특집 01

기계를 위한 비디오 부호화 표준개발 동향

윤용욱 · 김동하 · 김재곤 (한국항공대학교)

목 차

1. 서 론

감시 비디오, 스마트 시티, 사물인터넷, 자율주행 등 다양한 응용 분야에서 수집되는 비디오 데이터의 양이 급격히 증가하고 있으며, 인공지능 기술을 바탕으로 기계가 수집된 비디오 데이터를 분석하여 이벤트 또는 객체를 검출하고 인식하여 사람에게 알려주거나 능동적으로 대처하는 머신 비전(machine vision) 기반의 지능형 서비스 또한 지속적으로 증가하고 있다[1].

한편, 컴퓨터 비전 및 영상 처리 분야는 딥러닝 기술을 통해 비약적인 발전을 이뤘지만, 딥러닝 기반의 비디오 압축 분야는 아직 초기 단계에 있다. 방대한 양의 비디오 데이터를 효과적으로 저장/전송하기 위해서 비디오 부호화는 필수적인 요소이다. 일반적으로 비디오 소비의 주체는 사람이기 때문에, 기존 비디오 압축 기술은 압축 대비 가능한 높은 화질의 비디오를 제공하기 위하여 HVS(Human Vision System) 특성을 고려하여 설계된다. 그러나, 지능형 분석 등 기계의 머신 비전 임무(task) 수행을 위해 기존 비디오 코덱을 이용하여 부/복호화된 비디오를 사용할 경우, 지능형 분석을 위한 중요한 정보가 손실되거나 분석에 불필요한 정보가 전송될 수 있는 압축의 비효율성이 존재한다. 이에 따라, MPEG(Moving Picture Experts Group)은 2019년 7월 기계를 위한 비디오 부호화 표준 개발을 위하여 VCM(Video Coding for Machine)이라는 AHG(Ad-Hoc Group)을 결성하고 기계를 위한 보다 효율적인 새로운 비디오 부호화 표준 기술의 탐색 단계를 진행하고 있다. 본 고에서는 VCM의 개요 및 표준화 현황을 소개하고 표준화 주요 이슈 및 현재 논의 중인 제안 기술들에 대해 살펴본다.

2. MPEG VCM 개요

앞서 기술한 바와 같이 사람이 아닌 기계가 소비하는 비디오를 효율적으로 부호화 하기 위한 새로운 비디오 압축 표준 개발을 위해 2019년 7월 제127차 MPEG 회의에서 VCM에 대한 논의가 시작되었고 2019년 10월 표준화를 위한 MPEG VCM AHG이 결성되었다. 표준화의 준비 단계로 VCM 표준을 위한 사용사례(Use Cases) 수집 및 요구사항(requirements) 도출을 진행하였으며, 최근 제135차 MPEG 회의까지 CfE(Call for Evidense) 응답을 포함하여 제안된 다양한 기술에 대한 검토 및 성능평가 프레임워크 등 표준화를 위한 제반 사항들에 대한 논의가 진행되고 있다. 본 장에서는 현재까지 진행된 몇 차례의 회의를 거쳐 논의된 다양한 사용사례, 주요 임무, 요구사항, 성능평가 방법 등에 대해 살펴본다.

2.1 VCM 사용사례(Use case) 및 요구사항(Requirements)

VCM은 기계의 비디오 소비와 관련된 광범위한 AI 응용 시나리오를 다룬다. 즉, 기계에 제공되는 영상 또는 비디오 전송 및 소비와 관련한 대부분의 응용에서 VCM 표준이 적용될 수 있다. MPEG VCM 그룹은 VCM 기술이 사용될 수 있는 대표적인 사용사례를 아래 6가지를 정리하였으며, 각 사용사례는 객체검출, 객체추적, 객체분할과 같은 VCM이 제시하는 16가지의 주요 임무 중 다수의 임무를 포함하고 있다[2].

- Surveillance
- Intelligent Transportation
- Smart City
- Intelligent industry
- Intelligent Content
- Consumer Electronics

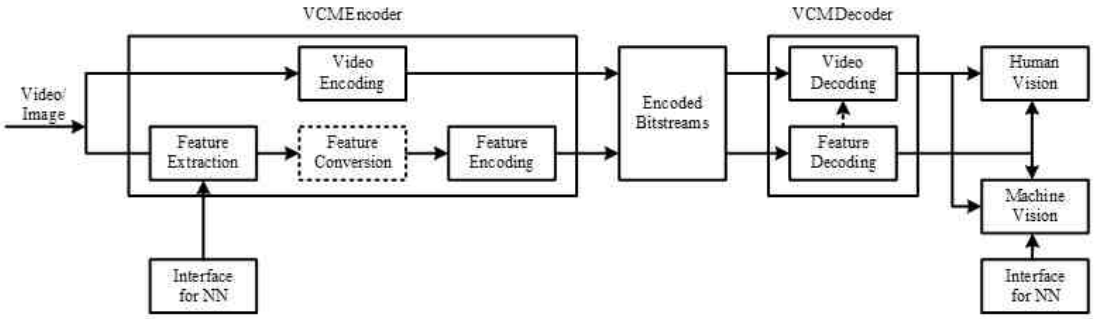
VCM은 위 사용사례를 지원할 수 있도록 VCM 표준이 충족해야 할 아래의 필수 요구사항을 도출하였다.

- 효율적인 압축 성능: 압축된 특징(feature)의 크기는 VVC(Versatile Video Coding)와 같은 최신 비디오 압축 기술을 사용하여 압축된 비디오 스트림보다 작아야 한다.
- 다중 임무에 대한 다양한 수준의 성능: 특정 요구사항으로 인한 임무별 다양한 부호화가 필요한 경우를 위하여 임무별 다양한 수준의 품질을 지원해야 한다.
- 하나 이상의 머신비전 임무 지원: 추출된 특징과 부호화된 비트스트림은 다일 또는 다중 임무에 대해 사용 가능하고 최적화되어야 한다.
- 머신비전 임무 전용 또는 하이브리드 머신/휴먼비전 임무 지원: 하나의 비트스트림은 기계 시작 전용 또는 기계 및 인간 소비에 사용돼야 한다.

