

# AI 기반 미디어 부호화 기술 표준화 동향

## Trends in the Standardization of AI-Based Media Coding Technology

강정원 (J.W. Kang, jungwon@etri.re.kr) 미디어부호화연구실 책임연구원/실장  
임성창 (S.-C. Lim, sclim@etri.re.kr) 미디어부호화연구실 책임연구원  
배성준 (S.J. Bae, sjbae@etri.re.kr) 미디어부호화연구실 책임연구원  
정세윤 (S.Y. Jeong, jsy@etri.re.kr) 미디어부호화연구실 책임연구원  
장인선 (I.S. Jang, jinsn@etri.re.kr) 미디어부호화연구실 책임연구원

### ABSTRACT

Global data traffic has surged due to the rise in internet users and the spread of smart devices and the Internet of Things. Audio and video content, which account for a large portion of network traffic, require efficient transmission and processing. Emerging challenges such as maintaining quality, optimizing for machine learning, and enabling real-time processing are difficult for traditional technologies to address. Artificial Intelligence (AI)-based media coding has emerged as a key solution, balancing data efficiency and quality while meeting the needs of both machines and humans. This study outlines the ongoing standardization efforts related to AI-based media coding technologies within the framework of MPEG media standards. Specifically, we focus on AI-driven coding techniques for both 2D and 3D video, as well as the standardization process for machine-centric video and audio coding technologies. The research explores the advancements and the collaborative initiatives underway to integrate AI into media compression, emphasizing the role of these technologies in shaping the future of media standards.

**KEYWORDS** AI, video coding, video coding for machine, audio coding for machine

## 1. 서론

인터넷 사용 인구의 증가와 스마트 기기 및 사물 인터넷의 확산으로 인해 전 세계적으로 데이터 트래픽이 폭발적으로 증가하고 있다. 특히, 오디오와 비디오 콘텐츠는 네트워크 트래픽의 상당 부분을

차지하며, 이러한 대규모 데이터의 효율적 전송과 처리는 필수적인 과제가 되었다. 현대의 복잡한 데이터 환경에서의 품질 유지, 기계학습을 위한 최적화, 실시간 처리 등 기존 기술로는 해결하기 어려운 요구사항들이 등장하고 있다. 이에 따라, AI 기반 미디어 부호화 기술은 데이터 효율성과 품질의 균

\* DOI: <https://doi.org/10.22648/ETRI.2025.J.400203>

\* 본 논문은 2025년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임[No. 2017-0-00072].



형을 유지하면서 기계와 인간의 다양한 요구를 충족시킬 수 있는 핵심기술로 부상하고 있다.

이에 본고에서는 MPEG 미디어 표준화와 관련하여 AI 기술을 적용하는 미디어 부호화 기술 표준화 현황에 대해 주로 살펴보고자 한다. 이를 위해 II장에서는 AI 기반 평면 및 입체 비디오 부호화를 위한 표준화 현황에 대해 설명하고, III장에서는 사람이 아닌 기계를 위한 비디오와 오디오 부호화를 위한 표준화 현황에 대해 소개하고자 한다.

## II. AI 기반 평면 및 입체 비디오 부호화 기술

### 1. AI 기반 평면 비디오 부호화 기술

최신 평면 비디오 부호화 국제 표준인 VVC(Versatile Video Coding)[1]를 비롯한 기존의 평면 비디오 부호화 표준들은 블록 기반 하이브리드 구조에 신호처리 기반의 다양한 부호화 기술을 추가하며 발전해 왔다. 그러나 이러한 전통적인 접근 방식은 압축률 향상에서 기술적 한계에 도달하고 있다. 이에 ITU-T Q6/SG21 VCEG(Video Coding Experts Group)과 ISO/IEC JTC 1/SC 29 MPEG(Moving Picture Experts Group)의 공동 비디오 부호화 표준화 단체인 JVET(Joint Video Experts Team)는 지난 10년간 강력한 컴퓨팅 자원과 방대한 학습 데이터를 기반으로 컴퓨터 비전과 영상 처리 분야에서 혁신적인 성과

를 보여준 신경망 기술에 주목하게 되었다[2].

신경망 기반 비디오 부호화 기술들이 표준화 과정에서 처음으로 제안되어 검토된 것은 VVC CfP(Call for Proposal) 응답을 통해서였으나[3], 최종 표준에는 구현 복잡도를 대폭 낮춘 형태의 MIP(Matrix-based Intra Prediction) 기술이 포함되었다[4]. 그림 1에서 보듯이, JVET은 VVC 표준 제정 직후 ‘NNVC(Neural Network-based Video Coding)’이라는 이름으로 기술 탐색 실험(EE: Exploration Experiments)을 시작했으며[5], Huawei, InterDigital, Bytedance, Tencent, Qualcomm, Ericsson, Nokia, Oppo 등 글로벌 기업들을 중심으로 신경망 기술을 보다 적극적으로 활용하는 새로운 비디오 부호화 표준의 가능성을 지속적으로 모색하고 있다.

