기계를 위한 비디오 부호화 표준화 동향

□ 추현곤, 정원식, 서정일 / 한국전자통신연구원

요 약

오늘날 인터넷 트래픽의 80% 이상은 이미지와 비디오와 같은 영상 정보가 차지하고 있으며, 딥러닝 기술의 발전과 더불어 영상을 사람이 아닌 머신이 처리하는 경우가 점점 늘어가고 있다. 사람의 시각적 특성과 머신이 처리하는 특징이 다를 수 있다는 점을 고려하여 MPEG을 비롯한 표준화 단체에서 딥러닝 네트워크를 포함한 기계(머신)를 위한 비디오 부호화에 대하여 표준화를 진행 중에 있다. 본 기고에서는 MPEG에서 진행되고 있는 머신 비전을 위한 영상 부호화 표준화 동향에 대해 정리한다.

I. 서 론

머신 비전은 인간이 눈으로 보고 판단하는 것을 기계(머신)가 가지고 있는 카메라와 같은 하드웨어와 소프트웨어의 시스템을 통해 대신 처리하는 기술이다. 최근 딥러닝 기술을 이용한 머신 비전의 정확도가 사람이 처리할 수 있는 정확도의 범위를 뛰어넘게 되었으며, 이를 바탕으로 기존의 산업 자동화 이외에 자율주행자동차, 영상 보안 및 안전 등으로 점점 그 응용 대상을 넓혀가고 있다. 더욱이 모바일 통신 기술의 발달로 인한 비디오 데이터가 양산되고 있어, 앞으로 기계를 통한 비디오 처리에 대한 수요는 점점 늘어가고 있는 상황이다. 특히, 이미지 및 비디오 데이터가 전체 인터넷 트래픽의 80% 이상을 차지하고 있어 비디오 데이터 용량을 줄이기 위한 부호화 효율 향상에 대한 요구도 계속 진행되고 있다. 기존 비디오 부호화 기술은 사람에게 최상의 영상 품질을 제공하도록 최적화되어 있다. 이는 사람이 아닌 비디오 데이터를 처리하는 기계의 입장에서 볼 때는 현재의 비디오 부호화 기술이 최적이 아닐 수 있다는 의미로 해석될 수 있으며, 기계 입장에서 최적화된 부호화 기술이 필요함을 의미한다. 이러한 새로운 패러다임은 차량 간의 연결과 IoT 기기, 대규모 비디오 보안 감시 네트워크, 스마트 시티 및 품질 검사 등의 출현이 주도하고 있

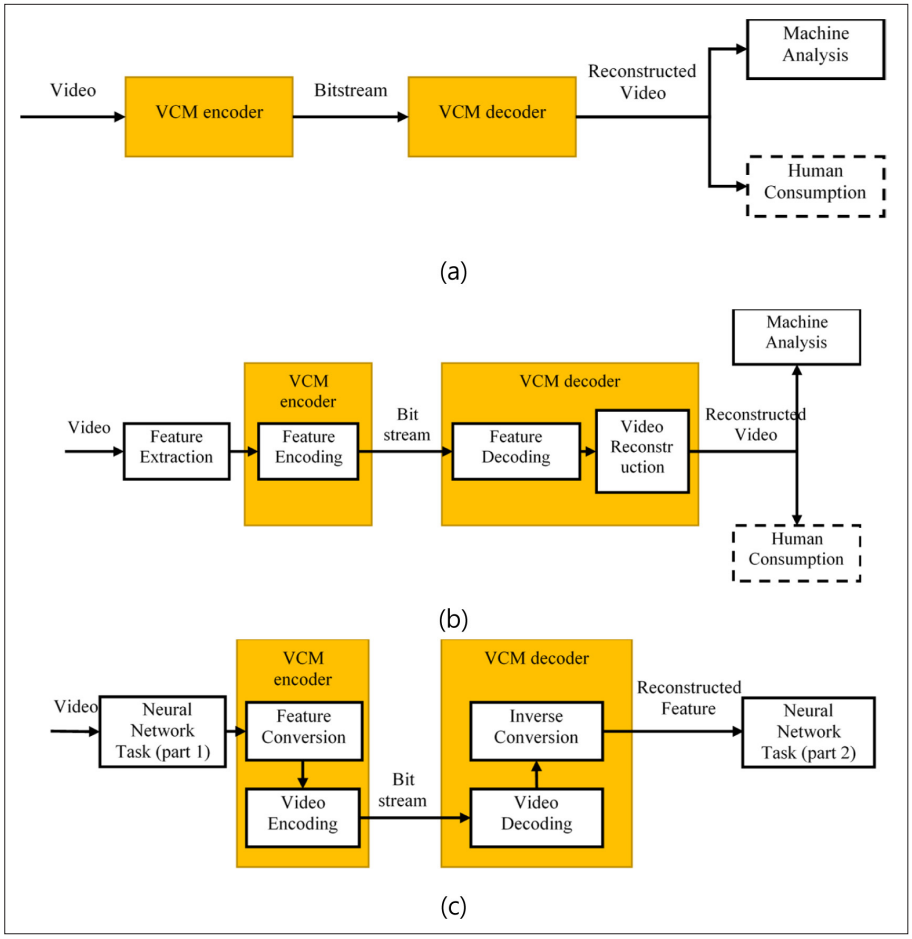
※ 본 논문은 과학기술정보통신부의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No. 2020-0-00011, 기계를 위한 영상 부호화 기술)

다. 이들은 모두 대기 시간과 규모에 대한 엄격한 요구사항을 가지고 있으며, 머신 비전을 목표로 하는 이미지/비디오 부호화에 대한 새로운 솔루션을 요구하고 있다.

이러한 머신 비전에 대한 요구사항은 기존의 시각적 부호화 연구로부터 상당한 변화를 나타내는 새로운 연구 방향과 접근방식에 영감을 주었다. 예를 들어, 다양한 분류와 추론 작업에 대한 최근의 딥러닝의 진보는 머신 비전 알고리즘의 적절한 압축 표현과 효과 사이의 관계에 대한 연구와 컴팩트한 특징에 대한 추가 연구를 유발하였다. 이러한 새로운 표현은 최첨단 압축 방식에 비해 전송 비용을 크게 절감하는 동시에 머신 비전 시스템이 대규모로 분산적으로 작동하는데 필요한 정보를 높은 정확도로 제공할 것으로 기대된다. 이와 관련하여 MPEG에서는 딥러닝을 포함한 머신 비전과 관련하여 VCM (Video Coding for Machines) 표준 기술에 대한 논의를 진행하고 있다[1]. 본 기고에서는 현재까지의 VCM의 표준화 활동에 대해서 분석하고 향후 표준화 진행 방향에 대하여 살펴본다.

