

특집논문 (Special Paper)

방송공학회논문지 제30권 제3호, 2025년 5월 (JBE Vol.30, No.3, May 2025)

<https://doi.org/10.5909/JBE.2025.30.3.320>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

VCM 전송을 위한 비트스트림 구조 설계

박 기 범^{a)}, 이 진 영^{b)}, 정 세 윤^{b)}, 김 규 현^{a)‡}

Design of a Bitstream Structure for streaming of Video Coding for Machines

Kibeom Park^{a)}, Jinyoung Lee^{b)}, Seyoon Jeong^{b)}, and Kyuheon Kim^{a)‡}

요 약

VCM(Video Coding for Machines)은 기계가 비디오 데이터를 효율적으로 이해하고 처리할 수 있도록 최적화된 부호화 방식으로, 객체, 장면, 움직임 정보를 효율적으로 표현하고 분석에 필요한 핵심 정보를 유지하면서 비트레이트(bitrate)를 절감한다. VCM은 자율 주행, 스마트 시티, 로봇 비전, 원격 감시 등 다양한 분야에서 활용되며, 현재 MPEG(Moving Picture Experts Group) 산하의 ISO/IEC JTC 1/SC 29/WG 4에서 표준화가 진행 중이다. VCM 비트스트림(bitstream) 구조는 기존 비디오 코딩 구조를 기반으로 VCM Unit, VCM NAL Sample Stream, VCM NAL Unit 형태로 구성되어 헤더가 중복되고 파서(parser) 복잡성과 관리 부담이 크다. 본 논문에서는 Sample Unit과 NAL Unit으로 단일화된 구조를 제안하여 중복 헤더를 제거하고 비트스트림 효율성을 향상하는 구조를 제안한다.

Abstract

Video Coding for Machines (VCM) is a video coding framework optimized for machines to efficiently interpret and process video data, effectively representing objects, scenes, and motion within complex video content while preserving essential information for analysis and recognition tasks and reducing bitrate. VCM is utilized in various fields such as autonomous driving, smart cities, robotic vision, and remote surveillance, and is currently under active standardization by ISO/IEC JTC 1/SC 29/WG 4 within the Moving Picture Experts Group (MPEG). The VCM bitstream structure is based on conventional video coding frameworks and consists of a VCM Unit, VCM NAL Sample Stream, and VCM NAL Unit, which results in duplicated headers and increases parser complexity and management overhead. This paper proposes a simplified structure that consolidates the format into only Sample Unit and NAL Unit, eliminating redundant headers and improving bitstream efficiency.

Keyword : Video Coding for Machines, Bitstream Structure, Standardization, Bitstream Efficiency

a) 경희대학교(Kyung Hee University)

b) 한국전자통신연구원(ETRI)

‡ Corresponding Author : 김규현(Kyuheon Kim)

E-mail: kyuheonkim@khu.ac.kr

Tel: +82-31-201-2442

ORCID: <https://orcid.org/0000-0003-1553-936X>

※ This research 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).

· Manuscript April 10, 2025; Revised May 9, 2025; Accepted May 9, 2025.

Copyright © 2025 Korean Institute of Broadcast and Media Engineers. All rights reserved.

“This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered.”

I. 서론

영상 데이터는 오랫동안 방송, 스트리밍, 저장 매체와 같은 시청자 중심의 콘텐츠 소비 환경에서 활용되어 왔으며, 이를 지원하기 위해 AVC/H.264^[1], HEVC/H.265^[2], VVC/H.266^[3] 등 다양한 비디오 코딩 표준이 개발되어 왔다. 이들 표준은 제한된 대역폭에서도 높은 시각적 품질을 제공하는 것을 목표로 하여, 주로 사람이 시청하는 것을 전제로 설계되었다. 이에 따라, 비디오 코딩 기술의 발전은 화질 개선, 압축 효율 향상, 그리고 오류 복원력을 높이는 방향으로 지속적으로 이루어지고 있다.

그러나 최근 인공지능 기술과 프로세서, 메모리 등 컴퓨팅 하드웨어 기술의 급격한 발전으로 인해, 영상 데이터를 사람이 아닌 기계가 분석하는 시대가 도래하였다. 특히, 자율 주행 차량, 스마트 시티, 원격 감시, 산업용 로봇 등 다양한 응용 분야에서 고해상도 영상 데이터를 기반으로 한 실시간 인식, 추적, 분석 작업이 요구되면서, 기존 사람 중심의 비디오 코딩 방식으로는 효율적인 처리가 어려워졌다. 사람을 위한 시각적 품질을 유지하는 것은 기계 학습 모델이 주로 활용하는 객체의 윤곽, 움직임, 특징점 등의 시맨틱(semantic) 정보뿐만 아니라 시각적 세부 묘사까지 포함되기 때문에, 정보 전달 측면에서는 불필요하게 방대한 데이터가 포함된다. 이러한 과도한 데이터량은 저장 및 처리 과정에서 높은 연산 자원을 요구하고, 네트워크 전송 시에도 대역폭이 비효율적으로 소모되어 실시간 데이터 처리 지연과 시스템 전체의 응답 속도 저하를 초래한다^{[5][6]}.

이러한 한계는 기계 학습 기반 시스템의 요구에 맞춘 새로운 비디오 코딩 패러다임의 필요성을 촉진시켰다. 이에 따라 기존과 다른 새로운 연구 방향과 접근 방식이 필요하여 등장한 것이 VCM(Video Coding for Machines)이다. VCM은 전통적인 화질 중심의 부호화 방식에서 벗어나, 객체 검출, 추적, 장면 분석 등 기계 학습 모델이 요구하는 시맨틱 정보를 효율적으로 보존하면서 전체 비트스트림(bitstream)의 크기를 줄이는 것을 목표로 한다^{[5][7]}. 이를 통해 VCM은 데이터 전송 지연을 최소화하고, 실시간 처리 및 대규모 데이터 환경에서도 높은 확장성을 제공한다. 현재 VCM은 MPEG(Moving Picture Experts Group) 산하 ISO/IEC JTC 1/SC 29/WG 4에서 표준화가 활발히 이루어지고 있다^[7].

