

특집논문 (Special Paper)

방송공학회논문지 제24권 제3호, 2019년 5월 (JBE Vol. 24, No. 3, May 2019)

<https://doi.org/10.5909/JBE.2019.24.3.401>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

SHVC 부호화 성능 개선을 위한 딥러닝 기반 계층간 참조 픽처 생성 방법

이우주^{a)}, 이종석^{a)}, 심동규^{a)}, 오승준^{a)†}

A Deep Learning based Inter-Layer Reference Picture Generation Method for Improving SHVC Coding Performance

Wooju Lee^{a)}, Jongseok Lee^{a)}, Dong-Gyu Sim^{a)}, and Seung-Jun Oh^{a)†}

요 약

본 논문에서는 SHVC 부호화 성능 개선을 위하여 딥러닝 기반 계층간 예측을 위한 참조 픽처 생성 방법을 제안한다. 새로운 참조 픽처를 생성하기 위하여 DCT-IF 기반 업샘플링 된 픽처를 VDSR 네트워크를 이용한 필터링을 진행하는 구조와 SHVC 계층간 참조 픽처를 생성하기 위한 트레이닝 방법에 대해 설명한다. 제안하는 방법은 SHM 12.0 기반으로 구현되어 있다. 성능 평가를 위하여 사전 학습을 이용하여 계층간 예측 픽처를 생성하는 방법과 비교를 진행하였다. 그 결과 상위 계층의 부호화 성능은 사전 학습을 이용한 방법 대비 최대 13.14%의 비트 감소, SHM 대비 최대 15.39%의 비트 감소율을 보였고, 평균 6.46%의 비트 감소율을 보였다.

Abstract

In this paper, we propose a reference picture generation method for Inter-layer prediction based deep learning to improve the SHVC coding performance. A description will be given of a structure for performing filtering using a VDSR network on a DCT-IF based upsampled picture to generate a new reference picture and a training method for generating a reference picture between SHVC Inter-layer. The proposed method is implemented based on SHM 12.0. In order to evaluate the performance, we compare the method of generating Inter-layer predictor by applying dictionary learning. As a result, the coding performance of the enhancement layer showed a bitrate reduction of up to 13.14% compared to the method using dictionary learning, a bitrate reduction of up to 15.39% compared to SHM, and a bitrate reduction of 6.46% on average.

Keyword : Scalable HEVC, CNN, Deep learning, Super resolution, Inter-layer prediction

a) 광운대학교(Kwangwoon University)

† Corresponding Author : 오승준(Seung-jun Oh)

E-mail: sjoh@kw.ac.kr

Tel: +82-2-940-5102

ORCID: <https://orcid.org/0000-0002-5036-3761>

※ 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터지원사업의 연구결과(IITP-2019-2016-0-00288)와 2018년도 광운대학교 교내 학술연구비 지원에 의해 연구되었음.

※ This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2019-2016-0-00288) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation) and the Research Grant of Kwangwoon University in 2018.

· Manuscript received March 18, 2019; Revised May 4, 2019; Accepted May 4, 2019.

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.”

1. 서론

최근 다양한 멀티미디어 장비와 기술의 급속한 발전에 따라, 동영상 정보를 포함한 다양한 멀티미디어 서비스에 대한 소비자의 요구가 증대되고 있으며, 4K UHD(Ultra High Definition) 이상의 고품질 방송 서비스가 등장하면서, 정보의 질적 향상에 더 많은 관심을 가지게 되었다. 이에 발 맞추어 ITU-T(International Telecommunication Union Telecommunication Standardization Sector)의 VCEG(Video Coding Experts Group)과 ISO/IEC(International Organization for Standardization/International Electrotechnical Commission Joint Technical Committee)의 MPEG(Moving Picture Experts Group)은 JCT-VC(Joint Collaborative Team on Video Coding)을 조직하고 8K 영상까지 지원하는 압축 표준인 HEVC(High Efficiency Video Coding)를 제정하였다. HEVC는 고정된 영상 포맷(format)과 비트율(bitrate)로 압축 및 전송하는 단일 계층 압축 부호화(single layer video coding) 방식이다. 하지만 단일 계층 압축 부호화 방식은 최근 무선 네트워크 기술의 발달로 다양한 환경에서 시변의 전송조건, 망의 특성, 사용자의 선호도, 그리고 단말기의 성능 등의 조건에 맞게 미리 부호화를 진행해야 하기 때문에 멀티미디어의 실시간 서비스 제공이 힘들다. 이러한 환경에서 실시간 비디오 전송 및 이종망(Heterogeneous) 서비스에 적합한 시간, 공간, 화질 등의 다양한 스케일러빌리티(scalability)를 지원하는 SVC(Scalable Video Coding) 기술은 멀티미디어의 전송 및 소비 환경 적응을 위한 해결책 중 하나다^[1]. 이에 발 맞추어 JCT-VC는 2014년에 공간적, 시간적, 화질적 스케일러빌리티를 제공하는 확장 표준인 SHVC(Scalable HEVC)를 제정하였다^[2]. HEVC와 같이 스케일러블(scalable)이 지원되지 않는다면, 사용자의 환경에 따라 원본 영상을 다양한 해상도/프레임율/화질로 부호화 하거나 이미 부호화 된 비트스트림을 각각의 환경에 적합하게 변환부호화(transcoding)하거나 복호화된 영상을 변환시키는 등 추가적인 처리가 필요하다. 그러나 SHVC는 하나의 비트스트림을 통하여 다양한 해상도/프레임율/화질 등을 지원하도록 계층적 방식으로 부호화를 진행하고 사용자의 환경에 따라 비트스트림의 일부를 이용하여 복호화 가능한 방식이다. SHVC의 구조는 시간적으로 동일한 위치에 있는 하위 계층의 픽처(picture)를 상위 계층의 참조 픽

