키는 연구는 거의 시도되지 않았다[5]. 본 논문에서는 딥러닝 비디오 프레임 생성 기법에 기반을 둔 화면 간 예측 부호화 알고리즘을 제안한다. 제안 알고리즘은 비디오 코덱의 참조프레임 리스트에 포함된 복원 프레임을 이용하여 현재 프레임의 가상 참조 프레임을 생성하고, 기존의 참조 프레임을 새로운 가상 프레임으로 대체하며 화면 간 예측에 이용한다. 추가적으로 제안 알고리즘에서는 HEVC의 움직임 벡터 예측 모드를 가상 참조프레임의 이용에 적합하게 개선하였다.

Ⅱ. 딥러닝 기반 비디오 프레임 생성 기법

비디오 프레임 생성 기법은 과거 비디오 프레임을 이용하여 아직 관측되지 않은 새로운 비디오 프레임을 예측하는 알고리즘이다. 최근에는 비디오 프레임 간 시간상의 연관성을 이용하는 순환신경망 구조 (RNN)에 기반을 두고 장기간의 비디오를 예측 생성하는 연구가 활발하게 진행이 되고 있으나, RNN의 학습에 오차 역전파 과정에서의 기울기 소실 문제를 완전히 극복하기 어려울 뿐 아니라 입력 영상을 은닉벡터로 변환하고 다시 복원하는 과정에서 많은 블러링이 발생하는 등의 문제가 있어 여전히 많은 연구가 필요하다[6]. 반대로 비교적 짧은 시간 거리에 있는 비디오 생성을 위해 RNN을 생략하고 CNN 구조를 이용하여 입력 비디오를 합성하고 출력하는 연구가 비디오 보간에 적용되어 우수한 보간 성능을 제공하고 있다[7]. 그림 1은 비디오 보간을 위해 입력 영상의 합성 커널 K 를 CNN을 이용하여 학습하고, 현재 시점의 영상을 생성하는 방식을 보인다. 복잡도 감소를 위하여 K는 두 입력 영상의 수평방향 합성 커널 및 수직방향 합성 커널로 분할하여 학습이 가능하다.

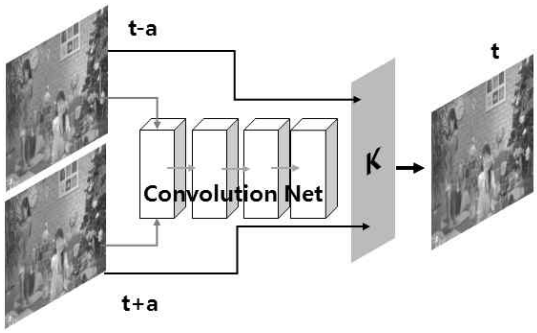


그림 1. CNN을 이용한 단기 비디오 보간법
Fig. 1. Short-term video interpolation using CNN [7]

Ⅲ. 제안 알고리즘

1. 딥러닝 기반 가상 참조프레임 생성을 이용한 화면 간 예측 부호화

제안 알고리즘에서는 딥러닝을 이용하여 가상 참조프레임을 생성하여 화면 간 예측 부호화에 이용한다. 그림2는 제안 알고리즘의 개요를 보인다. 복호 프레임의 저장 버퍼 (decoded picture buffer, DPB)에서 비디오 프레임 순서 (picture of count, POC)로 과거 시점 $t-a$의 전방향 예측 프레임 P_F과 및 미래 시점 $t+a$의 역방향 예측 프레임 P_B을 입력하여 현재 시점 t의 프레임과 동일한 시점에서의 가상 참조프레임을 생성하고 HEVC의 움직임벡터예측 (Advanced motion vector prediction, AMVP) 모드 및 머지 (Merge) 모드를 이용한 화면 간 예측 부호화에 이용한다. 그림 2의 제안 알고리즘 개요에서 가상 참조프레임을 생성하는 비디오 프레임 생성기 (SYN)는 Niklaus et al.의 딥러닝 기반 비디오 보간법[7]을 이용하였다. 이용된 딥러닝 기반 보간법 모델은 총 30여개의 합성곱 계층으로 이루어져 있으며 과거 시점의 영상과 미래 시점의 영상을 입력으로 받아 두 시점의 중간에 위치한 영상을 출력한다. 이때, 출력된 영상과 정답으로 사용하는 현재 시점의 이미지 간의 L_1 손실 값을 손실함수로 설정한다. 부호기와 복호기 모두 동일한 학습 파라메터를 갖는 비디오 프레임 생성기를 갖도록 설정하여 별도의 부가정보전송이 필요 없도록 하였다. 복호 프레임의 저장 버퍼에서 P_F와 P_B의 선택은 현재 프레임의 POC와 전방향 및 역방향 참조프레임과의 POC 차이

가 가장 작으면서 그 차이가 같도록 HEVC의 참조프레임 리스트 0 (RefPicList0)과 참조프레임리스트 1 (RefPicList1)에서 각각 한 장씩 사용한다.

