

# 비디오 비트율 제어 기술의 최신 동향 분석과 국방 및 항공우주 영상시스템의 비트율 제어 기술 발전 방향 전망

현 명 한<sup>\*,1),2)</sup>

<sup>1)</sup> 경상국립대학교 제어로봇공학과

<sup>2)</sup> 경상국립대학교 공학연구원

## A Study on the Latest Trends of Video Bit-Rate Control Technology and Prospects for the Development of Bit-Rate Control Technology in Defense and Aerospace Imaging Systems

Myunghan Hyun<sup>\*,1),2)</sup>

<sup>1)</sup> Control and Robot Engineering, Gyeongsang National University, Korea

<sup>2)</sup> Engineering Research Institute, Gyeongsang National University, Korea

(Received 19 August 2024 / Revised 20 October 2024 / Accepted 4 November 2024)

### Abstract

Video bit-rate control techniques are essential for efficiently transmitting videos over communication networks. These techniques have been pivotal in broadcasting and internet streaming services, and they could significantly enhance defense capabilities if applied to defense and aerospace imaging systems. Therefore, this paper first reviews the history and standard technologies of video coding, then describes the standard trends of the latest video coding technologies. Finally, it outlines the features of the latest video bit-rate control techniques and discusses their applicability in the defense and aerospace fields.

Key Words : Video Bit-Rate Control(비디오 비트율 제어), National Defense and Aerospace(국방 및 항공우주), Imaging Systems(영상시스템), Video Coding(비디오 영상 부호화)

### 1. 서론

인공지능(Artificial Intelligence, AI) 및 머신러닝(Machine Learning) 기술의 발달로 말미암아 전자공학,

컴퓨터공학뿐만 아니라 경제, 사회 등 전방위적인 분야에서 괄목할 만한 기술 발전 및 성장이 이루어지고 있다. 특히 무인 항공기(Unmanned Aerial Vehicle, UAV), 드론, 로봇, 무인 우주탐사, 자율비행, 국방 무인화 등의 국방 및 항공우주 시스템 개발 및 응용에 인공지능 기술이 적용됨으로 인하여 더욱 정밀하고 효율적인 시스템 개발이 가능하게 되었다. 이러한 국방 및

\* Corresponding author, E-mail: [drwise@gnu.ac.kr](mailto:drwise@gnu.ac.kr)

Copyright © The Korea Institute of Military Science and Technology

항공우주 시스템 개발에 필수적인 것이 영상데이터이며, 그 중요성은 점진적으로 증가하고 있어 이 분야에 특화된 영상시스템의 연구개발이 필요하다.

근래에는 사람의 관점에서 영상을 이해하고 판단/처리하는 영상처리기술이 아니라 기계의 입장에서 영상을 효율적으로 이해하고 처리할 수 있는 기술의 소요가 증가하고 있다. 왜냐하면, AI 기술의 발전으로 인하여 기계가 영상을 판단하고 처리하는 기술 소요가 많이 증가하고 있기 때문이다. 즉, 기존 영상처리의 관점은 어떻게 하면 최소비트 사용으로 최대화질을 얻어낼 것인가에 초점이 맞춰졌던 것이라면 현재에는 화질 성능보다는 객체탐지 및 추적 성능에 방점을 둔 기술의 연구가 필요한 시점이 되었다<sup>[1]</sup>.

이렇듯 기하급수적으로 늘어나는 영상데이터를 효율적으로 처리하는 영상데이터 부호화술은 실시간 영상처리를 위한 필수적인 기술로서 1990년대 초반의 원격화상회의(Teleconference) 기술을 위한 저용량 영상처리기술을 시작으로 현재에는 디지털 TV, 이동통신 기기, 스마트폰, UHD TV 등의 대용량 영상처리 기술로 발전했다. 즉, 실시간 영상처리를 위해 영상데이터를 효율적으로 부호화하는, 즉 최소비트로 최대화질을 내는 부호화 기법의 연구가 지속해서 이루어지고 있다. 또한, 실시간성 이용이 필요한 시스템에서는 한정된 통신 대역에서 일정한 비트율로 부호화를 하여 안정적인 전송을 가능케 하는 비디오 영상 비트율-제어(Rate Control, RC) 기술이 연구되고 있다.

하지만 앞으로는 기계 소비 중심의 영상처리 발전 방향이 두드러질 것이며, 그에 따른 영상 부호화 기술로의 연구전환이 이루어져야 할 것이다. 이를 위해 UAV, 드론, 무인 우주탐사, 자율비행 등의 기술 개발을 위한 국방 및 항공우주 영상시스템에서도 기계 관점의 영상처리 기술도입을 통하여 국방 전투력 향상이 이바지할 필요가 있을 것이다. 아울러 국방 및 항공우주 시스템에 사용되는 영상시스템은 전송 및 수신 간 실시간성이 유지되어야 할 필요성이 크므로 비디오 영상 비트율-제어 기술의 적용이 필수적일 것으로 판단된다.

따라서 본 논문의 다음 장에서는 비디오 영상 부호화 기술의 역사, 최신 비디오 영상 부호화 표준, 및 향후 기술 발전 방향을 제2장에서 소개하고, 제3장에서는 비디오 영상 비트율-제어 기술의 기본개념과 최신 기술을 소개한다. 이를 바탕으로 제4장에서는 국방 및 항공우주 영상시스템에 특화된 비디오 영상 비

트율-제어 기술 발전 방향에 대해 논의하고, 제5장으로 결론을 맺는다.

## 2. 비디오 영상 부호화 기술

### 2.1 역사

비디오 영상 부호화 기술은 화면예측기술을 통해 영상 내부(intra) 또는 영상 간(inter)에 존재하는 중복성을 제거하거나, 인간의 눈에 덜 민감한 신호를 제거하여 데이터의 양을 줄이는 기술이다. 비디오 영상의 크기를 실례로 들어보면, HD(High Definition)급의 영상은 해상도가 1280×720, 컬러색상비트 수는 24 bit, 초당 프레임 수는 30개로 표현되고, UHD(Ultra High Definition) 급의 영상은 해상도가 3840×2160, 컬러색상비트 수는 30 bit, 초당 프레임 수는 30개 이상으로 표현된다. 즉, HD급의 경우 초당 약 664 Mbits(=1280×720×24×30)가 필요하고, UHD급의 경우 초당 7.5 Gbits(=3840×2160×30×30)가 필요하다. 따라서 이렇게 큰 대용량의 비디오 영상데이터를 효율적으로 전송, 저장, 가공하는 데는 비디오 영상 부호화 기술이 필수적이라 할 수 있다.

비디오 영상 부호화 기술의 표준화는 ISO/IEC의 MPEG(Moving Picture Experts Group)과 ITU-T의 VCEG (Video Coding Experts Group)이 연합하여 기술 표준화를 진행하고 있으며, 최근 최신 비디오 영상 부호화 표준인 다용도 비디오 부호화(Versatile Video Coding, VVC) 기술이 표준화 완료되었다<sup>[2]</sup>. Fig. 1은 비디오 영상 부호화 표준의 역사를 보여준다. Fig. 1에서 볼 수 있듯이, 상기 두 표준화 그룹이 연합하여 총 4개의 비디오 영상 표준을 완성하였다.