위에 기술된 사용사례와 요구사항을 충족시킬 수 있는 VCM 기술의 성능평가를 위해 다수의 회의를 거쳐 평가방법 및 각 세부사항들이 논의되었다.

2.2 MPEG VCM 시스템 구조 및 파이프라인

VCM의 비디오 부호화 목표를 달성하기 위해서는 머신비전 뿐만 아니라 휴먼비전을 위한 기술이 필요하다. 앞서 설명한 사용사례와 요구사항을 반영한 잠정적인 VCM 시스템 구성도는 그림 1과 같다. VCM의 부호화기는 비디오 부호화기나 특징 부호화기 또는 두 부호화기 모두 포함할 수 있다. 입력 비디오를 직접 부호화하지 않고 머신비전 임무 수행에 활용될 수 있는 비디오의 특징을



(그림 1) VCM 시스템 구성도

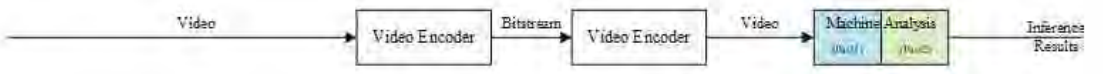
부호화하는 특징 부호화기는 비디오로부터 특징을 추출하는 특징 추출, 특징 변환, 특징 부호화 단계로 구성될 수 있다. VCM 복호화기로부터 복원된 비디오 또는 특징은 머신비전 임무뿐만 아니라 휴먼비전 임무에서도 사용될 수 있다. 또한 VCM 부/복호화기는 특징 추출을 위한 신경망(NN: Neural Network)과 머신비전 임무 수행을 위한 신경망을 포함할 수 있으며 이들 신경망의 압축을 위한 NNR(Neural Network Compression and Representation) 표준과의 인터페이스를 제공할 수도 있다.

그림 1과 같이 잠정적인 VCM의 시스템 구조에 따르면 다양한 VCM 코덱의 활용 시나리오가 파생될 수 있다. MPEG VCM 그룹은 VCM 시스템 구조로부터 파생될 수 있는 다양한 VCM 부/복호화 과정을 보다 명확히 하기 위해 그림 2와 같이 3가지 처리 파이프라인(pipeline)을 제시하고 있다[3]. 그림 2-(가)는 비디오를 입력으로 부/복호화하고 복호화된 비디오를 통해 머신비전 임무를 수행하는 과정을 나타낸 것이다. 여기서 비디오 부/복호화기는 기존 비디오 코덱 또는 딥러닝 기반의 비디오 코덱이 사용될 수 있다. 추후 언급될 VCM의 성능평가를 위해서 VVC를 비디오 코덱으로 사용한 파이프라인 1을 앵커(Anchor)로 사용한다. 즉 Anchor 대비 제안기법의 성능을 비교한다.

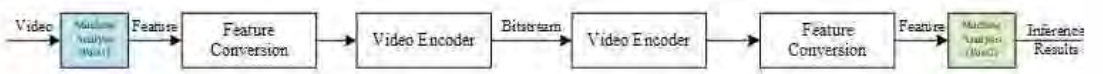
그림 2-(나)는 머신비전 임무를 위한 네트워크가 두 파트로 나뉜다. 파트 1 네트워크는 특징을 추출하는 서브 네트워크이고, 파트 2는 파트 1의 출력인 특징을 사용하여 임무를 수행하는 서브 네트워크이다. 파이프라인 2는 파트 1으로부터 추출되는 특징을 압축하는 구조로, 이 또한 기존 비디오 코덱 또는 딥러닝 기반의 비디오 코덱을 사용하여 압축할 수 있다. 파트 1으로부터 추출되는 특징은 일반적으로 입력 영상/비디오에 비해 큰 데이터 크기를 갖기 때문에 파트 1 네트워크의 어떤 계층에서 특징을 추출하고 압축할지에 대한 이슈가 있다. 또한, 기존 비디오 코덱을 사용하여 압축할 경우, 적절한 입력 형태로 변환될 필요가 있기 때문에 패킹(packing)과 같은 특징 변환 과정이 필요하다.

그림 2-(다)는 머신비전 임무에 추가적으로 휴먼비전에 대한 수요가 있는 경우 머신비전 임무를 위해 부호화된 정보를 활용하여 휴먼비전을 위하여 복호화된 영상/비디오의 품질을 개선할 수 있도록 한다. 파이프라인 3는 두 비트스트림을 제공하기 때문에, 하나의 비트스트림과 비교될 수 있는 압축률을 가져야한다.

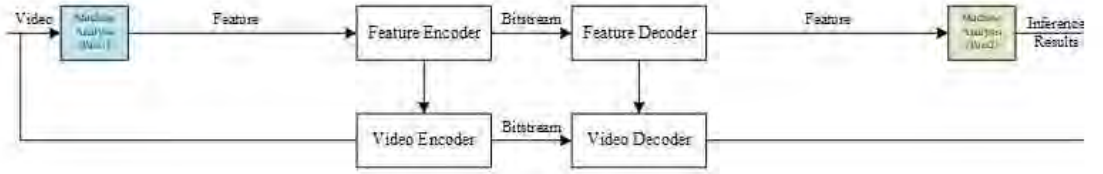
대표적인 3가지 파이프라인이 제시됐지만 아직



(가) 파이프라인 1



(나) 파이프라인 2



(다) 파이프라인 3

(그림 2) VCM의 파이프라인

본격적인 표준화를 위한 참조 모델의 파이프라인이 정해진 것은 아니다. 이 외에도 다양한 파이프라인이 제시되고 있으며, 하이브리드 비전 임무 수행, 다중 임무 처리 등을 고려한 성능평가를 통해 하나의 파이프라인이 결정될 가능성이 있다.

3. VCM 제안 기술의 성능평가

VCM의 표준 기술의 목표인 머신비전 임무 수행 성능을 저하시키지 않으면서 효율적인 압축을 달성하기 위해, 제안할 기술들의 비교를 위한 머신비전 임무, 머신비전 임무 네트워크, 평가 데이터셋(dataset), 평가 측도(metric) 및 기준(anchor)를 정의하였다[3].

