

다해상도 피쳐 맵 압축 손상 보상 방법

권나성¹ 이민훈¹ 최한솔¹ 박승진¹ 오승준¹ 김연희² 이주영² 정세윤² *심동규¹
 광운대학교¹ 한국전자통신연구원²
 {naseongkwon, minhun, hschoi95, promo, sjoh, *dgsim}@kw.ac.kr
 {kimyounhee, leeje1003, jsy}@etri.re.kr

Compression Error Compensation Method for Multi-Resolution Feature Map

Naseong Kwon¹ Minhun Lee¹ Hansol Choi¹ Seungjin Park¹ Seung-Jun Oh¹ Younhee Kim²
 Jooyoung Lee² SeYoon Jeong² *Donggyu Sim¹
 Kwangwoon University¹ Electronics and Telecommunications Research Institute²

요 약

본 논문에서는 다해상도 피라미드 피쳐 맵 압축 손상 보상 방법을 제안한다. 본 논문에서 제안하는 방법은 패킹된 C-레이어 피쳐 맵을 비디오 코덱으로 압축할 때, 저해상도 계층의 원본 피쳐 맵과 복원된 피쳐 맵 간의 차분 값을 구해 이를 고해상도 계층의 피쳐 맵에 더함으로써 부호화 과정에서 발생하는 오차를 보상하는 방법이다. 본 논문에서 제안하는 방법의 성능을 평가하기 위하여 OpenImageV6 데이터셋 중 1000 장에 대해 객체 검출 성능을 평가하였다. 본 논문에서 제안하는 피쳐 맵 압축 방법은 C-레이어 피쳐 맵 압축 방법 대비 bpp와 mAP의 BD-rate 관점에서 35.10%의 성능 향상을 보인다.

1. 서론

최근 지능형 산업, 스마트 시티, 사물인터넷 (IoT: Internet Of Things), 자율주행 자동차 등의 다양한 분야에 인공지능 기술을 접목시키는 연구 및 응용이 증가하고 있다. 다양한 분야에서 생성되는 방대한 양의 영상을 처리하기 위해 사람 대신 기계를 통한 영상 분석을 제공하는 서비스가 요구되고 있다. 이를 위해 HEVC (High Efficiency Video Coding)[1], VVC (Versatile Video Coding)[2]와 같은 영상 압축 표준 기술을 기반으로 하는 비디오 코덱을 사용하여 기계 간 영상 또는 피쳐 맵을 전송 및 저장할 수 있다. 이와 같은 영상 압축 표준 기술은 사람의 인지 시각 특성을 고려한 기술로, 객체 탐지, 객체 분할 등과 같은 비전 임무를 수행하는 기계의 특성을 고려한 압축 기술에 대한 요구가 대두되었다. 이에 따라 기계에서 소비될 영상을 압축하는 기술의 표준을 수립하기 위해 MPEG (Moving Picture Expert Group)은 2019년 7월 제 127차 회의에서 VCM (Video Coding for Machines) 기술의 표준화를 시작하여 최근까지 논의를

진행하고 있다.

VCM 분야에서는 영상을 압축하는 비디오 코덱 방식과 영상으로부터 추출한 피쳐 맵을 압축하는 피쳐 맵 코덱 방식에 대한 연구가 활발히 이루어지고 있다[3, 4].

VCM 표준화 과정에서 논의되고 있는 임무수행 네트워크 중에서, 객체 탐지 등의 특정 기계 시각 임무에서 객체 크기에 대한 불변성을 고려한 특징 추출 네트워크인 피쳐 피라미드 네트워크 (Feature Pyramid Network, FPN)[5]는 피라미드 구조로, 다해상도의 피라미드 피쳐 맵을 출력한다. 다해상도 피라미드 피쳐 맵을 추출하는 과정에서 고해상도 레벨의 피쳐 맵을 생성하기 위해 저해상도 레벨의 피쳐 맵과의 합 연산이 수행된다. 이 때, 부호화된 저해상도 레벨의 피쳐 맵을 고해상도 레벨의 피쳐 맵에 더함으로써 압축 손상이 전파되는 문제가 발생한다.

본 논문에서는 부호화 과정에서 발생하는 압축 손상이

다해상도 피라미드 피쳐 맵을 생성하는 과정에서 전파되는 오차를 보상하는 방법을 제안한다.

본 논문의 구성은 다음과 같다. 2 절에서는 제안하는 피쳐 맵 압축 손상 보상 방법에 대해 설명하고, 3 절에서는 제안하는 압축 손상 보상 방법의 성능을 기존 피쳐 맵 압축 방법과의 성능을 비교하여 제안하는 방법의 성능을 비교한다. 마지막으로 4 절에서는 본 논문에 대한 결론을 맺는다.