1995년도에 표준화 완료 된 H.262/MPEG-2 13818-2는 SD(Standard Definition) 급의 디지털 텔레비전 방송 및 DVD 비디오 부호화에 사용되었다<sup>[3]</sup>. 2003년도에 표준화 완료 된 H.264/MPEG-4 AVC는 기존 H.262 대

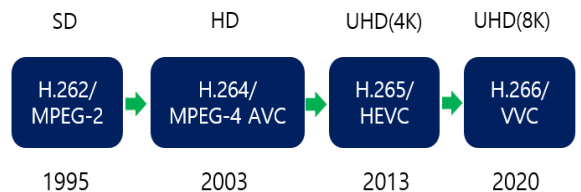


Fig. 1. A history of video coding standard

비 2배 이상의 높은 부호화 효율을 달성하였고, 블루레이 디스크, 인터넷 스트리밍, HDTV 방송 등의 다양한 응용 분야에 활용되었다<sup>[4]</sup>. 2013년도에 표준화 완료된 H.265/HEVC(High Efficiency Video Coding) 표준은 H.264/MPEG-4 AVC의 후속 비디오 영상 부호화 표준으로서, 이 또한 기존 부호화 효율 대비 2배의 이상의 부호화 효율 성능을 제공하며 4K급의 고해상도 비디오 영상 부호화에 특화되어 표준화가 완료되었다<sup>[5]</sup>. 가장 최근인 2020년도에 표준화 완료된 VVC는 H.265/HEVC 보다 2배 더 높은 부호화 효율 성능을 보이며 4k/8K급의 초고해상도 비디오 영상에 초점을 두고 개발된 표준이다<sup>[6]</sup>.

2.2 현존 최신 비디오 영상 부호화 표준

최신 비디오 부호화 표준인 다용도 비디오 부호화 표준(VVC)은, 이전 표준인 고효율 비디오 코딩(HEVC) 보다 고해상도 비디오 영상에 대해 더 높은 부호화 효율성을 나타낸다. 이는 적응형 해상도 비디오 스트리밍, 화면 콘텐츠 비디오 코딩, 초저지연 스트리밍, 고명암비(HDR) 및 넓은 색역(Wide Color Gamut, WCG) 비디오 코딩, 360° 몰입형 비디오 코딩 등 다양한 응용 분야에 적용될 수 있다<sup>[6]</sup>.

Fig. 2는 VVC 부호화기의 기본 구조이다. Fig. 2의 VVC 부호화기는 이전 세대의 비디오 영상 부호화기의 구조와는 큰 차이는 없다. 하지만 부호화 효율을 높이기 위한 다양한 내부/외부 화면예측기법이 적용되었다. 이를 위해, 내부 예측기법으로 a) 넓은 각도 모드 확장이 있는 67개의 내부 모드, b) 위치 의존 내부 예측 조합(Position-Dependent Prediction Combination,

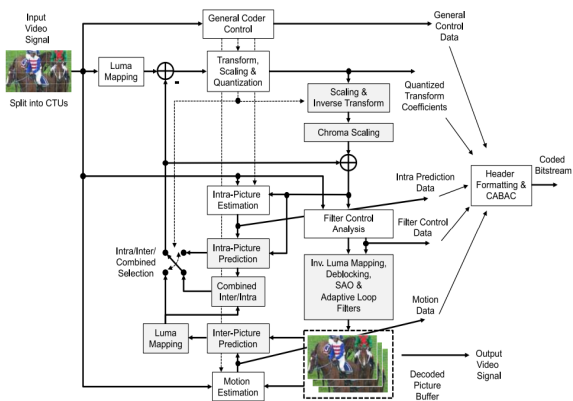


Fig. 2. Typical WC encoder<sup>[6]</sup>

PDPC), c) 교차 구성 요소 선형 모델 내부 예측(Cross-Component Linear Model, CCLM), d) 다중 참조 라인 내부 예측(Multi-Reference Line, MRL), e) 내부 부분 분할(Intra Sub-Partitioning, ISP), f) 행렬 가중치 내부 예측(Matrix-based Intra Prediction, MIP) 등이 사용되었다<sup>[6]</sup>.

외부 예측 모드로는 a) 확장 병합 예측 MVD(Motion Vector Difference)를 사용한 병합 모드(Merge Motion Vector Difference, MMVD), b) 대칭 MVD 신호를 사용하는 AMVP(Advanced Motion Vector Prediction) 모드, c) 아핀 운동 보상 예측, d) 서브 블록 기반 시간적 움직임 벡터 예측(Subblock-based Temporal Motion Vector Prediction, SbTMVP), e) 적응형 움직임 벡터 해상도(Adaptive Motion Vector Resolution, AMVR), f) 가중 평균을 사용한 양방향 예측(Bi-directional Weighted Averaging, BWA), g) 양방향 옵티컬 플로우(Bi-directional Optical Flow, BDOF), h) 디코더 측 움직임 벡터 정제(Decoder-side Motion Vector Refinement, DMVR), i) 삼각형 분할 예측, j) 결합된 상호 및 내부 예측(Combined Intra and Inter Prediction, CIIP) 등이 사용되었다<sup>[6]</sup>.

VVC가 고해상도 비디오 영상 부호화에 효율을 낼 수 있는 가장 큰 이유 중 하나는, 비디오 영상 부호화에 기본이 되는 화면예측성능을 높인 것이다. Fig. 3은 예측 성능을 높이기 위한 부호화 유닛의 크기와 종류를 나타낸다. Fig. 3에서 볼 수 있듯이, VVC에서는 부호화의 기본이 되는 코딩 트리 유닛(Coding Tree Unit, CTU)의 크기가 이전 비디오 영상 부호화 표준대비 커졌으며, 코딩유닛(Coding Unit, CU)의 형태도 다양해져서 부호화 효율을 크게 높일 수 있었다<sup>[7]</sup>.

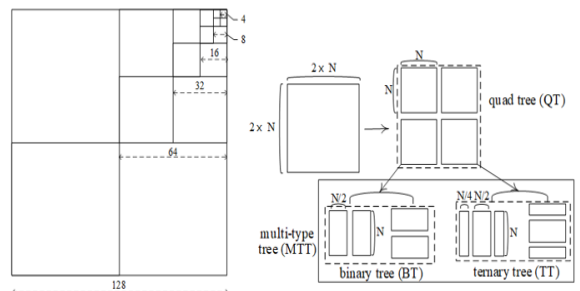


Fig. 3. Larger coding tree unit(CTU) for flexible coding unit(CU) types and structures<sup>[7]</sup>

2.3 향후 비디오 영상 표준화 발전 방향

Fig. 2와 같이 블록 기반의 화면예측을 사용한 비디오 영상 부호화 방법을 전통적인 비디오 영상 부호화 기법이라 할 수 있다. 하지만 최근 인공지능 기술의 발달로 기존의 방식과는 전혀 새로운 형태의 비디오 영상 부호화 기술에 관한 표준 연구가 시작되었다.

