일반논문 (Regular Paper)
방송공학회논문지 제24권 제1호, 2019년 1월 (JBE Vol. 24, No. 1, January 2019)
https://doi.org/10.5909/JBE.2019.24.1.132
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

초협대역 비디오 전송을 위한 심층 신경망 기반 초해상화를 이용한 스케일러블 비디오 코딩

김 대 은[a], 기 세 환[a], 김 문 철[a]‡, 전 기 남[b], 백 승 호[b], 김 동 현[c], 최 증 원[c]

Scalable Video Coding using Super-Resolution based on Convolutional Neural Networks for Video Transmission over Very Narrow-Bandwidth Networks

Dae-Eun Kim[a], Sehwan Ki[a], Munchurl Kim[a]‡, Ki Nam Jun[b], Seung Ho Baek[b], Dong Hyun Kim[c], and Jeung Won Choi[c]

요 약

매우 제한된 전송 대역을 이용하여 비디오 데이터를 전송해야 하는 필요성은, 광대역을 통한 비디오 서비스가 활성화되어 있는 현 시점에서도 꾸준히 존재한다. 본 논문에서는 초협대역 네트워크를 통한 저해상도 비디오 전송을 위해, 공간 확장형 스케일러블 비디오 코딩 프레임워크에서 기본 계층의 부호화된 프레임을 심층 신경망 기반 초해상화 기법을 이용하여 업스케일링 하여 향상 계층 부호화 시에 예측 영상으로 활용하여 부호화 효율을 높이는 방법을 제안한다. 기존의 스케일러블 HEVC (High efficiency video coding) 표준에서는 고정된 필터로 업스케일링을 하는데 비해, 본 논문에서는 초해상화 수행을 위해 학습된 심층신경망을 기존의 고정 업스케일링 필터를 대체하여 적용하는 스케일러블 비디오 코딩 프레임워크를 제안한다. 이를 위해 스킵 연결과 잔차 학습 기법 등이 적용된 심층 콘볼루션 신경망 구조를 제안하고, 비디오 코딩 프레임워크의 실제 응용 상황에 맞추어 학습시켰다. 입력 해상도가 352×288이고 프레임 율이 8fps인 영상을 110kbps로 부호화 하는 응용 상황에서, 기존의 스케일러블 HEVC 프레임워크에 비해 제안하는 스케일러블 비디오 코딩 프레임워크의 화질이 더 높고 부호화 효율이 우수함을 확인할 수 있었다.

Abstract

The necessity of transmitting video data over a narrow-bandwidth exists steadily despite that video service over broadband is common. In this paper, we propose a scalable video coding framework for low-resolution video transmission over a very narrow-bandwidth network by super-resolution of decoded frames of a base layer using a convolutional neural network based super resolution technique to improve the coding efficiency by using it as a prediction for the enhancement layer. In contrast to the conventional scalable high efficiency video coding (SHVC) standard, in which upscaling is performed with a fixed filter, we propose a scalable video coding framework that replaces the existing fixed up-scaling filter by using the trained convolutional neural network for super-resolution. For this, we proposed a neural network structure with skip connection and residual learning technique and trained it according to the application scenario of the video coding framework. For the application scenario where a video whose resolution is 352×288 and frame rate is 8fps is encoded at 110kbps, the quality of the proposed scalable video coding framework is higher than that of the SHVC framework.

Keyword : Convolutional Neural Network, Super-Resolution, Scalable Video Coding, High Efficiency Video Coding

Ⅰ. 서 론

최근 고화질 비디오의 수요가 계속해서 증가함에 따라, 비디오 코덱의 중요성은 더욱 커지고 있다. 이러한 요구에 발맞추어 2013년 1월에 ISO/IEC의 Moving Picture Experts Group (MPEG)과 ITU-T의 Video Coding Experts Group (VCEG)이 공동으로 조직한 Joint Collaborative Team on Video Coding (JCT-VC)가 High Efficiency Video Coding [1] 표준을 완료하였고, 최근에는 MPEG과 VCEG이 공동으로 Joint Video Exploration Team (JVET)을 구성하여 새로운 비디오 부호화 표준인 Versatile Video Coding (VVC)의 표준안을 만들어가고 있다[2]. 한편, 심층 콘볼루션 신경망 기술이 다양한 분야에서 성공적인 결과를 보이고 있는데, 심층 콘볼루션 신경망을 이용하여 이미지 또는 비디오 코덱의 압축 성능을 개선하려는 시도 역시 다양하게 진행되고 있다[3-9]. 완전히 새로운 심층신경망 구조를 이용하는 구조[3-5]와 전통적인 하이브리드 코덱의 구조를 유지하면서 일부 모듈을 심층신경망 구조로 대체하여 부호화 효율을 높이는 방법[6-9] 등이 제안되었다. 본 논문에서는 공간 확장형 스케일러블 비디오 부호화[10][11] 프레임워크에 심층 콘볼루션 신경망 기반 초해상화 기술을 적용하여 비디오 부호화 성능을 향상시킬 수 있는 구조를 제안한다.