VCM 표준에서 정의하는 현재의 비트스트림 구조는 V3C(Versatile Video Coding-based 3D Visual Content Coding)^[4] 기반의 계층적 설계로, VCM Unit, VCM NAL Sample Stream, VCM NAL Unit으로 구성된다. 이러한 구조는 유연성과 모듈성을 제공하지만, 계층마다 헤더가 중복으로 존재하여 비트스트림 오버헤드(overhead)가 증가하고, 파서(parser) 설계 복잡성과 관리 부담을 초래한다^[5].

본 논문에서는 이러한 문제를 해결하기 위한 하나의 방안으로 Sample Unit과 NAL Unit으로 단일화된 새로운 비트스트림 구조를 제안한다. 제안하는 구조는 중복된 헤더를 제거하여 비트스트림의 효율성을 높이고, 파싱 과정을 단순화함으로써 기계 중심의 비디오 전송 및 처리 성능을 향상시킬 수 있다. 이를 통해 VCM의 실용성을 높이고, 기계 학습 기반 시스템에서의 활용 가능성을 확대하는 데 기여하고자 한다.

II. VCM 표준 및 비트스트림 구조

다양한 기계 기반 응용 분야에서 영상 데이터의 수요가 급격히 증가하면서, 효율적인 비디오 코딩 기술이 요구되고 있다. 특히, 기계 학습 기반 시스템에서는 시각적 품질보다는 객체 인식, 추적, 분석 등 시맨틱 정보를 효과적으로 전달하는 것이 중요하다. 이러한 요구사항들을 충족하기 위해 등장한 것이 VCM(Video Coding for Machines) 기술이며, 국제 표준화 기구인 MPEG을 중심으로 관련 표준화가 진행되고 있다. 본 장에서는 2025년 1월에 열린 149차 MPEG 회의에서 논의된 내용을 기반으로 VCM 기술의 기본 개념을 설명하고, VCM 인코더(encoder) 및 디코더(decoder)의 동작 원리를 포함하는 VCM process를 중심으로 시간적 리샘플링(Temporal Resampling), 공간적 리샘플링(Spatial Resampling), 관심 영역 기반 프로세싱(RoI-based processing), 비트 심도 절감 처리(Bit Truncation), 내부 코덱(Inner Codec) 등 주요 처리 단계를 설명한다. 마지막으로 현재 VCM 표준에서 정의하는 비트스트림 구조를 설명하고자 한다^[10].

1. VCM Process

VCM process는 기계 학습 기반 영상 처리에서 필요로

하는 데이터 효율성과 처리 최적화를 달성하기 위해 설계된 일련의 처리 단계로 구성된다. 각 단계는 원본 비디오에서 불필요한 중복 데이터를 제거하고, 기계 학습 모델이 효과적으로 활용할 수 있는 정보를 유지하는 것을 목표로 한다. 주요 프로세스로는 시간적 리샘플링(Temporal Resampling), 공간적 리샘플링(Spatial Resampling), 관심 영역 기반 프로세싱(RoI-based processing), 비트 심도 절감 처리(Bit Truncation), 내부 코덱(Inner Codec)이 있다.

- 시간적 리샘플링(Temporal resampling): 입력 비디오의 시간 해상도를 조정하여 프레임 간 중복 정보를 줄이고, 전송 및 저장 효율성을 향상시키는 과정이다. 일반적으로 불필요한 프레임을 제거하거나 일정 간격으로 프레임을 선택하는 방식을 사용하여 기계 학습 모델이 요구하는 시간을 줄이는 방법이며, 복원시에는 해당 프레임을 보간(interpolation)한다.
- 공간적 리샘플링(Spatial resampling): 영상의 공간 해상도를 조절하여 기계 학습에 필요한 주요 특징을 유지하면서 비트스트림 크기를 줄이는 과정이다. Downsampling을 통해 불필요한 고해상도 데이터를 줄이고, 분석에 필요한 영상 정보를 효율적으로 제공한다.
- 관심 영역 기반 프로세싱(RoI(Region of Interest)-based processing): 영상 내 주요 분석 영역에 비트레이트를 집중하여 데이터 효율성을 극대화하는 방식이다. 주요 객체가 포함된 관심 영역에는 높은 비트레이트를 할당하고, 배경 영역은 낮은 비트레이트로 처리하여 기계 학습 모델의 인식 정확도를 향상시킨다.
- 비트 심도 절감 처리(Bit truncation): 복원 품질 손실을 최소화하면서 비트 깊이를 줄여 데이터량을 최적화하는 과정이다. 기계 학습 모델이 필요로 하는 핵심 정보를 유지하면서 불필요한 데이터를 줄여 전송 및 저장 효율성을 높인다.
- 내부 코덱(Inner codec): VCM 비트스트림 내 복원 데이터를 인코딩하는 기존 비디오 코덱을 의미하며, HEVC, VVC, AVC 등이 사용된다. Inner codec은 feature-level 데이터와 복원 데이터를 효과적으로 전달하여 VCM process 전반의 압축 효율성을 높인다.

이러한 VCM process는 정해진 순서에 따라 수행되며, 각 프로세스는 시나리오에 따라 선택적으로 활성화하거나 비활성화할 수 있다. 즉 시간적 리샘플링, 공간적 리샘플링, 관심 영역 기반 프로세싱, 비트 심도 절감 처리는 필요에 따라 개별적으로 적용 여부를 설정할 수 있으며, 마지막 단계인 내부 코덱은 항상 수행되어야 한다.

2. VCM 비트스트림 구조

VCM 비트스트림 구조는 MPEG V3C(Versatile Video Coding-based 3D Visual Content Coding) 비트스트림 구조와 유사하게 설계되어 있으며, 계층적 구조를 통해 유연성과 확장성을 제공한다. VCM 비트스트림은 그림 1과 같이 크게 VCM Unit, Sample Stream, 그리고 VCM NAL Unit의 세 가지 주요 계층으로 구성된다.

