

VCM을 위한 PCA 기반 피쳐 맵 압축 방법

박승진<sup>1</sup> 이민훈<sup>1</sup> 최한솔<sup>1</sup> 김민섭<sup>1</sup> 오승준<sup>1</sup> 김연희<sup>2</sup> 도지훈<sup>2</sup> 정세윤<sup>2</sup> \*심동규<sup>1</sup>  
 광운대학교<sup>1</sup> 한국전자통신연구원<sup>2</sup>  
 {promo, minhun, hschoi95, minsub20, sjoh, \*dgsim}@kw.ac.kr  
 {kimyounhee, jhdo, jsy}@etri.re.kr

A PCA-based feature map compression method applied to video coding for machines

Park, Seungjin<sup>1</sup> Lee, Minhun<sup>1</sup> Choi, Hansol<sup>1</sup> Kim, Minsub<sup>1</sup> Oh, Seung-Jun<sup>1</sup>  
 Kim, Younhee<sup>2</sup> Do, Jihoon<sup>2</sup> Jeong, Se Yoon<sup>2</sup> \*Sim, Donggyu<sup>1</sup>  
 Kwangwoon University<sup>1</sup> Electronics and Telecommunications Research Institute<sup>2</sup>

요약

인공지능 기반 머신 비전 응용이 증가함에 따라 사람이 아닌 기계에서 소비되는 영상 정보를 전송하는 요구가 발생하고 있다. 일반적으로 영상 정보를 전송할 때는 전송 비용을 고려하여 정보를 압축하며 기존 영상 압축 방법은 사람의 시각 인지적 특성을 반영하여 설계되었다. 따라서 기존 영상 압축 방법은 기계에서 소비되는 영상 정보를 압축하는 방법으로 적절하지 않다고 판단하여 2019년 7월, 기계를 위한 영상 부호화 기술의 표준화가 시작되었다. 본 논문에서는 머신 비전 태스크 중, 객체 탐지를 수행하는 네트워크의 피쳐 맵을 압축하는 방법을 제안한다. 제안하는 방법은 피쳐 맵의 채널 간 중복성을 제거하기 위해 PCA 기반의 변환을 적용하여 피쳐 맵의 차원을 축소하며 특히 해상도 계층 구조를 갖는 네트워크의 피쳐 맵을 압축하기 위해 각 해상도 계층 간 변환 기저를 예측하여 추가로 압축률을 높인다. 제안하는 방법을 적용하여 객체 탐지 결과의 큰 성능 하락 없이 약 92.3%에 데이터양 감소를 달성하였다.

1. 서론

지능형 교통, 지능형 산업, 지능형 콘텐츠, 스마트 시티, 감시/정찰 등의 분야에서 인공지능 기반 응용이 증가하고 있다. 최근 이러한 응용에서는 사람을 위해서뿐만 아니라 다양한 머신 비전 태스크를 수행하기 위해 기기 간 영상 데이터를 주고받는다. 따라서 기존의 영상 압축 기술이 목표하는 인지기적 화질뿐 아니라 머신 비전 태스크의 성능 또한 고려하는 압축 기술을 개발하기 위하여 2019년 7월, 제127차 MPEG 회의에서 기계를 위한 영상 부호화(Video coding for machine, VCM) 기술의 표준화가 시작되었다. VCM 표준화 과정에서 고려되는 머신 비전 태스크는 객체 검출, 객체 분할, 객체 추적 등을 포함하며 영상 혹은 피쳐 맵을 압축하는 구조가 고려되고 있다[1][2].

머신 비전 태스크를 위한 이미지 압축 방법으로 인공지능 기반 이미지 압축 방법을 적용하였을 때, 기존 HEVC 및 VVC와 같은 전통적인 압축 방법 대비 높은 압축 효율을 보이고 있다[3]. 하지만 이미지 압축은 서버에서 머신 비전 태스크에 사용하기 위해 다시 원본 이미지로 복원되어야 하므로 응용에 따라 개인 정보 유출에 위험이 있다. 반면, 이미지로부터 추출한 피쳐 맵을 압축하여 전송하면 머신 비전 태스크에 필요한

정보는 전부 전송되지 원본 이미지로의 복원이 어려우므로 개인 정보의 유출 위험이 적다. 하지만 현재까지 연구된 피쳐 맵 압축은 이미지 압축에 비하여 낮은 압축 성능을 보이고 있다[3].