처 리스트에 추가하여 상위 계층 부호화에서 화면 간 예측과 같이 사용할 수 있도록 지원하는 참조 인덱스 기반 프레임워크를 채택하였다^[3]. 계층간 예측을 위하여 하위 계층의 픽처를 DCT-IF(Discrete Cosine Transform-based Interpolation Filter) 기반 업샘플링(upsampling) 방법^[4]을 사용하여 확장 계층(Enhancement Layer)의 참조 픽처로 사용한다. 하위 계층에서 업샘플링 된 픽처가 상위 계층의 예측 신호로 사용된다면 상위 계층에서 예측이 힘든 블록의 경우 예를 들어, 새로운 텍스처가 발생하거나 영상 내 새로운 객체가 등장한 경우에는 기존 예측 방법보다 잔차 신호를 줄여 비트율을 감소시킬 수 있다. 하지만 DCT-IF 기반 업샘플링 방법은 고주파 성분을 온전히 재구성하기 힘들다는 ill-posed 문제를 가지고 있다. ill-posed 문제를 해결하기 위하여 SC(Sparse-Coding) 방식의 사전 학습(Dictionary Learning)을 이용한 SISR 접근법이 있다^{[5][6]}. 영상 데이터가 이전에 구성된 사전 원소들의 선형 조합으로 표현될 수 있다는 것을 이용하여 고해상도의 영상을 복원하는 방법이다. 하지만 저해상도의 복원된 영상에 발생하는 압축 열화가 매우 다양하게 발생하여 사전학습 기반 SISR의 경우는 모든 경우에 대하여 사전을 만들지 못하는 한계가 있다.

최근 SISR분야에서 CNN(Convolutional Neural Network)을 이용한 모델이 기존의 머신 러닝 기반 SISR방법의 성능보다 일반적으로 향상된 결과를 보이며 많은 연구가 꾸준히 이루어지고 있다^{[7][8][9][10]}. CNN 기반 SISR(Single Image Super Resolution)의 경우 많은 열화를 포함하는 데이터 셋을 사용함으로써 다양한 열화에 대한 고려가 가능하다는 장점이 있다. SRCNN^[7]은 SISR분야에서 CNN이 최초로 적용된 네트워크 구조이다. SRCNN의 수용 영역(receptive field)의 크기가 작아 높은 해상도에서 복원율이 좋지 않고 모든 배율에 해당하는 개별적인 네트워크를 학습시켜 사용을 해야 한다는 단점이 있다. 이 단점을 보완하기 위해 수용 영역의 크기를 증가시키고 여러 배율을 한번에 학습시켜 하나의 네트워크로 설계한 VDSR(Very Deep Super Resolution)^[8]이 있다. 또한, 잔차 신호에 대한 학습을 진행하기 때문에 SRCNN보다 학습 시간을 줄였다. VDSR의 경우 하나의 Global Skip-connection이 존재하기 때문에 지역적인 정보들을 충분히 사용하지 못하고 무시해버리는 문제점이 존재한다. 이를 개선하기 위해 잔차 블록과 Local Skip-connection을 사용하여 정확도를 높인 SRResnet^[9]과 이를

개선한 EDSR(Enhanced Deep Super Resolution)^[10]이 있다.

비디오 압축 분야에서도 기존의 SAO(Sample Adaptive Offset)를 통한 인루프(In-loop) 필터링 대신에 CNN을 기반으로 링잉 열화(Ringing artifact)에 대하여 학습시킨 IFCNN(In-loop Filtering CNN) 모델을 제안하여 약 2~3%의 비트율을 감소시켰다^[11]. 위 방법은 비디오 압축 분야에서 CNN을 이용하여 압축 효율을 올릴 수 있다는 증거를 제시하였다. 또한 Jens^[12]은 이러한 SC 방식의 사전 학습 기반의 SISR을 이용하여 SHVC에서 저해상도의 복원된 픽처를 업샘플링한 후 고주파 성분을 복원하여 상위 계층의 참조 픽처를 생성하는 방법을 제안했다. 그러나 위 논문에서 사용한 사전 학습 기반 SISR의 경우 다양한 압축 열화에 대하여 사전 원소를 구성하기 어렵고 화면내 예측이 진행되는 I 프레임에서만 사전 학습 기반의 SISR을 적용한 문제가 있다.