1. VCM 표준화 범위

MPEG VCM 표준화에서 정의하고 있는 VCM의 표준



<그림 1> VCM 구조의 예

화 범위는 다음과 같다[1].

"*MPEG-VCM aims to define a bitstream from encoding video, descriptors or features extracted from video that is efficient in terms of bitrate and performance of a machine task after decoding.*"

상기 표준화 범위에서 정의된 바와 같이 VCM 코덱의 입력과 출력은 비전 업무를 수행하기에 필요한 video, descriptor 및 feature와 같은 다양한 형태가 될 수 있다. 이러한 다양한 입력과 출력 조합에 따른 VCM 구조의 예는 〈그림 1〉에서와 같다.

VCM에 대한 Use cases로 〈그림 2〉에 나타나 있는 대표적으로 일곱 가지를 제시하고 있다[1]. 전통적으로 머신 비전이 많이 사용되는 보안(Surveillance) 및 지능형 산업, 스마트 시티, 지능형 자동차 외에 지능형 콘텐츠, 지능형 가전 등을 포함하고 있다.

2. 요구사항

2022년 10월 기준 MPEG 표준화 회의에서 논의된 VCM 요구사항[1]은 총 17개로서, 그중 반드시 만족해야 하는 의무 요구사항('shall)이 16개로 대부분을 차지한다. VCM 요구사항들을 나열하면 〈표 1〉에서와 같다.

〈표 1〉에서 보는 바와 같이 VCM의 17개 요구사항은 크게 VCM 전반에 대한 요구사항, Track 1에 해당하는 Feature coding에만 적용되는 요구사항, Track 2에 해당하는 Video coding에만 적용되는 요구사항 및 Feature/video coding 모두에 적용되는 요구사항으로 구별되며, 각 요구사항이 어디에 해당되는지도 명시되어 있다.

3. 성능 평가 지표

VCM을 위한 성능 평가 지표는 평가 구조 문서[2] 및 공통 실험 환경[3] 문서에 정의되어 있다. VCM 부호화기의 성능 평가를 위해 사용하는 성능 평가 지표는 동일(유사)한 데이터에서의 각각의 머신 비전 성능에 따라 정해진다. 현재 머신 비전 임무별로 사용되는 성능 지표는 다음과 같다. 객체 검출(Object Detection)의 성능 평가를 위해서 특정한 범위의 IoU (Intersection over Union)값에 대해 객체 클래스별 AP (Average

<그림 2> VCM의 Use cases

<표 1> VCM Requirements

No.	Requirements
1	VCM shall support video coding for machine task consumption purposes.
2	VCM shall support feature coding. (Feature coding)
3	VCM shall support a coding efficiency improvement for at least 30% BD-rate over the VVC standard on machine vision tasks. (Video/Feature coding)
4	VCM shall support a broad spectrum of encoding rates.
5	VCM shall support various degrees of delay configuration. (Video 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/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)
13	VCM 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.
17	VCM shall support privacy and security.

Precision)를 평균하여 계산한 mAP (mean AP)를 사용한다.

$$Precision(T_{IoU}) = \frac{TP(T_{IoU})}{TP(T_{IoU}) + FP(T_{IoU})}$$

여기서 IoU는 두 개의 사각형 영역(정답 사각형과 예측된 영역 사각형) 사이의 일치하는 비율을 의미한다.

$$\text{IoU} = \frac{Area\ \ of\ \ Overlap}{Area\ \ of\ \ Union}$$

VCM 성능평가에는 0.5부터 0.95까지 0.05 간격의 매 IoU값에 대해 얻어진 AP값(AP@[0.5:.095])들의 평균한 값을 사용한다. 정지영상의 객체 분할 임무의 성능 평가 지표의 경우에도 객체 검출 임무와 동일한 mAP를 사용한다. 다만 객체 검출에서 객체가 포함된 박스 영역으로 IoU를 계산했다면, 객체 분할의 경우, 객체의 모양에 따라 이진 맵을 만들고 이진 맵들 간의 영역의 중복성을 이용하여 계산한다. 동영상 객체 분할 임무의 성능 평가 지표는 분할 영역 유사도를 측정하는 J score (Jaccard Index score)와 객체 경계선 기반 유사도를 측정하는 *F* score를 평균한 J&F mean을 사용한다[2].

동영상 객체 추적 임무의 성능 평가 지표는 MOTA (MOT Accuracy)를 사용한다[2]. t 번째 frame에서 GT를 Ground Truth 라고 할 때 MOTA는 다음과 같이 계산할 수 있다.

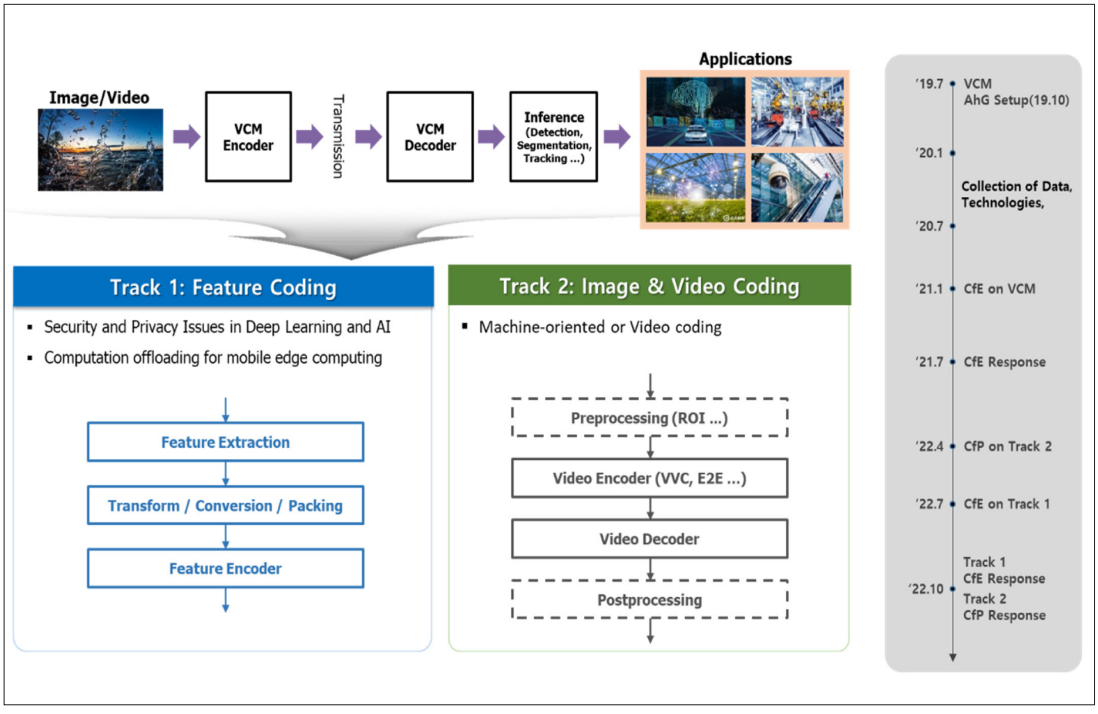
$$\text{MOTA} = 1 - \frac{\sum_t (FN_t + FP_t + IDSW_t)}{\sum_t GT_t}$$

여기서 FN은 False Negative 에러, FP는 False Positive 에러, IDSW (equivalently an identity switch) 에러는 두 개의 물체가 겹쳐질 경우에 동일하게 나타나지 않는 에러를 나타낸다. MOTA는 객체의 ID를 오인하거나 새로운 객체로 잘못 인식한 경우에 대한 에러를 고려함으로써, 정확하게 추적된 프레임 내의 물체의 궤도의 정확도를 계산하도록 고안되었다.