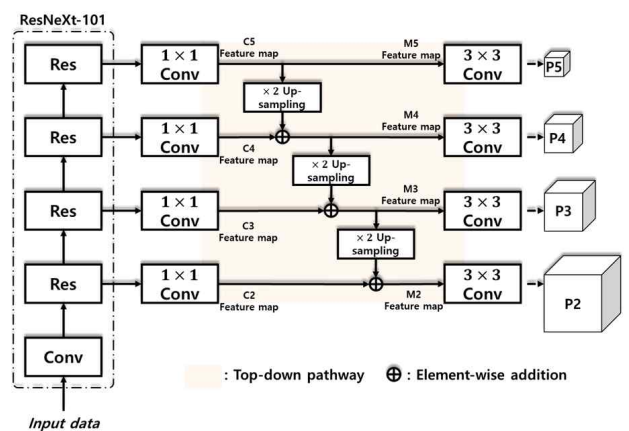


그림 1. Faster R-CNN X101-FPN 네트워크의 FPN 구조

VCM 표준화 과정에서 논의되고 있는 딥 러닝 네트워크에서 추출한 압축 대상의 피쳐 맵 중 하나는, 그림 1과 같이 주로 머신 비전 태스크가 객체 크기에 불변적인 특성을 가질 수 있도록 백본 네트워크 (Backbone network)에서 추출된 피쳐 맵을 피쳐 피라미드 네트워크 (Feature pyramid network, FPN)[4]를 통해 피라미드 형태로 재구성한 다수 개의 피쳐 맵들이다. 백본 네트워크에서 피쳐 맵의 데이터양은 입력단으로부터 레이어를 거칠수록 점차 감소하는 경향이 있지만, FPN에서는 백본 네트워크 내에 여러 레이어에서 출력된 피쳐 맵들을 취합하여 데이터양이 크게 증가한다. 따라서 기존에는 FPN에서 피라미드 형태로 취합되기 이전에 데이터양이 적은 피쳐 맵을 압축하는 방법이 주로 연구되었다. 하지만 입력단에 가까운 피쳐 맵을 압축할 경우 다시 복원된 피쳐 맵이 남은 네트워크를 거치며 압축 손실이 증폭되어 머신 비전 태스크의 정확도에 큰 영향을 미치는 문제점이 존재한다. 반면, 입력단으로부터 상대적으로 많은 레이어를 거쳐 추출된 피쳐 맵을 압축하면 정확도를 유지할 수 있다.

본 논문에서는 Faster R-CNN X101-FPN 네트워크[5]의 FPN에서 추출한 다중 해상도 피쳐 맵들을 효과적으로 압축하기 위하여 주성분 분석 (Principle component analysis, PCA)[6] 기반 변환을 통해 해상도 계층 및 채널 간 중복성을 제거하여 압축할 데이터의 양을 줄이는 방법을 제안한다. 제안하는 방법은 객체 탐지뿐만 아니라, 다양한 머신 비전 태스크에서 공통으로 사용되는 FPN 네트워크의 출력 피쳐 맵들을 효과적으로 압축할 수 있다.

2. 본론

피쳐 맵은 이미지에 분포된 특징을 나타내며 각 채널은 서로 다른 특징을 의미한다. 이때 각 특징은 고유하지 않고 서로 높은 상관관계를 갖기도 하며 피쳐 맵에 채널 간에는 중복성이 발생할 수 있다. 중복성은 압축을 위해 필수적으로 제거되어야 하며 이런 중복성은 PCA 기반 변환을 통해 효과적으로 제거할 수 있다[7]. 기존에 영상 압축 방법의 경우, 기저 벡터를 전송하지 않기 위해 DCT, DST와 같이 고정된 기저 벡터를 사용하는 변환이 주로 사용되지만[8], 이는 공간적 중복성을 제거하는데 효과적이며 피쳐 맵에 채널 간 중복성을 제거하는 방법으로는 적절하지 않다.

