특집논문 (Special Paper)
방송공학회논문지 제27권 제1호, 2022년 1월 (JBE Vol.27, No.1, January 2022)
https://doi.org/10.5909/JBE.2022.27.1.31
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

신경망 기반 비디오 압축을 위한 레이턴트 정보의 방향 이동 및 보상

김 영 웅[a)], 김 동 현[b)], 정 세 윤[b)], 최 진 수[b)], 김 휘 용[a)‡]

Latent Shifting and Compensation for Learned Video Compression

Yeongwoong Kim[a)], Donghyun Kim[b)], Se Yoon Jeong[b)], Jin Soo Choi[b)], and Hui Yong Kim[a)‡]

요 약

전통적인 비디오 압축은 움직임 예측, 잔차 신호 변환 및 양자화를 통한 하이브리드 압축 방식을 기반으로 지금까지 발전해왔다. 최근 인공 신경망을 통한 기술이 빠르게 발전함에 따라, 인공 신경망 기반의 이미지 압축, 비디오 압축 연구 또한 빠르게 진행되고 있으며, 전통적인 비디오 압축 코덱의 성능과 비교해 높은 경쟁력을 보여주고 있다. 본 논문에서는 이러한 인공 신경망 기반 비디오 압축 모델의 성능을 향상시킬 수 있는 새로운 방법을 제시한다. 기본적으로는 기존 인공 신경망 기반 비디오 압축 모델들이 채택하고 있는 변환 및 복원 신경망과 엔트로피 모델(Entropy model)을 이용한 율-왜곡 최적화(Rate-distortion optimization) 방법을 사용하며, 인코더 측에서 디코더 측으로 압축된 레이턴트 정보(Latent information)를 전송할 때 엔트로피 모델이 추정하기 어려운 정보의 값을 이동시켜 전송할 비트량을 감소시키고, 손실된 정보를 추가로 전송함으로써 손실된 정보에 대한 왜곡을 보정한다. 이러한 방법을 통해 기존의 인공 신경망 기반 비디오 압축 기술인 MFVC(Motion Free Video Compression) 방법을 개선하였으며, 실험 결과를 통해 H.264를 기준으로 계산한 BDBR (Bjøntegaard Delta-Bitrate) 수치(%)로 MFVC(-14%) 보다 두 배 가까운 비트량 감축(-27%)이 가능함을 입증하였다. 제안된 방법은 MFVC 뿐 아니라, 레이턴트 정보와 엔트로피 모델을 사용하는 신경망 기반 이미지 또는 비디오 압축 기술에 광범위하게 적용할 수 있다는 장점이 있다.

Abstract

Traditional video compression has developed so far based on hybrid compression methods through motion prediction, residual coding, and quantization. With the rapid development of technology through artificial neural networks in recent years, research on image compression and video compression based on artificial neural networks is also progressing rapidly, showing competitiveness compared to the performance of traditional video compression codecs. In this paper, a new method capable of improving the performance of such an artificial neural network-based video compression model is presented. Basically, we take the rate-distortion optimization method using the auto-encoder and entropy model adopted by the existing learned video compression model and shifts some components of the latent information that are difficult for entropy model to estimate when transmitting compressed latent representation to the decoder side from the encoder side, and finally compensates the distortion of lost information. In this way, the existing neural network based video compression framework, MFVC (Motion Free Video Compression) is improved and the BDBR (Bjøntegaard Delta-Rate) calculated based on H.264 is nearly twice the amount of bits (-27%) of MFVC (-14%). The proposed method has the advantage of being widely applicable to neural network based image or video compression technologies, not only to MFVC, but also to models using latent information and entropy model.

Keyword: Neural network based Video Coding, Learned Image Compression, Learned Video Compression

I. 서 론

비디오 압축이란 비디오 영상을 보고 듣는 사람들이 불편함을 느끼지 않을 범위에서 신호를 압축함으로써 저장공간의 문제, 전송의 문제, 방송전파의 효율적 사용 등을 목적으로 하는 기술이다. 비디오 압축은 영상 압축과 음성 압축을 한 번에 일컫는 말로도 쓰일 수 있는데, 여기서는 연속된 영상을 압축한다는 의미로 한정한다. 압축은 크게 '손실압축'과 '무손실압축'으로 나눌 수 있으며, 손실압축은 영상의 비트량을 크게 줄여 저장 또는 전송의 효율성을 높일 수는 있으나 원래의 정보를 온전히 보존할 수 없는 방법이고, 무손실압축은 원래의 정보를 그대로 복원할 수 있는 반면에, 압축 성능은 비교적 떨어지는 방법이다. 오늘날 전통적인 표준 코덱은 H.264/AVC[1], H.265/HEVC[2], H.266/VVC[3] 등이 있으며 모두 손실압축 코덱이다. 이러한 표준 비디오 압축 코덱들은 예측 부호화 및 변환 부호화 방식을 사용한 복합 부호화를 통해 높은 압축률을 자랑한다.

최근에는 인공 신경망을 활용한 기술들이 급속하게 발전함에 따라 인공 신경망 기반의 비디오 압축(Learned Video Compression) 기술 연구가 빠르게 진행되고 있으며, 전통적인 코덱의 방식과 비교할 수 없을 정도로 짧은 시간 안에 큰 발전을 이루었다. 현재 여러 신경망 기반의 비디오 압축 모델들은 계산 복잡도를 제외하면 현재 널리 쓰이고 있는 비디오 압축 표준인 HEVC보다 더 나은 성능을 보여주고 있다. 인공 신경망 기반 비디오 압축 모델의 구조는 크게 두 가지 방향성을 가지는데, 그중 하나는 전통적인 코덱의 일부 모듈을 DNN (Deep Neural Network) 모듈로 대체하는 것이고, 나머지 하나는 전체 비디오 압축 프레임워크를 인공 신경망을 통해 구현하여 종단간 학습 방식(End-to-end training)으로 최적화하는 것이다.

본 논문에서는 종단간 학습 방식의 인공 신경망 기반 비디오 압축 모델의 성능을 향상시킬 수 있는 새로운 방법을 제시한다. 먼저, 기존 인공 신경망 기반의 이미지 압축 모델과 비디오 압축 모델에서 핵심이 되는 레이턴트 정보(Latent information)[4] 개념과 엔트로피 모델의 확률분포 추정의 의미에 대한 새로운 관점을 제시한다. 다음으로, 엔트로피 모델의 확률분포 추정의 한계를 보이고 이를 해결하기 위한 LDS (Latent Directional Shifting) 방법과 LCM (Latent Compensation Module)을 소개한다. 마지막으로 기존의 인공 신경망 기반 비디오 압축 기술인 MFVC (Motion Free Video Compression)[5]의 성능을 향상시킨 실험결과를 제시한다.

II. 관련 연구

1. 신경망 기반의 이미지 압축

최근 인공 신경망 기반의 이미지 압축 모델은 전통적인 영상 압축 코덱[6]과 비교했을 때 상당한 경쟁력을 가지게 되었다. 그것들은 오토인코더(Auto-encoder) 형태의 변환 및 복원 신경망을 통해 화소 단위의 이미지를 양자화된 레이턴트 정보 $\hat{y}$로 변환한 후 다시 화소 단위의 이미지로 복원한다. 이와 동시에 복원된 영상의 화질은 유지하면서 $\hat{y}$의 비트량을 줄이기 위한 양자화, 엔트로피 부호화를 진행한다. 비트량은 무손실압축 기법인 엔트로피 부호화에 의해 최종 결정되는데, 변환 및 양자화를 거친 $\hat{y}$의 엔트로피를 낮추기 위하여 신경망 기반의 엔트로피 모델을 함께 학습시켜 확률분포 추정에 사용한다. 종단간 학습 방식으로 이러한 신경망 구조를 최적화하기 위한 손실함수 L로는 수식 (1)에 나타낸 바와 같이 비트율 지표 R과 왜곡지표 D를 가중합하여 사용한다. 이때 비트율 R은 레이턴트 정보의 엔트로피로 계산되며, 왜곡 D는 원본 영상과 복원 영상간의 PSNR(Peak Signal-to-Noise Ratio) 또는 MS-SSIM (Multi-Scale Structural Similarity Index Measure)[7] 을 이용하여 계산한다. λ는 비트율과 화질(복원 오차) 사이의 중

a) 경희대학교(KyungHee University)
b) 한국전자통신연구원(ETRI)
‡ Corresponding Author : 김휘용(Hui Yong Kim)
E-mail: hykim.v@khu.ac.kr
Tel: +82-31-201-3760
ORCID:https://orcid.org/0000-0001-7308-133X
※ 이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2017-0-00072, 초실감 테라미디어를 위한 AV 부호화 및 LF 미디어 원천기술 개발).
· Manuscript received November 24, 2021; Revised December 21, 2021; Accepted December 21.

