

VCM 을 위한 SAD 기반 특징맵 시퀀스 재배열

김동하, 윤용욱, 김재곤

한국항공대학교

donghakim@kau.kr, yuyoon@kau.kr, jgkim@kau.ac.kr

SAD-Based Reordering of Feature Map Sequence for VCM

Dong-Ha Kim, Yong-Uk Yoon, and Jae-Gon Kim

Korea Aerospace University

요 약

최근 머신비전 임무(machine vision task)를 위해 기계에 소비되는 비디오가 증가하면서 MPEG 은 기계를 위한 비디오 부호화 표준으로 VCM(Video Coding for Machine) 표준화 진행하고 있다. VCM 은 기계분석 네트워크에 입력되는 비디오 또는 특징(feature)을 부/복호화하여 압축 대비 임무 수행 정확도를 평가한다. 본 논문은 기계분석 네트워크에서 추출한 특징 데이터를 기존의 비디오 코덱을 사용하여 부/복호화를 진행할 때, 각 채널의 특징맵을 SAD(Sum of Absolute Difference) 기반으로 재배열하는 방법을 제안한다. 제안기법은 VCM 의 기준성능(anchor)에는 미치지 못하지만, 채널 재배열하지 않은 특징을 비디오 코덱으로 부호화 할 때 보다 개선된 성능을 보인다.

1. 서론

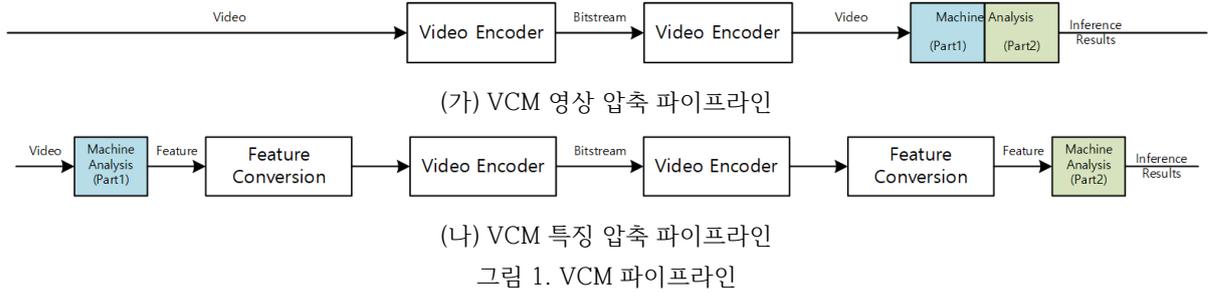
최근 비디오 데이터를 소비하는 딥러닝 기반의 기계들이 증가하면서 기계에 입력되는 비디오를 보다 효과적으로 부호화 하는 방법이 요구되고 있다. 이러한 머신 비전(machine vision)에 사용되는 비디오를 압축하기 위해 HVS(Human Visual System)을 목적으로 하는 기존의 코덱을 사용할 경우 비효율적이고 머신 비전에 필요한 정보가 손실될 수도 있다. 따라서, MPEG 은 머신 비전을 위한 새로운 비디오 부호화 표준인 VCM(Video Coding for Machine)의 표준화를 진행하고 있다. [1]

현재 VCM 은 머신비전 임무에 사용되는 비디오를 직접 압축하는 방법과 입력된 비디오로부터 추출된 특징(feature)을 압축하는 방법으로 나누어 표준화를 진행중이다. 본 논문에서는 기존의 코덱을 사용하여 비디오 특징을 압축하는 방법을 제안한다. 본 논문의 구성은 2 절에서 현재 VCM 의 특징의 특성과 특징 압축의 일반적인 구조를 설명하고, 3 절에서 제안하는 특징 압축 방법을 기술한다.

2. VCM 특징 압축 방법

현재 VCM 이 제시하는 머신비전에 사용되는 비디오를 압축하는 파이프라인(pipeline)은 그림 1 과 같다. 그림 1-(가)는 머신비전 네트워크에 입력되는 비디오를 압축하는 방법으로, VVC(Versatile Video Coding)를 사용한 경우의 압축 대비 임무 수행 정확도 성능을 VCM 의 기준 성능인 anchor 로 정의한다. [2] 그림 1-(나)는 VCM 이 제시하는 특징 압축 파이프라인으로, 입력된 비디오로부터 특징 데이터를 추출하는 머신분석(machine analysis) 파트 1 과 특징 데이터를 분석하는 머신분석 파트 2 로 구성된다. 이때, 추출한 특징 데이터는 머신분석 파트 1 입력 비디오 대비 방대한 크기의 특징 데이터를 포함한다.