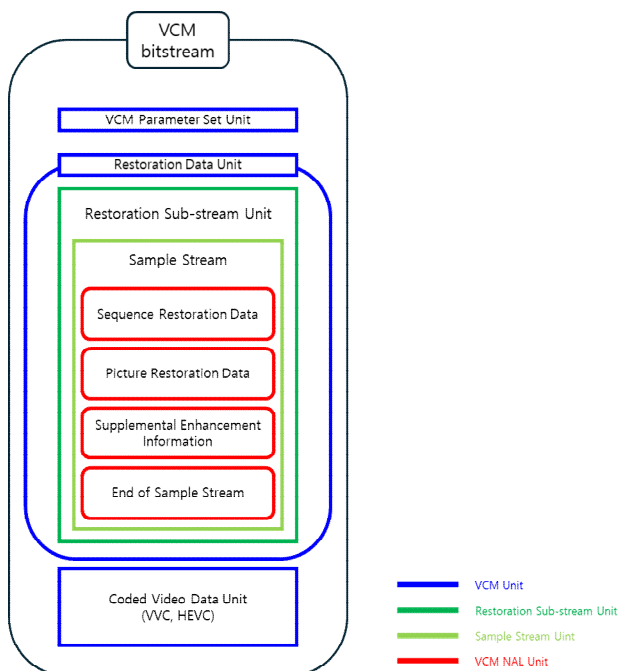


그림 1. VCM 비트스트림 구조
Fig. 1. VCM bitstream structure

VCM Unit: 전체 비트스트림의 최상위 계층으로, 비디오 부호화에 필요한 초기화 정보, 복원 데이터, 그리고 인코딩된 비디오 데이터를 포함한다. VCM Unit 내에는 VCM pa-

parameter set unit, restoration data units, coded video data units로 세분화되어 있다.

- **VCM Parameter Set Unit:** 비디오 코덱의 유형과 프로파일 등 복원 및 디코딩 과정에서 필요한 주요 초기화 파라미터 정보를 포함한다.
- **Restoration Data Units:** 복원 데이터와 관련된 정보를 포함하며, 해당 Unit의 하위 계층에는 실제 복원 데이터를 보다 효율적으로 처리하기 위한 Restoration Sub-Stream Unit이 존재한다.
- **Restoration Sub-Stream Unit:** 복원 데이터의 메타 정보와 함께, 복원 과정에서 필요한 각 Sample Stream NAL Unit으로 분할된 데이터를 전달한다. 해당 구조는 복원 데이터를 계층적으로 관리하여 디코딩 효율성을 높이고, 복원 프로세스의 확장성과 유연성을 제공한다.
- **Coded Video Data Units:** inner codec으로 인코딩된 원본 비디오 데이터가 포함된다. 이때, VCM Parameter Set Unit에서 지정된 코덱이 사용된다.

VCM Unit은 이러한 구조를 통해 복원 데이터와 원본 비디오 데이터를 효율적으로 함께 관리하고, 다양한 코덱 옵션을 지원할 수 있도록 설계되어 있다.

Sample Stream: Restoration Sub-Stream Unit 내에서 실제 복원 데이터를 전달하는 단위로, 효율적인 파싱과 전송을 위한 NAL Unit 단위의 Stream 구조이다.

- Sample Stream은 여러 개의 Sample Stream NAL Unit으로 구성된다.
- 각 Sample Stream NAL unit은 실제 복원 데이터를 포함하며, 이 데이터를 전달하기 위해 NAL Unit의 크기 정보와 복원 데이터의 payload를 포함한다.
- 이를 통해 디코더는 비트스트림 내에서 정확히 NAL Unit을 식별하고 효율적으로 파싱할 수 있다.

Sample Stream 구조는 복원 데이터의 전송 및 파싱 효율성을 높이고, 필요한 경우 스트림 내에서 동적으로 확장하거나 축소할 수 있도록 유연성을 제공한다.

VCM NAL Unit: Sample Stream 내에서 복원 데이터를 실제로 전달하는 단위로, 각 복원 단계별로 세부적인 정보

를 포함한다. VCM NAL Unit은 다음과 같은 세부 구조를 가진다.

- **Sequence Restoration Data:** 시퀀스 레벨 복원 정보를 포함.
- **Picture Restoration Data:** 프레임 단위의 복원 정보.
- **Supplemental Enhancement Information:** 추가적인 보조 정보를 제공.
- **End of Sample Stream:** Sample Stream의 종료를 나타내는 마커.

VCM NAL Unit은 복원 과정에 필요한 정보를 계층적으로 포함하여 디코더가 효율적으로 복원 데이터를 처리할 수 있도록 한다.

III. 제안된 VCM 비트스트림 구조

현재 VCM 비트스트림 구조는 V3C 기반의 계층적 설계를 적용하여 유연성과 확장성을 제공하고 있으며, 복원 데이터와 원본 비디오 데이터를 효율적으로 전송할 수 있도록 설계되어 있다. 그러나, VCM Unit과 NAL Unit 계층 모두에 헤더가 존재하는 구조로 인해 비트스트림 내 오버헤드가 발생할 수 있으며, 이로 인해 파싱 설계의 복잡성 증가 및 데이터 전송 효율성 측면에서 개선 여지가 존재한다.

본 논문에서는 VCM 비트스트림의 효율성을 더욱 향상시키기 위한 방안으로 새로운 비트스트림 구조를 제안하고자 한다. 제안하는 구조에서는 기존 구조에서 사용되었던 VCM Unit 계층을 제거하고, VCM Sample Unit과 VCM NAL Unit으로 단순화된 구조로 재구성하였다. 이를 통해 비트스트림 구성과 파싱의 복잡성을 줄일 수 있다. 또한, VCM process별로 Sequence-Level과 Picture-Level 데이터를 명확히 구분하여 중복된 데이터 전송을 방지하고, 필요에 따라 선택적으로 전송할 수 있도록 설계하였다. 여기서 Sequence-Level 데이터는 입력 영상의 전체 시퀀스에 적용되는 공통적인 파라미터를 의미하고, Picture-Level 데이터는 각 프레임 단위로 적용되는 개별적인 파라미터를 의미한다. 제안하는 VCM 비트스트림 구조를 시각적으로 표현하면 그림 2와 같다^[12].

