

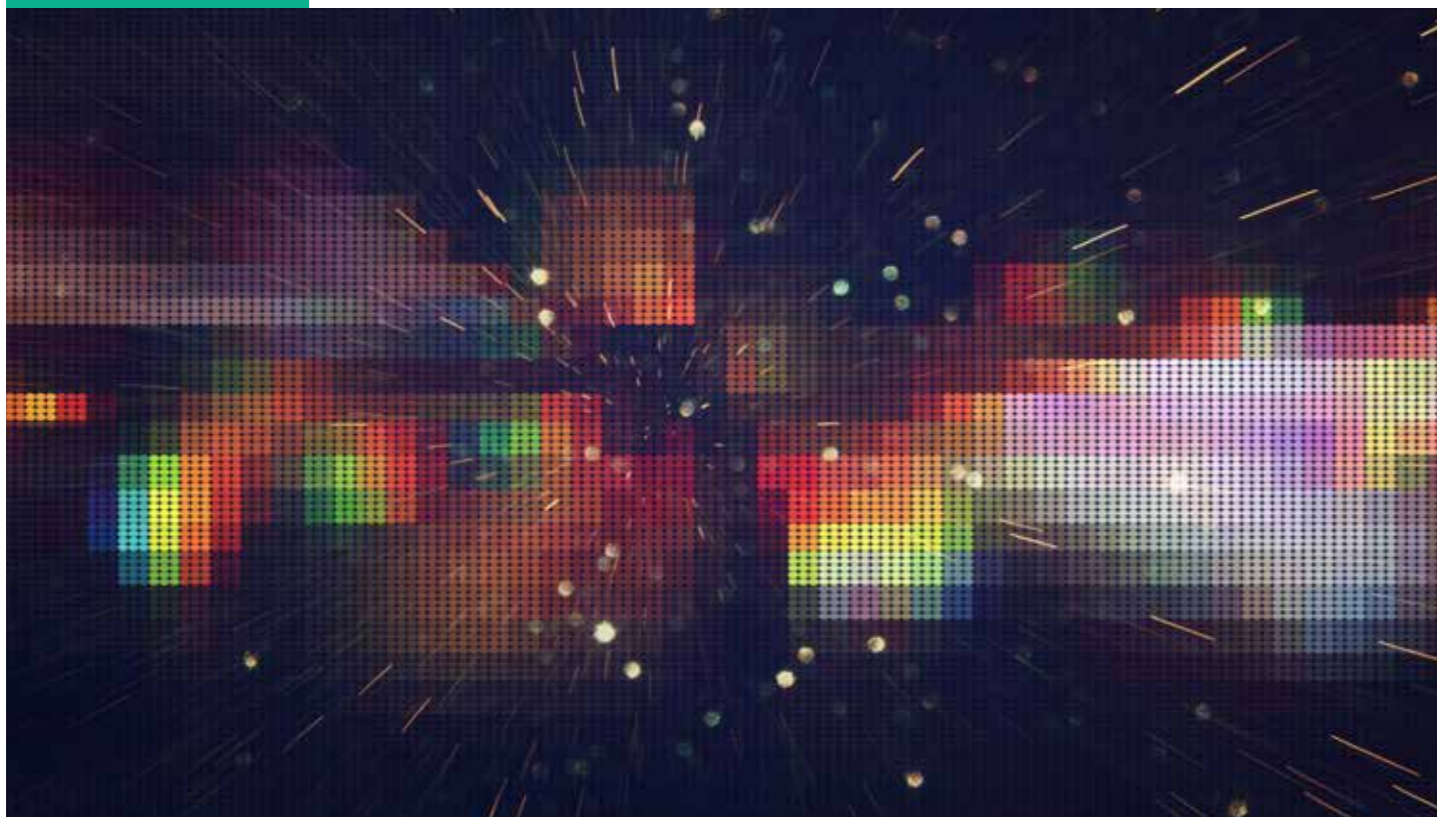
# MPEG VCM 표준화 동향

한국전자통신연구원

정순흥 책임연구원

## 1 서론

MPEG(Moving Picture Experts Group)에서는 2019년 7월, 5개 기관에서 공동으로 제안한 기계를 위한 영상 부호화 표준(VCM: Video Coding for Machines)에 대한 필요성을 받아들여 관련 표준화를 진행 중이다. 당시, ISO/IEC JTC 1/SC 29 산하 WG 11을 MPEG이라 칭하였으나, 2020년 7월 구조조정을 통해 WG 11이 폐지되고 7개의 새로운 WG(작업반)과 3개의 AG(자문반)으로 재구성되었다. 본 고에서는 MPEG에서 진행 중인 VCM 표준화 동향에 대해 소개한다.



ISO/IEC JTC 1/SC 29 산하 MPEG 조직 구성도	
WG 2	MPEG Technical Requirements
WG 3	MPEG Systems
WG 4	MPEG Video Coding
WG 5	Joint Video Coding Team with ITU
WG 6	MPEG Audio Coding
WG 7	MPEG Coding for 3D Graphics and Haptics
