물체 기반 비디오 압축

김명준, *이영렬
세종대학교
mjkim@sju.ac.kr, *yllee@sejong.ac.kr

Object based Video Compression

MyungJun Kim Yung-Lyul Lee
Sejong University

요 약

본 논문에서는 YOLO(You Only Look Once) 사물 인식 알고리즘을 활용하여 영상 압축에 적용한다. YOLO 는 물체의 일반화된 특징을 학습한 뉴럴 네트워크이다. 영상을 압축하는 동시에 YOLO 를 활용하여, 영상 내의 사물을 인식한다. 사물이 인식된 영역을 영상 압축을 할 때, 더 구체적으로 예측을 하는 방법을 제안한다. 본 논문에서 제안하는 방법은 QP(Quantization Parameter)를 조절하여, YOLO 로부터 인식된 사물을 더 정교하게 사물을 부호화/복호화한다. VVC(Versatile Video Coding) 기반에서 Rate-Control 를 사용하며, QP 를 조절한다. QP 는 CTU-Level 단위로 조절하며, 사물이 포함된 CTU 는 더 낮은 QP 를 바탕으로 효율적인 화질을 가져온다. 본 논문에서 제안하는 방법은 VVC 기반으로 한 Rate-Control 보다 주관적 화질이 선명한 것으로 보인다.

1. 서론

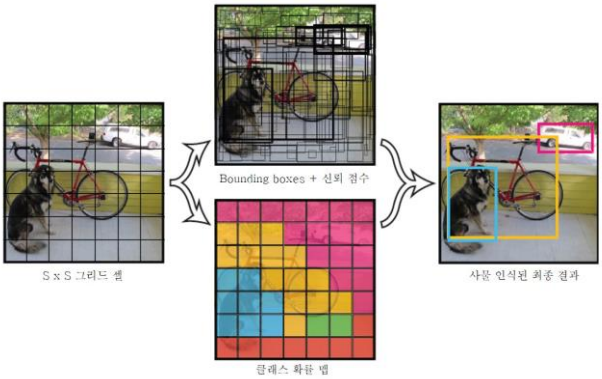
차세대 비디오 부호화(Future video Coding) 표준 개발을 위하여 구성된 ITU-T VCEG(Video Coding Expert Group) 와 ISO/IEC MPEG(Moving Picture Experts Group)의 협력 팀인 JVET(Joint Video Experts Team)에 의해 공동으로 VVC 표준화를 진행하였다. JVET 은 2018 년 4 월에 진행된 제 10 차 샌디에고 회의에서 기술제안요청(CfP, Call for Proposal)에 응답한 23 개의 제안 기술들을 바탕으로 검토되었으며, 그 결과를 바탕으로 VVC 의 규격초안(WD, Working Draft)와 실험모델(VTM, Test Model of VVC)이 나왔다. VVC 는 HEVC 의 QT(Quad-Tree) 모델과 JEM(Joint Exploration Model)과 기관에서 제안한 BT(Binary Tree)와 TT(Ternary Tree) 구조를 결합시켰으며, HEVC 의 기존의 일부 툴을 제거한 뒤 VVC 에 사용하였다. VVC[1]는 HEVC 에서 사용된 기술 이외에 다양한 기술들이 채택되었다. 화면 내 예측에서는 대표적으로 MIP(Matrix-based Intra Prediction), Wide-angle Intra Prediction 등이 개발되었으며, 화면 간 예측 에서는 Affine Prediction 과 BDOF(Bi-Directional Optical Flow)와 Symmetric MVD(Motion Vector Difference) 등이 채택되었다. 변환에서는 기존 HEVC 에서는 DCT(Discrete Cosine Transform)-II 와 DST(Discrete Sine Transform)-VII 커널을 사용하였으나, VVC 에서 DCT-VIII 이 추가되었다. 또한, 양자화에서는 Dependent Quantization 이라는 기술이 추가적으로 채택되었다. 이에 따른 부호화 성능이 VTM7.0 기준으로 HEVC 대비 Random access configuration 에서 35% 비트 절감의 성능 향상을 가져왔으며, 부호화에서는 9 배와 복호화에서는 1.8 배의 속도 증가를 가져왔다[2].