신경망 기반 비디오 부호화 탐색 실험에서는 두 가지 방향의 신경망 기술 활용을 검토하고 있다. 첫 번째는 전체 부호화/복호화 과정을 종단간(End-to-End) 최적화된 단일신경망으로 구축하는 방식이고, 두 번째는 기존 블록 기반 하이브리드 구조에 신경망 기반 부호화 기술을 접목하는 방식이다. 두 번째 방식은 화면 내 예측, 화면 간 예측, 루프 필터 등의 기존 구성 요소들을 신경망 기반 기술로 대체하거나 후처리 필터, 초해상화 등의 새로운 신경망 기반 구성 요소를 추가하는 방식이다. 현재는 압축률 향상에 더 큰 잠재력을 보이는 두 번째 방식을 중심으로 탐색 실험이 진행되고 있다.

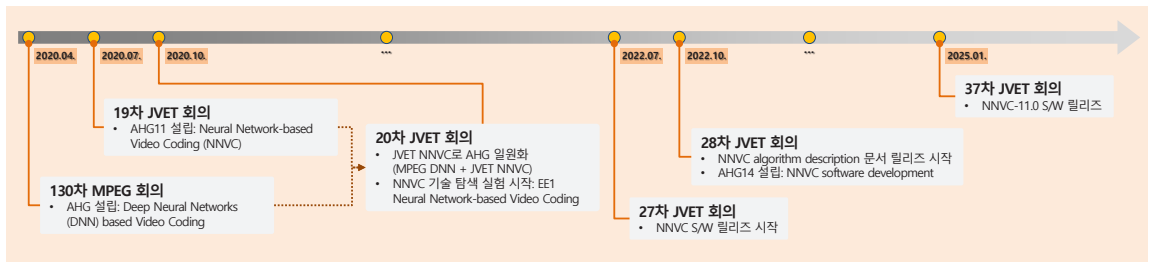


그림 1 JVET의 신경망 기반 비디오 부호화 표준화 진행 경과

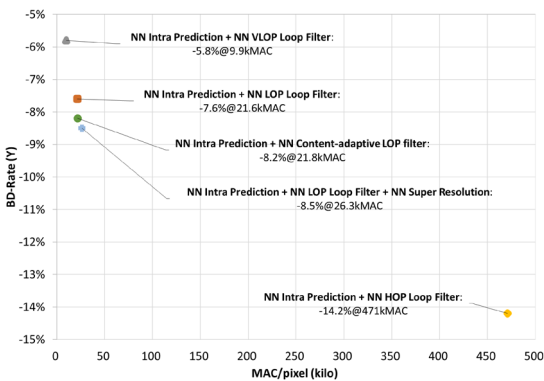
탐색 실험의 이름과 동일한 ‘NNVC’는 여러 JVET 표준화 회의를 거쳐 압축률 향상과 구현 복잡도 감소 측면에서 효과적인 기술들을 선별하여 구축된 탐색 실험 모델을 의미하기도 한다. NNVC는 알고리즘 설명 문서와 탐색 실험 모델 소프트웨어로 제공되며, 다음과 같은 다양한 신경망 기반 부호화 기술들을 포함하고 있다[6,7].

- 신경망 기반 루프 필터(Neural Network-based Loop Filter): 기존 블록 기반 하이브리드 구조 내 루프 필터 프로세스에 컨볼루션 신경망 기반 필터를 추가하여 복원된 영상의 객관적/주관적 화질 개선하는 기술로 다음의 3가지 방식 고려
  - ① LOP(Low complexity Operation Point): 낮은 복잡도로 구현된 루프 필터
  - ② VLOP(Very Low complexity Operation Point): 매우 낮은 복잡도로 경량화된 루프 필터
  - ③ HOP(High complexity Operation Point): 고성능을 목표로 하는 복잡한 구조의 필터
- 신경망 기반 화면 내 예측(Neural Network-based Intra Prediction): 주변 참조 샘플들로부터 복잡한 패턴을 예측하기 위해 완전 연결 신경망을 사용하는 화면 내 예측 방법

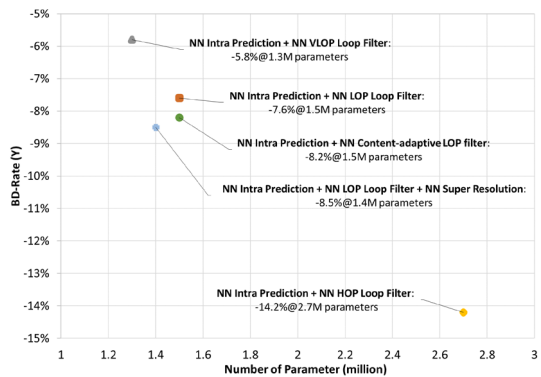
- 콘텐츠 적응형 신경망 필터(Content-adaptive Neural Network Filter): 각 영상 특성에 오버피팅(Overfitting)이 가능한 신경망 루프 필터 및 후처리 필터
- 신경망 기반 초해상도(Neural Network-based Super Resolution): 신경망을 통해 고해상도 영상을 생성하기 위한 후처리 필터

NNVC 소프트웨어는 11.0 버전까지 배포되었으며[8], 경량화된 신경망 추론을 위한 전용 라이브러리인 SADL(Small Ad-hoc Deep Learning)을 기반으로 구현되어 있다. 또한, 탐색 실험을 위한 명확한 평가 기준을 제시하기 위해 공통 실험 조건 및 평가 절차를 정립하고, NNVC 탐색 모델 소프트웨어 설정, 학습 및 실험 데이터셋을 포함한 학습/추론 조건, 압축 성능 측정 방법, 학습 및 추론 정보와 각 과정의 복잡도 측정 방법, 학습의 교차 검증 절차 등의 내용을 지속적으로 개선하고 있다[9].