2.3.1 기계를 위한 비디오 부호화

인공지능의 발달로 기계의 관점에서 영상을 이해하고 처리하는 소요가 늘어나고 있다. 즉, 일상위험감지(Surveillance), 지능형 교통시스템(Smart Traffic System), 스마트시티(Smart City), 스마트 제조(Intelligent Industry), 지능형 콘텐츠 등의 기계 비전(Machine Vision) 애플리케이션 수행을 위해 기계 입장에서 다양한 객체검출(Object Detection), 객체추적(Object Tracking)을 위한 영상처리가 필수적이다<sup>[1]</sup>.

따라서 앞에서 살펴본 기존의 비디오 영상 부호화 표준은 인간의 시각 시스템에 적합한 형태로 표준화가 진행되었다면, 앞으로는 기계의 입장에서 효율적인 임무 수행을 위한 비디오 영상 부호화 표준이 필요하게 되었다. 이에 MPEG 표준화 그룹에서는 MPEG-VCM(Video Coding for Machines) 그룹을 결성하여 비디오 영상 부호화 표준 개발을 시작하였다<sup>[1]</sup>.

Fig. 4는 기계를 위한 비디오 영상 부호화 표준화 요구사항에서 명시하는 VCM의 구조의 일례를 나타낸다. 그림에서 볼 수 있듯이, 기계를 위한 특징추출(Feature Extraction)을 Neural Network Task(part 1)에서 수행하고, VCM 부호화/복호화를 거친 복원특징(Reconstructed Feature)을 가지고 Neural Network Task(part 2)에서 원하는 기계 비전 업무를 수행한다.

또한, VCM 표준화는 두 트랙(Track)으로 표준화가 진행되고 있으며, 트랙 1은 특징부호화트랙(Feature Coding Track)이며, 트랙 2는 이미지/비디오 영상부호화트랙(Image and Video Coding Track)이다. 트랙 1은 영상 또는 비디오로부터 기계 비전 네트워크의 출력

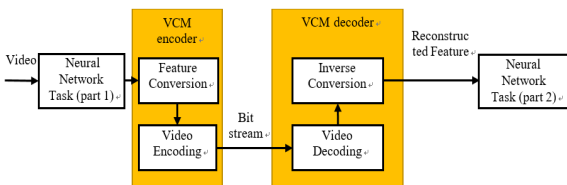


Fig. 4. An example of possible VCM architecture<sup>[1]</sup>

인 특징 지도(Feature Map)를 부호화하는 기술을 의미하며, 트랙 2는 영상 또는 비디오 입력을 받아서 압축/복원을 수행한 복원 영상을 이용하여 기계 비전 업무를 수행하는 부호화 기술을 의미한다. 더 나아가, 트랙 2의 기술은 세부적으로 전처리 기반의 영역기반 접근방식(ROI based Approach)과 딥러닝 기반 종단 간 영상 압축(End-to-End Image Compression) 방식이 연구되고 있다.

2.3.2 신경망 기반의 종단 간 영상 부호화 기술

신경망 기반의 종단 간 영상 부호화 기술은 원본 영상데이터를 입력으로 받아서 부호화된 압축데이터를 생성하고, 이를 다시 복호화하여 원본과 유사한 영상을 출력하는 전체 과정을 신경망 모델이 처리하는 기술이다. Fig. 5는 신경망 기반 영상 부호화 방법의 구조 예시를 보여준다. 특히 영상 부호화에 사용되는 변환(Transform)처리 부분을 신경망(Neural Network)으로 구성하는 기술이 최초제안 되었고, 기본 신경망 구조는 3단계의 구조를 가지며 각 단계에서는 합성곱(Convolution), 풀링(Pooling), 활성화(Activation) 과정이 수행된다<sup>[8]</sup>. 이러한 구조는 입력 영상을 저차원 잠재공간(Latent Space)으로 변환한 뒤, 이를 다시 원본 영상으로 복원하는 과정의 오토인코더(Autoencoder) 구조를 가지며 이러한 과정을 통해 영상 압축이 수행된다. 하지만 양자화(Quantization) 수행으로 연속신호가 불연속 신호로 변화되어 미분할 수 없게 되는 문제점이 있다. 즉 신경망 학습에 중요하게 사용되는 역전파(Backpropagation)가 불가능하여 학습을 수행할 수 없게 된다. 따라서 상기 신경망 기반의 영상 부호화 방법에 새로운 양자화 근사화 방법을 제안하여 학습할

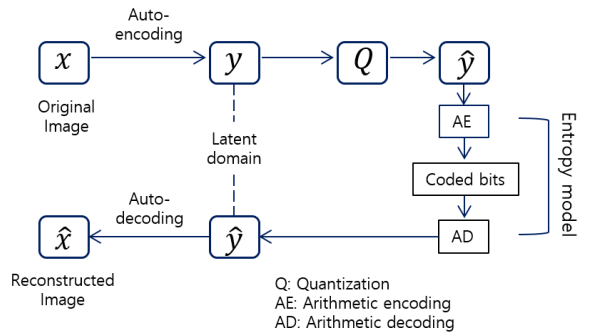


Fig. 5. An structure of neural network based end-to-end image compression

수 있게 하였고, 양자화된 심볼과 그 확률을 기반으로 산술부호화(Arithmetic Encoding)를 수행하여 최종 부호화 비트를 생성하였다. 반대로 복호화의 과정은 생성된 부호화 비트를 산술복호화(Arithmetic Decoding), 역양자화를 수행 후 이를 이미지합성(Synthesis) 신경망을 통해 최종 복원 영상을 만들어 내므로 영상 압축복원을 수행한다.

이 기술은 추후 영상 부호화에서 비디오 영상 부호화로 확장되었고, 화면 간 예측에 중요한 역할을 하는 움직임 벡터(Motion Vector) 예측에 신경망 구조가 적용되어 비디오 영상 부호화가 진행되었다<sup>9)</sup>.

### 2.3.3 암시적 신경망 표현

암시적 신경망 표현(Implicit Neural Representation, INR)은 신경망을 통해 데이터를 암시적으로 표현하는 방법이다. 이 방법은 영상의 2차원 좌표값을 받아 해당 위치의 화소값을 출력할 수 있으며, 데이터를 신경망의 파라미터로 표현하므로 저장공간을 줄일 수 있는 특징이 있다. 따라서 영상 및 비디오 영상을 신경망 모델을 이용한 부호화를 통해 압축할 수 있고, 다시 신경망 모델을 이용한 복호화가 가능하여 차세대 비디오 표준화의 요소기술로 적용될 수 있다. 뿐만 아니라 3차원 데이터처리에 적용될 시 3차원 데이터의 효율적인 저장과 렌더링에 응용될 수 있다. 이러한 기술적인 특징 및 장점으로 비디오 영상 부호화 측면에서의 표준화 기술 개발이 시작되었고, 2차원뿐만 아니라 3차원 영상의 표현 및 압축 성능검증에 지속적인 연구가 이루어지고 있다<sup>10)</sup>.

다시 말하면, 암시적 신경망 표현은 데이터를 근사할 수 있는 함수 표현의 파라미터 값을 저장하여 그 파라미터로 데이터를 복원하는 것에 초점을 둔 기술이다. 반면 2.3.2절의 신경망 기반의 종단 간 영상부호화기술은 영상의 공간적/시간적 중복성을 줄이려는 방안으로 오토인코더 구조를 사용하였고, 전통적인 블록 기반 압축기법 대비 압축 성능을 향상하는 것에 초점을 둔 기술이다.