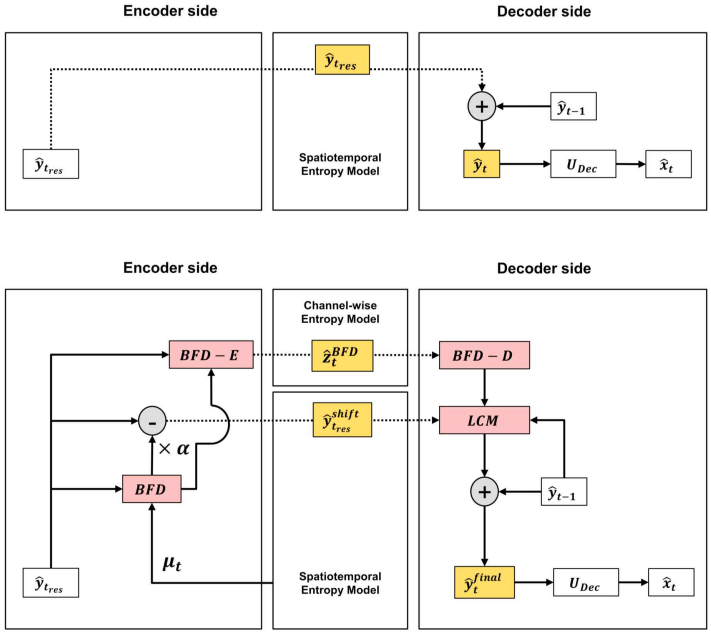

그림 1. 첫 번째 그림은 기존 MFVC 프레임워크에서 레이턴트 정보의 잔차 정보 $\hat{y}_{t_{res}}$ 를 전송하는 과정을 간략히 나타낸 것이고, 두 번째 그림은 본 논문에서 제안하는 방법을 간략히 나타낸 것으로, U_{Dec}는 통합 오토인코더(Unified auto-encoder)의 디코더를 의미한다. 특히 두 그림 모두 시공간 엔트로피 모델(Spatiotemporal entropy model)에서 사용하는 $\hat{z}_t$의 전송과정은 생략했다

Fig. 1. The first figure briefly shows the process of transmitting residual signals in the existing MFVC framework, and the second figure briefly shows the method proposed in this paper, and U_{DEC} means the decoder of the Unified auto-encoder. In particular, both figures omitted the transmission process for $\hat{z}_t$ used in the spatiotemporal entropy model

요도에 따라 그 비율을 결정해주는 상수로, λ값이 증가할수록 복원 왜곡이 줄어들지만 비트율이 증가하게 된다.

$$L = R + \lambda \times D(x, \hat{x}) \tag{1}$$

이러한 기본적인 틀을 마련한 Ballé의 논문[8]에서는 영상 압축에 효과적인 GDN (Generalized Divisive Normali- zation) 계층을 변환 신경망에 추가로 사용하였으며, $\hat{y}$의 각 성분 엔트로피를 독립적으로 최소화할 수 있는 엔트로피 모델을 사용하였다. Ballé의 이후 논문[4]에서는 초사전정보(Hyper-prior)라는 개념을 도입하여 입력 영상에 따라 가변적인 엔트로피 모델을 각각의 성분마다 평균이 0인 정규분포로 추정하는 방법을 제안했다. 이는 인코더에서 레이턴트 정보 $\hat{y}$ 와 함께 초사전정보 $\hat{z}$를 추가로 디코더에 전송함으로써 $\hat{y}$의 엔트로피를 더 효율적으로 낮추는 방법이다. Minnen의 논문[9]에서는 PixelCNN[10]을 사용하여 레이턴트 정보의 공간적인 중복성을 재귀적으로 제거하여 논문[4]에서 제시한 엔트로피 모델의 성능을 향상시켰다. 이러한 Minnen의 모델은 HEVC 비디오 코덱 기반의 이미지 코덱인 BPG[6]를 PSNR과 MS-SSIM 관점에서 모두 능가했다. Cheng의 논문[11]에서는 Minnen 모델의 변환 신경망을 잔차 블록(Residual block) 기반의 신경망[12]으로 대체하고, 이미지 복원 모델에서 사용되는 어텐션 모듈(Attention module)[13]을 추가하여 변환 성능을 높였다. 또한, 엔트로피 모델로 가우시안 혼합 모델[14]을 사용하여 더 정확한 확률분포 추정이 가능하게 했다. 앞선 결과들은 변환 신경망의 최적화와 더불어 $\hat{y}$의 확률분포를 정확하게 추정하는 것이 이미지 압축의 성능을 효과적으로 향상시킬 수 있음을 보여주고 있다. 그러나 이러한 방법들은 아직 이미지 압축 기술에 관한 것이고, 특히 성능 향상을 이룬 Minnen의 방법과 Cheng의 방법은 재귀적인 방법을 사용함에 따라 병렬처리가 어려운 한계가 존재한다.

2. 신경망 기반의 비디오 압축

DVC (Deep Video Compression)[15]는 최초의 종단간 학습 방식의 인공 신경망 기반 비디오 압축 모델로, 전통적인 비디오 압축 방법과 유사하게 변환 부호화, 움직임 예측 및 보상 등의 기술들을 CNN (Convolutional Neural Network)을 통해 구현했다. 해당 모델은 신경망 기반의 이미지 압축 모델들이 사용하고 있는 인트라 코딩(intra-frame coding)을 사용함과 동시에, 전통적 비디오 압축 코덱에서의 P-frame을 위한 인터 코딩(inter-frame coding)을 위한 추가적인 구조를 포함한다. P-frame을 고려한 구조를 추가하는 것은 이전 영상의 정보를 이용해 시간적인 중복성을 제거하여 효율적인 영상 압축을 수행하겠다는 것으로, 이미지 압축 모델에서 비디오 압축 모델로 이어지는 도약이라고 할 수 있다. DVC는 움직임 예측 모듈을 신경망 기반으로 구현하기 위해 광학 흐름 신경망(Optical flow network)인 SpyNet[16]을 사용하였으며, 인코더 측에서는 이전에 복원된 영상과 현재 영상 사이의 움직임 정보를 계산한 후 움직임 정보와 잔차 신호를 두 개의 변환 신경망으로 각각 변환 및 양자화하여 디코더에게 전송하고, 디코더는 변환된 신호를 통해 영상을 다시 복원한다. DVC 역시 전통적 비디오 코덱에서 사용하는 지표인 PSNR 또는 MS-SSIM와 엔트로피 기반의 손실 함수를 이용하였으며, 해당 모델은 널리 사용되는 비디오 압축 코덱인 H.264/AVC 보다 PSNR 과 MS-SSIM 관점에서 더 높은 성능을 기록하며 인공 신경망 기반의 비디오 압축 모델이 발전할 수 있는 기반이 되었다.