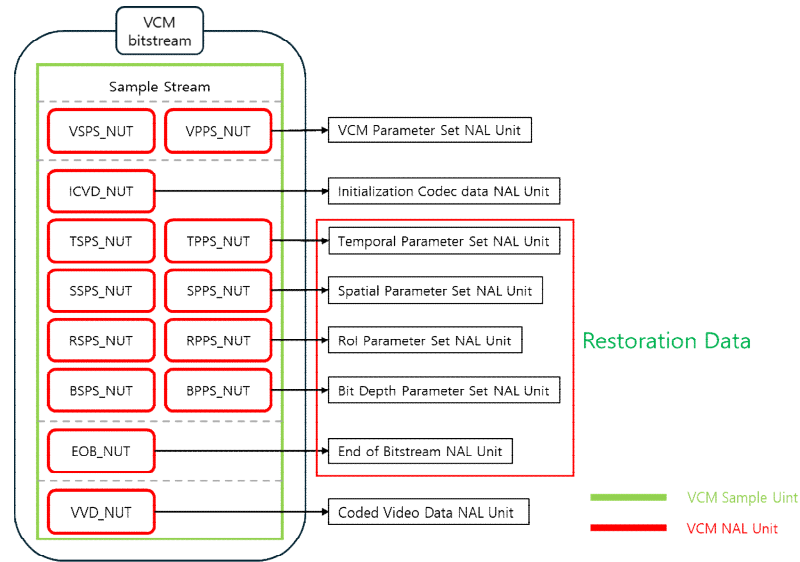


그림 2. 제안하는 VCM 비트스트림 구조

Fig. 2. Suggested VCM bitstream structure

표 1. 제안하는 VCM NAL 단위 타입

Table 1. Proposed VCM NAL Unit Type

nal_unit_type	Name of nal_unit_type	Content of NAL unit and RBSP syntax structure	NAL unit type class
0	ICVD_NUT	Initialization codec video data initialization_codec_video_data()	non-VCL
1	VVD_NUT	VCM video data vcm_video_data()	VCL
2	VSPS_NUT	Video sequence parameter set video_sequence_parameter_set_rbsp()	non-VCL
3	VPPS_NUT	Video picture parameter set video_picture_parameter_set_rbsp()	non-VCL
4	TSPS_NUT	Temporal sequence parameter set temporal_sequence_parameter_set_rbsp()	non-VCL
5	TPPS_NUT	Temporal picture parameter set temporal_picture_parameter_set_rbsp()	non-VCL
6	SSPS_NUT	Spatial sequence parameter set spatial_sequence_parameter_set_rbsp()	non-VCL
7	SPPS_NUT	Spatial picture parameter set spatial_picture_parameter_set_rbsp()	non-VCL
8	RSPS_NUT	Rol sequence parameter set roi_sequence_parameter_set_rbsp()	non-VCL
9	RPPS_NUT	Rol picture parameter set roi_picture_parameter_set_rbsp()	non-VCL
10	BSPS_NUT	Bit depth sequence parameter set bit_depth_sequence_parameter_set_rbsp()	non-VCL
11	BPPS_NUT	Bit depth picture parameter set bit_depth_picture_parameter_set_rbsp()	non-VCL
12	EOB_NUT	End of bitstream end_of_bitstream_rbsp()	non-VCL
13~63	RSV_NVCL	Reserved for future parameter set	

그림 2에서 제안하는 비트스트림 구조는 VCM 프로세스별로 생성된 NAL unit이 VCM Sample Unit 형태로 구성되어 있음을 보여준다. 제안하는 VCM 비트스트림은 VCM Sample Unit과 VCM NAL Unit의 두 가지 주요 구성 요소로 이루어진다.

- VCM Sample Unit: 각 NAL Unit의 길이 정보와 NAL Unit를 포함하여, 디코더가 비트스트림 내에서 빠르게 NAL Unit을 정확히 식별하고 파싱할 수 있도록 한다.
- VCM NAL Unit: 헤더(header)와 페이로드(payload)로 구성되며, 헤더는 NAL Unit의 타입과 식별 정보를 제공하고, 페이로드는 복원 데이터 또는 파라미터 정보를 담는다. 헤더 내에는 표 1과 같이 각 프로세스와 기능에 따라 정의된 VCM NAL Unit Type이 존재한다.

표 1에서 제시된 각 NAL Unit들은 다음과 같은 역할을 수행한다:

- ICVD_NUT(Initialization Codec Video Data): VCM 디코딩 프로세스 시작 이전에 내부 코덱(HEVC, VVC 등)의 동작을 위한 초기 설정 정보를 포함하는 NAL Unit이다. 예를 들면 코덱 프로파일, 해상도, 프레임 속도 등 VCM 레벨에서 사전에 코덱에 필요한 파라미터 값들을 전달해주는 역할을 한다.
- VVD_NUT(VCM Video Data): VCM 비트스트림 내에서 실제 압축된 영상 데이터를 전달하는 VCL NAL Unit으로, 내부 코덱을 통해 인코딩된 비디오 데이터를 포함한다. 해당 NAL Unit의 페이로드(payload)에는 내부 코덱에서 정의한 NAL Unit들을 포함하고 있다.
- EOB_NUT(End of Bitstream): VCM 비트스트림의 종료를 명시하는 NAL Unit으로, 디코더는 해당 NAL Unit을 통해 파싱 종료 여부를 판단하며, 잔여 데이터를 무시하거나 안전하게 스트림 종료 처리를 할 수 있다.
- VSPS_NUT/VPPS_NUT: 비디오 전처리에 대한 공통 설정 또는 특정 프레임 파라미터들을 포함하고 있다. VSPS_NUT의 경우 영상 전체에 적용되는 파라미터(해상도, 샘플 포맷 등)를 정의하고 있으나,

VPPS_NUT은 현재 정의되어 있지 않다.

