

신경망 특징맵 부호화를 위한 특징맵 재배열 방법

*한희지 **곽상운 **윤정일 **정원식 **서정일 *최해철[†]

*한밭대학교 정보통신전문대학원 멀티미디어공학과

**한국전자통신연구원

choihc@hanbat.ac.kr[†]

Feature map reordering for Neural Network feature map coding

Heeji Han, Sangwoon Kwak, Joungeil Yun, Won-Sik Cheong, Jeongil Seo and Haechul Choi

Hanbat National University, Graduate School of Information and Communications
Multimedia Engineering

Electronics and Telecommunications Research Institute

요약

최근 IoT 기술이 대중화됨에 따라 커넥티드 카, 스마트 시티와 같은 machine-to-machine 기술의 활용 분야가 다양화되고 있다. 이에 따라, 기계 지향 비디오 처리 및 부호화 기술에 대한 연구분야에 산업계와 학계의 관심 역시 집중되고 있다. 국제 표준화 단체인 MPEG은 이러한 추세를 반영하여 기존 비디오 부호화 표준을 개선할 새로운 표준을 수립하기 위해 Video Coding for Machines (VCM) 그룹을 구성하여 기계 소비를 대상으로 하는 비디오 표준의 표준화를 진행하고 있다. 이에 본 논문에서는 VCM이 기계 소비를 대상으로 진행하고 있는 특징맵 부호화의 부호화 효율을 개선하기 위해 특징맵을 시간적, 공간적으로 재정렬 하는 방법을 제안한다. 실험 결과, 제안 방법이 CityScapes의 검증 세트 내 일부 이미지에 대해 시간적 재정렬을 수행한 결과 random access 조건에서 최대 1.48%의 부호화 효율이 향상됨이 확인되었다.

1. 서론

최근 IoT 기술이 대중화됨에 따라 CCTV, 의료 모니터링, 수송 및 포장 등의 machine-to-machine (M2M) 활용 분야가 다양화되어 기계 지향 비디오 처리 및 부호화 기술에 관한 연구 분야가 급성장하고 있다. Cisco가 발표한 2020년 연례보고서에는 M2M이 2023년에 전 세계 연결 디바이스 및 회선에서 차지하는 비중이 절반을 차지할 것으로 전망하고 있으며, 비디오 응용프로그램의 배포 증가로 인해 M2M 연결의 트래픽이 연결 수보다 급증하고 있는 것으로 추정했다. 이는 기계를 위한 비디오 데이터의 최적화가 급증하는 트래픽 문제의 해결방안이 될 수 있음을 의미한다[1]. 또한, 오늘날의 비디오는 대부분 기계에 의해 분석되며 기계의 분석 결과를 인간이 소비하는 형태를 띠고 있는데, 이는 H.264(AVC), H.265(HEVC), H.266(VVC)와 같은 이전 세대의 비디오 부호화 표준에서 지향하는 바와 확연한 차이를 보인다[2]-[4]. 이에 따라, MPEG은 2019년 7월 스웨덴 예테보리에서 열린 제127차 MPEG 회의에서 이전 세대의 비디오 부호화 표준을 개선할 새로운 표준을 수립하기 위해 Video Coding for Machines (VCM) 그룹을 구성하였다[5]. VCM 표준화에서는 2020년 10월 온라인에서 진행된 회의의 결과로 기술조사요청서(Call for Evidence, CfE) 문서의 초안을 발간하였으며, 차기 회의에서는 CfE 응답에 대한 평가를 진행할 예정이다[6]. 공표된 VCM 표준화의 CfE 문서에는 세부적인 진행일정과 함께 전체 시스템 구조, 평가를 진행할 주요 임무와 평가 방법 등이 명시되어 있다[7].

이전 세대의 비디오 부호화 표준이 인간이 비디오를 소비하는 것에 중점을 두는 것과 달리 VCM 표준화에서는 비디오를 소비하는 대상이 인간과 기계 각각인 경우와 두 가지 대상을 모두 포함하는 경우 모두를 만족시키기 위한 총 세 가지 파이프라인을 제시하고 있다[8]. 현재까지의 연구는 기계를 소비 대상으로 하는 파이프라인에 대한 것이 대부분이다. 파이프라인에서 VCM의 부호화 과정은 입력된 이미지 또는 이미지에서 추출된 특징에 대해 Ffmpeg[9]을 이용한 데이터 전처리 과정(resolution scaling, format conversion)을 수행하고 VVC와 같은 기존 비디오 코덱을 이용해 압축하여 전송하는 순서로 진행되는 것이 일반적이며, VCM의 앵커 역시 이 파이프라인 구조를 따른다.