### 2.3.4 신경망 모델 압축표준

신경망 모델 구조는 다양한 활용 분야에서 기존의 전통적인 방법대비 큰 성능개선이 이루어지고 있다. 하지만 신경망의 계층이 깊어짐에 따라 신경망 학습에 필요한 파라미터의 수가 기하급수적으로 증가하게 되어 소형 모바일 내장형 기기에서의 사용에 큰 제약

이 있다. 또한, 실시간 응용 프로그램에서 동작하기 위해서는 모델의 소형/경량화를 통한 저장공간 절약 및 네트워크 전송 효율을 높여야 할 필요성이 있다.

따라서 신경망 모델을 소형/경량화시키면서도 신경망 모델의 성능을 최대한 유지할 수 있는 기술표준의 소요가 있었고, MPEG 그룹에서 Compression of Neural Networks(NNC)에 대한 연구가 진행되었다<sup>11)</sup>. 이를 위해 신경망 모델에 사용되는 파라미터의 정밀도를 낮추어 신경망 모델의 크기를 줄이는 양자화 기술, 신경망 내부의 비중요 뉴런을 제거하는 프루닝(Pruning)기술 등이 주로 사용되어 신경망 모델의 압축 성능을 향상했다. 이러한 신경망 모델 압축기술은 모바일 내장형 기기, 데이터 클라우드, IoT(Internet of Things) 기기 등에 배포되어 다양한 애플리케이션 개발에 활용될 수 있는 특징이 있어 앞으로도 지속적인 기술연구가 필요한 분야이다.

## 3. 비디오 영상 비트율-제어 기술

### 3.1 기본개념

비디오 영상 비트율-제어는 영상 부호화 이후에 발생하는 부호화 비트의 양이나 복호화된 영상의 화질에 따라 고정 비트율(Constant Bit-Rate, CBR) 제어 또는 가변 비트율(Variable Bit-Rate, VBR) 제어로 나눌 수 있다. CBR의 경우 비디오 영상 전체에 걸쳐 일정한 비트율을 유지하는 방식으로서, 부호화 비트량 예측이 쉬우며 통신 대역폭이 일정한 환경에 적용이 유리하다. 반면 비디오 영상의 복잡도에 따라 화질에 변동이 있으며, 복잡한 영상의 장면에서는 화질이 떨어질 수 있다.

VBR은 비디오 영상의 복잡도에 따라 비트율을 가변적으로 조절하는 방식으로서 복잡한 장면에서는 비트율을 높이고, 단순한 장면에서는 비트율을 낮추게 된다. 즉 본 방식은 화면 간 화질이 일정할 필요가 있는 분야에 적합한 기술이다. 또한, 최근에는 통신 네트워크 상황에 따라 실시간으로 비트율을 조정하여 비디오 영상 스트리밍에 최적화된 적응형 비트율(Adaptive Bit-Rate, ABR) 제어 방식이 사용되는데, 이 방식은 비디오 영상 수신 측면에서 화질의 변화는 있더라도 비디오 영상의 지속적인 수신을 가능케 하는 영상 스트리밍 서비스에 적용될 수 있다.

비디오 영상 비트율-제어를 위해서는 부호화되는

부호화 비트를 예측하거나 복호화된 영상과 원본 영상과의 화질 왜곡(Distortion, D) 정도를 예측하는 기술이 필수적이다. 기본적으로 부호화 비트량과 화질 왜곡의 수준 및 정도는 비디오 영상 부호화 과정(예측, 변환, 양자화, 엔트로피 코딩 등) 중 양자화 과정에서 선택되는 양자화 계수(Quantization Parameter, QP)에 따라 결정되기 때문에 비트-QP(R-Q) 관계 또는 화질 왜곡-QP(D-Q) 관계 때문에 부호화 비트 또는 화질 왜곡 값이 계산될 수 있다. 다음의 하부 절에서는 기본적인 비디오 영상 비트율-제어 모델에 관해 기술한다.

### 3.1.1 R-Q 모델

MPEG-4 표준에서는 비디오 영상 부호화 과정을 거쳐 생성된 양자화 변환계수(Quantized Transformed Coefficients, QTC)를 테일러 시리즈(Taylor Series)로 전개하면 다음의 이차식 모델을 만들 수 있다<sup>[12]</sup>.

$$R = c_1 Q^{-1} + c_2 Q^{-2} \quad (1)$$

여기서  $c_1$ 와  $c_2$ 는 비디오 콘텐츠 특성에 따른 모델 파라미터를 나타내고, Q는 QP를 나타내며, R은 특정 부호화 레벨에 대한 목표 비트량을 나타낸다.

이전의 H.264/MPEG-4 AVC<sup>[4]</sup> 및 HEVC<sup>[5]</sup>와 같은 비디오 영상 부호화 표준에서는 비트율 추정 정확도를 높이기 위해 픽셀 강도의 기울기를 활용하였다<sup>[13-17]</sup>. 또한, 변환계수(Transformed Coefficient, TC) 값을 모델링 하기 위해 Laplacian, Cauchy 및 Gaussian과 같은 다양한 확률 밀도 함수(Probability Density Function, pdf)가 R-Q 모델의 정확도를 높이기 위해 고려되었다<sup>[18,19]</sup>. R-Q 모델의 특별한 경우로서, TC 값 pdf 모델의 QTC 중 0의 백분율인  $\rho$ 를 선형 함수로 활용하는 방식이 있는데, 이를 R-p 모델이라 하고 다음의 수식으로 표현된다<sup>[20]</sup>.

$$R = \theta \cdot (1 - \rho) \cdot N \quad (2)$$

여기서  $(1-\rho)$ 는 프레임에서 0이 아닌 QTC의 백분율을 나타내며,  $\theta$ 는 모델 매개 변수이고, N은 프레임의 전체 픽셀 수를 나타낸다.

HEVC 표준화 초기 단계에서, R-Q 곡선이 다양한 CU의 동일 깊이 수준에서 거의 유사하다는 관찰에 기반하여 픽셀 단위의 통합 비트율 양자화(Unified Rate Quantization, URQ) 모델이 제안되었다<sup>[21]</sup>. 반면,

다양한 CU 깊이 수준에서의 pdf 특성을 활용하여 R-Q 혼합 모델에 대한 여러 방안이 제안되었다<sup>[22-25]</sup>. 아울러, Kullback-Leibler 발산을 최소화하는 합성 pdf 모델이 제안되었고, 그 합성 pdf 모델을 R-p 모델과 결합하여 R-D 추정 정확도를 높였다<sup>[26]</sup>.

