VCM 의 MSFC 기반 특징 압축을 위한 Min-Max 시그널링을 제외한 특징맵 생성 기법

김동하, 윤용욱, 김재곤
한국항공대학교
{donghakim, yuyoon}@kau.kr, jgkim@kau.ac.kr

A Feature Map Generation Method for MSFC-Based Feature Compression without Min-Max Signaling in VCM

Dong-Ha Kim, Yong-Uk Yoon, and Jae-Gon Kim
Korea Aerospace University

요 약

MPEG-VCM(Video Coding for Machines)에서는 머신비전(machine vision) 네트워크의 백본(backbone)에서 추출된 이미지/비디오 특징 압축을 위한 표준화를 진행하고 있다. 현재 VCM 표준기술 탐색 과정에서 가장 좋은 압축 성능을 보이는 MSFC(Multi-Scale Feature compression) 기반 압축 네트워크 모델은 추출된 멀티-스케일 특징을 단일-스케일 특징으로 변환하여 특징맵으로 구성하고 이를 VVC 로 압축한다. 본 논문에서는 MSFC 기반 압축 모델에서 Min-Max 값 시그널링을 제외한 최소-최대(Min-Max) 정규화를 포함한 개선된 특징맵 생성 기법을 제시한다. 즉, 제안기법은 VCM 디코더에서의 특징맵 복원을 위한 Min-Max 값을 학습 기반으로 생성함으로써 Min-Max 시그널링의 비트 오버헤드 절감뿐만 아니라 별도의 시그널링 기제를 생략한 보다 단순한 전송 비트스트림 구성을 가능하게 한다. 실험결과 제안기법은 이미지 앵커(Anchor) 대비 BPP-mAP 성능에서 83.24% BD-rate 이득을 보이며, 이는 기존 MSFC 보다 1.74%정도 다소 떨어지지만 별도의 Min-Max 시그널링 없이도 기존의 성능을 유지할 수 있음을 보인다.

1. 서론

딥러닝 기반의 머신 비전(machine vision) 임무를 수행하는 기계를 위한 VCM(Video Coding for Machines) 표준은 입력되는 영상이나 머신비전 백본(backbone) 네트워크에서 추출된 특징(feature)을 압축하는 두 트랙으로 표준화가 진행되고 있다[1]. 올 10 월 MPEG-VCM 회의에서 입력 영상으로부터 추출한 특징을 압축하는 Track 1 과 입력 영상을 직접 압축하는 Track 2 에 대한 각각 CfE(Call for Evidence)와 CfP(Call for Proposal)에 대한 응답기술을 검증하는 단계를 진행중이다[2]. 그림 1 은 Track 1 에서 가장 좋은 압축 대비 임무(task) 성능을 보이는 VVC(Versatile Video Coding) 코덱을 활용한 MSFC(Multi-Scale Feature compression) 기반의 특징 압축 모델이다[3]. 백본에서 추출한 멀티-스케일의 특징을 MSFF(Multi-Scale Feature Fusion) 모듈을 이용해 단일-스케일의 특징으로 정렬 및 채널을 축소하고, 채널이 축소된 특징을 특징맵으로 변환하여 VVC 로 압축 복원한 후 MSFR(Multi-Scale Feature Reconstruction) 통해 멀티-스케일의 특징으로 재구성한다.

특징맵을 VVC 로 부호화 할 때, 32-비트 텐서(tensor)의 특징맵을 8-비트의 영상으로 변환하기 위한 최소-최대 정규화(Min-Max normalization)를 포함한다. 그리고 부호화된 특징맵을 복호화 하기 위해선 정규화에 사용한 Min/Max 값은 부호화한 특징맵과 함께 전송되어 멀티-스케일 특징 복원에 사용된다. 따라서, MSFC 기반의 멀티-스케일 특징 압축을

위해서는 Min/Max 값의 시그널링 비트 오버헤드(overhead)뿐만 아니라 기존의 VVC 비트스트림 이외에 별도의 Min-Max 시그널링을 위한 비트스트림 설계가 요구된다.

따라서, 본 논문에서는 MSFC 기반의 멀티-스케일 특징 압축에서 요구되는 Min-Max 값 시그널링을 제외한 특징맵 생성 기법을 제안한다.

