특집논문 (Special Paper)
방송공학회논문지 제27권 제1호, 2022년 1월 (JBE Vol.27, No.1, January 2022)
https://doi.org/10.5909/JBE.2022.27.1.56
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

PCA 기반 변환을 통한 다해상도 피처 맵 압축 방법

박 승 진[a)], 이 민 훈[a)], 최 한 솔[a)], 김 민 섭[a)], 오 승 준[b)], 김 연 희[c)], 도 지 훈[c)], 정 세 윤[c)], 심 동 규[a)‡]

A Feature Map Compression Method for Multi-resolution Feature Map with PCA-based Transformation

Seungjin Park[a)], Minhun Lee[a)], Hansol Choi[a)], Minsub Kim[a)], Seoung-Jun Oh[b)], Younhee Kim[c)], Jihoon Do[c)], Se Yoon Jeong[c)], and Donggyu Sim[a)‡]

요 약

본 논문에서는 VCM을 위한 다해상도 피처 맵에 대한 압축 방법을 제안한다. 제안하는 압축 방법은 PCA 기반의 변환을 통해 다해상도 피처 맵의 채널 및 해상도 계층 간 중복성을 제거하며 변환에 사용된 기저 벡터와 평균 벡터 그리고 변환을 통해 얻어진 변환 계수를 각각의 특성에 따라 VVC 기반 부호화기와 DeepCABAC을 통하여 압축한다. 제안하는 방법의 성능을 측정하기 위하여 OpenImageV6와 COCO 2017 validation set에 대하여 객체 검출 성능을 평가하며, MPEG-VCM 앵커 및 본 논문에서 제안하는 피처 맵 압축 앵커 대비 bpp와 mAP를 BD-rate 관점에서 비교한다. 실험 결과, 제안하는 방법은 OpenImageV6에서 피처 맵 압축 앵커 대비 25.71%의 BD-rate 성능 향상을 보이며, 특히 COCO 2017 validation set의 크기가 큰 객체들에 대해서 MPEG-VCM 앵커 대비 최대 43.72%의 BD-rate 성능이 향상됨을 보인다.

Abstract

In this paper, we propose a compression method for multi-resolution feature maps for VCM. The proposed compression method removes the redundancy between the channels and resolution levels of the multi-resolution feature map through PCA-based transformation. According to each characteristic, the basis vectors and mean vector used for transformation, and the transformation coefficient obtained through the transformation are compressed using a VVC-based coder and DeepCABAC. In order to evaluate performance of the proposed method, the object detection performance was measured for the OpenImageV6 and COCO 2017 validation set, and the BD-rate of MPEG-VCM anchor and feature map compression anchor proposed in this paper was compared using bpp and mAP. As a result of the experiment, the proposed method shows a 25.71% BD-rate performance improvement compared to feature map compression anchor in OpenImageV6. Furthermore, for large objects of the COCO 2017 validation set, the BD-rate performance is improved by up to 43.72% compared to the MPEG-VCM anchor.

Keyword : Video coding for machine, Principal component analysis, Feature map compression, Moving picture expert group

Copyright ⓒ 2022 Korean Institute of Broadcast and Media Engineers. All rights reserved.
“This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (http://creativecommons.org/licenses/by-nc-nd/3.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered.”

Ⅰ. 서 론

모바일 기기의 보급과 SNS(Social network service)의 일반화 등 4차 산업 혁명에 힘입어 생성되고 소비되는 데이터 양은 폭발적으로 증가하는 추세이다[1]. 특히 영상 데이터는 인터넷 트래픽의 대부분을 차지할 정도로 그 양이 방대하며, 풍부한 정보를 담고 있어 의료, 제조, 자동차 등의 다양한 산업 분야에서 활용하려는 움직임이 활발하다. 하지만 영상이 가진 정보의 양이 거대한 만큼 영상을 효율적으로 분석하고 처리하기 위해 다양한 방법이 연구되어오고 있다. 최근에는 딥 러닝(Deep learning) 기반의 인공지능을 활용한 영상 분석 및 처리 기술이 높은 정확도를 보이고 있다[2].

딥 러닝 기술의 발전과 더불어, 영상을 입력으로 하는 기계 시각 임무 수행 기술이 보편화되고 있다. 이러한 응용에서는 영상의 취득 장치와 처리 장치가 다른 경우가 많으며, 전송 비용을 고려하면 영상의 압축은 필수적이다. 일반적으로 영상 압축을 위해서는 HEVC(High efficiency video coding)[3], VVC(Versatile video coding)[4]와 같은 영상 압축 표준 기술을 기반으로 하는 부호화기를 사용한다. 이러한 기술들은 효과적으로 영상을 압축하기 위해 인간의 시각적 특징을 고려해 인간이 인식할 수 없을 정도의 정보를 제거함으로써 압축을 수행한다. 따라서 기존 영상 부호화 기술로 압축된 영상으로 기계 시각 임무를 수행하는 경우, 실제 임무 수행에 사용되는 중요한 정보가 제거되어 임무 수행의 정확도 측면에서 비효율적일 수 있다. 이에 따라 기계 시각 임무 수행을 위한 효율적인 영상 압축 기술에 대한 필요성이 제기되었다.

다양한 산업 분야에서의 요구와 기존 영상 압축 방법의 한계점이 맞물려, MPEG(Moving picture expert group)은 2019년 7월 제127차 회의에서 기계에서 소비될 영상을 압축하는 기술의 표준을 수립하기 위해 VCM(Video coding for machines)이라는 이름의 AhG(Ad-hog group)을 구성하였다. MPEG은 본격적인 표준화에 앞서 MPEG-VCM의 표준 기술 후보를 탐색하기 위해 CfE(Call for evidence)의 발행과 CfE 응답을 위해 실험 조건, 평가 방법 및 과정과 요구사항을 정의하였다. MPEG-VCM은 기계 시각 임무의 성능을 유지하면서 영상 또는 영상으로부터 추출한 피처 맵(Feature map)을 효율적으로 압축하는 비트스트림(Bitstream) 표현에 대한 기술 제정을 목표로 한다. 이를 위해 MEPG-VCM에서 고려 중인 세 가지 파이프라인(Pipeline)은 그림 1과 같다. 그림 1 (a)와 (b)는 기계 시각 임무 수행을 목표로 영상 또는 피처 맵을 압축하는 구조이고, 그림 1의 (c)는 기계 시각 임무 수행뿐만 아니라, 인간에게 소비될 응용까지 동시에 고려하는 하이브리드(Hybrid) 시각을 목표로 영상을 압축하는 구조이다. MPEG-VCM에서 상정하는 사용 사례에는 감시, 지능형 교통, 스마트 시티, 지능형 산업, 지능형 콘텐츠 그리고 전자 제품에서의 사용과 같이 6개 항목이 논의되었다. 각 항목에서 활용될 다양한 기계 시각 임무 중, 성능 평가의 베이스라인(Baseline)으로 사용하기 위한 앵커(Anchor) 생성을 위해, 공통으로 사용되는 다섯 가지 주요 임무인 객체 탐지, 객체 분할, 객체 추적, 행동 인식, 자세 추정에 대해 각 임무에 대응되는 딥 러닝 네트워크 및 데이터 셋을 정의하였고, 그림 1 (a)의 파이프라인에 따라 최신 영상 데이터 압축 표준 기술 VVC의 참조 소프트웨어인 VTM12.0(VVC test model 12.0)[5]을 통해 압축 및 복원한 영상으로부터 기계 시각 임무의 성능을 측정하는 것을 앵커 생성 절차로 정의하였다[6].