하이브리드 비전과 같이 사람이 보는 화질 평가를 위해서는 기존의 비디오 부호화에서 사용한 PSNR 및 SSIM값을 이용한다. 다만 기존의 비디오 부호화 개발에 사용한 주관적 화질 평가의 경우, VCM에서 사용하지 않는다. 또한 비디오 부호화 표준화에 있어서 계산 복잡도 평가를 위해 부복호화 수행시간도 평가에 고려될 수 있다.

4. 표준화 진행사항

2019년 7월 구텐버그 회의에서 기계가 시각 기반 임무를 수행할 때 머신 비전의 성능을 유지하면서 영상 데이터를 압축하는 비트스트림 표현에 대한 기술에 대한 논의가 시작되어, 2019년 9월에 Video coding for machines (VCM)란 이름의 새로운 비디오 부호화 방식에 대한 논의를 위해 MPEG VCM AhG (adhoc group)가 구성되었다. 2021년 4월 회의에서는 VCM 표준화 기술에 대한 CfE (Call-for-Evidence)[4]를 진행하여 이미지 부호화에 있어서 가능성을 검증하였으며, 이를 바탕으로, 2022년에 4월에 Image/Video Coding 트랙에 대한 CfP (Call-for-Proposal)[5]가 발간되어 10월에 이에 대한 기술 제안을 진행하였다. 또한 2022년 7월에는



<그림 3> 표준화 진행사항

Feature Coding 트랙에 대한 CfE[6]를 발간하고, 10월에 기술 평가를 진행하였다. 다음 장에서 각 Track별 진행사항 및 주요 기술에 대해 분석한다.

II. VCM Image and Video Coding Track

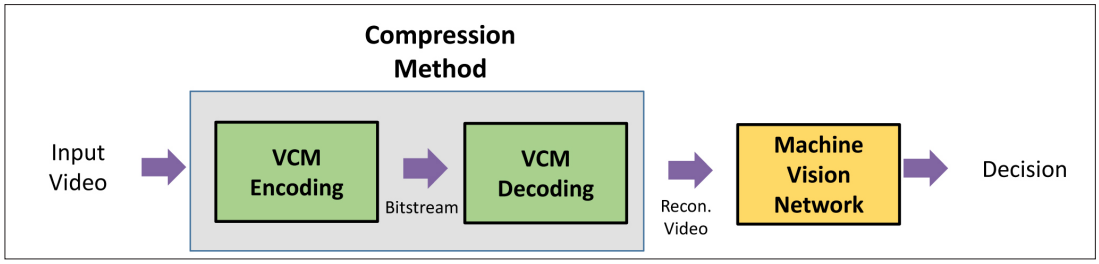
본 절에서는 MPEG VCM 그룹의 두 번째 트랙인 Image and Video Coding 트랙에 대해서 살펴본다. Image and Video Coding 트랙은 〈그림 4〉와 같이 영상 또는 비디오 입력을 받아서 압축된 비트스트림을 생성하고, 이를 다시 원래의 비디오와 같은 형태의 비디오로 복원한 후, 복원된 영상을 이용하여 머신 비전 네트워크의 입력으로 사용하는 형태의 부호화기를 의미한다.

Image and Video Coding 트랙의 경우, 2021년 4월에 Call-for-Evidence (CfE)에 대한 평가를 진행하였으며, 2022년 10월에 Call for Proposals (CfP)에 대한 제안 평가를 진행하였다. CfP에서의 평가는 Object Detection, Instance Segmentation, Object Tracking 세 개의 임무를 대상으로 성능을 비교하였으며, 각 임무에 대한 데이터셋은 다음과 같이 정해져 있다[2][3].

Image and Video Coding 트랙에 제안된 기술은 크게 전처리를 기반으로 한 영역기반 접근방식(ROI-based Approach)과 End-to-End 딥러닝 네트워크 방식으로 구분할 수 있다.

1. 영역기반 접근방식

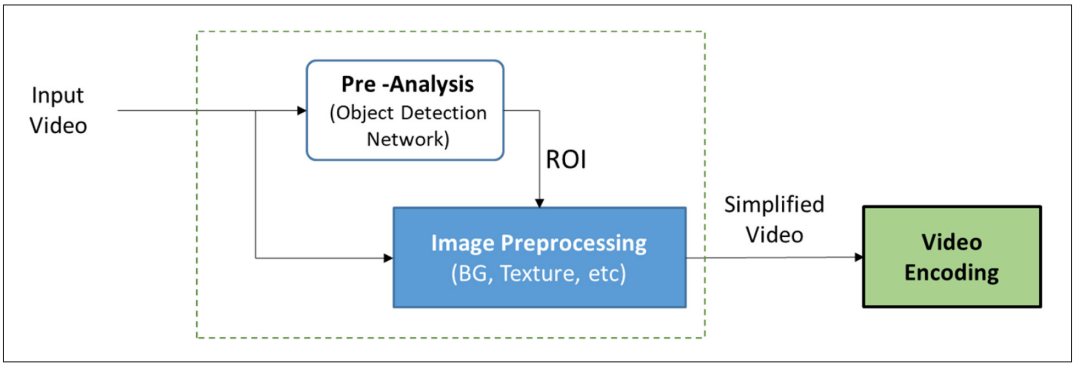
VCM에서 ROI (Region-of-Interest)는 머신 비전에서 주요 특징을 추출하는 데 사용되는 영역을 의미한다. 입력 비디오에 대해 머신 비전에 사용되는 네트워크 또는 네트워크 일부를 사용하여, 비디오 내의 주요 객체가 존재하는 영역 또는 프레임을 검출하고 해당하는 영역 또는 프레임만을 부호화하거나, 해당 영역 또는 프레임