따라서 본 논문에서는 SHVC의 저해상도의 복원된 영상에 대한 다양한 열화 및 계산 복잡도를 고려하여 CNN기반의 VDSR 네트워크를 이용한 계층간 참조 픽처 생성방법을 제안한다. 본 논문에서 사용하는 VDSR 네트워크는 기존 배율에 대한 열화가 아닌 비디오 부/복호화에서 발생하는 열화에 대하여 학습을 진행하였다. 제안하는 방법은 DCT-

IF를 이용하여 하위 계층의 픽처를 업샘플링 하고 사전에 학습된 VDSR 네트워크를 통해 계층간 참조 픽처를 생성한다. 또한 I 프레임뿐 만 아니라 화면 간 예측을 진행하는 P, B 프레임에 대해서도 제안하는 방법을 적용하였다. CNN 기반 업샘플링 방법을 사용함으로써 하위 계층의 복원된 픽처가 가지는 부호화 열화와 다운샘플링으로 인한 고주파성분의 손실을 보상할 수 있다.

이후 본 논문의 구성은 다음과 같다. 2장에서는 SHVC의 구조와 VDSR 네트워크 구조에 대해 설명하고 3장에서는 제안하는 방법의 구조와 학습 방법을 설명하고 4장에서는 제안하는 방법과 이전 연구들과 객관적, 주관적 측면에서 비교하여 성능을 평가하고 5장에서 결론을 맺는다.

II. 관련 이론

1. SHVC(Scalable HEVC)의 구조

그림 1은 SHVC 부호화기 흐름도이다. SHVC는 그림 1과 같이 원본 해상도의 영상을 계층적으로 다운샘플링 하

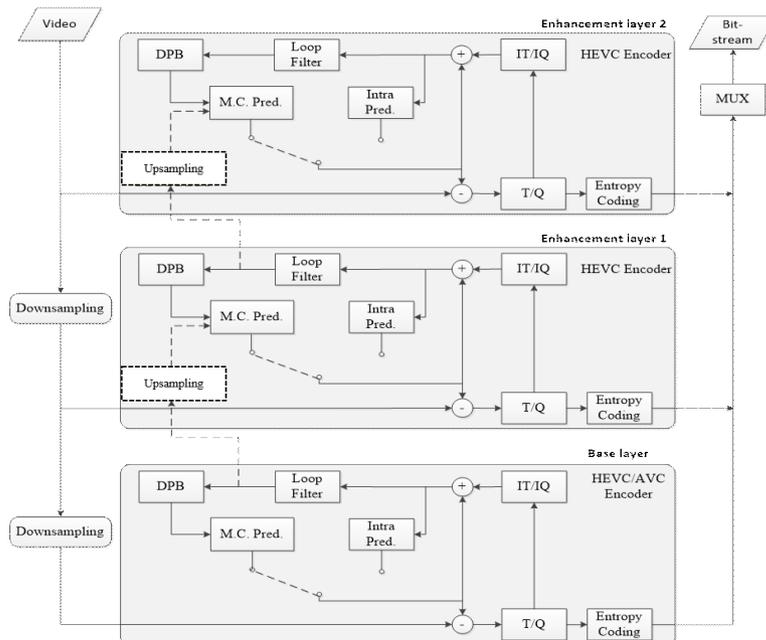


그림 1. 기존 SHVC 부호화기 흐름도^[13]
 Fig. 1. The conventional SHVC encoder flowchart^[13]

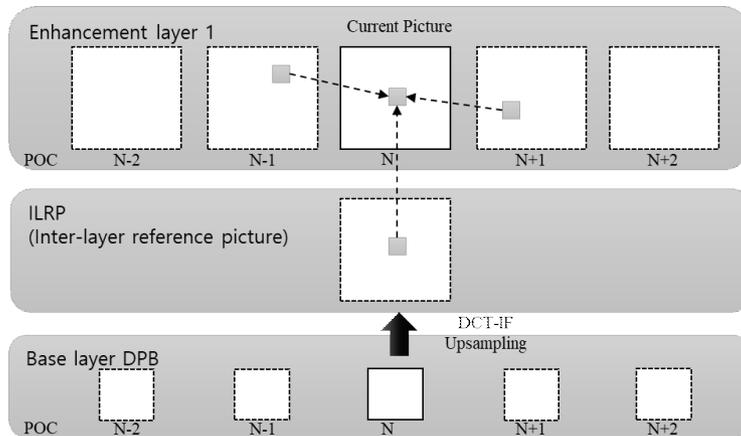


그림 2. SHVC의 계층간 참조 방법의 예시
 Fig. 2. An example of SHVC Inter-Layer reference picture method

여 각 계층의 입력하여 부호화 하는 다 계층 구조이다. 각 계층들은 하위 계층에서 부호화 한 정보를 이용하여 상위 계층의 참조 픽처로 활용함으로써 부호화 효율을 향상시킬 수 있다. SHVC에서는 HEVC 표준의 수정 없이 계층간 예측에 사용하기 위해서 하위 계층의 복호화 된 픽처를 상위 계층의 해상도에 맞게 DCT-IF 기반 업샘플링을 진행한 픽처를 상위 계층 예측을 위한 참조 픽처로 사용한다. 이러한 방법을 통해 상위 계층에서는 계층간 참조 픽처를 추가적인 예측 정보로 활용이 가능하다. 그림 2는 SHVC의 계층간 참조 방법의 예시이다. SHVC의 상위 계층의 픽처를 예측하기 위하여 상위 계층의 화면간 예측 픽처 리스트에 하

위 계층의 동일한 POC(Picture Order Count) 픽처를 상위 계층의 해상도에 맞도록 업샘플링하여 추가적인 픽처로 사용하여 예측을 진행한다. 또한, 계층간 참조 픽처를 사용하여 부호화하는 경우 움직임 벡터(Motion Vector)가 (0, 0)로 동일한 위치에 존재하는 블록만을 예측 블록으로 제한함으로써 부호화 속도를 높였다.