2. 본론

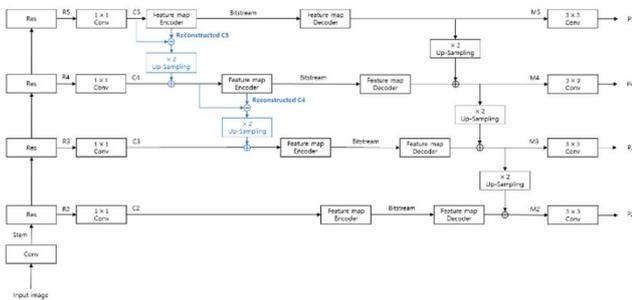


그림 1. C-레이어 피쳐 맵 압축 손상 보상 구조도

다해상도 피라미드 피쳐 맵을 추출하는 과정에서 고해상도 계층의 피쳐 맵을 생성하기 위해 저해상도 계층의 피쳐 맵과의 합 연산을 수행할 때, 저해상도 계층의 피쳐 맵의 압축 손상이 고해상도 계층의 피쳐 맵에 전파되는 문제점이 있다. 이 문제를 해결하기 위해, 본 논문에서는 고해상도 계층의 피쳐 맵을 구하는 과정에서 저해상도 계층의 피쳐 맵의 압축 손상을 보상하여 압축 손상에 의한 오차 전파를 감소시키는 방법을 제안한다. 그림 1 은 FPN 구조에서 압축 손상을 보상하는 과정에 대한 블록도이다. 부호화를 수행하기 전의 N 레벨의 C-레이어 피쳐 맵과 부/복호화된 N 레벨의 C-레이어 피쳐 맵 간의 차분 값을 구한 후, 이를 업 샘플링한 값을 (N-1) 레벨의 C-레이어 피쳐 맵에 더하여 부/복호화 과정에서 발생하는 오차를 보상한다. 업 샘플링 방법은 피쳐 피라미드 네트워크 상에서 서로 다른 계층의 피쳐 맵 간 합 연산을 수행할 때 사용된 방법과 동일한 최근접 이웃 화소 보간법을 사용하였다.

C-레이어 피쳐 맵 부호화 과정은 다음과 같다. Faster R-CNN X101-FPN[6] 네트워크에 영상을 입력하여 중간 단계에서 추출된 C-레이어 피쳐 맵을 비디오 부호화기로 압축시키기 위해 10-bit 로 Min-Max 양자화를 수행하였다. 각 계층마다 256 개 채널로 구성되어 있는 양자화된 피쳐 맵을 비디오 부호화기를 사용하여 효율적으로 압축시키기 위해 동일 계층의 피쳐 맵끼리 하나의 영상으로 패키징을 수행하였다. 비디오 부호화기의 입력은 YUV 형식이므로, 패키징된 PNG 형식 영상을 YUV 4:0:0 형식으로

변환한 후, 비디오 부호화기를 사용하여 압축하였다.

복호화 과정은 비디오 복호화기를 사용하여 영상을 복호화한 후, YUV 형식 영상을 PNG 형식으로 변환을 수행하였다. 복호화된 영상을 역 패키징하여 피쳐 맵으로 분리한 후, 역양자화를 수행하였다. 복원된 피쳐 맵을 사용하여 기계 시각 임무 분석 결과를 도출하였다.

3. 실험 방법 및 실험 결과

본 논문에서 제안하는 방법의 성능을 평가하기 위하여 객체 검출 임무를 수행하는 Detectron2[7] 프레임워크를 사용하였다. 피쳐 맵을 압축하기 위해 VVC 참조 소프트웨어인 VTM-12.0[8]을 사용하였고 VVC CTC (Common Test Condition)의 AI (All Intra)조건에서 실험을 진행하였다. 양자화 파라미터 (Quantization Parameter, QP)는 41, 43, 45, 47, 49 을 사용하였다. 실험 영상은 OpenImageV6[9] 중 1000 장의 정지 영상을 사용하여 실험을 진행하였다. YUV 형식 영상과 PNG 형식 영상 변환을 위해 FFmpeg[10]을 사용하였다. 성능 평가를 위해 비교할 C-레이어 피쳐 맵 압축 방법은 [11]의 방법을 차용하여 C-레이어 피쳐 맵을 각 계층마다 하나의 영상으로 패키징하여 10-bit Min-Max 양자화한 후, VTM12.0 을 통해 부/복호화된 피쳐 맵을 사용해 기계 시각 임무 성능을 평가하는 것으로 정의하였다. C-레이어 피쳐 맵 압축 방법과의 압축성능 비교를 위해 bpp (bit per pixel)와 mAP (mean Average Precision)를 측정하여 R-P (Rate-Performance) 곡선을 그리고 BD-rate[12]를 비교하였다. 그림 2 는 C-레이어 피쳐 맵 압축 방법과 제안하는 방법의 R-P 곡선으로, 파란색 점선은 C-레이어 피쳐 맵 압축 방법, 주황색 점선은 제안하는 방법의 R-P 곡선을 나타낸다. C-레이어 피쳐 맵 압축 방법 대비 제안하는 방법이 bpp 와 mAP 의 BD-rate 관점에서 35.10%의 성능 향상을 보였다.