WG 8	MPEG Genomic Coding
AG 2	MPEG Technical Coordination
AG 3	MPEG Liaison and Communication
AG 5	MPEG Visual Quality Assessment

## 2 MPEG VCM 표준화 현황

### 2.1 배경

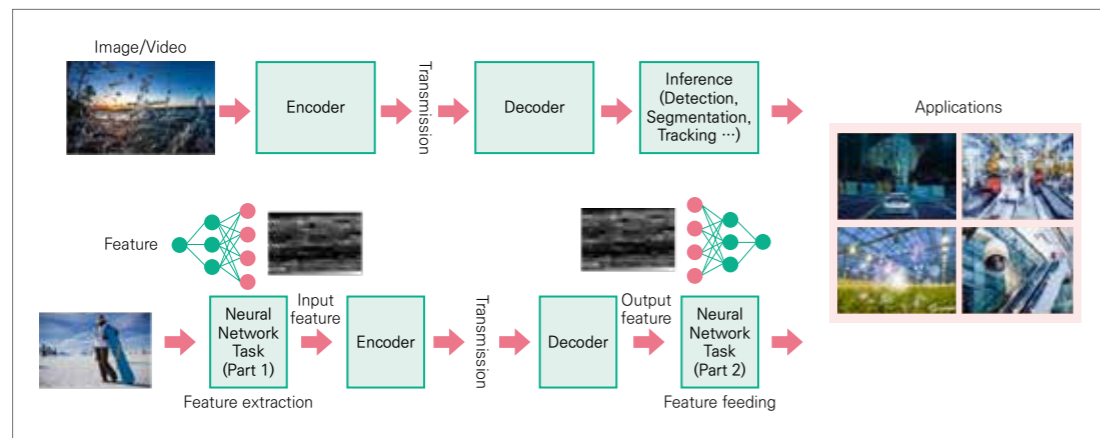
인공지능 기술과 IoT(Internet of Things) 기기의 발전 및 M2M(Machine-to-Machine) 시장의 성장으로 기계가 새로운 영상 소비 주체로 부상하고 있다. 특히, 머신 비전에 인공지능 기술이 적용되면서 영상을 통해 사람이 분석할 수 있는 능력을 뛰어넘는 성능을 제공하는 분야가 늘어나고 있다. 이는 머신 비전 기술을 탑재한 기계들의 활용 분야 확장으로 이어지고 있다. CISCO는 2018년부터 2023년까지 인터넷 장치 및 연결 수에 관해서 M2M 연결의 성장률이 19%로 가장 높아, 2023년도에 전체 연결 수의 50%를 차지할 것으로 예상하였다. 이러한 추세에 따라, M2M 연결에서 발생하는 영상 트래픽이 폭증하고 있다.