지난 수 차례 MPEG 회의에서, MPEG-VCM을 위한 다양한 후보 기술들이 CfE에 대한 응답 기술로써 제시되었다. 그중, 기계 시각 임무를 위해 그림 1 (a)와 같이, 영상을 압축하는 방법인 비디오 코덱(Codec)과 그림 1 (b)와 같이, 영상으로부터 추출한 피처 맵을 압축하는 방법인 피처 맵 코덱에 대한 논의가 활발하게 이루어졌다. 비디오 코덱 방식은 크게 전통적인 영상 부호화 기술 기반의 압축 방법과

a) 광운대학교 컴퓨터공학과(Department of Computer Engineering, Kwangwoon University)
b) 광운대학교 전자공학과(Department of Electronic Engineering, Kwangwoon University)
c) 한국전자통신연구원(Electronics and Telecommunications Research Institute)
‡ Corresponding Author : 심동규(Donggyu Sim)
E-mail: dgsim@kw.ac.kr
Tel: +82-2-940-5470
ORCID:https://orcid.org/0000-0002-2794-9932
※ 이 논문은 2021년도 광운대학교 우수연구자 지원 사업과 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2020-0-00011, (전문연구실)기계를 위한 영상부호화 기술).
※ This work was supported by the Excellent researcher support project of Kwangwoon University in 2021 and Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2020-0-00011, Video Coding for Machine).
· Manuscript received December 2, 2021; Revised December 15, 2021; Accepted December 16, 2021.

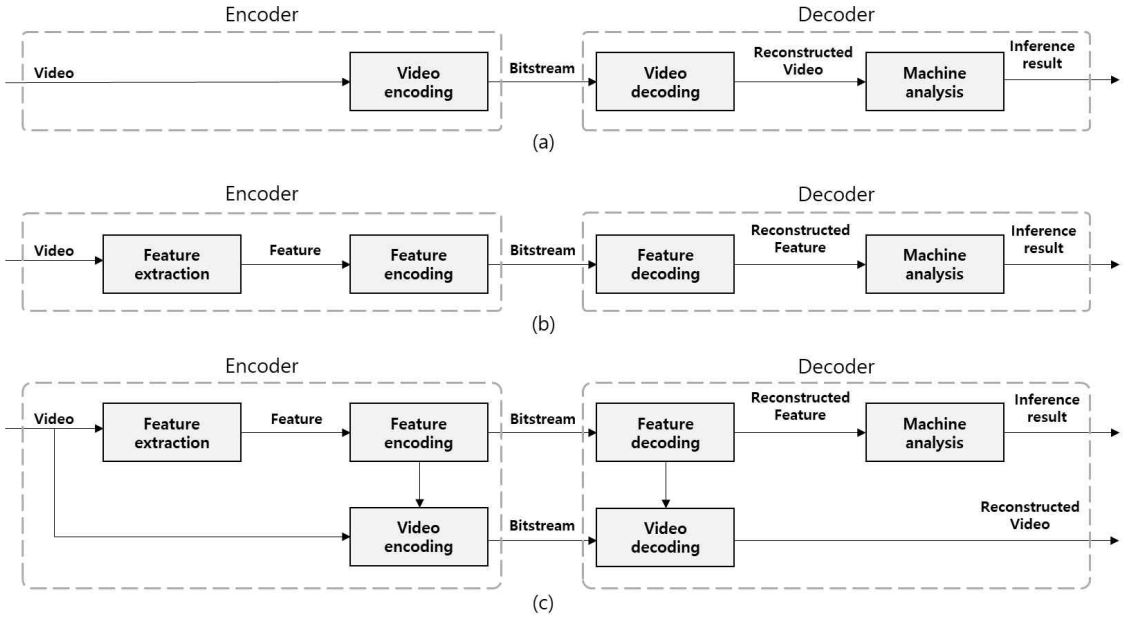


그림 1. MPEG-VCM의 파이프라인 구조 예시 (a) 비디오 코덱; (b) 피처 맵 코덱; (c) 하이브리드 코덱
Fig. 1. Examples of pipeline structure in MPEG-VCM (a) Video codec; (b) Feature codec; (c) Hybrid codec

딥 러닝 기반 압축 방법으로 분류되어 다양한 기술들이 제시되었다. 전통적인 영상 부호화 기술 기반의 압축 방법은 HEVC 또는 VVC와 같은 기존 영상 부호화기에서 사용되는 다양한 기술들이 기계 시각 임무에 어떤 영향을 미치는지 분석하고, 이를 기반으로 압축 효율을 높이는 방법[7][8][9]과, 전/후처리 과정을 통해 압축 효율을 높이는 방법[10], 등이 제시되었다. 반면, 딥 러닝 기반 압축 방법으로는 종단간 압축 네트워크(End-to-end compression network) 기반의 압축 방법이 주로 논의되고 있으며, 이때 압축 파라미터를 학습하기 위한 손실 함수에 기계 시각 임무의 성능을 반영하여 최적화를 수행함으로써 MPEG-VCM 앵커 대비 높은 압축 성능을 보이고 있다[11][12].

피처 맵 코덱 방식은 크게 전통적인 영상 부호화기 기반 압축 방법[13][14]과 딥 러닝 기반 압축 방법[15]으로 분류되어 다양한 기술들이 제시되었다. 피처 맵을 전통적인 영상 부호화기로 압축하기 위해서는 몇 가지 절차가 수반된다. 피처 맵은 보통 실수 값들로 구성되지만, 영상 부호화기는 정수형 영상을 입력으로 받기 때문에 양자화가 필수적이다. 또한 한 개 혹은 세 개의 채널을 갖는 일반적인 영상과 다르게 피처 맵은 수십, 수백 개의 채널로 구성되기 때문에 효과적으로 영상 부호화기를 통해 영상을 압축을 위해서는 피처 맵의 채널들을 한 개 또는 다수 개의 영상으로 패킹(Packing)하는 과정이 필요하다[16]. 또한 피처 맵 압축을 위해 딥 러닝 네트워크를 구성하는 수많은 레이어(Layer)의 피처 맵 중 압축할 대상을 선택해야 한다. 압축되는 피처 맵이 입력 단과 가까울수록 압축 과정에서 발생하는 손실이 남은 네트워크를 거치며 증폭되어 기계 시각 임무의 정확도가 낮아지는 경향을 보이며[17], 각 레이어의 피처 맵이 각기 다른 크기를 가지므로 적당한 레이어의 피처 맵을 선택하는 작업 역시 피처 맵 코덱의 성능에 큰 영향을 미친다. 앞서 언급한 VCM의 주요 임무에 따라 결정된 딥 러닝 네트워크 중, 많은 네트워크가 스케일(Scale)에 대한 불변성을 위해 FPN(Feature pyramid network) 구조를 포함한다. 기계 시각 임무 성능의 보존을 위하여 FPN 구조로부터 추출한 피처 맵들의 압축을 수행하는 것이 효율적이지만, FPN의 피처 맵은 다수 개의 피처 맵들로 이루어져 있기 때문에 압축 이전 데이터양이 매우 크다는 문제점이 존재한다. 하지만 FPN의 피처 맵은 다해상도의 피처 맵들을 구성하기 위하여 고해상도 피처 맵의 생성 과정에서 저해상도 피처 맵을 업 샘플링(Up-sampling) 하여 이용하기 때문

에 각 피처 맵 간의 중복성이 발생한다.