<그림 4> Image and Video Track의 Pipeline

<표 2> Key tasks and Target Dataset for VCM Track2 CfP

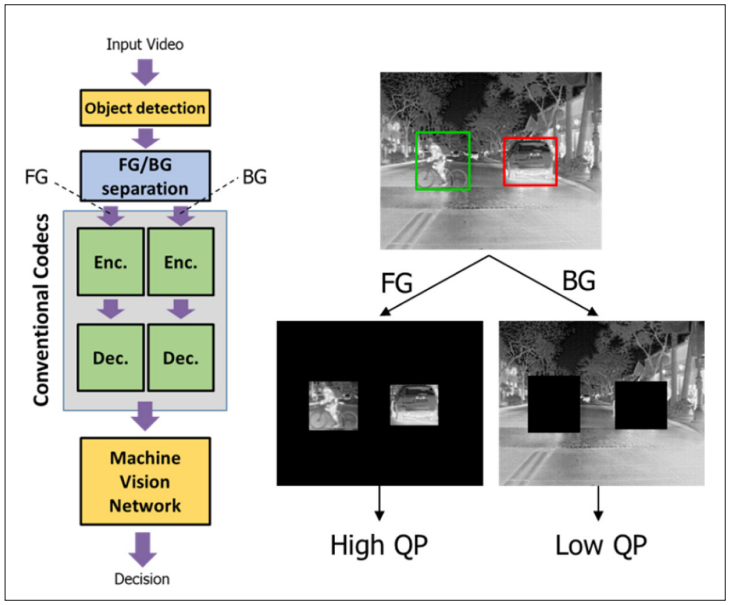
Target Tasks		Object Detection	Instance Segmentation	Object Tracking
Network		Faster-RCNN	Mask R-CNN	JDE-1088
Dataset	Image	FLIR, TVD, Openimages v6	TVD, Openimages v6	-
	Video	SFU-HW		TVD
Performance Metric		mAP/Bpp	mAP/Bpp	MOTA/Bpp



<그림 5> 영역기반 접근방식

에 더 많은 데이터를 부가하도록 하는 방법을 의미한다.

가장 기본적인 영역기반 부호화 방법은 ROI를 찾아서 해당하는 부분만을 전송하는 방법이다. 포츠난 대학에서는 ROI를 검출하여 배경을 단순화한 영상을 기존의 비디오 부호화를 이용하여 압축하는 방식[7, 8, 9]을 제안하였으며, ETRI와 명지대에서는 머신 비전을 이용하여 얻어진 결과에 대한 서술자와 Descriptor와 객체 정보 영상을 이용하여 다른 머신 비전 태스크인 객체 추적에 적용하는 기술[10]에 대해 제안하였다. 알리바바와 홍콩시립대학에서는 Yolov7 기반의 객체 검출 네트워크를 이용하여 물체의 경계를 추출하고, 이를 바탕으로 시간적, 공간적으로 정보를 줄여서 비트율을 낮추는 방식[11]을 제안하였으며, Florida Atlantic 대학과 OP Solutions에서는 머신 비전을 이용하여 물체가 포함된



<그림 6> 중요도에 따라 서로 다른 화질로 전송하는 방식의 예[14]

영역을 찾고, 이 영역만을 이용하여 영상을 재구성하여 전송하고, 원래의 영상으로 복원하는 방식의 부호화 기술[12]에 대해 제안하였다.

ROI 영역만을 보내지 않고, ROI에 대해 서로 다른 Bitrate을 부여하는 방법도 제안되었다. Ericsson에서는 머신 비전을 이용하여 물체가 포함된 영역을 찾고, 이 영역에 대한 정보를 기존의 VVC Encoder에 입력하여 각 영역의 화질인자(QP, Quality Parameter)를 조절하여 부호화 효율을 높이는 방법[13]을 제안하였다. ETRI와 건국대는 객체 인지 중요도에 따라 서로 다른 성능 지표를 할당하여 전송하는 방법을 제안하였다. CfE 기술 제안에서는 배경과 전경을 구분한 후, 이를 두 개의 스트림으로 나누어서 전송하는 부호화 방식[14]을 제안하였으며, CfP 기술 응답에서는 배경과 물체를 새로운 프레임으로 합성한 후, 서로 다른 화질로 부호화는 방안[15]을 제안하였다.

비디오 영상에 대한 샘플링을 위한 방법도 제안되었다. CAS-ICT와 차이나 텔레콤에서는 시간적으로 Frame sample만을 이용한 비디오 부호화 기술을 제안하였다[16]. 이 방법은 복호화기에서 기존의 프레임을 생성하기 위한 Interpolation 네트워크를 이용하여 중간 프레임을 재생성하였다. Tencent와 우한대학에서는 시공간 샘플링 기술을 이용하여 데이터의 용량을 줄이고, 복호화 시 심층신경망을 기반으로 한 Post Filtering을 이용한 솔루션에 대해 기술을 제안하였다[17, 18].

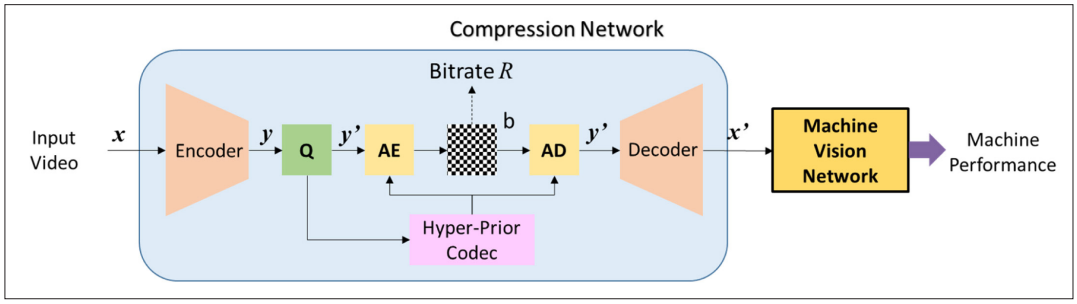
2. 딥러닝 네트워크 기반 압축 방식

딥러닝 기술의 발달로 인해 딥러닝 네트워크를 이용한 비디오를 비롯한 다양한 멀티미디어 압축 기술도 활발히 연구되고 있다. 딥러닝 기반 압축 기술은 심층신경망에 이미지 또는 비디오를 입력하여 제한된 형태의 은닉벡터를 추출하여 부호화한다. 일반적인 영상 압축의 경우, 압축 효율을 높이기 위해 심층신경망은 복원 영상의 화질은 높이면서 은닉벡터가 적은 비트로 표현될 수 있도록 학습된다. VCM에서는 딥러닝 네트워크를 이용하여 부호화 시 머신 비전 네트워크의 에러 함수를 같이 사용한다. 이를 통해 네트워크가 머신 비전에 더 유리한 방향으로 훈련이 될 수 있도록 한다.