한편, 영상은 인터넷 트래픽의 80% 이상을 차지하는 것으로, 압축 부호화를 통해 이를 줄일 수 있는 기술 개발이 지속적으로 진행 중이다. 하지만, 전통적인 영상 압축 기술은 인간을 대상으로 개선된 화질의 서비스를 제공하는 방향으로 발전하였고, 기계를 대상으로 영상 속에 담긴 정보를 정확하게 이해하고 추출할 수 있는 성능에 대한 고려가 부족하였다. 또한 커넥티드카, 영상 감시, 스마트 시티 등 대규모 데이터의 실시간 처리 필요성을 고려하면, 사람이 개입하는 기존의 파이프라인으로 처리하기에는 비효율성과 비현실성이 존재한다. 2019년 7월 MPEG에서는 인공지능 기술의 발전으로 기계 간 폭증하는 트래픽을 효율적으로 압축 처리할 수 있는 표준에 대한 필요성이 제기되었다. MPEG에서는 이러한 필요성을 받아들여 2019년 10월 AHG(AdHoc Group)을 수립하여 서비스 시나리오, 요구사항, 공통 실험 조건, 성능 평가 방법, 코덱 구조 등에 대한 논의를 시작하였다.

2.2. 표준화 범위 및 요구사항

ISO/IEC JTC 1/SC 29/WG 2 MPEG Technical Requirements 그룹에서 VCM의 요구사항, 공통 실험 조건, 성능 평가 방법에 대한 논의가 표준화 초기에 활발하게 진행되었다. 당시, 전통적인 영상의 소비 대상이었던 사람의 요구사항은 잘 정립이 되어있으나, 기계가 영상을 소비하는 다양한 형태에 따른 요구사항은 그에 비해 미흡한 상황이었다.

그림 1 Video Coding for Machines 개념도



[그림 1]은 VCM 개념도를 나타내며, 부호화기 입력으로 영상(Image/Video)뿐만 아니라 영상에서 얻어지는 특징(Feature)을 포함한다. 영상을 입력 받는 경우에는 압축된 비트스트림을 생성하고, 복호화기를 통해 이를 복원한 후, 복원된 영상을 이용하여 머신비전 네트워크의 입력으로 사용하는 형태의 부호화 기술을 의미한다. 영상으로부터 머신비전 네트워크 중간 출력을 통해 얻어지는 feature를 입력받는 경우에는 압축된 비트스트림을 생성하고, 복호화기를 통해 복원 후, 복원된 feature를 이후 머신비전 네트워크의 입력으로 사용하는 형태의 부호화 기술을 의미한다. MPEG Technical Requirements 그룹에서는 2021년 10월 회의에서 두 입력에 따른 진행 상황을 고려하여 Feature Extraction and Compression Tasks(track 1)와 Image and Video Compression Tasks(track 2)의 두 개의 트랙으로 분리하여 진행하도록 결정하였고, 2022년 4월에 서비스 시나리오 및 요구사항 문서를 발행하였다. 해당 문서에서 표준화의 범위를 부호화 비트율 대비 기계 임무 수행 성능 관점에서 효율적인 영상 또는 영상으로부터 추출되는 서술자 및 특징을 부호화한 비트스트림을 정의하는 것으로 정리하였다. 그리고, 표준화 완료 후 적용 가능한 7종의 서비스 시나리오(Surveillance, Intelligent Transportation, Smart City, Intelligent Industry, Intelligent Content, Consumer Electronics, VCM multi-task with descriptors)와 이를 위한 17개의 요구사항을 제시하였다. 요구사항은 [표 1]과 같으며, 입력이 영상 및 특징에 해당되는 요구사항에 대해서 각각 Video Coding, Feature Coding으로 표시하였다.

