

특집논문 (Special Paper)

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

<https://doi.org/10.5909/JBE.2025.30.3.331>

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

기계를 위한 특징 부호화 표준화 동향

한희지^{a)}, 이성은^{a)}, 서예원^{a)}, 정순흥^{b)}, 최해철^{a)*}

Feature Coding for Machines: Emerging Standards and Trends

Heeji Han^{a)}, Seongeun Lee^{a)}, Yewon Seo^{a)}, Soon-heung Jung^{b)}, and Haechul Choi^{a)*}

요약

최근 감시, 산업 자동화, IoT 응용의 확산으로 기계가 분석하는 영상 데이터의 양이 인간 소비를 초과하고 있으며, 이에 따라 기계 소비를 위한 영상 트래픽이 급격히 증가하고 있다. 이러한 변화에 대응하기 위해, 연산 부담을 에지 디바이스와 클라우드 서버 간에 분산하는 분산 신경망 실행 방식이 등장하였으며, 이러한 방식에서는 에지 디바이스에서 생성된 중간 특징의 효율적 전송이 요구된다. 그러나 기존 비디오 부호화 기술은 인간 시각 품질을 기준으로 설계되어 있어, 신경망 중간 특징과 같은 비 시각적 데이터에 최적화되어 있지 않다. 또한, 대규모 데이터 전송이 요구되는 스마트 시티 및 자율주행 환경에서는 원본 영상이나 특징 데이터를 그대로 전송할 경우 네트워크 대역폭이 극심하게 소모된다. 이에 따라, 중간 특징을 효율적으로 압축하면서도 신경망 표현의 중요한 정보를 유지할 수 있는 새로운 부호화 기술의 필요성이 대두되고 있다. 이에, ISO/IEC JTC 1/SC 29 MPEG Video에서는 기계를 위한 특징 부호화(feature coding for machines, FCM) 표준화를 추진하여, 신경망 특징 데이터를 기계 소비에 최적화된 방식으로 압축하는 기술을 개발하고 있다. 본 논문에서는 FCM의 표준화 동향을 소개하고, 부호화 구조 및 성능 평가 조건과 함께 테스트 모델의 성능을 분석한다.

Abstract

The expansion of surveillance, industrial automation, and IoT applications has led to a significant increase in machine-analyzed video traffic, exceeding that intended for human consumption. This shift requires split inference, which shares the computational load between edge devices and cloud servers, making the transmission of intermediate features generated at the edge necessary. However, conventional video coding technologies, designed to optimize human visual perception, are not well suited for compressing non-visual data such as intermediate features. Moreover, in scenarios such as smart cities and autonomous driving, where large volumes of data must be transmitted, directly sending original videos or uncompressed features leads to severe network bandwidth consumption. Thus, there is a growing need for new coding technologies that can efficiently compress intermediate features while preserving important information needed for machine analysis. In response, ISO/IEC JTC 1/SC 29 MPEG Video has launched standardization efforts on Feature Coding for Machines (FCM), which aims to efficiently compress neural network features optimized for machine vision applications. This paper introduces the FCM standardization progress, coding framework, and evaluation conditions, and analyzes test model performance.

Keyword : Feature Coding, Machine Vision, Video Coding, Neural Network, MPEG Standard

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

최근 딥러닝 기반 비전 알고리즘, 모바일 네트워크, 센서 기술의 발전에 따라, 기계가 인간의 시각적 인식을 대체하는 사례가 증가하고 있으며, 이에 따라 머신 비전의 활용 범위 또한 빠르게 확대되고 있다. 머신 비전은 기계가 카메라 및 소프트웨어 시스템을 이용하여 영상을 분석하고 의사 결정을 수행하는 기술로, 산업 자동화, 자율주행, 영상 보안, 스마트 시티 등 다양한 응용 분야에서 핵심적인 역할을 하고 있다. 특히, 기계가 직접 영상을 해석하고 처리하는 사례가 증가하면서, 인간의 인지 능력을 초과하는 수준의 정확도를 보이는 AI 기반 비전 시스템이 도입되고 있다.

기존 영상 부호화 기술은 인간의 시각적 경험을 최적화하는 방향으로 발전해 왔기 때문에, 기계가 소비하는 비디오 데이터의 특성을 충분히 반영하지 못한다. 기계 소비 작업에서는 전체 영상의 시각적 품질보다 분석 및 추론에 필요한 공간적·구조적 특징의 정확한 전달이 더욱 중요하다. 그러나 기존 화질 중심 부호화 방식은 이러한 기계적 요구를 만족시키기에 한계가 있으며, 결과적으로 분석 성능 저하 및 데이터 전송 비효율을 초래할 수 있다.

한편, 모바일 통신 기술의 발전과 네트워크에 연결된 디바이스 수의 급증으로 인해, 현재 인터넷 트래픽의 82% 이상이 영상 데이터로 구성되고 있으며^[1], 기계 간 통신이 차지하는 비중 또한 지속해서 증가하고 있다^[2]. 이러한 변화에 의해, 대기 시간(latency)과 연산 부담을 줄이면서 대량

의 영상 데이터를 효율적으로 처리하는 기술의 필요성이 증가하고 있다. 그러나 대규모 데이터를 저장하고 전송하는 과정은 네트워크 및 연산 자원에 상당한 부담을 초래할 수 있으며, 특히 기계 소비를 위한 영상 트래픽이 증가함에 따라 이러한 문제가 더욱 심각해지고 있다. 따라서, 기계 학습 및 비전 시스템이 필요로 하는 핵심 정보를 보존하면서도 전송 및 저장 효율을 동시에 개선할 수 있는 새로운 부호화 기술이 요구된다.

이러한 요구를 충족하기 위해, ISO/IEC JTC 1/SC 29 MPEG Video에서는 기계를 위한 특징 부호화(Feature Coding for Machines, FCM) 표준화 프로젝트를 추진하고 있다^[3]. FCM은 기계 학습 및 비전 알고리즘에서 활용되는 특징 데이터를 직접 압축하여 전송 비용을 절감하는 것을 목표로 한다. 이는 단순한 픽셀 기반 압축이 아닌, 기계가 필요로 하는 핵심 정보를 유지하면서 데이터 크기를 최소화하는 새로운 접근 방식을 제시한다. 본 논문에서는 MPEG FCM 표준화 동향과 주요 응용 사례 및 요구사항을 정리하고, v6.1 릴리스를 기준으로 전체 파이프라인 구조와 성능 평가 방법을 소개한다. 또한, FCM 참조 소프트웨어의 부호화 효율 분석을 통해 현재까지의 성능을 평가하고, 향후 표준화 진행 방향에 대해 살펴본다.

