MPEG에서의 AI 기반 미디어 부호화: 차세대 압축을 위한 표준화 동향과 향후 방향

AI-Based Media Coding in MPEG: Standardization Trends and Future Directions for Next-Generation Compression

강정원 (J.W. Kang, jungwon@etri.re.kr) 미디어부호화연구실 책임연구원/실장
방건 (G. Bang, gbang@etri.re.kr) 미디어부호화연구실 책임연구원
임성창 (S.-C. Lim, sclim@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

The continuous expansion of the Internet-using population, coupled with the proliferation of smart devices, the Internet of Things, and intelligent services, is driving a rapid surge in global data traffic. As audio and video content constitute a large share of network traffic, efficient media compression has become increasingly important. However, emerging requirements for machine learning, including real-time processing, quality preservation, and optimization, pose significant challenges for conventional signal-processing-based coding technologies. In this context, artificial intelligence (AI)-based media coding has emerged as a key solution capable of balancing compression efficiency and quality, while supporting both human- and machine-oriented consumption.

This article reviews ongoing standardization efforts for AI-based media coding within the MPEG framework, focusing on AI-driven coding technologies for 2D and immersive videos, as well as emerging standards for machine-centric video and audio coding. By examining recent technical progress and collaborative standardization activities, this study highlights the growing role of AI-based approaches in shaping next-generation media compression standards.

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

I. 서론

최근 인터넷 사용 인구의 지속적인 증가와 스마트 기기, 사물인터넷(IoT), 지능형 서비스의 확산으로 전 세계 데이터 트래픽이 빠르게 증가하고 있다. 특히 오디오와 비디오 중심의 미디어 데이터는 전체 네트워크 트래픽에서 높은 비중을 차지하고 있으며, 실감형 콘텐츠(VR · AR) 및 초고화질(8K) 스트리밍 서비스의 확산은 이러한 증가 추세를 더욱 가속화하고 있다.

* DOI: https://doi.org/10.22648/ETRI.2026.J.410203
* 본 논문은 2025년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No. 2017-0-00072)

이에 따라 대규모 미디어 데이터를 효율적으로 전송 · 저장 · 처리하기 위한 부호화 기술의 중요성이 지속적으로 확대되고 있다. 동시에 실시간 처리, 고품질 유지, 기계학습 활용 등 새로운 요구사항이 등장하면서, 기존 신호처리 기반 미디어 부호화 기술의 성능적 · 구조적 한계가 점차 부각되고 있다.

이러한 환경 변화에 대응하기 위해 인공지능(AI)을 활용한 미디어 부호화 기술이 대안으로 주목받고 있다. AI 기반 접근 방식은 데이터의 구조적 특성을 학습 기반으로 모델링함으로써 압축 효율과 복원 품질을 동시에 향상시킬 수 있는 가능성을 제시한다. 더 나아가 최근에는 기계와 알고리즘이 주요 소비 주체로 부상하면서 기계를 위한 비디오 · 오디오 부호화 기술에 대한 요구 또한 본격화되고 있다.

이에 국제 표준화 기구인 MPEG에서는 AI 기반 비디오 및 입체 미디어 부호화뿐만 아니라, 기계를 위한 비디오 · 오디오 부호화 기술의 표준화를 활발히 추진하고 있다. 본고는 이러한 흐름을 바탕으로, AI 기술이 적용된 미디어 부호화 표준화의 최근 동향과 기술적 특징을 정리하고, 관련 표준화 활동의 진행 현황과 향후 방향을 살펴보고자 한다.

II. AI 기반 영상 부호화 표준화 동향

1. 신경망 기반 평면 비디오 부호화의 표준화 현황

2020년에 제정된 평면 비디오 부호화 국제 표준인 VVC(Versatile Video Coding)[1]를 비롯한 기존 비디오 부호화 표준은 블록 기반 하이브리드 구조를 중심으로, 다양한 신호처리 기반 기술을 단계적으로 확장 · 고도화하며 발전해 왔다. 그러나 수십년 동안 지속되어 온 이러한 전통적인 접근 방식은 최근에는 구조적 복잡성 증가와 함께 압축 성능 향상 폭이 점차 둔화되는 기술적 한계에 직면하고 있다.

이와 같은 상황에서 ITU-T Q6/SG21 산하 VCEG(Video Coding Experts Group)과 ISO/IEC JTC 1/SC 29 MPEG(Moving Picture Experts Group)이 공동 운영하는 JVET(Joint Video Experts Team)은 압축 성능 향상을 위한 대안으로 신경망 기반 접근 방식에 주목하고 있다[2]. 최근 컴퓨터 비전 및 영상 처리 분야에서 딥러닝 기술이 비약적인 성과를 보임에 따라, 이를 비디오 부호화 구조에 접목하려는 시도가 활발히 이루어지고 있다. 신경망 기반 부호화는 이미지 분야에서는 2017년경부터, 비디오 분야에서는 2019년 이후 본격적으로 연구가 진행되었으며, 복잡한 영상 패턴과 시공간적 상관관계를 비선형적으로 모델링함으로써 기존 방식 대비 높은 압축 성능 향상 가능성을 제시하고 있다.