2. DCT-IF(Discrete Cosine Transform-based Interpolation Filter) 기반 업샘플링

SHVC에서는 하위 계층의 복원된 픽처를 상위 계층에서

표 1. 휘도 성분에 대한 업샘플링 필터 계수^[3]
 Table 1. Upsampling filter coefficients for luminance component^[3]

위상 p	보간 필터 계수								
	fL[p,0]	fL[p,1]	fL[p,2]	fL[p,3]	fL[p,4]	fL[p,5]	fL[p,6]	fL[p,7]	
0	0	0	0	64	0	0	0	0	
1	0	1	-3	63	4	-2	1	0	
2	-1	2	-5	62	8	-3	1	0	
3	-1	3	-8	60	13	-4	1	0	
4	-1	4	-10	58	17	-5	1	0	
5	-1	4	-11	52	26	-8	3	-1	
6	-1	3	-9	47	31	-10	4	-1	
7	-1	4	-11	45	34	-10	4	-1	
8	-1	4	-11	40	40	-11	4	-1	
9	-1	4	-10	34	45	-11	4	-1	
10	-1	4	-10	31	47	-9	3	-1	
11	-1	3	-8	26	52	-11	4	-1	
12	0	1	-5	17	58	-10	4	-1	
13	0	1	-4	13	60	-8	3	-1	
14	0	1	-3	8	62	-5	2	-1	
15	0	1	-2	4	63	-3	1	0	

참조 픽처로 사용하기 위해서 해상도에 맞도록 업샘플링이 필요하다. 업샘플링 하기 위해 사용되는 필터는 표 1에서처럼 휘도 성분에 대하여 8 탭 필터 계수를 사용한다. 필터의 계수는 6비트의 정밀도를 가지며 기본적으로 HEVC의 움직임 보상 보간시 사용하는 DCT-IF 와 동일하다. 또한, 계층간 해상도 배율을 다양하게 지원하도록 모든 위상에 대한 필터 계수를 추가하여 설계하였다. 그러나 DCT-IF 기반 업샘플링은 다른 선형 보간법이 갖는 고주파 성분 복원이 어렵다는 한계점을 가지고 있다.

3. SRCNN과 VDSR(Very Deep Super Resolution) 네트워크

SRCNN 네트워크^[7]는 3개의 계층으로 구성되어 있고 수용 영역(receptive field)은 13x13을 사용하는 구조이다. SRCNN은 사용하고자 하는 배율에 따라 네트워크를 변경해야 하는 단점이 존재한다. 그림 3은 제안하는 방법에서 사용한 VDSR 네트워크 구조이다. SRCNN은 사용하고자 하는 배율에 따라 각각의 네트워크에 학습이 필요하지만, VDSR은 여러 배율의 영상을 같이 학습시켜 하나의 모델 파라미터로 다양한 배율(x2, x3, x4)의 SISR이 가능하도록 설계가 되었다. 네트워크의 깊이를 20개의 계층으로 구성하였고, 수용 영역이 41x41로 SRCNN의 수용 영역 13x13 보다 10 배 정도 넓기 때문에 정보의 손실이 많은 고배율 영상 복원에 PSNR 측면에서 더 높은 성능을 보였다. SRCNN의 경우 RGB 색상에 대하여 학습을 진행하지만 VDSR의 경

우 YCbCr 색상으로 변환 후 Y 색상에 대해서만 SISR을 적용하였다. 또한, 잔차 신호가 예측 신호로 출력되기 때문에 예측 신호의 출력 범위가 좁아 빠르게 수렴하여 학습 시간을 줄일 수 있다.

다음 3장에서는 2장에서 설명한 기존 SISR 방법을 SHVC에 적용하여 계층간 참조 영상을 생성하는 방법에 대하여 보다 자세히 설명한다.

III. VDSR을 이용한 계층간 참조 픽처 생성 방법

1. 계층간 참조 픽처 생성 방법

그림 4는 제안하는 방법의 SHVC 부호화기 및 복호화기의 블록도이다. SHVC 부호화기 및 복호화기에서 입력 픽처를 DCT-IF 기반 방법으로 다운샘플링 하여 하위 계층의 입력 픽처로 사용한다. 상위 계층의 예측을 위해 하위 계층의 픽처를 업샘플링하고 업샘플링 된 픽처를 VDSR 네트워크를 통해 계층간 참조 픽처를 생성하여 상위계층에서 계층간 예측 신호로 사용한다. SHVC에서 하위계층의 복원된 픽처는 다양한 배율로 다운샘플링 될 수 있고 다양한 QP 값으로 부호화 될 수 있기 때문에 기존의 DCT-IF 기반 업샘플링 방법으로는 다양한 손실에 대한 고려가 불가능하다. 따라서 본 논문에서 제안하는 방법은 기존 배율에 대한 학습이 아닌 QP 값에 의한 에러를 VDSR 네트워크에 학습시킨 방