VCM 의 특징 데이터 부호화 시 비디오 코덱을 사용할 경우, 각 채널의 특징 데이터를 특징맵으로 구성하고 전체 채널의 특징맵들을 하나의 시퀀스로 만든다. 즉, 시퀀스는 특징을 구성하는 각 채널을 하나의 프레임으로 사용하는 특징맵들의 시퀀스이다. 본 논문은 추출한 특징을 특징맵 시퀀스로 변환하여 비디오 코덱을 활용할 때, 특징맵 시퀀스의 부호화 효율을 높이고자 SAD(Sum of Absolute Difference)기반 특징맵 시퀀스 재배열 기법을 제안한다.



3. SAD 기반 특징맵 시퀀스 재배열

SAD 기반 특징맵 시퀀스 재배열 기법은 머신분석 파트 1로부터 추출한 다수의 채널로 구성된 특징을 비디오 코덱에 사용할 수 있는 특징맵 시퀀스로 변환하기 전에, 채널의 순서를 SAD 기반으로 재배열하는 것이다.

$$SAD = \sum_{i=0}^n |x_i - y_i| \quad (1)$$

x_i 는 현재 채널 특징맵의 i 번째 픽셀의 특징 값, y_i 는 특정 채널 특징맵의 i 번째 픽셀의 특징 값을 의미하고, n 은 채널을 구성하는 전체 픽셀 수를 의미한다. SAD 기반 특징맵 시퀀스 재배열 기법은 채널의 특징맵의 평균값이 가장 낮은 값을 가진 채널의 특징맵을 첫 번째 특징맵으로 배치한다. 다음으로, 이미 배치된 채널의 특징맵과 현재 채널의 특징맵을 제외한 나머지 채널의 특징맵과 SAD의 값을 계산하여 SAD 값이 가장 작은 채널의 특징맵을 다음 특징맵으로 배치한다. 이를 반복하여 전체 특징맵을 재배열한다. 특징맵은 복호화 후 다시 기존의 채널 순서로 배치되어야 함으로 재배열 정보가 전송되어야 한다.

추출된 특징은 32 비트의 부동소수점 값을 갖기 때문에, 최소 최대 정규화를 통해 특징 채널을 8 비트의 영상 데이터 형태로 변환하고 최종적으로 특징맵 시퀀스로 변환한다.

$$\hat{x} = \frac{x - \min(x)}{\max(x) - \min(x)} \ll n \quad (2)$$

x 는 변환될 특징, n 은 변환하고자 하는 영상의 표현 비트 수, 그리고 \hat{x} 은 변환된 특징을 의미한다. 이때, 정규화에 사용된 최소, 최대값은 복호화에 사용될 수 있도록 전송되어야 한다. 그림 2는 제안기법 유무에 따른 특징맵 시퀀스의 비교를 예시로 나타낸 것이다.



(가) 특징맵 시퀀스



(나) SAD 기반 재정렬된 특징맵 시퀀스

그림 2. 제안기법 유무에 따른 특징맵 시퀀스 예시

4. 실험결과

제안하는 SAD 기반 특징 시퀀스 재배열 방법은 VCM의 객체탐지(Object Detection) 임무의 평가 네트워크로 정의된 Detectron2의 R-CNN X101-FPN의 백본 네트워크로부터 추출된 stem 특징을 사용하여 성능을 평가했다[3]. 이때, 데이터셋은 COCO validation 2017을 사용했다. 특징맵 시퀀스는 VVenC(Versatile Video Encoder) [4]를 사용하여 압축했다. 추출된 stem 특징은 특징맵 시퀀스로 변환되어 VVenC에 입력되며 QP {22, 27, 32, 37}에 대해 임의접근(RA: Random Access) 부호화하였다.

표 1은 재배열 기법 적용 유무에 따른 BPP(Bit Per Pixel) 대비 객체탐지 임무 수행 정확도인 mAP(mean Average Precision) 성능을 보여준다. 이때, 기존의 특징맵 평균값을 기반으로 오름차순 재배열한 평균(Mean)[5]기반 재배열 기법도 같이 표기하여 현재 제안기법과 성능 비교하였다. 표 2는 제안기법과 VCM의 앵커의 성능 비교를 나타낸다. 제안기법을 적용할 경우가 적용하지 않은 경우보다 더 높은 mAP 성능이 도출되는 것을 확인할 수 있었다. 또한, 평균기반 재배열 기법 보다 SAD 기반 재배열 기법에서 더 높은 mAP 성능을 얻었다. 그림 3과 그림 4는 표 1과 표 2의 실험 결과값을 그래프로 나타낸 것이다.