DVC 프레임워크가 공개된 후에는 양방향 움직임 예측을 사용한 모델[17], 양방향 움직임 예측 및 단방향 움직임 예측에 가중치를 사용한 HLVC (Hierarchical Leaned Video Compression)[18] 등이 등장하여 성능 향상을 이루었다. 하지만 이러한 DVC 프레임워크 기반의 모델들은 전통적 비디오 코덱의 구조를 기반으로 만들어졌기 때문에, 고질적인 문제인 오류 전파(Error propagation)가 발생하게 된다, 오류 전파란 복원 영상이 참조 영상으로 거듭 쓰이면서, 정보 손실이 누적되어 영상 내 심한 열화(Artifact)가 생기는 현상을 일컫는다. 이후 이러한 오류 전파 문제를 해결하기 위해서 다양한 방법이 등장했는데, 그 중 SSF(Scale-Space Flow) 모델[19]은 DVC 모델에서 광학 흐름 신경망에 의해 발생하는 잘못된 움직임 예측을 보완할 수 있도록 Scale-space flow라는 개념을 제시했다. 잘못된 움직임 예측을 보완하여 오류 전파를 다소 해결한 SSF 모델은 당시 SOTA(State Of The Art) 성능을 기록했다. 이후 Google AI Research의 2021년 논문[20]에서는 SSF 모델과 GAN (Generative Adversarial Network)[21] 기반의 이미지 압축 모델 HiFiC[22]을 결합한 모델을 제시하였다. 해당 모델에서는 오류 전파를 억제하기 위해 Randomized Shifting이라는 기술을 추가로 사용하였고, 사용자 평가(User study)에서 HEVC보다 높은 선택을 받았다. 하지만 오류 전파 문제를 해결하기 위해 복잡한 학습 방식을 선택했음에도 불구하고 역시 오류 전파 문제를 완전히 극복하진 못했다는 한계가 존재한다.

비슷한 시기에 Zhenhong의 논문[5]에서는 움직임 예측을 사용하는 전통적 비디오 코덱 기반의 방법 자체에 대한 문제 제기와 함께 MFVC (Motion Free Video Compression)를 제안했다. MFVC는 움직임 예측을 사용하지 않고, 인터 코딩을 수행할 때 엔트로피 모델의 시공간적인 중복성을 제거할 수 있도록 새로운 구조를 제안했다. 이를 위해 기존 영상 압축 모델[23] 기반의 통합 오토인코더(Unified Auto-encoder)를 변환 및 복원에 사용하였으며, 추가로 시공간 엔트로피 모델(Spatiotemporal entropy model) 구조를 사용했다. 그림 2에서 볼 수 있듯이 인트라 코딩 과정에서는 파란 선을 따라 단순히 압축 및 복원을 수행하지만, P-frame을 위한 인터 코딩을 수행할 때는 이전 영상과 현재 영상의 레이턴트 정보 $\hat{y}_{t-1}$, $\hat{y}_t$의 차이인 $\hat{y}_{t_{res}}$ 만을 디코더에 전송한다. $\hat{y}_{t_{res}}$의 확률분포를 추정하기 위해서는 이전 영상의 레이턴트 정보인 $\hat{y}_{t-1}$와 현재 영상의 레이턴트 정보 $\hat{y}_t$ 정보를 초사전정보 $\hat{z}_t$로 변환하여 디코더로 추가 전송하며, 이미 디코더가 가지고 있는 정보인 $\hat{y}_{t-1}$를 함께 이용한다. 디코더에서는 다시 $\hat{y}_t = \hat{y}_{t-1} + \hat{y}_{t_{res}}$를 계산하여 통합 오토인코더의 디코더를 통해 복원 영상 $\hat{x}_t$를 생성한다. 결과적으로 인코더에서 디코더에 전송해야 하는 정보는 레이턴트 공간(Latent space)에서의 잔차 신호인 $\hat{y}_{t_{res}}$ 와 초사전정보 $\hat{z}_t$이다. $\hat{y}_{t_{res}}$는 $\hat{z}_t$와 이전 시점의 latent인 $\hat{y}_{t-1}$를 이용한 성분별(Element-wise) 엔트로피 모델에 의해 엔트로피 부호

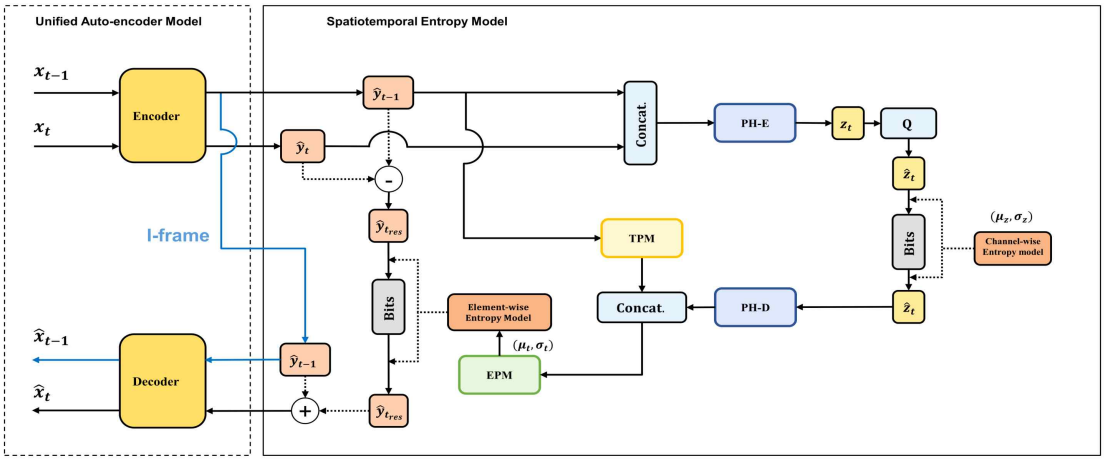

그림 2. MFVC (w/o SPM)의 기본 구조로, 영상 압축 및 복원을 위한 통합 오토인코더와 시공간적 중복성을 제거하여 엔트로피를 최소화하는 시공간 엔트로피 모델로 구성된다. 시공간 엔트로피 모델은 TPM (Temporal Prior Module), EPM (Entropy Parameter Module), PH-E/D (P-frame Hyper-prior Encoder/Decoder)를 포함한다. 특히 원래 MFVC 논문에서 제시한 SPM (Spatial Prior Module)은 제외한 구조이다

Fig. 2. The basic structure of the MFVC framework (w/o SPM) consists of a unified auto-encoder model for transform and inverse transform and a Spatiotemporal entropy model that minimizes entropy by eliminating spatiotemporal redundancy. It includes the Temporal Prior Module (TPM), the Entropy Parameter Module (EPM), and PH-E/D (P-frame Hyper-prior Encoder/Decoder). In particular, it is a structure that excludes the SPM (Spatial Priority Module) originally proposed in the MFVC paper

화가 이루어지고, $\hat{z}_t$는 학습 가능한 채널별(Channel-wise) 파라메터 μ_z, σ_z를 가진 엔트로피 모델에 의해 엔트로피 부호화가 이루어진다. 엔트로피 코딩에 실제로 사용하는 확률분포는 라플라스 분포로, 정규 분포를 사용하는 것과 유사하게 동작하며 수렴 속도가 빠르다[5]. 이러한 방법으로 최종 엔트로피와 손실 함수는 (2), (3)과 같이 나타낼 수 있다. 식 (2)는 각각 $\hat{y}_{t_{res}}$와 $\hat{z}_t$에 대한 추정 확률 모델이며, 식 (3)은 이러한 확률 모델을 통해 엔트로피 부호화가 이루어질 때 계산되는 엔트로피이다. 이렇게 계산되는 엔트로피는 엔트로피 부호화 이후의 비트스트림이 실제 발생시키는 비트량에 근사하게 된다. 식 (2)의 첨자 i는 레이턴트 정보의 각 성분을 의미하며, 레이턴트 정보의 각 모든 성분의 확률 모델이 독립적이라는 가정하에 전체 레이턴트 정보의 확률 모델은 각 성분의 확률 모델로 인수분해(Factorization)[8]한 것으로 표현된다. 이렇듯 MFVC는 인터 코딩을 수행할 때 엔트로피 모델의 시공간적 중복성만을 제거하기 때문에 복원 영상에서의 오류 전파 문제가 전혀 발생하지 않게 되며, 그럼에도 비디오 압축 성능은 PSNR 기준으로 SOTA 성능에 근접하고, MS-SSIM 기준으로는 SOTA 성능을 기록했다. 본 논문에서는 이러한 MFVC를 기반으로 새로운 기술을 적용한 실험을 진행하였다.