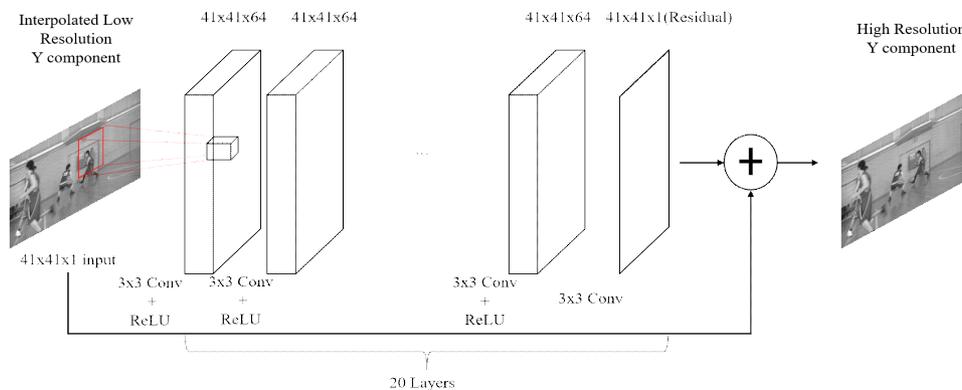


그림 3. 제안하는 방법에서 사용한 VDSR 네트워크 구조^[8]
 Fig. 3. The VDSR network architecture in the proposed method^[8]

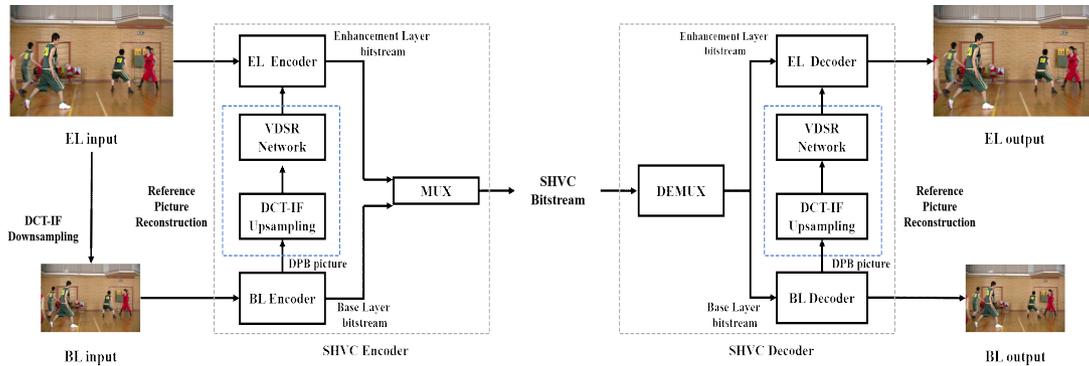


그림 4. 제안하는 방법의 부호화기 및 복호화기 블록도
 Fig. 4. The proposed codec flowchart

법과 이를 이용하여 SHVC 상위 계층에서 참조 픽처로 사용하는 방법을 제안한다. 본 논문에서 사용되는 VDSR 모델의 파라미터 필터 크기 및 개수, 계층의 깊이는 [8]에서 사용한 값을 차용하였다. VDSR은 입력 픽처의 고주파 성분에 대한 잔차 신호가 어느정도 복원되기 때문에 DCT-IF 업샘플링 방법에 비해 높은 PSNR 결과를 보인다. 이는 VDSR로 생성된 계층간 참조 픽처를 예측 샘플로 사용하였을 경우 기존 DCT-IF 보다 적은 잔차 신호가 발생한다. 따라서 계층간 예측 모드가 선택된다면 압축 효율이 증가하게 된다.

2. 학습 방법

본 논문에서 VDSR을 SHVC에 적용하기 위하여 다음과 같은 방법으로 학습을 수행하였다. 먼저, Ultra-Video Group의 120fps FHD 테스트 시퀀스를 SHM CTC(Common Test Conditions)^[14] 임의 접근(Random Access) 환경에서 QP 22, QP 27, QP 32, QP 37으로 각각 부호화 하고 하위 계층에서 복원된 픽처를 DCT-IF를 통해 업샘플링하여 VDSR의 입력을 생성하였다. VDSR의 정답으로는 원본 영상을 사용하였다. 저 지연(Low Delay) 환경에서 복원된 영상의 화질은 POC(Picture of Count)가 증가할수록 감소하기 때문에 결과적으로 화질에 대한 학습은 여러 QP 값에 대한 학습으로 보완이 가능하기 때문에 임의 접근 모드로 학습 데이터셋을 구성하였다. 또한 학습 데이터가 120fps 영상이기 때문에 CTC 환경에서는 128의 intra period를 갖는다. 따라서 64 프레임마다 한 장씩 추출하여 I 프레임과 B 프레임을 1:1 비율을 가지는 학습 데이터를 구성하였다. 학습에 사용

표 2. VDSR 학습에 사용한 시퀀스
 Table 2. Video sequences for training VDSR

FHD급 시퀀스 종류	패치 크기	학습 데이터의 수
Bosphorus	41x41	98,720
HoneyBee	41x41	98,720
Jockey	41x41	98,720
ShakeNDry	41x41	98,720
YachtRide	41x41	98,720
Total		462,750