II. FCM 표준화

1. 표준화 동향

기계가 시각 기반 임무를 수행할 때 머신 비전의 성능을 유지하면서 특징 데이터를 효율적으로 압축할 수 있는 비트스트림 표현의 개발을 목표로, 2019년 7월 제127차 MPEG 회의에서 ‘기계를 위한 비디오 부호화(Video Coding for Machines, VCM)’에 대한 논의가 시작되었다^[4]. FCM은 초기에 VCM 프로젝트의 일부로써, 영상 압축과 특징 압축이 구분 없이 함께 논의되었으며, 이후 압축 방식에 따라 트랙이 분리되면서 현재의 FCM으로 발전하였다. VCM Adhoc Group(AhG)은 2019년 10월 제128차 MPEG 회의에서 결성되었으며, 이 회의에서 보다 구체적인 서비스 시나리오, 요구사항, 초기 코덱 구조 및 성능 평가 방법

a) 한밭대학교 소프트웨어융합대학원 지능미디어공학과(Department of Intelligent Media Engineering, Graduate School of Software Convergence, Hanbat National University)

b) 한국전자통신연구원(Electronics and Telecommunications Research Institute)

‡ Corresponding Author : 최해철(Haechul Choi)

E-mail: choihc@hanbat.ac.kr

Tel: +82-42-821-1149

ORCID: <https://orcid.org/0000-0002-7594-0828>

※ 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), and by Korea Research Institute for defense Technology planning and advancement(KRIT) grant funded by the Korea government(DAPA(Defense Acquisition Program Administration)) (KRIT-CT-22-047, Space-Layer Intelligent Communication Network Laboratory, 2022)

• Manuscript March 31, 2025; Revised April 30, 2025; Accepted April 30, 2025.

등에 대한 논의가 진행되었다. 초기 단계에는 휴먼 비전, 머신 비전, 그리고 두 가지를 모두 고려하는 하이브리드 비전의 세 가지 파이프라인이 코덱 구조의 후보로써 제안되었으며^[5], 이는 각각 휴먼 비전을 위한 영상 압축, 머신 비전을 위한 특징 압축, 그리고 두 접근 방식을 결합하는 하이브리드 방식을 포괄할 수 있도록 설계되었다. 이후 VCM은 본격적인 표준화를 추진하기에 앞서, 후보 기술의 가능성과 성능을 평가하기 위해 2020년 10월 기술조사요청서(Call for Evidence, CfE)의 초안을 발행하였다. 그에 앞서, 2020년 7월까지의 회의에서는 CfE 응답 기술에 대한 성능 평가를 위해 앵커의 구성 방식이 논의되었으며, 신경망과 데이터셋의 종류에 따라 다양한 앵커 구성이 제안되었다. 2021년 10월에는 영상을 직접 압축하는 접근 방식이 상대적으로 빠르게 표준화가 진행됨에 따라, VCM은 특징 압축을 담당하는 Track 1과 영상 압축을 담당하는 Track 2로 분리되어 각각 독립적으로 표준화를 진행하게 된다.

신경망으로부터 추출된 특징은 일반적으로 서로 다른 공간 해상도를 갖는 복수의 특징 맵으로 구성되며, 이러한 구성은 다중 스케일 특징으로 불린다. VCM Track 1에서 초기에 제안되었던 앵커 구성은 이러한 다중 스케일 특징을 Versatile Video Coding(VVC)으로 압축하기 위해, 복수의 특징 맵을 하나의 단일 특징 프레임 형태로 패키징하고, 이를 양자화한 후 VVC에 입력하는 방식으로 설계되었다. 그러나 다중 스케일 특징은 전체 데이터 크기가 원본 영상보다 큰 경우가 많아, 부호화 효율을 향상시키기 위한 다양한 경량화 및 축소 기법이 함께 논의되었다. 예를 들어, 주성분 분석을 기반으로 한 차원 축소 기법^[6], 패키징된 특징을 리사이징 하는 방법^[7] 등이 제안되었다. 하지만, 당시 특징 데이터를 압축하는 방식은 특징의 크기가 여전히 매우 컸기 때문에 본격적인 표준화 진행을 위해 그 효과성과 필요성을 입증할 수 있는 근거가 더 필요한 상황이었다. 2022년 1월, 다중 스케일 특징을 단일 스케일로 형태로 융합하는 신경망 기반 특징 축소 기술^[8]이 소개되어, 전체 특징의 크기를 효과적으로 줄이면서도 임무 성능을 유지할 수 있음이 확인되었다. 이후 제안 기술들은 다중 스케일 특징 자체를 직접 압축하는 비 신경망 기반 특징 축소 기술과, 추출된 특징을 신경망 모델을 통해 융합하는 방식의 신경망 기반 특징 축소 기술로 구분되어 발전하게 되었으며, 이들 중 일부는

기존 VVC 테스트 모델(VTM)을 활용한 픽셀 기반 압축 방식에 비해, 동일한 임무 성능에서 더 낮은 비트율을 달성하여 임무 성능 대비 비트율 효율면에서 우수한 결과를 보였다.

이러한 근거를 바탕으로, 2022년 7월 VCM 그룹은 특징 압축 기술에 대한 기술조사요청서^[9]를 발행하였으며, 같은 해 10월 회의에서 제안 기술에 대한 평가가 진행되었다. 이 시점을 기준으로 프로젝트 명칭은 Feature Coding for Video Coding for Machines(FCVCM)으로 변경되었고, 본격적인 표준화를 준비하기 위해 기술제안요청서(Call for Proposal, CfP) 이후의 평가 방식에 대한 보다 정교한 논의가 함께 이루어졌다. 그 결과, CfE 응답 기술들이 VVC 대비 최대 -98.01%의 부호화 효율을 달성함이 확인되었으며, 이를 바탕으로 2023년 4월 기술제안요청서(Call for Proposal, CfP)가 공표되었다. 2023년 10월 MPEG 회의에서는 기술제안요청서에 대한 응답 기술들의 평가가 이루어졌으며, 최적의 성능을 보인 기술들을 기반으로 참조 소프트웨어인 Feature Compression Test Model(FCTM)^[11]의 첫 번째 버전이 개발되었다. 기계 학습 기반 비디오 압축 실험을 위한 오픈소스 플랫폼인 CompressAI-Vision^[12]이 기술 통합을 위한 베이스라인 플랫폼으로 채택되었으며, 이에 따라 응답 기술 중 우수한 성능을 보인 주요 기술들^[13,14]이 이 플랫폼 위에 통합되는 형식으로 FCTM이 구축되었다. 이와 함께, FCVCM 프로젝트는 WG 2(Technical Requirements)에서 WG 4(Video)로 이관되어, 표준화가 본격화됨에 따라 프로젝트 명칭 또한 FCVCM에서 FCM으로 변경되었다.

이후 회의부터는 주요 기술의 성능 개선과 안정화가 지속적으로 진행되었다. 2024년 1월 회의에서는 기존 베이스라인 모델보다 향상된 성능을 보이는 새로운 특징 축소 모델^[15]이 채택되었으며, 4월 회의에서는 위원회 초안(Committee Draft, CD) 발간을 준비하기 위한 Preliminary Working Draft(PWD)의 첫 번째 버전^[16]이 공개되었다. 이후 채택된 기술들은 대부분 특징 변환(Feature Conversion) 과정에서 발생하는 손실을 줄이기 위한 기술들이 주를 이루고 있다^[17-20]. 2025년 1월에는 기존 베이스라인 모델의 경량화 버전^[21]이 채택되어 기술 최적화와 안정성 강화가 함께 이루어졌으며, 비트스트림 구조를 포함한 하イレ벨 선택스(HLS)의 설계가 중점적으로 논의되었다.