스케일러블 비디오 부호화 프레임워크는 다양한 대역폭, 단말기 등의 다채로운 서비스 환경을 효율적으로 처리하기 위하여 필요한 기술로서, 한번의 부호화만으로 다양한 서비스 시나리오를 수용할 수 있다. 특히, 공간적 확장의 (spatial-scalability) 경우에 기본 계층 (base layer)의 복원 영상을 향상 계층 (enhancement layer)의 예측 (prediction)으로 이용하여 부호화 효율을 높이는 계층간 예측 (inter-layer prediction)을 지원한다. HEVC 표준에서 지원하는 스케일러블 비디오 부호화[10](Scalable HEVC, SHVC) 역시 계층간 예측을 지원하고 있으며, 이를 가능하도록 하기 위해서는 기본 계층의 복원 영상의 해상도를 향상 계층의 해상도와 동일하도록 높여주는 업스케일링 (up-scaling) 알고리즘이 필요하다. SHVC 표준에서는 업스케일링 알고리즘이 휘도 신호와 색차 신호에 대해 각각 8탭, 4탭의 고정된 계수를 이용하는 필터로 정의되어 있다.

본 논문에서는 해당 업스케일링 알고리즘을 심층 콘볼루션 신경망 구조로 대체하여 부호화 효율을 향상시키는 새로운 프레임워크인 deepSHVC 프레임워크를 제안한다. 제안하는 deepSHVC 프레임워크의 우수성을 입증하기 위해 협대역 서비스 시나리오를 가정하여 기존의 SHVC 프레임워크와 deepSHVC 프레임워크의 부호화 효율을 비교한다. 협대역 서비스 시나리오에 대해서는 이어지는 4장에서 자세히 설명된다. 본 논문의 구성은 다음과 같다. 2장에서는 SHVC 프레임워크에 대해 설명한다. 3장에서는 제안하는 deepSHVC 프레임워크의 핵심인 심층 콘볼루션 신경망 기반의 초해상화 네트워크 구조에 대해 설명한다. 4장에서는 실험 환경을 기술하고 실험 결과를 제시한다. 5장에서는 본 논문의 결론을 맺는다.

Ⅱ. 스케일러블 HEVC (SHVC) 개요

스케일러블 비디오 부호화는 한번의 부호화만으로 다양한 서비스 수준을 제공할 수 있도록 하는 기술이다. 비디오 서비스 시나리오에는 대역폭 상황 등의 다양한 네트워크 환경과, 소비자의 단말기의 종류에 따른 다양한 수준의 서비스 형태가 있을 수 있다. 일반적으로 스케일러블 부호화에서는 시간적 확장성 (temporal scalability), 공간적 확장성 (spatial scalability), 및 화질 확장성 (quality scalability)을 지원한다. 스케일러블 비디오 부호화기를 통해 부호화하면 서비스 공급자가 원하는 형태의 시간적, 공간적, 화질의 서비스 수준으로 부호화된 비트스트림이 생성된다. 서비스 수준은 계층 (layer)이라고 표현하는데, 일반적으로 가장 상위 계층은 입력 비디오와 동일하며, 낮은 계층일수록 비디오의 자료 양이 적다. 예를 들어 입력 비디오가 3840×2160@60fps라고 하면, 계층 1 비디오는 1920×

a) 한국과학기술원 전기및전자공학부(The School of Electrical Engineering, Korea Advanced Institute of Science and Technology)
b) LIG 넥스원(LIG Nex1)
c) 국방과학연구소(Agency for Defense Development)
‡ Corresponding Author : 김문철(Munchurl Kim)
E-mail: mkimee@kaist.ac.kr
Tel: +82-42-350-7519
ORCID: http://orcid.org/0000-0003-0146-5419
※본 연구는 국방과학연구소가 지원하는 "초협대역 고효율 영상압축 기술개발" 사업의 일환으로 수행하였음(UC170016ED).
· Manuscript received October 24, 2018; Revised December 21, 2018; Accepted December 21, 2018.

1080@30fps으로, 계층 2 비디오는 1920×1080@60fps, 계층 3 비디오는 입력 비디오와 동일하게 3840×2160@60fps으로 설정하는 것 등이 가능하다. 예로 든 3개 계층의 입력 비디오가 동시에 스케일러블 비디오 부호화기에 입력되면, 하나의 비트스트림으로 부호화된다. 서비스시에는 비트스트림으로부터, 원하는 품질 수준의 계층을 추출하는 것이 가능하며, 복호화 하여 비디오를 계층 1, 2, 또는 3의 비디오를 복원한다.

본 논문에서는 특별히 공간적 확장성을 갖는 스케일러블 부호화 알고리즘에, 심층신경망 구조를 결합하여 부호화 성능을 향상시키는 프레임워크를 제안하였다. 그림 1은 공간적 확장성을 갖는 스케일러블 부호화의 개념을 나타낸다.