$$p_{\hat{y}_{t_{res}}}(\hat{y}_{t_{res}}|\hat{y}_{t-1},\hat{z}_t)=\prod_{i=1}\left(\int_{\hat{y}_i-\frac{1}{2}}^{\hat{y}_i+\frac{1}{2}} Lap(\hat{y}_{t_{res}};\mu_t,e^{\sigma_t})\,dy\right)$$
$$p_{\hat{z}_t}(\hat{z}_t)=\prod_{i=1}\left(\int_{\hat{z}_i-\frac{1}{2}}^{\hat{z}_i+\frac{1}{2}} Lap(\hat{z}_t;\mu_z,e^{\sigma_z})\,dz\right) \quad (2)$$

$$L_p=R_{\hat{y}_{t_{res}}}+R_{\hat{z}_t}$$
$$=E_{x\sim p_x}[-\log_2 p_{\hat{y}_{t_{res}}}(\hat{y}_{t_{res}})]+E_{x\sim p_x}[-\log_2 p_{\hat{z}_t}(\hat{z}_t)] \quad (3)$$

Ⅲ. 본 론

1. 문제 제기

$\hat{y}_{t_{res}}$의 확률분포를 추정하는 엔트로피 모델은 라플라스 분포의 정적분 형태로, 이 라플라스 분포의 입력 파라메터인 (μ_t, σ_t)는 $\hat{y}_{t-1}$, $\hat{z}_t$ 정보와 여러 CNN 기반 모듈에 의해 계산된다. 그림 4에서 볼 수 있듯이, 실제 발생한 $\hat{y}_{t_{res}}$와 가

깝게 μ_t를 추정한다면 발생하는 비트량은 적게 발생하며, 더 멀리 추정할수록 비트량이 크게 발생한다. 특히, 엔트로피 모델이 양자화를 고려하여 설계되었기 때문에, 추정된 μ_t와 실제 발생한 $\hat{y}_{t_{res}}$ 값의 차이가 0.5보다 작을 경우 비트량을 거의 발생시키지 않고, 0.5보다 클 경우 비트량을 크게 발생시킨다. σ_t는 예측에 대한 확신(confidence)으로 생각할 수 있으며, μ_t와 함께 비트량에 일부 관여한다. 하지만 역시 가장 중요한 것은 실제 발생한 $\hat{y}_{t_{res}}$와 가깝게 μ_t를 추정하는 것이다.

MFVC는 앞서 설명했듯이 오류 전파 문제를 완전히 해결한 구조를 가지며, 추가로 레이턴트 정보의 잔차 신호만을 전송하기 때문에, $\hat{y}_{t_{res}}$에서 발생하는 값은 대부분 -1, 0, 1 값 중 하나로, 발생하는 값의 범위를 줄이는 역할을 했다. 하지만 여전히 기존 비디오 압축 프레임워크의 엔트로피 모델과 동일하게 사전정보 $\hat{z}_t$를 사용하며, $\hat{z}_t$는 그림 2와 같이 PH-E (P-frame Hyper-prior Encoder) 모듈에 $\hat{y}_{t-1}$와 $\hat{y}_t$를 입력으로 넣어 출력한다. $\hat{z}_t$의 너비와 높이는 $\hat{y}_{t_{res}}$보다 16배 작아서, $\hat{z}_t$의 정보를 가지고 $\hat{y}_{t_{res}}$의 모든 성분의 값을 정확히 추정하기 어렵게 되고, 그림 4의 두 번째 행의 첫 번째 그림에서 볼 수 있듯이 일부 성분에서 비트량을 크게 발생시키게 된다.

2. 제안된 방법

이러한 문제를 해결하기 위해 생각할 수 있는 간단한 방법은 사전정보 $\hat{z}_t$의 너비와 높이를 크게 만들어 디코더 측에서 고해상도의 정보를 이용할 수 있도록 해주는 것이다. 하지만 그렇게 되면 $\hat{y}_{t_{res}}$의 발생 비트량인 $R_{\hat{y}_{t_{res}}}$는 줄어들지만 $\hat{z}_t$의 발생 비트량인 $R_{\hat{z}_t}$가 더 많이 증가하기 때문에 결론적으로는 비트량 이득을 얻을 수 없게 된다. 본 논문에서는 사전정보를 전송하여 더 정확한 확률 모델을 만드는 일반적인 방법이 아닌, 인코더 측에서 $\hat{y}_{t_{res}}$를 추정된 확률 모델에 적합한 방향으로 이동시켜 디코더로 전송하는 방법을 제안한다. 이때 확률 모델에 적합한 방향이란, 비트량을 많이 발생시키는 성분을 추정된 μ_t에 가깝게 이동시키는 것을 말하며, 본 논문에서는 LDS (Latent Directional Shift)라고 한다. LDS를 수행하는 방법은 여러 가지가 있지만, 한 가지 예시로 BFD_t (Big Fault Direction)라는 개념을 사용한다. BFD_t의 수식은 (4)와 같이 표현된다.

$$BFD_t = \begin{cases} 0 & \text{if } Abs(\hat{y}_{t_{res}} - \mu_t) < 0.5 \\ Sign(\hat{y}_{t_{res}} - \mu_t) & elsewhere \end{cases} \tag{4}$$

$$\begin{aligned} & p_{\hat{y}_{t_{res}}}(\hat{y}_{t_{res}} | \hat{y}_{t-1}, \hat{z}_t) \\ &= \prod_{i=1} \left(\int_{\hat{y}_i - \frac{1}{2}}^{\hat{y}_i + \frac{1}{2}} Lap(\hat{y}_{t_{res}} - \alpha BFD_t ; \mu_t, e^{\sigma_t})\, dy \right) \\ &= \prod_{i=1} \left(\int_{\hat{y}_i - \frac{1}{2}}^{\hat{y}_i + \frac{1}{2}} Lap(\hat{y}_{t_{res}}^{shift} ; \mu_t, e^{\sigma_t})\, dy \right) \end{aligned} \tag{5}$$

식을 살펴보면 현재 시점의 BFD_t는 실제 $\hat{y}_{t_{res}}$ 값과 추정값 μ_t의 차이가 0.5 미만인 경우에는 0, 그렇지 않은 경우 1 또는 -1의 값을 가진다. 즉, 확률 모델의 입장에서 μ_t 추정에 크게 실패한 성분에 대한 방향 정보라고 할 수 있다. 이러한 BFD_t는 $\hat{y}_{t_{res}}$ 값과 μ_t를 알고 있으면 계산할 수 있으며, 일반적으로 인코더는 알 수 있지만, 디코더는 알 수 없는 정보이다. 인코더 측에서 디코더에서 $\hat{y}_{t_{res}}$를 전송할 때 그림 1의 하단 그림과 같이 BFD_t를 일정량(α) 상수 배 곱하여 이동시킨 후 디코더에게 전송하면, 확률 모델에서 μ_t 추정이 크게 빗나간 성분의 비트량을 감소시킬 수 있다. 실제 감소한 비트량은 그림 4의 두 번째 행에서 확인할 수 있으며, α가 0.5인 경우와 α가 1인 경우 모두 비트량이 크게 감소했다. 이렇게 LDS를 거친 레이턴트 정보의 의 확률 모델은 식 (5)와 동일하게 생각할 수 있다.