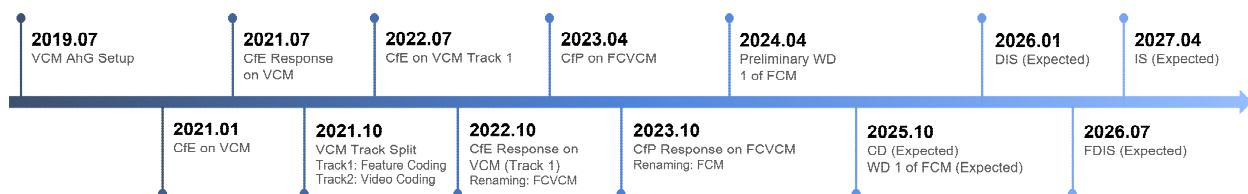


그림 1. FCM 표준화 주요 일정 및 예상 일정

Fig. 1. Timeline of FCM Standardization Activities

또한, FCM의 표준화 일정에 대해서도 논의가 이루어져, 2025년 10월 CD 발간과 2027년 4월 최종 국제표준 (International Standard, IS) 발간을 목표로 주요 일정이 수립되었다. 향후 회의에서는 베이스라인 구조의 대규모 변경보다는 구현 최적화, HLS 설계, 특징 변환 기술 등을 중심으로 논의가 이루어질 예정이다.

2. 응용 서비스 및 요구사항

기계가 주체가 되는 소비 환경에서는, 인간의 시각적 품질 인식과는 다른 기준에 따라 영상이 처리되므로, 표준화 과정에서는 이러한 특수성을 반영한 응용 시나리오와 기술적 요구사항의 정의가 중요하다. 2022년 4월까지 기준으로 정의된 문서에서는 서비스 시나리오를 Surveillance, Intelligent Transportation, Smart City, Intelligent Industry,

Intelligent Content, Consumer Electronics의 여섯 가지로 분류하였다. 또한, VCM은 다양한 기계 임무를 고려하여, 17가지 기술적 요구사항을 결정하였다. 요구사항들은 영상 압축과 특징 압축을 모두 포함하는 전체 VCM 시나리오를 기반으로 설정되었다. 이 중, VCM Track 1에서 다루는 특징 압축에 대한 기술 요구사항으로 13가지 필수 요구사항과 1가지 권장사항이 정의되었다^[4]. 표 1은 VCM Track 1 특징 압축의 요구사항을 정리한 것이다.

지금까지 FCM의 표준화 동향과 함께, 서비스 시나리오, 시각 임무, 기술적 요구사항을 정리하였다. 다양한 요구사항 중에서도 가장 핵심적인 요소는 압축 효율이며, 이를 평가하는 기준의 정립이 중요한 과제로 다루어졌다. FCM에서는 압축률과 임무 성능 간의 균형을 고려해야 하므로, 평가 지표와 절차에 대한 논의가 지속되어 왔다. 이에 따라, 3장에서는 부호화 파이프라인과 성능 평가 방법을 자세히 기술한다.

표 1. VCM Track 1 특징 압축 관련 요구사항 (필수/권장)^[4]Table 1. Requirements for VCM Track 1 (Mandatory/Recommended)^[4]

Category	Coding Target (feature/video)	Requirements
Mandatory Requirements (shall)	common	VCM shall support video coding for a machine task consumption purposes.
	feature coding	VCM shall support feature coding.
	common	VCM shall support a broad spectrum of encoding rates.
	common	VCM shall support various degrees of delay configuration.
	common	VCM shall be agnostic to network models.
	common	VCM shall be agnostic to machine task types.
	feature coding	VCM shall provide description of the meaning or the recommended way of using the decoded data.
	common	A single VCM bitstream shall support any number of instances of machine tasks.
	common	VCM complexity shall allow for feasible implementation within the constraints of the available technology at the expected time of usage.
	common	VCM shall support rectangular picture format up to 7680x4320 pixels (8K).
	common	VCM shall support fixed and variable rational frame rates for video inputs.
	common	VCM shall support any input source from video or image.
Recommended Requirements (should)	common	VCM shall support privacy and security (Mandated by ISO).
	common	VCM should support the use and inclusion of information such as descriptors in its bitstream.

III. FCM 부호화 구조 및 테스트 모델

1. FCM 부호화 파이프라인

앞서 설명한 서비스 시나리오와 요구사항을 반영하여 설계된 FCM의 중간 특징 압축 파이프라인은 그림 2와 같다. FCM은 분할 추론 시나리오를 기반으로, 신경망 모델의 중간 단계에서 생성되는 중간 특징(intermediate feature) 데이터를 효율적으로 압축하는 데 목적이 있다. 분할 추론이란 신경망 모델을 두 부분으로 분리하여, 첫 번째 부분(NN part1)은 에지 디바이스에서 중간 특징을 추출하고, 두 번째 부분(NN part2)은 클라우드나 서버에서 이를 입력으로 받아 객체 검출, 추적, 의미론적 분할 등의 기계 비전 임무를 수행하는 추론 구조를 의미한다. FCM 부호화기는 NN part1과 NN part2 사이에 위치하여, NN part1으로부터 추출된 중간 특징을 효율적으로 압축하고, NN part2가 이를 활용할 수 있도록 복원하는 기능을 수행한다.

중간 특징은 신경망 내부에서 추출된 중간 산출물로, 단일 또는 다중 스케일의 3차원 특징 맵으로 구성된다. 각 특징 맵은 입력 영상에 대한 신경망의 중간 표현을 담고 있으며, 채널 수와 공간 해상도(높이와 너비)는 신경망의 구조와 분할 지점에 따라 달라질 수 있다. FCM은 다양한 인공지능 모델과 분할 추론 시나리오에 대응할 수 있도록, 중간 특징의 형태와 구성에 관계없이 이를 효율적으로 처리할

수 있도록 설계되었다. 현재는 객체 검출, 객체 추적, 의미론적 분할 등의 시나리오를 중심으로 실험이 진행되고 있으며, 이들은 FCM의 구조적 유연성과 다양한 적용 가능성을 고려하여 공통 학습 및 실험 조건(Common Training and Test Conditions, CTTC)으로 선정된 작업들이다^[24]. 또한, 보다 다양한 신경망 모델과 분할 사례를 CTTC에 수용하기 위한 논의도 FCM 그룹 내에서 지속적으로 이루어지고 있다^[25].

