

특집논문 (Special Paper)

방송공학회논문지 제29권 제6호, 2024년 11월 (JBE Vol.29, No.6, November 2024)

<https://doi.org/10.5909/JBE.2024.29.6.879>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

VCM을 위한 공간적 리샘플링 기술에 관한 성능 분석

안은빈^{a)}, 김아영^{b)}, 곽상운^{b)}, 정순흥^{b)}, 정원식^{b)}, 추현곤^{b)}, 서광덕^{a)†}

Performance Analysis on the Spatial Resampling Technology for VCM

Eunbin An^{a)}, Ayoung Kim^{b)}, Sangwoon Kwak^{b)}, Soon-heung Jung^{b)}, Won-Sik Cheong^{b)},
Hyon-Gon Choo^{b)}, and Kwang-deok Seo^{a)†}

요약

최근 괄목할 만한 발전을 이룬 머신 비전은 다양한 산업 및 연구 분야에서 활용되고 있다. 특히, 머신 비전에서 활용되는 영상이 급격하게 늘어나면서 기존 인간의 시각 인지를 기반으로 발전된 비디오 압축 방식이 아닌 기계의 특성을 고려한 새로운 압축 방식인 VCM에 대한 논의가 이뤄지고 있다. 특히, VCM에서 채택된 MI-RPR 기반 적응형 공간적 리샘플링 기술은 입력 영상의 주요 객체 식별 가능성을 고려하기 위하여 사전 분석을 진행하고, 분석된 정보를 토대로 영상의 해상도를 적응적으로 적용하여 부호화하기 때문에 기계 임무의 정확도를 최대한 유지하면서도 비교적 안정적으로 부호화 성능을 높일 수 있다. 본 논문에서는 MI-RPR 기반의 적응형 공간적 리샘플링 기술을 소개하고, 146차 회의 이후에 보고된 CEO 보고서를 기반으로 공간적 리샘플링 기술의 적용이 VCM 성능에 미치는 영향을 상세히 분석한다. 공간적 리샘플링 기술의 활성화를 통해 VCM-RS v0.8.1과 v0.9에서 각각 BD-rate 6.08%, 4.27% 만큼 VCM의 성능 개선 효과를 얻을 수 있다.

Abstract

Recent advancements in machine vision have led to significant applications across various industries and research fields. With the exponential growth in the volume of video data utilized in machine vision, there has been increasing discussion around a new coding technology, Video Coding for Machines (VCM), which is tailored to the machine consumption rather than traditional video compression methods designed for human visual perception. The MI-RPR-based adaptive spatial resampling adopted in VCM conducts pre-analysis to assess the likelihood of effectively identifying key objects in the input video and adaptively applies resolution during encoding based on the analyzed information. This approach allows for maintaining the accuracy of machine tasks while relatively stabilizing and enhancing the coding performance. This paper introduces the adaptive spatial resampling based on MI-RPR and analyzes its impact on VCM performance, as detailed in the summary report of CEO following the 146th meeting. The implementation of this spatial resampling technique results in performance improvements of 6.08% and 4.27% in BD-rate for VCM-RS versions 0.8.1 and 0.9, respectively.

Keyword : VCM, MPEG, Spatial resampling, RPR, Object occupancy distribution

Copyright © 2024 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.”

1. 서론

최근 머신 비전의 비약적인 발전은 비디오의 객체나 이벤트를 신속하게 인식 및 분석 가능하게 함으로써 다양한 응용 분야의 기존 한계를 넘어선 새로운 기술적 도약의 기반이 되고 있다. 이에 따라 전 세계에서 다양한 목적을 가진 방대한 양의 비디오 데이터가 매일같이 생성되고, 또 이를 머신 비전을 통해 분석하여 새로운 데이터로 가공하는 작업이 계속되고 있다. 그러나 수행하는 작업의 종류나 장치의 역할에 따라서 비디오를 수집하는 영상 획득 장치와 머신 비전을 수행하는 장치는 분리된 경우가 많기 때문에 통상적으로 수집된 영상을 부호화하여 전송 또는 저장, 복호화한 뒤에 임무에 활용하는 방법이 고려되고 있다. 가령 감시카메라를 통해 거리를 촬영하고 범죄자를 색출하는 시나리오를 상정하자. 감시카메라의 성능 및 가격, 전력 소모량, 부피 등을 고려했을 때, 범죄자를 탐지하는 행위는 수집된 영상을 분리된 장치로 전달한 뒤에 머신 비전을 수행하는 것이 합리적이다. 또한, 감시카메라로부터 수집된 방대한 양의 영상 데이터를 머신 비전이 수행되는 장치에게 효율적으로 전달하기 위해서는 계획된 머신 비전 작업 성능 저하를 최소화하면서도 최대한 높은 효율로 압축하는 기술이 필요하다.

