



일반논문 (Regular Paper)

방송공학회논문지 제31권 제3호, 2026년 5월 (JBE Vol.31, No.3, May 2026)

<https://doi.org/10.5909/JBE.2026.31.3.434>

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

VCM의 도구 간 상호작용으로 인한 평가 불안정성 개선을 위한 공간적 재표본화 스케일 보정 기법

김민석^{a)}, 신호승^{a)}, 박세희^{a)}, 서광덕^{a)†}

Spatial Resampling Scale Adjustment Method for Improving Evaluation Instability Caused by Tool Interactions in VCM

Minsuk Kim^{a)}, Hyoseung Shin^{a)}, Sehee Park^{a)}, and Kwang-deok Seo^{a)†}

요 약

본 논문은 VCM-RS 기반 환경에서 Spatial Resampling과 RoI-based processing을 동시에 적용할 때 발생하는 도구 간 상호작용을 분석하고, BD-rate/BD-mAP 산출 불안정성을 해결하는 객체 크기 기반 스케일 계수 보정 기법을 제안한다. 기존 MPEG VCM의 CE0 테스트의 단독 비활성화 실험의 성능 보고 방식의 한계를 보완하기 위해 단독 활성화 및 부호화 툴(tool) 간 상호작용 실험을 수행하였으며, 상호작용 조건에서 RD-curve의 성능 범위 불일치와 곡선 교차로 인한 성능지표 계산 불가의 문제가 발생함을 확인하였다. 제안 기법은 최소 검출 객체 크기를 기준으로 Spatial Resampling 스케일을 적절히 상향 보정해 누적 해상도 축소 현상을 억제한다. 적용 후 CTC(common test condition)에 따른 RA/LD/AI 전체 테스트 환경과 전체 테스트 시퀀스의 클래스에서 성능 지표가 정상 산출되었고, 과도한 BD-rate 증가가 완화되어 평가 신뢰성과 해석 가능성이 향상되었다.

Abstract

This paper analyzes tool interaction effects when Spatial Resampling and RoI-based processing are jointly applied in a VCM-RS environment and proposes an object size-based scale adjustment to resolve BD-rate/BD-mAP instability. We performed various experiments not only at the single-tool deactivation test but also at the interactive test between VCM coding tools to cope with the limitations found at the reporting template of the CE0 experiment in MPEG VCM. According to the extensive experiments, we found that there exist mismatches in the scope of RD-curve performance and problems of uncomputable evaluation metrics due to RD-curve crossing. The proposed method stabilizes RD-curve evaluation across RA/LD/AI modes by suppressing cumulative object downscaling. By applying the proposed method, the evaluation metrics could produce normal output values for all the RA/LD/AI modes according to the CTC, and accordingly we could obtain interpretable test results and achieve reliability on the performance evaluation by mitigating excessive BD-rate increase.

Keyword : Video coding for Machine, Spatial Resampling, Tool Interaction, Scale Adjustment

Copyright © 2026 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 인공지능 기술의 발전과 함께 영상 데이터를 기반으로 한 객체 탐지, 분류, 추적과 같은 기계 인식 응용이 다양한 산업 분야에서 핵심 기술로 활용되고 있다^[1]. 특히 자율주행, 스마트 시티, 지능형 감시 시스템과 같은 응용 환경에서는 대규모 영상 데이터를 실시간으로 처리하고 분석해야 하며, 영상 데이터는 인간의 시각적 감상보다 기계 학습 및 추론을 위한 입력 정보로서 중요성이 점차 증가하고 있다. 이러한 변화는 영상 부호화 기술이 인간 중심의 시각 품질 최적화에서 벗어나, 기계 인식 성능을 고려한 새로운 설계 방향으로 확장되어야 함을 시사한다. 반면 기존의 비디오 부호화 기술은 인간의 시각적 인지 특성을 기반으로 설계되어 왔다^[2,4]. 이러한 접근은 시각적 품질 관점에서는 효과적이었으나, 기계 인식 관점에서는 객체의 크기, 형태, 경계와 같은 구조적 특징 정보의 보존 여부가 핵심적인 요소로 작용한다. 특히 공간 해상도 축소 과정에서 객체의 크기가 감소할 경우, 객체 탐지기의 안정 동작 범위를 벗어나 성능이 급격히 저하될 수 있다. 이러한 문제는 단순한 화질 열화를 넘어, 기계 인식 태스크의 안정성을 직접적으로 저해하는 요인이다.