신경망 기반 비디오 부호화 기술은 VVC 표준화 과정에서 CfP(Call for Proposal) 응답을 통해 처음으로 공식 검토되었으나[3], 당시에는 구현 복잡도와 실용성 측면의 제약으로 인해 구현 복잡도를 크게 완화한 형태의 MIP(Matrix-based Intra Prediction) 기술만이 표준에 반영되었다[4]. 이에 따라, JVET은 VVC 표준 제정 직후 'NNVC(Neural Network-based Video Coding)'라는 명칭하에 신경망 기반 기술의 적용 가능성을 체계적으로 검증하기 위한 탐색 실험(EE: Exploration Experiments)에 착수하였다[5]. 현재는 Huawei, Inter Digital, ByteDance, Tencent, Qualcomm, Ericsson, Nokia, Oppo 등 글로벌 주요 기업을 중심으로, 신경망 기술을 더욱 적극적으로 활용하는 차세대 신경망 기반 비디오 부호화 표준의 가능성에 대한 논의와 기술 검토가 지속적으로 이루어지고 있다.

신경망 기반 비디오 부호화 탐색 실험에서는 신경망 기술 활용 방식을 크게 세 가지로 구분하여 검토하고 있다. 첫 번째는 부호화와 복호화 전 과정을 단일 종단 간(End-to-End) 신경망으로 구성하는 방

식이다. 이 접근법은 기존 블록 기반 구조를 완전히 대체하는 개념으로, 이론적으로는 높은 압축 성능을 기대할 수 있으나 계산 복잡도 및 구현 부담이 크다는 한계가 있다. 두 번째는 기존 블록 기반 하이브리드 부호화 구조를 유지하면서 신경망 기반 부호화 기술을 부분적으로 접목하는 방식이다. 이 방식은 화면 내 예측, 화면 간 예측, 루프 필터 등 부호화 요소를 신경망 기반 기법으로 대체하거나, 후처리 필터 및 초해상화와 같은 새로운 기능을 추가하는 접근이다. 세 번째는 종단 간 신경망 부호화를 통해 압축 · 복원된 영상을 기존 블록 기반 하이브리드 부호화 구조에서 참조 영상(Reference Picture)으로 활용하는 방식이다. 이 접근법은 신경망 기반 기술의 장점을 간접적으로 활용하면서도 기존 표준 구조와의 호환성을 유지할 수 있는 특징을 갖는다. 현재 JVET에서는 실용성과 구현 가능성을 고려하여, 기존 블록 기반 하이브리드 구조에서 압축 성능 향상을 도모할 수 있는 두 번째와 세 번째 방식을 중심으로 탐색 실험을 진행하고 있다.

NNVC는 탐색 실험의 명칭이면서 동시에 이러한 탐색 실험을 위한 참조 모델의 명칭으로, 다수의 JVET 회의를 통해 압축 성능 향상과 구현 복잡도 측면에서 유효성이 검증된 신경망 기반 기술이 통합되어 있다. NNVC는 다음과 같은 핵심 기술을 포함한다[6,7].

신경망 기반 루프 필터(Neural Network-based Loop Filter)는 기존 블록 기반 하이브리드 구조의 루프 필터 과정에서 컨볼루션 신경망을 적용하여 복원 영상의 객관적 · 주관적 화질을 개선하는 기술로 복잡도 수준에 따라 LOP(저복잡도), VLOP(초저복잡도), HOP(고복잡도)의 세 가지 필터가 존재한다. 신경망 기반 화면 내 예측은 주변 참조 샘플을 입력으로 완전 연결 신경망을 활용하여 복잡한 공간 패턴을 예측하는 기술이다. 콘텐츠 적응형 신경망 필터는 영상 특성에 맞춰 학습된 모델을 오버피팅(Overfitting)하여 적용하는 루프 필터 및 후처리 필터 기술이다. 신경망 기반 초해상도는 저해상도 영상을 고해상도로 복원하는 후처리 필터 기술이며, 딥 참조 영상은 학습 기반으로 생성된 참조 영상을 화면 간 예측에 활용한다. 또한 하이브리드 종단 간 화면 내 부호화는 종단 간 신경망으로 복원된 영상을 기존 하이브리드 부호화 구조의 화면 내 영상으로 활용하는 프레임워크를 제공한다.

NNVC 소프트웨어는 15.0 버전까지 배포되었으며[8], 경량 신경망 추론 라이브러리 SADL(Small Ad-hoc Deep Learning)을 기반으로 구현되었다. 또한, 탐색 실험의 객관성과 재현성을 확보하기 위해 공통 실험 조건과 평가 절차가 수립되었으며, NNVC 소프트웨어 설정을 비롯해 학습 · 실험 데이터셋을 포함한 학습 및 추론 조건, 압축 성능과 계산 복잡도 측정 방법, 학습 과정의 교차 검증 절차 등을 지속적으로 고도화되고 있다[9].