본 논문에서는 VCM의 다양한 기계 시각 임무를 위한 딥 러닝 네트워크에서 공통으로 사용되는 FPN으로부터 추출한 피처 맵을 압축하는 방법을 제안한다. FPN 구조에서 추출된 피처 맵은 상대적으로 딥 러닝 네트워크의 중반부에 위치하여, 보다 입력단에 가까운 레이어로부터 추출한 피처 맵을 압축하였을 때와 비교하여 압축 손실의 전파되는 양이 적기 때문에 기계 시각 임무의 성능을 높게 유지할 수 있다는 장점이 존재한다. 반면, 상대적으로 입력단에 가까운 레이어로부터 추출한 피처 맵에 비해 데이터양이 많아 효과적으로 압축하기 위한 방법이 필요하다. 제안하는 방법은 FPN 구조에서 추출한 피처 맵의 채널 간 중복성을 제거하기 위해 PCA(Principal component analysis) 기반의 변환을 적용하여 데이터의 양을 줄인다. 또한, FPN 구조 내의 해상도 계층 간 중복성을 제거하기 위하여 고해상도 피처 맵의 변환 기저를 예측함으로써 데이터의 양을 더욱 효과적으로 감소시키며 PCA 기반의 데이터 축소 과정에서 발생하는 기저 벡터와 평균 벡터, 그리고 변환 계수를 각각의 특성에 따라 전통적인 부호화기를 통해 압축함으로써 기계 시각 임무의 정확도를 유지하며 효과적인 비트-율(Bit-rate) 감소를 달성한다. 제안하는 방법의 성능을 평가하기 위해 OpenImageV6[18] 및 COCO2017 validation set[19] 데이터 셋을 사용하였으며, 객체 검출 임무의 성능을 기반으로 하는 BD-rate[20]를 다른 앵커들과 비교하였으며 실험 결과 본 논문에서 제시한 피처 맵 압축 앵커보다 우수한 부호화 효율을 보였으며 피처 맵 대비 약 80배 적은 데이터양을 갖는 이미지를 압축하는 MPEG-VCM 앵커와 유사한 부호화 효율을 보였다.

본 논문의 구성은 다음과 같다. 2장에서는 피처 맵에 적용하는 PCA 기반 변환 및 변환의 기저 벡터, 평균 벡터, 그리고 변환 계수를 압축하기 위해 사용한 전통적인 부호화기에 대하여 설명한다. 그리고 3장에서는 제안하는 방법의 압축 성능을 평가하고 4장에서 결론을 맺는다.

Ⅱ. 제안하는 피처 맵 압축 방법

기존의 영상 압축 방법을 통한 압축 과정에서는 영상 내 공간적 중복성 제거를 위해 주로 DCT(Discrete cosine transform) 혹은 DST(Discrete sine transform)와 같이 고정된 기저를 사용하는 변환이 사용된다. 하지만 FPN 구조로부터 추출한 피처 맵은 입력 영상 대비 매우 큰 데이터양을 갖고, 이에 따라 피처 맵 내의 중복성 제거를 위해 고정된 기저를 사용하는 것은 효율적이지 않을 수 있다. 따라서 본 논문에서는 피처 맵 내의 채널 간 중복성을 제거하기 위해 PCA를 사용하여 입력 데이터로부터 적응적인 기저 벡터를 추출하고, 원본 피처 맵 데이터의 정보를 가장 잘 표현하는 소수의 기저 벡터만을 사용하여 피처 맵 변환을 수행하는 방법을 제안한다. 또한, 제안하는 방법은 FPN 구조의 특성을 기반으로 해상도 계층 간 중복성을 제거하여 보다 적은 양의 기저 벡터로 피처 맵을 표현하며, 효과적인 압축을 위해 변환 후 각 데이터의 특성에 따라 추가적인 압축을 수행한다.

그림 2는 제안하는 피처 맵 압축 방법의 블록도이다. PCA를 통한 기저 벡터를 추출하는 과정에서 피처 맵의 해상도 간 중복성을 기반으로 공통 기저 벡터를 추출하고, 이를 통해 각 피처 맵의 변환을 수행한다. 본 논문에서 제안하는 PCA 기반의 변환은 데이터 셋에 적응적인 기저 벡터를 추출하여 수행하기 때문에 원래 데이터로의 복원을 위해서는 복호화기는 변환 계수와 함께 기저 벡터를 필요로 한다. 영상을 입력으로 하는 딥 러닝 네트워크에서는 입력 단과 가까운 피처 맵은 여전히 영상적 특징이 남아있으며, 이러한 피처 맵으로부터 PCA를 통해 추출된 기저 벡터 역시 영상적 특징이 가지고 있을 수 있다. 따라서 제안하는 방법에서는 피처 맵의 기저 벡터를 압축하기 위해 전통적인 영상 부호화기를 이용한다. 반면, 변환 계수는 공간적으로 연속적인 영상의 특징을 갖지 않으므로 변환 계수는 엔트로피(Entropy) 부호화기를 통해 압축한다. 이때 엔트로피 부호화기는 압축 대상의 분산이 작을수록 부호화 효율이 좋으므로 각 기저 벡터에 해당하는 계수들의 평균값이 0이 되도록 변환 이전에 피처 맵에서 평균 벡터를 제거한다. 또한, 원본 피처 맵으로 복원을 위해서는 평균 벡터 역시 필요하며 평균 벡터는 기저 벡터와 같이 영상적인 특성을 가지므로 기저 벡터와 함께 전통적인 영상 부호화기를 통해 압축을 수행한다.

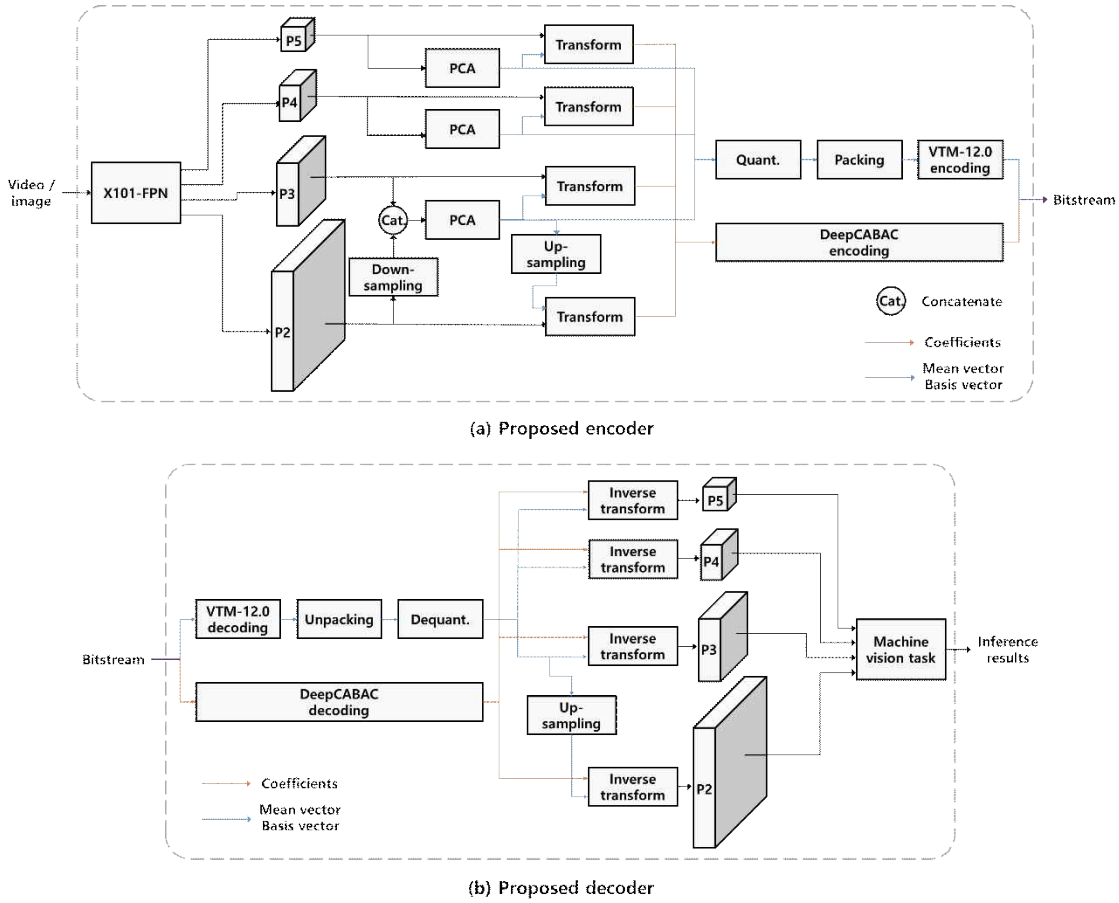


그림 2. 제안하는 피처 맵 압축 방법의 파이프라인 구조
Fig. 2. Pipeline architecture of the proposed feature map compression method