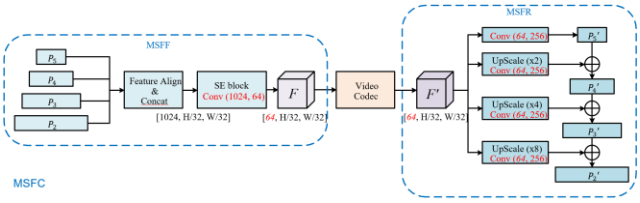


그림 1. VVC 를 활용한 MSFC 기반 특징 압축 모델 구조

2. MSFC 기반 Min-Max 생성 모듈

그림 1 의 MSFF 에서 추출 한 64 개의 채널로 구성된 특징을 VVC 로 방법은 그림 2 와 같다. 특징 채널들을 래스터 배열(raster scan) 순서로 한 프레임에 나열하여 하나의 특징맵으로 구성하고 Min-Max 정규화를 통해 0 ~ 1 사이의 값으로 매핑된 8-비트 특징맵으로 변환한다. 32-비트 텐서값인 Min/Max 값은 복호화된 특징맵을 32-비트 특징으로 복원하는 역정규화에 사용된다. 복원된 64 채널로 구성된 특징은 MSFR 에 입력되어 멀티-스케일의 특징으로 재구성된다.

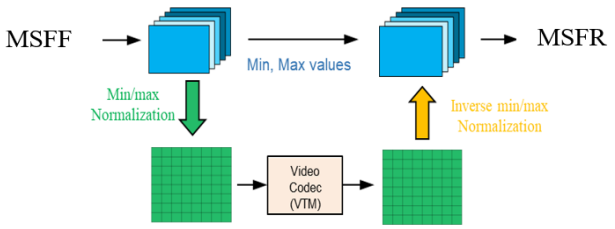


그림 2. 특징맵 부호화 과정

그림 3 은 제안기법인 MSFC 기반 Min-Max 생성 모델을 나타낸 것이다. MSFF 에서 추출한 특징맵을 Min-Max 정규화를 통해 0 ~ 1 의 값으로 매핑하고, VVC 로 압축 복원된 64 채널의 특징을 Min/Max 생성 모듈에 입력한다. Min/Max 생성 모듈은 0 ~ 1 로 매핑된 64 개 채널에서 GAP(Global Average Pooling)을 통해 각 채널의 평균값을 추출한다. 추출된 64 개의 평균값으로부터 3 개의 FC(Fully-Connected) 계층을 통해 2 개의 32-비트 텐서값을 Min/Max 값으로 사용하여 복원된 64 채널의 특징에 역정규화를 수행한다.

MSFC 기반의 압축 모델은 머신비전 네트워크를 동결하고 VVC 를 제외한 채 머신비전 네트워크의 임무손실(task loss)만을 사용하여 학습한다. Min-Max 생성 모듈은 MSFR 전단에 두고 MSFC 와 함께 학습하여 역정규화 할 수 있는 최적의 Min/Max 32-비트 텐서값을 출력하도록 학습한다.

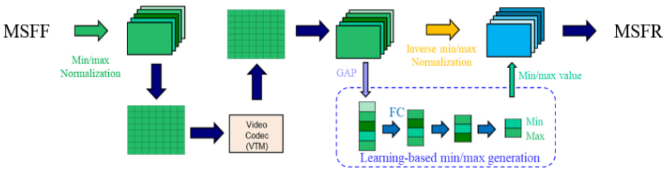
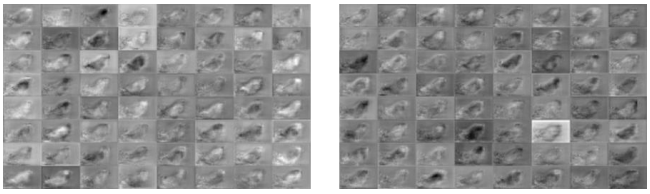


그림 3. MSFC 기반 Min-Max 생성 모듈

그림 4 는 동일한 입력 영상에 대한 기존 MSFC 압축 모델의 MSFF 에서 추출한 특징맵과 MSFC 기반 Min-Max 생성 모듈을 적용했을 때의 MSFC 압축 모델의 MSFF 에서 추출한 특징맵을 비교한 것이다. 두 특징맵은 서로 다른 채널을 강조하지만 유사한 형태를 보인다.



가. 기존 MSFC 압축 모델 나. 제안 MSFC 압축 모델

그림 4. Min-Max 생성 모듈 유무에 따른 특징맵 비교

3. 실험결과

머신비전 네트워크는 객체감지를 수행하는 Detectron2 Faster R-CNN 으로[4] Min-Max 생성 모듈은 MSFC 의 MSFR 전단에 두고 전체 머신비전 네트워크를 동결한 채 MSFC 와 함께 학습하였다. COCO train2017[5] 데이터셋을 학습에 사용하고, VCM 의 CTC(Common Test Condition)[6]에 따라 OpenImage V6[7] 검증 이미지 5 천장을 사용하여 학습한 모델을 평가했다. 배치 크기는 2, 30,000 번 반복과 0.0005 학습율로 학습하였다.

표 1 은 Min-Max 시그널링 유무에 따른 MSFC 기반 압축 모델의 압축 대비 머신비전 임무 성능을 나타낸 것이다.