그림 1에는 공간적으로 기본 계층 (base layer)와 향상 계층 (enhancement layer)의 확장성을 갖는 부호화기의 구조가 나타나 있다. 그림 1에서 볼 수 있듯이, 기본 계층의 부호화 결과는 업스케일링 (up-scaling) 되어 향상 계층의 예측 (prediction) 영상으로 이용된다. 이 때, 기존의 SHVC 프레임워크에서는 휘도 신호에 대해 8탭 필터, 색차 신호에 대해 4탭 필터를 사용한다. 본 논문에서는 해당 업스케일링 알고리즘을 심층 콘볼루션 신경망 기반 초해상화 네트워크로 대체하여 부호화 효율을 높이는 방법을 제안한다. 업스케일링 블록을 심층 콘볼루션 신경망 기반 초해상화 블록으로 교체하는 것이므로 엔트로피 코딩을 수정할 필요는 없지만, 디코더에서도 역시 해당 업스케일링 블록을 교체하는 것이 필요하다.

Ⅲ. 심층신경망 업스케일링 기반의 스케일러블 비디오 부호화 프레임워크

본 논문에서는 스케일러블 비디오 부호화 프레임워크에 심층신경망 구조의 업스케일링 기술을 결합하여 부호화 효율을 높일 수 있는 기술을 제안한다. 그림 2는 제안하는 스케일러블 비디오 부호화 프레임워크에서 사용하는 심층신경망 기반의 업스케일링 알고리즘의 구조를 나타낸다. 심층신경망은 10개의 합성곱 계층으로 구성되어 있고, Leaky ReLU를 활성화 함수로 이용하였다. 잔차 학습 (residual learning) 기법을 이용하여, 영상의 고주파 성분을 잘 학습

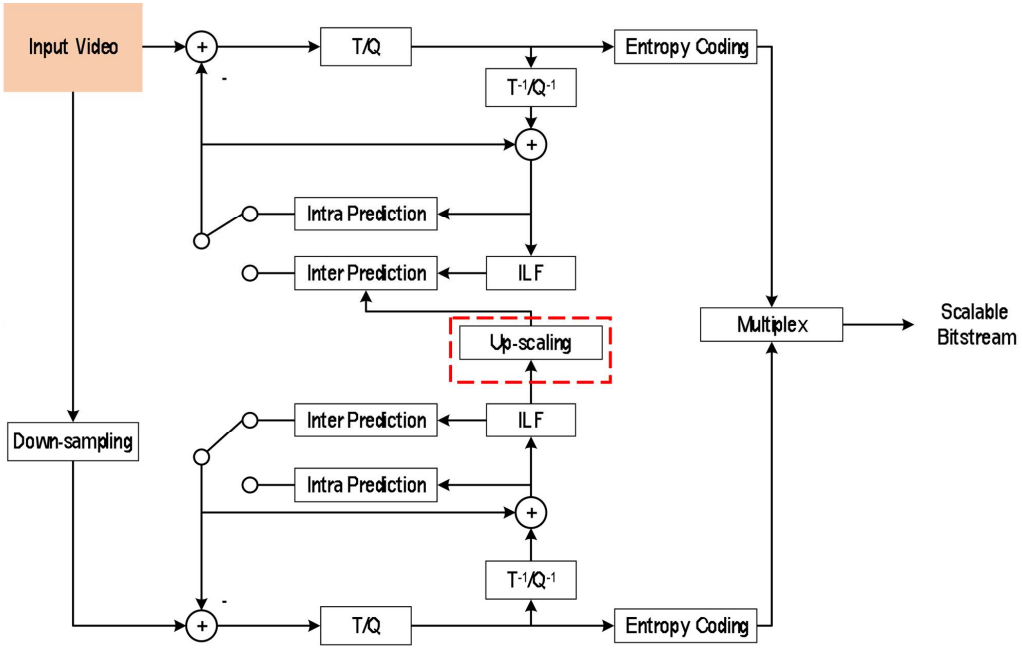


그림 1. 공간적 확장 스케일러블 부호화 알고리즘 개념도
Fig. 1. A block diagram of a scalable video coding with spatial scalability

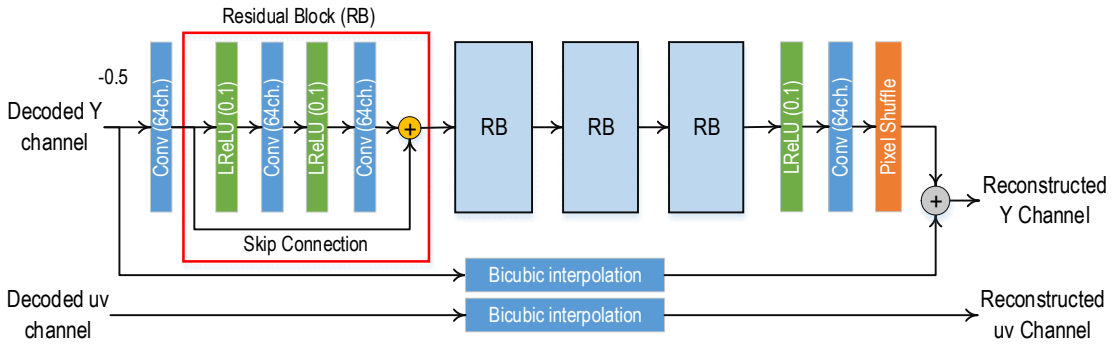


그림 2. 심층신경망 기반 업스케일링 네트워크 구조
Fig. 2. Structure of proposed up-sampling network based on a convolutional neural network