1. PCA 기반 피처 맵 채널 간 중복성 제거

피처 맵의 각 채널은 서로 다른 커널(Kernel)을 통해 원본 영상이 가진 임의 성분을 공간 도메인(Domain) 상에 표현한다. 하지만 각각의 커널은 서로 고유하지 않으며 피처 맵의 채널 간에는 중복성이 발생할 수 있다[21]. 제안하는 방법에서는 피처 맵의 높은 압축 효율을 위해 전처리 과정으로 PCA 기반의 변환을 적용하여 피처 맵의 차원을 축소한다.

PCA는 N개의 차원으로 표현되는 고차원 데이터 집합 안에서 상호 연관성을 분석하여 각 데이터가 가지는 정보를 최대한 보존하며 원래의 데이터 집합과 선형 관계에 있는 저차원 공간을 찾아내는 방법이다. PCA를 통해 찾아낸 저차원 공간은 서로 직교하는 M개의 기저 벡터로 표현되며 고차원의 데이터는 기저 벡터들을 이용하여 저차원 공간으로 사영할 수 있다. 이때 하나의 데이터를 표현하기 위해 필요한 변량은 N개에서 M개로 차원 축소가 일어난다. 일반적으로 PCA를 통해 구해지는 기저 벡터의 수는 분석하고자 하는 데이터 집합의 차원 크기와 표본 수 중 작은 것과 같으며, 각 기저 벡터가 데이터를 표현할 때 가지는 에너지를 고려하여 전체 기저 벡터 중 일부만 선별하여 사용한다. 고차원 데이터 집합의 중복성이 클수록 적은 수의 기저 벡터로 에너지가 집중되며 더 낮은 차원에서 데이터를 효과적으로 표현할 수 있다.

본 논문에서 제안하는 방법은 피처 맵의 채널 간 중복성을 제거하기 위해 PCA 기반 변환을 적용한다. 이를 위해

피처 맵의 각 채널을 고차원 데이터 집합에서의 하나의 표본으로 간주하여 PCA를 수행한다. 이때 피처 맵의 너비와 높이, 채널의 수가 각각 W, H, C라고 할 때, 일반적으로 PCA는 1차원 데이터에 대하여 정의하므로 2차원 데이터인 피처 맵 표본을 1차원으로 재배열한다. 이때 피처 맵은 (W×H)×C 크기를 갖는 2차원 행렬로 표현된다. 이 행렬에 PCA를 수행할 경우 보통의 피처 맵에서 W×H는 C에 비하여 큰 값을 갖기 때문에 최대 C개의 W×H 차원 기저 벡터를 얻게 되며 그중 C'개의 기저 벡터를 선택한다. C'을 크게 설정할수록 원본 데이터를 정확하게 표현할 수 있기 때문에 기계 시각 임무의 정확도를 높게 유지할 수 있는 반면 추가로 압축해야 할 기저 벡터와 변환 계수의 수가 증가하기 때문에 압축율이 감소하므로 목표 비트-율이나 임무의 정확도에 따라 최적의 C' 값을 설정해야 한다. PCA를 기반으로 결정한 C'개의 기저 벡터를 이용하여 다시 2차원으로 재정렬된 피처 맵을 변환하면 C'×C 크기의 변환 계수를 추출할 수 있다.

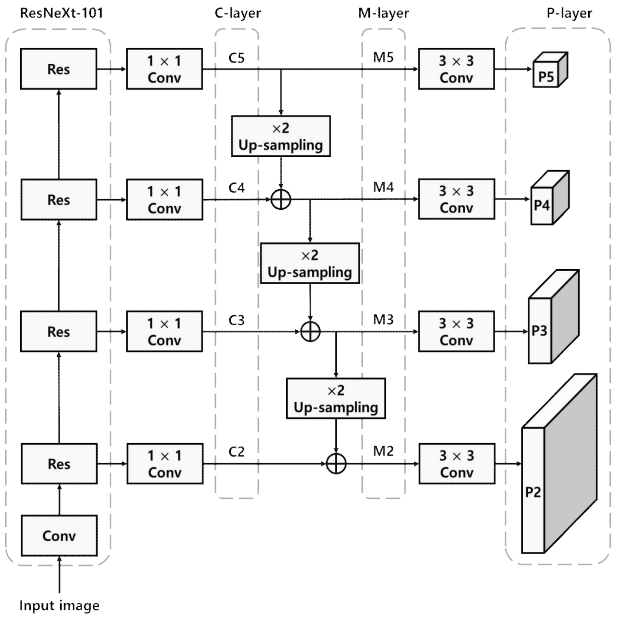


그림 3. 객체 검출을 위한 Faster R-CNN X101-FPN의 FPN 구조
Fig. 3. FPN structure of Faster R-CNN X101-FPN for object detection

2. 해상도 계층간 중복성 제거

MPEG-VCM의 표준화 과정에서 객체 검출과 객체 분리 및 객체 추적 임무를 위해 FPN 구조를 백본(Backbone) 네트워크로 사용하는 딥 러닝 네트워크를 선정하였다[22]. FPN 구조에서 입력 영상으로부터 추출된 피처 맵은 임무 수행을 위한 후반부의 네트워크에서 사용된다. 그중, 객체 검출을 위한 딥 러닝 네트워크인 Faster R-CNN X101-FPN 내의 FPN 구조는 그림 3과 같이, ResNeXt-101[23] 네트워크

표 1. 해상도 계층에 따른 피처 맵의 데이터양과 전체 데이터양에서 차지하는 비중
Table 1. The amount of data in the P-layer feature map according of the resolution layer

Resolution level	Data amount of feature map	Ratio of data amount (%)
P5	W×H×C	1.18
P4	2W×2H×C	4.71
P3	4W×4H×C	18.82
P2	8W×8H×C	75.29

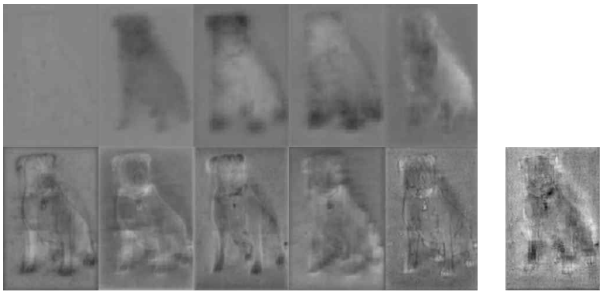
표 2. 입력 영상의 크기가 640×480×3 일 때, P-레이어 피처 맵의 PCA 기반 변환 인자의 데이터양 (C'=25)
Table 2. Data amount of PCA-based transformation parameters of the P-layer feature map when the size of the input image is 640×480×3 (C'=25)