공통 실험 조건과 평가 절차에 따라 NNVC 기술의 압축 성능과 구현 복잡도를 평가한 결과는 그림 1에 제시되어 있다[10]. 비교 대상으로는 VVC 표준의 참조 소프트웨어인 VTM(VVC Test Model)을 사용하였으며, 압축 성능은 BD-Rate 기반의 비트율 감소로 평가하였다. 신경망 기반 화면 내 예측, 루프 필터, 콘텐츠 적응형 필터, 초해상도, 딥 참조 영상 기술을 조합한 실험 결과, VVC 대비 최소 5% 이상의 비트율 감소가 가능함을 확인하였다.

이는 기존 신호처리 기반 부호화 기술 대비 매우 우수한 성능이나, 픽셀당 MAC(Multiply-Accumulate) 연산 수와 모델 파라미터 수로 측정되는 구현 복잡도는 여전히 주요 과제로 남아 있다. 현재 JVET은 탐색 실험을 통해 신경망 기반 비디오 부호화 기술의 표준화 가능성을 다각도로 검토하고 있으며, 향후 표준화 단계에서는 구현 복잡도 저감을 위한 기

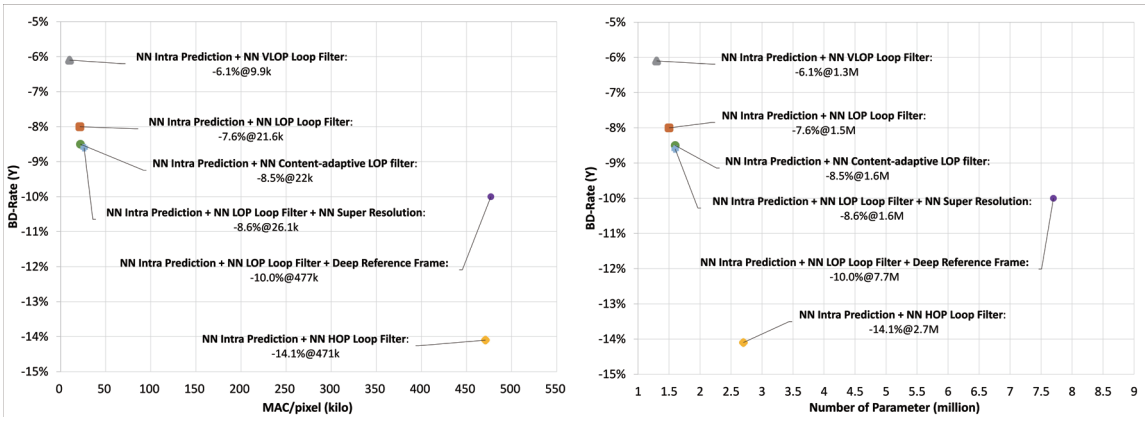

그림 1 신경망 기반 비디오 부호화 기술의 성능: (a) 압축 성능 대비 계산 복잡도, (b) 압축 성능 대비 파라미터 수

술 개발에 중점을 둘 것으로 예상된다. 구현 복잡도 측면에서 충분한 개선이 이루어질 경우, 일부 신경망 기반 부호화 기술이 새로운 표준에 반영될 가능성이 있는 것으로 전망된다.

2. AI 기반 입체 비디오 부호화 표준화 동향

메타버스, 디지털 트윈, 자율주행, 실감형 콘텐츠 산업의 확산으로 입체 비디오 및 3차원 콘텐츠 수요가 증가하고 있으며, 포인트 클라우드 등 대용량 3D 데이터의 실시간 전송 · 저장을 위한 고효율 압축 기술의 중요성이 커지고 있다. 이에 따라 국제 표준화 기구인 MPEG에서는 전통적인 신호처리 기반 압축 기술을 넘어, 인공지능을 활용한 차세대 입체 비디오 부호화 기술의 표준화를 본격적으로 추진하고 있다.

입체 비디오 부호화 기술은 초기에는 MVC(Multi-view Video Coding)[11]와 같이 다수의 카메라로 획득한 영상을 효율적으로 압축하는 다시점 영상 부호화를 중심으로 발전해 왔다. 이후 자유 시점 영상과 볼류메트릭 비디오로 응용이 확대되면서, 그림 2와 같은 메쉬와 포인트 클라우드 등 3차원 그래픽 표현 방식이 핵심 기술로 주목받았으며, 관련 부호화 기술은 현재 MPEG WG7(Coding of 3D Graphics and Haptics)에서 중점적으로 논의되고 있다.