하지만 이러한 방법을 사용하게 되면 그림 2의 MFVC 구조에서, $\hat{y}_t = \hat{y}_{t-1} + \hat{y}_{t_{res}}$ 연산과 통합 오토인코더의 디코

$$\begin{aligned} L_p &= R_{\hat{y}_{t_{res}}} + R_{\hat{z}_t} + R_{\hat{z}_t^{BFD}} + \lambda \times D(x_t, U_{DEC}(\hat{y}_t^{final})) \\ &= E_{x \sim p_x}[-\log_2 p_{\hat{y}_{t_{res}}}(\hat{y}_{t_{res}})] + E_{x \sim p_x}[-\log_2 p_{\hat{z}_t}(\hat{z}_t)] + E_{x \sim p_x}[-\log_2 p_{\hat{z}_t^{BFD}}(\hat{z}_t^{BFD})] + \lambda \times D(x_t, U_{DEC}(\hat{y}_t^{final})) \end{aligned} \tag{6}$$

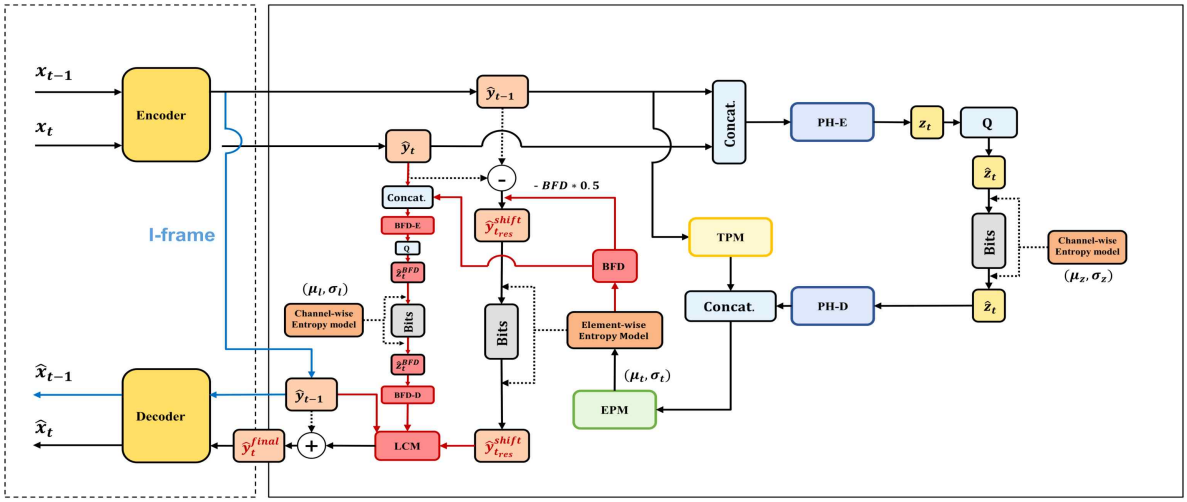

그림 3. 기존 MFVC에 본 논문에서 제시하는 방법을 추가한 전체 구조로, BFD (Big Fault Direction)를 추가로 계산하며 BFD-E/D (BFD Encoder/Decoder), LCM (Latent Compensation Module)이 추가되었다

Fig. 3. With the structure of adding the method presented in this paper to the existing MFVC framework, BFD (Big Fault Direction) is additionally calculated and BFD-E/D (Encoder/Decoder) and Latent Compensation Module (LCM) are added

터를 이용하여 복원 영상 $\hat{x}_t$를 생성할 수 없게 된다. 따라서 그림 3과 같이 LCM과 BFD-E/D (BFD Encoder/Decoder) 모듈을 도입하여, BFD_t에 의해 손실된 정보를 추가로 전송하여 보상하게 된다. BFD-E/D는 PH-E/D와 같은 구조로 설계하였으며, 자세한 네트워크 구조는 표 1에서 찾아볼 수 있다. BFD-E는 현재 시점의 BFD_t 정보와 $\hat{y}_t$ 정보를 입력으로 받아 변환 및 양자화를 거쳐 최종적으로 $\hat{z}_t^{BFD}$를 출력한다. 이때도 $\hat{z}_t$와 마찬가지로 채널별 엔트로피 모델을 사용하여 디코더에게 전송하게 되고, BFD-D는 인코더로부터 받은 추가 정보 $\hat{z}_t^{BFD}$를 복원하여 LCM의 입력으로 사용한다. 디코더에서는 이렇게 전송된 추가 정보 $\hat{z}_t^{BFD}$와 함께 $\hat{y}_{t_{res}}^{shift}$, $\hat{y}_{t-1}$를 LCM의 입력으로 넣고, 출력을 $\hat{y}_{t-1}$와 더해서 $\hat{y}_t^{final}$을 계산한다. 마지막으로 통합 오토인코더의 디코더에 $\hat{y}_t^{final}$를 입력으로 주어 복원 영상 $\hat{x}_t$를 생성한다. 따라서, 총 엔트로피는 식 (6)과 같다. 식 (6)에서 U_{DEC}는 통합 오토인코더의 디코더를 의미하며, $D(\bullet,\bullet)$은 MSE (Mean Square Error) 또는 MS-SSIM을 사용한다. 전체적인 모델 구조는 그림 3에서 확인할 수 있으며, 핵심 구조 요약은 그림 1에서 확인할 수 있다.

한 가지 문제점은 t=3부터 다음 영상의 압축 및 복원을 위해 $\hat{y}_{t-1}$가 아닌 $\hat{y}_{t-1}^{final}$를 참조해야 한다는 것이다. 하지만 $\hat{y}_{t-1}^{final}$는 이미 LCM에 의해서 노이즈가 섞인 신호이기 때문에 $\hat{y}_{t-1}^{final}$를 참조하여 연속적으로 영상을 압축 및 복원할 경우 좋은 성능을 기대하기 어렵다. 따라서 $\hat{y}_{t-1}^{ref} =$

표 1. 모듈 상세 구조

Table 1. Module detail structure

P-HE	P-HD	TPM	EPM	BFD-E	BFD-D	LCM
Conv: 3 × 3, N, s1 Leaky ReLU Conv: 5 × 5, N, s2 Leaky ReLU Conv: 5 × 5, N, s2	TConv: 5 × 5, N, s2 Leaky ReLU TConv: 5 × 5, N, s2 Leaky ReLU Conv: 3 × 3, 2N, s1	Conv: 5 × 5, 4N/3, s1 Leaky ReLU Conv: 5 × 5, 5N/3, s1 Leaky ReLU Conv: 5 × 5, 2N, s1	Conv: 1 × 1, 10N/3, s1 Leaky ReLU Conv: 1 × 1, 8N/3, s1 Leaky ReLU Conv: 1 × 1, 2N, s1	Conv: 3 × 3, N, s1 Leaky ReLU Conv: 5 × 5, N, s2 Leaky ReLU Conv: 5 × 5, N, s2	TConv: 5 × 5, N, s2 Leaky ReLU TConv: 5 × 5, N, s2 Leaky ReLU Conv: 3 × 3, 2N, s1	Conv: 3 × 3, 3N, s1 Leaky ReLU Conv: 3 × 3 2N, s1 Leaky ReLU Conv: 1 × 1 N, s1

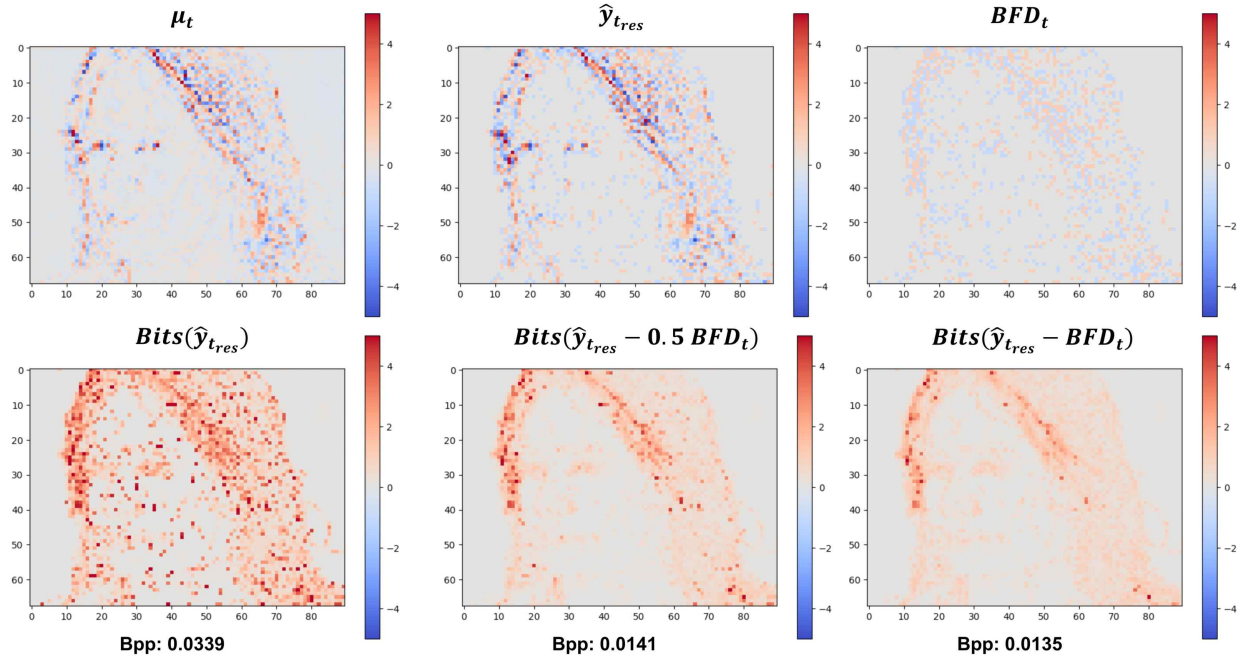

그림 4. 첫 번째 행은 UVG 데이터셋 중 비디오 'Beauty'의 두 번째 영상에 대한 $\hat{y}_{t_{res}}$, μ_t, BFD_t의 19번째 채널을 시각화한 것이고, 두 번째 행은 동일한 확률 모델을 사용하면서 BFD_t 정보를 이용했을 때와 이용하지 않았을 때의 발생 비트량을 시각화한 것이다. 비트량은 전체 채널에서 발생한 비트량의 총합을 구한 후 픽셀의 수로 나누어 Bpp (Bitrate per pixel)를 따로 계산하여 아래 표시했다