NNVC 소프트웨어의 신경망 기반 부호화 기술들에 대한 압축 성능과 구현 복잡도를 공통 실험 조건과 평가 절차에 따라 평가한 결과는 그림 2와 같다[10]. VVC 표준의 참조 소프트웨어인 VTM(VVC



(a)



(b)

그림 2 신경망 기반 비디오 부호화 기술의 성능: (a) 압축 성능 대비 계산 복잡도, (b) 압축 성능 대비 파라미터 수

Test Model)을 성능 평가 비교 대상으로 사용했으며, 압축 성능에 대응하는 비트율 감소는 BD-Rate로 측정하였다. 신경망 기반 화면 내 예측, 루프 필터, 콘텐츠 적응형 신경망 필터, 신경망 기반 초해상도 기술을 조합한 실험 결과를 통해 VVC 표준 대비 최소 5% 이상의 비트율 감소가 가능함을 확인했다. 이는 신호처리 기반 비디오 부호화 기술과 비교해도 우수한 압축 성능이다. 다만 픽셀당 MAC(Multiply and ACcumulate) 연산 수와 모델 파라미터 수로 측정된 구현 복잡도는 신호처리 기반 부호화 기술보다 현저히 높은 상황이다.

현재 JVET는 신경망 기반 비디오 부호화 기술들의 표준화 가능성을 다각도로 검토하고 있으며, 기술 탐색 실험을 통해 우수한 압축 성능을 확인했다. VVC 이후의 새로운 표준화에서는 구현 복잡도를 낮추기 위한 기술 개발에 집중할 것으로 예상된다. 구현 복잡도가 개선된다면, 새로운 표준에 하나 이상의 신경망 기반 부호화 기술이 도입되어 압축 성능 향상에 크게 기여할 것으로 전망된다.

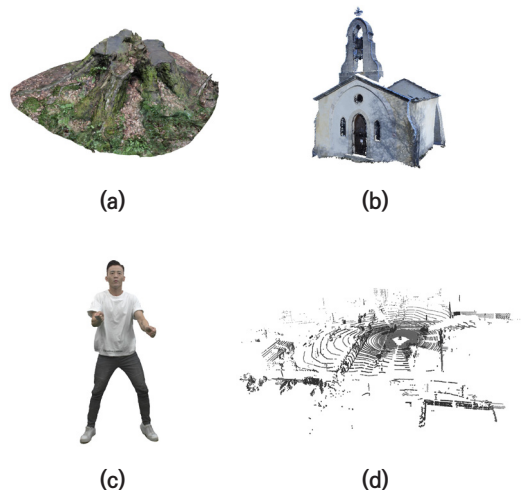
## 2. AI 기반 입체 비디오 부호화 기술

입체 비디오 부호화 기술 분야에서도 AI 기반의 부호화 기술 표준화가 추진되고 있다. 입체 비디오 부호화 기술은 전통적으로 MVC(Multiview Video Coding)[11]와 같은 다시점 비디오 부호화 표준기술을 기반으로 시작되었으며, 그래픽스 어플리케이션에 활용하기 위한 메쉬 및 포인트클라우드 기반의 부호화 기술이 최근 MPEG WG7(Coding of 3D Graphics and Haptics)을 중심으로 부상하고 있다. 특히, 포인트클라우드 기반의 볼류메트릭 객체 데이터를 2D 비디오 스트림으로 변환하여 압축하는 표준인 V-PCC(Video-based Point Cloud Compression) [12]와 3D 기하학적 공간에서 직접 압축하는 표준

인 G-PCC(Geometry-based Point Cloud Compression) [13]를 MPEG WG 7에서 제정하였다.

기존의 신호처리 기반 부호화 표준과 더불어 AI 기반 부호화 기술이 발전함에 따라, 2021년 7월 135차 MPEG 회의부터 MPEG WG 7에서 AI 기반 포인트클라우드 데이터를 압축하기 위한 AI-GC (AI-based Graphics Coding) 표준화 추진을 결정하고, 그림 3과 같은 이머시브 콘텐츠 및 라이다 콘텐츠 등에 대해 표 1과 같은 주제들로 기술탐색(EE)을 진행하였다.

