RoI 추출 방법에 따른 기계를 위한 영상 압축 성능 비교

이예지, 김신, *윤경로
건국대학교
*yoonk@konkuk.ac.kr

Comparison of Image Compression Performance based on RoI Extraction Methods for Machines Vision

Yegi Lee, Shin Kim, Kyoungro Yoon
Konkuk University

요 약

기존 RDO(Rate Distortion Optimization) 기반 압축 방식은 압축 성능에 초점을 두기 때문에 영상 내 인지 특성이 무시될 수 있다. 따라서 RoI(Region of Interest)을 기반으로 압축률을 조절하는 연구가 고안[1, 2, 3, 4] 되었으며, HVS(Human Visual System) 관점에서 영상 내 중요한 부분에 대해 더 높은 품질로 영상을 압축하는 연구가 대부분이다. 최근 인공지능 기술이 발전함에 따라 지능형 영상 분석에 대한 수요가 증가하고 있으며, 이에 따라 머신 비전을 위한 영상 부호화 및 효율적인 전송에 대한 필요성이 대두되고 있다. 본 논문에서는 VVC(Versatile Video Coding)의 dQP(delta Quantization Parameter)를 활용하여 RoI(Region of Interest) 기반 압축 방법을 제안하고, 두가지의 RoI 추출 방식을 소개한다. Detectron2 Faster R-CNN X101-FPN [5]의 첫번째 탐지기를 통해 후보 영역 기반 RoI 을 추출하고, 두번째 탐지기를 통해 객체 기반 RoI 을 추출하여, 영상 내 객체 부분과 비객체 부분으로 나누어 서로 다른 압축률로 압축을 수행하였으며, 이에 따른 성능을 비교하고자 한다.

1. 서론

HEVC(High Efficiency Video Coding)나 VVC 같은 RDO 기반 압축 방식은 압축 성능에만 초점을 두어 최적화를 수행하기 때문에 영상 내 인지 특성이 무시 될 수 있다. 스포츠 영상이나 감시 영상에 경우, 사람은 움직이는 사람이나 사물에 더 많은 관심을 가진다. 또한 대화형 영상에 경우, 사람은 얼굴 영역과 같은 특정한 영역에 관심을 가진다. 따라서 영상의 주관적 품질은 높이기 위해 RoI 를 기반으로 CTU(Coding Tree Unit) 또는 CU(Coding Unit) 단위로 영상 내 압축률을 조절하는 연구가 고안되었으며, 활발히 연구되고 있다.

최근 인공지능 기술이 발전하고, 감시, 자율주행자동차, 스마트 시티 등 다양한 분야에서 딥러닝 기술이 적용됨에 따라 사람의 콘텐츠 소비의 대한 수요보다 머신을 위한 지능형 영상 분석에 대한 수요가 높아지고 있다. 이에 따라 머신 비전을 위한 영상 부호화 및 전송에 대한 필요성이 대두되고 있으며, MPEG(Moving Picture Experts Group)에서는 이러한 요구사항에 따라 VCM (Video Coding for Machine) 그룹을 구성하여 영상 압축 및 영상 특징(feature) 압축에 대한 표준화가 진행중이다. 본 논문에서는 기존 사람의 인지 품질을 높이기 위한 RoI 기반 압축 방식에서 아이디어를 착안하여, 머신의 임무 수행을 위한 RoI 기반 압축 방식을 제안한다.

본 논문에서는 RoI 기반 압축 방식으로 VVC 의 dQP 를 활용하여 CU 단위로 영상 내 압축률을 조절한다. 또한 RoI 추출을 위한 도구로는 detectron2 의 Faster R-CNN 을 사용하여, 첫번째 탐지기를 통해 후보영역기반 RoI 을 추출하고, 두번째 탐지기를 통해 객체 기반 RoI 을 추출한다. 그 후 추출된 영역에

대해 영상 내 객체 부분과 비객체 부분으로 나누어 서로 다른 압축률로 압축을 수행하였으며, 이에 따른 성능을 비교하고자 한다.

본 논문의 구성은 다음과 같다. 2 장 1 절에서는 실험 환경에 대해 살펴본 후, 2 장 2 절에서는 RoI 추출 방법에 대해 서술한다. 2 장 3 절에서는 RoI 기반 압축 방식에 대해 서술하며, 3 장에서는 제안한 방법에 실험 결과를 확인한다. 마지막으로 4 절에서는 본 논문에 대한 결론을 맺는다.