Fig. 4. The first row is a visualization of the 19th channel of $\hat{y}_{t_{res}}$, μ_t, and BFD_t for the second image of the video 'Beauty' in the UVG datasets. And the second row is a visualization of the amount of bits generated when BFD_t is used and not used. The bits was calculated by calculating the total amount of bits generated in the entire channel and dividing it by the number of pixels, and then Bpp (Bitrate per pixel) was calculated and displayed below

$U_{ENC}(\hat{x}_{t-1}) = U_{ENC}(U_{DEC}(\hat{y}_{t-1}^{final}))$를 참조하도록 한다. 이때, U_{ENC}는 통합 오토인코더의 인코더이다. 이렇게 다시 추가적인 변환을 수행하면 계산량이 추가로 요구되지만, $\hat{y}_{t-1}^{final}$에 존재하는 노이즈를 없애 연속적으로 영상을 압축 및 복원할 수 있게 된다.

Ⅳ. 실험 및 결과

MFVC는 통합 오토인코더와 시공간 엔트로피 모델로 구성되며, 통합 오토인코더를 먼저 학습한 후 파라메터를 고정한 뒤 시공간 엔트로피 모델을 학습한다[5]. MFVC를 제시한 논문[5]에서는 Minnen의 모델[9]을 기반으로 조건부 컨볼루션 층(Conditional convolutional layer)[24]을 적용하여 여러 화질 수준에 모두 대응할 수 있는 한 개의 모델을 통합 오토인코더로 사용하였다. 본 논문에서는 신경망 기반의 이미지 압축 모델들을 API 형태로 제공하고 있는 Compressai[25]의 사전 학습된 Cheng의 모델[11]을 통합 오토인코더로 선택했으며, 학습 속도를 고려하여 어텐션 모듈을 사용하지 않는 가벼운 모델을 사용하였다. 또한, 총 6개의 화질 수준에 따라 6개의 모델을 각각 사용하였다. 시공간 엔트로피 모델은 직접 학습하여 사용하였으며, 모듈 상세 구조는 표 1에서 확인할 수 있다. 모델의 채널 수인 N은 낮은 화질 수준을 갖는 3개의 모델에 대해서는 128, 높은 화질 수준을 갖는 3개의 모델에 대해서는 192로 설정하였다. 원래 논문[5]에서는 하나의 시공간 엔트로피 모델을 여러 화질 수준에 동시에 최적화하였지만, 본 논문에서는 더 정확한

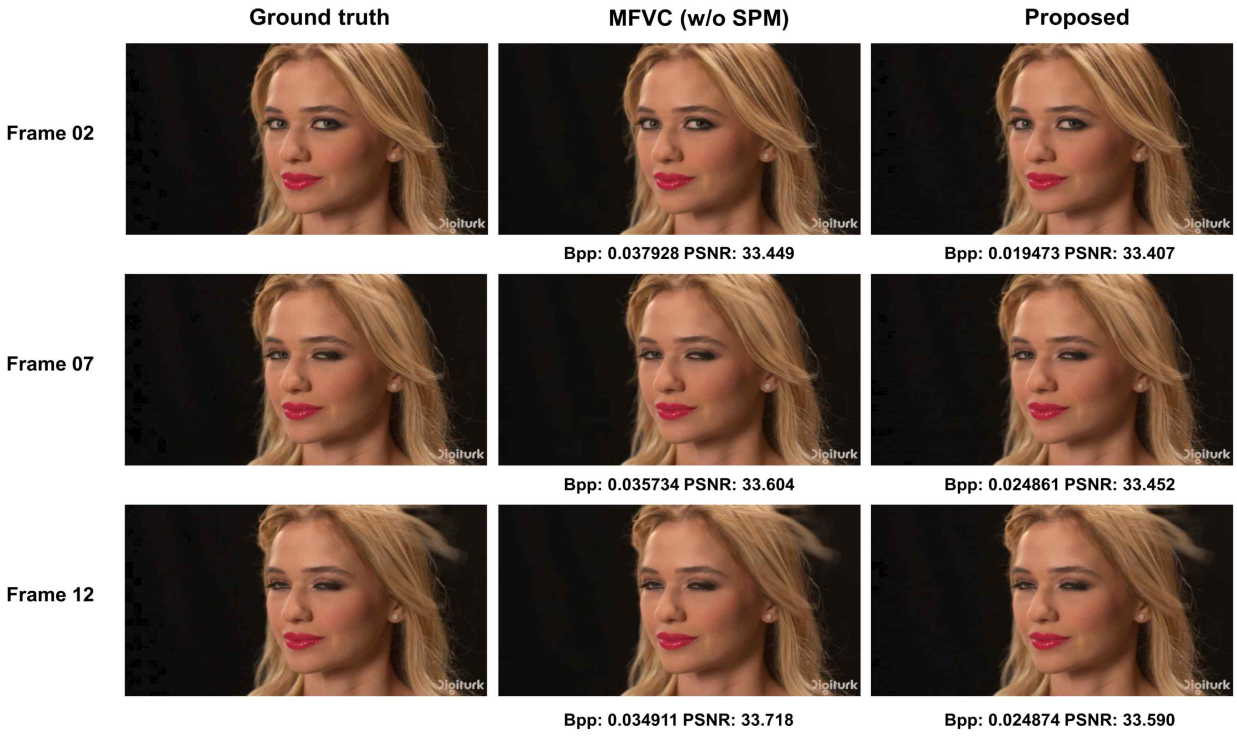

그림 5. UVG 데이터셋의 'Beauty' 비디오의 2번, 7번, 12번 영상을 압축 및 복원한 결과로, 제안된 모델은 PSNR 관점에서 화질 저하가 약간 존재하지만, 비트량을 크게 줄인 것을 확인할 수 있다
Fig. 5. The result of compressing and reconstructing the 2nd, 7th, and 12th images of UVG dataset's 'Beauty' video, it can be seen that the proposed model has a slight decrease in image quality from a PSNR perspective, but significantly reduces the bit amount

성능 비교를 위하여 화질 수준에 맞는 모델을 각각 학습시켜서 사용하였다. 총 6개의 화질은 rate-distortion의 비율을 결정해주는 상수 λ에 따라서 결정되며, Compressai[25]에서 제시한 값을 따랐다.