- TSPS_NUT/TPPS_NUT: 시간적 리샘플링 과정에 필요한 파라미터들을 포함한다.
- SSPS_NUT/SPPS_NUT: 공간적 리샘플링 과정에 필요한 파라미터들을 포함한다.
- RSPS_NUT/RPPS_NUT: 관심 영역 처리와 관련된 파라미터들을 포함한다.
- BSPS_NUT/BPPS_NUT: 비트 심도 절감 처리에 필요한 파라미터들을 포함한다.

이때, 전체 시퀀스에 동일하게 적용되는 파라미터들은 각 프로세스별로 정의된 시퀀스 파라미터 세트(Sequence Parameter Set) NAL Unit에 포함되어 있으며, 프레임별로 적용해야 하는 파라미터들을 픽처 파라미터 세트(Picture Parameter Set) NAL Unit에 포함된다. 각 NAL Unit은 VCL(Video Coding Layer)과 non-VCL(Non-Video Coding Layer)로 구분되는데, VCL은 실제 영상의 픽셀 데이터와 압축된 영상 데이터를 포함하는 계층을 의미하며, non-VCL은 비디오 데이터를 디코딩하기 위한 메타 데이터 및 파라미터 정보를 제공하는 계층을 의미한다.

제안한 VCM 비트스트림 구조는 VCM 내의 각 프로세스와 향후 기능 확장을 효율적으로 지원하기 위해 다양한 NAL Unit Type이 세분화되어 있으며, 각 프로세스에 필요한 시퀀스 레벨 및 픽처 파라미터들은 별도의 NAL Unit으로 독립적으로 관리된다. 디코더는 이러한 정보를 기반으로 효율적으로 복원 작업을 수행할 수 있다.

이러한 구조는 기존의 V3C 기반의 계층 구조와 다르게 단일 계층으로 구성되어 있어 파서의 복잡도를 감소시키고, 각 NAL Unit만의 헤더를 유지함으로써 중복된 헤더 구조로 인한 비트스트림 크기 증가 문제를 해결할 수 있다. 또한, 새로운 기능이 추가되더라도 기존 구조를 변경하지 않고 NAL Unit Type만 추가하여 확장할 수 있어 높은 유연성과 호환성을 제공한다.

IV. 제안한 VCM 비트스트림 구조를 적용한 시나리오

본 4장에서는 제안한 VCM 비트스트림 구조가 실제 VCM 프로세스에서 어떻게 구성되고 동작하는지 설명하

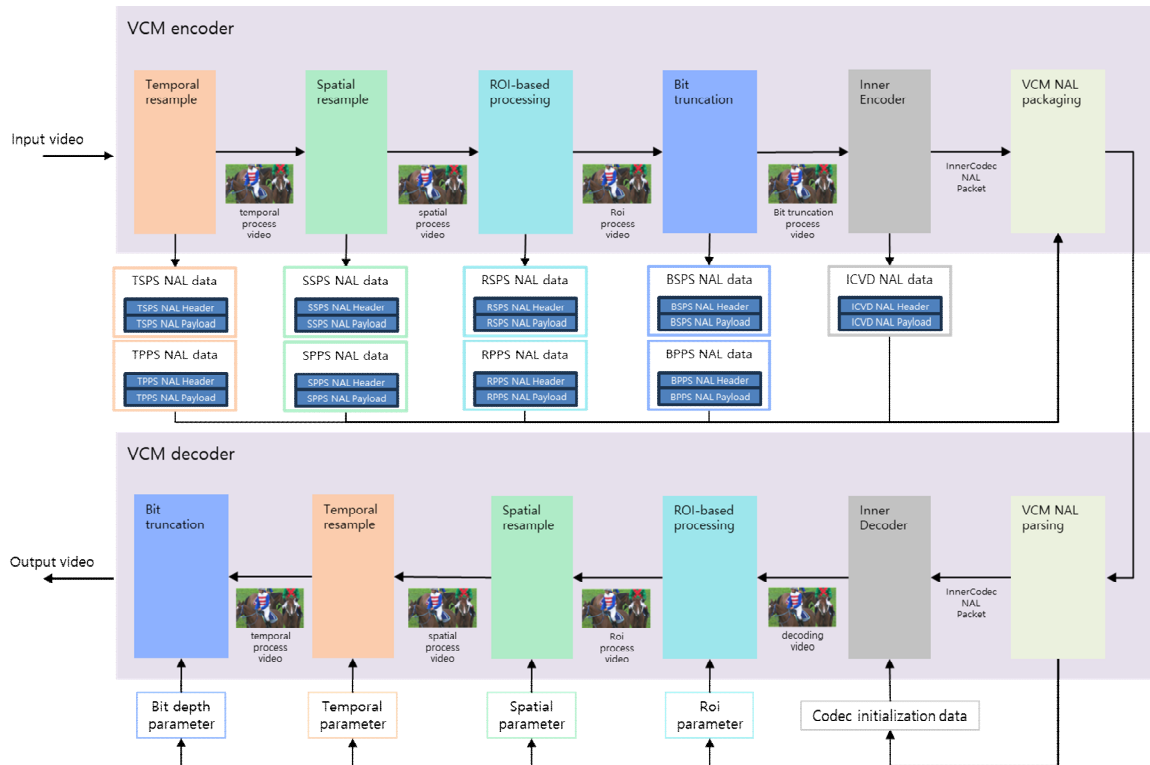


그림 3. 새로운 VCM 비트스트림을 적용한 VCM 프로세스
Fig. 3. VCM process with new VCM bitstream

고, 이를 기반으로 다양한 시나리오를 제시한다. 제한한 VCM 비트스트림 구조를 실제 VCM 프로세스에 적용하면 그림 3과 같이 인코더와 디코더 단계로 표현할 수 있다. 인코더 단계에서는 시간적 리샘플링, 공간적 리샘플링, 관심 영역 기반 프로세싱, 비트 심도 절감 처리, 내부 인코더의 각 프로세스를 거친 후, VCM NAL 패키징 단계에서 프로세스별로 생성된 NAL unit들이 VCM 비트스트림으로 구성된다. 각 NAL unit은 각 프로세스의 시퀀스 및 픽처 레벨의 파라미터를 독립적으로 포함하여 디코더에서 개별적으로 처리할 수 있도록 한다. 디코더 단계에서는 VCM NAL 파싱을 통해 전달된 각 NAL unit을 파싱한 후 내부 디코더를 거쳐 복원 데이터가 생성된다. 이후 각 프로세스의 역순으로 데이터를 처리하여 원본에 가깝게 복원된 비디오 데이터를 얻는다.

