VCM 의 객체추적을 위한 다중스케일 특징 압축 기법

윤용욱, 한규웅, 김동하, 김재곤
한국항공대학교
{yuyoon, woong2614, donghakim}@kau.kr, jgkim@kau.ac.kr

A Method of Multi-Scale Feature Compression for Object Tracking in VCM

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

요 약

최근 인공지능 기술을 바탕으로 지능형 분석을 수행하는 기계를 위한 비디오 부호화 기술의 필요성이 요구되면서, MPEG 에서는 VCM(Video Coding for Machines) 표준화를 시작하였다. VCM 에서는 기계를 위한 비디오/이미지 압축 또는 비디오/이미지 특징 압축을 위한 다양한 방법이 제시되고 있다. 본 논문에서는 객체추적(object tracking)을 위한 머신비전(machine vision) 네트워크에서 추출되는 다중스케일(multi-scale) 특징의 효율적인 압축 기법을 제시한다. 제안기법은 다중스케일 특징을 단일스케일(single-scale) 특징으로 차원을 축소하여 형성된 특징 시퀀스를 최신 비디오 코덱 표준인 VVC(Versatile Video Coding)를 사용하여 압축한다. 제안기법은 VCM 에서 제시하는 기준(anchor) 대비 89.65%의 BD-rate 부호화 성능향상을 보인다.

1. 서론

최근 인공지능 기술을 바탕으로 기계가 수집한 비디오 데이터를 분석하여 객체/이벤트를 검출 및 추적하여 사용자에게 알려주거나 능동적으로 대처할 수 있는 머신비전(machine vision) 응용프로그램이 지속적으로 증가하면서, 방대한 양의 비디오 데이터를 처리 및 전송할 수 있는 표준 기술이 요구되고 있다. 기존 HVS(Human Vision System) 특성을 고려하여 설계된 비디오 압축 기술을 기계의 머신비전 임무 수행을 위해 사용할 경우, 지능형 분석을 위한 중요한 정보가 손실되거나 분석에 불필요한 정보가 전송되면서 비디오가 비효율적으로 압축될 수 있다. 이에 사람이 아닌 기계가 소비하는 비디오를 효율적으로 부호화하기 위한 새로운 비디오 압축 표준 개발을 위해 MPEG 에서는 MPEG-VCM 그룹을 결성하여 표준화를 활발히 진행하고 있다[1].

VCM 의 부호화기는 비디오 부호화기나 특징(feature) 부호화기 또는 두 부호화기를 모두 포함할 수 있음에 따라, 다양한 조합의 시스템 구조가 될 수 있다. 이러한 잠정적인 시스템 구조를 바탕으로 크게 다음 두 가지 접근방식에 따른 트랙(track)을 나누어 표준화를 진행하고 있다.

- Track 1: Feature extraction and compression tasks
- Track 2: Image and video compression tasks

트랙 1 은 기계 임무를 위한 특징 추출과 압축 방법을 포함하고 있으며, 일반적으로 머신비전 네트워크로부터 출력되는 특징을 압축하는 기술이 논의되고 있다. 트랙 2 는 이미지 또는 비디오를 입력으로 머신비전 성능을 고려한 딥러닝 기반의 종단간 압축하는 기술과 관심영역을 구분하여 압축하는 기술 등이 논의되고 있다. 최근 10 월 MPEG-VCM 회의에서는 트랙 1 에 대해 CfE(Call for Evidence), 트랙 2 는 CfP(Call for Proposal)