기존 비디오 코덱의 예측 기술은 공간적 또는 시간적으로 인접한 화소 간의 중복성을 제거하여 잔차 신호의 크기를 줄임으로써 압축 효율을 높였다. VCM의 주요 임무에 대한 특징맵은 대부분 컨볼루션 신경망의 2차원적 가중치 구조로 인해 유사하거나 반복적인 특징이 같은 특징 공간에서 산발적으로 발생하는 형태로 표현된다. 따라서 본 논문에서는 머신비전 임무 수행 시 특징맵의 채널 간 유사성을 기반으로 특징맵을 시간적 또는 공간적으로 재배열함으로써 부호화 효율을 높이는 방법을 제안한다.

2. 특징맵 재배열

본 장에서는 제안 방법인 특징맵 재배열 과정에 대해 소개한다. 제

안 방법은 VCM의 부호화 파이프라인 중 입력 이미지에서 추출된 특징에 대한 데이터 전처리 과정 전에 수행된다. 먼저, 입력 이미지로부터 추출된 몇 가지 특징맵 중 재배열하고자 하는 특징맵을 선택하여 정보를 획득하는 특징맵 정보 획득 단계가 수행되고, 이후 특징맵의 채널간 유사도를 비교하여 인덱싱하는 특징맵 분류 단계, 특징맵 데이터를 인덱스에 따라 실제로 재배열하는 특징맵 재배열 단계, 마지막으로 부호화를 위한 특징맵의 실수 정수 변환 단계 순서로 수행된다.

특징맵 정보 획득 단계에서는 입력 이미지에 대해 추출된 특징맵 중 재배열할 특징맵을 가로, 세로 크기가 각각 n, m 이고 k 개의 채널을 가지는 특징맵으로 정의하고 k 개의 채널 각각에 대하여 $0 \sim k-1$ 의 채널 인덱스를 부여한다. 이후 수행되는 특징맵 분류 단계는 그림 1과 같이 유사한 특성을 가지는 채널간 채널 그룹을 구성하고 각 그룹 내에서 채널간 유사도를 판단하여 유사도에 따라 채널 인덱스를 새로 부여한다.

특징맵 재배열 과정에서는 새로 부여된 채널 인덱스에 따라 특징맵의 데이터를 시간적 또는 공간적으로 재배열 될 수 있다. 재배열 과정에 따라 VVC 인코더의 입력 데이터 형태가 결정된다. 시간적 재배열은 특

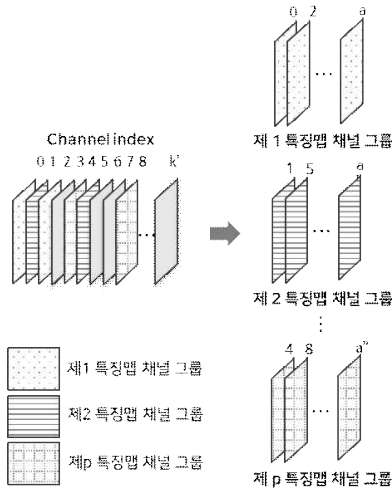


그림 1. 특징맵 분류 과정

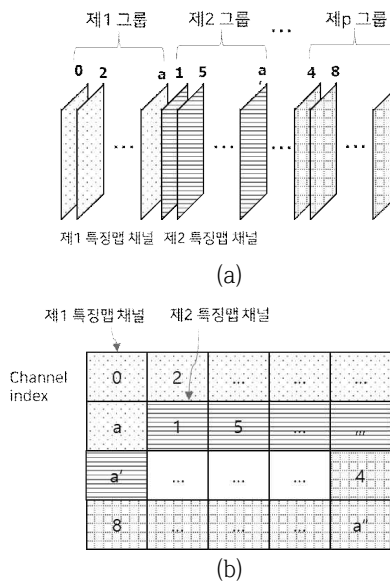


그림 2. 특징맵의 (a) 시간적 재배열과 (b) 공간적 재배열

징맵을 유사도에 따라 $n \times m$ 크기의 k 개 프레임을 갖는 시퀀스 형태로 재배열 하는 것이고, 공간적 재배열은 $n \times m$ 크기의 k 개 특징맵을 공간적으로 재배열한 한 장의 시퀀스 형태로 재배열하는 방법이다. 그림 2(a),(b)는 특징맵을 시간적, 공간적으로 재배열한 형태를 나타낸 것이다. 이와 같이 재배열된 특징맵은 양자화 과정을 수행하게 되고 양자화된 특징맵은 VVC 파이프라인에 따라 전처리 과정을 거친 후 최종적으로 VVC 인코더에 입력으로 들어가게 된다.