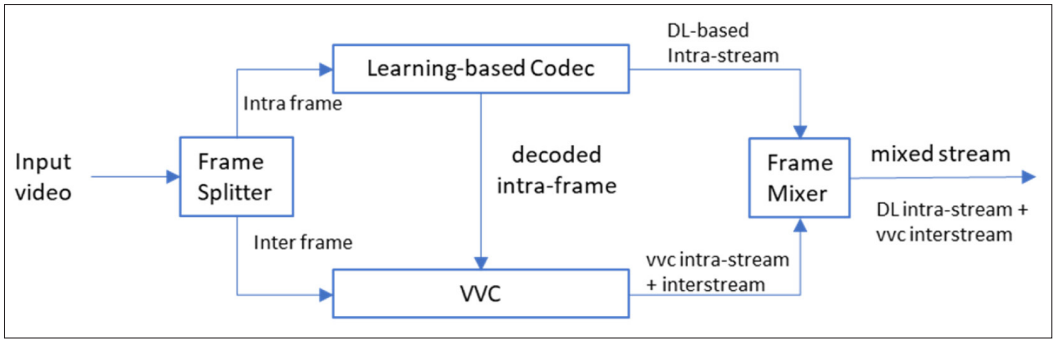
예를 들어 객체 검출 네트워크를 대상으로 할 경우 압축 네트워크에 대한 비용함수(Loss)는 다음과 같이 비트율(R)과 영상의 에러, 그리고 검출 네트워크의 비용을 조합하여 다음과 같이 표현할 수 있다.

$$L_{overall} = R + \lambda_{mse} L_{mse} + \lambda_{detect} L_{detect}$$

중국의 저장대학에서는 Key frame Encoder (MIKEnc)와 Key frame Decoder (MIKDec)로 구성된 심층신경



<그림 7> 딥러닝 네트워크 기반 압축 방식



<그림 8> Nokia의 Hybrid Codec 구조

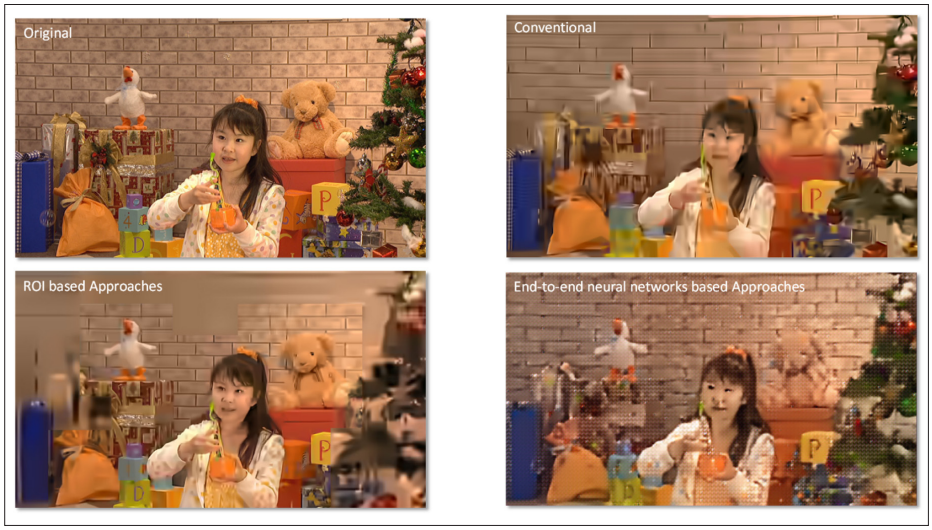
망 기반 부호화 기술[19]을 제안하였다. 이 방법에서 부호화기인 MIKEnc는 영상의 크기 조절 및 정규화를 수행하는 전처리 네트워크인 PreP, 이에 대한 특징을 추출하는 NN-FE 및 다시 특징을 압축하는 NN-FC 모듈로 구성되었으며, 부호화기인 MIKDec는 특징을 복호화하는 NN-FR, 특징벡터로부터 원래의 영상을 재구성하는 NN-IR, 그리고 영상의 크기 조절 및 정규화를 수행하는 후처리 네트워크로 구성하였다. Tencent와 우한대학에서는 Variable-rate Intra Coding과 Scale space flow Coding[22]을 이용한 End-to-end Learning based Solution을 제안하였다[20]. Intra Coding Network을 위해서 기존의 Cheng2020 with attention 모델[21]에 가변 bitrate 지원을 위한 Scalingnet을 추가하였다. 노키아에서는 Intra-frame coding에는 딥러닝 기반의 압축 부호화 기술을 이용하고, Inter-frame coding에는 VVC 코딩을 이용하는 하이브리드 형태의 부호화 기술을 제안하였다[23][24].

Track2에 대한 Call for Proposals (CfP)에 대한 평가 결과는 CfP test report 문서와 CfP response report에 제공되어 있다[25][26]. 제안 결과보고서를 통해 각 임무별 성능은 〈표 3〉과 같다. 임무별로는 VVC 대비 성능 비교 결과, 객체 검출 임무에서 39%, 객체 분할 임무에서 45%, 객체 추적 임무에서 57%의 성능 개선이 있은 것으로 나타났다. 〈표 3〉의 결과는 개별 임무에 대한 최고 성능에 대해 나타난 것으로 통합된 성능의 경우, 현재의 결과보다 낮게 나올 것으로 예상된다.

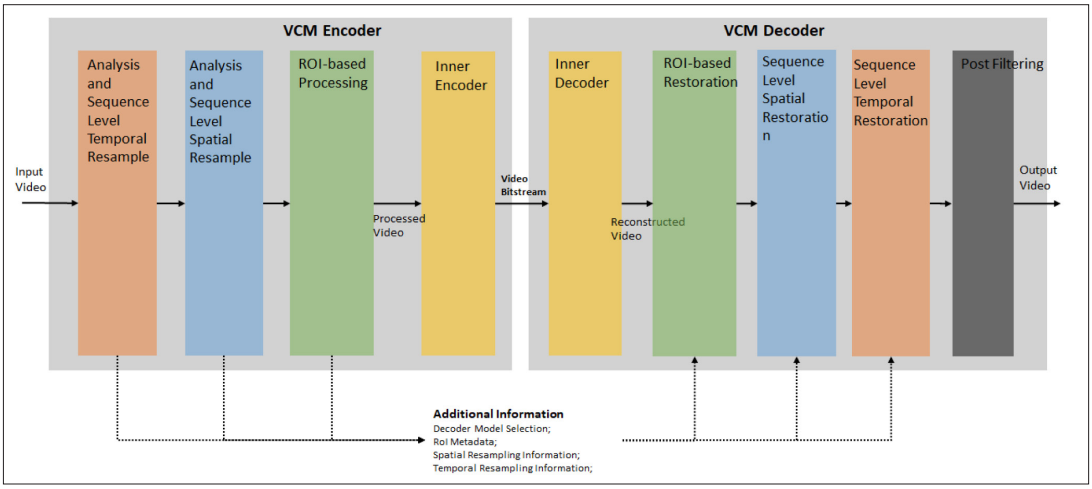
<표 3> CfP Best Performances

Task	Dataset	BD-rate change
Object tracking	TVD (videos)	-57%
Instance segmentation	OpenImages	-51%
	TVD	-57%
Object detection	OpenImages	-47%
	FLIR	-53%
	TVD	-65%
	SFU (videos)	-36%