이러한 배경에서 MPEG에서는 Video Coding for Machine(VCM) 표준화를 추진하며, 기계 인식 성능을 고려한 영상 부호화 프레임워크를 정의하고 있다^[5-10]. VCM 표준에서는 Spatial Resampling, RoI-based processing 등 다양한 전처리 도구를 통해 부호화 효율과 기계 인식 성능

간의 균형을 달성하고자 하며, 참조 소프트웨어인 VCM-RS를 통해 성능을 검증한다^[11-12]. 성능 평가는 BD-rate 및 BD-mAP과 같은 기계 인식 중심 지표를 사용하여 수행된다^[13]. VCM-RS를 기반으로 수행되어 온 Core Experiment 0(CEO)에서는 각 도구의 성능 기여도를 분석하기 위해, 분석 대상 도구 하나만을 비활성화하고 나머지 도구는 모두 활성화하는 단독 비활성화 조건에서 실험 결과가 주로 보고되어 왔다^[14]. 이러한 실험 방식은 실제 파이프라인 환경을 유지한 상태에서 개별 도구의 상대적 기여도를 평가하는 데에는 효과적이다. 그러나 이와 같은 접근에서는 도구들이 실제로 활성화된 상태에서, 또는 복수의 도구가 동시에 활성화된 상태에서 발생하는 상호작용 효과를 직접적으로 분석하는 데에는 한계가 있다.

특히 Spatial Resampling^[15-16]과 RoI-based processing^[17-18]은 모두 영상의 공간 해상도와 객체 크기에 영향을 미치는 도구로서, 동시에 활성화될 경우 해상도 축소 효과가 누적될 가능성이 존재한다. 본 논문에서는 이러한 한계를 보완하기 위해, 각 도구를 개별적으로 활성화한 단독 활성화 조건과, Spatial Resampling과 RoI-based processing이 동시에 활성화된 조건에서 실험을 수행하였다. 이를 통해 단일 도구 활성화 시의 성능과, 복수 도구 결합 시 발생하는 도구 간 상호작용에 따른 성능 변화를 구분하여 분석한다. 실험 결과, 각 도구를 개별적으로 활성화한 단독 활성화 조건에서는 전반적으로 안정적인 성능 결과가 관측된 반면, Spatial Resampling과 RoI-based processing이 동시에 활성화된 조건에서는 일부 시퀀스에서 성능이 과도하게 저하되거나 평가 지표 산출이 불안정해지는 현상이 확인되었다. 이는 두 도구 간 상호작용에 기인한 결과로 해석할 수 있으며, 기존 단독 비활성화 실험만으로는 충분히 드러나지 않았던 특성이자, 본 논문은 이러한 도구 간 상호작용으로 인해 발생하는 성능 불안정성을 해결하기 위해 객체 크기 기반 스케일 계수 보정 기법을 제안한다. 제안 기법은 객체 크기 기반 조건을 고려하여 공간 해상도 축소 비율을 조정함으로써, Spatial Resampling과 RoI-based processing의 결합 적용 시 발생하는 과도한 객체 크기 축소를 완화하는 것을 목표로 한다. 이를 통해 도구 결합 환경에서도 성능 지표가 안정적으로 산출될 수 있는 평가 조건을 확보하고자 한다.

a) 연세대학교 소프트웨어학부(Yonsei University)

‡ Corresponding Author : 서광덕(Kwang-deok Seo)

E-mail: kdseo@yonsei.ac.kr

Tel: +82-33-760-2788

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

※이 논문 또는 저서는 4단계 BK21 사업(관리번호 133561)에 의하여 지원되었음.