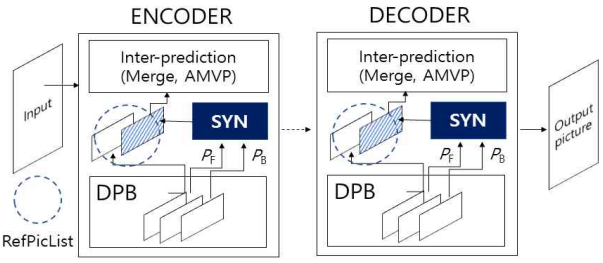


그림 2. 제안 알고리즘 블록 다이어그램
Fig. 2. Block diagram of the proposed algorithm

참조프레임리스트의 구성은, 참조프레임리스트 0과 참조프레임리스트 1에 채워진 첫 번째 인덱스 (Idx0)에 위치한 참조 프레임은 기존과 동일하게 유지하되, 두 번째 인덱스 (Idx1)에 위치한 참조 프레임은 가상 참조프레임을 기존 참조프레임을 대체한다. 가상참조프레임을 생성하는 경우 POC 차이가 가까울수록 비디오 보간이 용이하므로 임의접근구조 (Random Access)의 마지막 시간 계층 (temporal layer)에 한하여 적용하였다.

2. 움직임 벡터 예측 모드

가상참조 프레임은 과거와 미래의 복원 프레임을 보간하여 생성되므로 그 원리 상 현재 프레임과 같은 시간에서의 움직임 정보를 포함할 수 있다. 따라서 HEVC의 움직임 벡터 예측 모드를 가상 참조프레임의 이용에 적합하게 그림 3과 같

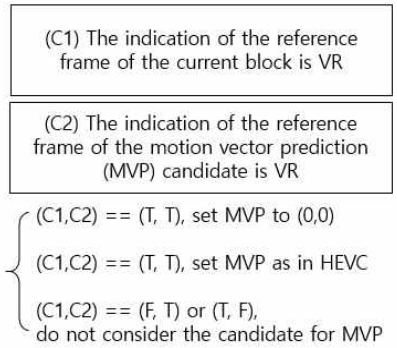


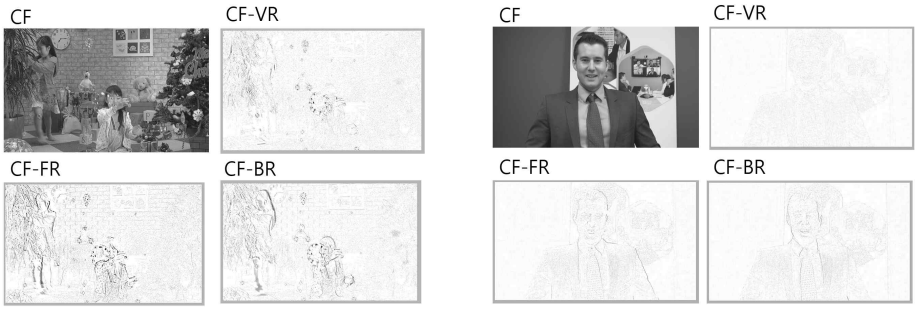
그림 3. 가상 참조프레임을 사용하는 경우 움직임 벡터 예측 기법 변경
Fig. 3. Modified AMVP when using the virtual reference frame

이 수정한다. 우선 현재 블록의 참조프레임 인덱스가 가상 참조프레임의 인덱스와 동일한지 확인 후 움직임벡터후보의 참조프레임이 가상프레임이면 움직임 벡터를 (0,0)의 영벡터로 예측하고, 그렇지 않으면 후보 선택을 생략한다. 또한 시간적으로 인접한 움직임예측벡터후보와 공간적으로 인접한 움직임예측벡터후보 간의 탐색 순서를 변경한다. 현재 블록의 참조프레임과 움직임벡터 후보의 참조 프레임이 모두 가상 참조프레임이 아니면 기존의 HEVC AMVP 모드를 사용한다.

IV. 실험 결과

제안 알고리즘은 HEVC 참조 코덱인 HM-16.9 코덱을 이용하여 구현하였다. 실험은 공통 실험 조건(Common Test Condition)을 참고하여 화면 간 압축의 Random Access 환경에서 수행하였고 HM16.9와 대비하여 부호화 성능을 측정하였다. 제안 알고리즘의 효율성을 검증하기 위해 참조 프레임 수를 최대 두 장으로 제한하여 실험하였다. 표 1에서와 같이 다양한 해상도를 가지는 테스트 비디오를 사용하였으며, 비디오 프레임 생성에 사용한 딥러닝 모델은 “Tom Scott”, “Casey Neistat”, “Linus Tech Tips” 그리고 “Austin Evans” 같은 Youtube 채널에서 취득한 비디오 데이터로 학습하여 테스트 비디오와 학습 비디오가 중복되지 않는다. 가상참조프레임의 생성은 Pytorch를 이용하여 구현하였고 실험은 CUDA를 이용하여 GPU 환경에서 수행하였다.