표 1. Stem 특징에 대한 실험결과(BPP vs. mAP)

QP	Non Reordering		Mean Reordering		SAD Reordering	
	BPP	mAP	BPP	mAP	BPP	mAP
22	4.36	38.21	4.35	38.62	4.28	38.78
27	1.74	28.58	1.68	29.35	1.63	29.73
32	0.68	15.51	0.64	15.88	0.61	16.29
37	0.29	5.70	0.26	5.80	0.25	6.16

표 2. Anchor 와 제안기법 비교(BPP vs. mAP)

QP	Anchor		SAD reorder	
	BPP	mAP	BPP	mAP
22	1.60	41.56	4.28	38.78
27	1.01	40.54	1.63	29.73
32	0.58	37.93	0.61	16.29
37	0.31	32.55	0.25	6.16

재배열 보다 개선된 성능을 보였다. 하지만, 비디오 특징 데이터의 부호화는 주어진 입력 비디오 크기 보다 절대적으로 증가된 데이터 크기로 인하여 VCM 앵커의 성능에는 미치지 못한다. 따라서 머신비전의 임무수행 성능을 유지하면서 특징 데이터의 크기를 줄인다면 BPP 대비 mAP 성능을 개선할 수 있을 것으로 예상된다.

Acknowledgement

이 논문은 산업통상자원부 국가표준기술원에서 시행한 국가표준기술력향상사업[20011687]의 지원을 받아 수행된 연구임.

참고 문헌(References)

- [1] 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, m57671, Jul. 2021.
- [2] M. Rafie, Y. Zhang, and S. Liu, "Evaluation Framework for Video Coding for Machines," ISO/IEC JTC 1/SC 29/WG 2, m57658, Jul. 2021.
- [3] detectron2, <https://github.com/facebookresearch/detectron2>
- [4] VVenC, <https://github.com/fraunhoferhhi/vvenc>
- [5] Y. Yoon, D. Kim, and J. Kim, "Compression of reordered feature sequences based on channel means for object detection," ISO/IEC JTC 1/SC 29/WG 2, m57497, Jul. 2021.

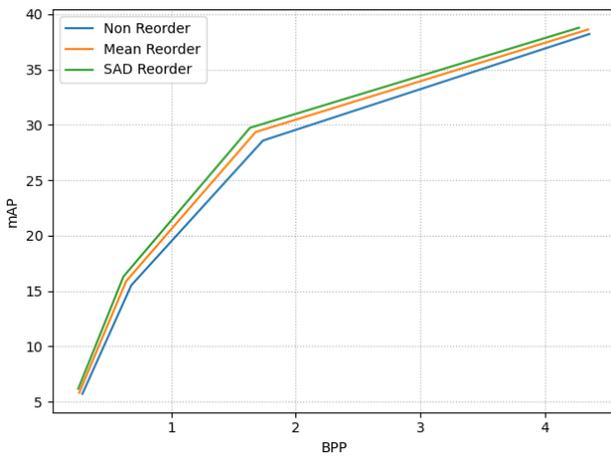


그림 3. 특징맵 시퀀스의 압축 대비 임무수행 정확도 성능 비교

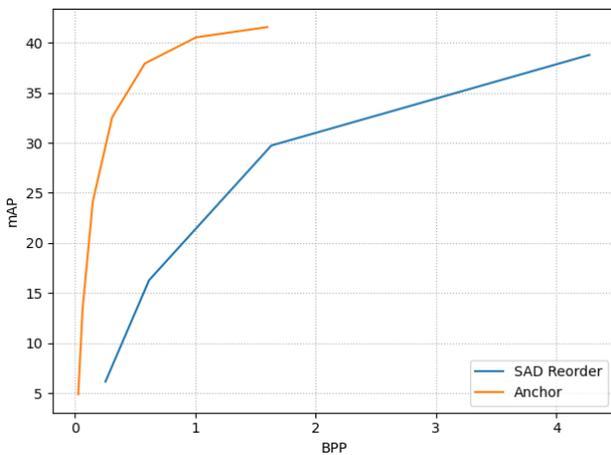


그림 4. Anchor 대비 제안기법 성능 비교

5. 결론

본 논문은 머신비전 네트워크로 추출한 특징맵을 SAD 기반으로 재배열한 특징맵 시퀀스를 VVC 로 부호화하는 기법을 제시하였다. 제안기법은 stem 특징의 각 채널의 특징맵 시퀀스 부호화에서 재배열을 하지 않은 경우뿐만 아니라 기존에 평균기반의 특징맵