피쳐 맵에서 한 개 채널을 하나의 표본으로 보고 PCA를 수행하면 일반적으로 피쳐 맵의 채널 수는 포인트 수보다 적기 때문에 채널 개수만큼의 기저 벡터와 한 개의 평균 벡터가 추출된다. 추출된 기저 벡터 중 표본을 변환하였을 때 분산이 높은 기저 벡터들을 선택하여 피쳐 맵을 변환하면 각 기저 벡터에 대한 피쳐 맵에 변환 계수를 얻을 수 있다. PCA를 기반으로 변환을 수행할 경우 역변환을 위해서는 계수뿐 아니라 기저 벡터와 평균 벡터가 필요하다.

| 해상도 계층 | / P5 | / All |
|--------|------|-------|
| P5     | 1    | 0.012 |
| P4     | 4    | 0.047 |
| P3     | 16   | 0.189 |
| P2     | 64   | 0.753 |

표 1. 해상도 계층에 따른 데이터양의 비율

FPN을 거쳐 출력된 피쳐 맵들은 각기 다른 해상도를 갖는 피쳐 맵으로 구성된다. 그림 1과 같이, FPN에서 각 해상도 계층에서 피쳐 맵의 데이터양은 가장 낮은 해상도를 갖는 피쳐 맵 및 전체 피쳐 맵의 데이터양과 비교하면 표 1과 같다. P2-P5 피쳐 맵에 PCA 기반 변환을 수행할 경우 모든 계층이 같은 수의 채널로 구성되기 때문에 변환 계수의 데이터양은 해상도 계층과 관계없이 같지만, 기저 벡터와 평균 벡터의 데이터양은 각 해상도 계층에서 피쳐 맵의 데이터양과 같은 비율을 갖는다. 앞서 언급한 것과 같이, 일반적으로 피쳐 맵에서 포인트의 수는 채널의 수보다 크기 때문에 PCA 기반 변환을 수행하였을 때 기저 벡터의 데이터양은 변환 계수와 평균 벡터의 데이터양보다 많다. 따라서 저해상도 계층에 피쳐 맵의 기저 벡터를 사용하여 고해상도 계층의 피쳐 맵의 기저 벡터를 예측할 수 있다면 매우 높은 비율로 데이터양을 줄일 수 있다.

그림 1에 FPN 구조와 같이 같이 가장 낮은 해상도 계층의 피쳐 맵을 제외한 각 해상도 계층에 피쳐 맵은 그보다 한 단계 낮은 해상도 계층의 피쳐 맵을 업 샘플링 (Up-sampling)하여 구성한다. 따라서 FPN의 출력 피쳐 맵 또한 계층 간에 중복성을 갖는다. PCA 기반 변환을 이용하여 FPN의 피쳐 맵을 압축할 때, 해상도 계층 간 중복성을 제거하기 위해 FPN에서 피쳐 맵을 생성하는 방법과 같이 낮은 계층의 기저 벡터를 업 샘플링 하여 높은 해상도 계층의 기저 벡터를 예측하여 사용할 수 있다. 본 논문에서는 PCA 기반 변환을 수행했을 때 데이터양의 대부분을 차지하는 P2 계층 피쳐 맵의 기저 벡터를 P3 계층 피쳐 맵의 기저 벡터로 예측하여 사용함으로써 데이터양을 압축하는 방법을 제안한다. 제안하는 압축 방법의 블록도는 그림 2와 같다.