기술탐색 결과를 바탕으로, 2024년 7월에 열린 148차 회의에서 CfP[14] 응답 기술들을 평가하여 실험 모델(TMAP) v.0를 선정하고, 이를 통해 본격적인 표준화 작업이 시작되었다. 실험 모델의 부호화 및 복호기 구조는 그림 4와 같다[15]. 기하정보는 AI 기반으로 압축하기 위해 희소 콘볼루션(Sparse Convolution)을 단계적으로 적용하여 고해상도의 복잡한 포인트클라우드를 최대한 단순화한 형태로 변



출처 Reproduced from VVCSoftware VTM, under the BSD License

그림 3 AI-GC 실험 대상 콘텐츠:

- (a) 메쉬추출형(Dense Static),
- (b) 정적포인트클라우드(Sparse Static),
- (c) 동적포인트클라우드(Dense Dynamic),
- (d) 동적희득(라이다)(Sparse Dynamic)

표 1 AI-GC의 주요 기술탐색 주제

EE분야	주제
EE5.0	AI Tools for Point Cloud Compression and Analysis
EE5.1	Deep Octree Coding
EE5.2	Deep Feature based PC Coding
EE5.3	AI-based Dynamic PC Coding
EE5.4	AI-based Attribute Coding
EE5.5	Hybrid Framework for AI-PCC Related Coding Experiments
EE5.6	Data Set Selection for CfP
EE5.7	Subjective Evaluation of Geometry-Only Compression
EE5.8	Anchors Generation

환한 후 이를 무손실 압축하고, 복호화 과정에서는 이로부터 다시 AI 기반의 예측을 통하여 원래 해상도의 포인트클라우드를 복원하는 방식을 적용하였고, 텍스처 정보는 G-PCC 표준에서도 적용된 신호 처리 기반의 RAHT 방식을 적용하였다. 실험 모델인 TMAP v.0의 성능은 V-PCC 대비 43.7%의 기하 정보량 감축 및 24.6%의 텍스처 정보량 감축 성능을 보인다[16].

AI-GC는 아직 표준화 초기 단계에 있으며, 앞으로 텍스처 정보 압축에도 AI 기반 부호화 기술이 적용될 가능성이 높을 것으로 전망된다.

### III. AI 기반 기계를 위한 미디어 부호화 기술

#### 1. 기계를 위한 비디오 부호화 기술

AI 기술의 발전에 따라, 기계와 알고리즘이 주요 소비 주체가 되는 비디오 트래픽 또한 급격히 증가하고 있다. 예를 들어, 2026년 말까지 전 세계 비디오 감시 카메라의 수는 약 10억 대에 이를 것으로 전망되는데, 이는 객체 인식 및 추적을 위한 AI 기술의 진보와 더불어 홈 보안 카메라, 공장 자동화 센서, 사물인터넷 기기 등 다양한 분야에서의 폭넓은 보급이 그 배경에 있다[17].

일반적으로 감시 카메라와 같은 저비용 기기는 컴퓨팅 리소스나 배터리 용량이 부족해 엣지 컴퓨팅 또는 클라우드 환경에서 AI 모델을 실행하는 것이 적합하다. 따라서, 기계를 위한 비디오 부호화 메커니즘의 필요성이 점점 증가하고 있다.

기계를 위한 비디오 부호화의 핵심은 AI 알고리즘과 인간의 시각적 관심 대상이 다르다는 점이다. 인간은 프레임 전체를 분석하여 맥락적으로 정보를 이해하지만, AI는 객체의 모양, 크기, 그리고 기하학적 이상 여부를 판단할 수 있는 프레임 일부 정보

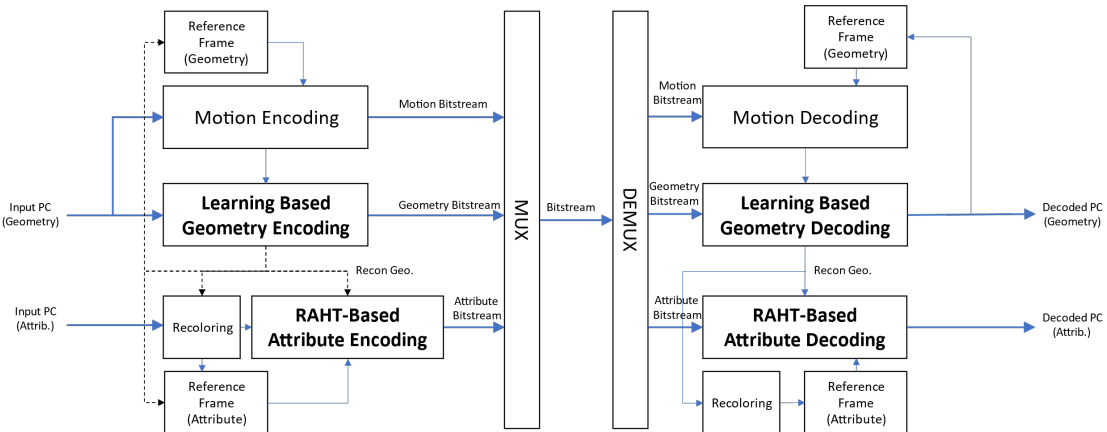


그림 4 AI-GC 부호화 및 복호화 과정

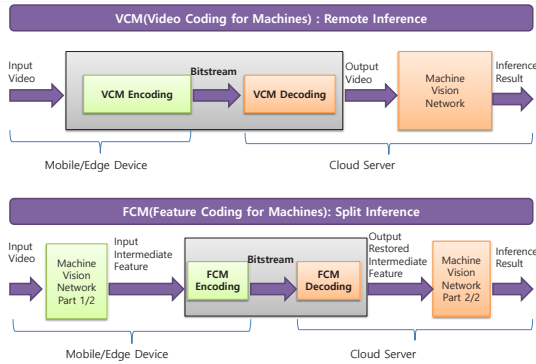


그림 5 VCM과 FCM 프레임워크

만 필요로 한다. 이러한 차이를 기반으로 기존 방식보다 압축률을 높이기 위한 새로운 기계 중심의 비디오 부호화 기술 개발이 요구되고 있다.