2. 제안 방법

2-1. 실험 환경

실험에 사용된 데이터 세트는 TVD(Tencent Video Dataset)[6]와 FLIR[7]로 VCM 앵커로 채택된 데이터 세트이다. TVD 데이터 세트는 1920x1080 해상도의 166 개의 RGB 이미지로 이루어져 있으며, 객체 탐지 및 분할에 대한 annotation 을 제공한다. FLIR 데이터 세트는 640x512 해상도의 300 장의 IR(InfraRed) 이미지로 이루어져 있으며, 객체 탐지에 대한 annotation 을 제공한다.

VCM 에 평가체제 문서[8]에 따라 객체 탐지 신경망은 detectron2 의 Faster R-CNN X101-FPN 을 사용하였으며, 객체 분할 신경망은 Mask R-CNN X101-FPN 을 사용하였다. 이미지 부/복호화 도구는 VTM(VVC Test Model) 12.0 을 사용하였으며, 부호화 시 VCM 평가 체제에 따라 all intra, 10 비트로 압축을 수행하였다. 이미지 포맷 변환 및 해상도 조절은 FFMPEG 4.2.2 를 사용하였다.

2-2. RoI 추출 방법

객체 탐지 및 객체 분할과 같은 임무 수행에 있어 머신은 사람이나 자동차 같은 객체 영역에 대해 관심이 있다. 따라서 본 논문에서는 객체 탐지 신경망인 Faster R-CNN X101-FPN 이용하여 RoI 추출 방법을 제안한다. Faster R-CNN 은 대표적인 2-stage detector 로 첫번째 탐지기에서 물체가 있을 만한 후보 영역(Region proposal)을 추출[9]하고, 두번째 탐지기에서 최종적으로 객체 분류(Classification)를 수행한다. 본 논문에서는 이러한 신경망의 특성을 활용하여 두가지의 RoI 추출 방법을 제안한다.

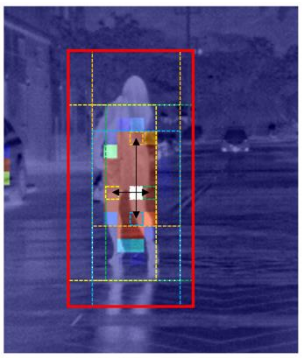

그림 1. 후보영역에서 RoI 추출 방법

RoI 영역을 추출하는 첫번째 방법은 후보 영역 단계에서 RoI 을 추출하는 것이다. 이는 [9]에서 서술한 방법과 같으며, 그림 1 은 FLIR 데이터 셋 내 이미지에서 P4 특징 맵에 대해 objectness map 을 나타낸 결과이다. 백본 신경망으로부터 받은 P2 부터 P6 까지의 특징 맵 내 픽셀들은 각 위치마다 앵커를 이용하여 후보 영역을 생성한다. 본 논문에서는 최종 객체 탐지 결과에서 역추적하여 후보영역에서 탐지된 영역을 찾고, 좌, 우, 상, 하로 이동하면서 물체가 있을 만한 확률(objectness score)이 99% 일 때까지 관심영역을 확장한다.

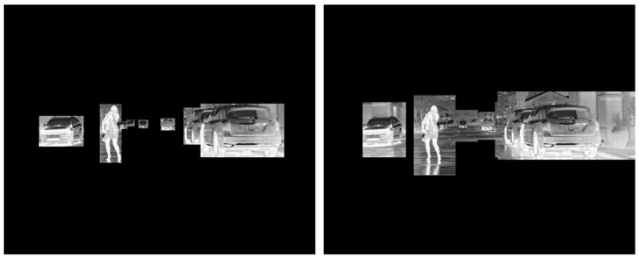

그림 2. RoI 추출 방법에 따른 객체 영역 비교
(좌: 객체 분류 기반, 우: 후보 영역 기반)

두번째 방법은 두번째 탐지기인 객체 분류 단계에서 신경망의 결과로 나온 객체 탐지 결과을 RoI 영역으로 간주하는 방식이다. 그림 2 는 RoI 추출 하는 방법에 따른 객체 영역을 비교한 그림이다. 객체 분류 단계에서 RoI 을 추출하는 경우 객체가 있는 영역만 추출되지만 후보 영역 단계에서 추출하는 경우 앵커 박스에 크기와 객체가 있을 확률에 따라 계산되기 때문에 실제 객체 영역보다 크게 추출된다.