WG7에서는 다양한 응용 환경을 고려하여 두 가지 대표적인 포인트 클라우드 압축 표준을 개발해 왔다. V-PCC(Video-based Point Cloud Compression)

출처 Reproduced from CTC on AI-based PCC under BSD Lincense

그림 2 AI-GC 실험 대상 콘텐츠

[12]는 3차원 포인트 클라우드를 2차원 영상 패치로 투영한 후 기존 영상 부호화 기술을 적용하는 방식으로, 안정적인 압축 성능과 기존 영상 인프라와의 높은 호환성을 특징으로 한다. 반면, G-PCC (Geometry-based Point Cloud Compression)[13]는 3차원 공간에서 기하/속성 정보를 직접 부호화하는 방식이다.

한편, 딥러닝 기술의 발전으로 데이터의 구조적 특성을 학습 기반으로 모델링할 수 있게 되면서, 기존 신호처리 방식의 한계를 넘어 압축 효율과 복원 품질을 동시에 개선할 가능성이 확대되었다. 이러한 흐름에 따라 MPEG WG7은 2021년 7월 제135차 MPEG 회의를 기점으로 AI 기반 포인트 클라우드 압축을 목표로 하는 AI-PCC(AI-based Point Cloud Compression) 표준 개발을 공식화하였다.

이후 그림 2와 같이, 메쉬 추출형 정적 콘텐츠, 희소 · 밀집 포인트 클라우드, 동적 볼류메트릭 비디오, 라이다(LiDAR) 기반 획득 데이터[14] 등 실제 산업 현장에서 활용되는 다양한 유형의 콘텐츠를 대상으로 기술 탐색 실험이 진행되어 왔다.

이러한 탐색 실험 결과를 바탕으로, 2024년 7월 제148차 MPEG 회의에서는 CfP를 통해 제출된 기술들을 종합적으로 평가하였으며, 이를 토대로 AI 기반 포인트 클라우드 압축을 위한 초기 실험 모델인 TMAP(Test Model for AI-based Point Cloud Compression) v0가 확정되었다. 이는 AI 기반 그래픽 부호화 기술이 개별 기술 검증 단계를 넘어, 통합된 표준 프레임워크로 발전하기 시작했음을 의미한다.

그림 3[15]에 제시된 AI-PCC 실험 모델의 구조를 살펴보면, 기하 정보 부호화 과정에서 희소 컨볼루션(Sparse Convolution)을 단계적으로 적용하여 고밀도 포인트 클라우드를 저차원의 잠재 표현(Latent Representation)으로 변환한 뒤 이를 무손실 방식으로 압축한다. 디코더에서는 학습된 AI 모델을 활용해 원본 해상도의 포인트 클라우드를 재구성함으로써, 기하 정보의 정밀도와 압축 효율을 동시에 확보한다. 속성(텍스처) 정보의 경우, 초기에는 검증된 신호처리 기반 RAHT 방식을 사용하였으나, 최근에는 AI 기반 초해상화 기법과 V-PCC 속성 부호화를 결합한 VSS-PCC(Video Super Sampling based Point Cloud Compression) 방식이 도입되어, 속성 품질 개선과 전송 대역폭 절감 가능성이 함께 검토되고 있다.

성능 측면에서 초기 TMAP v0 모델은 기존 V-PCC 대비 기하 정보 부호화에서 43.7%, 텍스처

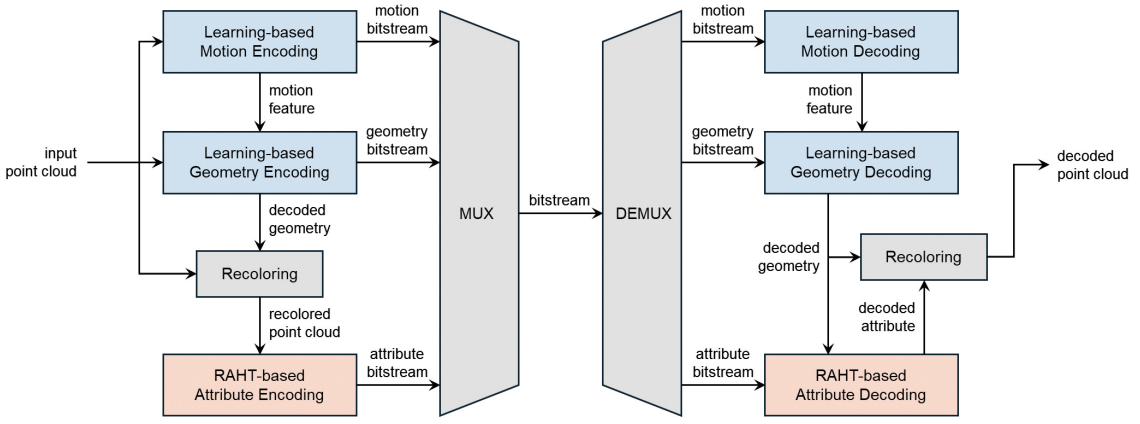

그림 3 AI-GC 부호화 및 복호화 과정

표 1 AI-GC의 주요 기술탐색 주제

구분	기술탐색(EE) 주제
EE5.2	Learning-based attribute coding
EE5.3	Cross-platform reproducibility
EE5.5	Performance analysis
EE5.6	High Level Syntax
EE5.7	Inter prediction
EE5.8	Study of AIS(Advanced Inverse Scaling)