QP가 증가함에 따라, 부호화된 비트에서 비텍스처 비트(ex. 헤더 비트)의 비율이 HEVC 표준에서 급격히 증가하게 되었다<sup>[23]</sup>. 따라서 기존의 R-Q 모델은 비텍스처 비트를 거의 다루지 않기 때문에 비트율 추정 성능이 저하되는 단점이 있다. 이러한 문제를 극복하기 위해 텍스처 비트와 비텍스처 비트를 각각의 모델로 별도로 추정하는 방법이 제안되었다<sup>[23,25]</sup>. 비록 이 방법들은 HEVC 표준에서는 성공적으로 동작했으나, Fig. 3에서 이미 언급한 바와 같이 최신 비디오 영상 부호화 표준인 VVC에서는 부호화 깊이 수준이 많이 증가하여 비디오 영상 부호화를 위한 계산 부담이 가중되는 단점이 있다. 더욱이, R-D 성능 향상을 위해 방사형 기저 함수(Radial Basis Function, RBF) 네트워크를 사용하는 라플라시안 혼합 모델(Laplacian Mixture Model, LMM)을 적용하면 계산 복잡도가 더욱 많이 증가한다<sup>[25]</sup>. 따라서 다중 깊이 CU를 위한 R-Q 모델은 더 깊은 다수의 CU에서 얻은 잔차(Residues)를 처리하는 데에 계산적인 부담이 있어 실시간 구현에 큰 어려움이 있다.

### 3.1.2 R- $\lambda$ 모델

기존의 R-Q 모델은 H.262/MPEG-2 13818-2<sup>[3]</sup>와 H.264/MPEG-4 AVC<sup>[4]</sup>를 위해 개발된 것으로, 비디오 영상 부호화 기술의 발전에 따라 새로운 표준화 기술에 적합한 비트율 추정 방법이 필요하게 되었다. HEVC<sup>[5]</sup> 표준에서는 인코딩된 비트가 다양한 화면 내 및 화면 간 예측 모드의 부호화 매개 변수에 영향을 받는 특징이 있으므로 QP가 최종 부호화 비트량을 결정하는 유일한 요소라 보기 어렵다. 따라서 R-Q 모델을 대신 하여, QP(또는 비트율)와 R-D 곡선의 기울기를 나타내는 라그랑지안 승수  $\lambda$  사이의 관계를 이용한 다양한 모델이 제안되었다<sup>[27-29]</sup>.

비트율-왜곡 최적화(Rate Distortion Optimization, RDO)<sup>[30]</sup> 관점에서 화질 왜곡(D)은 주어진 비트 예산(Rate Budget, Rb) 내에서 비트율(R)을 최소화하도록 계산되어야 하며 수식으로 표현하면 다음과 같다.

$$\min D \quad s.t. \quad R \leq R_b \quad (3)$$

라그랑지 승수 방법을 통해<sup>[31]</sup>, (3)은 다음과 같이 비구속 문제(Unconstrained Problem)로 표현될 수 있다.

$$J = D + \lambda \cdot (R - R_b) \quad (4)$$

여기서  $J$ 는 R-D 비용 함수이고,  $\lambda$ 는 라그랑지 승수이다. 또한, Mallet 등은 R-D 곡선이 직사각형 쌍곡선 함수로 표현될 수 있음을 확인했고, R-D의 관계를 다음의 수식으로 표현하였다<sup>[32]</sup>.

$$D(R) = \phi \cdot R^{-\gamma} \quad (5)$$

여기서  $\phi$ 와  $\gamma$ 는 모델 매개 변수이다. R-D 곡선은 미분 가능하므로, (5)는 다음 식으로 재표현될 수 있다.

$$\lambda = -\frac{\partial D}{\partial R} = \phi\gamma R^{-\gamma-1} \quad (6)$$

(6)을 정리하여 유도한 R- $\lambda$  모델은 다음과 같다.

$$\lambda = \alpha \cdot R^\beta \quad (7)$$

여기서  $\alpha$ 와  $\beta$ 는 R- $\lambda$  모델의 매개 변수를 나타낸다. 또한 (7)의 R은 텍스처와 비텍스처 비트를 모두 포함하는 반면, (1)의 R은 텍스처 비트만 포함한다.

상기 R- $\lambda$  모델의 정확한 R-D 모델링 성능 덕분에, R- $\lambda$  모델 기반 RC 알고리즘이 HEVC Reference SW Model(HM)에 채택되었다<sup>[27,29]</sup>. 그러나 (7)의 모델 매개 변수는 이전에 부호화된 데이터를 통해 추정되는 구조이므로, 이전에 부호화된 데이터가 비선형 특성을 가지면 R-D 모델링 성능이 저하될 가능성이 크다.

### 3.2 최신 기술

#### 3.2.1 파티클 필터링 기반의 R-Q 모델

기본적으로 RC 모델의 성능을 좌우하는 것은 비트 예측 성능이다. 즉 최종 부호화 비트를 잘 예측하면 비트율-제어 성능을 높일 수 있다. 비디오 영상 부호화 효율성을 높이기 위해서 2.2절에서 설명한 것처럼 다양한 비디오 영상 부호화 기술들이 접목되고 있다. 그에 따라 비디오 영상 부호화 결과로 나오는 비트나 화질의 값이 굉장히 무작위적이고 불안정해지는 특성이 있어서 최종 부호화 비트의 예측에 어려움이 증가하고 있다.

파티클 필터링은 위치추정이나 시계열예측 같은 공학 및 경제 분야에 사용되는 예측기법으로, 비디오 영상 부호화의 R/D 모델에 적용되어 큰 효율을 나타내었다<sup>[7]</sup>. 파티클 필터링을 통한 예측을 수행하기 위해서는 임의의 파티클을 무작위로 생성하고 초기 예측을 수행하게 되는데, 특히 순차적 중요 재 표본화(Sequential Importance Resampling, SIR) 알고리즘을 사용하여 초기예측값의 확률을 업데이트하는 일련의 과정을 반복수행하여 예측 성능을 향상하는 특징이 있다. 또한, 이전의 필터링 방법(ex. 칼만 필터)과 비교하여, 예측을 위한 제약조건(cf. 다른 기법의 경우 가우시안 같은 사전 pdf 모델이 필요)이 없으므로 무작위로 생성되고 표본화된 입자로부터 예측값을 얻는데 매우 유용한 확률적 추정 방법이다.

파티클 필터링에 사용되는 마르코프 과정(Markov Process)과 베이저안 이론(Bayesian Theorem)에 의해 사전확률과 사후 확률이 다음의 수식으로 표현된다.

$$p(\bar{x}_k | \bar{y}_k) = \int p(\bar{x}_k | \bar{x}_{k-1}) p(\bar{x}_{k-1} | \bar{y}_{k-1}) d\bar{x}_{k-1} \quad (8)$$

$$p(\bar{x}_k | \bar{y}_k) = \frac{p(\bar{y}_k | \bar{x}_k) p(\bar{x}_k | \bar{y}_{k-1})}{p(\bar{y}_k | \bar{y}_{k-1})} \quad (9)$$

여기서 (8)은 사전확률, (9)는 사후 확률을 나타내며,  $\bar{x}_k$ 는 시간 k에서의 무작위 파티클 벡터,  $\bar{y}_k$ 는 SIR 알고리즘에 의해 가중치가 업데이트된 파티클 벡터이다. 또한  $p(\bar{x}_k | \bar{x}_{k-1})$ 는 시간 (k-1)과 시간 k 사이의 상태 변화 pdf,  $p(\bar{x}_k | \bar{y}_{k-1})$ 는 시간 (k-1)에서의 사후 확률,  $p(\bar{y}_k | \bar{x}_k)$ 는 우도 확률(likelihood pdf),  $p(\bar{y}_k | \bar{y}_{k-1})$ 는 상수이다. 파티클 필터링은 (8)과 (9)의 과정이 상호 보완되는 과정을 통하여 예측 성능을 높이는 구조임을 확인할 수 있다.