표 1. Min-Max 시그널링 유무에 MSFC 기반 압축 모델 성능

	Existing MSFC-based		Proposed MSFC-based	
QP	BPP	mAP	BPP	mAP
Uncom.	-	78.55	-	78.11
QP22	0.177	78.710	0.160	78.166
QP27	0.117	78.554	0.101	78.250
QP32	0.064	78.811	0.051	78.027
QP37	0.028	76.675	0.021	73.136
QP42	0.012	67.665	0.009	62.275
QP47	0.005	53.202	0.003	45.912
BD-rate	-85.98%		-83.24%	

BD-rate 은 VCM 의 이미지 앵커(anchor) 대비 BPP-mAP 성능을 나타낸 것이다. 본 논문에서 제안한 Min-Max 생성 모듈을 포함한 MSFC 기반의 압축 모델이 1.74%의 BD-rate 손실을 보인다. 그림 5 는 표 1 결과의 성능 곡선이다.

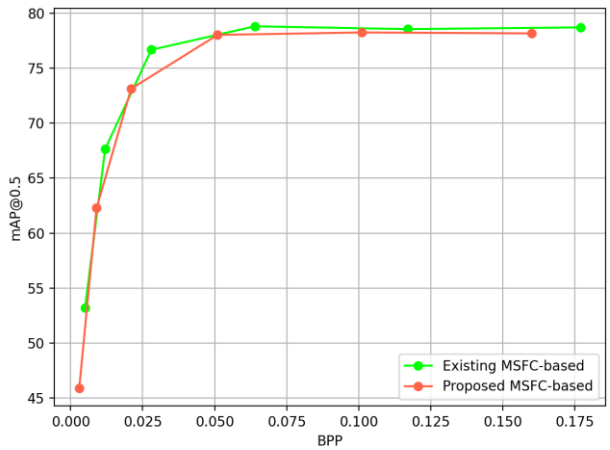


그림 5. Min-Max 시그널링 유무에 MSFC 기반 압축 모델 성능

4. 결론

본 논문에서는 MSFC 기반 압축 모델에서 Min-Max 값 시그널링을 제외한 최소-최대(Min-Max) 정규화를 포함한 개선된 특징맵 생성 기법을 제시한다. 즉, 제안기법은 VCM 디코더에서의 특징맵 복원을 위한 Min-Max 값을 학습 기반으로 생성함으로써 Min-Max 시그널링의 비트 오버헤드 절감뿐만 아니라 별도의 시그널링 기제를 생략한 보다 단순한 전송 비트스트림 구성을 가능하게 한다.

본 논문은 VCM 에서 탐색중인 MSFC 기반 특징 압축 모델에서 Min/Max 생성 모듈을 추가하여 Min-Max 정규화를 위한 Min-Max 값 시스널링을 생략한 특징맵 생성 기법을 제안하였다. 제안기법은 이미지 앵커(anchor) 대비 BPP-mAP 성능에서 83.24% BD-rate 이득을 보여 기존 MSFC 보다 1.74% 정도 다소 BD-rate 성능이 떨어지지만 거의 성능을 유지할 수 있음을 보였다. 본 제안기법은 특징맵 복원을 위한 Min-Max 시그널링 제거함으로써 시그널링의 비트 오버헤드 감소뿐만 아니라 별도의 시그널링 기제를 생략한 보다 단순한 전송 비트스트림 구성을 가능하게 한다. 향후 Min-Max 생성 모델 및 학습방법의 개선이 가능할 것으로 보인다.

Acknowledgement

본 논문은 이 논문은 산업통상자원부 국가표준기술원에서 시행한 국가표준기술력향상사업의 지원을 받아 수행된 연구임 (20011687, 머신러닝 기반 자율주행영상 특징정보 표현 국제표준 개발).

참 고 문 헌(Reference)

[1] “Use cases and Requirements for Video Coding for Machines” ISO/IEC JTC 1/SC 29/WG 2, N00190, Online, Apr. 2022.

[2] “Evaluation Framework for Video coding for Machines,” ISO/IEC JTC 1/SC 29/WG 2 N00162, Online, Jan. 2022.

[3] D. kim, Y. Yoon, J. Kim, J. Lee, Y. Kim, and S. Jeong “[VCM Track1] Compression of FPN Multi-Scale Features for Object Detection Using VVC,” ISO/IEC JTC 1/SC 29/WG2, m59562, Apr. 2022.

[4] “detectron2”, [Online]. Available: https://github.com/facebookresearch/ detectron2

[5] “COCO dataset”, 2017, [Online]. Available: https://cocodataset.org/#download

[6] “Common Test Conditions and Evaluation Methodology for Video Coding for Machines,” ISO/IEC JTC 1/SC 29/WG 2 N00231, Online, Jul. 2022.

[7] “OpenImages”, V6, [Online]. Available: https://storage.googleapis.com/openimages/web/index.html