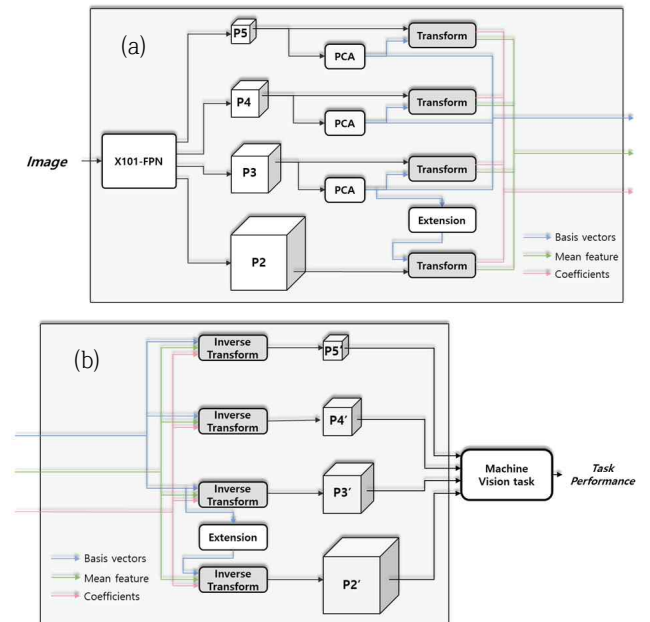


그림 2. 제안하는 압축 방법 (a) 부호화 블록도 (b) 복호화 블록도

그림 2-(a)와 같이 P5, P4와 P3 계층에 피쳐 맵은 PCA 기반 변환을 수행하며 P2 계층에 피쳐 맵을 변환하기 위해 P3 계층에 피쳐 맵의 기저 벡터를 업 샘플링 하여 예측한다. 변환 계수는 그 분포가 0 근처일 때 압축이 용이하므로 P2 계층의 피쳐 맵은 평균을 차감하고 예측된 기저 벡터로 변환한다. 제안하는 압축 방법의 복호화 블록도는 그림 2-(b)와 같다. P5, P4와 P3 계층에 피쳐 맵은 각각의 기저 벡터, 변환 계수 그리고 평균 벡터로 역변환하고 P2 계층의 피쳐 맵은 변환 계수와 평균

벡터 그리고 P3 계층 피쳐 맵의 기저 벡터를 업 샘플링한 기저 벡터로 역변환한다.

### 3. 실험 방법 및 실험 결과

본 논문에서 제안하는 방법의 성능 평가를 위해 ResNexXt-101 기본 네트워크와 P5~2의 해상도 계층을 갖는 FPN으로 구성되며 객체 검출 태스크를 수행하는 Facebook의 Detectron2[9] 네트워크를 사용하였다. Detectron2는 COCO 데이터 셋을 이용하여 학습되었으며 실험은 COCO validation 2017 데이터 셋 중 100장을 선택하여 진행하였고 추가적인 학습은 수행하지 않았다. PCA 기반으로 변환하였을 때 데이터의 양을 조절하기 위해 전체 기저 벡터 중 일정 비율의 기저 벡터를 변환에 이용할 것인지를 결정하는 파라미터  $r$  ( $0 < r \leq 1$ )을 사용하였다. 성능의 비교를 위해 PCA 기반 변환만을 수행했을 때, P5~P3 해상도 계층에 피쳐 맵을 각각  $r$ 만큼의 기저 벡터를 사용하고 P2 해상도 계층에 피쳐 맵 기저 벡터는 P3 계층에 피쳐 맵 기저 벡터로 예측하여 사용했을 때(제안 방법 1), 마지막으로 P5와 P4 해상도 계층에 피쳐 맵은  $r$ 만큼의 기저 벡터를 사용하고 P3 해상도 계층에 피쳐 맵은  $2r$ 만큼의 기저 벡터를 사용하였을 때(제안 방법 2)에 성능을 실험하였다. 압축 효율을 확인하기 위해 데이터양 대비 객체 검출 태스크의 mAP@50 성능을 실험하였으며 그 결과는 표 2와 그림 3과 같다.

| $r$  | 원본 피쳐 맵 |            | PCA 기반 압축 |           |
|------|---------|------------|-----------|-----------|
|      | mAP (%) | 데이터양       | mAP (%)   | 데이터양      |
| 0.10 |         |            | 32.24     | 1,977,374 |
| 0.12 |         |            | 34.76     | 2,343,744 |
| 0.14 | 42.53   | 18,496,000 | 36.00     | 2,710,114 |
| 0.16 |         |            | 38.27     | 3,076,484 |
| 0.18 |         |            | 39.49     | 3,442,854 |

| $r$  | 제안 방법 1 |         | 제안 방법 2 |           |
|------|---------|---------|---------|-----------|
|      | mAP (%) | 데이터양    | mAP (%) | 데이터양      |
| 0.10 | 30.88   | 562,074 | 36.74   | 915,774   |
| 0.12 | 33.34   | 657,344 | 37.27   | 1,080,704 |
| 0.14 | 34.69   | 751,714 | 37.72   | 1,258,746 |
| 0.16 | 35.51   | 846,984 | 38.23   | 1,424,676 |
| 0.18 | 36.20   | 940,454 | 37.99   | 1,589,606 |