정보에서 24.6%의 비트율 절감 효과를 달성하며 AI 기반 부호화의 잠재력을 입증하였다. 이후 TMAP v4 모델[16]에서는 압축 성능의 추가 개선보다는 실제 산업 적용을 고려해 구조를 단순화하고 연산 복잡도 저감에 중점을 두어, 초기 모델 대비 약 35%의 계산 복잡도 절감 성과를 보고하였다. 이는 AI 기반 압축 기술이 연구 단계를 넘어 실시간 처리와 대규모 서비스로 확장될 수 있음을 시사한다.

현재 AI-PCC 표준은 2025년 10월 WD(Working Draft)가 공개된 상태이며, 표 1과 같이 속성 부호화, 상호 예측, 고급 보간 샘플링(AIS) 등 세부 기술 요소를 대상으로 한 탐색 실험이 병행하여 진행되고 있다.

향후 2027년 10월 표준화 완료를 목표로 표준화 작업이 진행될 예정으로, AI 기반 입체 비디오 부호화 기술은 향후 실감형 콘텐츠 산업과 데이터 중심 산업 전반의 핵심 인프라 기술로 자리매김할 것으로 기대된다.

III. AI 중심 미디어 부호화 표준화 동향

1. 기계를 위한 비디오 부호화: VCM과 FCM 표준화 현황

AI 기술의 확산과 함께, 영상 데이터의 주요 소비 주체가 인간에서 기계와 알고리즘으로 빠르게 이동하고 있다. 특히 감시, 교통, 산업 자동화와 같은 응용 분야에서는 영상 자체의 시각적 품질보다 기계의 임무 수행 정확도와 처리 효율이 핵심 요구사항으로 주목받고 있다[17]. 이러한 변화는 기존의 인간 중심 비디오 부호화 기술만으로는 효율적인 데이터 전달이 어렵다는 한계를 드러내고 있다.

기계 기반 영상 분석 환경에서는 모든 화소 정보가 동일한 중요도를 갖지 않는다. 인간은 프레임 전체를 시각적으로 해석하며 장면의 맥락을 이해하지만, 기계는 객체의 형태, 위치, 크기, 이동 특성과 같이 임무 수행에 직접적으로 이바지하는 정보에만 선택적으로 의존한다. 그러므로, 기계가 실제로 활용하지 않는 정보까지 동일한 비트량으로 부호화하는 기존 방식은 불필요한 데이터 증가를 초래한다.

이러한 문제 인식을 바탕으로 MPEG WG4에서는 기계를 주요 소비 대상으로 하는 비디오 부호화 기술을 새로운 표준화 방향으로 설정하고, 두 가지 다른 접근 방식인 VCM(Video Coding for Machines)과 FCM(Feature Coding for Machines)을 중심으로 표준화를 추진하고 있다. 두 방식은 모두 기계 중심 응용을 목표로 하지만, 시스템 분할 방식과 데이터 전달 구조에서 차이를 가진다.

VCM은 입력 비디오를 기계 분석에 적합한 형태로 변환한 뒤, 기존 비디오 부호화 기술을 활용해 압축 · 전송하는 구조를 따른다. 이 방식에서는 추론 연산이 엣지 장치 외부의 서버나 클라우드 환경에서 수행되므로, 센서 단말의 연산 부담을 최소화할 수 있다는 장점이 있다. 그림 4는 VCM 참조 소프트

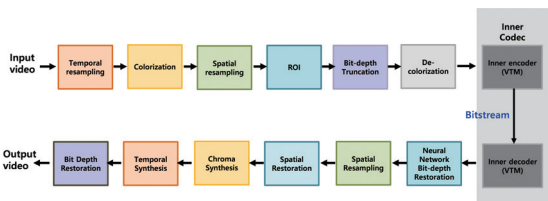

그림 4 VCM 참조 소프트웨어(VCMRS) 구조

웨어 구조도이다[18]. VCM 참조 소프트웨어는 이러한 개념을 구현하기 위해 내부 코덱으로 기존 비디오 부호화 표준을 활용하고, 그 전후 단계에서 기계에 불필요한 데이터 양을 줄이기 위한 다양한 처리 과정을 포함한다[18,19].