하게 함은 물론, skip 연결을 이용하여 학습을 용이하도록 하였다. 마지막으로 서브픽셀 셔플 계층[12]을 구성하여 계산 양을 줄이는 구조를 이용하였다.

그림 2에서 보듯이 제안한 심층신경망 기반 업스케일링 네트워크 구조는 휘도신호에 대해 적용되며, 색차신호는 단순히 겹삼차 보간(bicubic interpolation)에 의해 업스케일링된다. 휘도 신호에 대한 업스케일링 구조는 입력 저해상도 입력 휘도 영상을 겹삼차 보간에 의해 업스케일링하여 콘볼루션 계층을 통해 업스케일링하지 않고 동일한 해상도를 유지하여 처리하고 픽셀셔플 계층에서 업스케일링 배수의 제곱만큼 출력 채널 수를 출력하고 이를 최종 재정렬하여 고해상도 잔차 신호를 구성한다. 이렇게 구성된 고해상도 잔차신호는 겹삼차 보간된 입력 영상에 더해져 최종 고해상도 휘도 영상이 출력된다. 그림 2에서 보듯이 제안한 심층신경망 기반 업스케일링 구조는 4개의 잔차 블록(residual block)이 직렬로 연결되어 있고, 저해상도 입력 휘도 영상(decoded Y channel)는 곧바로 첫번째 잔차 블록에 입력되지 않고 1개의 콘볼루션 계층을 통과한 후 출력인 64채널 특징맵이 첫번째 잔차 블록에 입력된다. 4번째 잔차 블록의 출력인 64채널 특징맵은 활성함수(Leaky ReLU)를 통과한 후 마지막 콘볼루션 계층을 통과한 후 스케일 배수의 제곱 수의 특징맵이 출력되고 최종 픽셀셔플 계층을 통해 재정렬되어 고해상도 휘도 잔차신호를 구성한다. 앞에서 언급했듯이 색차 영상은 신호의 변화량이 휘도 영상에 비해 매우 낮으므로 복잡도를 고려하여 단순 겹삼차 보간(bicubic interpolation)에 의해 업스케일링된다.

제안된 심층신경망 기반 업스케일링 네트워크에서 (l+1)-번째 잔차 블록(RB, residual block)의 입력(x_l)과 출력(x_{l+1}) 관계는 다음과 같이 정의될 수 있다.

$$x_{l+1} = F(x_l) + x_l \tag{1}$$

여기서 $F(\bullet)$는 잔차 블록의 매핑 과정을 나타낸다. 식 (1)에서 보듯이 잔차 블록은 입력이 출력으로 스킵 연결(skip connection)되어 있으므로 입력(x_l)과 잔차 블록의 매핑 출력인 $F(x_l)$의 합이 잔차 블록의 최종 출력이 된다. 잔차 블록에 사용된 비선형 활성함수인 leaky ReLU는 다음과 같이 정의 된다[13].

$$LReLU(x) = \max(0.1x, x) \tag{2}$$

제안한 심층신경망 기반 업스케일링 네트워크를 훈련하기 위해 L1 손실 함수(loss function)를 아래와 같이 정의하여 사용한다.

$$L = \frac{1}{N}\sum_{p=0}^{N-1}\frac{1}{n}\sum_{i=0}^{n-1}\left|y_{i,p} - g(x_p)_{i,p}\right| \tag{3}$$

여기서 N은 미니배치(mini-batch)의 크기를 나타내며, x_p는 p-번째 저해상도(LR, low resolution) 영상 패치(image

patch)를 나타내며, $g(x_p)_{i,p}$는 입력 x_p에 대해 심층신경망 기반 업스케일링 네트워크에서 최종 출력된 고해상도(HR, high resolution) 영상 패치의 i-번째 화소 값을 나타낸다. 마지막으로 $y_{i,p}$는 저해상도 입력 x_p에 대응되는 레이블 영상 패치의 i-번째 화소 값을 나타낸다. 제안한 심층신경망 기반 업스케일링 네트워크에서 사용된 콘볼루션 필터의 크기는 모두 3×3이 사용되었다.