※This research was supported by the Regional Innovation System & Education(RISE) program through the Gangwon RISE Center, funded by the Ministry of Education(MOE) and the Gangwon State(G.S.), Republic of Korea (2026-RISE-10-006), and by the National Research Foundation(NRF), Korea, under project BK21 FOUR (NO. 133561).

· Manuscript January 27, 2026; Revised February 19, 2026; Accepted February 19, 2026.

II. 관련 연구

1. Spatial Resampling

Spatial Resampling은 VCM에서 사용되는 주요 전처리 모듈 중 하나로, 영상의 공간 해상도를 낮춤으로써 비트율을 절감하는 동시에 기계 인식에 필수적인 객체의 크기 및 형태 정보를 최대한 유지하는 것을 목표로 한다. VCM-RS에서 정의된 Spatial Resampling은 전체 처리 과정을 기준으로 사전 분석(pre-analysis stage) 단계와 압축(inner coding stage) 단계의 두 부분으로 구성된다. 사전 분석 단계에서는 입력 프레임에 포함된 객체의 점유 특성을 분석하여 프레임 단위의 해상도 축소 비율을 결정하며, 이후 압축 단계에서는 사전 분석 결과에 따라 영상의 해상도를 실제로 조정한다. MI-RPR 기반 Spatial Resampling의 전체 처리 흐름은 그림 1에 나타나 있다.

사전 분석 단계는 세부적으로 Valid Object Filtering, Optimal Scale Decision, 그리고 Scale Adjustment의 세 과정으로 구성된다. 먼저 Valid Object Filtering 단계에서는 객체 검출 결과 중 신뢰도 임계값 이상을 만족하는 객체만을 유효 객체로 선택함으로써, 검출 오류나 노이즈로 인한 영향을 최소화한다. 이후 Optimal Scale Decision 단계에서는 선별된 유효 객체들을 대상으로 복수의 후보 스케일 계수 x (예: $x=\{90\%, 70\%, 50\%\}$)에 대해 Object Occupancy Distribution(OOD)을 계산한다. OOD는 프레임 내에서 유효 객체가 차지하는 공간 분포를 나타내는 지표이다. 각 후보 스케일 계수에 대해 계산된 OOD_x 는 원본 해상도(100%)에서의 기준 분포인 OOD_{100} 과 비교되며, 두 분포간의 상관도(cross-correlation)를 통해 스케일 축소에 따른 객체 분포 왜곡 정도가 정량적으로 평가된다. 이러한 비교

과정을 통해 영상 해상도 축소가 객체 공간 분포에 미치는 영향을 고려한 스케일 선택이 이루어진다. 사전 분석 단계의 마지막 과정인 Scale Adjustment에서는 앞선 과정에서 도출된 스케일 계수를 대상으로 추가적인 보정이 수행된다. 이 단계에서는 단순한 OOD 기반 판단에 그치지 않고, 영상 해상도와 객체 점유 특성을 동시에 고려하여 최종 스케일을 결정한다. 이를 위해 해상도 정보와 객체 점유율(Object Area Ratio, OAR)을 각각 정량화하고, 두 요소를 결합한 Quality Score(QS)를 산출한다. QS는 영상 해상도가 낮거나 프레임 내 객체의 크기가 상대적으로 작은 경우 더 큰 값을 가지며, QS 값이 클수록 스케일 계수는 상향 조정된다.

이후 압축 단계는 MI-RPR(Machine Inference-based Reference Picture Resampling) 구조를 기반으로 동작한다. 이 단계에서는 사전 분석 과정에서 산출된 프레임 단위 스케일 계수를 입력으로 받아, 해당 계수에 맞추어 영상의 공간 해상도를 리샘플링한다.