하지만 기존의 부호화기는 HVS(Human Visual System) 특성을 고려하여 설계되었기 때문에 기계 임무 수행에 필요한 영상 정보가 아님에도 인간 시각 인지에 중요한 특성 정보를 유지하거나 압축률을 높이기 위하여 데이터의 손실을 단순히 늘림으로써, 오히려 기계 인지 능력이 감퇴하는

a) 연세대학교 소프트웨어학부(Division of Software, Yonsei University)

b) 한국전자통신연구원(Electronics and Telecommunications Research Institute)

‡ Corresponding Author : 서광덕(Kwang-deok Seo)

E-mail: kdseo@yonsei.ac.kr

Tel: +82-33-760-2788

ORCID: <http://orcid.org/0000-0001-5823-2857>

* 이 논문의 결과 중 일부는 한국방송·미디어공학회 2024년 하계학술대회에서 발표한 바 있음.

‡ This research was supported by the Institute of Information and Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2020-0-00011, Video Coding for Machine), the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2021R1F1A1048404), and the MIST(Ministry of Science, ICT), Korea, under the National Program for Excellence in SW, supervised by the IITP in 2023 (2019-0-01219).

• Manuscript September 4, 2024; Revised October 31, 2024; Accepted November 1, 2024.

문제가 발생할 수 있다. 따라서 인간이 아닌 기계의 소비를 대상으로 특화된 부호화 기법에 대한 논의가 필요하다. 나아가 이러한 기법에 대한 논의에는 장치나 데이터셋, 머신 비전의 종류 등을 특정하지 않고 포괄적으로 압축 가능한 방식에 대한 고민이 포함된다.

기계에서 소비되는 영상을 대상으로 압축 효율을 높이기 위한 방안을 마련하기 위하여 ISO/IEC JTC1 SC29 산하 MPEG(Moving Picture Experts Group)은 2019년 VCM(Video Coding for Machines) Ad hoc Group(AhG)을 결성하고, 기계 임무 수행에 초점을 맞춘 새로운 비디오 부호화 기술 표준화를 시작했다. 제안된 후보 기술들 사이에서 전통적인 비디오 압축 방식과 특징맵 기반의 압축 방식이 구분됨에 따라, 2021년에 전통적인 비디오 압축 방식 기반의 VCM과 특징맵 기반의 압축 방식을 표준화하는 FCM(Feature Coding for Machines), 두 개의 트랙으로 분리하여 표준화를 계속하고 있다. 현재에 이르러 MPEG VCM은 크게 네 개의 도구 1) 시간적 리샘플링(Temporal resampling), 2) 공간적 리샘플링(Spatial resampling), 3) ROI 기반 처리(ROI-based processing), 4) 비트 깊이 절삭(bit depth truncation) 기술을 채택하였다¹⁻³⁾.

그림 1은 MPEG VCM 부호화기의 기본적인 구성 및 수행 과정을 보여준다. VCM에 입력된 원시 비디오 데이터(YUV)는 앞서 소개한 네 개의 도구를 거친 이후에, 내장 코덱(Inner Codec)을 차례로 수행하여 최종적인 비트스트림을 생성한다. 이때 내장 코덱은 다양한 종류의 부호화기가 사용될 수 있으며, 현재 VCM 참조 소프트웨어는 VVC(Versatile Video Coding)를 활용하고 있다. 시간적 리샘플링으로부터 도출된 결과는 ROI 기반 처리와 비트 깊이 절삭 기술을 거쳐 전처리된 비디오(Preprocessed video)를 생성한다. 또한, 시간적 리샘플링으로부터 도출된 결과에 대하여 공간적 리샘플링 기술의 사전 분석(Pre-analysis) 과정을 수행함으로써 OOD(Object Occupancy Distribution) 기반의 다운스케일링 크기(Optimal Scale Factor)를 결정한다. 도출된 전처리 비디오와 최적 스케일링 크기는 내장 코덱에 전달되어 최종적인 비트스트림을 출력한다^{4,5)}.

특히, 2023년 7월 143차 회의에서는 VVC의 RPR(Reference Picture Resampling)을 머신 소비에 알맞게 사용할 수 있도록 재설계된 MI-RPR(Machine inference-