Ⅳ. 실험 환경 및 실험 결과

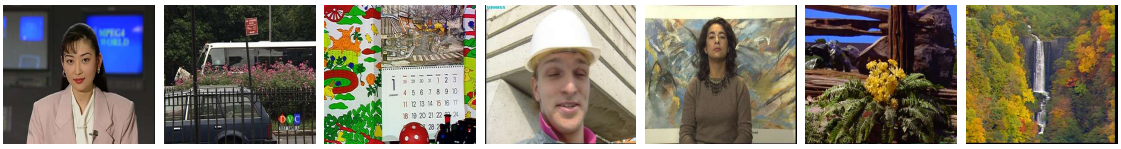
제안하는 deepSHVC 프레임워크의 부호화 성능을 검증하기 위하여 기존의 스케일러블 HEVC (SHVC) 프레임워크와 부호화 성능을 비교하였다.

1. 심층신경망 기반의 업스케일링 네트워크 학습

심층 콘볼루션 신경망 기반 초해상화 네트워크는 학습 기반의 알고리즘으로, 실제로 적용되는 상황을 가정하여 적절한 데이터 세트를 구성하여 학습해야 한다. 본 논문의 목적 응용 분야인 초협대역을 통한 비디오 전송을 위해 부호화 성능을 평가하기 위한 프로파일으로서 협대역 상황을 가정하였다. 총 110kbps의 대역폭내에서 기본 계층과 향상 계층이 전송되도록 설계 하였다. 초협대역 비디오 전송을 위한 프로화일로서, 실험 영상 해상도는 CIF(352×288)이고 프레임율은 8fps로 설정 되었다. 이러한 환경에 맞는 훈련 영상을 확보하기 위해, 온라인에서 다운로드 받은 영상들(Full HD, 1920×1080)[15][16]과 직접 촬영한 영상(Full HD, 1920×1080)을 포함하여 총 54 종류의 비디오를 수집한 후, 352×288@8fps에 맞게 변환 하였다. 이 후에 기본 계층 입력 비디오를 만들기 위해 352×288@8fps의 실험 비디오를 SHM 12.3 소프트웨어를 이용하여 176×144@8fps의 비디오로 다운샘플링하였다. 그림 1에서 업스케일 블록을 대체할 수 있도록 하기 위해서는 심층신경망 기반의 업스케일링 네트워크의 입력으로는 압축 왜곡을 포함하는 176×144@8fps 영상이, 출력으로는 압축 왜곡이 없는 352×288@8fps의 영상이 대응되어야 한다. 따라서 초해상화 네트워크를 학습하기 위해서 176×144@8fps 영상을 55kbps 비트율로 율 제어하여 부호화한 후, 각 프레임에 해당하는 352×288@8fps 영상을 라벨(정답) 영상으로 대응시켜 학습하도록 하였다. 이 때, 식 (3)의 L1 손실 함수를 아담 최적화기(Adam optimizer)[14]로 최소화 하도록 학습하였다. 학습을 위해 입력 영상에서 120×120 크기의 패치, 정답 영상에서 240×240 크기의 패치를 사용하였다. 학습을 위한 미니 배치(mini batch)의 크기는 2이고, 1 에폭(epoch)당 2,000번의 반복 (iteration)을 100 에폭까지 수행 하였다. 이때, 학습율 (learning rate)은 첫 50 에폭까지는 10^{-4}으로 설정하고, 50 에폭 이후에는 10^{-5}으로 설정하였다.

2. 부호화 성능 평가

제안하는 deepSHVC 프레임워크의 부호화 성능을 검증하기 위하여 협대역 상황의 전송 프로파일을 가정하였다. 해당 프로파일에서 영상의 해상도는 CIF(352×288)이고 프레임 율은 8fps 이며, 전송 목표 비트율은 110kbps이다. 제안하는 부호화기의 성능을 검증하기 위해 7 종의 실험 영상을 이용하였다. 실험 영상은 CIF의 해상도를 갖는 *akiyo*, *bus*, *calendar*, *foreman*, *silent*, *tempete*, *waterfall* 영상을 사용하였다. 그림 3에는 실험 영상들이 나타나 있다. 이 영상들은 프레임율이 25fps 또는 30fps 등인데, 8fps가 되도록 프레임을 서브샘플링 하였다. 전송 프로파일의 110 kbps 대역폭을 기본 계층과 향상 계층의 비트스트림의 비율에



(a) *akiyo* (b) *bus* (c) *calendar* (d) *foreman* (e) *silent* (f) *tempete* (g) *waterfall*

그림 3. 성능 평가에 이용된 실험 영상
Fig. 3. Test sequences for performance evaluation

따라 성능의 변화를 보기 위해 110kbps의 기본 계층과 향상 계층에 각각 (기본 계층 비트율, 향상 계층 비트율)로서 (25 kbps, 85kbps), (45kbps, 65kbps), (65kbps, 45kbps), (85 kbps, 25kbps)로 구성하여 실험하였다. 부호화 구조 (configuration)는 임의 접근(random access)으로 설정하였다. I-픽쳐 주기는 32이고 GOP의 크기는 8이다. 성능 비교 대상은 SHVC의 참조 소프트웨어인 SHM 12.3을 이용하였다.