파티클 필터링 기반의 R-Q 모델에서는 생성하려는 파티클을 임의의 비트값으로 생성하고, 이를 필터링 과정을 수행하여 최종 부호화된 비트값을 예측하는 것이다. 또한, 임의의 비트값 파티클을 이용하여 다음 프레임의 비트값을 예측하기 위해서는 프레임 간 화질 왜곡 값의 비율을 이용하며 다음의 수식으로 표현된다.

$$\Delta D_i = \alpha \cdot (D_i - D_{i-j}) / D_{i-j} \quad (10)$$

여기서  $D$ 는 평균 제곱 오차(Mean Square Error, MSE)를 의미하고,  $i$ 와  $(i-j)$ 는 프레임 번호를 의미한다.  $\alpha$ 는 비트율 조정 파라미터로 상숫값이다.

따라서 다음 프레임 예측을 위해 생성된 비트값 파티클은 다음의 수식으로 표현된다.

$$\bar{b}_{i+1} = B_i \cdot \underbrace{[1 \ 1 \ 1 \ \dots \ 1]}_N^T + \Delta D_i \cdot \bar{b}_i \quad (11)$$

여기서  $\bar{b}_{i+1}$  및  $\bar{b}_i$ 는 벡터로 표현된  $(i+1)$  및  $i$  번째 프레임의  $N$ 개의 비트값 파티클이며,  $B_i$ 는 비디오 영상 부호화 결과 얻어진  $i$  번째 프레임의 스칼라 비트값이다. (11)의 비트값 파티클들은 SIR 및 확률 정규화(Normalization) 과정을 거쳐 다음의 (12)와 같이 표현된다.

$$c_{k,i+1} = p(b_{k,i+1}) / \sum_{l=1}^N p(b_{l,i+1}), \quad k = 1, 2, \dots, N \quad (12)$$

여기서  $p(b_{k,i+1})$ 는  $\bar{b}_{i+1}$ 의 pdf를 의미하며,  $\bar{c}_{i+1}$ 는 확률 정규화 과정을 거쳐 생성된  $(i+1)$  번째 프레임의 파티클( $N$ 개)의 확률이다. 따라서 이렇게 생성된 (11)과 (12)의 비트값 파티클 및 해당 확률값의 내적으로 최종 예측 비트값이 다음의 수식으로 계산된다.

$$\tilde{B}_{i+1} = \bar{c}_{i+1} \cdot \bar{b}_{i+1} \quad (13)$$

(13)으로 계산된 최종 예측 비트값을 부호화의 목표 비트로 할당하여 부호화하면  $R$ - $\lambda$ - $Q$  관계에 의해 비트율-제어가 이루어진다.

### 3.2.2 라그랑지 계수 조절을 통한 비트율-제어

전통적인 비트율-제어 방법에서는 QP 값과 비트값의 관계로 비트율-제어를 수행하는데, 프레임 간 텍스처의 형태가 다양하여 프레임 간 비트 변화량이 큰 경우가 있다. 이러한 조건에서 더 효율적으로 비트율-제어를 하는 방법으로 비디오 영상 부호화의 RDO 단계에 사용되는 라그랑지 계수  $\lambda$ 를 조정하여 비트율-제어 효과를 극대화하기도 하였다<sup>[33]</sup>.

이 방법은 구조적으로는 3.2.1절 사용된 방법과 동일하게 파티클 필터링을 수행하였으나 파티클 필터링에 적용한 파티클의 값을 비트율( $R$ )이 아닌 화질 왜

곡( $D$ ) 값으로 선정된 점이 다르다. 즉, (10)~(12)의  $D$ 를 대신하여  $R$ 을 사용하여 파티클 필터링을 수행하였고, 최종 예측된 화질 왜곡 값은 다음의 수식으로 계산된다.

$$\tilde{D}_{i+1} = \bar{q}_{i+1} \cdot \bar{d}_{i+1} \quad (14)$$

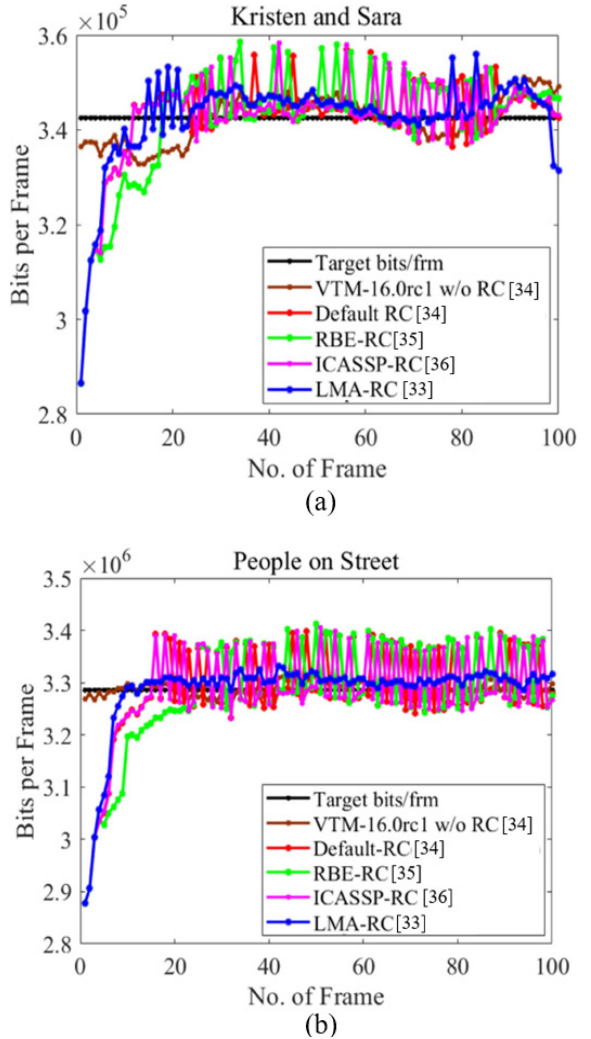


Fig. 6. Rate control performances of bit-rate consistency based on the state-of-the-art method<sup>[33]</sup>; (a) *Kristen and Sara* (1280×720) with a target-bits per frame( $T_B$ ) of 342.5 Kbits for QP = 22, *People on Street* (2560×1600) with  $T_B = 3285.7$  Kbits for QP = 22



여기서  $\bar{q}_{i+1}$ 는 확률 정규화 과정을 거쳐 생성된 (i+1) 번째 프레임의 N개의 화질 왜곡 파티클의 확률이며,  $\bar{d}_{i+1}$ 는 화질 왜곡 값 파티클의 벡터표현이다. (14)로 예측된 최종 예측 화질 왜곡 값은 비디오 영상 부호화의 RDO에 사용되는  $\lambda$ 를 정밀하게 조절하는 조정값 계산에 사용되며 조정값은 다음과 같이 계산된다.