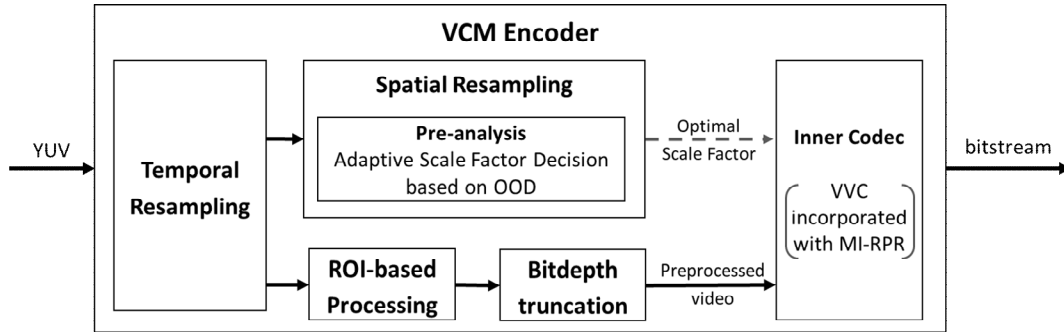


그림 1. MPEG VCM 부호화기의 흐름도
 Fig. 1. Flowchart of MPEG VCM encoder

based RPR)을 이용한 적응형 공간적 리샘플링 기술이 처음 제안되었고, 2024년 1월 145차 회의에서 CE(Core Experiment) 3: Spatial resampling을 통해 최종 채택되었다. MI-RPR 기반의 적응형 공간적 리샘플링 기술은 각 픽처에 대하여 사전 분석을 실시하고, 분석된 결과를 기반으로 픽처의 해상도를 적응적으로 샘플링하기 때문에 기계 임무 완수에 부정적인 영향을 최소화하고 압축의 효율을 높이는 데 있어 비교적 안정적인 성능 효과를 얻을 수 있었다^[6-10].

VCM AhG은 채택된 네 개의 기술 적용에 따른 VCM의 성능 변화를 확인하기 위하여 145차 회의부터 CE0: Tool experiments를 실시하였다. 본 논문에서는 현재 채택된 4개의 기술 중에서 공간적 리샘플링 기술을 소개하고, CE0에 보고된 내용을 토대로 공간적 리샘플링 기술이 VCM의 압축 성능에 미치는 영향을 상세히 분석한다^[11-14].

II. VCM을 위한 공간적 리샘플링 기술

MI-RPR 기반의 적응형 공간적 리샘플링 기술은 그림 2

와 같이 1) OOD 기반 최적 다운스케일 크기 결정을 위한 전처리 과정과 2) MI-RPR이 이식된 VVC를 통한 내부 부호화 과정으로 구성된다. 전처리 과정에서는 각 픽처에 대하여 OOD를 생성하고, 이를 기반으로 픽처마다 최적의 스케일 크기를 도출한다. 반면, 전달된 최적의 스케일 크기가 적용된 다양한 해상도를 가지는 비트스트림을 생성하기 위하여 MI-RPR이 이식된 VVC의 부호화 과정이 수행된다.

1. OOD 기반 최적 다운스케일 크기 결정

기계 임무를 수행함에 있어 영상의 주요 객체 식별은 임무의 성공 여부를 가르는 중요한 요소이다. 영상의 주요 객체 크기가 작은 경우, 객체 사이의 겹침 정도가 심한 경우, 객체를 식별하기 어려울 정도로 영상의 품질이 낮은 경우 등에서는 기계가 주요 객체를 식별 또는 인지할 확률이 낮아지며, 이는 기계 임무의 성능을 크게 저해할 수 있다. 따라서 주요 객체가 식별될 가능성이 높은지를 판단할 수 있다면 기계 임무의 정확도 또한 높을 것임을 유추할 수 있다.

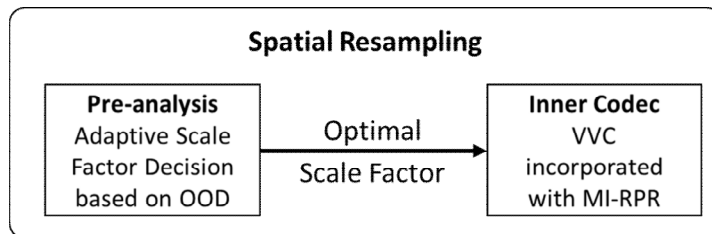


그림 2. VCM 공간적 리샘플링 기술의 개념적 동작 과정
 Fig. 2. Conceptual process of VCM spatial resampling

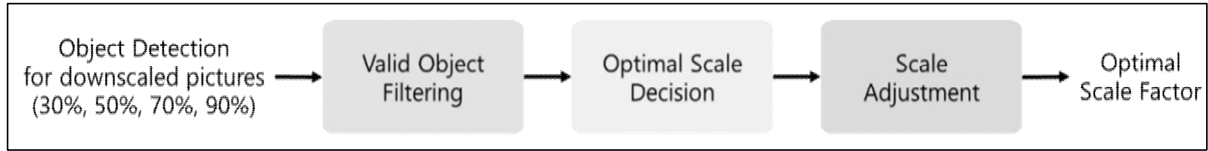


그림 3. VCM의 공간적 리샘플링 기술을 위한 최적 다운스케일링 크기 결정 과정
 Fig. 3. Decision process for the optimal down-scaling factor for VCM spatial resampling

OOD는 사진 분석에서 수행된 객체 탐지 결과를 토대로, 탐지된 모든 객체에 대하여 영상의 해상도 대비 객체가 차지하는 비율의 분포를 의미한다. 따라서 객체의 수와 객체의 크기 정보를 기반으로 생성된 OOD를 이용하면 머신 비전의 성능을 최대한 유지하면서도 공간적 해상도를 최소화할 수 있는 최적 다운스케일링 크기를 결정할 수 있다.