표 2는 FCM CTTC에서 정의된 주요 신경망 임무와 이에 대응하는 신경망 모델 및 분할 지점을 요약한 것이다. 각 작업에 사용된 분할 지점은 해당 네트워크의 구조적 특성과 실험 목적을 고려하여 설정되었다. 예를 들어, 객체 검출 및 인스턴스 분할 작업에서는 대표적인 다중 스케일 특징을 생성하는 Faster R-CNN 또는 Mask R-CNN의 p-layer 계층이 분할 지점으로 사용되며, JDE 네트워크의 경우에는 평가 데이터셋(TVD, HiEve)에 따라 서로 다른 계층이 분할 지점으로 지정된다. 이러한 분할 지점들은 네트워크의 기능적 흐름을 기준으로 설정되며, 중간 특징의 정보 보존과 압축 효율 간의 균형을 고려하여 선택된다. 다양한 신경망 구조에서 생성되는 중간 특징을 효과적으로 압축하기 위해, FCM은 일련의 부/복호화 단계를 거치며, 그 구조는 그림 3에 설명된다. FCM 특징 부호화기는 중간 특징 정보를 압축해 데이터의 크기를 줄이고, 복호화기에서 다시 복원하는 구조를 가진다. 이를 위해 먼저 중간 특징을 추출한

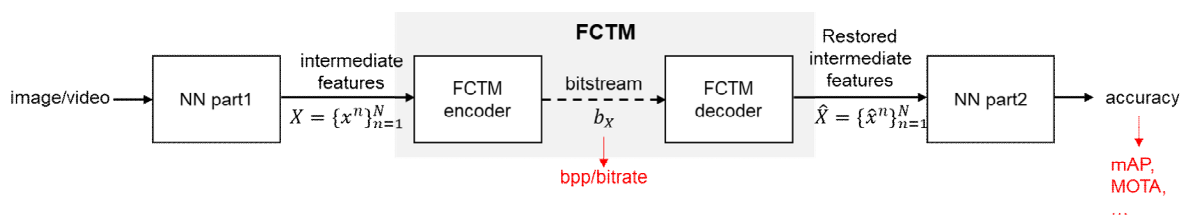


그림 2. 중간 특징 압축을 위한 FCTM의 입력력 파이프라인^[23]

Fig. 2. Input and Output of FCTM for intermediate feature compression^[23]

표 2. FCM 임무별 신경망 모델 및 분할 지점의 구성

Table 2. Neural networks and split points for FCM

Task	Neural Network	Split Point	
Object Detection and Instance Segmentation	Faster/Mask R-CNN ^[26]	p-layer	{p2, p3, p4, p5}
Object Tracking	JDE network ^[27]	darknet 53	{d36, d61, d72}
		alternative 1	{d105, d90, d75}

뒤, 몇 가지 단계를 거쳐 압축 가능한 형태로 변환하며, 전체 과정은 특징 축소(Feature Reduction), 특징 변환(Feature Conversion), 그리고 내부 코덱(Inner Codec)을 통한 특징 내부 부호화(Feature Inner Coding)의 세 가지 단계로 구성된다.

먼저, 특징 축소(Feature Reduction) 단계는 중간 특징의 공간적·시간적 크기를 줄이는 것을 목적으로 한다. 이 단계는 시간적 다운샘플링(Temporal Down-sampling), 특징 변환(Feature Transform), 특징 채널 조정(Feature Channel Adjustment)의 세 가지 처리로 구성된다. 이 중 시간적 다운샘플링은 연속된 특징 프레임 중 일부를 드롭하여 전체 처리해야 할 특징 데이터의 양을 줄이는 방식으로, 기본적으로는 사용되지 않도록 설정되어 있다. 이후 적용되는 특징 변환은 추출된 다중 스케일 특징을 단일 스케일 형태로 융합하고, 신경망 기반 인코더 구조를 통해 보다 간결한 표현으로 변환함으로써 특징 정보의 크기를 줄인다. 마지막으로, 특징 채널 조정 단계에서는 단일 특징 맵 형태의 데이터로부터 중요도가 낮은 채널을 제거하여 부호화할 전체 특징 데이터의 크기를 추가로 축소시킨다. 특징 변환(Feature Conversion) 단계에서는 축소된 특징을 내부 코덱에 입력할 수 있는 형태로 가공한다. 구체적으로, 먼저 3차원 특징 맵을 2차원 형태로 정렬하는 패킹(Packing) 과정을 통해 2차원 특징 프레임 형태로 변형시키고, 이후 정규화(Normalization)를 통해 특징 데이터의 값의 범위를 [0, 1] 범위로 조정한다. 마지막으로, 양자화(Quantization) 과정을 통해 데이터 타입을 32비트 부동소수점에서 10비트 정수형으로 변환하여 부호화가 가능한 형태로 만든다. 특징

내부 코딩(Feature Inner Coding) 단계는 부/복호화 과정을 모두 포함하며, 인코더 측에서는 일련의 변환 과정을 거친 특징 데이터를 내부 코덱을 통해 비트스트림 형태로 부호화한다. FCTM에서는 JM, HM, VTM, 신경망 기반 코덱 등 다양한 코덱이 내부 코덱으로써 활용될 수 있으며, 현재 FCTM은 CTTC를 따라 VTM-23.3을 내부 코덱으로 활용하고 있다. 이후 생성된 비트스트림은 복호화기로 전달되어 이후 처리에 활용된다.

FCM의 복호화기는 부호화기에서 생성된 비트스트림을 입력으로 받아, 이를 다시 중간 특징의 형태로 복원하는 역할을 수행한다. 전체 과정은 특징 내부 코딩의 특징 내부 복호화(Feature Inner Decoding), 특징 역변환(Feature Inverse Conversion), 그리고 특징 복원(Feature Restoration)의 세 가지 단계로 구성된다. 내부 복호화(Inner Decoding) 단계에서는 내부 부호화 단계에서와 마찬가지로, 내부 코덱으로 VVC를 사용해 비트스트림을 복호화함으로써 이를 다시 부호화 이전의 특징 형태로 재구성한다. 특징 역변환 단계에서는 복호화된 데이터를 원래의 중간 특징 형식으로 복원하는 과정이다. 역변환 단계는 역양자화(Dequantization), 언패킹(Unpacking), 축소된 특징 정제(Reduced Feature Refinement) 과정으로 구성된다. 먼저 역양자화를 통해 정수형 데이터를 0과 1 사이의 부동소수점 형식으로 복원하고, 특징 역패킹을 통해 2차원 특징 프레임 형태의 특징 데이터를 다시 3차원 텐서 구조로 재구성한다. 이어서 수행되는 축소된 특징 정제 단계에서는 인코더 측으로부터 시그널링된 축소된 특징에 대한 통계 정보(평균 및 표준편차)를 활용하여, 복원된 특징의 분포를 보정함