표 1은 제안 기법의 부호화 성능을 보이며 비디오 압축에서 HM16.9 대비 평균적으로 1.9% BD-rate 감소의 우수한 부호화 효율을 보인다. 가상 참조프레임 생성시간을 모두 고려한 결과, 부호기 및 복호기의 복잡도는 각각 103%, 218%로 늘어났다. GPU 환경에서 연산하여 비디오 생성은 부호기 복잡도에 큰 영향을 미치지 않는다. 하지만 CPU 환경에서는 부호기 및 복호기의 복잡도는 279%, 1,562%로 다소 크게 늘어난다. 각 비디오 시퀀스 별 결과를 분석해보면, “BasketballPass” 및 “BQSquare” 등과 같이 작고 움직임이 빠른 오브젝트를 포함한 비디오에서 우수한 부호화 효율을 보였고, “Johnny” 등과 같이 움직임이 비교적 적어 과거 및 미래 비디오 프레임과 현재 프레임 사이 시간적 연관성이 큰 비디오는 상대적으로 부호화 효율이 적었다. 위 비교



(A) Blowing Bubbles (B) Johnny

그림 4. (a) Blowing Bubble 비디오와 (b) Johnny 비디오의 현재프레임 대비 가상참조프레임, 전방참조프레임, 후방참조프레임 간 비교
Fig. 4. Comparisons between the current frame (CF) and virtual reference (VR) frame, forward reference (FR) frame, and backward reference (BR) frame on (A) Blowing Bubbles video and (B) Johnny video [5]

표 1. HM-16.9 대비 제안 기법의 부호화 성능
Table 1. Coding gain of the proposed algorithm VS HM-16.9

Test sequence		BD-rate (Y)
416x240	RaceHorses	-3.4%
	BlowingBubbles	-2.8%
	BasketballPass	-3.2%
	BQSquare	-2.9%
832x480	RaceHorses	-2.3%
	BQMall	-2.6%
	PartyScene	-2.1%
	BasketballDrill	-1.0%
1920x1080	BasketballDrive	-1.5%
	BQTerrace	-0.8%
	Cactus	-1.8%
	Kimono1	-2.0%
1280x720	ParkScene	-2.0%
	Johnny	-0.4%
	FourPeople	-0.9%
	KristenAndSara	-0.6%
Average		-1.9%

를 통해 가상 참조프레임은 전방향 및 역방향 참조프레임과 현재 비디오 프레임의 시간적 연관성이 적을수록 시간 간 예측 부호화에서 보다 효율적으로 사용되었음을 알 수 있다. 그림 4는 “Blowing Bubble” 시퀀스의 현재프레임과 가상 참조프레임의 차분치와 전방향/역방향 참조프레임 간 차분치를 비교하여 보인 결과로 중앙에 위치한 사람의 움직임이 가상참조 프레임에서 보다 정확히 추정됨을 확인할 수 있다.

V. 결 론

본 논문에서는 효율적인 비디오 부호화를 위한 딥 러닝 기반 화면 간 예측 기법을 제안하였다. 제안 기법은 과거 및 미래 비디오 신호로부터 현재 비디오 신호를 생성하여 가상의 참조 프레임으로 사용하고 예측 부호화를 수행한다. 향후 연구로는 고효율 저복잡도의 비디오 생성기를 개발하여 부호화 효율 및 복잡도를 개선하는 것이다.

참 고 문 헌 (References)

[1] S. Liu, L. Wang, P. Wu, and H. Yang, “JVET AHG report 9: Neural Networks in Video Coding (AHG9)” in ISO/IEC/JTC1/SC29/ WG11 and ITU-T SG16 Q.6, Apr. 2018.
[2] S.H Cho, Y. H. Kim, W. Lim, H.W. Kim, and C.S. Choi, “A Technical Analysis on Deep Learning based Image and Video Compression", Journal of Broadcast Engineering, Vol. 23, No. 3, 383-394, May 2018
[3] W. Park and M. Kim, "CNN-based in-loop filtering for coding efficiency improvement," IEEE Image, Video, and Multidimensional Signal Processing Workshop, 2016.
[4] J. Kang, S. Kim, and K. M. Lee, “Multi-modal Multi-scale Convolutional Neural Network based In-loop Filter Design for Next Generation Video Codec,” IEEE International Conference on Image Processing, 2017.
[5] J.K. Lee and J.-W. Kang, “Video coding technique based on deep learning”, 2018 KIBME Summer Conference.
[6] B. D. Brabandere, X. Jia, T. Tuytelaars, and L. V. Gool, "Dynamic filter networks," Neural Information Processing Systems (NIPS). 2016.
[7] S. Niklaus, L. Mai, and F. Liu. "Video frame interpolation via adaptive separable convolution," International Conference on Computer Vision, 2017.
[8] HM16.9 software, available at: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.9
[9] F. Bossen, “JCTVC-L1100: Common test conditions and software reference configurations” in ISO/IEC/JTC1/SC29/WG11 and ITU-T SG16 Q.6, Jan.2013.