〈그림 9〉는 제안된 기술을 이용하여 영상을 복원한 예를 보여준다. CfP response 평가 결과를 바탕으로 표준화 진행을 위해 VCM은 〈그림 10〉과 같이 표준 모델 초안을 바탕으로 테스트 모델(Test model)을 만들고 있다. 표준 모델의 초안에 표기된 Inner codec (encoder, decoder)는 AVC, HEVC, VVC 등 기존의 비디오 코딩 기술 또는 그와 유사한 기능을 하는 비디오 부호화 기술을 뜻한다.



<그림 9> CfP 기술을 이용한 영상 복원 결과 비교

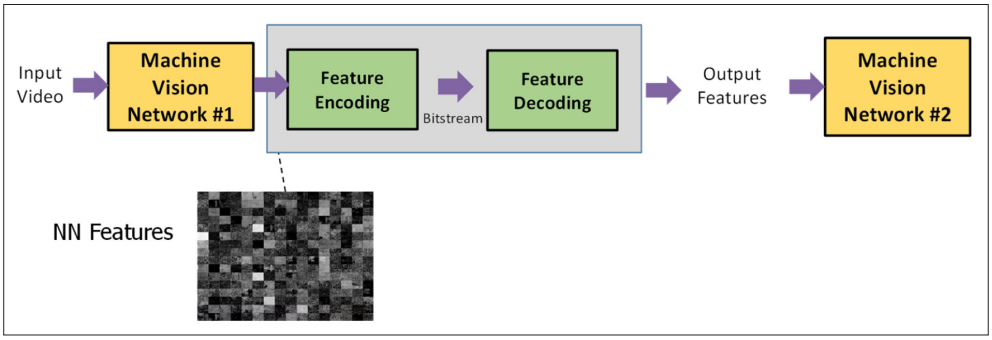


<그림 10> VCM 표준 모델 초안

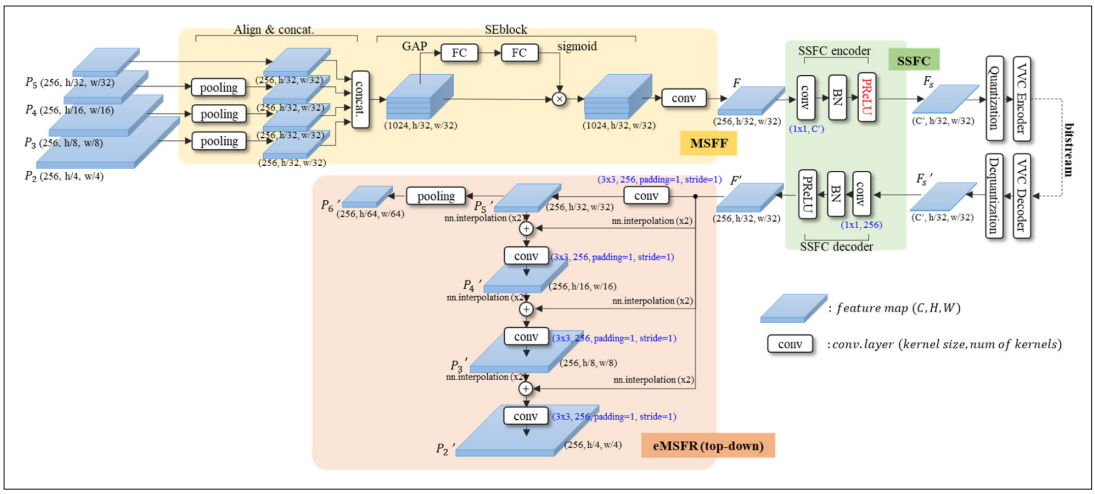
III. VCM Feature Coding Track

본 절에서는 MPEG VCM 그룹의 첫 번째 트랙인 Feature Coding 트랙에 대해서 살펴본다. Feature Coding 트랙은 〈그림 11〉과 같이 영상 또는 비디오로부터 1차적인 머신 비전 네트워크의 출력을 받아서 이를 Feature Map과 데이터를 입력받아 부호화하는 기술을 의미한다.

MPEG VCM Track1에서는 2022년 7월 CfE[6]를 발행하였다. Track1에서는 CfE를 위한 필수 임무로 객체 추적과 객체 분할이 선택되었으며, 선택 임무로 객체 검출이 선택되었다. 한밭대학과 ETRI는 피라미드 형태의 심층신경망의 각 계층의 특징을 입력받아 계층을 결합하는 결합 네트워크(MSFF: Multi-scale feature fusion)



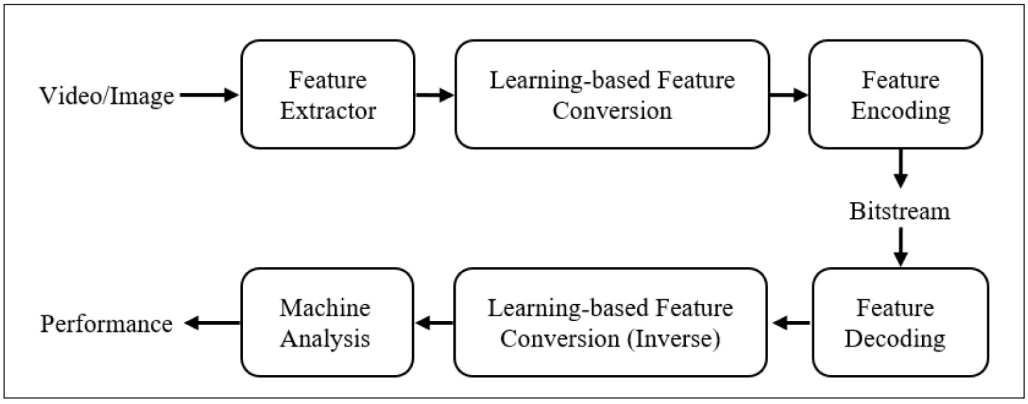
<그림 11> Feature coding Track의 Pipeline



<그림 12> MSFC 기반 VCM Feature codec 구조[27]

와 결합된 특징 맵의 차원을 줄여서 데이터를 줄이는 특징 부호화 네트워크(SSFC: Single-stream feature codec), 복원된 특징을 원래의 피라미드 형태의 심층신경망의 특징 계층으로 복원(MSFR: Multi-scale feature reconstruction)하는 복수 계층 특징 압축(MSFC: Multi-scale feature compression) 프레임워크를 기반하는 파이프라인 구조[27]를 제안하였다. VVC 코덱과의 연동을 위해, SSFC 부호화기에서 출력되는 특징을 패킹, 양자화 후 VVC 부호화기의 입력에 맞는 포맷으로 변경하는 과정이 있으며, VVC 복호화기의 출력을 포맷 변경, 역양자화, 언패킹 과정을 거쳐 SSFC 복호화를 수행 후, 이를 복수 계층의 특징으로 복원하여 임무를 수행할 수 있도록 하였다.