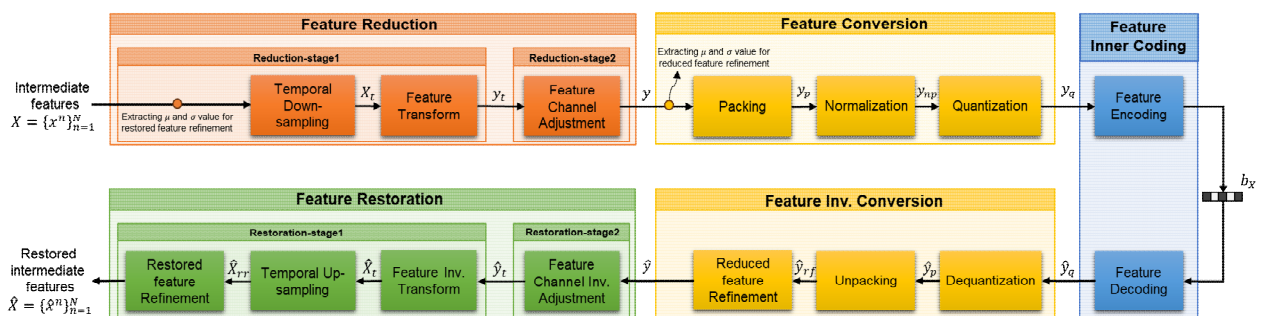


그림 3. FCTM의 파이프라인^[23]

Fig. 3. FCTM pipeline^[23]

로써 특징의 분포를 보다 원본에 가까운 형태로 되돌린다. 마지막으로, 특징 복원 단계에서는 인코더에서 축소되었던 구조적 정보들을 복원한다. 채널 복원(Inverse Feature Channel Adjustment) 단계에서는 축소된 채널 수를 원래 크기로 복원하고, 특징 역변환(Feature Inverse Transform)을 통해 단일 스케일로 변환되었던 특징을 원래의 다중 스케일 또는 다층 구조로 되돌린다. 이후, 시간적 업샘플링(Temporal Up-sampling)을 통해 다운샘플링 과정에서 생략되었던 프레임들 보간하여 복원하고, 마지막 복원된 특징의 정제 단계(restored feature refinement)에서는 인코더 측의 입력 특징(restored feature)의 통계 정보를 기반으로 전체 특징의 품질을 향상시킨다. 이와 같은 복호화 과정을 통해 최종적으로는 인코더에서 생성된 중간 특징과 동일한 구조를 갖는 출력이 재구성되며, 이는 NN part2로 전달되어 후속 신경망 임무 수행에 활용된다.

2. 성능 평가 조건 및 방법

이번 절에서는 FCM에 대한 공통 실험 및 학습 조건(CTTC)을 바탕으로, 실험 환경 및 조건, 추론에 사용되는 주요 데이터셋, 실험 수행 시 고려해야 할 사항, 그리고 평가 방법에 대해 기술한다. FCM의 평가용 데이터셋은 OpenImages V6, SFU-HW-Object-v3.2, Tencent Video Dataset(TVD), 그리고 HiEve 영상 데이터로 구성되며, 각각 객체 탐지(object detection), 인스턴스 분할(instance segmentation), 그리고 다중 객체 추적(multi-object tracking) 임무에 활용된다. 각 데이터셋과 임무 조합에 따라 선정된 신경망 구조와 분할 지점을 기준으로 성능이 평가되며, 이

에 대한 구성은 표 3에 요약되어 있다. 성능 평가는 임무별로 정의된 성능 지표에 따라 이루어진다. 객체 탐지 및 인스턴스 분할의 경우, $mAP@0.5$ ^[28] 및 $mAP@0.5 - 0.95$ ^[29]가 사용되며 여기서 mAP는 예측 결과와 실제 정답 간의 겹침 정도를 나타내는 Intersection over Union(IoU)를 기반으로 계산된다. $mAP@0.5$ 는 IoU 임계값 0.5에서의 평균 정밀도를, $mAP@0.5 - 0.95$ 는 IoU 임계값을 0.5부터 0.95까지 0.05 간격으로 변화시키며 계산한 평균 정밀도를 의미한다. 객체 추적은 Multi-Object Tracking Accuracy(MOTA)^[30] 지표를 기반으로 평가되며, 이는 탐지 오류(오탐, 미탐)와 탐지 객체의 ID 스위칭을 모두 고려하여 추적 성능을 종합적으로 측정하는 지표이다. 또한, 비트레이트는 이미지의 경우 bits per pixel(BPP), 영상의 경우 kbps 단위로 측정된다.

FCM의 성능 평가는 성능 포인트(Performance Points, PPs)를 기반으로 수행된다. 총 네 개의 성능 포인트(PP1부터 PP4까지)가 정의되며, 각 포인트는 네트워크의 임무 수행 성능을 기준으로 결정된다. 예를 들어, PP1은 분할되지 않은 전체 네트워크의 기준 성능을 나타내고, PP4는 가장 낮은 성능 지점을 나타낸다. FCM의 모든 제안 기술은 각 성능 포인트에서의 결과가 해당 성능 범위 내에 위치해야 한다. 이를 통해 BD-rate 계산이 가능한 곡선 중첩(overlapping curve)을 확보할 수 있다. 특히 BD-rate 계산 시에는 결과가 단조성(monotonicity)을 만족해야 하며, 일부 지점의 성능이 허용 범위 내에 있더라도 곡선의 형태가 적절하지 않으면 계산이 불가능할 수 있다. BD-rate는 동일한 성능 수준에서의 비트레이트 차이를 기반으로 압축 효율을 비교하는 지표이므로, 각 성능 포인트에서의 결과는

표 3. FCM의 데이터셋, 신경망 임무, 신경망 모델 및 평가 지표의 구성
Table 3. Configuration of FCM dataset, task, model, and evaluation metric

Dataset	Task	Network	Bitrate metric	Performance metric
Openimages V6 (5000 images)	Object Detection / Instance Segmentation	Faster R-CNN X101-FPN / Mask-CNN X101-FPN	BPP	$mAP@0.5$
SFU-HW	Object Detection	Faster R-CNN X101-FPN	kbps	$mAP@0.5-0.95$
TVD	Multi Object Tracking	JDE-1088x608 (DN53)	kbps	MOTA
HiEve	Multi Object Tracking	JDE-1088x608 (ALT1)	kbps	MOTA

8은 OpenImages V6(OIV6), HiEve, Tencent Video Dataset (TVD) 데이터셋 및 분할점의 조합에 대한 성능 변화를 나타낸 그래프로, 평가 조건 변화가 상대적으로 적었던 사례를 기반으로 성능을 비교한다. SFU 데이터셋은 v2.0.0 릴리스에서 평가 metric이 $mAP@0.5$ 에서 $mAP@0.5 - 0.95$ 로 변경되었고, v4.0 릴리스에서부터는 개선된 ground truth로 평가됨에 따라 버전 간 성능 편차가 크게 나타나 본 비교에서는 제외하였다.