3. 실험결과

본 장에서는 제안 방법을 통해 재배열된 특징맵을 부호화한 결과를 부호화 효율 측면과 정확도 측면에서 설명한다. 실험에는 VCM의 주요 임무 중 객체 탐지 임무에 사용되는 Faster R-CNN (X101+FPN)의 네트워크 구조가 사용되었으며 재배열 대상 특징맵으로는 FPN (Feature Pyramid Network) 구조의 C2 특징맵을 선택하였다. 실험에 사용되는 이미지는 VCM의 주요 임무에 사용되는 Cityscapes의 검증 세트에 포함된 이미지(2048x1024) 중 임의의 3개 이미지가 사용되었다.

실험은 VVC 참조 소프트웨어 버전은 VTM 9.0으로 앵커의 경우 참조 소프트웨어를 기반으로 공통 실험 조건인 All Intra (AI), Random access (RA) 구성에 대하여 Quantization Parameter가 22, 27, 32, 37인 조건으로 진행되었으며, 2장에서 설명한 바와 같이 시간적, 공간적으로 재배열 된 특징맵을 하나의 시퀀스 형태로 구성하여 전체 채널에 대해 부호화하는 형식으로 실험을 진행하였다.

표 1. 시간적, 공간적 재배열 방법의 실험결과

Test condition		Y BD-rate	
Test	Image	AI	RA
temporal (256x336x168)	frankfrut	0.00%	-1.10%
	lindau	0.00%	-1.36%
	munster	0.00%	-1.99%
average (temporal)		0.00%	-1.48%
spatial (1x5376x2688)	frankfrut	-0.11%	-0.11%
	lindau	-0.11%	-0.13%
	munster	-0.16%	-0.12%
average (spatial)		-0.13%	-0.12%
average (all)		-0.06%	-0.80%

실험 결과 특징맵 재배열 방법은 입력 이미지에 대한 원본 특징맵을 부호화 한 것과 비교하여 Y성분에 대해 AI 조건에서 평균 0.06%, RA 조건에서 평균 0.80%의 BD-rate 감소를 보였다. 그 중 시간적 재배열 방법은 AI, RA 조건에서 각각 0.00%, 1.48%의 BD-rate 감소를 보였고 공간적 재배열 방법은 AI, RA 조건에서 각각 0.13%, 0.12%의 BD-rate 감소를 보였다. 본 실험은 특징맵 부호화 시 특징맵을 시간적으로 재배열하는 것이 최대 부호화 효율을 달성할 수 있음을 보였다.

4. 결론

본 논문은 특징맵에서 유사하고 반복적인 특징을 가지는 채널이 산발적으로 발생한다는 특징을 이용하여 특징맵 부호화 시 특징맵을 재배열하는 방법을 제안하였다. 특징맵 재배열 방법으로 시간적 재배열과 공

간직 재배열 방법을 제안하였으며, 실험결과 특징맵을 시간적으로 재배열할 때 RA 조건에서 최대 1.48%의 부호화 성능이 향상되었다. 또한, 다른 네트워크의 특징맵에 대해서도 동일한 방법으로 적용 가능하기 때문에 제안 방법이 특징맵 부호화가 적용되는 여러 임무 수행 과정에서 활용될 수 있을 것으로 보인다.

감사의 글

본 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2020-0-00011, 기계를 위한 영상 부호화 기술 개발)

참고 문헌

- [1] Cisco, "Cisco Annual Internet Report (2018-2023) White Paper", Mar. 2020.
- [2] ISO/IEC JCT1/SC29/WG11 and ITU-T SC16, "Advanced video coding for generic audio-visual services, ITU-T rec. H.264 and ISO/IEC 14496-10 (AVC)", May 2003.
- [3] ISO/IEC JCT1/SC29/WG11 and ITU-T SC16, "ISO/IEC 23008-2 MPEG-H Part 2 and ITU-T H. 265", Jan, 2013.
- [4] ISO/IEC JCT1/SC29/WG11, "Versatile Video Coding (Draft 9)", JVET-R2001, Apr. 2020.
- [5] Y. Zhang, "Video Coding for Machines", ITU Workshop on "The future of media", Oct. 2019.
- [6] ISO/IEC JCT1/SC29/WG2, "Friday report", Oct. 2020.
- [7] ISO/IEC JCT1/SC29/WG2, "Call for Evidence for Video Coding for Machines", m55605, Oct. 2020.
- [8] ISO/IEC JCT1/SC29/WG2, "Evaluation Framework for Video Coding for Machines", m55604, Oct. 2020.
- [9] FFmpeg, "version 4.2.2". FFmpeg Git., Jan. 2020.