YOLO(You Only Look Once)는 2015 년에 나온 사물 인식 알고리즘이다[3]. 본 논문에서는 YOLO 사물 인식 알고리즘을 바탕으로, 사물이 있는 위치의 영상에 효율적인 화질 향상을 시키는 방법을 제안한다. YOLO 를 활용하여, 영상 내의 사물을 인식하여, CTU-Level 단위로 QP 를 조절하며 효율적인 화질을 가져온다. 제안된 방법은 VVC 시험 모델인 VTM8.2 에서

실험하였으며, 사물이 인식된 블록에 주관적 화질의 향상을 기대해볼 수 있다.

본 논문의 구성은 다음과 같다. 2 절에서는 본 논문에서 YOLO 를 활용하여 VVC 에 적용한 방법과 3 절에서는 실험 조건 및 결과, 4 절에서는 결론을 맺는다.

2. 사물 인식 알고리즘을 활용한 VVC

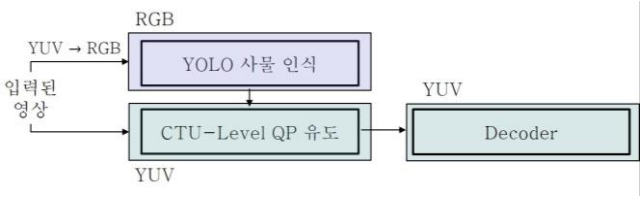


(그림 1) YOLO 모델에 대한 예시[3]

(그림 2) Cactus 영상에 사물 인식한 결과

YOLO 의 네트워크 구조는 이미지 분류를 위한 GoogleNet 모델로부터 영감을 얻었다. YOLO 네트워크는 24 개의 convolutional layers 와 2 개의 Fully-connection layers 로 이루어져, 1x1 reduction layers 를 사용하고 3x3 convolutional layers 를 사용한다.

YOLO 의 기본적인 방식은 (그림 1)과 같이 이미지를 입력받아 여러 개의 Bounding Box 를 회귀하여 최종 물체를 예측한다. (그림 1)에서 이미지의 입력되면 NxN 으로 사이즈를 조정한 후, SxS 그리드 셀(grid cell)으로 나누어 준다. 각각의 사각형 블록은 바운딩 박스(B)의 구성은 바운딩 박스의 x, y 좌표와 가로세로 길이와 박스의 신뢰 점수로 구성되어 있다. 이때, 예측된 박스의 신뢰 점수는 실제 물체와 겹치는 면적에 따라 점수를 구하게 된다. 따라서, 각각의 그리드 셀마다 구한 물체의 바운딩 박스들을 SxSx(B*5+C)의 텐서로 구성하여 NMS(Non Maximum Suppression) 알고리즘을 통하여 최종 결정된 사물 정보를 얻을 수 있다. (그림 2)는 VTM 표준 시퀀스 Cactus 영상에 YOLO 를 통해서 인식한 결과를 나타낸다.



(그림 3) 제안하는 방법의 프레임워크

(그림 3)는 본 논문에서 제안하는 방법의 프레임워크이다. 입력받은 영상을 RGB 칼라 포맷으로 변경 후, 사물 인식을 한다. 인식된 사물의 위치의 CTU 단위의 블록에 (1)의 식과 같이 QP 의 값을 변경해준다.

$$QP' = isObjectDetected\ ?\ QP - THR : QP \qquad (1)$$

사물이 존재하는 경우 *isObjectDetected* 가 1 로 설정이 되어 *QP* 의 값을 임계치(THR)를 설정하여 변경한다. 본 논문에서는 THR 을 2 로 주어 실험을 진행했다.