VCM에서 적용되는 주요 처리 방식은 기계 임무 수행 관점에서의 효율성에 초점을 둔다. 먼저 객체 인식이나 추적 성능에 상대적으로 중요하지 않은 영역을 축소하거나 제거하는 관심 영역 기반 처리를 통해 공간적 데이터 양을 줄인다. 또한 입력 영상의 시간적 또는 공간적 해상도가 기계 분석 요구 수준을 초과하는 경우, 이를 낮춘 상태로 부호화한 뒤 복호화 단계에서 복원하는 방식을 사용한다. 더 나아가, 기계 분석에 필요하지 않은 과도한 비트 심도를 조정함으로써 추가적인 데이터 절감을 달성할 수 있다.

VCM 표준화 주요 이력은 다음과 같다. 2022년 10월 제140차 MPEG 회의에서 CfP 응답기술 평가를 진행하였고, 2023년 10월 제144차 MPEG 회의에서 WD가 발행되었으며, 2025년 4월 CD(Committee Draft) 및 2026년 1월 DIS(Draft International Standard) 단계를 거쳐, 2026년 10월 FDIS(Final DIS) 단계로 표준화가 완료될 예정이다[20].

반면, FCM은 영상 전체를 부호화 대상으로 삼지 않고, 영상 분석을 위한 신경망 처리 과정에서 생성되는 중간 피처 데이터를 직접 전송하는 방식을 채택한다. 이 구조에서는 센서 단말에서 신경망의 일부 계층이 실행되며, 이후 생성된 피처를 압축하여 서버로 전달한 뒤 나머지 추론 과정을 수행한다. 이러한 분할 추론 구조는 VCM 대비 더 높은 압축 효율을 제공할 수 있으며, 피처 표현의 특성상 원본 영상의 복원이 어렵기 때문에 개인정보 보호 측면에서도 이점을 가진다.

그림 5는 FCM 참조 소프트웨어(FCTM) 구조도

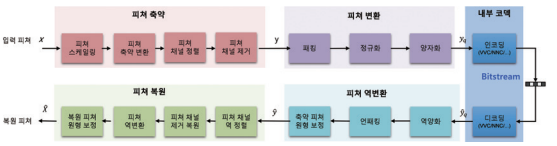

그림 5 FCM 참조 소프트웨어(FCTM) 구조

이다[21]. FCTM은 서로 다른 해상도와 특성을 갖는 다계층 피처를 입력으로 받아, 계층 간 중복성이 높은 구조를 단일 계층 표현으로 축약한 뒤 내부 코덱을 통해 부호화한다[21]. 복호화 단계에서는 해당 정보를 다시 다계층 피처 형태로 복원하여 후속 추론에 활용한다. 이 과정에서 피처 축약과 복원 단계는 전체 압축 성능과 복잡도에 가장 큰 영향을 미치는 핵심 요소로 작용한다.

FCM 표준화 주요 이력은 다음과 같다. 2023년 10월 제144차 MPEG 회의에서 CfP 응답기술 평가를 진행하였고, 2026년 1월에는 PWD(Preliminary WD)가 발행되었다[22]. 이후 MPEG-AI Part 4로서, 2026년 4월 CD 단계와 7월 DIS 단계를 거쳐, 2027년 1월에 최종 표준안인 FDIS 단계로 표준화가 완료될 예정이다.

이처럼 VCM과 FCM은 각각 원격 추론(Remote Inference)과 분할 추론(Split Inference)이라는 상이한 시스템 설계를 기반으로 기계 중심 영상 분석 환경의 요구사항을 충족하고자 한다. 두 기술은 기존 인간 중심 비디오 부호화 방식과는 다른 관점에서 데이터 압축 효율성을 극대화함으로써, 향후 다양한 AI 기반 산업 응용에서 핵심적인 역할을 수행할 것으로 기대된다.

2. 기계를 위한 오디오 부호화: ACoM 표준화 동향

기계를 위한 음향 데이터는 오디오의 공간적 분

포 정보를 포함한 다채널 형태로 대규모로 생성되는 경우가 많다. 이에 따라 사람 개입을 전제로 한 기존 오디오 처리 방식은 지연 시간과 처리 효율 측면에서 한계를 보이고 있으며, 지능형 플랫폼 지원을 위한 고압축 · 저지연의 기계 지향 오디오 압축 기술에 대한 요구가 커지고 있다.

이러한 배경에서 2022년 10월 제140차 MPEG 회의에서 FhG IDMT는 기계를 위한 오디오 부호화(ACoM: Audio Coding for Machines)를 새로운 표준화 아이템으로 제안하여 MPEG WG6(MPEG Audio Coding)은 제150차 회의에서 CfP를 발간하고, 현재 ACoM 표준화를 본격적으로 추진하고 있다[23-25].