FCTM은 매 릴리스마다 BD-rate 성능이 꾸준히 개선되는 경향을 보이고 있다. CFP 당시 앵커는 추출된 중간 특징을 그대로 압축하는 구조로, 압축해야 할 특징 데이터량이 상대적으로 많은 편이었다. Remote inference는 입력 이미지 혹은 비디오를 VVC로 직접 부호화하는 방식으로, CFP 앵커보다는 특징 데이터량이 적지만 여전히 상당한 수준의 비트량이 관찰된다. FCTM v1.0.1에서는 CFP에서 채택되었던 두 개의 베이스라인 모델이 각각 객체 탐지·세분화 임

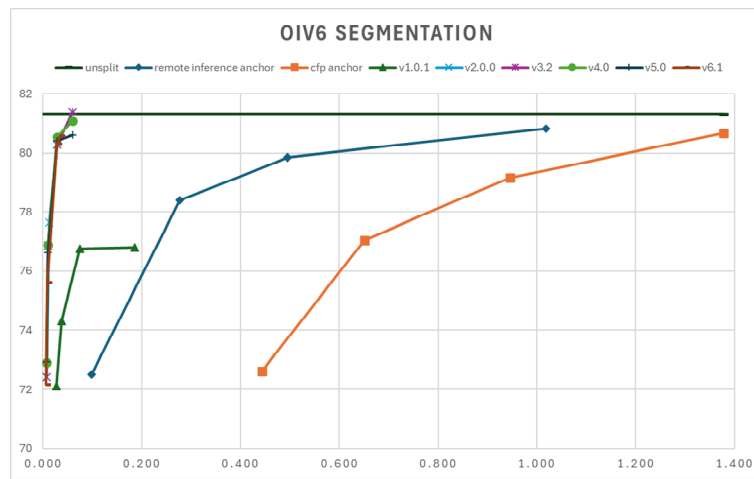


그림 4. FCTM의 버전별 OIV6 데이터셋에 대한 인스턴스 분할 임무의 RP-curve 비교
Fig. 4. RP-Curve Comparison of OIV6 Instance Segmentation Task Across FCTM Versions

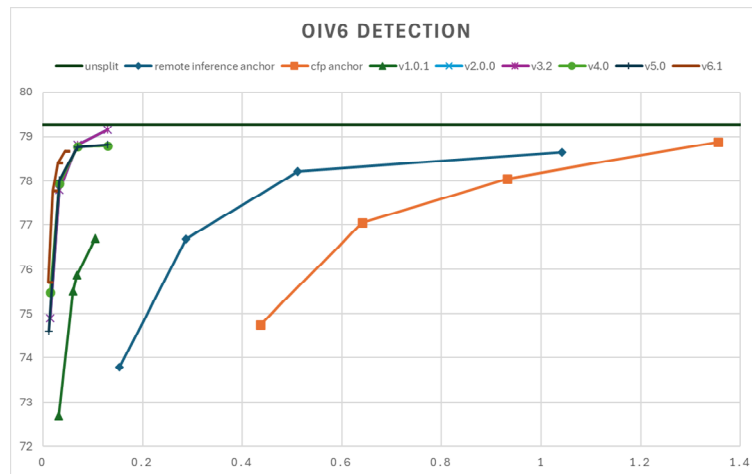


그림 5. FCTM의 버전별 OIV6 데이터셋에 대한 객체 검출 임무의 RP-curve 비교
Fig. 5. RP-Curve Comparison of OIV6 Object Detection Task Across FCTM Versions

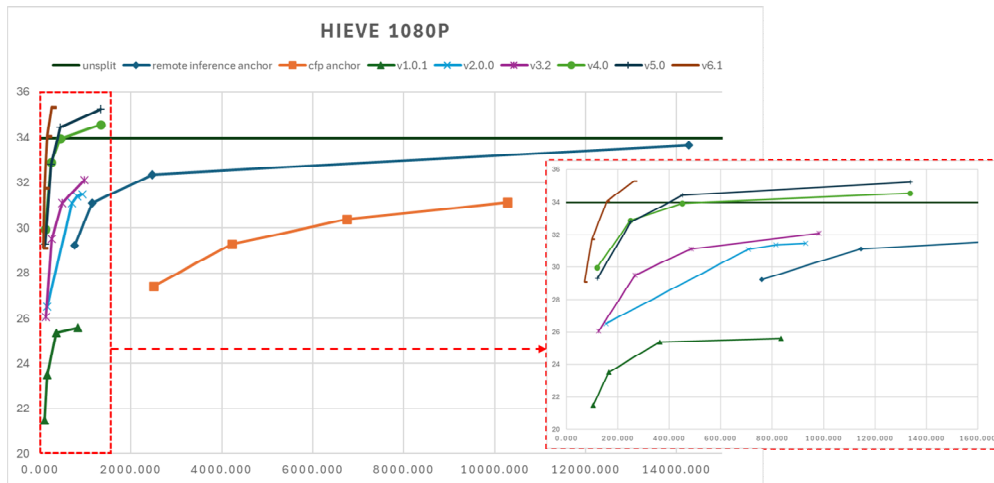


그림 6. FCTM의 버전별 HiEve1080p 데이터셋에 대한 객체 추적 임무의 RP-curve 비교

Fig. 6. RP-Curve Comparison of HiEve 1080p Object Tracking Task Across FCTM Versions

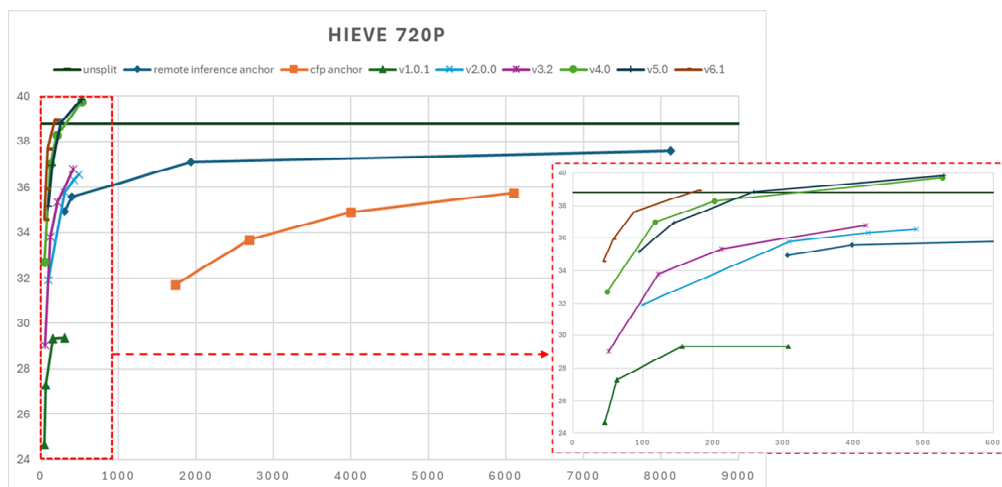


그림 7. FCTM의 버전별 HiEve720p 데이터셋에 대한 객체 추적 임무의 RP-curve 비교

Fig. 7. RP-Curve Comparison of HiEve720p Object Tracking Task Across FCTM Versions

무^[15]와 객체 추적임무^[14]에 적합하게 나뉘어 통합되었으며, 이후 베이스라인 모델의 변경은 총 두 차례 이루어졌다. 첫 번째는 v1.0.1에서 v2.0.0으로의 전환으로, 모델의 성능 향상을 목적으로 채택되었으며, 실제로 신경망 임무 수행 성능이 뚜렷하게 향상되었다^[16]. 두 번째는 v5.0에서 v6.1로의 전환으로, 모델 경량화를 목적으로 이루어졌으며, 기존 성능을 거의 유지하면서도 비트량은 절반 가까이 감소하였다^[22]. 그림에는 표현되지 않았지만, 인코딩 및 디코딩 복잡도 역시 유사한 비율로 단축되었다.