3. 실험 조건 및 결과

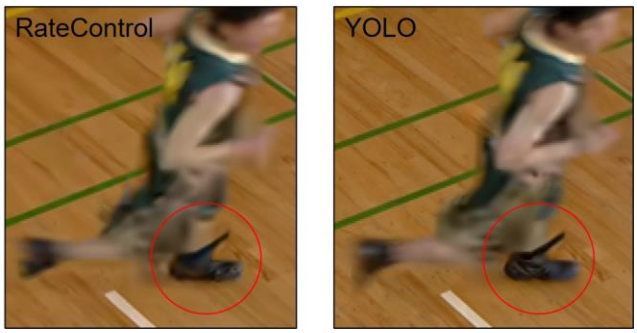
본 논문은 VTM 참조 소프트웨어인 VTM 8.2[4]에서 구현되었다. RA configuration 에 맞추어 실험을 진행하였으며, Rate-Control[5]을 사용하였다. Rate-Control 에서의 첫번째 initalQP 는 실험에 처음 설정되는 QP 를 사용하였다. QP 는 {22, 27, 32, 37}을 사용하였으며, 실험에 사용된 시퀀스들은 클래스 B, C, D 의 1 초 분량이다. 각각의 B, C, D 해상도는 1080p, 832x420, 416,240 에 해당된다. YOLO 는 두번째 버전(YOLOv2)을 활용하였다. 본 실험에서는 YOLO 를 CPU 를 활용하여 실험을 진행하였다. <표 1>은 각각의 실험에서 VTM 대비 인코딩 속도를 나타낸다. <표 1>에서는 Rate-Control 을 킨 상태의 실험과 Rate-Control 과 YOLO 를 접목한 실험을 나타낸 결과이다[6]. 디코딩 속도는 같기 때문에 생략했다. <표 1>에서 클래스 B 와 C 는 할당된 타겟비트(Targetbit)가 VTM8.2 보다 적었기 때문에 더 빠른 속도로 인코딩 되었다. VTM-YOLO 가 VTM-RC 보다 각

클래스에서 약 20% 정도 인코딩 시간을 소모하는 것을 확인할 수 있다.

<표 1>

클래스	VTM-RC	VTM-YOLO
B	90%	95%
C	87%	108%
D	115%	137%
Averages	96%	111%

(그림 4)는 BasketballDrill 영상 13 번째 프레임 비교 예시이다. (그림 4)의 (가)는 VTM8.2 에서 Rate-Control 을 적용한 영상이며, (그림 4)의 (나)는 YOLO 를 적용한 그림이다. 농구선수의 신발을 보면 더 화질이 선명하게 보이는 것을 확인할 수 있다.



(가) (나)

(그림 4) BasketballDrill 영상 13 번째 프레임 비교 예시

4. 결론

본 논문에서는 YOLO 사물 인식 알고리즘을 활용하여 영상 압축에 접목한 방법을 제안하였다. QP 는 CTU-Level 단위로 조절하며, 사물이 포함된 CTU 는 더 낮은 QP 를 바탕으로 효율적인 화질을 가져온다. 본 논문에서 제안하는 방법에서 VTM-YOLO 가 122% 인코딩 시간을 더 소모하는 것으로 보인다. 다음 연구에는 단순히 QP 를 조절하는 것이 아니라 Rate-Control 에서 사용하는 Lamda 식과 관계를 조사하여, 더 효율적인 방법을 제시할 것이다.

감사의 글

이 논문의 일부는 2020 년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. NRF-2018R1D1A1B07045156)

참고문헌

[1] B. Bross, J. Chen, S. Liu, Y.-K. Wang, “Versatile Video Coding (Draft 9),” document JVET -R2001, April. 2020.

[2] F. Bossen, X. Li, K. Sühring, “JVET AHG report: Test model software development (AHG3),” document JVET -Q0003, Jan. 2020.

[3] J. Redmon, S. Divvala, R. Girshick, “You Only Look Once: Unified, Real-Time Object Detection,” The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788.

[4] Y.-H. Chao, Y.-C. Sun, J. Xu, X. Xu, “JVET common test conditions and software reference configurations for non-4:2:0 colour formats,” document JVET -R2013, April. 2020.

[5] T. Chiang, Y.-Q. Zhang, A new rate control scheme using quadratic rate distortion model, IEEE Trans. Circuits Syst. Video Technol. 7 (1) (1997) 246-250.

[6] YOLO, “YOLO: Real-Time Object Detection,” https://pjreddie.com/ (accessed Sept, 2016).