표 1 VCM 요구사항[1]

No.	Requirements
1	VCM shall support video coding for machine task consumption purposes. Description: The specification shall enable bitstream to be decompressed to perform a machine task of various kinds.
2	VCM shall support feature coding. (Feature coding) Description: VCM shall support the input of feature map.
3	VCM shall support a coding efficiency improvement for at least 30% BD-rate over the VVC standard on machine vision tasks. (video coding)
4	VCM shall support a broad spectrum of encoding rates.
5	VCM shall support various degrees of delay configuration. Description: Ultra-low latency of less than one picture interval could require encoding and decoding operation without picture re-ordering and at units less than a picture. Streaming applications or similar could enable hierarchical coding of pictures and hence incur picture re-ordering delay in encoding and decoding. (Video coding, need further discussion on feature coding)
6	VCM shall be agnostic to network models. (Video/Feature coding)
7	VCM shall be agnostic to machine task types. (Video/Feature coding)
8	VCM shall provide description of the meaning or the recommended way of using the decoded data. (Feature coding)
9	VCM should support the use and inclusion of information such as descriptors in its bitstream. (Feature/Video coding)
10	A single VCM bitstream shall support any number of instances of machine tasks. (video coding/feature coding)
11	VCM shall support at least the following colour formats; monochrome, RGB, and YUV (YCbCr). (Video coding)
12	VCM shall support at least the following input bit depths: 8-bit and 10-bit. (Video coding) Note: Other bit depth supports are not prohibited. YCbCr color spaces with 4:0:0, 4:2:0 and 4:4:4 sampling, 8 and 10 bits per component shall be supported. RGB with 4:4:4 sampling, 8 and 10 bits per component shall be supported. BT709 and BT2100 color gamuts shall be supported.
13	VCM complexity shall allow for feasible implementation within the constraints of the available technology at the expected time of usage.
14	VCM shall support rectangular picture format up to 7680x4320 pixels(8K).
15	VCM shall support fixed and variable rational frame rates for video inputs.
16	VCM shall support any input source from video or image. Note: For example, the source content may be camera-captured or may have text or graphics overlaid onto a camera-captured scene.
17	VCM shall support privacy and security. (Mandated by ISO)

2.3. 표준화 진행 현황

2021년 10월 회의에서 입력에 따른 두 트랙으로 분리가 결정된 이후, 영상을 입력으로 받는 트랙 2는 VCM 명칭을 사용하고 있으며, 특징을 입력으로 받는 트랙 1은 FCVCM(Feature Compression for Video Coding for Machines) 명칭을 사용하다가 2023년 10월 회의에서 FCM(Feature Coding for Machines)으로 변경하였다. VCM 및 FCM은 MPEG Technical Requirements 그룹에서 2022년 10월과 2023년 10월에 각각 CfP(Call for Proposal) 대응 기술 평가를 거친 후, 비디오 부호화 표준 개발을 담당하는 WG 4: MPEG Video Coding 그룹으로 표준화 활동을 이관하여, VCM AHG 및 FCM AHG에서 표준화가 진행되고 있다.