표 1은 기존의 SHVC 프레임워크와 제안하는 deep-SHVC 프레임워크의 부호화 성능 비교를 나타낸다. 표 1에서, BL은 기본 계층의 부호화 결과, EL은 향상 계층의 부호화 결과를 나타내며 EL (SHM)은 기존의 SHVC 프레임워크로 부호화된 향상 계층, EL(deep)은 deepSHVC 프레임워크로 부호화된 향상 계층을 나타낸다. 기본 계층 부호화의 경우, 제안 방법과 기존의 방법 간에 차이가 없다. 표1에서 볼 수 있듯이, 제안 방법과 기존 방법의 향상 계층 부호화 결과를 비교하면, 거의 유사한 비트율로 부호화 되었음에도 불구하고 모든 실험 결과에서 제안 방법의 화질이 높은 것으로 나타났다(빨간 색으로 표시함). 파란색으로 표시한 부분은 제안 방법이 기존 방법에 비해 비트율을 더 적게 쓴 경우인데, 이 경우에도 제안 방법의 화질이 더 높았다.

표 1. 기존 SHVC 프레임워크와 제안 deepSHVC 프레임워크의 부호화 성능
Table 1. Coding performance of the conventional SHVC and the proposed deepSHVC framework

sequence		Bitrate (kbps)	Y-PSNR (dB)	Bitrate (kbps)	Y-PSNR (dB)	Bitrate (kbps)	Y-PSNR (dB)	Bitrate (kbps)	Y-PSNR (dB)
akiyo	BL	24.9358	43.0798	45.0903	46.0719	63.7136	47.9287	77.0638	48.9725
	EL (SHM)	84.327	43.8570	64.4488	42.9441	44.7906	41.5143	24.8671	39.0445
	EL (deep)	84.288	44.1635	64.9733	43.4997	44.6174	42.6068	25.0045	41.3146
bus	BL	29.0545	23.9282	45.5223	26.6642	65.7844	29.0068	84.736	30.9134
	EL (SHM)	84.7695	26.3697	65.0240	25.1711	47.6069	24.1405	29.3303	22.7313
	EL (deep)	84.2834	26.9856	64.4038	26.3330	44.7345	26.1270	24.8640	25.7896
Calendar	BL	26.9237	25.6719	45.9762	28.8066	65.0068	30.7991	85.1372	32.5718
	EL (SHM)	84.9225	26.8809	64.9475	25.6911	46.0020	23.9286	25.0576	21.4826
	EL (deep)	84.8390	27.1844	64.7227	26.6874	44.8757	26.0373	24.9054	25.0666
Foreman	BL	25.024	32.5017	45.8365	35.8586	65.2347	38.1267	85.1528	39.76
	EL (SHM)	85.0060	34.529	64.9748	33.5318	45.0068	32.2489	24.8554	29.9245
	EL (deep)	84.9647	34.8864	64.9795	34.6591	44.8507	34.5341	25.6983	33.915
Silent	BL	25.0537	34.3653	45.2129	37.8991	65.3167	40.4857	84.9514	42.4654
	EL (SHM)	85.4650	36.1513	65.2433	34.9187	45.1138	33.359	25.1770	31.1984
	EL (deep)	85.4548	36.6126	65.3643	36.316	45.0443	35.9862	25.0131	35.3174
tempete	BL	25.5225	28.3468	45.4139	31.2099	65.0844	33.0498	85.0758	34.604
	EL (SHM)	85.0028	29.1757	65.1601	28.1929	45.0524	26.7871	25.0655	24.9273
	EL (deep)	85.0551	29.5122	65.2503	29.0046	45.0839	28.4806	25.0465	27.7986
waterfall	BL	26.6402	33.6929	46.4297	36.511	65.6883	38.5037	85.3057	39.9708
	EL (SHM)	85.0461	34.6758	65.0123	33.5599	44.6594	32.0489	24.9952	29.902
	EL (deep)	85.0245	35.2398	65.6279	34.6349	45.4247	33.9184	24.4868	32.936
평균	BL	26.1649	31.6552	45.6403	34.7173	65.1184	36.8429	83.9175	38.4654
	EL (SHM)	84.9341	33.0913	64.9730	32.0014	45.4617	30.5753	25.6212	28.4587
	EL (deep)	84.8442	33.5121	65.0460	33.0192	44.9473	32.5272	25.0027	31.7340
	EL 차이	-0.0899	0.4208	0.073	1.0178	-0.5144	1.9519	-0.6185	3.2753

(a) SHM result (PSNR-Y: 23.2133 dB) (b) proposed result (PSNR-Y: 25.3815 dB)

그림 4. *bus* 영상의 향상 계층 65 kbps 부호화 결과 영상 (41번째 프레임)
Fig. 4. Decoded frame of *bus* of the enhancement layer at 65 kbps (41st frame)