그림 3은 OOD 기반의 최적 다운스케일링 크기 결정 과정을 보여준다. 픽처에 대하여 30%, 50%, 70%, 그리고 90% 해상도로 공간적 다운스케일링을 수행한 뒤에, 다운스케일링된 각 픽처에 대하여 객체 탐지를 수행하고 탐지 신뢰도가 높은 객체를 선별한다. 최적 스케일 판정 과정을 통해 다운스케일링된 픽처에 대한 OOD를 생성하고, 원본 해상도 OOD와 비교하여 유사도가 높은 스케일 크기들을 최적 다운스케일 후보로 설정한다. 이 중에서 가장 작은 후보값이 잠재적인 최적 다운스케일로 선정된다. 마지막으로 양자화 강도를 고려한 스케일 보정 과정을 거친 후에 최종적인 최적 다운스케일링 크기를 도출한다.

2. MI-RPR

VVC는 참조 픽처의 공간적 리샘플링을 제공한다. 그림 4는 기존 VVC RPR의 동작 과정을 개념적으로 보여준다. 기존의 RPR은 입력된 픽처에 대하여 다운스케일링 후 원래 크기로 업스케일링을 수행한 뒤에 원본과 리샘플링된 결과의 PSNR(Peak Signal-to-Noise Ratio)을 계산한다. 만일 PSNR이 충분히 크다면 리샘플링된 결과가 원본의 품질과 큰 차이가 없다고 판단하여 다운스케일링된 픽처를 참조 픽처로 부호화한다. 반면, PSNR이 작다면 리샘플링된 픽처의 품질이 원본의 품질에 비하여 차이가 크다고 판단하고 참조 픽처를 원본 해상도로 부호화한다.

VCM의 공간적 리샘플링은 그림 5와 같이 MI-RPR이 이식된 VVC 부호화기를 내장 코덱으로 사용한다. 인간의 시각 인지를 고려하여 PSNR을 기준으로 RPR 적용 여부를 결정하는 기존의 RPR과 달리, MI-RPR은 앞서 계산된 스케일링 크기를 입력 받아 픽처마다 다양한 해상도를 부여한다. 따라서 MI-RPR이 이식된 VVC 부호화기는 VCM의 목적 달성을 위하여 기계의 성능 저하를 최소화 하면서도

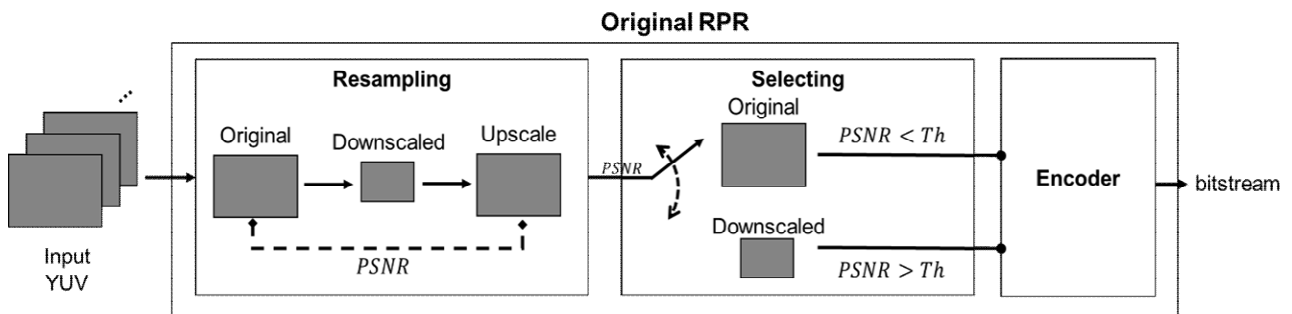


그림 4. RPR 동작 개념도
 Fig. 4. Pipeline of the conventional RPR

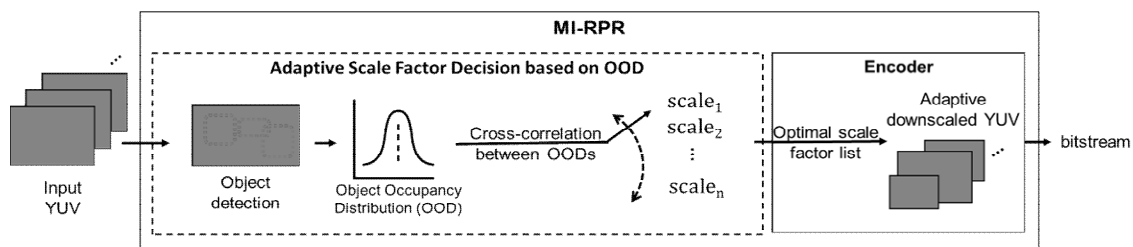


그림 5. OOD를 기반으로 최적 스케일 크기를 이용하는 MI-RPR 흐름도
 Fig. 5. MI-RPR flowchart with Optimal Scale Factor derived based on OOD

최대의 압축 효율을 얻을 수 있도록 다양한 해상도로 구성된 비트스트림을 출력한다.