한 패치 크기는 41x41을 사용하였다. 추가적으로 과적합 문제가 발생하지 않도록 VDSR 네트워크의 학습 방법과 마찬가지로 데이터 확장(data augmentation)을 적용하였다. 학습에 사용한 데이터의 개수는 462,750개를 사용하여 학습을 진행하였다. 표 2에 VDSR 학습에 사용한 학습데이터를 정리하였다. 기존 계층간 업샘플링 방법과 성능 비교를 위하여 휘도 성분에 대해서만 학습을 수행하였다. 학습은 식 (1) L2 손실 함수(Loss function)를 사용하여 원본 영상과 QP 값에 따라 부호화 된 후 DCT-IF 기반 업샘플링 된 픽처를 사용하여 진행하였다. 학습에 진행된 파라미터는 배치 사이즈(batch size)는 64를 사용하였고, 초기 학습률(learning rate) 0.0001 값을 사용하고 40 epoch 마다 0.1 배를 적용하였다. 학습에 사용된 옵티마이저(Optimizer)는 Adam(Adaptive Moment Estimation) 옵티마이저를 사용하였고, Adam 옵티마이저의 파라미터는 $\beta_1 = 0.9$, $\beta_2 = 0.999$ 를 사용하였다.

$$Loss = \sum_{i=1}^n (y_i - f(x_i))^2 \quad (1)$$

IV. 실험 및 결과

본 논문에서 제안하는 방법의 성능을 평가하기 위해 Windows 10 환경에서 Inter i7-8700k 3.7GHz 프로세서, GTX-1080Ti GPU 1개를 사용하였고 C++ 기반 SHM에서 Python 기반의 VDSR 네트워크를 호출하는 방식으로 실험을 진행하였다. 제안하는 부호화기의 성능을 평가하기 위

해서 MPEG 표준 테스트 시퀀스 표준 영상을 SHM CTC Spatial 2x 환경, RA(Random Access), AI(All Intra), LP (Low delay P)에 대하여 SHM 12.0을 이용하여 부호화를 진행하였다. 본 논문에서 제안하는 방법의 부호화 효율 평가를 위해 표준 SHM, 사전 학습 방법을 통하여 고주파 성분을 복원하는 방법^[12]과 비교하였다. 부호화 효율을 평가하기 위한 지표로 BD-BR(Bjontegaard-Delta Bitrate)을 사

표 3. MPEG 테스트 시퀀스에 대한 SHM 대비 사전학습 방법과 제안하는 방법의 BD-비트율 비교

Table 3. The average BD-rate reduction of the dictionary learning and the proposed methods compared to SHM in MPEG test sequences

구분	사전 학습[12]			제안하는 방법		
	AI	RA	LD	AI	RA	LD
BQTerrace	-2.25%	-0.89%	-0.33%	-15.39%	-7.62%	-5.56%
BasketballDrive	-1.28%	-1.37%	-1.19%	-8.75%	-6.91%	-5.39%
Cactus	-1.88%	-1.04%	-0.59%	-11.55%	-7.14%	-4.41%
Kimono	-0.58%	-0.59%	-0.35%	-4.12%	-3.32%	-3.55%
ParkScene	-1.26%	-1.05%	-0.52%	-6.07%	-3.87%	-2.72%
AVG	-1.45%	-0.99%	-0.60%	-9.18%	-5.63%	-4.56%

표 4. QP 별 테스트 시퀀스에 대한 SHM 대비 제안하는 방법의 성능비교

Table 4. Performance analysis between SHM and the proposed method in QP-based MPEG test sequences

시퀀스 종류		All Intra				Random Access				Low Delay P			
		EnT(%)		DeT(%)		EnT(%)		DeT(%)		EnT(%)		DeT(%)	
		CPU	GPU	CPU	GPU	CPU	GPU	CPU	GPU	CPU	GPU	CPU	GPU
BQTerrace	37	205	162	11546	4876	135	122	10971	4636	129	114	11879	4790
	32	193	154	10186	4395	118	113	10791	4425	121	111	11037	4436
	27	173	139	10399	3899	111	105	10069	3998	110	104	9682	3882
	22	167	135	9840	3707	112	104	10142	3965	110	107	9374	3767
	AVG	184	148	10493	4220	119	111	10493	4256	117	109	10493	4219
Basketball Drive	37	184	140	12834	5432	140	125	13224	5113	130	116	13511	5778
	32	178	141	12802	5238	129	116	12750	5034	124	113	12569	5286
	27	160	143	12232	4868	121	113	12309	4847	113	109	11316	4739
	22	143	120	11537	4362	109	106	11122	4394	112	107	9999	4065
	AVG	166	136	12351	4975	125	115	12351	4847	120	112	11849	4967
Cactus	37	192	162	15389	5878	141	121	14104	5853	129	117	16057	6294
	32	186	156	14285	5557	138	118	14271	5630	126	114	14860	5859
	27	176	137	13101	5288	125	113	14014	5326	113	104	13122	5183
	22	144	111	11510	4862	106	104	11894	4349	113	110	10244	4075
	AVG	174	141	13571	5396	128	114	13571	5290	120	111	13571	5353
Kimono	37	196	152	15667	6051	136	117	15418	6330	124	114	16554	6544
	32	191	149	14640	5758	130	113	14759	5994	117	114	15306	6105
	27	184	141	13989	5675	129	111	13984	5570	115	107	13507	5422
	22	168	153	13403	5591	116	108	13234	4987	111	103	12027	4859
	AVG	185	149	14425	5769	128	112	14349	5720	117	110	14349	5733
ParkScene	37	188	152	12308	5079	132	114	11488	4963	126	116	13059	5124
	32	182	138	11197	4465	128	109	10996	4574	120	117	11643	4609
	27	180	147	10405	4106	127	112	10897	4171	116	108	10135	4034
	22	151	137	9389	3661	114	107	9918	3586	113	104	8463	3384
	AVG	175	143	10825	4328	125	110	10825	4323	119	111	10825	4288
Total	177	143	11568	4938	126	112	10493	4887	119	111	12318	4912	