베이스라인 모델 변경 외에도, 또한 v4.0에서는 특징 정제(feature refinement) 도구^[18,19]가 도입되었으며, 이는 특징 변환 및 내부 코딩 단계에서 수행되는 두 차례의 양자화 과정에서 발생한 정보 손실을 보완함으로써, 특히 HiEve 데이터셋에서의 임무 성능 향상으로 이어졌다. v6.1에서는 BD-rate 측면에서 v5.0 대비 추가적인 효율 개선이 관찰되었으며, 이는 전체적인 압축 성능 향상으로 이어졌다. 반면, 일부 조건에서는 성능 포인트 1에서 v5.0 대비 임무 수행 성능의 하락이 관찰되었다. 이는 우수한 성능을 보이는 복

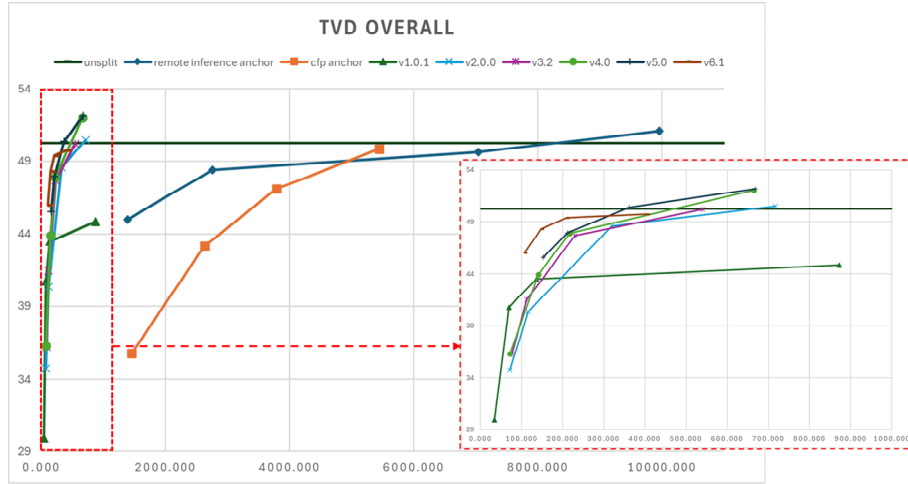


그림 8. FCTM의 버전별 TVD 데이터셋에 대한 객체 추적 임무의 RP-curve 비교
Fig. 8. RP-Curve Comparison of TVD Object Tracking Task Across FCTM Versions

수의 학습 기반 기술들^[22,31]이 충분한 융합 학습 없이 통합된 데에 기인한 것으로 보인다. 또한, 제한된 학습 기간으로 인해, 학습이 충분히 이루어지지 않아 발생한 문제일 가능성도 배제할 수 없다. 향후 유사한 기술 통합이 이루어질 경우, 기술 간 상호작용에 대한 사전 검토와 체계적인 융합 전략 마련이 필요할 것으로 판단된다.

IV. 결 론

기존 비디오 압축 기술은 인간의 시각적 품질에 최적화되어 발전해 왔으나, 영상 데이터를 기계가 주요 소비 주체로 활용하는 환경이 확산됨에 따라, 기계 소비에 특화된 새로운 압축 기술의 필요성이 대두되고 있다. 이에 MPEG에서는 기계 학습 기반 비전 시스템의 효율적인 운용을 위한 특징 부호화 기술인 FCM 표준화를 진행해 왔으며, 특히 분할 추론 기반 시나리오에 적합한 부호화 구조와 효율적인 전송 체계를 중심으로 기술이 고도화되었다. 표준화 측면에서 FCM은 CFP 이후 다양한 기술 제안을 통합하며 참조 소프트웨어인 FCTM을 구축하고, 성능 향상을 위한 베이스라인 모델 개선, 특징 보정 기술, 특징 채널 변환 기술들을 단계적으로 도입해 왔다. 이 과정에서 압축 효율 향상이 꾸준히 보고되었으며, 평가 체계 또한 성능 포인트 기반의 정량적 지표를 중심으로 점차 정교화되고 있다. 다만 일

부 릴리스에서는 신경망 수행 성능 저하가 관찰되었으며, 이는 복수 학습 기반 기술 도입 시 발생할 수 있는 구조적 간섭 및 통합 전략의 미비에서 비롯된 것으로 분석된다. 이러한 경험은 향후 신경망 기반 표준화 과정에서 기술 간 상호작용을 고려한 통합 및 최적화 전략의 필요성을 시사한다. 본 논문에서는 FCM의 구조, 성능, 평가 방법 및 표준화 동향을 정리하고, 테스트 모델의 부호화 효율을 분석함으로써 기계 소비 중심의 차세대 영상 부호화 기술 개발을 위한 기초 자료를 제공하고자 하였다.

참 고 문 헌 (References)

- [1] Cisco, Global Device Growth and Traffic Profiles, Cisco Visual Networking Index (VNI) Forecast Highlights, https://www.cisco.com/c/dam/m/en_us/solutions/service-provider/vni-forecast-highlights/pdf/Global_Device_Growth_Traffic_Profiles.pdf (accessed Oct. 22, 2022).
- [2] Cisco, Cisco Annual Internet Report (2018 - 2023) White Paper, <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html> (accessed Mar. 9, 2020).
- [3] Convenor of ISO/IEC JTC 1/SC 29/AG 03, "Press release of 144th MPEG meeting," in ISO/IEC JTC 1/SC 29/AG 03 N0129, October 2023.
- [4] ISO/IEC JTC 1/SC 29 WG 2, "Use cases and requirements for Video Coding for Machines," ISO/IEC JTC 1/SC 29 WG 2 w21545, Apr. 2022.
- [5] H.J. Kwon et al., "Standardization Trends in Video Coding for