III. VCM을 위한 공간적 리샘플링 기술의 성능 분석

본 논문에서는 VCM의 공간적 리샘플링 기술 성능 평가를 위하여 146차 회의 이후에 실시된 CE0의 결과를 바탕으로 공간적 리샘플링 기술 활성화 여부에 따른 성능 변화를 분석했다. 성능 변화 분석을 위해 시행된 실험은 VCM CTC(Common Test Conditions)에서 제시하는 조건을 따랐으며, 객체 검출을 위하여 Faster R-CNN 모델(X101-FPN)을, 객체 추적을 위하여 JDE(Joint Detection and Embedding) 모델(JDE-1088x608)을 각각 사용했다. 또한 데이터셋은 객체 검출 작업에 SFU-HW 데이터셋을, 객체 추적 작업에 TVD(Tencent Video dataset)를 각각 사용했다. 공간적 리샘플링 기술의 전처리 과정에서 영상의 객체 검출을 위하여 활용된 모델은 YOLOv7이다^[15-19].

146차 회의에서 보고된 CE0에서는 VCM-RS(VCM Reference Software) v0.8.1이 사용되었으며 두 종류의 실험이 수행됐다. 표 1은 공간적 리샘플링 기술의 성능을 확

표 1. 공간적 리샘플링 기술 분석을 위한 실험 기준 및 테스트 대상
 Table 1. The experimental anchor and target for the analysis of spatial resampling

| Experiment | Anchor | Target |
|------------|----------------------|-------------------------------|
| 1 | Deactivate all tools | Activate spatial resampling |
| 2 | Activate all tools | Deactivate spatial resampling |

인하기 위하여 실험별 기준 및 대상을 보여준다. 실험 1을 통해서 VCM의 다른 기술이 미치는 영향을 제외하고 순수하게 공간적 리샘플링 기술이 동작했을 때의 성능을 확인할 수 있으며, 실험 2를 통해서 공간적 리샘플링 기술이 VCM의 최종적인 성능에 미치는 영향을 분석할 수 있다.

표 2는 기계 임무의 종류 및 압축 모드에 따라서 표 1의 실험 1과 2 결과를 보여준다. 압축 모드는 사용 사례에 따라 다양하게 구성될 수 있는 비트스트림의 GOP 구조 또는 키프레임의 출현 빈도를 정의하는 것을 의미하며, 실험에서 제시하는 RA(Random Access), LD(Low Delay), AI(All Intra)는 VVC의 참조 소프트웨어버전 20.0에서 제공되는 설정과 동일하다. 객체 검출과 객체 추적의 정확도 계산을 위하여 mAP(mean Average Precision)와 MOTA(Multiple Object Tracking Accuracy) 지표를 각각 사용한다. 또한 같은 기계 임무 성능 대비 압축 효율 성능을 비교하기 위하여 고안된 기계 임무 성능 기반의 BD-Rate(이하 BD-Rate) 성능을 계산한다.

표 2. VCM-RS v0.8.1의 공간적 리샘플링 기술 적용 여부에 따른 BD-Rate 성능

Table 2. BD-Rate performance of VCM-RS v0.8.1 according to whether spatial resampling is activated

| Machine Tasks | Coding Configuration | Experiment 1 (BD-rate) | Experiment 2 (BD-rate) |
|------------------|----------------------|------------------------|------------------------|
| Object Detection | RA | -3.14% | 2.35% |
| | LD | -7.08% | 13.50% |
| | AI | -10.94% | 6.34% |
| Object Tracking | RA | -1.01% | 1.93% |
| | LD | 12.18% | 11.93% |
| | AI | -8.66% | 0.43% |
| Average | - | -3.11% | 6.08% |

실험 1의 결과를 통해 공간적 리샘플링 기술이 활성화됨으로써 VCM의 BD-Rate가 -3.11% 향상되는 것을 확인할 수 있다. 반면, 실험 2에서 공간적 리샘플링 기술의 비활성화로 인하여 BD-Rate가 6.08% 악화된 것을 확인할 수 있다. 따라서 공간적 리샘플링 기술이 VCM 성능에 긍정적인 영향을 미치고 있음을 확인할 수 있다. 하지만, 객체 추적 임무에서 실험 1 LD에서는 공간적 리샘플링 기술의 활성화가 악영향을 미쳤음을 확인할 수 있다. 특히 테스트 시퀀스 TVD-02-1과 TVD-03-3에서 성능이 크게 나빠진다. 이러한 결과에는 세 가지 원인을 고려할 수 있다. 먼저 1) 시퀀스의 내용에 따라 공간적 리샘플링 기술의 효과가 다를 수 있다. 가령 장면 전환 중에 발생할 수 있는 페이드 인, 페이드 아웃, 또는 카메라의 이동 등으로 인하여 영상의 품질이 저하된다면 머신의 정확도가 낮아질 수 있다. 또한 탐지할 주요 객체가 없는 경우, 공간적 리샘플링 기술의 전처리 과정에서 적절한 다운스케일 크기를 찾지 못했을 수 있다. 둘째로, 2) 실험 기준과 대상의 비트레이트, mono-MOTA, 그리고 MOTA에 대하여 각각 압축률과 상대 에러 값을 계산했을 때, 문제가 되는 시퀀스에서 MOTA 보다 BD-rate 계산에 이용되는 mono-MOTA의 상대 오차율이 크게 관찰된다. 실제로 시퀀스 TVD-02-1과 TVD-03-3의 MOTA 상대 오차율과 mono-MOTA 상대 오차율을 계산하여 비교하면 mono-MOTA의 오차율이 더 커진 것을 확인할 수 있다. 마지막으로 3) 양자화가 강하게 적용되었을 때 공간적 리샘플링이 적용되면 영상에서 객체의 식별이 어려워져 머신 성능을 크게 저하시킬 수 있다. TVD-02-1과 TVD-03-3 모두 가장 양자화가 강하게 부호화된 사례에서 MOTA의 오차율이 가장 컸다.