Resolution level	Data amount			
	Feature map (W×H×C)	Basis vector ((W×H)×C')	Mean vector (W×H)	Coefficients (C×C')
P5	34×25×256 =217,600	34×25×25 =21,250	34×25 =850	256×25 =6,400
P4	68×50×256 =870,400	68×50×25 =85,000	68×50 =3,400	256×25 =6,400
P3	136×100×256 =3,481,600	136×100×25 =340,000	136×100 =13,600	256×25 =6,400
P2	272×200×256 =13,926,400	272×200×25 =1,360,000	272×200 =54,400	256×25 =6,400

로부터 추출한 여러 해상도의 피처 맵을 종합하여 P-레이어의 피처 맵을 생성하며, 인접한 해상도 계층의 피처 맵 간에는 두 배의 해상도 차이를 보인다. 따라서 가장 높은 해상도 계층의 피처 맵인 P2는 가장 낮은 해상도 계층의 피처 맵인 P5 보다 64배 높은 해상도를 가지며 표 1과 같이, P-레이어 피처 맵의 전체 데이터양 중에서 약 75%를 차지한다. 또한, 표 2는 각 P-레이어 피처 맵의 PCA 기반 변환에 따르는 데이터양과 그 비율을 나타낸다. 앞서 언급한 것과 같이 너비와 높이, 채널의 수가 각각 W, H, C인 피처 맵에 대하여 PCA 기반 변환을 수행하여 피처 맵의 차원을 C'으로 축소할 때, 기저 벡터와 평균 벡터 그리고 변환 계수의 크기는 각각 (W×H)×C', W×H, C×C'이고, 일반적인 피처 맵에서 (W×H)는 C보다 크기 때문에 PCA 기반 변환을 수행하였을 때 기저 벡터와 평균 벡터가 데이터의 대부분을 차지한다. 표 2와 같이 P2의 변환 인자들의 데이터양은 다른 피처 맵보다 매우 크며, 특히 P2의 기저 벡터의 데이터양은 다른 피처 맵에 비해 매우 크다.

따라서 본 논문에서는 P-레이어의 서로 다른 해상도 계층의 피처 맵 간의 공통의 기저 벡터를 추출하여 변환에 사용함으로써 해상도 계층 간 피처 맵의 중복성을 제거하여 데이터의 양을 효과적으로 줄이는 방법을 제안한다. 특히, 데이터양의 대부분을 차지하는 P2와 그보다 한 단계 낮은 해상도를 갖는 P3의 공통 기저 벡터를 통해 추출한다. 그림 2와 같이 P-레이어 피처 맵은 M-레이어의 피처 맵이 한 층의 컨볼루션(Convolution) 레이어를 거치며 생성되고 M-레이어의 각 피처 맵은 한 단계 낮은 해상도의 피처 맵을 최근접 이웃 보간법(Nearest neighbor interpolation)으로 업 샘플링 한 후 잔차 피처 맵과 합산하여 생성한다. 이때 잔차 피처 맵의 에너지는 매우 작기 때문에 P-레이어 피처 맵은 M-레이어 피처 맵의 강한 영향을 받는다. 따라서 P-레이어 역시 M-레이어의 업 샘플링으로 인한 해상도 계층 간 높은 중복성을 보인다. 이를 기반으로 P2와 P3의 공통의 기저 벡터를 찾기 위해 P2를 P3 크기에 맞춰 쌍입방 보간법(Bicubic interpolation)으로 다운 샘플링(Down sampling) 하여 P3와 채널 축으로 연결(Concatenate) 후, PCA를 수행한다. 이러한 과정에서 유도된 공통의 기저 벡터를 통해 P3의 변환 계수를 도출한다. 그리고 P2의 변환 계수를 얻기 위해 P3와 동일한 해상도를 갖는 공통의 기저 벡터가 P2와 동일한 해상도를 갖도록 피처 맵 생성에서 쓰인 방법과 동일하게 업 샘플링을 수행한 뒤, P2의 변환 계수를 도출한다.

3. 기저 벡터 및 평균 벡터 압축



(a) (b)

그림 4. 시각화 된 기저 벡터와 평균 벡터의 예시 (a) 기저 벡터; (b) 평균 벡터

Fig. 4. An example of visualized basis vectors and mean vector (a) Basis vectors; (b) Mean vector

그림 4와 같이 기저 벡터와 평균 벡터는 다시 원래의 모양으로 재정렬하였을 때 공간적 중복성이 존재하는 영상적 특징을 갖기 때문에 예측을 기반으로 압축을 수행하는 영상 압축 표준 기술인 VVC를 이용하여 압축한다. 기존 영상 부호화기의 입력인 디지털 영상은 각 화소 값이 양자화되어 있지만, 기저 벡터와 평균 벡터는 실수 값들로 구성된다. 또한, 기저 벡터와 평균 벡터를 VVC를 통해 효율적으로 압축하기 위해서는 하나 또는 다수 개의 프레임(Frame)으로 구성하기 위한 패킹을 고려할 수 있다. 이때 시간 축으로의 배열과 공간 축으로의 배열로써 패킹하는 방법이 있는데, 기저 벡터는 서로 직교하므로 시간적 예측은 효과가 떨어질 수 있다. 따라서 본 논문에서는 각 해상도 계층의 기저 벡터와 평균 벡터를 회색조 영상으로 한 공간상에 배열한 단일 프레임 영상으로 패킹을 수행한다. 또한, 취합된 각 단일 프레임 영상은 영상 단위로 값의 범위를 고려하여 8-bits로 균등 양자화(Uniform quantization)한 뒤, VVC를 이용하여 압축한다.

4. 변환 계수 압축

그림 5와 같이 변환 계수는 기저 벡터, 평균 벡터와는

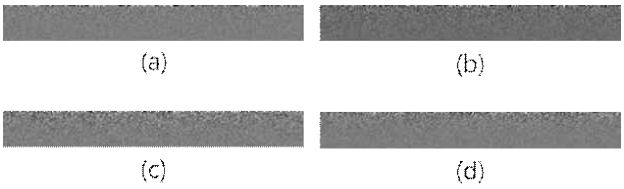


그림 5. 시각화 된 P-레이어 피처 맵의 변환 계수 예시 (a) P2; (b) P3; (c) P4; (d) P5
Fig 5. An example of visualized transformation coefficients of the P-layer feature map (a) P2; (b) P3; (c) P4; (d) P5

다르게 공간상에서 값이 연속적인 영상적 특징을 갖지 않으므로 공간적 예측을 통한 압축은 효율적이지 않다. 따라서 본 논문에서는 통계적 예측을 기반으로 하는 엔트로피(Entropy) 부호화기인 DeepCABAC[24]을 채택하여 변환 계수를 압축한다.

DeepCABAC은 문맥 기반의 적응적 이진 산술 부호화기(Context-based adaptive binary arithmetic coder)의 일종으로 딥 러닝의 실수형 파라미터(Parameter)를 압축하도록 고안되었다. DeepCABAC은 그림 6과 같은 순서로 변환 계수를 압축한다. 일반적으로 엔트로피 부호화는 무손실 압축을 수행하지만 DeepCABAC은 실수형 데이터에 대해 압축 효율을 높이기 위해 먼저 균등 양자화를 수행한다. 양자화된 입력은 다시 이진화하여 산술 부호화를 진행하는데, 이때 이미 입력된 n개의 값을 이용해 현재 부호화하려는 값을 확률적으로 예측하여 부호화에 사용한다. 본 논문에서는 각 해상도 레벨의 변환 계수 간에는 상관도가 낮다고 가정하고 서로 독립적으로 DeepCABAC을 통해 압축을 수행하였다.