제한한 구조는 다양한 시나리오에 따라 효율적으로 운용될 수 있으며, 구체적으로 초기화 패킷(Initialization Packet), 단일 프로세스 적용 사례, 두 가지 이상의 프로세

스 적용 사례로 나누어 설명한다.

1. 초기화 패킷

초기화 패킷은 비디오 데이터를 본격적으로 전송하기 전에 디코더가 비디오 코덱과 관련된 필수 초기화 정보를 제공받을 수 있도록 설계된 구조이다. 그림 4는 초기화 패킷 구조를 나타내며, ICVD_NUT(Initialization Codec Video Data) 타입의 NAL Unit을 통해 내부 코덱 초기화 데이터가 전달된다. 초기화 패킷을 통해 디코더는 이후 전송되는 비디오 데이터를 정확히 디코딩할 준비를 마치게 된다.

2. 단일 VCM 프로세스

단일 VCM 프로세스가 적용된 경우는 특정 하나의 프로세스(시간적 리샘플링, 공간적 리샘플링, 관심 영역 기반 프로세싱, 비트 심도 절감 처리 등)만 수행하여 비트스트림을

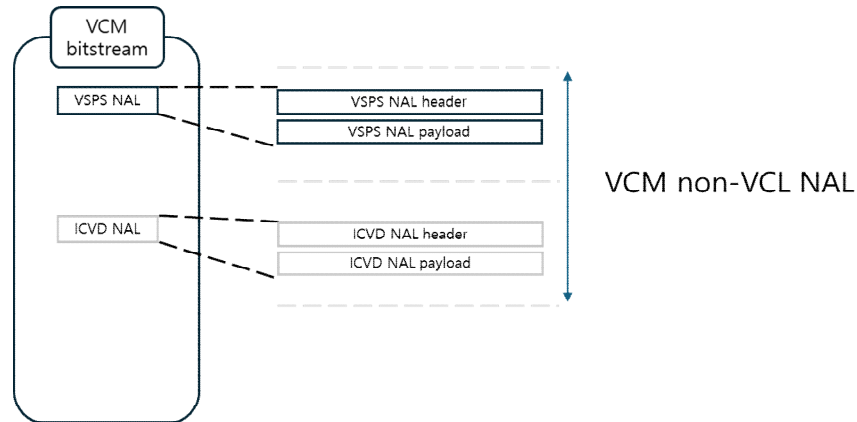


그림 4. 초기화 VCM 비트스트림 구성
Fig. 4. Initialization VCM Bitstream Configuration

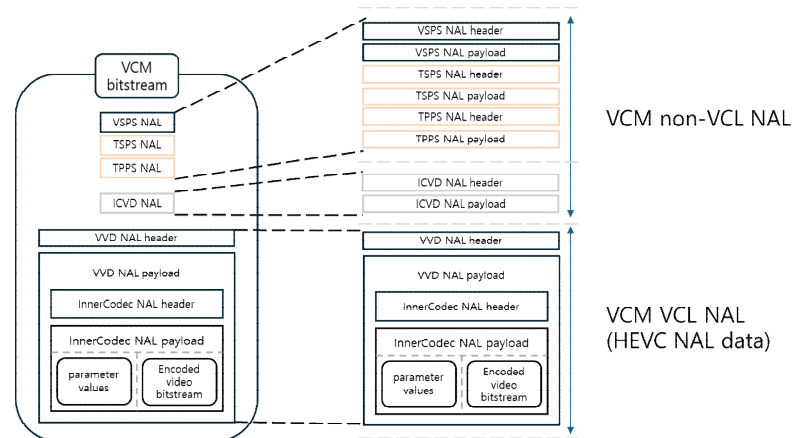


그림 5. 단일 VCM 프로세스(Temporal)의 비트스트림 구성
Fig. 5. Bitstream Configuration for a Single VCM Process (Temporal)

구성하는 시나리오를 의미한다. 그림 5에서는 시간적 리샘플링만 수행된 사례를 예시로 보여준다. 이 경우 비트스트림은 ICVD_NUT 초기화 패킷과 시간적 리샘플링을 위한 TSPS_NUT(Temporal Sequence Parameter Set) 및 TPPS_NUT(Temporal Picture Parameter Set)으로 구성된다. 이 시나리오는 단순한 VCM 환경에서 매우 효율적으로 작동할 수 있으며, 최소한의 파라미터 세트로 비트스트림을 구성함으로써 전송 효율성을 극대화할 수 있다.

3. 둘 이상의 VCM 프로세스

VCM 환경에서는 다양한 프로세스를 조합하여 비디오

데이터를 처리할 수 있다. 이러한 복합적 시나리오의 예시는 그림 6을 통해 표현되며, 각각 두 가지 이상의 프로세스가 함께 동작하는 상황을 나타낸다.

그림 6은 전체 프로세스가 적용된 사례이다. 이 경우 비트스트림은 초기화 패킷(ICVD_NUT)부터 모든 프로세스에 해당하는 시퀀스 및 픽처 레벨의 NAL들과 내부 코덱을 위한 관련 파라미터로 구성된다. 이를 통해 디코더는 시간적 리샘플링 처리 후 압축된 비디오 데이터를 효율적으로 복원할 수 있다.