용하여 상위 계층의 비트율을 비교하였다. 표 3은 SHM 대비 비트율을 비교한 결과이다. 좌측은 사전 학습 방법을 이용한 결과이고 우측은 제안하는 방법의 결과이다. 기존 사전 학습을 이용 방법은 최대 2.25%의 비트율을 감소시켰다. 본 논문에서 제안하는 방법은 최대 15.39%의 비트율이 감소가 되었고, 평균 6.46%의 비트율 감소를 보였다. 또한 낮은 에너지의 고주파 성분이 넓게 분포되어 있는 특성을 갖는 Kimono와 ParkScene 영상에 대해서도 평균 4%의 감소율을 보였다. 표 4는 SHM대비 제안하는 방법의 실행 시간을 QP 값에 따라 비교한 결과이다. 실행 시간은 딥러닝 모듈 계산에 사용한 하드웨어에 따라 나누어 보여주었다. 제안하는 방법의 VDSR 네트워크는 부호화 및 복호화 과정에서 소요되는 시간은 동일하나 부호화 시간이 복호화 시간에 비해 오래 걸리기 때문에 복호화 과정에서 수행시간의 증가율이 높게 측정되었다. 실행 시간 측면에서 제안하는 방법은 부호화 과정과 복호화 과정 모두 동일한 시간이 소모되지만, 복호화 과정의 시간이 부호화 과정 시간에 비해 매우 짧기 때문에 복호화 시간이 증가되었다. VDSR 네트워크를 이용하여 필터링을 진행하였을 때 비트율 측면에서 감소되는 이득을 얻었다. 화면 간 예측이 진행되지 않는 AI 모드에서 높은 비트율 감소를 보였다.

V. 결론

본 논문에서는 SHVC 부호화 성능 개선을 위하여 딥러닝 기반 참조 픽처 생성 방법을 제안하였다. SHVC에서 DCT-IF 다운 샘플링과 부호화 열화된 픽처를 DCT-IF 업샘플링과 VDSR 네트워크를 이용하여 필터링을 진행하고 고주파 성분이 복원된 픽처를 상위 계층의 참조 픽처로 사용하는 방법이다. 기존 방법에서 화면 내 예측이 이루어지는 I 프레임에서만 사전 학습을 이용한 SISR이 적용된 문체를 보완하기 위해 본 논문에서 제안하는 방법은 화면 간 예측이 이루어지는 P, B 프레임에도 VDSR을 이용한 SISR을 사용을 하였다. VDSR 네트워크를 이용하여 필터링을 적용하였을 때 비트율 측면에서 감소되는 이득을 얻을 수 있었다. 결과적으로 기존 사전 학습을 이용한 방법 대비 상위 계층의 비트율을 최대 13.14%, 평균 6.46% 감소시키는 결과를

보였다. 본 논문에서 제안하는 방법은 영상에서 잡음과 같이 낮은 에너지의 고주파 성분이 넓게 분포되어 있는 특성으로 인하여 고주파 성분이 복원되지 못하는 영상 Kimono와 ParkScene에 대해서도 기존 사전 학습을 이용한 방법과 비교하여 AI 모드에서 평균 3%의 비트 감소율을 보였다. 제안하는 방법은 VDSR 네트워크를 이용하여 휘도 성분에 대해서 고주파 성분을 재구성한 픽처를 SHVC 상위 계층의 참조 픽처로 이용하는 방법이다. 추가적으로 색차 성분에 대해서도 딥러닝 기반 SISR을 적용 한다면 비트율을 더 감소시킬 수 있을 것이라고 예측된다. 그러나 딥러닝 네트워크를 사용함으로써 크게 증가되는 계산 복잡도에 대해서는 개선해야할 여지가 존재한다.

참고 문헌 (References)

- [1] H. Choi, K. Lee, J. Kang, S. Bae, and J. Yoo, "Overview and Performance Analysis of the Emerging Scalable Video Coding", *Journal of Broadcast Engineering* Vol.12, no. 6, pp. 542-554, November, 2007, <http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE00926642>.
- [2] J. M. Boyce, Y. Ye, J. Chen and A. K. Ramasubramanian, "Overview of SHVC: Scalable Extensions of the High Efficiency Video Coding Standard", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 26, no. 1, pp. 20-34, July, 2016, <https://doi.org/10.1109/TCSVT.2015.2461951>.
- [3] J. Lee, J. Kang, H. Lee and J. Choi, "SHVC Standard Technology Trend," *TTA Journal*, vol.152, pp. 60-65, March, 2015, <https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE06391006>.
- [4] Y. Kim, J. Choi and H. Choi, "Performance Analysis of Scalable HEVC Coding Tools", *JBE* Vol.20, no. 4, pp. 497-508, July, 2015, <https://doi.org/10.5909/JBE.2015.20.4.497>.
- [5] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via a sparse representation", *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861 - 2873, November 2010, <https://doi.org/10.1109/TIP.2010.2050625>.
- [6] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang, "Coupled dictionary training for image super-resolution", *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3467 - 3478, Aug 2012, <https://doi.org/10.1109/TIP.2012.2192127>.
- [7] D. Chao, L. Chen, H. Kaiming and T. Xiaou, "Image super-resolution using deep convolutional networks", *IEEE transactions on pattern analysis and machine intelligence*, Vol. 38, No.2. pp. 295-307, February, 2016, <https://doi.org/10.1109/TPAMI.2015.2439281>.
- [8] J. Kim, J. Lee and K. Lee, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, USA*, pp. 1646-1654, June, 2016, <https://doi.org/10.1109/CVPR.2016.182>.
- [9] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acost