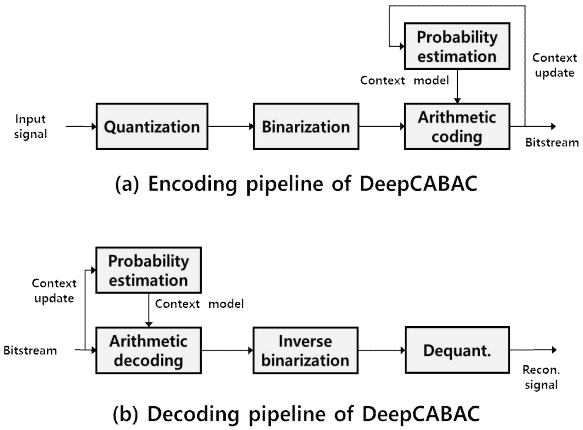


그림 6. DeepCABAC의 부/복호화기의 파이프라인 구조
Fig. 6. Encoding and decoding pipeline structure of DeepCABAC

Ⅲ. 실험 환경 및 결과

본 논문에서 제안하는 방법의 성능을 평가하기 위해 객체 검출 임무를 수행하는 Detectron2 프레임워크(Framework)를 사용하였으며 기저 벡터와 평균 벡터 그리고 변환 계수의 압축을 위해 VTM12.0과 DeepCABAC을 사용하였다. 실험 데이터 셋으로는 OpenImage V6와 COCO 2017 validation set 중 VCM에서 공통된 실험을 위해 추려낸 각 5000장의 이미지를 사용하였다. 변환에 사용된 기저 벡터는 최대 기저 벡터의 수의 약 10% 수준으로 24개가 사용되었으며 평균 벡터와 함께 5×5 배열로 단일 프레임 영상을 구성하였다. 기저 벡터와 평균 벡터의 단일 프레임 영상은 VTM12.0으로 압축하기 위하여 8-bits의 회색조 영상으로

표 3. 제안하는 방법을 사용하였을 때 QP 별 OpenImageV6 데이터 셋의 bpp 및 mAP 성능
Table 3. Bpp and mAP of the proposed method according to QPs on OpenImageV6

QP	Image compression anchor		Feature-map compression anchor		Proposed method	
	bpp	mAP (%)	bpp	mAP (%)	bpp	mAP (%)
22	0.86	78.93	23.94	79.16	1.84	78.08
27	0.51	77.99	13.20	79.12	1.25	78.03
32	0.29	77.26	6.42	79.00	0.69	77.98
37	0.15	73.96	2.83	79.05	0.36	77.07
42	0.08	68.84	1.15	78.77	0.20	74.67
47	0.04	58.02	0.45	75.86	0.14	71.13

양자화하였으며 변환 계수는 DeepCABAC에서 20을 간격으로 양자화한 후 압축을 수행하였다.

VCM에서는 성능 평가를 위해 MPEG-VCM 앵커만을 정의하여 사용하고 있지만, 이는 피처 맵이 아닌 영상을 압축하여 복원된 영상으로부터 기계 시각 임무의 성능을 측정한 결과이므로 피처 맵 압축 방법의 비교 대상으로 적합하지 않다. 따라서 본 논문에서는 피처 맵을 단순하게 압축하는 방법을 추가로 비교하였다. 피처 맵 압축 앵커의 생성 절차는 제안하는 방법에서 기저 벡터와 평균 벡터에 적용한 것과 같이, P-레이어 피처 맵을 하나의 영상으로 취합하고 8-bits로 균등 양자화하여 VTM12.0을 통해 부/복호화한 후, 기계 시각 임무의 성능을 평가하는 것으로 정의하였다.

표 3과 표 4, 그리고 그림 7은 OpenImage V6 데이터 셋에 대해 MPEG-VCM 앵커와 피처 맵 압축 앵커 그리고 제안하는 방법에 대한 실험 결과이다. 피처 맵 압축 앵커는 조금 더 기계 시각 임무 네트워크에 종단에 가까운 P-레이어 피처 맵을 압축하므로 압축 손실의 전파가 적어 전반적으로 mAP(Mean average precision)가 높은 것을 확인할 수 있지만 그 데이터의 양이 매우 많기 때문에 bpp(Bit per pixel)가 높으며 실험한 다른 방법들과 비교 가능한 수준의 bpp에서는 mAP 성능이 급격하게 감소하는 것을 볼 수 있다. 반면 제안하는 방법은 피처 맵 압축 앵커와 같이 P-레이어 피처 맵을 압축함에도 데이터양을 크게 줄이는 전처리 작업을 수행하기 때문에 압축 효율이 높아 피처 맵 압축 앵커 대비 bpp와 mAP의 BD-rate 관점에서 28.11%의 성능 향상을 보였다. 또한 P-레이어 피처 맵이 원본 이미지에 비해 약 80배에 많은 데이터로 구성됨에도 제안하는 방법이 MPEG-VCM 앵커와 근접한 압축 성능을 보였다.

표 4. OpenImageV6 데이터 셋에서 앵커 대비 제안하는 방법의 BD-rate
Table 4. BD-rate of the proposed method with mAP on OpenImageV6

Anchor	BD-rate with mAP (%)
Feature-map compression anchor	-28.11
MPEG-VCM anchor	25.71

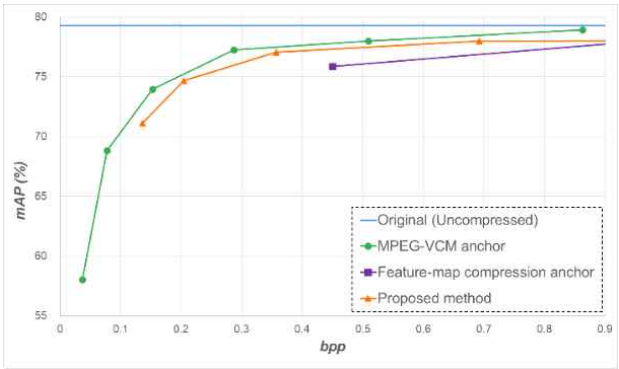


그림 7. OpenImageV6 데이터 셋에서의 rate-precision 곡선
Fig. 7. Rate-precision curves on OpenImageV6

표 5-7과 그림 8은 위와 같은 방법들을 COCO 2017 validation set에 대해 실험한 결과이다. 이 데이터 셋은 객체를 크기에 따라 세 가지 범주로 분류하여 각 크기 범주에 대한 검출 성능을 평가할 수 있다. 크기가 큰 객체에 대해서는 제안하는 방법으로 피처 맵을 압축하였을 경우 제안하는 방법의 MPEG-VCM 앵커 대비 43.72% 만큼 높은 압축 효율을 보이며 피처 맵 압축 앵커 대비 압축 효율이 36.04% 높아 모든 크기의 객체에 대해서 전반적으로 압축 효율이 상승하는 것을 보였다.

표 5. 제안하는 방법을 사용하였을 때 QP 별 COCO 2017 validation set의 Bpp 및 mAP 성능
Table 5. Bpp and mAP of the proposed method according to QPs on COCO 2017 validation set

QP	Image compression anchor		Feature-map compression anchor		Proposed method	
	bpp	mAP (%)	bpp	mAP (%)	bpp	mAP (%)
22	1.31	35.27	68.13	39.11	5.54	35.66
27	0.93	34.57	37.50	39.05	3.46	35.40
32	0.59	32.64	18.22	39.05	1.88	34.57
37	0.32	28.70	7.93	38.42	0.95	32.63
42	0.15	21.79	3.20	36.29	0.54	28.70
47	0.06	12.91	1.24	29.07	0.36	22.84

표 6. 제안하는 방법을 사용하였을 때 QP 별 COCO 2017 validation set의 객체 크기에 따른 Bpp 및 mAP 성능
Table 6. Bpp and precision of the proposed method according to QPs on COCO 2017 validation set (AP_S, AP_M and AP_L)