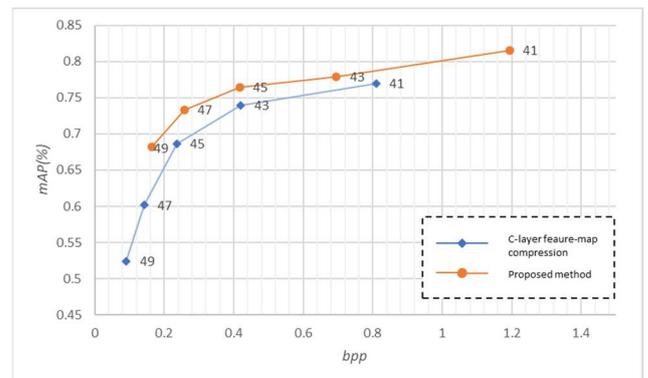


그림 2. C-레이어 피쳐 맵 압축 방법과 제안하는 방법 R-P 곡선 비교

표 1. C-레이어 피쳐 맵 압축 방법과 제안하는 방법을 사용한 경우의 QP 별 OpenImageV6 데이터셋의 bpp 및 mAP 성능 비교

QP	C-layer feature map compression anchor		Proposed method	
	bpp	mAP (%)	bpp	mAP (%)
41	0.81	0.77	1.19	0.82
43	0.42	0.74	0.70	0.78
45	0.24	0.69	0.42	0.76
47	0.14	0.60	0.26	0.73
49	0.09	0.52	0.16	0.68

4. 결론 및 향후 연구 방향

본 논문에서는 다해상도 피라미드 피쳐 맵의 압축 손상을 보상하는 압축 방법을 제안하였다. 본 논문에서 제안하는 방법은 C-레이어 피쳐 맵 압축 방법 대비 35.10%의 BD-rate 성능 향상을 확인할 수 있었다. 다만 C2 피쳐 맵에 압축 손상을 보상하는 과정을 추가할 경우, 성능 하락의 문제가 있었다. 따라서 향후 C2 피쳐 맵에 압축 손상을 보상하는 문제 분석 및 성능 향상을 위한 연구를 진행할 예정이다.

감사의 글

이 논문은 2022 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2020-0-00011, (전문연구실)기계를 위한 영상부호화 기술)

참조문헌

- [1] G. Sullivan, J. Ohm, W. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," IEEE Transactions on circuits and systems for video technology, Vol.22, No.12, pp. 1649-1668, Dec. 2012.
- [2] B. Bross, J. Chen, S. Liu, and Y.-K. Wang, "Versatile Video Coding (Draft 10)," JVET-S2001, Jul. 2020.
- [3] M. Lee, H. Choi, S. Park, M. Kim, S. Oh, D. Sim, Y. Kim, J. Lee, J. Do, and S. Jeong, "Advanced feature map compression based on optimal transformation with VVC and DeepCABAC," ISO/IEC JTC 1/SC 29/WG 2, m58787, Online, Jan. 2022.
- [4] S. Park, M. Lee, H. Choi, M. Kim, S. Oh, Y. Kim, J. Do, S. Jeong, and D. Sim, "A Feature Map Compression Method for Multi-resolution Feature Map with PCA-based Transformation", Journal of Broadcast Engineering 27.1, pp. 56-68, 2022.
- [5] T. Lin et al, "Feature Pyramid Networks for Object

Detection," Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2117-2125, 2017.

[6] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," Advances in neural information processing systems 28, 2015.

[7] Detectron2, <https://github.com/facebookresearch/detectron2>

[8] VTM12.0, https://vcgit.hhi.fraunhofer.de/jvet/VVC_Software_VTM/-/tree/VTM-12.0

[9] OpenImageV6, <https://storage.googleapis.com/openimages/web/download.html>

[10] FFmpeg, <https://ffmpeg.org/documentation.html>

[11] M. Lee, H. Choi, S. Park, M. Kim, S. Oh, D. Sim, Y. Kim, J. Lee, J. Do, and S. Jeong, "P-layer feature map anchor generation for object detection on OpenImageV6 dataset," ISO/IEC JTC 1/SC 29/WG 2, m58786, Online, Jan. 2022.

[12] G. Bjøntegaard, "Calculation of average PSNR differences between RDcurves," Tech. Rep. VCEGM33, Video Coding Experts Group (VCEG), 2001.