이러한 다양한 시나리오를 통해 제안된 VCM 비트스트림 구조는 기존의 V3C 기반 구조 대비 구조적 단순성과 계층 축소에 따른 파싱 효율 향상, 중복 헤더 제거에 따른

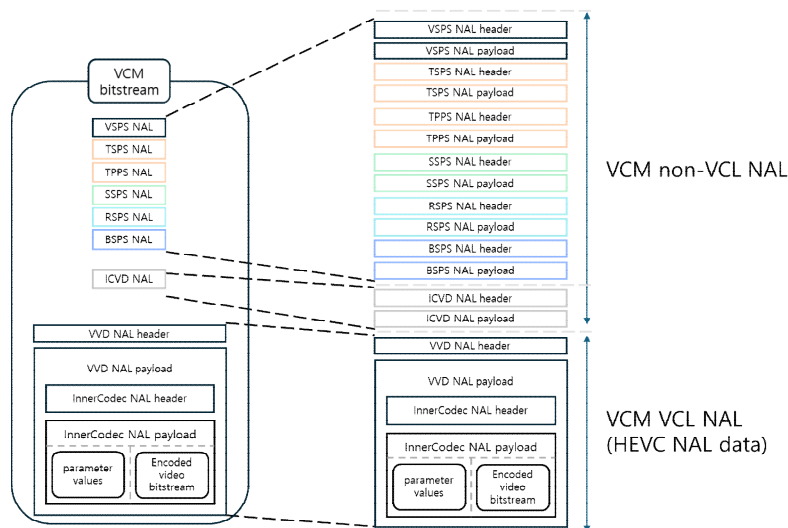


그림 6. 여러 VCM 프로세스의 비트스트림 구성
Fig. 6. Bitstream Configuration for Multiple VCM Processes

전송 오버헤드 절감이라는 측면에서 효율성을 갖춘 것을 확인할 수 있다. 예를 들어, 단일 프로세스 시나리오에서는 필수적인 최소한의 NAL Unit만 포함하여 전송 용량을 절감할 수 있으며, 복합 프로세스 시나리오에서도 각 프로세스에 대응하는 NAL Unit이 독립적으로 관리되기 때문에 디코더는 필요한 프로세스만 선택적으로 파싱할 수 있다.

또한, NAL Unit 단위의 독립적 구성 방식은 추후 새로운 프로세스가 추가되는 경우에도 기존 구조를 수정하지 않고 새로운 NAL Unit Type을 정의하여 손쉽게 확장할 수 있다는 점에서 유연성을 확보한다. 이러한 구조적 유연성은 실제 MPEG 회의에서 논의 중인 150차 버전에서의 적용 가능성을 뒷받침할 수 있으며, 향후 다양한 실험 환경 및 네트워크 조건에서의 전송 실험을 통해 전송률, 파싱 속도, 복원 정확도 등의 수치 기반 평가로도 구조의 우수성을 입증할 수 있을 것이다.

V. 결 론

본 논문에서는 기계를 위한 비디오 부호화를 보다 효율적이고 유연하게 처리하기 위한 새로운 VCM 비트스트림 구조를 제안하였다. 기존 구조의 중복된 헤더 오버헤드를

제거하고 VCM Unit 계층을 생략함으로써 Sample Unit과 NAL Unit으로 구성된 간결한 2계층 구조로 재구성하였다. 이를 통해 비트스트림의 효율성을 높이고 파싱 복잡성을 낮출 수 있었다.

제안한 구조는 프로세스별 독립적인 시퀀스 및 픽처 레벨 파라미터 구성을 통해 데이터 전송과 복원을 효율적으로 수행할 수 있게 한다. 다양한 시나리오를 통해 단일 프로세스와 복합적 프로세스 환경에서도 유연하고 효과적인 비디오 데이터 처리가 가능함을 입증하였다. 또한 향후 추가적인 프로세스가 필요할 경우에도 기존 구조의 변경 없이 간단히 NAL Unit Type만 추가하여 확장할 수 있는 유연성을 확보하였다.

상기 2장과 3장에서 논의된 내용을 기반으로, 향후 MPEG 150차 회의에서는 더욱 발전된 형태의 VCM 비트스트림이 도출될 것으로 기대된다. 향후 연구로는 새로운 프로세스가 VCM에 추가될 경우, 이에 따라 새롭게 구성되는 NAL Unit을 정의하고, 해당 NAL의 syntax와 semantics를 정의하여 구현할 예정이다. 또한, MMT 또는 DASH와 같은 전송 표준을 적용하여 실제 전송 환경에서 더욱 최적화된 비트스트림 구조와 시나리오를 도출하는 연구를 진행할 계획이다. 이에 따라 제안한 비트스트림 구조는 향후 VCM 기반의 다양한 응용 환경에서 높은 성능과 확장성을

제공할 것으로 기대된다.

참 고 문 헌 (References)