이러한 배경에서 MPEG WG 4(MPEG Video Coding)는 2019년부터 기계를 위한 비디오 부호화 표준화를 추진하여, 그림 5[32]와 같이 VCM(Video Coding for Machines)과 FCM(Feature Coding for Machines)이라는 두 가지 방식으로 표준화를 진행 중이다.

VCM은 비디오 데이터를 압축하여 전송하고, AI 작업은 엣지 서버나 클라우드에서 수행하는 원격 추론(Remote Inference) 방식이다. 반면, FCM은 AI 작업을 두 단계로 나누어, 센서 기기에서 생성된 중간 데이터(Intermediate Features)를 압축하여 전송하고, 클라우드에서 나머지 AI 작업을 수행하는 분할 추론(Split Inference) 방식이다. 이 방식은 VCM보다 높은 압축률을 제공해 대역폭을 절약할 수 있지만, 센서 기기에서 일부 AI 모델을 실행해야 하므로 약간의 에너지 비용이 더 발생한다. FCM의 주요 장점은 VCM보다 높은 압축률을 제공하며, 중간 데이터인 피쳐 자체가 가지는 특성으로 인해 개인정보 보호와 보안 기능도 기본적으로 제공한다는 점이다.

VCM 표준화는 2022년 10월 140차 MPEG 회의에서 CfP 응답기술 평가를 통해 본격적으로 시작되었으며, 2023년 10월 144차 MPEG 회의에서

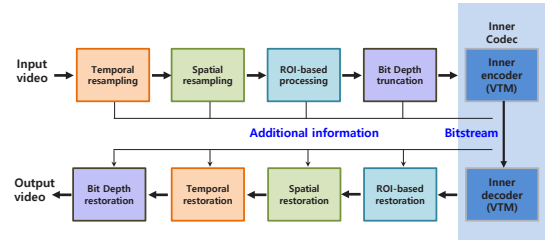


그림 6 VCM 참조 소프트웨어(VCM-RS) 구조

WD(Working Draft)가 발행되었다. 이후 MPEG-AI Part 2로서, 2025년 1월 CD(Committee Draft) 단계와 2025년 7월 DIS(Draft International Standard) 단계를 거쳐, 2026년 1월에 최종 표준안인 FDIS(Final Draft International Standard) 단계로 표준화가 완료될 예정이다[18].

그림 6[19]은 VCM 참조 소프트웨어인 VCM-RS(VCM-Reference Software) 구성도를 나타내며, VCM의 요소 기술 구성을 보여준다. 입력된 비디오에서 기계에 불필요한 데이터를 제거하는 전처리 과정 후, VVC와 같은 기존 압축 기술을 내부 코덱으로 사용해 비디오를 압축한다. 이후 복호화와 복원 과정을 거치는 단계로 구성된다[19-21].

VCM에서 주요 전/후처리 기술은 다음과 같다.

- ROI 전/후처리: 성능에 중요한 ROI(Region Of Interest) 영역을 제외한 불필요한 부분을 제거하거나 스케일 변환 등의 전처리를 통해 데이터를 압축 및 전송한 후, 복호화하여 원래 스케일로 복원하는 기술
- 시간적/공간적 샘플링: 목표 성능에 필요한 시간적 또는 공간적 해상도보다 입력 해상도가 높은 경우, 이를 낮춰 압축한 뒤 복호화 과정에서 원래 해상도로 복원하는 기술
- 비트 심도 샘플링: 신호의 비트 심도가 기계에 불필요하게 높은 경우, 이를 낮춰 압축하고 복호화 후 복원하는 기술

FCM 표준화는 초기에는 VCM 표준화의 일부로 시작되었으나, 2022년 7월 139차 MPEG 회의에서 FCM을 위한 CFe(Call for Evidence) 발행을 계기로 분리되었다. 이후, 2023년 10월 144차 MPEG 회의에서 CFe 응답기술 평가를 통해 본격적인 표준화가 시작되었고, 2024년 4월에는 PWD(Preliminary Working Draft) 초안이 발행되었다. 향후 MPEG-AI Part 4로서, 2025년 7월 CD 단계와 2026년 1월 DIS 단계를 거쳐, 2026년 7월에 최종 표준안인 FDIS 단계로 표준화가 완료될 예정이다[22].

그림 7[23]은 FCM 참조 소프트웨어인 FCTM(FCM Test Model) 구성도로, FCM의 요소 기술 구성을 보여준다. 입력된 중간 피쳐는 서로 다른 해상도를 가지는 다계층 피쳐로, 계층 간 중복성이 높아 이를 단일 계층으로 변환하는 피쳐 리덕션 기술, 내부 코덱이 요구하는 프레임 포맷으로 피쳐를 변환하는 기술, 내부 코덱, 복호화된 프레임을 피쳐로 역변환하는 기술, 그리고 단일 계층을 다계층으로 복원하는 피쳐 복원 기술로 구성된다.