3.1 VCM 머신비전 임무 및 네트워크

VCM의 사용사례에서의 머신비전 임무는 매우 다양하며 MPEG VCM 그룹은 이들 16가지의 주요 머신비전 임무 중 핵심 임무 5가지를 아래 표 1과 같이 선정하였다. 이들 핵심 임무에 따라 사용되는 네트워크 및 데이터셋을 표 1과 같이 정의하고 이에 따라 제안기술의 평가를 진행한다.

3.2 평가 데이터셋

딥러닝 학습을 위한 다양한 데이터셋이 있지만, 많은 데이터셋들이 비상업적 사용만 허용하기 때문에, MPEG VCM표준화에서 사용하지 못하는 이슈가 있다. 따라서 MPEG VCM그룹에서는 상업적 사용이 가능하면서 VCM의 성능평가에 적합한 5가지 데이터셋을 선별하였다.

(1) OpenImages-v6[8]: OpenImages-v6는 객체검출 및 객체분할에 사용된다. 평가 데이터셋에는 20,000개 이상의 영상이 있으며 학습 시간을 줄이기 위해 5000개만 선택된다. 객체검출을 위한 5000개의 데이터셋은 객체분할을 위한 데이터셋과 동일하지 않게 시험 데이터셋으로 사용된다.

(2) FLIR[9]: FLIR는 자율주행 및 첨단 운전자 지

〈표 1〉 머신비전 임무에 따른 평가 네트워크 및 데이터셋

Task	Network Architecture	Link	Training Dataset
Object Detection	Faster R-CNN with ResNeXt-101 backbone	[4]	OpenImages-v6 FLIR SFU-HW-Objects TVD
Object Segmentation	Mask R-CNN with ResNeXt-101 backbone	[4]	OpenImages-v6 TVD
Object Tracking	JDE-1088x608	[5]	HiEve-10 TVD
Action Recognition	Slowfast	[6]	HiEve-10
Pose Estimation	HRNet	[7]	HiEve-10

원(ADAS: Autonomous Driving and Advanced Assistance) 시나리오에 사용되는 객체검출에 적합한 RGB 영상과 적외선 영상이 모두 포함되어 있다. 시뮬레이션은 적외선 영상이 저조도 조건에서 RGB 영상보다 더 나은 검출 성능을 달성함을 보여준다. 따라서 적외선 영상만 VCM 평가에 사용된다.

(3) HiEve-10[10]: HiEve 데이터셋에는 많은 수의 포즈(pose), 복잡한 이벤트 동작, 긴 지속기간의 궤적이 포함되며 객체추적, 동작인식, 포즈추정 평가에 적합하다. HiEve 데이터셋 내에서 10개의 비디오로 구성된 서브 데이터셋을 HiEve-10이라 불리고 상업적 목적으로 사용이 가능하며 VCM 평가에 사용된다.

3.3 평가 측도(metric)

VCM의 성능평가를 위해 비전 임무 수행 평가에 사용되는 아래의 측도를 정의하여 부호화 효율을 평가한다.

객체검출과 객체분할 임무의 성능평가를 위해서 IoU(Intersection over Union)에 따른 객체 클래스별 AP(Average Precision)를 평균한 mAP (mean AP)를 사용한다. 객체검출의 경우 객체 영역을 박스 형태로 예측하여 IoU를 계산하고, 객체분할의 경우 예측한 박스 내에서 각 픽셀이 해당 객체에 속하는지 여부를 나타내는 임의의 분할맵으로 IoU를 계산하는 차이가 있다. 동작인지 임무의 성능평가는 각 프레임의 mAP로 평가되고, 포즈예측 임무의 경우 AP로 성능평가에 활용한다.

동영상 객체추적 임무의 성능평가를 위해 MOTA(MOT Accuracy)를 사용한다. MOTA는 객체 영역 박스를 잘 검출하였지만, 해당 객체의 ID를 오인하거나 새로운 객체로 인식한 경우엔 해당 프레임의 추적을 틀린 것으로 간주한다. 따라서, 잘못 추적한 프레임을 제외하여 옳게 추적한 객체의 추적궤도의 프레임만으로 정확도를 계산한다.

압축률을 측정하기 위해 객체검출과 객체분할 임무와 같이 영상에 대한 임무를 수행하는 경우, BPP(Bit Per Pixel)을 식 (1)과 같이 계산하여 압축률을 측정한다. 기존의 HVS에 기반하여 설계된 코덱의 압축성능평가는 인지화질인 PSNR을 측정하여 BD(Bjøntegaard Delta)-PSNR과 BD-rate를 성능평가에 사용하는데 양자화 파라미터를 변화시켜 PSNR을 측정하고 율왜곡(RD: Rate Distortion) 곡선을 사용하여 평균 PSNR 및 비트

율 차이를 계산하여 성능을 비교한다. VCM은 복원된 데이터의 인지화질보다 머신비전 임무의 결과를 우선시하기 때문에, 임무 수행 성능을 측정하여 RP(Rate Performance) 곡선으로 성능을 비교한다.

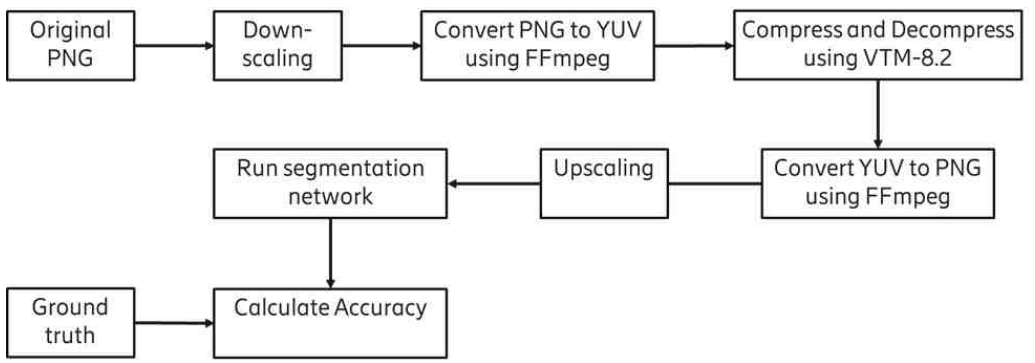
$$BPP = \frac{Total\ bitstream\ size\ in\ bits}{number\ of\ pixels\ in\ the\ source\ image} \quad (1)$$