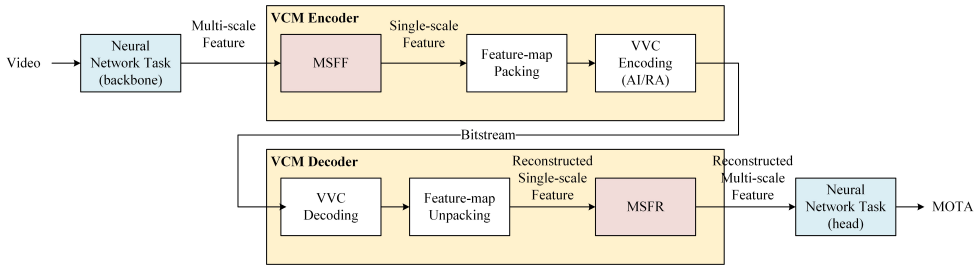

그림 1. 제안기법의 다중스케일 특징 압축 파이프라인

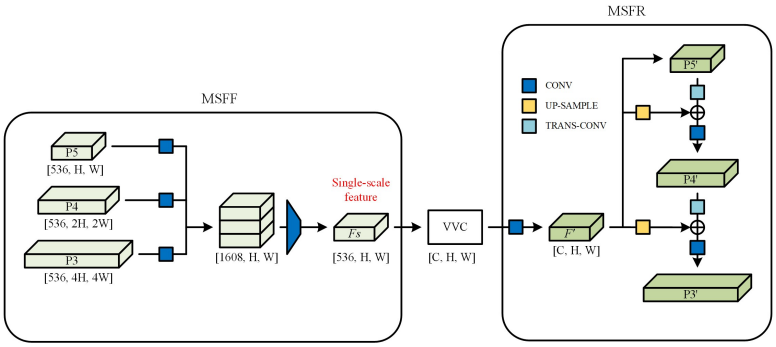

그림 2. 제안기법의 다중스케일 특징 압축 네트워크 구조

응답 기술 검토를 진행하였다[2], [3].

본 논문에서는 VCM 의 트랙 1 에서 논의되고 있는 특징 압축을 위해, 객체추적(object tracking) 임무(task)에 대한 효율적인 다중스케일(multi-scale) 특징 비디오 압축 기법을 제시한다.

2. 제안기법의 구조 및 네트워크 모델

MPEG-VCM 에서는 제안기술들의 성능평가를 위한 머신비전 임무, 머신비전 임무 네트워크, 평가 데이터셋, 평가측도 및 기준(anchor)을 정의한다. VCM 의 다양한 사용사례 만큼 머신비전 임무 또한 다양하여 MPEG-VCM 은 다양한 머신비전 임무 중 핵심임무 3 가지를 선정하였다. 또한 이들 핵심임무에 따라 사용되는 참조 네트워크 및 평가 데이터셋을 표 1 과 같이 정의하여 이에 따라 제안기술의 평가를 진행한다[4].

표 1. VCM 머신비전 임무 평가환경

Task	Network architectures	Eval. datasets
Instances Segmentation	Mask R-CNN X101-FPN	OpenImages-v6 TVD
Object Detection	Faster R-CNN X101-FPN	OpenImages-v6 FLIR, TVD
Object Tracking	JDE-1088x608	TVD

각 임무에 대한 참조 네트워크는 서로 다른 백본(backbone) 네트워크로 구성되어 있기 때문에, 추출되는 다중스케일 특징의 크기가 다르다. 객체추적을 위해 사용하는 참조 네트워크는 JDE-1088x608 은 DarkNet-53 백본 네트워크와 YOLO-v3 헤드(head) 네트워크로 구성되어 있으며, 백본 네트워크로부터 크기가 다른 세 개의 다중스케일 특징이 추출된다[5]. 표 2 는 백본 네트워크로부터 추출되는 다중스케일 특징의 크기를 보여준다. 객체추적 네트워크는 입력을 비디오로 하기 때문에, 비디오 시퀀스의 각 프레임에 대하여 다중스케일 특징이 추출된다. 따라서 추출되는 특징 또한 특징 시퀀스 형태로 변환될 수 있다.