ACoM은 오디오 및 다차원 스트림, 또는 이들로부터 추출된 특징을 효율적으로 압축하기 위한 비트스트림과 데이터 포맷을 정의하며, 비트율 및 데이터 크기 효율성을 목표로 한다. 또한, 복호화 이후에도 기계 작업 성능 저하를 최소화하고, 기계 단독 및 기계-인간 혼합 소비 시나리오를 모두 지원하도록 설계된다. 아울러 신호의 획득 방식과 생성 조건을 기술하는 메타데이터를 포함해 데이터의 활용성과 재사용성을 높이는 것을 표준화 범위로 하고, 이를 위해 MPEG WG 6에서는 산업체, 의료, 미디어 서비스 등 다양한 응용 분야에 대한 ACoM 사용 사례를 수집하였다[26].

ACoM 표준화는 Phase 1과 Phase 2로 구분하여 추진되고 있다. Phase 1에서는 응용 분야에 종속되지 않도록 데이터를 거의 무손실로 부호화하여, 표준 기반 비트스트림을 활용한 범용 데이터 교환 환경을 구축하는 것을 목표로 한다. Phase 2에서는 특징 추출 및 응용 분야별 최적화를 도입해 표준을 확장함으로써, 초기에는 범용성을 확보하고 이후 단계적으로 응용 특화 기능을 강화하는 구조를 갖는다.

현재 표준화가 진행 중인 ACoM Phase 1 시스템의 입력은 오디오 신호, 메타데이터, 라이선스 정보로 구성된다. 제149차 및 제150차 MPEG 회의에서 발간 · 개정된 ACoM CfP 문서에서는 Phase 1 응답 기술에 대해 오디오, 메타데이터, 라이선스를 포함하는 비트스트림 구조를 정의하고, 무손실 오디오 복원과 메타데이터의 정확한 보존을 필수 요구사항으로 규정하였다[25,27]. 또한, 액세스 유닛 길이, 계산 복잡도, 메모리 사용량과 같은 추가 요구사항이 제시되었으나, 이는 CfP 단계에서는 평가 대상에서 제외되며 향후 CE 절차에서 고려될 예정이다.

ACoM CfP 응답 기술의 성능 평가는 기준 코덱(Anchor) 대비 상대적 성능 향상을 기준으로 수행되었으며, 오디오 무손실 압축의 기준 앵커로는 MPEG-4 SLS가 사용되었다. 최종 앵커 비트스트림은 오디오 에센스, 메타데이터, 라이선스 파일을 하나의 파일로 통합한 후 info-zip 3.0을 이용해 무손실 압축한 형태로 구성되었다.

제153차 MPEG 회의에서는 CfP에 응답한 ETRI, Dolby의 기술에 대한 성능 평가 결과와 향후 작업 계획이 논의되었다[28,29]. 두 기술 모두 필수 요구사항을 충족하고 기준 앵커 대비 성능 향상을 보였으며, 이 중 ETRI는 통계적으로 유의미한 수준에서 더 우수한 성능을 제시하였다(그림 6). 무손실 오디오 압축과 관련하여 두 기술 모두 ISO/IEC 23003-8 및 ITU-T T.261(BWC, Biomedical Waveform Coding)을 기반으로 하며, ETRI 기술의 경우 preLPC가 추가로

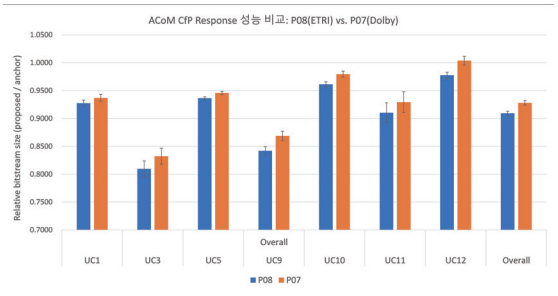

그림 6 ACoM CfP 제안기술 성능 비교

도입되었다. 이에 따라 ETRI의 제안 기술을 BWC와 병합하는 방안이 검토되었고, MPEG WG6와 ITU-T Q6/21 간 공동회의를 통해 해당 병합을 CE 제안으로 통합할 가능성을 검토하기로 합의하였다. 한편, 메타데이터 압축에 대해서는 ETRI에서 제안된 방법이 채택되었다.

ACoM은 2026년 7월 WD 단계를 시작으로 2027년 1월 CD, 4월 DIS 단계를 거쳐, 2027년 10월에 최종 표준안인 FDIS 단계로 표준화가 완료될 예정이다.

IV. 결론

본고에서는 AI 기반 미디어 부호화 기술을 중심으로 MPEG 표준화 동향을 살펴보았다. 먼저 신경망 기반 평면 비디오 부호화와 입체 비디오 · 포인트 클라우드 압축 분야에서는 기존 신호처리 기반 접근 방식의 한계를 극복하기 위한 다양한 탐색 실험이 진행되고 있으며, NNVC와 AI-PCC 등 초기 실험 모델을 통해 의미 있는 압축 성능 향상 가능성이 확인되고 있다. 다만, 이러한 기술들이 실제 표준으로 채택되기 위해서는 연산 복잡도와 구현 효율성에 대한 지속적인 개선이 필수적인 과제로 남아 있다.