3.4 평가 기준(anchor) 성능

MPEG VCM은 제안기술의 부호화 효율 평가를 위해 기준이 되는 앵커(anchor)를 생성하였다. 앵커는 그림 2-(가)의 파이프라인을 따르며, 그림 3은 앵커 생성의 상세한 과정을 보여주는 것으로, 영상 스케일링, 컬러 포맷 변환, VVC 부/복호화, 컬러 포맷 역변환, 스케일링, 비전비전 임무 수행, 성능평가의 순으로 진행된다. 이 때, 스케일링은 {100%, 75%, 50%, 25%} 4가지 해상도에 대해 수행하고 비디오 압축은 VVC 참조소프트웨어 VTM(VVC Test Model) 8.2를 사용한다. 서로 다른 비트율에 따른 임무 성능 변화를 측정하기 위해 6개의 QP(Quality Parameter) 값 {22, 27, 32, 37, 42, 47}을 사용하여 부호화하고 복호화된 영상에 대해 머신비전 임무를 수행한 후 그 성능을 측정한다. 앵커 생성을 위한 컬러 포맷 변환 및 VTM의 상세 부호화 환경 설정 등을 정의하고 있다[3].

4. MPEG VCM 제안기술 및 동향

MPEG VCM그룹은 지난 7월 제135차 MPEG 회의까지 다양한 서비스 시나리오, 요구사항, 시스템 구조, 성능평가 방법에 대한 논의가 진행되었으며, 또한 기술탐색 단계에서 다양한 기술들이 제안되었으며 이에 대한 평가 및 검토가 진행되어 왔다. 표 2는 VCM의 주요 제안기술들을 입력 데이터의 종류와 사용하는 코덱에 따라 분류하고 앵커 대비 현재의 성능에 대한 경향을 나타낸 것이다. 본 절에서는 다음과 같이 주요 제안기술들을 (1) 특징 추출, (2) 기존 비디오 코덱을 이용한 특징 부호화, (3) 딥러닝 기반 영상 및 특징 압축, (4) 종단간 학습 기반 영상 압축, (5) 특징 압축을 위한 새로운 접근으로 분류하여 기술한다.



(그림 3) 앵커 생성 파이프라인

〈표 2〉 VCM 주요 제안기술 범주 및 앵커 대비 성능

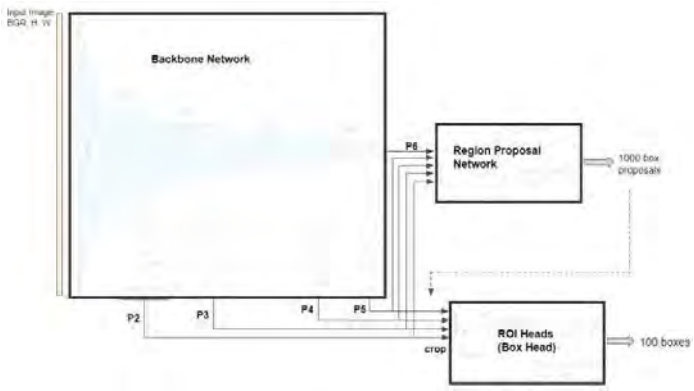
Input \ Codec	Video Codec (HEVC, VVC)	Neural Network based Codec (Autoencoder)		New Codec
		Local training	E2E training	
Image/Video	Anchor	(3) Comparable	(4) Better	-
Feature (1)	(2) Worse	(3) Comparable	Not reported yet	(5) Worse

4.1 특징 추출

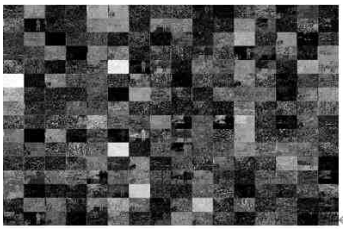
특징을 부호화하기 위해, 먼저 머신비전 네트워크의 중간 계층으로부터 특징을 추출해야 한다. 특징 부호화는 그림 2-(나)의 파이프라인 2를 따르며, 최근 제135차 MPEG 회의까지 많은 기술제안이 있었다[11-24]. 특히 객체검출 임무에 대한 특징맵(feature map) 부호화 기술이 다수 제안되었으며, VCM의 머신비전 임무의 평가 네트워크로 정의된 Faster R-CNN의 백본(backbone) 네트워크로부터 추출된 특징을 부호화한다. 그림 4는 VCM의 객체검출 및 객체분할 임무 수행을 위해 정의된 R-CNN FPN의 구조를 보여준다. 백본 네트워크의 다양한 계층에서 특징맵이 추출될 수 있으며, Stem 특징맵[11-13], 다중 크기를 갖는 전체 특징맵[14], C2 특징맵[15]을 부호화하는 방법이 제안되었다. 일반 영상/비디오보다 절대적인 크기가 큰 특징맵을 압축하기 때문에 압축 효율 측면에서 좋지 않은 결과를 보여주고 있다. 또한 특징맵을 추출하는 계층의 위치에 따라 비전 임무 성능에 미치는 영향이 달라지기 때문에, 압축 효율과 비전 임무 성능을 모두 고려해야하는 이슈가 존재한다.