$$\gamma_{i+1} = \theta \cdot (\bar{D}_{i+1} - D_{i-j}) / D_{i-j} \quad (15)$$

여기서  $\gamma_{i+1}$ 는 (i+1)번째 프레임을 위한 라그랑지 계수  $\lambda$  조정값이며, j는 프레임 간격을 조절하는 요소이다. 그리고  $\theta$ 는 화질 왜곡 변화를 완만하게 조절하는 상수값으로서 실험적 수치이다. (15)의  $\gamma_{i+1}$ 를 비디오 영상 부호화 RDO 단계의  $\lambda$  조정에 사용하였고, 다음 프레임을 위한 부호화용  $\lambda$ 는 다음과 같이 계산된다.

$$\lambda_{i+1} = (\gamma_{i+1} + 1) \cdot \lambda_i \quad (16)$$

Fig. 6은 라그랑지 계수 조정을 통한 비트율-제어의 결과를 보여준다<sup>[33]</sup>. Fig. 6에서 알 수 있듯이, 라그랑지 계수 조정을 통한 비트율-제어가 기존의 다른 방법들<sup>[34-36]</sup> 대비 비트 향상성을 높여, 밴드제약이 고려되는 통신환경에서 적합하게 사용될 수 있음을 확인할 수 있다.

#### 4. 국방 및 항공우주 영상시스템의 비트율-제어 기술 발전 방향

과거에는 사람이 직접 전장 시스템 운용에 참여하여 국방 무기체계의 목표 정밀도를 높여 왔다. 예를 들어 국방 무기체계 운용에 MITL(Man-In-The-Loop) 개념을 도입하여 사람이 직접 타겟 영상을 보면서 무기체계를 제어하는 방식으로 정밀타격 효율성을 증대해왔다. 하지만 현대에서는 다수의 목표물과 표적을 실시간으로 분석하고 처리하며 최적 교전 시나리오를 작성하여 교전을 수행해야 하므로 기계의 힘을 이용해야 한다. 이렇듯 기계의 입장에서 전장 환경 영상을 정확하고 신속하게 판단하고 처리할 방법에 관한 집중적인 연구가 필요하게 되었다.

제2장에 기술한 비디오 부호화 기술의 발전 방향 및 동향을 분석해 보았을 때도, 영상 소비의 형태가

사람 중심의 영상 소비에서 기계 중심의 영상 소비로 바뀌고 있으며 그에 따른 방향으로 기술이 발전되고 있음을 확인할 수 있었다. 이는 자율비행, 자율주행과 같은 애플리케이션에서 영상인식, 객체탐지 등과 같은 기술의 수요가 늘어나고 있으며, 거기에 맞는 비디오 영상 신호처리 기술이 필수적인 요소로 등장하고 있음을 확인할 수 있었다.

특히 국방 및 항공우주체계의 영상시스템에서는 전장 분석 및 무기체계 정밀성을 높이기 위해 획득 영상의 한 장면 한 장면(scene by scene)의 중요성이 커지기 때문에 화면 내 부호화 비트율-제어(Intra Coding RC)의 활용도가 높아질 것으로 예상된다. 그리고 통신자원이 제한되고 대역이 협소한 전장 통신, 심우주 통신 같은 초 협대역의 상황에서도 비트를 안정적으로 제어하는 비트율-제어 기법의 연구가 꾸준히 필요할 것으로 예상된다. 아울러, 화면 내 부호화의 경우 제약된 화면 내 정보만을 이용하여 부호화하므로 부호화 발생 비트의 변동량이 커질 수 있는 경향이 있다. 그러므로 전체적으로 비트 발생량을 평활화하는 기법의 적용을 통하여 통신 대역의 효율성을 높여야 할 필요가 있다.

또한, 앞서 서술한 기계 중심의 영상 소비 패턴에 부합한 비디오 영상 신호처리 기술로서 VCM 관련 표준화가 진행 중이지만 현재까지의 표준화 기술 연구 진행 상황으로 볼 때에, VCM의 내부(inner) 코덱으로서 기존의 비디오 영상 부호화 표준(ex. VVC)이 사용될 가능성이 크다. 이는 VCM이 특징부호화트랙(Feature Coding Track)의 트랙 1 또는 이미지/비디오 영상부호화트랙(Image and Video Coding Track)의 트랙 2가 선정되더라도 기존에 RC에 사용되었던 기술들이 차세대 비디오 영상 부호화 표준에서도 사용될 가능성이 크다는 의미이다. 따라서, 본 논문의 제3장에서 소개한 최신 RC 기술들이 새로운 비디오 영상 부호화 표준에도 충분히 적용될 수 있기에 앞으로도 전통적인 RC 기술들에 관한 지속적인 기술연구가 필수적일 것으로 판단된다.

내부 코덱을 기존의 비디오 영상 표준을 사용하지 않게 되더라도, 통신 전송을 위해서 발생하는 부호화 비트의 제어는 필수적이다. 따라서 화질 왜곡과 비트량의 상관관계를 최적으로 예측하는 모델이 개발된다면 조금 더 효율적인 비디오 영상 비트율-제어가 가능할 것으로 판단한다. 즉, 비트율을 효과적으로 예측할 수 있는 비트율 예측 심층신경망을 구성하는 것도 RC 효율을 높일 수 있는 또 다른 대안이라고 판단된다.

## 5. 결 론

본 논문에서는 국방 및 항공우주 영상시스템을 위한 비디오 비트율-제어 기술에 대한 필요성 및 발전 방향을 소개하였다. 이를 위해 먼저 비디오 영상 부호화 기술의 역사와 표준기술을 소개하고 최신 영상 부호화 기술의 표준동향에 관해서도 기술하였다. 아울러 비디오 영상 비트율-제어 기술의 특성 및 최신 영상 부호화 표준에 적용된 비트율-제어 기술의 특징을 소개하였다. 특히 지금은 인공지능 기술의 발전과 더불어 비디오 영상의 소비 주체가 인간에서 기계로 바뀌는 시점이다. 이에 따라 새롭게 표준화되는 새로운 비디오 영상 부호화 표준기술은 기계의 입장에서 객체 탐지 및 추적에 활용성이 높은 표준이다. 따라서 해당 표준기술은 무인 항공기, 드론, 무인 로봇탐사와 같은 국방 및 항공우주 시스템에 꼭 필요한 기술이라 할 수 있다. 아울러 새로운 표준의 특성에 따른 내부 코덱으로 기존의 전통적인 부호화 표준의 적용이 유력하므로 전통적인 비트율-제어 기법이 사용될 가능성이 큼을 확인하였다. 따라서 새로운 비디오 영상 부호화 표준기술이 개발되더라도 국방 및 항공우주 영상 시스템을 위해 특화된 전통적인 비트율-제어 기반의 신기술에 대한 지속적인 연구개발이 필요할 것으로 판단된다.