(a) SHM result (PSNR-Y: 23.7715 dB) (b) proposed result (PSNR-Y: 25.8363 dB)

그림 5. *calendar* 영상의 향상 계층 45 kbps 부호화 결과 영상 (42번째 프레임)
Fig. 5. Decoded frame of *calendar* of the enhancement layer at 45 kbps (42nd frame)

(a) SHM result (PSNR-Y: 30.0045 dB) (b) proposed result (PSNR-Y: 31.3335 dB)

그림 6. *waterfall* 영상의 향상 계층 85 kbps 부호화 결과 영상 (64번째 프레임)
Fig. 6. Decoded frame of *waterfall* of the enhancement layer at 85 kbps (64th frame)

그림 4-6은 제안 방법과 기존 방법의 결과 영상을 비교하여 나타낸다. 그림 4에서 기존 방법의 결과 영상의 버스 앞쪽이 분리되어 보이는 반면 제안 방법의 결과 영상은 버스 앞쪽이 정상적으로 보인다. 배경도 기존 방법이 훨씬 블러(blur)한 것을 볼 수 있다. 그림 5에서는 장난감 기차 앞쪽 공이 기존 방법의 결과 영상에서는 찌그러져 있지만, 제안 방법의 결과 영상에서는 정상적으로 나타나 있음을 볼 수 있다. 마지막으로 그림 6에서도 폭포수 주변의 나무를 살펴보면, 제안 방법의 결과 영상에 비해 기존 방법의 결과 영상이 더 뭉개져 있는 것을 볼 수 있다.

Ⅴ. 결 론

본 논문에서는 심층 콘볼루션 신경망 구조의 초해상화를 이용한 스케일러블 비디오 부호화 (deepSHVC) 프레임워크를 제안하였다. 공간적 확장 스케일러블 비디오 부호화 프레임워크에는 기본 계층의 부호화 결과를 업스케일링 하여 향상 계층의 예측 영상으로 이용할 수 있다. 기존의 SHVC 표준에는 이 업스케일링 알고리즘을 고정된 계수의 필터로 처리하였지만, 본 논문에서는 심층 콘볼루션 신경망 구조의 초해상화 네트워크로 대체하였다. 제안하는 스케일러블 비디오 부호화 프레임워크의 성능을 측정하기 위해 협대역 상황의 프로파일을 가정하였고, 352×288@8fps의 7종류의 실험 영상에 대해 동일한 비트율일 때, 제안하는 스케일러블 비디오 부호화 프레임워크의 향상 계층의 결과 영상이 기존의 스케일러블 비디오 부호화 프레임워크의 향상 계층의 결과 영상에 비해 높은 객관적 화질을 나타냄을 확인할 수 있었다.

참 고 문 헌 (References)

[1] G. J. Sullivan, J.-R. Ohm, W.-J. Han, T. Wiegand, “Overview of the High Efficiency Video Coding (HEVC) standard,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, pp. 1648-1667, Dec. 2012.

[2] B. Bross, Working Draft 1 of Versatile Video Coding, document JVET-J1001, Joint Video Experts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, Apr. 2018.

[3] E. Agustsson, F. Mentzer, M. Tschannen, L. Cavigelli, R. Timofte, L. Benini, L. V. Gool, “Soft-to-hard vector quantization for end-to-end learning compressible representations,” *Proceeding of Advances in Neural Information Processing Systems*, Long beach, California, pp. 1141-1151, 2017.

[4] J. Ballé, V. Laparra, E. P. Simoncelli, “End-to-end optimized image compression,” *Proceeding of International Conference on Learning Representations*, Toulon, France, 2017.

[5] C.-Y. Wu, N. Singhal, P. Krähenbühl, “Video compression through image interpolation,” *Proceeding of European Conference on Computer Vision*, Munich, Germany, 2018.

[6] W.-S. Park, M. Kim, “CNN-based In-loop Filtering for Coding Efficiency Improvement,” *Proceeding of IEEE Image Video and Multidimensional Signal Processing (IVMSP) workshop*, Bordeaux, France, pp. 1-5, 2016.

[7] N. Yan, D. Liu, H. Li, F. Wu, “A convolutional neural network approach for half-pel interpolation in video coding,” *Proceeding of International Symposium on Circuits and Systems*, Baltimore, Maryland, pp. 1-4, 2017.

[8] D. Liu, H. Ma, Z. Xiong, F. Wu, “CNN-based DCT-like transform for image compression,” *Proceeding of International Conference on Multimedia Modeling*, Bangkok, Thailand, pp. 61-72, 2018.

[9] Z. Liu, X. Yu, Y. Gao, S. Chen, X. Ji, D. Wang, “CU partition mode decision for HEVC hardwired intra encoder using convolution neural network,” *IEEE Trans. Image Processing*, vol. 25, no. 11, pp. 5088-5103, Nov. 2016.