LD 부호화 설정에서 테스트 시퀀스 TVD-02-1의 실험 결과를 객체 추적 정확도(MOTA)에 대한 비트율-왜곡 곡선으로 보여주는 그림 6을 통해 보다 상세한 원인 분석이 가능하다. 그래프에서 확인할 수 있듯이 원본(Origin) 데이터의 MOTA에 대한 비트율-왜곡 곡선은 반드시 우상향을 그리지 않기 때문에 관찰된 데이터로부터 보간된 그래프가 비단조적(Non-monotonic)인 형태를 띌 수 있다. 이러한 비단조적인 형태의 비트율-왜곡 곡선에서는 BD-Rate의 계산이 어렵기 때문에 VCM AhG은 단조적인 형태의 비트율-왜곡 곡선을 만들기 위한 다양한 곡선 피팅 접근법을 논의하였다^[20-22]. 이에 따라 곡선 피팅이 적용된 모노(Mono) 데이터를 생성할 수 있으며, 그림 6에서도 모노 데이터로 그려진 곡선 ‘Anchor Mono’와 ‘Test Mono’를 관찰할 수 있다. 하지만 원본 데이터 값이 약간의 차이가 나더라도 곡선 피팅 과정에서는 큰 왜곡이 발생할 수 있으며, 이는 BD-Rate의 값이 크게 변하는 요인이 될 수 있다.

또한 두 실험의 모노 결과를 관찰했을 때, 모두에서 비트레이트가 실험 기준(Anchor) 대비 상당히 줄어든 것을 관찰할 수 있다. 공간적 리샘플링 기술은 머신 임무의 정확도를 최대한 유지하면서 비트레이트를 최소화하기 위해 고안되었기 때문에 비트레이트의 큰 감소는 매우 긍정적인 결과이다. 하지만 실험 1의 낮은 비트레이트 구간에서 MOTA의 값 또한 크게 낮아진다. 강한 양자화 및 공간적 리샘플링이 동시에 적용된다면 영상의 품질 하락이 커질 수 있고, 머신 비전 성능 하락으로 이어질 수 있다. 이 점은 그림 6에서 실험 2의 BD-Rate 값이 우수한 반면에 실험 1의 BD-Rate 값이 크게 저하된 또 다른 요인으로 볼 수 있다.

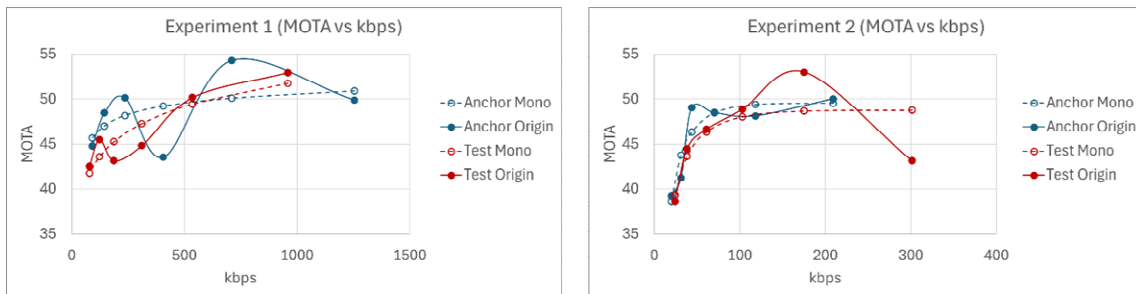


그림 6. LD 부호화 설정에서 객체 추적 임무에 TVD-02-1 테스트 시퀀스의 실험 1과 2의 비트율-왜곡 곡선
 Fig. 6. R-D curve for test sequence TVD-02-1 for object tracking under LD configuration: Experiments 1 and 2

147차 회의에서 진행된 CE0에서는 VCM-RS v0.9가 이용되었으며, VCM의 모든 기술이 활성화되었을 때의 결과를 실험 기준으로 설정하고 각 기술이 비활성화됐을 때의 성능 하락을 측정하였다. 이때, 공간적 리샘플링 기술이 비활성화되었을 때의 성능은 표 3에서 확인할 수 있으며, BD-rate 4.27% 나빠진 결과가 나오는 것을 확인할 수 있다. 146차 회의에서 제출된 결과보다 더 낮은 값이지만 소프트웨어 업데이트 이후에도 공간적 리샘플링 기술이 VCM의 성능 향상에 기여한다는 것을 확인할 수 있다.