한편, 기계를 위한 비디오 및 오디오 부호화 분야에서는 VCM, FCM, ACoM과 같이 인간 중심 미디어 소비를 전제로 하지 않는 새로운 부호화 패러다임이 본격적으로 논의되고 있다. 이들 기술은 원격 추론, 분할 추론, 기계 중심 데이터 소비 특성을 고려한 압축 구조를 통해 기존 방식 대비 높은 압축 효율과 새로운 활용 가능성을 제시하고 있으며, 감시 · 보안, 자율주행, 산업 자동화, 지능형 플랫폼 등 다양한 응용 분야에서의 활용이 기대된다.

종합적으로 볼 때, AI 기반 미디어 부호화 기술은 단순한 성능 개선을 넘어 미디어 처리 구조 자체의 변화를 이끄는 핵심 요소로 자리 잡고 있다. 향후 표준화 과정에서는 압축 성능과 구현 복잡도 간의 균형, 실시간 처리 가능성, 다양한 응용 환경에 대한 확장성이 주요 논점이 될 것으로 예상된다. 이러한 흐름 속에서 국내 연구기관과 산업계의 지속적인 표준화 참여와 핵심 기술 확보는 향후 글로벌 미디어 기술 경쟁력 강화 측면에서 중요한 의미를 가질 것으로 판단된다.

용어해설

MPEG ISO/IEC 산하 국제 표준화 기구로, 디지털 비디오·오디오 및 멀티미디어 데이터의 압축·전송·표현 표준을 개발

인루프 필터(in-loop filter) 복호화 과정 중 예측 루프 내부에 적용되는 필터로, 압축 과정에서 발생하는 왜곡을 줄이고 참조 영상의 품질을 개선하기 위한 기술

초해상화 저해상도 영상으로부터 고해상도 영상을 복원하는 기술

포인트클라우드 3차원 공간상의 객체나 환경을 수많은 점(point)의 집합으로 표현한 데이터 형식

참고문헌

[1] Versatile Video Coding, Standard ISO/IEC 23090-3, ISO/IEC JTC 1, July 2020.

[2] 최진수 외, "인공지능 기반 비디오 부호화 기술 동향," 주간기술동향 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 13 and software version 15 in neural network-based video coding (NNVC)," 40th Joint Video Experts Team (JVET) meeting, JVET-AN2019, Oct. 2025.

[7] 최기호, "JVET 신경망 기반 비디오 코딩 기술 연구 동향," 방송과 미디어, 제28권 제1호, 2023. 1, pp. 29-37.

[8] NNVC 소프트웨어. 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 (AHG11)," 41st Joint Video Experts Team (JVET) meeting, JVET-AO0011, Jan. 2026.

[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] CTC on AI-based point cloud coding, N1326, ISO/IEC JTC1/SC 29/WG 7, Dec. 2024.

[15] Working draft of AI-based point cloud coding, N1327, ISO/IEC JTC 1/SC 29/WG 7, Nov. 2025.

[16] TMAP v4 for AI-based point cloud coding, N1328, ISO/IEC JTC 1/SC 29/WG 7, Nov. 2025.

[17] F. Racape, H. Choi, "Video Coding for Machines: The Need for Compression," Sep. 2024. https://www.interdigital.com/post/video-coding-for-machines-the-need-for-compression

[18] Algorithm description of tools in VCM reference software, N731, ISO/IEC JTC 1/SC 29/WG 4, Dec. 2025.

[19] Preliminary text of ISO/IEC DIS 23888-2 Video coding for machines, N729, ISO/IEC JTC 1/SC 29/WG 4, Dec. 2025.

[20] Report of 21st Meeting, N721, ISO/IEC JTC 1/SC 29/WG 4, Dec. 2025.

[21] Algorithm description of FCTM, N737, ISO/IEC JTC 1/SC 29/WG 4, Dec. 2025.

[22] PWD of feature coding for machines, N739, ISO/IEC JTC 1/SC 29/WG 4, Nov. 2025.

[23] Workplan on Audio Coding for Machines, N269, ISO/IEC JTC1 SC 29/WG 6, July 2024.

[24] Workplan on Audio Coding for Machines, N309, ISO/IEC JTC1 SC 29/WG 6, Jan. 2025.

[25] Call for proposals on Audio Coding for Machines, N337, ISO/IEC JTC1 SC 29/WG 6, Apr. 2025.

[26] Use cases and requirements on audio coding for machines, N343, ISO/IEC JTC1 SC 29/WG 6, Apr. 2025.

[27] Updated call for proposals on Audio Coding for Machines, N364, ISO/IEC JTC1 SC 29/WG 6, July 2025.

[28] Report on ACoM call for proposals, N399, ISO/IEC JTC1 SC 29/WG 6, Jan. 2026.

[29] Workplan on audio coding for machines, N400, ISO/IEC JTC1 SC 29/WG 6, Jan. 2026.