표 2. 기저 벡터 사용비율에 따른 객체 검출 성능과 데이터양

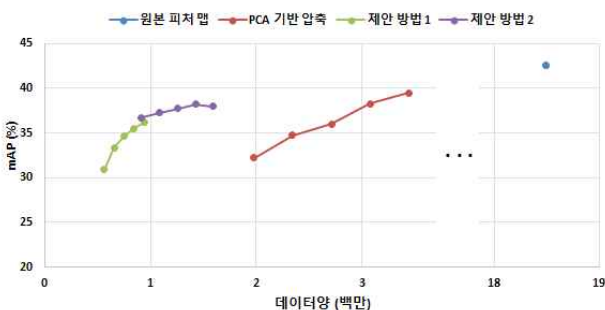


그림 3. 기저 벡터 사용비율에 따른 객체 검출 성능과 데이터양

실험을 통해 제안하는 방법으로 피쳐 맵을 압축하였을 때 원본 피쳐 맵 대비 데이터양은 약 92.3%에서 최대 97.0%까지 감소하였으며 이때 머신 비전 태스크에서 성능은 대략 4.3%p에서 11.65%p까지 감소한 것을 확인하였다. 또한, 기저 벡터의 예측 없이 PCA 기반 변환만을 이용해 피쳐 맵을 압축하는 방법과 비교하였을 때 객체 검출 성능을 유지하며 데이터양을 절반 아래로 감소시키는 효과를 확인하였다.

### 4. 결론 및 향후 연구 방향

본 논문에서는 VCM을 위한 압축 방법 중 피쳐 맵을 압축하기 위해 피쳐 맵을 압축하는 방법을 제안하였다. 제안된 방법을 사용하여 머신 비전 태스크에서 큰 성능 하락 없이 약 92.3%만큼 데이터양을 감소시키는 것을 확인할 수 있었다. 다만, 본 논문에서 제안하는 방법만으로는 충분한 압축 효율을 기대할 수 없으며 추가적인 양자화, 엔트로피 부호화 등의 압축 방법이 연구되어야 한다. 따라서 향후 이 방법과 함께 더욱 압축 효율을 높일 방안을 연구할 예정이다.

### 5. 감사의 글

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원 (No.2020-0-00011, (전문연구실) 기계를 위한 영상부호화 기술) 및 한국연구재단의 지원을 받아 수행된 기초연구사업 (NRF-2021R1A2C2092848)의 지원을 받아 수행된 연구임

### 6. 참고문헌

- [1] ISO/IEC JTC1/SC29/WG02/w20705, "Use cases and requirements for Video Coding for Machines," Aug. 2021.
- [2] ISO/IEC JTC1/SC29/WG02/w20706, "Draft Evaluation Framework for Video Coding for Machines," Aug. 2021.
- [3] ISO/IEC JTC1/SC29/WG02/w20708, "Collection of Evidences Table for Video Coding for Machines," Aug. 2021.
- [4] T. Lin et al., "Feature Pyramid Networks for Object Detection," CVPR, 2017
- [5] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE, VOL. 39, NO. 6, June 2017
- [6] S. Wold, K. Esbensen, and P. Geladi, "Principal Component Analysis," Chemometrics Intell. Lab. Syst., vol. 2, no. 1, pp. 37-52, Aug. 1987
- [7] ISO/IEC JTC1/SC29/WG02/m54414, Hansol Choi et al., "[VCM] A result of feature data reduction using PCA for object detection," Oct. 2020
- [8] J. Mok, J. Ma, Y. Ahn, D. Sim, "Implementation of Fast HEVC Inverse Transform using AVX2 Instruction Set," The Korean Society of Broad Engineers, July, 2015
- [9] Detectron2: <https://github.com/facebookresearch/detectron2>