## References

- [1] Use cases and requirements for Video Coding for Machines, ISO/IEC JTC1/SC29/WG2 N190, 2022.04
- [2] B. Bross, J. Chen, S. Liu, and Y. Wang, "Versatile video coding editorial refinements on draft 10," in Proc. 20th Meeting, Teleconference, pp. 1-114, 2020.
- [3] MPEG-2 Test Model 5, ISO/IEC JTC1/SC29 WG11/93-400, 1993.
- [4] T. Wiegand, G. J. Sullivan, G. BjØntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," IEEE Trans. Circuits Syst. Video Technol., Vol. 13, No. 7, pp. 560-576, Jul. 2003.
- [5] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding(HEVC) standard," IEEE Trans. Circuits Syst. Video Technol., Vol. 22, No. 12, pp. 1649-1668, Dec. 2012.
- [6] B. Bross et al., "Overview of the versatile video coding(VVC) standard and its applications," IEEE Trans. Circuits Syst. Video Technol., Vol. 31, No. 10, pp. 3736-3764, Oct. 2021.
- [7] M. H. Hyun, "A study on particle filtering based prediction for rates and distortions of video coding and an application to rate control," Ph.D. dissertation, Dept. of Elec. Engine., KAIST, Daejeon, Republic of Korea, 2021.
- [8] J. Ballé, V. Laparra, and E. P. Simoncelli, "End-to-end optimized image compression," arXiv:1611.01704v3, 2017.
- [9] G. Lu, W. Ouyang, D. Xu, X. Zhang, C. Cai, and Z. Gao, "Dvc: An end-to-end deep video compression framework," Proceedings of the IEEE/CVF CVPR, 2019.
- [10] G. Bang, "Neural network video representation technology and MPEG-INVR standardization trends," Broadcast and Media Magazine, Vol. 28, No. 3, pp. 45-54, 2023.
- [11] W. Bailer, et al., "Text of ISO/IEC FDIS 15938-17 Compression of neural networks for multimedia content description and analysis," ISO/IEC JTC1/SC29/WG4, N20331, Jun. 2021.
- [12] T. Chiang and Y.-Q. Zhang, "A new rate control scheme using quadratic rate distortion model," IEEE Trans. Circuits Syst. Video Technol., Vol. 7, No. 1, pp. 246-250, Feb. 1997.
- [13] Y. Chang and M. Kim, "Hybrid codec-based intra-frame joint rate control for stereoscopic video," IEEE Signal Process. Lett., Vol. 18, No. 10, pp. 543-546, Oct. 2011.
- [14] X. Jing, L. P. Chau, and W. C. Siu, "Frame complexity-based rate-quantization model for H.264/AVC intraframe rate control," IEEE Signal Process. Lett., vol. 15, pp. 373-376, Mar. 2008.
- [15] M. H. Wang, K. N. Ngan, and H. L. Li, "An efficient frame-content based intra frame rate control for high efficiency video coding," IEEE Signal Process. Lett., Vol. 22, No. 7, pp. 896-900, Jul. 2015.
- [16] M. Karczewicz and X. Wang, "Intra frame rate

- control based on SATD,” JCTVC-M0257, 13th meeting, Incheon, KR, 2013.
- [17] B. Yan and M. H. Wang, “Adaptive distortion-based intra-rate estimation for H.264/AVC rate control,” *IEEE Signal Process. Lett.*, Vol. 16, No. 1-3, pp. 145-148, Mar. 2009.
- [18] E. Y. Lam and J. W. Goodman, “A mathematical analysis of the DCT coefficient distribution for images,” *IEEE Trans. Image Process.*, Vol. 9, No. 10, pp. 1661-1666, Oct. 2000.
- [19] N. Kamaci, Y. Altunbasak, and R. Mersereau, “Frame bit allocation for the H.264/AVC video coder via Cauchy-density-based rate and distortion models,” *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 15, No. 8, pp. 994-1006, Aug. 2005.
- [20] Z. He and S. K. Mitra, “Optimum bit allocation and accurate rate control for video coding via  $\rho$ -domain source modeling,” *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 12, No. 10, pp.840-849, Oct. 2002.
- [21] H. Choi, J. Yoo, J. Nam, D. Sim, and I. V. Bajic, “Pixel-wise unified rate-quantization model for multi-level rate control,” *IEEE J. Sel. Topics Signal Process.*, Vol. 7, No. 6, pp. 1112-1123, Dec. 2013.
- [22] B. Lee and M. Kim, “Modeling rates and distortions based on a mixture of Laplacian distributions for inter-predicted residues in quadtree coding of HEVC,” *IEEE Signal Process. Lett.*, Vol. 18, No. 10, pp. 571-574, Oct. 2011.
- [23] B. Lee, M. Kim, T. Q. Nguyen, “A frame level rate control scheme based on texture and nontexture rate models for high efficiency video coding,” *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 24, No. 3, pp. 465-479, Mar. 2014.
- [24] C. W. Seo, J. H. Moon, and J. K. Han, “Rate control for consistent objective quality in high efficiency video coding,” *IEEE Trans. Image Process.*, Vol. 22, No. 6, pp. 2442-2454, Jun. 2013.
- [25] Z. W. Zhang, T. Jing, J. N. Han, Y. W. Xu, and F. Zhang, “A new rate control scheme for video coding based on region of interest,” *IEEE Access*, Vol. 5, pp. 13677-13688, Mar. 2017.
- [26] W. Gao, S. Kwong, H. Yuan, and X. Wang, “DCT coefficient distribution modeling and quality dependency analysis based frame-level bit allocation for HEVC,” *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 26, No. 1, pp. 139-153, Jan. 2016.
- [27] B. Li, H. Li, L. Li, and J. Zhang, “ $\lambda$  domain rate control algorithm for high efficiency video coding,” *IEEE Trans. Image Process.*, Vol. 23, No. 9, pp. 3841-3854, Sep. 2014.
- [28] B. Li, D. Zhang, H. Li, and J. Xu, QP determination by lambda value, document JCTVC-I0426, Geneva, CH, May 2012.
- [29] B. Li, H. Li, L. Li, and J. Zhang, rate control by R-lambda model for HEVC, document JCTVC-K0103, Shanghai, China, Oct. 2012.
- [30] G. J. Sullivan and T. Wiegand, “Rate-distortion optimization for video compression,” *IEEE Signal Process. Magazine*, Vol. 15, No. 6, pp. 74-90, Nov. 1998.
- [31] S. Boyd and L. Vandenberghe, “Unconstrained minimization,” in *Convex Optimization*, 7th ed., NY, USA: Cambridge University Press, pp. 457-520, 2009.
- [32] S. Mallat and F. Falzon, “Analysis of low bit rate image transform coding,” *IEEE Trans. Signal Process.*, Vol. 46, No. 4, pp. 1027-1042, Apr. 1998.
- [33] M. H. Hyun, B. Lee, and M. Kim, “A VVC intra rate control with small bit fluctuations using a Lagrange multiplier adjustment,” *IEEE Trans. Multimedia*, Vol. 26, pp. 6811-6821, Jan. 2024.
- [34] A. Browne, Y. Ye, and S. Kim, “Algorithm description for versatile video coding and test model 16(VTM16),” in *Proc. 25th Meeting, Teleconference*, pp. 1-256, 2022.
- [35] M. H. Hyun, B. Lee, and M. Kim, “A frame-level constant bit-rate control using recursive bayesian estimation for versatile video coding,” *IEEE Access*, Vol. 8, pp. 227255-227269, 2020.
- [36] Y. Chen et al., “Intra frame rate control for versatile video coding with quadratic rate-distortion modelling,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, pp. 4422-4426, 2020.