표 3. VCM-RS v0.9의 공간적 리샘플링 기술 적용 여부에 따른 BD-Rate 성능
 Table 3. The experimental anchor and target for the analysis of spatial resampling

| Machine Tasks | Coding Configuration | BD-rate |
|------------------|----------------------|---------|
| Object Detection | RA | 5.59% |
| | LD | 15.91% |
| | AI | 2.11% |
| Object Tracking | RA | -0.06% |
| | LD | 3.35% |
| | AI | -1.28% |
| Average | - | 4.27% |

IV. 결론

본 논문에서는 VCM의 공간적 리샘플링 기술에 대하여 소개하고 해당 기술의 적용 여부에 따른 VCM 성능 변화를 분석했다. VCM에 채택된 MI-RPR 기반의 적응형 공간적 리샘플링 기술은 각 픽처에 대하여 사전 분석을 실시하여 최적 다운스케일 크기를 구하고, MI-RPR이 내장된 VVC를 통해 픽처의 해상도를 적응적으로 샘플링하여 부호화하기 때문에 기계 임무 수행에 부정적인 영향을 최소화하고 압축의 효율을 높이는데 있어 비교적 안정적인 성능 효과를 얻는다. 공간적 리샘플링 기술이 비활성화됐을 때, VCM-RS v0.8.1와 v0.9에서 각각 BD-rate 6.08%, 4.27% 하락하는 것을 확인했다. VCM의 공간적 리샘플링 기술은 VCM에 채택된 4개의 기술 중에서 사실상 가장 마지막에 수행되기 때문에 다른 기술의 영향을 많이 받을 수 있다.

따라서 소프트웨어의 업그레이드에 맞춰 기술의 성능 변화를 확인하고 개선점을 확인할 필요가 있다. 또한, 양자화의 강도가 큰 경우에 BD-rate가 크게 하락하는 문제를 해결하기 위하여 스케일 크기 보정에 대한 추후 연구가 필요하다.

참고 문헌 (References)

- [1] L. Duan, J. Liu, W. Yang, T. Huang and W. Gao, "Video coding for machines: A paradigm of collaborative compression and intelligent analytics," *IEEE Transactions on Image Processing*, Vol.29, pp.8680-8695, 2020.
- [2] W. Gao, S. Liu, X. Xu, M. Rafie, Y. Zhang, I. Curcio, "Recent standard development activities on video coding for machines," arXiv:2105.12653, 2021. doi: <https://doi.org/10.48550/arXiv.2105.12653>
- [3] WG 04, "Working draft 3 of video coding for machines," ISO/IEC JTC1/SC29/WG4 output document N0467, April 2024.
- [4] B. Bross, et al., "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. doi: <https://doi.org/10.1109/TCSVT.2021.3101953>
- [5] WG 4, "Common test conditions for video coding for machines," ISO/IEC JTC1/SC29/WG4 output document N00467, February 2024.
- [6] K. Andersson, J. Ström, R. Yu, "AHG9/AHG8: On reference picture resampling," *ISO/IEC JTC 1/SC 29/WG 5*, m54034, June 2021.
- [7] K. Andersson, J. Ström, R. Yu, "AHG10: GOP-based RPR encoder control," *ISO/IEC JTC 1/SC 29/WG 5*, m67944, July 2021.
- [8] K. Andersson, J. Ström, R. Yu, "AHG8: MI-RPR for machine consumption," *ISO/IEC JTC 1/SC 29/WG 5*, m67944, July 2021.
- [9] A. Kim, et al, "[VCM] Adaptive spatial resampling based on MI-RPR for VCM," *ISO/IEC JTC 1/SC 29/WG 4*, m64124, July 2023.
- [10] A. Kim, et al, "Performance Improvement and Modification of VVC RPR Technology for Machine Vision Systems," *JBE*, 28(7), 911-917, 2023.
- [11] R. Chernyak, et al, "[VCM] Summary report of CE0: Tool experiments," *ISO/IEC JTC 1/SC 29/WG 4*, m67931, April 2024.
- [12] E. An, et al, "[VCM] Report of CE0.1.4 and CE0.2.4: spatial resample tool test," *ISO/IEC JTC 1/SC 29/WG 4*, m67944, April 2024.
- [13] R. Chernyak, et al, "[VCM] Summary report of CE0: Tool experiments," *ISO/IEC JTC 1/SC 29/WG 4*, m69230, July 2024.
- [14] E. An, et al, "[VCM] CE0.4 spatial resampling tool test," *ISO/IEC JTC 1/SC 29/WG 4*, m68892, July 2024.
- [15] Ren, S., et al, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol.39, pp.1137-1149, 2017.
- [16] Z. Wang, et al, "Towards Real-Time Multi-Object Tracking," *The European Conference on Computer Vision (ECCV)*, July 2020.
- [17] H. Choi, et al, "SFU-HW-Objects-v1: Object labelled dataset on raw

video sequences,” 2020.

doi: <https://doi.org/10.17632/hwm673bv4m.1>

- [18] W. Gao, X. Xu, M. Qin, S. Liu, “An Open Dataset for Video Coding for Machines Standardization,” IEEE International Conference on Image Processing (ICIP), 2022.
- [19] W. Chien-Yao, A. Bochkovski, and H. M. Liao. “YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,” Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), pp. 7464-7475, 2023.