표 2. 객체추적 네트워크 백본의 출력 특징 크기

	Feature size (ch/h/w)
Input resolution	1088x608
P5	536/19/34
P4	536/38/68
P3	536/76/136

2.1. 제안기법의 프레임워크

본 논문에서는 핵심임무 중 객체추적에 대해 효율적인 특징 압축 기법을 제시한다. 그림 1 과 그림 2 는 각각 제안기법의 파이프라인과 네트워크 구조를 포함한 VCM 인코더와 디코더를 보여준다. 제안하는 네트워크 모델은 다중스케일 특징을 단일스케일 특징으로 변환하는 MSFF(Muli-Scale Feature Fusion)와 단일스케일 특징을 다시 다중스케일 특징으로 복원하는 MSFR(Multi-Scale Feature Reconstruction)로 구성되어 있다.

MSFF 는 참조 네트워크의 백본으로부터 추출된 다중스케일

특징을 단일스케일(single-scale) 특징으로 변환한다. 각 스케일 특징은 표 2 와 같이 동일한 채널 수를 갖고, 서로 다른 높이와 너비를 갖기 때문에, 크기 변환을 통해 단일스케일 특징으로 변환한다. 제안기법은 기존 비디오 압축 코덱 VVC 를 이용하여 특징을 압축하기 때문에, MSFF 로부터 변환된 단일스케일 특징은 이미지 형태의 특징맵(feature map)으로 패킹(packing)되어야 한다. 그림 3 은 패킹된 특징의 예를 보여준다. 패킹된 특징은 입력 비디오 시퀀스의 한 프레임에 대한 특징으로, 모든 프레임에 대해 추출된 특징들은 변환과정을 통해 특징 프레임으로 변환되고, 최종적으로 특징 시퀀스를 형성한다. 형성된 특징 시퀀스는 VVC 로 압축되며, 임의접근(RA: Random Access) 환경으로 압축된다. VVC 를 통해 재구성된 단일스케일 특징은 MSFR 모듈에서 특징 복원 단계를 거쳐, 원래 크기의 다중스케일 특징으로 복원되고, 참조 네트워크의 헤드 네트워크로 입력되어 머신비전 임무를 수행한다.

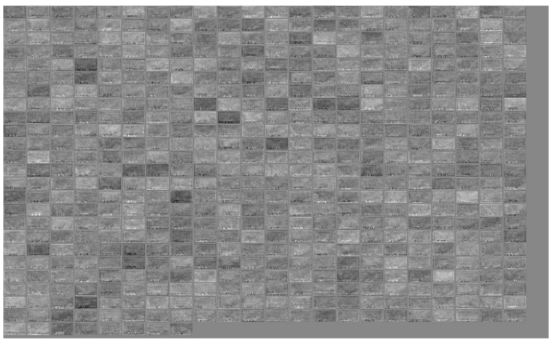

그림 4. 특징맵의 예

2.2. 제안기법의 네트워크 모델

전술한 바와 같이 제안하는 네트워크는 MSFF 와 MSFR 모듈로 구성된다. MSFF 는 표 2 와 같이 서로 다른 크기의 특징을 접합하기 위해 각 스케일 특징에 대해 합성곱(convolutional) 계층(layer)을 구성하여 다운 스케일링(down-scaling)을 수행한다. 접합된 특징은 합성곱 계층 거쳐 채널이 감소되어 단일스케일 특징으로 변환된다. MSFR 은 단일스케일 특징을 업스케일링(up-scaling) 및 탑-다운(top-down) 방식의 구조를 이용하여 원래 크기의 다중스케일 특징으로 복원한다. 제안기법의 사용된 계층은 모두 단일 계층으로 구성되어 있으며, 활성 함수 및 배치 정규화는 사용되지 않는다.