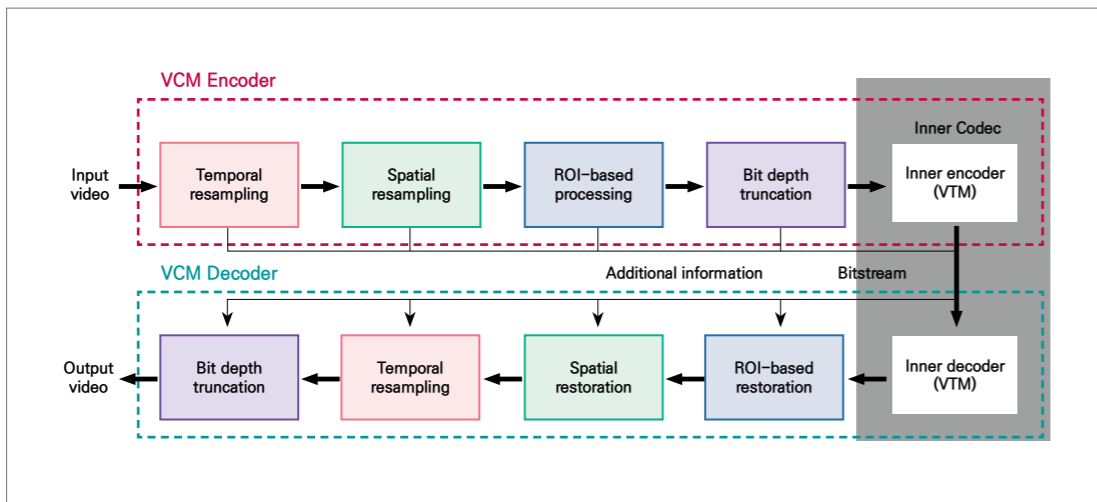
2.3.1. VCM 표준화 진행 현황

VCM은 2022년 10월 CfP 대응 기술 평가 후 2023년 10월 Working Draft 초안이 작성되었고 참조 소프트웨어와 함께 업데이트되고 있다. [그림 2]는 VCM 참조 소프트웨어의 구조도를 나타낸다. VCM 부호화기는 영상을 입력받아 시간 해상도 재표본화, 공간 해상도 재표본화, 관심영역 처리, 비트 심도 절삭



후 VTM 기반 inner encoder를 통해 부호화된 비트스트림과 부가정보를 생성할 수 있는 구조로 구현되어 있다. 시간 해상도 재표본화 및 공간 해상도 재표본화에서는 머신 임무 수행 성능을 유지하면서 줄일 수 있는 시공간 해상도를 결정하고, 이를 통해 부호화될 영상의 시공간 해상도를 줄임으로써 압축 성능을 높인다. 관심 영역 처리에서는 머신 임무 수행에 중요한 관심 영역의 정보는 유지하고, 배경 영역은 적은 비트율로 부호화할 수 있도록 하는 처리가 이루어진다. 비트 심도 절삭에서는 휘도와 색차를 나타내는 데이터의 비트 심도를 절삭함으로써, 부호화 효율 개선이 이루어진다.

그림 2 VCM 참조 소프트웨어(VCMRS: VCM Reference Software) 구조도

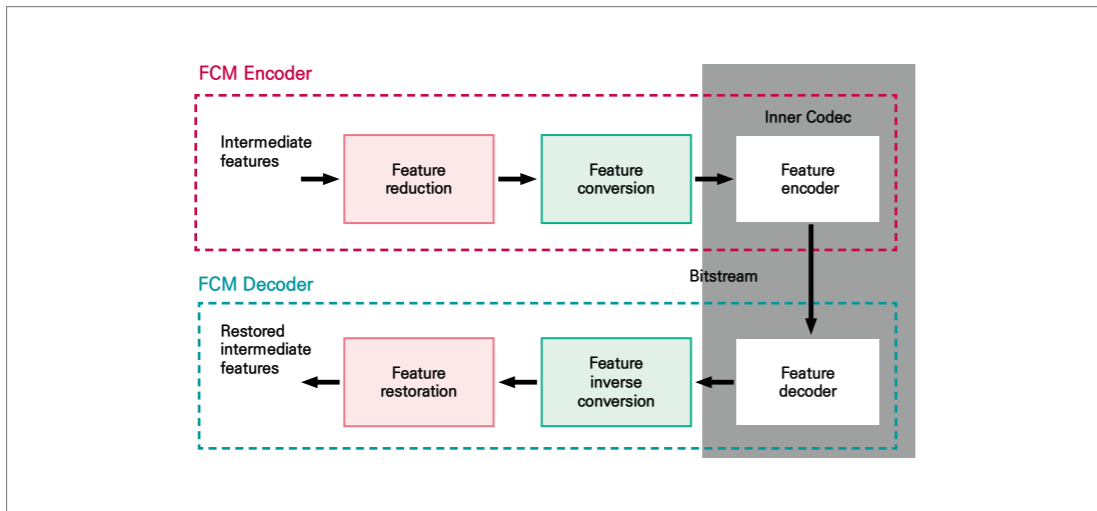


VCM 복호화기에서는 해당 비트스트림과 부가정보를 받아서 영상을 복원하는 구조를 가진다.