FCM은 VCM과 유사하게 내부 코덱을 사용하는 구조이지만, 피쳐 리덕션 과정에서 핵심적인 압축이 이루어지는 점이 특징이다[23-25].

VCM과 FCM 표준화를 위해 ETRI와 대학들이 공동연구를 통해 초기 단계부터 협력해 왔으며, VCM CFe와 FCM CFe에서 각각 공동 1위를 달성하고 주요 핵심기술을 채택하면서 표준화를 주도하고 있다.

특히 FCM의 경우, 2024년 1월 145차 MPEG 회의에서 ETRI와 경희대학교가 공동 개발한 L-MSFC(Learnable Multi-scale Feature Compression)[26] 기술이 FCTM 2.0의 핵심기술인 피쳐 리덕션 기술로 채택되었다. 또한, 2025년 1월 149차 MPEG 회의에서는 ETRI와 경희대학교가 공동 개발한 복잡도를 줄이면서도 성능을 크게 개선한 LightFCTM 기술이

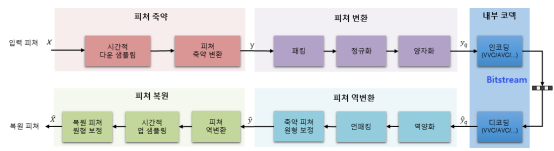


그림 7 FCM 참조 소프트웨어(FCTM) 구조

FCTM 6.0의 핵심기술로 채택되었다.

VCM과 FCM은 각각 원격 추론과 분할 추론의 장점을 기반으로 하는 기계를 위한 비디오 부호화 기술로, 기존의 인간을 위한 비디오 부호화 기술들보다 혁신적인 압축률 제공을 가능하게 하여 다양한 산업 응용에서 큰 변화를 가져올 것으로 기대된다.

## 2. 기계를 위한 오디오 부호화 기술

기계를 위한 음향 데이터는 오디오의 공간적 분포를 중요한 정보로 포함하고 있으며 다채널 형태로 제공되는 경우가 많다. 이런 방대한 음향 데이터가 상시 생성되면서, 기존의 수작업 중심 처리 방식은 지연 시간과 효율성 면에서 한계를 드러내고 있다. 이에 따라, 지능형 플랫폼을 지원하기 위한 고압축·저지연 오디오 압축 기술의 필요성이 점차 커지고 있다.

이러한 배경으로 2022년 10월 140차 MPEG 회의에서 Fraunhofer IDMT가 기계를 위한 오디오 부호화(ACoM: Audio Coding for Machines)를 새로운 표준화 아이템으로 제안하였으며 MPEG WG 2(Technical Requirements)와 MPEG WG 6(Audio Coding)에서 ACoM의 사용 사례와 요구사항에 대한 공동 논의가 시작되었다[27,28]. 2024년 4월 146차 MPEG 회의에서 공동 논의를 MPEG WG 6으로 이관하여 표준화 기술 탐색(Exploration) 단계가 시작되었으며 [29], 2024년 7월 147차 MPEG 회의에서는 CFe를 진행하기로 결정하고 관련 작업 계획을 구체화하는

등 표준화가 점차 본격화되고 있다[30].

MPEG WG 6에서 논의 중인 ACoM은 오디오, 다차원 스트림 또는 해당 신호에서 추출된 특징(Feature)을 압축하기 위한 비트스트림과 데이터 포맷을 정의하며, 이러한 비트스트림과 데이터 정의에 있어서 비트율과 데이터 크기 측면에서의 효율성을 목표로 한다. ACoM은 복호화된 다음에도 기계의 작업 성능 저하를 최소화하면서 다양한 작업을 수행할 수 있도록 한다. 복호화된 데이터는 기계 단독 또는 기계와 인간의 혼합 소비에 사용할 수 있도록 지원하며, 또한 포맷에는 오디오나 다차원 스트림의 획득 방법을 설명하는 메타데이터도 포함되어 데이터 활용성을 높이고자 한다.

ACoM 표준화는 Phase 1과 Phase 2로 구분하여 진행될 예정이다. Phase 1은 응용 분야에 구애받지 않도록 데이터를 거의 무손실로 부호화하여, 표준 기반 비트스트림을 통해 데이터 교환을 간소화함으로써 산업에서 유용하게 활용될 수 있다. Phase 2에서는 특징 추출 방식을 추가하고, 해당 특징들을 응용 분야에 맞게 최적화하는 것을 포함한다. 일반적인 ACoM 시스템 구조는 부호화기와 복호화기로 구성되며 시스템의 입력은 다음 중 하나와 메타데이터가 될 수 있다.