2. RoI-based processing

VCM 표준에 채택된 RoI-based processing은 객체가 포함된 중요 영역의 구조적 정보를 보존하면서, 객체가 존재하지 않는 배경 영역의 공간 해상도를 선택적으로 축소하여 비트율을 효율적으로 절감하기 위한 전처리 모듈이다. 이 기법은 객체 중심의 공간 해상도 제어를 통해 기계 인식 성능 저하를 최소화하는 것을 목표로 하며, 전체 처리 과정은 RoI Extraction, Margin Dilation, Grouping, Accumulation, Retargeting의 다섯 단계로 구성된다. 그림 2는 RoI-based processing의 전체 처리 과정을 나타낸다.

첫 번째 단계인 RoI Extraction에서는 RoI Generation Network로 지정된 객체 검출 모델을 이용하여 프레임 단위

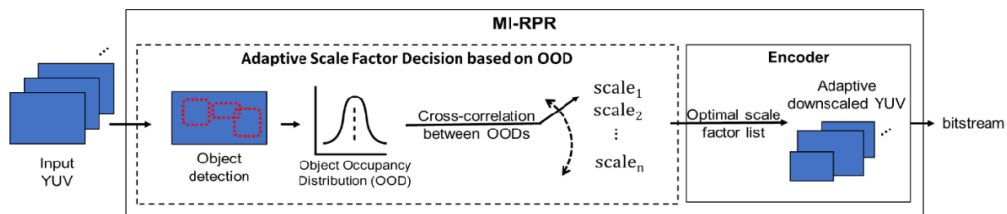


그림 1. MI-RPR 기반 Spatial Resampling의 전체 처리 과정
Fig. 1. Overall Processing Flow of MI-RPR - Based Spatial Resampling

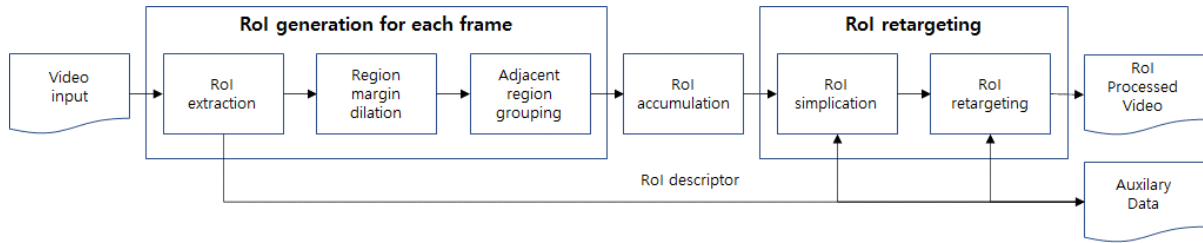


그림 2. RoI-based processing의 전체 처리 과정
 Fig. 2. Overall Processing Flow of RoI-Based Processing

로 객체의 위치와 크기를 추정하고, 이에 대응하는 RoI 바운딩 박스 정보를 생성한다. 이 단계에서 추출된 RoI는 프레임별로 독립적으로 존재하며, 이후 단계에서 안정적인 공간적 처리를 수행하기 위한 기초 입력으로 활용된다.

Margin Dilation 단계에서는 검출된 각 RoI 바운딩 박스의 상·하·좌·우 방향으로 약 20픽셀의 여유 영역을 확장 적용한다. 이는 다운샘플링 및 리타겟팅 과정에서 발생할 수 있는 객체 경계 손실을 방지하고, 객체 주변의 중요한 구조적 정보가 손상되는 것을 최소화하기 위한 안정화 조치이다. 특히 VCM-RS 환경에서는 해상도 축소가 비균일적으로 적용되며, RoI 외부 영역의 제거와 축소가 동시에 발생할 수 있기 때문에 객체 경계 보존은 전체 부호화 품질과 기계 인식 성능을 좌우하는 중요한 요소로 작용한다.