2-3. VVC 기반 RoI 압축 방법

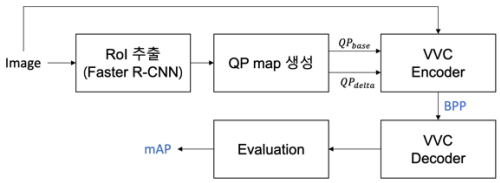

그림 3. VVC 기반 RoI 압축 방법

본 논문에서는 2-2 에서 추출한 RoI 을 기반으로 VVC 의 dQP 를 이용하여 영상 내 압축률을 조절하는 방법을 제안하며, 이는 그림 3 과 같다. 이미지를 입력으로 2-2 절에서 서술한 방법을 통해 영상 내 RoI 영역을 추출하고 이에 따라 각 픽셀 별 QP 맵을 생성한다. QP 맵 생성 시 각 픽셀 별 객체 영역은 $QP_{base} \in \{22, 27, 32, 37, 42, 47\}$으로 설정하였으며, 비객체 영역은 $QP_{delta} = QP_{base} + 15$ 로 설정하였다. 이렇게 생성된 QP 맵은 입력 이미지와 함께 VVC 인코더의 입력으로 들어가며 압축이 수행된다. 부호화 시 CU 영역 내의 QP_{base}픽셀이 하나라도 있는 경우 CU 영역 전체를 QP_{base}로 압축하였으며, 그 외의 경우에는 QP_{delta}로 압축 하였다. 그 후 VVC 디코더를 통해 복호화 후 각 데이터 세트에 맞는 객체 탐지 및 객체 분할 신경망을 사용하여 머신의 성능을 측정하였다.

3. 실험 결과

본 논문에서는 VCM 평가 체제 문서에 나와 있는 FLIR 와 TVD 앵커 결과와 VVC 기반 RoI 압축 방법의 성능을 비교하고자 한다. QP 22 에 경우, 100% 해상도의 결과 보다 75%의 결과가 압축 및 머신의 성능이 더 높게 나오고, 머신 성능 결과가 더 큰 범위로 나오기 때문에 75%에 대해서만 성능 비교를 하였다.

표 1 은 FLIR 데이터 세트에 대해 객체 탐지 성능을 비교한 결과이다. 실험 결과, 객체 분류 기반 압축을 사용하면 bpp 가 절반 이하로 떨어지지만 mAP(mean Average Pricision)도 같이 2~3%p 떨어지는 것을 확인 할 수 있다. 반면, 후보영역 기반 압축 방식은 객체 분류 기반보다는 bpp 가 올라가지만 mAP 도 커지는 것을 확인 할 수 있다. 75% 해상도에서 앵커 결과와 두가지의 RoI 기반 압축 기법에 대해 BD-rate 를 계산한 결과, 객체 분류 기반 압축 방식에서는 1.60%으로 압축 성능이 앵커보다 좋지 않은 것을 보여주는 반면에 후보 영역 기반 압축 기법은 BD-rate 가 -2.52%로 앵커보다 압축 성능이 좋은 것을 확인할 수 있다.

표 1. FLIR 데이터 세트 객체 탐지 결과 (75% 해상도)

	앵커		후보 영역 기반		객체 분류 기반	
QP	bpp	mAP	bpp	mAP	bpp	mAP
22	0.886	40.340	0.435	39.591	0.377	38.592
27	0.399	39.641	0.254	38.663	0.231	37.652
32	0.189	36.626	0.149	34.669	0.140	34.530
37	0.098	30.878	0.085	28.603	0.082	27.881
42	0.049	19.025	0.042	17.422	0.041	15.988
47	0.022	7.601	0.019	7.312	0.018	5.951
BD-rate			-2.52%		1.60%	

표 2 은 TVD 데이터 세트에 대해 객체 탐지 성능을 비교한 결과이며, 표 3 은 객체 분할 결과를 보여준다.