3. 실험결과

제안기법의 성능평가를 위해 VCM 의 CTC(Common Test Condition) 및 평가 프레임워크에 따랐다[4], [6]. 객체추적 임무에 대한 성능평가가 이뤄졌으며, 제안기법의 네트워크 모델은 객체추적 임무의 참조 네트워크와 함께 학습되었다. 이때, 참조 네트워크의 파라미터는 고정되어 제안기법의 네트워크 모델의 파라미터만 학습된다. 특징 압축을 위해 VVC 의 참조 소프트웨어 VTM12.0 을 사용하였으며 [7], RA 환경에서 부호화 하였다.

제안기법은 두 가지 기준 성능과 비교하여 평가되었다. 첫번째 기준성능은 이미지 기준성능(image anchor)이며, 이는 입력 이미지를 VVC 로 압축하여 재구성된 이미지에 대해 머신비전 임무를 수행한 결과이다. 두번째로, 참조 네트워크의 백본 네트워크로부터 추출된 특징을 VVC 로 압축하여 재구성된 특징에 대해 머신비전 임무를 수행한 결과인 특징 기준성능(feature anchor)이다. 표 3 은 제안기법의 실험결과를 보여준다. 제안기법은 이미지 기준성능 대비 47.43% BD-rate 향상을 보여주고, 특징 기준성능 대비 89.65% BD-rate 향상을 보여준다.

표 3. 제안기법의 실험결과

QP	Bitrate (kbps)	MOTA	BD-rate (vs. Image anchor)	BD-rate (vs. feature anchor)
QP #1	1956.819	50.01	-47.43%	-89.65%
QP #2	1401.751	50.01		
QP #3	1018.806	50.00		
QP #4	650.706	48.27		
QP #5	460.849	45.53		
QP #6	354.940	42.23		

4. 결론

제안기법은 VCM 에서의 효율적인 특징 압축을 위한 다중스케일 특징 압축 기법을 제시했다. 제안기법은 이미지 기준 성능 대비 47.43%, 특징 기준 성능 대비 89.65%의 BD-rate 부호화 효율 향상을 보여줬다. 최근 MPEG-VCM 에서 다중스케일 특징 압축 기반의 많은 제안기술이 발표되면서 [8]-[11], 높은 압축 성능을 보여주고 있다. 가능성을 보여주는 만큼 앞으로 더 활발한 표준화 활동이 진행될 것으로 보이며, 다중스케일 특징 압축을 통해 머신비전 뿐만 아니라 휴먼비전을 고려한 네트워크 설계도 이뤄질 것으로 기대된다.

Acknowledgement

본 논문은 2022 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No. 2017-0-00486).

참 고 문 헌(References)

[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] "Call for Proposals on Video Coding for Machines," ISO/IEC JTC 1/SC 29/WG 2, N00220, Jul. 2022.

[3] "Call for Evidence on Video Coding for Machines," ISO/IEC JTC 1/SC 29/WG 2, N00215, Jul. 2022.

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

[5] Z. Wang, *et al*, "Towards Real-Time Multi-Object Tracking," The European Conference on Computer Vision (ECCV) 2020, Jul. 2020.

[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] VTM 12.0 Available at: https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM/-/tags

[8] 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.

[9] D. Kim, Y.-U. Yoon, J.-G. Kim, J. Lee, S.-Y. Jeong, "[VCM-Track1] Performance of the Enhanced MSFC with Bottom-Up MSFF," ISO/IEC JTC 1/SC 29/WG2, M60130, Jul. 2022.

[10] H. Han, M. Choi, H. Choi, S.-H. Jung, S. Kwak, H.-G. Choo, W.-S. Cheong, and J. Seo, "[VCM Track 1] Experimental results of multi-scale feature compression anchor generation on m59576," ISO/IEC JTC 1/SC 29/WG2, M60130, Jul. 2022.

[11] Y.-U. Yoon, D. Kim, J.-G. Kim, J. Lee, S. Jeong, and Y. Kim, "[VCM] Response to VCM CfE: Multi-scale feature compression with QP-adaptive feature channel truncation," ISO/IEC JTC 1/SC 29/WG2, M60799, Oct. 2022.