- 오디오 신호(1차원 혹은 다차원)
- 다차원 스트림(예: 의료데이터)
- 추출된 특징(Phase 2만 해당)

MPEG WG 6에서는 산업체, 의료, 미디어 서비스 등 다양한 응용분야에 대한 ACoM의 사용 사례를 모으고(표 2), 각 시장 규모를 조사하는 한편, 이를 기반으로 한 요구사항을 수립하고 있다. 요구사항으로는 비트스트림 압축 효율성, 다중임무 수행, 근사 무손실 부호화, 개인정보 보호, 메타데이터 부호화, 기계와 사람의 하이브리드 소비, 엣지컴퓨팅

표 2 ACoM 사용 사례

Use Cases		Application
UC1	Predictive Maintenance	Industrial
UC2	Process Control	
UC3	In-line Testing	
UC4	End-of-line Testing	
UC5	Traffic Monitoring and Control	Site Monitoring
UC6	Construction Site Monitoring	
UC7	Speech Recognition and Acoustic Scene Analysis	Life-logging Contents Service
UC8	Timed Medical Data	Medical
UC9	Flexible Medical Data	
UC10	UGC Analysis	Contents Service
UC11	Live Stream Content Analysis	
UC12	Artistic creation	

및 타임스탬프, 주파수 대역폭, 채널 수 등에 대한 세부 사항이 포함되어 있다[31]. 148차 MPEG 회의에서는 ACoM 표준화 기술 탐색을 위해 부호화기 입력을 위한 메타데이터 서술 방법과 오디오(특히, 다채널) 데이터 수집 계획을 구체화하고, 이를 바탕으로 CFE 작업 계획을 업데이트하였다. ACoM CFE에 대한 응답으로는 아래의 사항이 포함되어야 한다. 첫째, 제출된 기술이 다채널 오디오의 최신 무손실 코딩(MPEG ALS(Audio Loseless Coding), SLS(Scalable Lossless Coding) 등)과 비교하여, 랜덤 접근, 확대, 부분 추출 등의 기능을 고려했을 때, 사람 중심 성능 관점에서 호환 가능한 성능을 제공하는지를 확인해야 한다. 동시에, 기계 소비 관점에서는 해당 기술이 기계의 임무 수행 성능 기준을 충족하는지도 평가해야 한다. 둘째, 제출된 메타데이터 서술 방식이 텍스트, 자막, 센서 노드, 마이크 유형, 좌표, 방향, 오디오 테크 등 다양한 유형의 메타데이터를 효율적으로 포함할 수 있는 공통 메타데이터 형식을 제공하는지를 평가해야 한다.

149차 MPEG 회의에서는 WG 2와의 공동회의를



통해 ACoM 요구사항 및 사용 사례 등을 검토하고 이를 바탕으로 CFe 발간 여부를 논의할 예정이다.

## IV. 결론

MPEG에서 진행되고 있는 AI 기반의 비디오와 오디오 부호화 표준화 동향에 대해 살펴보았다. 평면비디오 부호화 분야에서는 신경망 기반 부호화 기술들에 대한 지속적인 기술 탐색을 통해 구현 복잡도를 낮추는 노력이 계속되고 있으며, 이는 향후 추진될 새로운 표준에 도입될 가능성을 높이고 있다. 또한, 입체 비디오 부호화 분야에서는 포인트 클라우드 데이터 압축을 위한 AI-GC 표준화가 진행되고 있으며, 이를 기반으로 다양한 AI 기반 입체 비디오 부호화 표준화로 확장될 것으로 예상된다.

기계를 위한 비디오 부호화 표준인 FCM은 2026년 7월 표준화 완료로 목표로 AI 기반의 부호화 기술 표준화가 진행 중이며, 기계를 위한 오디오 부호화 기술은 본격적인 표준화 추진을 위한 기술 탐색을 진행하고 있다.

AI 기반의 미디어 부호화 기술 개발은 AI 기술의 발전과 함께 더욱 활발하게 발전하고 있으며, 이를 통해 더 효율적이고 혁신적인 미디어 처리와 전달이 가능해질 것으로 기대한다.

### 약어 정리

ACoM	Audio Coding for Machines
AI-GC	AI-based Graphics Coding
ALS	Audio Lossless Coding
BD	Bjontegaard Delta
CD	Committee Draft
CE	Core Experiment
CFe	Call for Evidence
CfP	Call for Proposal
DIS	Draft International Standard
EE	Exploration Experiments

FCM	Feature Coding for Machines
FCTM	FCM Test Model
FDIS	Final Draft International Standard
G-PCC	Geometry-based Point Cloud Compression
HOP	High complexity Operation Point
JVET	Joint Video Experts Team
L-MSFC	Learnable Multi-scale Feature Compression
LOP	Low complexity Operation Point
MAC	Multiply and ACcumulate
MIP	Matrix-based Intra Prediction
MPEG	Moving Picture Experts Group
MVC	Multiview Video Coding
NNVC	Neural Network-based Video Coding
PWD	Preliminary Working Draft
RAHT	Region-Adaptive Hierarchical Transform
ROI	Region of Interest
SADL	Small Ad-hoc Deep Learning
SLS	Scalable Lossless Coding
VCEG	Video Coding Experts Group
VCM	Video Coding for Machines
VCM-RS	VCM Reference Software
VLOP	Very Low complexity Operation Point
V-PCC	Video-based Point Cloud Compression
VTM	VVC Test Model
VVC	Versatile Video Coding
WD	Working Draft