- a, A. Aitken, A. Tejani, J. Totz, Z. Wang and W. Shi, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network", Proceedings of the the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Hawaii, USA, pp.4681- 4690, July, 2017, <https://arxiv.org/abs/1609.04802>.
- [10] B. Lim, S. Son, H. Kim, S. Nah, and K. Lee, "Enhanced Deep Residual Networks for Single Image Super-Resolution", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Hawaii, USA, pp.136-144, July, 2017, <https://arxiv.org/abs/1707.02921>.
- [11] W. Park and M. Kim, "CNN-based in-loop filtering for coding efficiency improvement", Proceedings of the IEEE Image, Video, and Multidimensional Signal Processing Workshop (IVMSP), IEEE, Bordeaux, France, July, 2016, <https://doi.org/10.1109/IVMSPW.2016.7528223>.
- [12] J. Schneider, J. Sauer, and M. Wien. "Dictionary learning based high frequency inter-layer prediction for scalable HEVC", Proceeding of IEEE Visual Communications and Image Processing (VCIP). IEEE, St. Petersburg, USA, pp.1-4, December, 2017, <https://doi.org/10.1109/VCIP.2017.8305019>.
- [13] J. Chen, J. Boyce, Y. Ye, M. Hannuksela and G. Barroux, "SHVC Test Model 11 (SHM 11)", Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, 113th Meeting, JCTVC-N15778, Geneva, Switzerland, pp. 1-12, October, 2015, <https://mpeg.chiariglione.org/standards/mpeg-h/high-efficiency-video-coding/n15778-scalable-hevc-shvc-test-model-11-shm-11>.
- [14] V. Seregin and Y. He, "Common SHM test conditions and software reference configuration", Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 17th Meeting, JCTVC-Q1009, Valencia, Spain, pp. 1-4, March, 2014, http://phenix.it-sudparis.eu/jct/doc_end_user/current_document.php?id=9106.

저 자 소 개



이 우 주

- 2018년 2월 : 광운대학교 전자공학과 학사
- 2018년 3월 ~ : 현재 광운대학교 전자공학과 석사과정
- ORCID : <https://orcid.org/0000-0002-3388-3765>
- 주관심분야 : 영상압축, 컴퓨터비전, 고해상도 위성영상 처리



이 종 석

- 2016년 2월 : 광운대학교 전자공학과 학사
- 2018년 2월 : 광운대학교 전자공학과 석사
- 2018년 3월 ~ 현재 : 광운대학교 컴퓨터공학과 박사과정
- ORCID : <https://orcid.org/0000-0001-8045-0244>
- 주관심분야 : 영상압축, 컴퓨터비전, 고해상도 위성영상 처리



심 동 규

- 1993년 2월 : 서강대학교 전자공학과 공학사
- 1995년 2월 : 서강대학교 전자공학과 공학석사
- 1999년 2월 : 서강대학교 전자공학과 공학박사
- 1999년 3월 ~ 2000년 8월 : 현대전자 선임연구원
- 2000년 9월 ~ 2002년 3월 : 바로비전 선임연구원
- 2002년 4월 ~ 2005년 2월 : University of Washington Senior research engineer
- 2005년 3월 ~ 현재 : 광운대학교 컴퓨터공학과 교수
- ORCID : <https://orcid.org/0000-0002-2794-9932>
- 주관심분야 : 영상신호처리, 영상압축, 컴퓨터비전

저 자 소 개



오 승 준

- 1980년 2월 : 서울대학교 전자공학과 학사
- 1982년 2월 : 서울대학교 전자공학과 석사
- 1988년 5월 : 미국 Syracuse University 전기/컴퓨터공학과 박사
- 1982년 3월 ~ 1992년 8월 : 한국전자통신연구원 멀티미디어연구실 실장
- 1986년 7월 ~ 1986년 8월 : NSF Supercomputer Center 초청 학생연구원
- 1987년 5월 ~ 1988년 5월 : Northeast Parallel Architecture Center 학생연구원
- 1992년 3월 ~ 1992년 8월 : 충남대학교 컴퓨터공학부 겸임교수
- 1992년 9월 ~ 현재 : 광운대학교 전자공학과 교수
- 2002년 3월 ~ 2017년 12월 : SC29-Korea 전문위원회 대표위원
- ORCID : <https://orcid.org/0000-0002-5036-3761>
- 주관심분야 : 비디오데이터처리, 영상압축, 컴퓨터비전