시공간 엔트로피 모델을 학습할 때는 널리 사용되는 Vimeo-90k 데이터셋[26]을 사용하였으며, 연속된 2장의 영상에서 256×256 크기의 패치(Patch)를 임의로 잘라서 (Cropping) 사용하였다. optimizer로는 Adam[27]을 사용하였으며, 학습률은 $\{1e^{-4},5e^{-5},1e^{-5},5e^{-6},1e^{-6}\}$을 300K, $400K$, $450K$, $475K$ 마다 바꿔가며 총 $500K$ 동안 학습시켰다. LCM, BFD-E/D를 학습할 때에는 시공간 엔트로피 모델의 파라메터를 고정한 후 추가로 해당 모듈들만 학습시켰으며 학습 조건은 수렴 속도에 따라 반복(Iteration) 수를 $100K$이내에서 조정한 것을 제외하면 시공간 엔트로피 모델을 학습할 때와 같다. 모델 평가는 논문[5]과 동일하게 UVG (Ultra Video Group) 데이터셋[28]의 7개 영상을 앞에서부터 100장씩 사용했다. GOP (Group of picture) 또한 동일하게 12로 설정하여 실험하였으며, 실험결과는 그림 5에서 확인할 수 있다. All I-frame은 P-frame을 사용하지 않은 것으로 통합 오토인코더의 인트라 코딩만을 사용한 것이고, MFVC (w/o SPM)는 MFVC에서 SPM (Spatial Prior Module)을 제외한 모델을 말한다. SPM은 Minnen의 모델[9]과 마찬가지로 재귀적인 방식을 사용하기 때문에 병렬처리가 어려우며, 디코딩 속도를 크게 증가시키기 때문에 제외하였다. 실험을 통해 세 가지 모델을 Rate-Distortion (PSNR) 관점에서 비교하였으며, GOP 12인 조건에서 비교했을 때 기존 MFVC (w/o SPM) 모델보다 본 논문에서 제안한 모델이 더 좋은 성능을 보여주고 있다. 또한, 표 2에서는 MFVC 논문[5]에서 사용한 H.264 성능에 대한 BDBR (Bjøntegaard Delta-Rate) 수치(%)를 확인할 수 있다. 통합

표 2. H.264를 기준으로 BDBR 측정한 결과
Table 2. The result of BDBR measurement against H.264

Model	BDBR(%)
All-intra (compressai)	15.55
MFVC (w/o SPM)	-14.78
MFVC (w/o SPM) + LDS/LCM (Proposed)	-27.13

오토인코더 모델만을 사용한 All-intra 조건의 경우 15.55% 높은 수치를 기록했으며, MFVC 프레임워크만을 사용했을 때는 -14.78%, 그리고 MFVC 프레임워크에 LDS/LCM을 사용한 모델은 -27.13%의 BDBR 이득을 보였다. 그림 5와 그림 6에서는 평가 데이터 중 일부를 시각화하였으며, 제안된 방법이 비트량을 효과적으로 줄인 것을 확인할 수 있다. 비트량이 줄어든 것과 동시에 PSNR이 다소 하락했지만, 줄어든 비트량에 비하면 크게 하락하지 않았다.

V. 결 론

본 논문에서는 신경망 기반의 비디오 압축 모델에서 성능 향상을 위해 사용할 수 있는 새로운 방법을 제시했다. 해당 방법은 엔트로피 모델이 정확히 추정할 수 없는 정보를 레이턴트 정보의 잔차 정보 $\hat{y}_{t_{res}}$에 미리 연산하여 전송함으로써 마치 엔트로피 모델이 $\hat{y}_{t_{res}}$의 확률분포를 정확하게 추정한 것과 같은 효과를 낸다. 또한, 레이턴트 공간에서의 보상 모듈인 LCM을 제시함으로써 레이턴트 공간에서의 후처리(Postprocessing) 모듈이 잘 동작함을 보였다. 이러한 방법을 통해 기존 MFVC의 성능을 향상시키는 실험을 보였으며 H.264를 기준으로 평가한 BDBR 기준으로 두 배 가까운 성능 향상을 보였다. 또한, LDS 방법과 LCM 모듈은 레이턴트 정보와 엔트로피 모델을 사용하는 신경망 기반의 이미지 또는 비디오 압축 기술에 광범위하게 적용

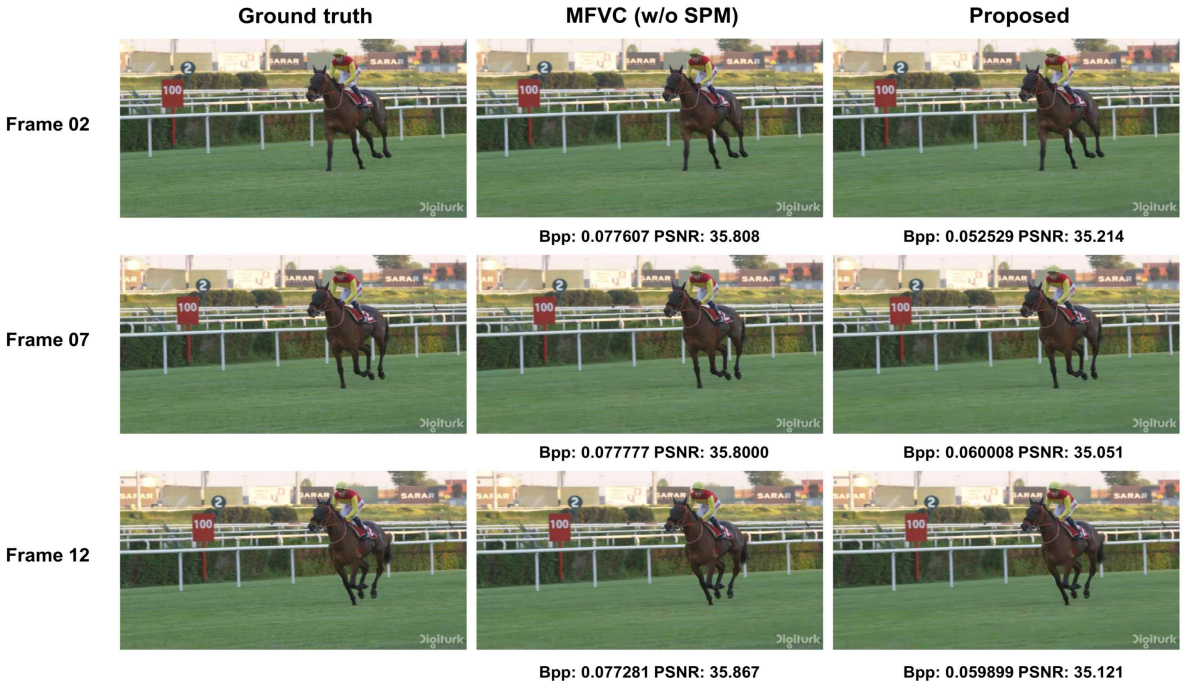

그림 6. UVG 데이터셋의 'Jockey' 비디오의 2번, 7번, 12번 영상을 압축 및 복원한 결과로, 높은 번호의 영상일수록 비트량이 높고 화질 저하가 다소 발생함을 알 수 있다. 하지만 여전히 기존 방법인 MFVC (w/o SPM)에 비해 좋은 압축 효율을 보여주고 있다
Fig. 6. The result of compressing and reconstrictiong the 2nd, 7th, and 12th images of UVG Dataset's 'Jokey' video, the higher the number of images, the higher the bit amount and the lower the image quality. However, it still shows better compression efficiency than the existing method, MFVC (w/o SPM)

할 수 있다는 장점이 있다. 그림 5에서 제시한 실험결과에서 전통적 비디오 코덱인 H.264, H.265와의 비교를 진행했는데, H.264에 비하면 좋은 성능을 보였지만 H.265에 비교했을 때는 비교적 부족한 성능을 보인다. 본 논문에서는 기존에 없던 방법인 레이턴트 정보 이동과 보상 모듈을 소개하는 것에 초점을 맞췄기 때문에 1) Multi-frame optimization[20], 2) DM backbone block[29], 3)Asymmetric Encoder/Decoder[29] 등의 일반적인 방법들을 적용할 수 있는 여지가 많이 남아있다.

참 고 문 헌 (References)

[1] Thomas Wiegand, Gary J Sullivan, Gisle Bjontegaard, and Ajay Luthra, "Overview of the h. 264/avc video coding standard," *IEEE Transactions on circuits and systems for video technology*, 13(7):560 – 576, 2003.

[2] Gary J Sullivan, Jens-Rainer Ohm, Woo-Jin Han, and Thomas Wiegand, "Overview of the high efficiency video coding (hevc) standard," *IEEE Transactions on circuits and systems for video technology*, 22(12):1649 – 1668, 2012.

[3] Jens-Rainer Ohm and Gary J Sullivan, "Versatile video coding – towards the next generation of video compression," *In Picture Coding Symposium*, volume 2018, 2018.