### 참고문헌

- [1] ISO/IEC, "Versatile Video Coding, Standard ISO/IEC 23090-3, ISO/IEC JTC 1," Jul. 2020.
- [2] 최진수 외, "인공지능 기반 비디오 부호화 기술 동향," 주간 기술동향 1970호, 2020. 10, pp. 15-29.
- [3] S. Liu et al., "JVET AHG report: Neural Networks in Video Coding(AHG9)," 10th Joint Video Experts Team(JVET) meeting, JVET-J0009, Apr. 2018.
- [4] B. Bross et al., "Overview of the Versatile Video Coding(VVC) Standard and its Applications," IEEE Trans. Circuits Syst. Video Technol., vol. 30, no. 10, 2021, pp. 3736-3764.

- [5] E. Alshina et al., "Description of Exploration Experiments on NN-based video coding," 20th Joint Video Experts Team(JVET) meeting, JVET-T2023, Oct. 2020.
- [6] F. Galpin et al., "Description of algorithms version 9 and software version 11 in neural network-based video coding(NNVC)," 36th Joint Video Experts Team(JVET) meeting, JVET-AJ2019, Nov. 2024.
- [7] 최기호, "JVET 신경망 기반 비디오 코딩 기술 연구 동향," 방송과 미디어, 제28권 제1호, 2023. 1, pp. 29-37.
- [8] VVCSoftware Homepage. [https://vcgit.hhi.fraunhofer.de/jvet-ahg-nnvc/VVCSoftware\\_VTM](https://vcgit.hhi.fraunhofer.de/jvet-ahg-nnvc/VVCSoftware_VTM)
- [9] E. Alshina et al., "Common test conditions and evaluation procedures for neural network-based video coding technology," 36th Joint Video Experts Team(JVET) meeting, JVET-AJ2016, Nov. 2024.
- [10] E. Alshina et al., "JVET AHG report: Neural network-based video coding," 37th Joint Video Experts Team(JVET) meeting, JVET-AK0011, Jan. 2025.
- [11] ISO/IEC 14496-10:2008/Amd 1:2007 Information technology — Coding of audio-visual objects — Part 10: Advanced Video Coding — Amendment 1: Multiview Video Coding
- [12] ISO/IEC 23090-5:2021 Information technology — Coded representation of immersive media — Part 5: Video-based Point Cloud Compression
- [13] ISO/IEC 23090-9:2021 Information technology — Coded representation of immersive media — Part 9: Geometry-based Point Cloud Compression
- [14] Call for Proposals for AI-based Point Cloud Coding, N365, ISO/IEC JTC 1/SC 29/WG 2, April 2024.
- [15] Preliminary working draft of AI-based point cloud coding, N1059, ISO/IEC JTC 1/SC 29/WG 7, Nov. 2024.
- [16] TMAP v0 for AI-based point cloud coding, N1060, ISO/IEC JTC 1/SC 29/WG 7, Nov. 2024.
- [17] F. Racape, "Video Coding for Machines: The Need for Compression," interdigital, Sep. 2024. <https://www.interdigital.com/post/video-coding-for-machines-the-need-for-compression>
- [18] <https://sd.iso.org/projects/project/88879/overview>
- [19] Algorithm description of tools in VCM reference software, N593, ISO/IEC JTC 1/SC 29/WG 4, Dec. 2024.
- [20] Working draft 5 of video coding for machines, N591, Dec. ISO/IEC JTC 1/SC 29/WG 4, 2024.
- [21] <https://git.mpeg.expert/MPEG/Video/VCM/VCM-RS>
- [22] C. Rosewarne and Y. Zhang, "BoG on Feature Coding for Machines," m71631, Jan. 2025.
- [23] Algorithm description of FCTM, N599, ISO/IEC JTC 1/SC 29/WG 4, Dec. 2024.
- [24] Preliminary WD of feature coding for machines, N601, ISO/IEC JTC 1/SC 29/WG 4, Dec. 2024.
- [25] <https://git.mpeg.expert/MPEG/Video/fcm/fctm>
- [26] Y.W. Kim et al., "End-to-End Learnable Multi-Scale Feature Compression for VCM," IEEE Trans. Circuits Syst. Video Technol., vol. 34, no. 5, May. 2024, pp. 3156-6167.
- [27] T. Sporer, "Proposal for New Work Item: Audio Coding for Machines," m61162, Oct. 2022.
- [28] Use Cases and Requirements for Audio Coding for Machines I(ACoM), N252, ISO/IEC JTC1 SC 29/WG 2, Oct. 2022.
- [29] Use Cases and Requirements on Audio Coding for Machines, N252, ISO/IEC JTC1 SC 29/WG 6, April 2024.
- [30] Workplan on Audio Coding for Machines, N269, ISO/IEC JTC1 SC 29/WG 6, July 2024.
- [31] Use cases and requirements on Audio coding for machines, N287, ISO/IEC JTC1 SC 29/WG 6, Nov. 2024.
- [32] 조용성 외, "초실감메타버스 기술 동향 및 발전 전망," 전자통신 동향분석, 제39권 제2호, 2024. 4, pp. 1-11.