- [20] H. Wang, et al, “[VCM] Improvements of the BD-rate model using monotonic curve-fitting method,” ISO/IEC JTC 1/SC 29/WG 4, m63692, July 2023.
- [21] J. Liu, et al, “[VCM] Dealing with non-monotonic distributed R-D points,” ISO/IEC JTC 1/SC 29/WG 4, m64389, July 2023.
- [22] D. Ding, et al, “[VCM] A Curve Fitting Approach to Transform Non-Monotonic Test Data for BD-Rate Calculation,” ISO/IEC JTC 1/SC 29/WG 4, m65531, October 2023.

저 자 소 개

안 은 빈



- Aug. 2016 : B.S. degree, Division of Computer and Telecommunications, Yonsei University
- Mar. 2017 ~ currently : Ph.D. Candidate, Division of Software, Yonsei University
- ORCID : <https://orcid.org/0000-0001-7681-4682>
- Research interests : Video coding for machine, immersive media, multimedia communication systems

김 아 영



- Feb. 2016 : B.S. degree, Division of Computer and Telecommunications, Yonsei University
- Feb. 2024 : Ph.D., Division of Software, Yonsei University
- Aug. 2024 : Postdoctoral Researcher, Division of Software, Yonsei University
- Sept. 2024 ~ Currently : Postdoctoral Researcher, ETRI
- ORCID : <https://orcid.org/0000-0002-3793-1365>
- Research interests : Video coding for machine, immersive media, multimedia communication systems

곽 상 운



- Feb. 2012 : B.S., Department of Electrical and Electronic Engineering, Yonsei University
- Feb. 2014 : M.S., Department of Electrical Engineering, KAIST
- Feb. 2014 ~ currently : Senior researcher, ETRI
- ORCID : <https://orcid.org/0000-0002-9295-8966>
- Research interests : Video coding for machines, Neural radiance field, View synthesis

저 자 소 개

정 순 흥



- Feb. 2001 : B.S., Electronics, Pusan National University
- Feb. 2003 : M.S., Electrical Engineering, KAIST
- Feb. 2016 : Ph.D., Electrical Engineering, KAIST
- Mar. 2003 ~ Mar. 2005 : Assistant Researcher, LG Electronics
- Aug. 2019 ~ Aug. 2020 : Visiting Scholar, Indiana University Bloomington
- Apr. 2005 ~ currently : Principal Researcher, ETRI
- ORCID : <https://orcid.org/0000-0003-2041-5222>
- Research interests : Immersive media, computer vision, video coding, realistic broadcasting system

정 원 식



- Feb. 1992 : B.S., Department of Electronic and Electrical Engineering, Kyungpook National University
- Feb. 1994 : M.S., Department of Electronic and Electrical Engineering, Kyungpook National University
- Feb. 2000 : Ph.D., Department of Electronic and Electrical Engineering, Kyungpook National University
- May 2000 ~ currently : Principal member of research staff, ETRI
- ORCID : <http://orcid.org/0000-0001-5430-29697>
- Research interests : 3DTV broadcasting system, light field imaging, video and image coding, video coding for machines and deep learning based signal processing

추 현 곤



- Feb. 1998 : B.S., Department of Electronic engineering, Hanyang University
- Feb. 2000 : M.S., Department of Electronic engineering, Hanyang University
- Feb. 2005 : Ph.D., Department of Electronic communication engineering, Hanyang University
- Feb. 2005 ~ currently : Principal Researcher, Electronics and Telecommunications Research Institute (ETRI)
- Jan. 2015 ~ Feb. 2017 : Director of the Digital Holography Section, ETRI
- Sep. 2017 ~ Aug. 2018 : Visiting Researcher, Warsaw University of Technology, Poland
- Feb. 2023 ~ currently : Director of the Digital Holography Section, ETRI
- ORCID : <https://orcid.org/0000-0002-0742-5429>
- Research interests : video coding for machines, holography, multimedia protection and 3D broadcasting technologies

서 광 덕



- Feb. 1996 : B.S., Department of Electrical Engineering, KAIST
- Feb. 1998 : M.S., Department of Electrical Engineering, KAIST
- Aug. 2002 : Ph.D., Department of Electrical Engineering, KAIST
- Aug. 2002 ~ Feb. 2005 : Senior research engineer, LG Electronics
- Sep. 2012 ~ Aug. 2013 : Courtesy Professor, Univ. of Florida, USA
- Mar. 2005 ~ currently : Professor, Yonsei University
- ORCID : <http://orcid.org/0000-0001-5823-2857>
- Research interests : Video coding, Visual communication, digital broadcasting, multimedia communication system