Grouping 단계에서는 Margin Dilation이 반영된 RoI들 간의 공간적 관계를 분석하여, 인접한 RoI들을 하나의 그룹으로 병합한다. 구체적으로, RoI 간의 가로 방향 또는 세로 방향 간격이 전체 영상의 가로폭 또는 세로폭의 1/60 이하인 경우 동일한 그룹으로 처리된다. 이러한 병합 과정은 인접 객체가 여러 개의 RoI로 분리될 경우 리타겟팅 격자가 과도하게 복잡해지는 문제를 방지하며, 이후 단계에서 공간 해상도 축소 비율 계산이 보다 안정적으로 수행될 수 있도록 한다.

Accumulation 단계에서는 RoIAccumulationPeriod로 정의된 시간 구간 동안 프레임별 RoI 정보를 누적하여, 해당 기간에 존재한 모든 RoI를 포함하는 누적 RoI 마스크를 생성한다. 이 과정은 객체 검출 모델이 프레임 간에 보일 수 있는 검출 변동성을 완화하고, 일정 구간 내에서 일관된 RoI 기반 해상도 제어가 가능하도록 설계되었다. Random Access(RA) 및 All Intra(AI) 구성에서는 기간 단위 누적 방식이 사용되며, Low Delay(LD) 구성에서는 과거 프레임만을 포함하는 슬라이딩 윈도우 방식이 적용된다.

마지막 **Retargeting** 단계에서는 누적된 RoI 정보와 클래스별 또는 영역별로 정의된 목표 축소 정책을 기반으로 각 RoI에 대해 **Scaling Factor Index**가 할당된다. **Scaling Factor Index**는 0부터 15까지의 정수 범위를 가지며, 값이 작을수록 원본 해상도 보존에 가까운 축소를 의미하고, 값이 클수록 더 큰 축소 비율이 적용되도록 설계되어 있다. 각 **Index** 값은 미리 정의된 분수 비율에 대응하여 RoI별 실제 공간 해상도 축소 비율을 결정한다.

III. 단일 도구 활성화 및 도구 간 상호작용에 따른 실험 결과 분석

본 장에서는 VTM을 실험 기준으로 설정한 조건에서 Spatial Resampling과 RoI-based processing의 성능을 비교하고, 두 도구가 동시에 활성화될 경우 발생하는 도구 간 상호작용이 성능 및 평가 지표 산출에 미치는 영향을 고찰한다. 모든 실험은 SFU-HW-Objects-v3.2 데이터셋^[19] 전체를 대상으로 수행되었으며, SFU 시퀀스를 RA, AI, LD 구성에서 평가하였다.

동일한 실험 구성은 CTC에 포함된 TVD(Tencent Video Dataset)^[20]에 대해서도 수행되었으며, RD-curve 범위 불일치 및 곡선 교차로 인한 성능 지표 산출 불안정 현상이 SFU와 동일한 구조로 재현됨을 확인하였다. 본 논문의 목적은 도구 간 상호작용으로 인해 발생하는 RD-curve 구조 왜곡의 원인을 분석하고 이를 안정화하는 방법을 제시하는 데 있으므로, 현상이 대표적으로 관찰된 SFU-HW 데이터셋을 중심으로 상세 분석을 기술한다.

본 논문에서는 Spatial Resampling과 RoI-based processing 두 도구만을 분석 대상으로 한정하였다. 이는 두 도구

가 모두 영상의 공간 해상도를 축소함으로써 비트율 절감을 수행하는 전처리 모듈이라는 공통점을 가지기 때문이다. 이러한 설정은 VCM 표준에서 정의된 모든 도구를 항상 적용하기 어려운 환경, 특히 연산 자원과 메모리가 제한된 엠티 디바이스와 같은 저성능 기기에서 제한된 도구 조합만으로 비트율 절감을 수행해야 하는 상황에서 효과적일 수 있다. 이에 따라 본 장에서는 Spatial Resampling 단독 활성화, ROI-based processing 단독 활성화, 그리고 Spatial Resampling과 ROI-based processing의 결합 활성화 조건에서의 실험 결과를 순차적으로 분석한다.