[10] H. Schwarz, D. Marpe, and T. Wiegand, “Overview of the scalable video coding extension of the H.264/AVC standard,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 9, pp. 1103 - 1120, Sep. 2007.

[11] J. M. Boyce, Y. Ye, J. Chen, A. K. Ramasubramonian, “Overview of SHVC: Scalable extensions of the High Efficiency Video Coding (HEVC) standard,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 20-34, Jan. 2016.

[12] W. Shi, J. Caballero, F. Huszár, J. Totz, A.P. Aitken, R. Bishop, D. Rueckert, Z. Wang, “Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, Nevada, pp. 1874-1883, 2016.

[13] A. L. Maas, A. Y. Hannun, A. Y. Ng, “Rectifier nonlinearities improve neural network acoustic models,” *Proceeding of International Conference on Machine Learning*, Atlanta, Georgia, p. 3, 2013.

[14] D. P. Kingma, J. L. Ba, “Adam: A method for stochastic optimization,” *Proceeding of International Conference for Learning Representations*, San Diego, California, pp. 1-41, 2015.

[15] Ultra Video Groupm, http://ultravideo.cs.tut.fi/#testsequences (accessed Jan. 2, 2019).

[16] SJTU 4K Video Sequence, http://medialab.sjtu.edu.cn/web4k/index.html (accessed Jan. 2, 2019).

저 자 소 개

김 대 은

- 2011년 2월 : 고려대학교 전기전자전파공학부 학사
- 2014년 2월 : 한국과학기술원 전기및전자공학과 석사
- 2014년 3월 ~ 현재 : 한국과학기술원 전기및전자공학부 박사과정
- ORCID : hjttp://orcid.org/0000-0003-0948-7049
- 주관심분야 : HDR 영상 처리, 비디오 부호화

기 세 환

- 2015년 2월 : 경북대학교 전기전자공학부 학사
- 2017년 2월 : 한국과학기술원 전기및전자공학과 석사
- 2017년 3월 ~ 현재 : 한국과학기술원 전기및전자공학부 박사과정
- ORCID : https://orcid.org/0000-0002-3809-7886
- 주관심분야 : 주관적 영상 압축 처리, 딥러닝 기반 영상 화질 개선

김 문 철

- 1989년 2월 : 경북대학교 전자공학과 학사
- 1992년 12월 : University of Florida, Dept. of Electrical and Computer Engineering, 석사
- 1996년 8월 : University of Florida, Dept. of Electrical and Computer Engineering, 박사
- 1997년 1월 ~ 2001년 1월 : 한국전자통신연구원 방송비디어 연구부, 선임연구원
- 2001년 2월 ~ 2009년 2월 : 한국정보통신대학교 공학부 조교수/부교수
- 2009년 3월 ~ 현재 : 한국과학기술원 전기및전자공학부 부교수/정교수
- ORCID : hjttp://orcid.org/0000-0003-0146-5419
- 주관심분야 : Perceptual Video Coding, SDR/HDR Image/Video Quality Assessment and Modeling, Super-Resolution, Image/Video Analysis and Understanding, Pattern Recognition, Machine Learning

전 기 남

- 2005년 2월 : 한양대학교 전자컴퓨터공학부 학사
- 2007년 2월 : 한양대학교 컴퓨터공학과 석사
- 2007년 1월 ~ 현재 : LIG 넥스원 수석연구원
- ORCID : https://orcid.org/0000-0003-1414-5297
- 주관심분야 : 센서 네트워크, 신호처리, 정보융합, 영상처리, 영상압축

백 승 호

- 2002년 2월 : 한국외국어대학교 정보통신공학부 학사
- 2004년 2월 : 한국외국어대학교 컴퓨터정보통신공학부 석사
- 2004년 3월 ~ 2009년 2월 : 한국전자통신연구원 연구원
- 2009년 3월 ~ 현재 : LIG 넥스원 수석연구원
- ORCID : https://orcid.org/0000-0002-6506-8861
- 주관심분야 : 센서 네트워크, 무인 지상감시센서, 영상압축, 무선통신

저 자 소 개

김 동 현

- 2009년 2월 : 연세대학교 전기전자공학부 학사
- 2011년 2월 : 연세대학교 전기전자공학부 석사
- 2011년 6월 ~ 현재 : 국방과학연구소 제2기술연구본부
- ORCID : hjttp://orcid.org/0000-0002-2136-5944
- 주관심분야 : 전술통신, 영상전송시스템, 데이터링크

최 증 원

- 1989년 2월 : 충남대학교 계산통계학과 학사
- 1993년 8월 : 충남대학교 계산통계학과(전산학) 석사
- 1997년 8월 : 충남대학교 전산학과 박사
- 1997년 7월 ~ 현재 : 국방과학연구소 수석연구원
- 2013년 9월 ~ 현재 : 과학기술연합대학원대학교 부교수
- ORCID : http://orcid.org/0000-0002-3642-2323
- 주관심분야 : 전술통신, 위성통신, 인지무선통신, 바이오통신, 정보융합 등