4.2 비디오 코덱을 이용한 특징 부호화

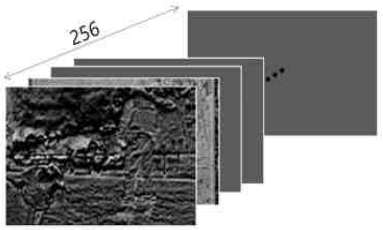
기존 비디오 코덱을 이용하여 중간 계층 특징을 부호화하기 위해서는 특징을 비디오 코덱에 적합한 형태로 변환이 필요하다. 그림 5-(가)와 같이 다채널의 특징을 하나의 프레임으로 변환하여 VVC[20]와 HEVC[21]로 압축하는 방법이 제안되었다. [22, 23]에서는 그림 5-(나)와 같이 다채널



(그림 4) R-CNN FPN 구조

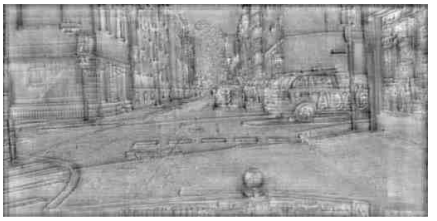


(가) 공간적 패킹

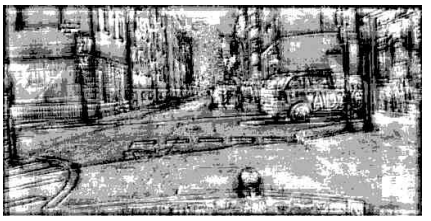


(나) 시간적 패킹

(그림 5) 특징맵 변환의 예



(가) 8비트 표현 특징맵



(나) 2비트 표현 특징맵

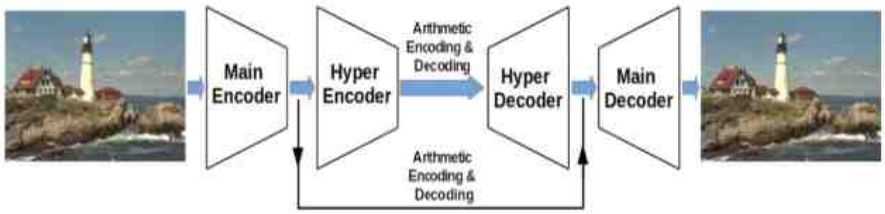
(그림 6) 특징맵 표현의 예

의 특징을 다중 프레임 시퀀스로 변환하여 VVC를 이용한 비디오압축을 수행했다. 하지만 일반 영상/비디오에 비해 특징 데이터의 절대적으로 큰 부분과, HVS에 기반하여 설계된 비디오 코덱의 적합하지않은 입력 형태로 인해 압축 효율이 크게 감소한다. 이에 특징 데이터의 크기를 줄이고자 [16-19]에서는 양자화 방법에 따른 머신비전 임무 성능평가 결과를 제시하였다. [16, 17]은 중간 계층 특징맵을 균등 양자화한 임무 성능을 제시하였다. [18]에서는 정규화 과정을 거쳐 구간을 나눠 구간마다 대표값을 지정하여 적은 비트로 특징맵을 표현함으로써 특징맵의 데이터를 크게 줄이는 방법을 제안하였다. 그림 6은 적은 비트로 표현된 특징맵의 예를 보여준다. 또한, PCA(Principal Component Analysis)를 통해 특징맵의 크기를 효과적으로 줄이는 방법이 제안되었다[19].

중간 계층 특징맵의 압축은 부호화하는 특징맵이 대부분의 비전 임무에서 공통적으로 사용될 수 있다면 압축에 의한 성능 저하가 크지 않다는 가정하에 공통적으로 사용될 수 있는 코덱이 될 수 있다. 하지만 특징을 추출하는 네트워크에 따라 복잡도의 문제가 따를 수 있고, 일반 영상/비디오를 압축한 비트스트림보다 특징맵을 압축한 비트스트림의 크기가 상당히 크다는 문제가 있다.

4.3 딥러닝 기반의 영상 및 특징 압축

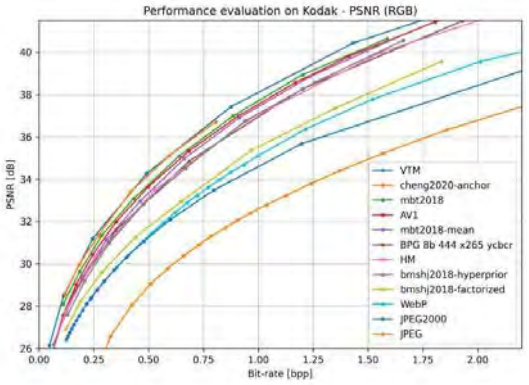
최근 딥러닝 기반의 영상 압축이 활발히 연구되면서, VCM에서도 기술 제안이 이뤄지고 있다. 오토인코더(autoencoer) 프레임워크는 영상의 분포를 추정하는 Hyper Prior 인코더에 의해 영상을 은닉 벡터(latent vector)로 변환하고 비트스트림으로 압축된다. 디코더는 비트스트림을 복호화하여 재구성된 은닉 벡터로부터 영상을 재구성한다. 그림 7은 오토인코더 네트워크의 구조를 보여준다. CompressAI는 최근 몇 년간 연구된 여러 학습 기반의 압축 알고리즘을 포함하고 있으며, 해당 분야의 연구를 위한 플랫폼으로서 활용되고 있다[35]. 그림 8-(가)는 CompressAI에 구현된 학습



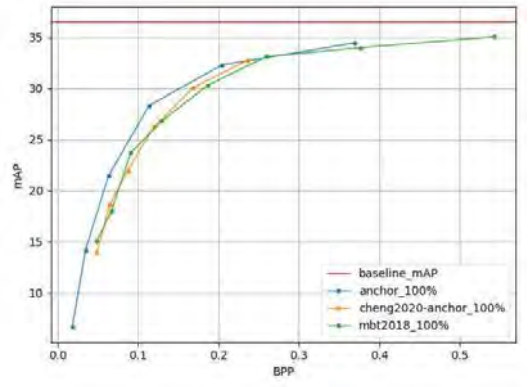
(그림 7) 오토인코더 네트워크의 프레임워크

기반의 영상 압축 알고리즘과 기존 코덱과의 성능 비교를 보여준다.

[25]는 입력 영상에 대해 CompressAI에 구현된 모델 중 VVC와 가장 유사한 성능을 보이는



(가) 압축 기술의 영상 압축성능 비교



(나) 딥러닝 기반 압축의 머신비전 성능 평가

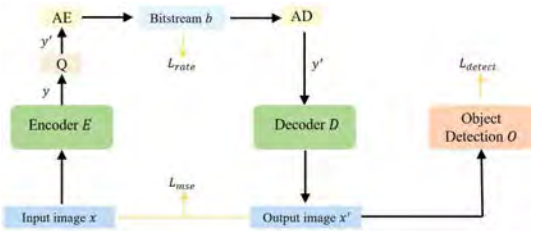
(그림 8) 압축기술에 따른 성능 비교

두 모델 mbt2018과 cheng2020을 이용하여 객체 분할 임무에 대한 성능평가를 하였다. 그림 8-(나)와 같이 앵커 성능과 유사한 임무 수행 성능을 보여줬다.