2.3.2. FCM 표준화 진행 현황

FCM은 2023년 10월 CfP 대응 기술 평가 후 2024년 8월 Preliminary Working Draft가 작성되었고 참조 소프트웨어와 함께 업데이트될 예정이다. [그림 3]은 FCM 참조 소프트웨어의 구조도를 나타낸다.

그림 3 FCM 참조 소프트웨어(FCM: Feature Compression Test Model) 구조도



FCM 부호화기는 머신 비전 네트워크 중간에서 출력되는 intermediate features를 입력받아 feature reduction, feature conversion 후 inner codec의 feature encoder를 통해 압축된 비트스트림을 생성하고, feature decoder, feature inverse conversion, feature restoration 과정을 통해 입력된 intermediate features를 복원하는 구조를 가진다. 공통 실험 조건에서 FCM 부호화기의 입력으로 들어가는 feature는 다계층 다채널 특징 맵으로 구성된다. Feature reduction에서는 다계층 다채널 특징 맵에서 머신 비전 임무 수행에 중요한 특징들을 선택하고, 중요도가 떨어지는 특징들은 버림으로써 feature의 양을 줄인다. Feature conversion에서는 feature 값을 VVC, HEVC 등의 기존 영상 코덱의 입력으로 사용될 수 있는 값으로 양자화하고, 이를 한 장의 영상 형태로 packing 한다. 이렇게 영상 형태로 packing된 데이터는 영상 코덱의 입력으로 들어가 압축된다.

2.4. 향후 일정

VCM은 2024년 7월 회의 결과로 Working Draft 4를 작성 중에 있으며, 2024년 11월 회의 후 Committee Draft, 2025년에 Draft International Standard가 작성될 것으로 예상된다. FCM은 2024년 8월 Preliminary Working Draft가 작성되었고, 이후 회의 결과를 기반으로 Working Draft가 작성 및 업데이트될 계획이다. VCM의 진행 일정으로 추정해보면, FCM은 2025년 10월 회의 후 Committee Draft, 2026년도에 Draft International Standard가 작성될 것으로 예상된다.

3 결론

2019년부터 MPEG에서 진행 중인 VCM, FCM 표준화는 기계 간 폭증하는 트래픽을 효율적으로 압축 처리할 수 있는 표준 개발을 목표로 하고 있다. MPEG 회의는 보통 연간 4회 개최되는데, VCM, FCM 표준화를 위해 15개 내외의 기관과 50명 내외의 전문가가 매번 참석하여 활발하게 진행되고 있다. 2024년 9월 현재 VCM 및 FCM 참조 소프트웨어의 부호화 비트율 대비 기계 임무 성능은 VVC 대비 30% 이상의 성능 개선을 달성하였다. 일정대로 표준화가 진행되면, 2026년 이후 Intelligent Transportation, Smart City, Intelligent Industry 등 적용가능한 서비스에서 VCM, FCM 표준이 활용될 수 있을 것으로 기대된다.

ACKNOWLEDGMENT

This work was supported by Institute of Information & communications. Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT)(No. 2020-0-00011, Video Coding for Machine)

참고문헌

- [1] Use cases and requirements for Video Coding for Machines, ISO/IEC JTC 1/SC 29/WG 2 N190, 2022.04
- [2] Call for Proposals on Video Coding for Machines, ISO/IEC JTC 1/SC 29/WG 2 N191, 2022.04.
- [3] Call for Proposals for Feature Compression for Video Coding for Machines, ISO/IEC JTC 1/SC 29/WG 2 N282, 2023.04.
- [4] Common test conditions for video coding for machines, ISO/IEC JTC 1/SC 29/WG 4 N543, 2024.08.
- [5] Common test and training conditions for FCM, ISO/IEC JTC 1/SC 29/WG 4 N497, 2024.06.
- [6] 정순홍, 추현근, "기계를 위한 영상 부호화 기술 동향 및 발전 전망", 한국방송-미디어공학회 하계학술대회, 2024.06.