QP	Image compression anchor			Feature-map compression anchor			Proposed method		
	AP_S (%)	AP_M (%)	AP_L (%)	AP_S (%)	AP_M (%)	AP_L (%)	AP_S (%)	AP_M (%)	AP_L (%)
22	17.69	38.17	48.40	22.66	42.30	51.15	16.93	41.02	50.32
27	17.31	37.25	48.17	22.59	42.27	51.02	16.99	41.22	49.61
32	15.25	35.52	45.28	22.37	42.53	51.28	14.16	41.04	49.61
37	12.43	31.40	40.50	21.31	42.25	51.37	10.04	40.08	50.11
42	7.78	23.90	33.65	17.12	41.52	51.11	3.97	35.42	48.74
47	3.14	12.85	23.12	6.98	34.67	47.04	0.61	24.41	44.80

표 7. COCO 2017 validation set에서 앵커 대비 제안하는 방법의 객체 크기 별 BD-rate
Table 7. BD-rate of the proposed method according to target object size on COCO 2017 validation set

Anchor	BD-rate by target object size (%)			
	All (mAP)	Small (AP_S)	Medium (AP_M)	Large (AP_L)
Feature-map compression anchor	-36.04	-22.93	-53.56	-86.70
MPEG-VCM anchor	76.91	391.29	28.49	-43.72

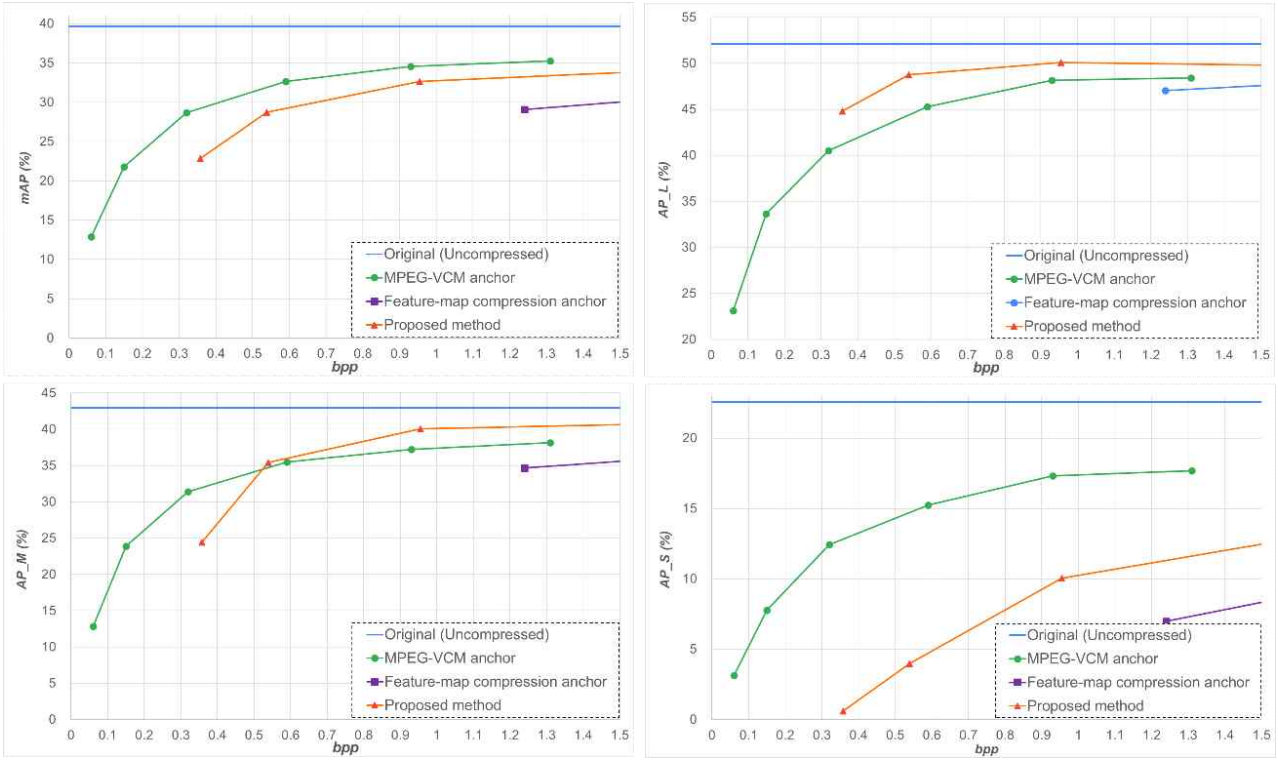


그림 8. COCO 2017 validation set 상에서의 rate-precision 곡선
Fig. 8. Rate-precision curves according to precision on COCO 2017 validation set

Ⅳ. 결 론

본 논문에서는 VCM을 위한 다해상도 계층 구조로 이루어진 피처 맵을 효과적으로 압축하기 위해서 PCA 기반의 변환을 이용하여 피처 맵의 채널 및 해상도 계층 간 중복성을 제거하는 방법을 제안하였다. 변환에 사용된 기저 벡터와 평균 벡터는 영상적 특징을 갖기 때문에 전통적인 영상 부호화기를 사용하여 압축하였으며 기저 벡터와 평균 벡터를 전통적인 영상 부호화기로 압축하기 위해서 수반되는 양자화 및 패킹 방법을 제안하였다. 반면 변환을 통해 생성된 변환 계수는 공간상에서 불연속적이기 때문에 공간적 예측 기반의 영상 부호화 방법을 통해 변환 계수를 압축하는 것은 효율적이지 않으므로 변환 계수는 통계적 특성을 활용하여 분포를 예측하는 엔트로피 부호화를 통해 압축하였다. 제안하는 방법의 성능 평가를 위하여 OpenImageV6와 COCO 2017 validation set에 대하여 객체 검출 성능을 실험한 결과 피처 맵 압축 앵커 대비 bpp와 mAP의 BD-rate 관점에서 각 테스트 셋에 대하여 28.11%, 36.04% 만큼 압축 효율의 상승을 보였다. 특히 크기가 큰 객체에 대한 검출 성능은 제안하는 방법으로 압축하였을 때 피처 맵 앵커 대비 86.70%, MPEG-VCM 앵커 대비 43.72%의 BD-rate 성능이 크게 향상되는 것을 확인하였다. 반면 크기가 작은 객체에 대해서는 MPEG-VCM 앵커보다 낮은 압축 효율을 보이는데, 이는 가장 높은 해상도를 갖는 P2를 변환하는 데 있어 업 샘플링 된 기저 벡터를 사용했기 때문으로 해석된다. 따라서 추후 크기가 작은 객체에 대한 검출 성능을 개선하기 위하여 고해상도 피처 맵의 기저 벡터를 예측하는 방법에 대한 연구가 필요하다.

참 고 문 헌 (References)

[1] M. F. Mahmood, N. Hussin, “Information in Conversion Era: Impact and Influence from 4th Industrial Revolution,” International Journal of Academic Research in Business and Social Sciences, Vol. 8, No. 9, pp. 320-328, 2018

[2] Y. LeCun, Y. Bengio, and G. E. Hinton, “Deep learning,” Nature, vol. 512, pp. 436-444, 2015.

[3] 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.

[4] B. Bross, Y. K. Wang, Y. Ye, S. Liu, and J. Chen, Overview of the versatile video coding (VVC) standard and its applications,” IEEE Transactions on Circuits and Systems for Video Technology, Vol 31, No 10, pp. 3736-3764, 2021.

[5] VTM12.0, https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM/-/tree/VTM-12.0 (accessed Nov. 26, 2021).

[6] M. Rafie, Y. Zhang, and S. Liu, “[VCM] Evaluation Framework for Video Coding for Machines,” ISO/IEC JTC 1/SC 29/WG 2, m58385, Online, Oct. 2021.

[7] 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/WG 2, m56681, Online, Apr. 2021.