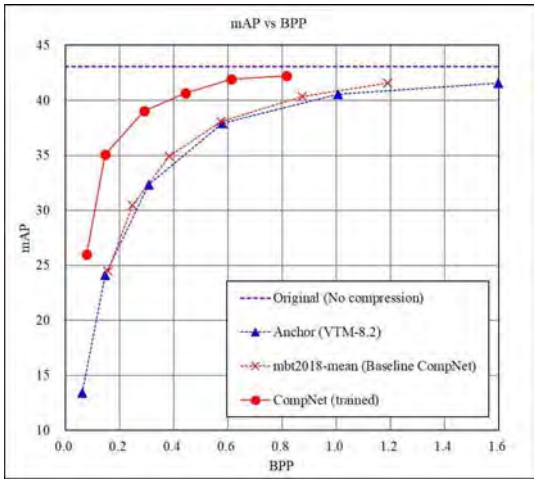
[13]은 특징맵을 입력으로 압축 네트워크를 학습하였다. 특징맵을 기존 비디오 코덱을 이용하여 압축했을 때는 입력 특징맵의 크기와 HVS에 기반하여 설계된 비디오 코덱으로 인해 압축 효율이 떨어졌지만, 특징맵에 기반하여 학습된 압축 네트워크는 앵커와 유사한 성능을 보여줬다. 따라서, 특징맵을 입력으로 압축 네트워크와 머신비전 네트워크를 결합하여 종단간 학습했을 때의 압축 효율 또한 기대되고 있다.

4.4 종단간 학습 기반 영상 압축

HVS 기반으로 설계된 기존 비디오 코덱과 달리, 딥러닝 기반의 압축 네트워크는 머신비전 임무 수행 성능 또한 고려한 학습이 가능하다. 이에 [26-28]에서는 딥러닝 기반의 압축 네트워크와 머신비전 네트워크를 결합하여 종단간 학습을 수행한다. 그림 9는 종단간 학습을 위한 파이프라인의 예를 보여준다. 손실함수를 머신비전 임무에서의 오류와 압축 네트워크로부터의 비트율을 결합하여 구성함으로써 머신비전 임무와 압축률을 함께 최적화한다. 그림 10은 종단간 학습 기반 영상 압축에 의한 성능을 보여준다. 객체검출 임무에서



(그림 9) 종단간 학습을 위한 파이프라인

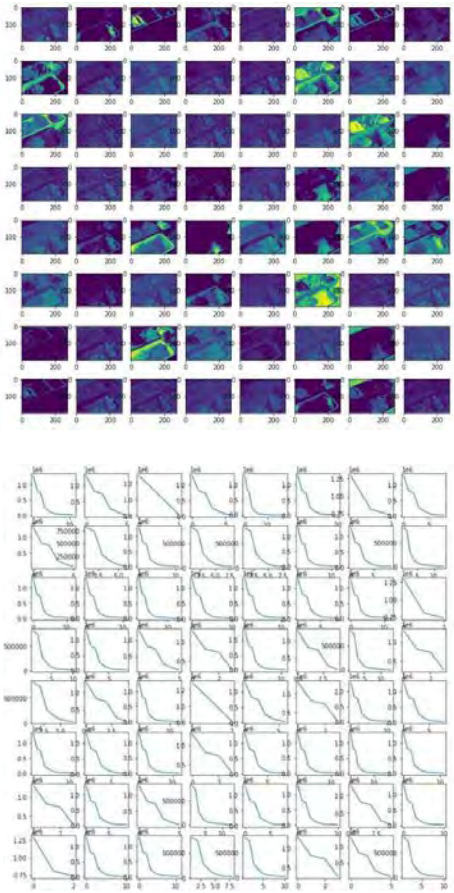


(그림 10) 종단간 학습 기반 영상 압축에 의한 머신비전 압축성능

앵커 성능보다 우수한 성능을 보여주면서 머신비전 임무에 최적화된 딥러닝 기반의 압축 네트워크의 가능성을 보여주었으며 VCM 기술탐색 단계에서 주요 기술로 논의될 것으로 기대된다.

4.5 특징 압축을 위한 새로운 접근

기존 비디오 코덱과 딥러닝 기반의 압축을 사용하지 않고 특징을 압축하는 기술도 몇 차례 제안되었다. [17, 24]에서는 특징을 추출하여 양자화 한 후 zip 부호화와 허프만 부호화를 통해 특징을 압축한다. [32]는 그림 11와 같이 특징맵의 채널간 상관성, 히스토그램 등을 분석하여 채널간



(그림 11) Stem 특징의 채널별 특징맵 및 히스토그램 분석

중복성을 줄이고 이진산술부호화를 통해 특징을 압축한다.

5. CfE(Call for Evidence) 응답

2021년 1월 제133차 MPEG 회의에서 MPEG VCM은 CfE를 공표하고, 제134차 MPEG 회의에서 5개의 CfE 응답을 받았다[28-32]. 본 절에서는 CfE 응답과 관련 기술들을 소개한다.

[28]은 종단간 압축 네트워크를 제안하였다. 압축 네트워크로는 Cheng2020이 사용되며 머신비

전 임무 네트워크는 VCM의 객체검출 임무를 위해 정의된 Faster R-CNN 네트워크를 사용한다. 객체검출 손실과 MSE 오류를 모두 고려하여 손실함수를 구성하여 두 네트워크가 학습된다. 이때, 객체검출 네트워크의 가중치는 모두 고정되고 압축 네트워크의 가중치만 학습된다. 실험결과에 따르면 OpenImage-v6 데이터셋에 대해 앵커 대비 22.80%의 비트율 감소를 보여주면서 높은 성능향상을 보여주었다. [27, 28]에서는 압축 네트워크만 다르고 동일한 접근 방식으로 성능을 평가하였고 이 또한 높은 성능향상을 보여주면서, 객체검출 임무에 대한 기술 검증을 제시하였다.