표 2. TVD 데이터 세트 객체 탐지 결과 (75% 해상도)

	앵커		후보 영역 기반		객체 분류 기반	
QP	bpp	mAP	bpp	mAP	bpp	mAP
22	0.311	55.913	0.207	54.291	0.191	51.515
27	0.179	52.074	0.127	52.023	0.118	49.898
32	0.098	49.988	0.077	46.451	0.073	44.624
37	0.051	42.668	0.044	39.712	0.043	37.497
42	0.025	28.295	0.022	26.711	0.022	24.584
47	0.012	14.061	0.010	12.306	0.010	13.622
BD-rate			-5.26%		5.08%	

표 3. TVD 데이터 세트 객체 분할 결과 (75% 해상도)

	앵커		후보 영역 기반		객체 분류 기반	
QP	bpp	mAP	bpp	mAP	bpp	mAP
22	0.311	44.163	0.207	45.058	0.191	42.613
27	0.179	41.945	0.127	40.777	0.118	39.771
32	0.098	37.950	0.077	35.940	0.073	35.247
37	0.051	33.226	0.044	31.374	0.043	30.927
42	0.025	21.454	0.022	20.903	0.022	21.550
47	0.012	13.167	0.010	11.990	0.010	13.545
BD-rate			-4.87%		-4.65%	

TVD 데이터 세트에 대하여 실험 결과, 객체 분류 기반 RoI 압축 방식에 대하여 객체 탐지에서 5.08%, 객체 분할에서 -4.65%의 BD-rate 결과가 나왔으며, 후보 영역 기반 RoI 압축

방식에 대하여 객체 탐지에서 -5.26%, 분할에서는 -4.87%의 BD-rate 결과가 나왔다.

RoI 기반 압축 방식은 앵커와 비교했을 때 bpp 가 크게 떨어지지만 dQP 의 영향으로 배경 부분 즉, 비객체 부분에 대해 mAP 가 앵커 결과보다는 작게 나온다. 또한 객체 분류 기반 RoI 압축 방식은 객체 탐지 결과에 따라 RoI 영역을 구분하기 때문에 후보 영역 기반 RoI 방식 보다는 mAP 가 떨어지는 것을 확인할 수 있다. BD-rate 의 경우, bpp 가 앵커 결과보다 약간 높게 나오지만 RoI 영역을 실제 객체 크기보다 크게 잡은 후보 영역 기반 RoI 압축 방식 앵커보다 약 5% 압축 성능이 좋은 것을 확인 할 수 있다.

4. 결론

본 논문에서는 기존 사람의 인지 품질을 높이기 위해 사용된 RoI 압축 방식에서 착안하여 머신의 지능형 영상 분석을 위한 RoI 기반 압축 방식을 소개하였으며, 2-stage detector 인 Faster R-CNN 에서 두 가지의 RoI 추출 방식을 사용하여 압축 성능을 비교하였다. RoI 가 아닌 배경 부분은 QP_{base} 보다 15 만큼 크게 압축을 수행하였기 때문에 앵커보다는 bpp 가 최대 절반 이하로 떨어지지만 mAP 도 약 2~3%p 떨어지는 것을 확인 할 수 있었다. 또한 두 가지의 RoI 압축 방법 중 실제 객체 크기보다 RoI 영역을 크게 잡은 후보 영역 기반 압축 방식은 DB-rate 가 약 -5%로 기존 앵커보다 압축 성능이 좋은 것을 확인할 수 있다.

감사의 글

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

참고 문헌

[1] Zhou, QiLui, Jiaying Liu, and Zongming Guo. "A multilevel region-of-interest based rate control scheme for video communication." MIPPR 2009: Remote Sensing and GIS Data Processing and Other Applications. Vol. 7498. SPIE, 2009.

[2] Zhu, Shiping, and Ziyao Xu. "Spatiotemporal visual saliency guided perceptual high efficiency video coding with neural network." Neurocomputing 275 (2018): 511-522.

[3] Song, Renjie, and Yuandong Zhang. "Optimized rate control algorithm of high-efficiency video coding based on region of interest." Journal of Electrical and Computer Engineering 2020 (2020).

[4] Sun, Xuebin, et al. "Content-aware rate control scheme for HEVC based on static and dynamic saliency detection." Neurocomputing 411 (2020): 393-405.

[5] Detectron2, https://github.com/facebookresearch/detectron2 (accessed May, 23, 2022).

[6] TVD dataset, https://multimedia.tencent.com/resources/tvd, (accessed May, 23, 2022).

[7] FLIR dataset, https://www.flirkorea.com/oem/adas/adas-dataset-form/, (accessed May, 23, 2022).

[8] ISO/IEC JCT1/SC29/WG2, “Evaluation Framework for Video Coding for Machine”, N162, January, 2022.

[9] Kim, Shin, et al. "Compression of thermal images for machine vision based on objectness measure." International Workshop on Advanced Imaging Technology (IWAIT) 2022. Vol. 12177. SPIE, 2022.