[4] Johannes Ballé, David Minnen, Saurabh Singh, Sung Jin Hwang, and Nick Johnston, "Variational image compression with a scale hyperprior," *In International Conference on Learning Representations*, 2018.

[5] Zhenhong Sun, Zhiyu Tan, Xiuyu Sun, Fangyi Zhang, Dongyang Li, Yichen Qian, Hao Li, "Spatiotemporal Entropy Model is All You Need for Learned Video Compression," *arXiv*, 2021, https://arxiv.org/abs/2104.06083 (accessed Oct. 24, 2021).

[6] F. Bellard, BPG image format, http://bellard.org/bpg/ (accessed: Jan. 30, 2017).

[7] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for imagequality assessment," in Signals, *Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on, IEEE*, vol. 2, 2003, pp. 1398 – 1402

[8] Johannes Ballé, Valero Laparra, and Eero P. Simoncelli, "End-to-end optimized image compression," *In International Conference on Learning Representations*, 2017.

[9] David Minnen, Johannes Ballé, and George D Toderici, "Joint autoregressive and hierarchical priors for learned imagecompression," *In Advances in Neural Information Processing Systems*, pages 10771 – 10780, 2018.

[10] Aaron Van den Oord, Nal Kalchbrenner, Lasse Espeholt, Oriol Vinyals, Alex Graves, et al., "Conditional image generation with PixelCNN decoders," *In Advances in neural information processing systems*, pages 4790 – 4798, 2016.

[11] Zhengxue Cheng, Heming Sun, Masaru Takeuchi, and Jiro Katto, "Learned image compression with discretized gaussian mixture likelihoods and attention modules," *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7939 – 7948, 2020.

[12] Z. Cheng, H. Sun, M. Takeuchi, J. Katto, "Deep ResidualLearning for Image Compression," *CVPR Workshop*, pp. 1-4, June 16-20, 2019.

[13] Y.Zhang, K. Li, K. Li, B. Zhong, Y. Fu, "Residual Nonlocal Attention Networks for Image Restoration," *International Conference on Learning Representations*, pp. 1-18, 2019

[14] Reynolds, Douglas. (2008), "Gaussian Mixture Models," *Encyclopedia of Biometrics,* 10.1007/978-0-387-73003-5_196.

[15] Guo Lu, Wanli Ouyang, Dong Xu, Xiaoyun Zhang, Chunlei Cai, and Zhiyong Gao, "DVC: An end-to-end deep video compression framework," *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages, 11006 – 11015, 2019.

[16] Anurag Ranjan and Michael J Black, "Optical flow estimation using a spatial pyramid network," *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4161 – 4170, 2017.

[17] Abdelaziz Djelouah, Joaquim Campos, Simone Schaub-Meyer, and Christopher Schroers, "Neural inter-frame compression for video coding," *In Proceedings of the IEEE International Conference on Computer Vision*, pages 6421 – 6429, 2019.

[18] Ren Yang, Fabian Mentzer, Luc Van Gool, and Radu Timofte, "Learning for video compression with hierarchical quality and recurrent enhancement," *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6628 – 6637, 2020.

[19] Eirikur Agustsson, David Minnen, Nick Johnston, Johannes Balle, Sung Jin Hwang, and George Toderici, "Scale-space flow for end-to-end optimized video compression," *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8503 – 8512, 2020.

[20] Fabian Mentzer, Eirikur Agustsson, Johannes Ballé, David Minnen, Nick Johnston and George Toderici, "Towards Generative Video Compression," *arXiv*, 2021, https://arxiv.org/abs/2107.12038 (accessed Aug. 26, 2021).

[21] Goodfellow, Ian J., Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville and Yoshua Bengio, "Generative Adversarial Nets," *NeurIPS*, 2014.

[22] Fabian Mentzer, George D Toderici, Michael Tschannen, and Eirikur Agustsson, "High-fidelity generative image compression," *Advances in Neural Information Processing Systems*, 33, 2020

[23] David Minnen, Johannes Ballé, and George Toderici, "Joint autoregressive and hierarchical priors for learned image compression," *In Advances in Neural Information Processing Systems*, pages 10771 – 10780, 2018.

[24] Yoojin Choi, Mostafa El-Khamy, and Jungwon Lee, "Variable rate deep image compression with a conditional autoencoder," *In Proceedings of the IEEE International Conference on Computer Vision*, pages 3146 – 3154, 2019.

[25] Compressai, https://interdigitalinc.github.io/CompressAI/index.html# (accessed Nov. 24, 2021).

[26] Tianfan Xue, Baian Chen, Jiajun Wu, Donglai Wei, and William T Free

man, “Video enhancement with task-oriented flow,” *International Journal of Computer Vision*, 127(8):1106 – 1125, 2019.

[27] Diederik P Kingma and Jimmy Ba, “Adam: A method for stochastic optimization,” *arXiv preprint* 2014, https://arxiv.org/abs/2107.12038 (accessed Nov. 24, 2021).

[28] Alexandre Mercat, Marko Viitanen, and Jarno Vanne, “Uvg dataset: 50/120fps 4k sequences for video codec analysis anddevelopment,” *In Proceedings of the 11th ACM Multimedia Systems Conference*, pages 297 – 302, 2020.

[29] Oren Rippel, Alexander G. Anderson, Kedar Tatwawadi, Sanjay Nair, Craig Lytle and Lubomir Bourdev, “ELF-VC: Efficient Learned Flexible-Rate Video Coding,” *arXiv preprint*, 2021, https://arxiv.org/abs/2104.14335 (accessed Nov. 24, 2021).

저 자 소 개

김 영 웅

- 2016년 3월 ~ 현재 : 경희대학교 응용수학과 학사과정
- ORCID : https://orcid.org/0000-0001-7378-7367
- 주관심분야 : 비디오 부호화, 영상처리, 딥러닝

김 동 현

- 2017년 8월 : 중앙대학교 융합공학부 디지털이미징학과 학사
- 2019년 8월 : 한국과학기술원 바이오및뇌공학과 석사
- 2019년 9월 ~ 현재 : 한국전자통신연구원 통신미디어연구소 미디어연구본부 미디어부호화연구실 연구원
- ORCID : https://orcid.org/0000-0002-1289-0667
- 주관심분야 : 비디오 부호화, 머신러닝

정 세 윤

- 1995년 2월 : 인하대학교 전자공학과 학사
- 1997년 2월 : 인하대학교 전자공학과 석사
- 2014년 8월 : KAIST 전기및전자공학과 박사
- 1996년 12월 ~ 현재 : ETRI 미디어부호화연구실 책임연구원
- ORCID : https://orcid.org/0000-0002-1675-4814
- 주관심분야 : 실감 방송, 비디오 코딩, 컴퓨터 비전

최 진 수

- 1990년 2월 : 경북대학교 전자공학과 공학사
- 1992년 2월 : 경북대학교 전자공학과 공학석사
- 1996년 2월 : 경북대학교 전자공학과 공학박사
- 1996년 5월 ~ 현재 : 한국전자통신연구원 책임연구원
- ORCID : https://orcid.org/0000-0003-4297-5327
- 주관심분야 : 영상부호화 및 영상처리, UHDTV방송, 3DTV방송

저 자 소 개

김 휘 용

- 1994년 8월 : KAIST 전기및전자공학과 공학사
- 1998년 2월 : KAIST 전기및전자공학과 공학석사
- 2004년 2월 : KAIST 전기및전자공학과 공학박사
- 2003년 8월 ~ 2005년 10월 : ㈜애드팍테크놀러지 멀티미디어팀 팀장
- 2005년 11월 ~ 2019년 8월 : 한국전자통신연구원(ETRI) 실감AV연구그룹 그룹장
- 2013년 9월 ~ 2014년 8월 : Univ. of Southern California (USC) Visiting Scholar
- 2019년 9월 ~ 2020년 2월 : 숙명여자대학교 전자공학전공 부교수
- 2020년 3월 ~ 현재 : 경희대학교 컴퓨터공학과 부교수
- ORCID : https://orcid.org/0000-0001-7308-133X
- 주관심분야 : 비디오 부호화, 딥러닝 영상처리, 디지털 홀로그램