[29]는 VCM의 객체검출 네트워크를 이용하여 영상의 객체 영역과 배경 영역을 구분한다. 객체 영역과 배경 영역 각각에 대해 VVC 코덱을 사용하여 서로 다른 설정으로 부호화한다. 실험결과에 따르면 FLIR 데이터셋에 대해 앵커 대비 30.76%의 비트율 감소를 보여준다.

[30]은 문맥적(context)으로 구조화된 영상 압축 기법을 제안하였다. 문맥적 구조화의 의미는 높은 수준의 특징과 낮은 수준의 특징을 각각 구조화하여 하나의 비트스트림으로 결합함을 의미한다. 여기서 높은 수준의 특징은 위치, 클래스 ID, 경계 박스와 같이 머신비전 네트워크로부터 출력된 결과를 의미하고, 낮은 수준의 특징은 높은 수준의 특징으로부터 추출된 객체를 의미한다. 머신비전 네트워크로부터 출력된 경계 박스와 같은 높은 수준의 특징만을 압축하여 전송하기 때문에, 비트스트림의 크기가 매우 작고, 머신비전 임무의 성능이 높은 수준을 유지할 수 있다.

[31]은 VVC의 개별 부호화 기술들이 머신비전 임무에 주는 영향을 분석하였다. 일부 부호화 툴(tool)들이 동작하지 않도록 했을 때 객체검출 임무에서 약간의 성능향상을 확인할 수 있었으며, 부호화 복잡도가 크게 감소하였다. [33, 34]에서는 추가적으로 IBC(Intra Block Copy)를 포함한 일부 부호화 기술들의 조합을 테스트하여 머신비전 임무에 대한 성능 변화를 확인하였다.

[32]에서는 머신비전 네트워크로부터 추출되는 중간 계층 특징맵을 기존 비디오 코덱이 아닌 새로운 방법으로 압축한다. Stem계층의 특징맵을 사용하여, 특징맵의 채널간 상관성 및 히스토그램을 분석하여 벡터 양자화와 이진산술부호화를 통해 압축한다. 실험결과에 따르면 앵커 성능에는 미치지는 못하고 있지만, 특징 압축을 위한 새로운 코덱 개발에 대한 여지를 주고 있어 관심 있게 논의되고 있다.

6. 결 론

MPEG VCM은 기계가 소비하는 비디오를 효율적으로 전송하고 저장하기 위한 표준을 위해 2019년 7월 결성되어 최근 135차 MPEG 회의까지 활발히 표준화가 진행되고 있다. 본 고에서는 MPEG VCM 표준화 현황과 제안된 기술 및 주요 이슈들에 대해 기술하였다. VCM은 궁극적으로 수요가 증가하고 있는 머신비전뿐만 아니라 기존의 휴먼비전도 함께 수용할 수 있는 보다 효율적인 머신/휴먼비전 용 비디오 부호화 표준 개발을 목표로 하고 있다. 현재 본격적인 표준화에 앞서 다양한 VCM의 후보 기술들이 제안되는 등 표준 기술탐색이 활발히 진행되고 있다. 특히 딥러닝 기반의 압축 기술 연구가 많은 관심을 받을 것으로 예상되며, 더불어 보다 효율적으로 머신비전과 휴먼비전을 함께 수용할 수 있는 새로운 압축 표준 개발도 진행될 것으로 예상된다. 국내에서도 다수의 기업, 연구소, 대학 등에서 VCM 표준화에 적극 참여하고 있으며, 더욱 증가하고 있는 머신비전 응용을 위한 새로운 비디오 부호화 표준에

대한 표준기술 확보 및 지재권 선점이 기대된다.

참 고 문 헌

[1] 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, 2020.

[2] Y. Zhang, L. Yu, J. Lee, M. Rafie, and S. Liu, “Use cases and requirements for Video Coding for Machines,” ISO/IEC JTC 1/SC 29/WG 2, N00103, Jul. 2021.

[3] M. Rafie, Y. Zhang, and S. Liu, “Evaluation Framework for Video Coding for Machines,” ISO/IEC JTC 1/SC 29/WG 2, N00104, Jul. 2021.

[4] Detectron2, https://github.com/facebookresearch/detectron2

[5] Towards-Realtime-MOT, https://github.com/Zhongdao/Towards-Realtime-MOT

[6] PySlowFast, https://github.com/facebookresearch/SlowFast

[7] Deep High-Resolution Representation Learning from Human Pose Estimation, https://github.com/leoxiaobin/deep-high-resolution- net.pytorch

[8] Open Images Dataset V6, https://storage.googleapis.com/openimages/web/index.html

[9] FLIR Thermal Dataset, https://www.flir.com/oem/adas/adas-dataset-form

[10] HiEve-10, http://humaninevents.org/

[11] S. Wang, Z. Wang, Y. Ye and S. Wang, “Image or video format of feature map compression for object detection,” ISO/IEC JTC 1/SC 29/WG2 input document, M55786, Jan. 2021.

[12] S. Wang, Z. Wang, Y. Ye and S. Wang, “Investigation on feature map layer selection for object detection and compression,” ISO/IEC JTC 1/SC 29/WG2 input document, M55787, Jan. 2021.

[13] J. Do, J. Lee, Y. Kim, S. Y. Jeong, J. Choi, “[VCM] Experimental Results of Feature Compression using CompressAI,” ISO/IEC JTC 1/SC 29/WG2 input document, M56716, Apr. 2021.

[14] S. Kim, M. Jeong, H. Jin, H. Lee, H. Choo, H. Lim, and J. Seo, “[VCM] A report on intermediate feature coding for object detection and segmentation,” ISO/IEC JTC 1/SC 29/WG2 input document, M55243, Oct. 2020.

[15] H. Han, H. Choi, S. Kwak, J. Yun, W.-S. Cheong, and J. Seo, “[VCM] Investigation on feature map channel reordering and compression for object detection,” ISO/IEC JTC 1/SC 29/WG2 input document, M56653, Apr. 2021.