항공대학과 ETRI는 MSFC 프레임워크 기반으로 MSFF후 출력되는 특징을 중요도 순서대로 정렬하여 MSFC-transformed-features를 구성하여 VVC 부호화를 수행하는 방식[28]을 제안하였다. MSFC-transformed-features VVC 부호화기에서 사용하는 비트율에 따라 적응적으로 선택되어 압축하게 되며, 복호화 과정에서 제거된 성능을 예측하는 방식으로 특징을 복원한다. Canon에서는 MSFC로부터 얻어진 특징을 PCA를 이용하여 차원을 줄여서 특징을 압축하



<그림 13> Learning-based feature 압축 구조 프레임워크

<표 4> 객체 추적 및 객체 분할에 대한 결과

제안 문서번호	Object Tracking		Instance Segmentation		Object Detection	
	BD-rate over Video	BD-rate over Feature	BD-Rate over Image	BD-Rate over Feature	BD-Rate over Image	BD-Rate over Feature
m60761	-87.44%	-97.58%	-79.21%	-95.56%	-81.11%	-94.15%
m60788	63.69%	-74.43%	-47.46%	-89.48%	-54.51%	-85.06%
m60799	-80.18%	-97.09%	-93.04%	-98.60%	-94.46%	-98.34%
m60803/ m60802	218.93%	-33.01%	-19.35%	-83.38%	-	-
m60821	-77.40%	-95.84%	-78.11%	-95.84%	-70.39%	-91.14%
m60925	-64.94%	-92.17%	-69.08%	-92.30%	-	-

는 방식[29]을 제안하였으며, 광운대학과 ETRI는 부호화 대상이 되는 특징에 대해서 PCA 기법을 사용하여 기저(basis)와 평균을 사전에 구하여 코덱에 사용하는 방식의 변환 기반의 특징 압축 기술[30]을 제안하였다. Tencent와 Wuhan 대학에서는 Learning-based Feature Conversion 모듈을 사용하여 부호화 효율을 개선하는 방식을 제안하였다[31]. Feature Encoding/Decoding으로 이미지에서 얻어진 특징 부호화 시 Cheng2020 기반의 심층신경망 기반 부호화 네트워크[21]를 사용하였고, 비디오에서 얻어진 특징 부호화 시 VVC 부호화기를 사용하였다.

CfE 제안 결과는 〈표 4〉와 같다[32]. 기존의 VVC 부호화기를 이용해 만든 Feature anchor 대비 객체 추적에서 97.58%, 객체 분할에서 98.60%, 객체 검출에서 98.34%의 성능 개선이 있는 것으로 나타났다. 한편, VVC anchor 대비로는 객체 추적에서 87.44%, 객체 분할에서 93.04%, 객체 검출에서 94.46%의 성능 개선이 있는 것으로 나타났다.

IV. 결론 및 향후 일정

MPEG VCM은 2년간의 회의를 거쳐 Image and Video Coding 트랙의 경우, CfP 기술 제안 평가를

마쳤으며, Feature Coding 트랙의 경우, CfE 기술 검증을 완료하였다. 각 트랙 모두 제안된 기술은 기존의 VVC 부호화를 이용하는 경우 보다 높은 성능을 보여주었다. 앞으로 Feature Coding 트랙의 경우, CfP 제안 기술에 대한 평가 방식 및 추가 데이터셋 등에 대해 논의를 거쳐 2023년 4월 CfP를 발간하고 2023년 10월 회의에서 기술 제안을 받아서 표준화를 시작할 예정이다.

Image and Video Coding 트랙의 경우, 표준 문서 생성을 위한 테스트모델(Test model)을 만들고 테스트 모델 내의 주요 알고리즘의 비교 평가를 위한 실험 절차(CE: Core Experiment)를 진행할 예정이다. 2022년 10월 기준, 다섯개의 CE가 만들어졌다.

- CE 1: RoI based coding methods : 관심영역을 추출하고 이를 통해 각 프레임을 재구성하여 비디오를 처리하는 방식에 대한 비교 실험
- CE 2: neural network based intra frame coding : 심층신경망을 이용한 Intra frame 코딩 알고리즘에 대한 비교 실험
- CE 3: frame level spatial resampling : 각각의 프레임을 resampling을 통해 크기를 줄여서 부호화하고, 복호화시 upsampling하는 방법에 대한 비교 실험
- CE 4 : temporal resampling : 관심영역 또는 타깃의 유무나 그 외 필터링 기술을 사용하여 일정 프레임을 삭제하여 부호화 후, 복호화 시 interpolation 네트워크 등의 방식으로 삭제된 프레임을 채워 넣는 알고리즘에 대한 비교 실험
- CE 5: post filtering: 디코딩된 비디오의 각 프레임을 해상도 및 화질을 높이기 위한 도구에 대한 비교 실험

향후 Image and Video Coding 트랙의 경우, CE 회의를 통해 작업문서(Working Draft)를 만들고, 2023년 10월 위원회 초안(Committee Draft)을 발간하고 2024년 하반기에 표준을 완료할 계획이다.

참 고 문 헌

[1] Use cases and requirements for Video Coding for Machines, ISO/IEC JTC1/SC29/WG2 N190, 2022.04

[2] Evaluation Framework for Video Coding for Machines, ISO/IEC JTC1/SC29/WG2 N193, 2022.04.

[3] Common Test Conditions and Evaluation Methodology for Video Coding for Machines, ISO/IEC JTC1/SC29/WG2 N193, 2022.04.

[4] Call for Evidence for Video Coding for Machines, ISO/IEC JTC1/SC29/WG2 N42, 2021.01.

[5] Call for Proposals for Video Coding for Machines, ISO/IEC JTC 1/SC 29/WG 2 N191, 2022.04

[6] Call for Evidence on Video Coding for Machines, ISO/IEC JTC 1/SC29/WG2 N215, 2022.07.

[7] Marek Domanski et. al., [VCM] Poznan University of Technology Proposal A in response to CfP on Video Coding for Machines, ISO/IEC JTC1/SC29/WG2/m60727, 2022. 10.

[8] Marek Domanski et. al., [VCM] Poznan University of Technology Proposal B in response to CfP on Video Coding for Machines, ISO/IEC JTC1/SC29/WG2/m60728, 2022. 10.