- [1] J.V.Team, "Advanced Video coding for generic audiovisual services," ITU-T Recommendation H.264 and ISO/IEC 14496-10 AVC, May 2005.
- [2] High Efficiency Video Coding, Recommendation ITU-T H.265 and ISO/IEC 23008-2 (HEVC), ITU-T and ISO/IEC JTC 1, Apr. 2013.
- [3] Versatile Video Coding, Recommendation ITU-T H.266 and ISO/IEC 23090-3 (VVC), ITU-T and ISO/IEC JTC 1, Aug. 2020.
- [4] Visual Volumetric Video-based Coding and Video-based Point Cloud Compression, ISO/IEC 23090-5, ISO/IEC JTC 1/SC 29, Nov. 2023.
- [5] GAO, Wen, et al. Recent standard development activities on video coding for machines. arXiv preprint arXiv:2105.12653, May. 2021. doi: <https://doi.org/10.48550/arXiv.2105.12653>
- [6] KWON, H. J., et al. Standardization trends in video coding for machines. Electronics and Telecommunications Trends, 35.5: 102-111, 2020. Oct. doi: <https://doi.org/10.22648/ETRI.2020.J.350509>
- [7] Seo, J., et al. Status of Video Coding for Machines standardization for M2M video communications. The Journal of Korean Institute of Communications and Information Sciences, 40.9: 58-67, 2023. Sep.
- [8] FISCHER, Kristian, et al. Video coding for machines with feature-based rate-distortion optimization. In: 2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP). IEEE, p. 1-6, Sep. 2020. doi: <https://doi.org/10.1109/MMSP48831.2020.9287136>
- [9] L. Duan, J. Liu, W. Yang, T. Huang, and W. Gao, "Video Coding for Machines: A Paradigm of Collaborative Compression and Intelligent Analytics," IEEE Transactions on Image Processing, Vol.29, pp. 8680-8695, Aug. 2020. doi: <https://doi.org/10.1109/TIP.2020.3016485>
- [10] ISO/IEC JTC 1/SC 29/WG 04 N0637. Working draft 6 of video coding for machines, Geneva, WG 04 MPEG Video coding, January 2025.
- [11] ISO/IEC JTC 1/SC 29/WG 04 N0642. VCM-CE 7: Bitstream formats, Geneva, WG 04 MPEG Video coding, January 2025.
- [12] Park, K., et. al., [VCM] Improved VCM Bitstream Structure Based on m70278 and Size Comparison Analysis, ISO/IEC JTC 1/SC 29/WG 04, m71463, Geneva, January 2025.
- [13] Hualong Yu et. al., [VCM] Design of VCM syntax structure, ISO/IEC JTC 1/SC 29/WG 04, m68250, Sapporo, July 2024.
- [14] Hualong Yu et. al., [VCM] Bitstream structure design for random access in VCM, ISO/IEC JTC 1/SC 29/WG 04, m71344, Geneva, January 2025.
- [15] Hualong Yu et. al., [VCM] Update of signaling and process of temporal restoration, ISO/IEC JTC 1/SC 29/WG 04, m70139, Kermer, November 2024.
- [16] Rhee, S., Lee, M., Kim, K. Analysis of Feature Map Compression Efficiency and Machine Task Performance According to Feature Frame Configuration Method. Journal of Broadcast Engineering, 27(3), 318-331, 2022. May. doi: <https://doi.org/10.5909/JBE.2022.27.3.318>
- [17] GAO, Wen, et al. An open dataset for video coding for machines standardization. In: 2022 IEEE International Conference on Image Processing (ICIP). IEEE. 4008-4012, Oct. 2022. doi: <https://doi.org/10.1109/ICIP46576.2022.9897525>
- [18] Lee, D., Jeon, S., Jeong, Y., Kim, J., & Seo, J. Exploring the Video Coding for Machines Standard: Current Status and Future Directions. Journal of Broadcast Engineering, 28(7), 888-903, 2023. Dec. doi: <https://doi.org/10.5909/JBE.2023.28.7.888>
- [19] Lee, Y., Yoon, K. Region of Interest Scaling Compression Techniques for VCMs. Journal of Broadcast Engineering, 29.4: 443-451, 2024. July. doi: <https://doi.org/10.5909/JBE.2024.29.4.443>
- [20] Y Lee, K Yoon. Compare the performance of VCM region-of-interest-based compression techniques. Journal of Broadcast Engineering, 29.6: 842-850, 2024. Nov. doi: <https://doi.org/10.5909/JBE.2024.29.6.842>
- [21] Rhee, S. Kim, K Study post-filter application analytics for VCM. Papers from the Korean Association of Broadcast Media Engineering Conference, 160-162, 2024. Jun.
- [22] Ahn, E., et al. Performance analysis of spatial resampling techniques for VCMs. Journal of Broadcast Engineering, 29.6: 879-887, 2024. Nov. doi: <https://doi.org/10.5909/JBE.2024.29.6.879>

— 저 자 소 개 —



박 기 범

- 2023년 2월 : 경희대학교 전자공학과 공학사
- 2025년 2월 : 경희대학교 전자정보융합공학과 공학석사
- 2025년 3월 ~ 현재 : 경희대학교 전자정보융합공학과 박사과정
- ORCID : <https://orcid.org/0009-0006-5256-3173>
- 주관심분야 : 딥러닝, 영상처리, 멀티미디어 시스템



이 진 영

- 1998년 5월 : B.S. EECS Michigan State University
- 1999년 12월 : M.S. EECS Michigan State University
- 2008년 12월 : Ph.D. EECS Michigan State University
- 2004년 3월 ~ 현재 : 한국전자통신연구원
- ORCID : <https://orcid.org/0000-0002-8718-1961>
- 주관심분야 : AI기반 영상처리, 멀티미디어 시스템, 메타데이터 처리



정 세 윤

- 1995년 2월 : 인하대학교 전자공학과 학사
- 1997년 2월 : 인하대학교 전자공학과 석사
- 2014년 8월 : KAIST 전기및전자공학과 박사
- 2014년 ~ 2015년 : USC 방문연구원
- 1996년 ~ 현재 : 한국전자통신연구원 미디어부호화연구실 책임연구원
- ORCID : <https://orcid.org/0000-0002-1675-4814>
- 주관심분야 : 실감 방송, 비디오 코딩, 컴퓨터 비전



김 규 현

- 1989년 2월 : 한양대학교 전자공학과 공학사
- 1992년 9월 : 영국 University of Newcastle upon Tyne 전기전자공학과 공학석사
- 1996년 7월 : 영국 University of Newcastle upon Tyne 전기전자공학과 공학박사
- 1996년 ~ 1997년 : 영국 University of Sheffield, Research Fellow
- 1997년 ~ 2006년 : 한국전자통신연구원 대화형미디어연구팀장
- 2006년 ~ 현재 : 경희대학교 전자정보대학 교수
- ORCID : <http://orcid.org/0000-0003-1553-936X>
- 주관심분야 : 디지털 방송, 영상처리, 멀티미디어 통신, 디지털 대화형 방송