[16] W. Zhang, P. Dong, L. Yang and B. Sun, “On the Feature Map Compression for Object Detection and Segmentation,” ISO/IEC JTC 1/SC 29/WG11 input document, M50984, Oct. 2019.

[17] P. Dong and W. Zhang, “Interframe and Intraframe Compression of Feature maps for Segmentation,” ISO/IEC JTC 1/SC 29/WG11 input document, M51847, Jan.2020.

[18] Y.-U. Yoon, D. Park, S. Chun, and J.-G. Kim, “[VCM] Results of feature conversion for object segmentation,” ISO/IEC JTC 1/SC 29/WG2 input document, M55153, Oct. 2020.

[19] E. Son, and C. Kim, “[VCM] CNN Intermediate feature coding for object detection,” ISO/IEC JTC 1/SC 29/WG11

output document, M54307, Jun. 2020.
[20] Y.-U. Yoon, D. Park, S. Chun, and J.-G. Kim, “[VCM] Results of feature map coding for object segmentation on Cityscapes datasets,” ISO/IEC JTC 1/SC 29/WG2 input document, M55152, Oct. 2020.
[21] S. Kim, M. Jeong, H. Jin, H. Lee, H. Choo, H. Lim, and J. Seo, “[VCM] A report on intermediate feature coding for object detection and segmentation,” ISO/IEC JTC 1/SC 29/WG2 input document, M55243, Oct. 2020.
[22] H. Han, H. Choi, S. Kwak, J. Yun, W.-S. Cheong, and J. Seo, “[VCM] Investigation on feature map channel reordering and compression for object detection,” ISO/IEC JTC 1/SC 29/WG2 input document, M56653, Apr. 2021.
[23] Y.-U. Yoon, D. Kim, and J.-G. Kim, “[VCM] Compression of reordered feature sequences based on channel means for object detection,” ISO/IEC JTC 1/SC 29/WG2 input document, M57497, Jul. 2021.
[24] W. Zhang, P. Dong, L. Yang and B. Sun, “On the Feature Map Compression for Object Detection and Segmentation,” ISO/IEC JTC 1/SC 29/WG11 input document, M50984, Oct. 2019.
[25] Y.-U. Yoon and J.-G. Kim, “[VCM] Evaluation results of object segmentation with deep learning-based image compression,” ISO/IEC JTC 1/SC 29/WG2 input document, M55960, Jan. 2021.
[26] S. Wang, Z. Wang, Y. Ye, and S Wang, “[VCM] End-to-end image compression towards machine vision for object detection,” ISO/IEC JTC 1/SC 29/WG2 input document, M56416, Apr. 2021.
[27] S. Cho, H. Lee, S. Y. Jeong, J. Lee, Y. Kim, J. Do, J. S. Choi, “[VCM] Image compression neural network optimized for object detection,” JTC 1/SC 29/WG2 input document, M56469, Apr. 2021.
[28] B. Zhu, L. Yu, D. Li, and Y. Pan, “[VCM] ZJU response to cfe: deep learning-based compression for machine vision,” ISO/IEC JTC 1/SC 29/WG2 input document, M56445, Apr. 2021.
[29] Y. Lee, S. Kim, K. Yoon, H. Lim, H.-G. Choo, W.-S. Cheong, and J. Seo, “[VCM] Response to CfE: Object detection results with the FLIR dataset,” ISO/IEC JTC 1/SC 29/WG2 input document, M56572, Apr. 2021.
[30] W. Gao, X. Xu, and S. Liu, “[VCM] Response to CfE: Investigation of VVC Codec for Video Coding for Machine,” ISO/IEC JTC 1/SC 29/WG2 input document, M56681, Apr. 2021.
[31] S. Sun, X. Jin, R. Feng, and Z. Chen, “[VCM] Evidence of VCM: Object Detection Evaluation on Semantically Structured Image Compression (SSIC),” ISO/IEC JTC 1/SC 29/WG2 input document, M56722, Apr. 2021.
[32] H.M. Wang, H. Wang, L.C. Wang, Y. Zhang, and X. Chen, “[VCM] Response to Call for Evidence of Video Coding for Machine: K-means and BAC based feature compression,” ISO/IEC JTC 1/SC 29/WG2 input document, M56749, Apr. 2021.
[33] C. Hollmann, P. Wennerstern, J. Strom, and L. Litwic, “[VCM] VCM-based rate-distortion optimization for VVC,” ISO/IEC JTC 1/SC 29/WG2 input document, M56634, Apr. 2021.
[34] S.-P. Wang, C.-C. Lin, C.-L. Lin, T.-H. Li, and Y.-C. Nie, “[VCM] Enable IBC in

VTM8.2 for VTM," ISO/IEC JTC 1/SC 29/WG2 input document, M56792, Apr. 2021.
[35] CompressAI, https://github.com/InterDigitalInc/CompressAI

저 자 약 력

윤 용 욱
이메일 : yuyoon@kau.kr

- 2017년 한국항공대학교 항공전자및정보공학과 (학사)
- 2019년 한국항공대학교 항공전자및정보공학과 (석사)
- 2019년~현재 한국항공대학교 항공전자및정보공학과 박사과정
- 관심분야 : 비디오 코딩, 딥러닝, 멀티미디어 응용

김 동 하
이메일 : donghakim@kau.kr

- 2021년 한국항공대학교 항공전자및정보공학과 (학사)
- 2021년~현재 한국항공대학교 항공전자및정보공학과 석사과정
- 관심분야 : 비디오 코딩, 딥러닝

김 재 곤
이메일 : jgkim@kau.ac.kr

- 1990년 2월 경북대학교 전자공학과 학사
- 1992년 2월 KAIST 전기 및 전자공학과 석사
- 2005년 2월 KAIST 전기 및 전자공학과 박사
- 1992년 3월~2007년 2월 한국전자통신연구원(ETRI) 선임연구원/팀장
- 2001년 9월~2002년 7월 Columbia University 연구원
- 2015년 12월~2016년 1월 UC San Diego, Visiting Scholar
- 2007년 9월~현재 한국항공대학교 항공전자정보공학부 교수
- 관심분야 : 비디오 부호화 표준, 비디오 신호처리, Immersive Video, Deep Learning