[8] S. Wang, C. Lin, C. Lin, T., and Y. Nie, "[VCM] Enable IBC in VTM8.2 for VCM," ISO/IEC JTC 1/SC 29/WG 2, m56792, Online, Apr. 2021.

[9] S. Wang, C. Lin, and C. Lin (ITRI), "[VCM] A study on impact of coding tools on machine vision performance and visual quality," ISO/IEC JTC 1/SC 29/WG 2, m56867, Online, Apr. 2021.

[10] B. Zhu, L. Yu, and D. Li, "[VCM] Deep learning-based compression for machine vision," ISO/IEC JTC 1/SC 29/WG 2, m57335, Online, Jul. 2021.

[11] Y. Lee, S., K. Yoon, H. Lim, H. Choo, W. Cheong, and J. Seo, "[VCM] Updated FLIR Anchor results for object detection," ISO/IEC JTC 1/SC 29/WG 2, m57375, Online, Jul. 2021.

[12] 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/WG 2, m57500, Online, Jul. 2021.

[13] 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/WG 2, m55243, Online, Oct. 2020.

[14] S Wang, Z. Wang, Y. Ye, and S Wang, "[VCM] Image or video format of feature map compression for object detection," ISO/IEC JTC 1/SC 29/WG 2, m55786, Online, Jan. 2021.

[15] J. Do, J. Lee, Y. Kim, S. Yoon J., and J. Choi, "[VCM] Experimental Results of Feature Compression using CompressAI," ISO/IEC JTC 1/SC 29/WG 2, m56716, Online, Apr. 2021.

[16] H. Han, H. Choi, S. Kwak, J. Yun, W. Cheong, and J. Seo, "[VCM] Investigation on feature map channel reordering and compression for object detection," ISO/IEC JTC 1/SC 29/WG 2, m56653, Online, Apr. 2021.

[17] S. Wang, Z. Wang, Y. Ye, and S. Wang, “[VCM] Investigation on feature map layer selection for object detection and compression,” ISO/IEC JTC 1/SC 29/WG 2, m55787, Online, Dec. 2020.

[18] OpenImageV6, https://storage.googleapis.com/openimages/web/download.html (accessed Nov. 26, 2021)

[19] COCO2017 validation set, https://cocodataset.org/#download (accessed Nov. 26, 2021).

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

[21] L. Jinhua, T. Zhang, and G. Feng. "Channel Compression: Rethinking Information Redundancy Among Channels in CNN Architecture," IEEE Access, Vol. 8, pp. 147265-147274, 2020.
[22] M. Rafie, Y. Zhang, and S. Liu, "[VCM] Call for Evidence for Video Coding for Machines," ISO/IEC JTC 1/SC 29/WG 2, m56995, Online, Apr. 2021.
[23] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, “Aggregated Residual Transformations for Deep Neural Networks,” arXiv, 2017.
[24] S. Wiedemann *et al.*, “DeepCABAC: A universal compression algorithm for deep neural networks,” IEEE J. Sel. Topics Signal Process., Vol. 14, No. 4, pp. 700-714, May 2020.

저 자 소 개

박 승 진

- 2021년 2월 : 광운대학교 컴퓨터공학과 학사
- 2021년 3월 ~ 현재 : 광운대학교 컴퓨터공학과 석사과정
- ORCID : https://orcid.org/0000-0002-5124-0253
- 주관심분야 : 영상신호처리, 영상압축, 컴퓨터비전

이 민 훈

- 2019년 2월 : 광운대학교 수학과, 전자공학과 (복수전공) 학사
- 2021년 2월 : 광운대학교 전자공학과 석사
- 2021년 3월 ~ 현재 : 광운대학교 컴퓨터공학과 박사과정
- ORCID : https://orcid.org/0000-0001-8165-5380
- 주관심분야 : 영상신호처리, 영상압축, 컴퓨터비전

최 한 솔

- 2018년 2월 : 광운대학교 컴퓨터공학과 학사
- 2020년 2월 : 광운대학교 컴퓨터공학과 석사
- 2020년 3월 ~ 현재 : 광운대학교 컴퓨터공학과 박사과정
- ORCID : https://orcid.org/0000-0001-7947-0723
- 주관심분야 : 영상신호처리, 영상압축, 컴퓨터비전

김 민 섭

- 2020년 2월 : 광운대학교 컴퓨터공학과 학사
- 2020년 3월 ~ 현재 : 광운대학교 컴퓨터공학과 석사과정
- ORCID : https://orcid.org/0000-0001-6837-4736
- 주관심분야 : 영상신호처리, 영상압축, 컴퓨터비전, 딥러닝

저 자 소 개

오 승 준

- 1980년 2월 : 서울대학교 전자공학과 학사
- 1982년 2월 : 서울대학교 전자공학과 석사
- 1988년 5월 : 미국 Syracuse University 전기/컴퓨터공학과 박사
- 1982년 3월 ~ 1992년 8월 : 한국전자통신연구원 멀티미디어연구실 실장
- 1986년 7월 ~ 1986년 8월 : NSF Supercomputer Center 초청 학생연구원
- 1987년 5월 ~ 1988년 5월 : Northeast Parallel Architecture Center 학생연구원
- 1992년 3월 ~ 1992년 8월 : 충남대학교 컴퓨터공학부 겸임교수
- 1992년 9월 ~ 현재 : 광운대학교 전자공학과 교수
- 2002년 3월 ~ 2017년 12월 : SC29-Korea 전문위원회 대표위원
- ORCID : https://orcid.org/0000-0002-5036-3761
- 주관심분야 : 비디오데이터처리, 영상압축, 컴퓨터비전

김 연 희

- 2000년 : 아주대학교 정보및컴퓨터공학과 학사
- 2002년 : 아주대학교 정보및컴퓨터공학과 석사
- 2009년 : George Mason University Computer Science 박사
- 2009년 ~ 현재 : 한국전자통신연구원 미디어부호화연구실 선임연구원
- ORCID : http://orcid.org/0000-0003-0658-6762
- 주관심분야 : 영상처리/압축, 기계학습, 컴퓨터 비전, 실감미디어 서비스

도 지 훈

- 2018년 2월 : 한국항공대학교 항공전자정보공학부 학사
- 2020년 2월 : 한국항공대학교 항공전자정보공학과 석사
- 2020년 3월 ~현재 : ETRI 미디어부호화연구실 연구원
- ORCID : https://orcid.org/0000-0002-8254-2481
- 주관심분야 : 실감 방송, 비디오 코딩, 컴퓨터 비전, 딥러닝

정 세 윤

- 1995년 2월 : 인하대학교 전자공학과 학사
- 1997년 2월 : 인하대학교 전자공학과 석사
- 2014년 8월 : KAIST 전기및전자공학과 박사
- 1996년 12월 ~ 현재 : ETRI 미디어부호화연구실 책임연구원
- ORCID : https://orcid.org/0000-0002-1675-4814
- 주관심분야 : 실감 방송, 비디오 코딩, 컴퓨터 비전

심 동 규

- 1993년 2월 : 서강대학교 전자공학과 공학사
- 1995년 2월 : 서강대학교 전자공학과 공학석사
- 1999년 2월 : 서강대학교 전자공학과 공학박사
- 1999년 3월 ~ 2000년 8월 : 현대전자 선임연구원
- 2000년 9월 ~ 2002년 3월 : 바로비젼 선임연구원
- 2002년 4월 ~ 2005년 2월 : University of Washington Senior research engineer
- 2005년 3월 ~ 현재 : 광운대학교 컴퓨터공학과 교수
- ORCID : https://orcid.org/0000-0002-2794-9932
- 주관심분야 : 영상신호처리, 영상압축, 컴퓨터비전