- Machines,” *Electronics and Telecommunications Trends*, Vol.35, No.5, pp102-111, Oct. 2022.
doi: <https://doi.org/10.22648/ETRI.2020.J.350509>
- [6] ISO/IEC JTC 1/SC 29 WG 2, “Evaluation Framework for Video Coding for Machines,” ISO/IEC JTC 1/SC 29 WG 2 w21825, Jul. 2022.
- [7] D. Gwak et al., “[VCM track 1] Feature data compression based on generalized PCA for object detection,” ISO/IEC JTC 1/SC 29 WG 2 m58785, Jan. 2022.
- [8] C. Kim et al., “[VCM track 1] Feature Compression by Arranging Each Channel as a Tile of VVC,” ISO/IEC JTC 1/SC 29 WG 2 m58782, Jan. 2022.
- [9] H. Han et al., [VCM] Investigation on deep feature compression framework for multi-task, ISO/IEC JTC 1/SC 29 WG 2 m58772 Jan. 2022.
- [10] ISO/IEC JTC 1/SC 29 WG 2, “Call for Evidence for Video Coding for Machines,” ISO/IEC JTC 1/SC 29 WG 2 N0125, Jul. 2022.
- [11] ISO/IEC JTC 1/SC 29 WG 2, “CfE response report for Video Coding for Machines,” ISO/IEC JTC 1/SC 29 WG 2 N0247, Oct. 2022.
- [12] Feature Compression Test Model (FCTM), <https://git.mpeg.expert/MPEG/Video/fcm/fctm> (accessed Oct. 19, 2023).
- [13] H. Choi et al., “[FCVCM] CompressAI-Vision: codebase and evaluation framework for MPEG FCVCM,” ISO/IEC JTC 1/SC 29 WG 2 m65437, Oct. 2023.
- [14] H. Wang et al., “Response to the FCVCM CfP from China Telecom,” ISO/IEC JTC 1/SC 29 WG 4 m65181, Oct 2023.
- [15] H. Han et al., “[FCVCM] Response to CfP: Enhanced Multi-scale Feature Compression for FCVCM,” ISO/IEC JTC 1/SC 29 WG 4 m65705, Oct 2023.
- [16] H. Jeong et al., “[FCM] CE 1.1.8. L-MSFC-v2 with fine-tuning, ISO/IEC JTC 1/SC 29 WG 4 m66341,” Jan. 2024.
- [17] ISO/IEC JTC 1/SC 29 WG 4, “Preliminary WD of feature coding for machines,” ISO/IEC JTC 1/SC 29 WG 4 N0517, Apr. 2024.
- [18] Md. E.H. Eimon et al., “[FCM] CE4.2c Report: Reconstruction Refinement(RefRef) on Input & Fused Feature Tensor,” ISO/IEC JTC 1/SC 29 WG 4 m68664, Sapporo, Jul. 2024.
- [19] H. Choi et al., “[FCM] CE4.2-Related: Simplified refinement method,” ISO/IEC JTC 1/SC 29 WG 4 m68391, Sapporo, Jul. 2024.
- [20] G. Han et al., “[FCM] ce4.2.3a: Combination of QACT (CE4.2.1) and RBICR (CE4.2.2) with Selective Learning Strategy,” ISO/IEC JTC 1/SC 29 WG 4 m71203, Geneva, Jan. 2025.
- [21] F. Racapé et al., “[FCM] Clarifications on the feature unpacking process in both PWD and reference software,” ISO/IEC JTC 1/SC 29 WG 4 m71625, Geneva, Jan. 2025.
- [22] D. Lim et al., “[FCM] CE1.1.4: LightFCTM,” ISO/IEC JTC 1/SC 29 WG 4 m71247, Geneva, Jan. 2025.
- [23] ISO/IEC JTC 1/SC 29 WG 4 “Algorithm description of FCTM,” ISO/IEC JTC 1/SC 29 WG 4 N0626, Geneva, Jan. 2025.
- [24] ISO/IEC JTC 1/SC 29 WG 4, MPEG video Coding, “Common Test and Training Conditions for FCM,” ISO/IEC JTC 1/SC 29 WG 4 N00625, Geneva, Jan. 2025.
- [25] ISO/IEC JTC 1/SC 29 WG 4, “FE2: FCM CTTC study,” ISO/IEC JTC 1/SC 29 WG 4 N0632, Geneva, Jan. 2025.
- [26] Detectron2, <https://github.com/facebookresearch/detectron2>
- [27] JDE network, <https://github.com/Zhongdao/Towards-Realtime-MOT>
- [28] M. Everingham et al., “The PASCAL Visual Object Classes (VOC) Challenge,” *International Journal of Computer Vision (IJCV)*, 2010. doi: <https://doi.org/10.1007/s11263-009-0275-4>
- [29] T.-Y. Lin et al., “Microsoft COCO: Common Objects in Context,” *European Conference on Computer Vision (ECCV)*, 2014. doi: https://doi.org/10.1007/978-3-319-10602-1_48
- [30] K. Bernardin et al., “Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics,” *EURASIP Journal on Image and Video Processing*, 2008. doi: <https://doi.org/10.1155/2008/246309>
- [31] G. Han et al., “[FCM] CE4.2.3a: Combination of QACT (CE4.2.1) and RBICR (CE4.2.2) with Selective Learning Strategy,” ISO/IEC JTC 1/SC 29 WG 4 m71203, Geneva, Jan. 2025

저 자 소 개



한 희 지

- 2018년 : 국립한밭대학교 정보통신공학 (학사)
- 2020년 : 국립한밭대학교 멀티미디어공학 (석사)
- 2020년 ~ 현재 : 국립한밭대학교 지능미디어공학 박사과정
- ORCID : <https://orcid.org/0000-0001-8187-2553>
- 주관심분야 : 비디오 부호화, 특징 부호화, 컴퓨터 비전, 딥러닝

저 자 소 개



이 성 은

- 2024년 : 국립한밭대학교 정보통신공학 (학사)
- 2024년 ~ 현재 : 국립한밭대학교 지능미디어공학 석사과정
- ORCID : <https://orcid.org/0009-0004-4009-8775>
- 주관심분야 : 비디오 부호화, 특징 부호화, 컴퓨터 비전, 딥러닝



서 예 원

- 2025년 : 국립한밭대학교 정보통신공학 (학사)
- 2025년 ~ 현재 : 국립한밭대학교 지능미디어공학 석사과정
- ORCID : <https://orcid.org/0009-0002-3443-4774>
- 주관심분야 : 비디오 부호화, 특징 부호화, 컴퓨터 비전, 딥러닝



정 순 흥

- 2001년 : 부산대학교 전자공학 (학사)
- 2003년 : 한국과학기술원 전기및전자공학 (석사)
- 2016년 : 한국과학기술원 전기및전자공학 (박사)
- 2005년 ~ 현재 : 한국전자통신연구원 책임연구원
- ORCID : <https://orcid.org/0000-0003-2041-5222>
- 주관심분야 : 실감미디어, 컴퓨터 비전, 머신러닝, 영상 부호화, 영상처리



최 해 철

- 1997년 : 경북대학교 전자공학 (학사)
- 1999년 : 한국과학기술원 전기및전자공학 (석사)
- 2004년 : 한국과학기술원 전기및전자공학 (박사)
- 2004년 ~ 2010년 : 한국전자통신연구원 선임연구원
- 2010년 ~ 현재 : 국립한밭대학교 지능미디어공학과 교수
- ORCID : <https://orcid.org/0000-0002-7594-0828>
- 주관심분야 : 비디오 부호화, 컴퓨터 비전, 딥러닝