[9] Marek Domanski et. al., [VCM] Poznan University of Technology Proposal C in response to CfP on Video Coding for Machines, ISO/IEC JTC1/SC29/WG2/m60729, 2022. 10.

[10] Sang-Kyun Kim et. al., [VCM] CfP response: Region-of-interest based video coding for machine, ISO/IEC JTC1/SC29/WG2/

m60758, 2022. 10.
[11] S. Wang et. al., [VCM] Video Coding for Machines CfP Response from Alibaba and City University of Hong Kong, ISO/IEC JTC1/SC29/WG2/m60737, 2022. 10.
[12] Hari Kalva et. al., [VCM] Response to VCM CfP from the Florida Atlantic University and OP Solutions, ISO/IEC JTC1/SC29/WG2/m60743, 2022. 10.
[13] Christopher Hollmann et. al., [VCM] Response to Call for Proposals from Ericsson, ISO/IEC JTC1/SC29/WG2/m60757, 2022. 10.
[14] Yegi Lee et. al., [VCM] Response to CfE: Object detection results with the FLIR dataset, ISO/IEC JTC1/SC29/WG11/M56572, 2021.04.
[15] Yegi Lee et. al., [VCM Track2] Response to VCM CfP: Video Coding with machine-attention, ISO/IEC JTC1/SC29/WG2/m60738, 2022. 10.
[16] Jianran Liu et. al., [VCM] Video Coding for Machines CfP Response from Institute of Computing Technology, Chinese Academy of Sciences (CAS-ICT) and China Telecom, ISO/IEC JTC1/SC29/WG2/ m60773, 2022. 10.
[17] Zlzheng Liu et. al., [VCM] Response to VCM Call for Proposals - an EVC based solution, ISO/IEC JTC1/SC29/WG2/ m60779, 2022. 10..
[18] Zlzheng Liu et. al., [VCM] Response to VCM Call for Proposals from Tencent and Wuhan University - an ECM based solution, ISO/IEC JTC1/SC29/WG2/ m60780, 2022. 10..
[19] Ke Jia et. al., [VCM] Response to the CfP on Video Coding for Machine from Zhejiang University, ISO/IEC JTC1/SC29/WG2/m60741, 2022. 10.
[20] Wen Gao et. al., [VCM]Response to VCM Call for Proposals from Tencent - an End-to-end Learning based Solution, ISO/IEC JTC1/SC29/WG2/m60777, 2022. 10.
[21] Cheng, Z., et. al., Learned image compression with discretized gaussian mixture likelihoods and attention modules. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7939-7948), 2020
[22] E. Agustsson, et. al., Scale-space flow for end-to-end optimized video compression, IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2020)
[23] Honglei Zhang et. al., [VCM] Response to the CfP of the VCM by Nokia (A), ISO/IEC JTC1/SC29/WG2/m60753, 2022. 10.
[24] Honglei Zhang et. al., [VCM] Response to the CfP of the VCM by Nokia (B), ISO/IEC JTC1/SC29/WG2/m60754, 2022. 10.
[25] C. Rosewarne, [VCM Track 2] CfP test report, ISO/IEC JTC1/SC29/WG2/m61010, 2022. 10..
[26] CfP response report for Video Coding for Machines, ISO/IEC JTC1/SC29/WG2/N248, 2022. 10.
[27] Heeji Han et. al.,, [VCM] Response from Hanbat National University and ETRI to CfE on Video Coding for Machines, ISO/IEC JTC1/SC29/WG2/m60761, 2022.10.
[28] Yong-Uk Yoon et. al.,, [VCM] Response to VCM CfE: Multi-scale feature compression with QP-adaptive feature channel truncation, ISO/IEC JTC1/SC29/WG2/m60799, 2022.10.
[29] C. Rosewarne, R. Nguyen, [VCM Track 1] Response to CfE on Video Coding for Machine from Canon, ISO/IEC JTC1/SC29/WG2/m60821, 2022.10.
[30] Minhun Lee, et. al.,, [VCM Track 1] Response to CfE: A transformation-based feature map compression method, ISO/IEC JTC1/SC29/WG2/m60788, 2022.10.
[31] Yong Zhang et. al.,, [VCM] Response to VCM Call for Evidence from Tencent and Wuhan University - a Learning-based Feature Compression Framework, ISO/IEC JTC1/SC29/WG2/m60925, 2022.10.
[32] CfE response report for Video Coding for Machines, ISO/IEC JTC 1/SC29/WG2 N247, 2022.10
[33] Hanming Wang, Zijun Wu, Tao Han, Yuan Zhang, [VCM][Response to CfE] An End-to-End Image Feature Compressing Method with Feature Fusion Module, ISO/IEC JTC1/SC29/WG2/m60802, 2022.10.
[34] Hanming Wang, Zijun Wu, Tao Han, Yuan Zhang, [VCM][Response to CfE] An End-to-End Video Feature Compressing Method with Feature Fusion Module, ISO/IEC JTC1/SC29/WG2/m60803, 2022.10.

필 자 소 개

추 현 곤

- 1998년 2월 : 한양대학교 전자공학과 (공학사)
- 2000년 2월 : 한양대학교 전자공학과 (공학석사)
- 2005년 2월 : 한양대학교 전자통신전파공학과 (공학박사)
- 2005년 2월 ~ 현재 : 한국전자통신연구원 책임연구원
- 2017년 9월 ~ 2018년 8월 : Warsaw University of Technology, Poland 방문연구원
- 주관심분야 : Computer vision, 3D imaging and holography, 딥러닝기반 신호처리, 멀티미디어표준화

정 원 식

- 1992년 2월 : 경북대학교 전자공학과 (공학사)
- 1994년 2월 : 경북대학교 대학원 전자공학과 (공학석사)
- 2000년 2월 : 경북대학교 대학원 전자공학과 (공학박사)
- 2000년 5월 ~ 현재 : 한국전자통신연구원 책임연구원
- 주관심분야 : 3DTV 방송 시스템, 라이트필드 이미징, 영상부호화, 딥러닝기반 신호처리, 멀티미디어 표준화

서 정 일

- 1994년 2월 : 경북대학교 전자공학과 (공학사)
- 1996년 2월 : 경북대학교 대학원 전자공학과 (공학석사)
- 2005년 8월 : 경북대학교 대학원 전자공학과 (공학박사)
- 1998년 2월 ~ 2000년 10월 : LG반도체 주임연구원
- 2010년 8월 ~ 2011년 7월 : 영국 Southampton University, ISVR 방문연구원
- 2000년 11월 ~ 현재 : 한국전자통신연구원 실감미디어연구실
- 주관심분야 : 오디오 신호처리, 실감음향, 디지털방송, 멀티미디어 표준화