1. Spatial Resampling 단독 활성화 조건에서의 성능 분석

Spatial Resampling만을 활성화한 단독 활성화 조건에서의 실험 결과를 분석하였다. 표 1은 Spatial Resampling 단독 활성화 조건에서의 실험 결과를 나타낸다.

표 1. Spatial Resampling 단독 활성화 조건에서의 성능 결과
Table 1. Performance Results under Single Tool-On Condition with Spatial Resampling

	Object Detection	End-to-End BD-Rate [%] mAP	End-to-End BD-mAP
Random Access	Class A	17.62%	-1.36
	Class B	0.69%	0.30
	Class C	#####	-4.32
	Class D	#####	-5.75
	Average	#####	-3.40
Low Delay	Class A	-6.13%	0.71
	Class B	-21.41%	1.27
	Class C	#####	-4.12
	Class D	#####	-5.91
	Average	#####	-2.96
All Intra	Class A	-8.83%	0.62
	Class B	-24.83%	1.12
	Class C	#####	-0.89
	Class D	#####	#####
	Average	#####	#####

실험 결과 Class A와 Class B와 같은 상대적으로 고해상도 시퀀스에서는 일부 조건에서 유의미한 비트율 절감 효과가 관측되었다. Class A에서는 최대 -8.83% 수준의 BD-rate 감소가 확인되었으며, Class B에서는 -24.83%까지 비트율이 절감되었다. BD-mAP 또한 Class A에서는 최

대 0.71, Class B에서는 1.27 수준으로 유지되거나 소폭 향상되는 경우가 관측되었다. 이는 고해상도 환경에서 Spatial Resampling의 프레임 해상도 축소가 배경 영역의 비트 소모를 효과적으로 줄일 수 있음을 보여준다. 반면, Class C와 Class D와 같은 저해상도 시퀀스에서는 Spatial Resampling 적용 시 성능 저하와 평가 지표 불안정성이 두드러지게 나타났다. 다수의 조건에서 BD-rate가 정상적으로 계산되지 않았으며, BD-mAP는 Class C에서 최대 -4.32, Class D에서는 -5.91까지 감소하였다. 이로 인해 SFU 전체 평균 기준에서도 BD-rate 평균은 도출되지 않았고, BD-mAP 평균은 -3.40 또는 -2.96 수준으로 나타났다.

이러한 현상은 Spatial Resampling의 스케일 결정 방식과 밀접한 관련이 있다. MPEG 제149차 회의 이전까지 Spatial Resampling은 QP로 인한 열화를 완화하기 위해 QP를 고려한 scale adjustment를 수행하였으나, 이후 non-monotonic bitrate 문제로 인해 표준 논의 과정에서 QP를 고려하지 않는 방식으로 변경되었다^[21]. 그 결과, QP가 증가할수록 객체 크기 축소 효과가 누적되며, 특히 객체 크기가 작은 저해상도 클래스에서는 객체 정보 손실이 급격히 증가한다. 이러한 조건에서는 실험 기준과 비교 대상 간 RD 곡선의 비교 가능한 성능 구간이 충분히 형성되지 않아 BD-rate와 같은 성능 지표가 정의되지 않는 문제가 발생한다.

2. Roi-based processing 단독 활성화 조건에서의 성능 분석

ROI-based processing만을 활성화한 단독 활성화 조건에서의 실험 결과를 분석하였다. 표 2는 ROI-based processing 단독 활성화 조건에서의 실험 결과를 나타낸다.

실험 결과, Class B와 같이 객체가 명확하고 배경 비중이 큰 시퀀스에서는 ROI-based processing 단독 적용 시 매우 큰 비트율 절감 효과가 관측되었다. RA 환경에서는 BD-rate가 -48.71%에 이르는 비트율 절감이 확인되어, ROI-based processing의 객체 중심 해상도 보존 전략이 효과적으로 작동함을 보여준다. 반면, Class A와 같은 고해상도 시퀀스에서는 ROI-based processing 단독 적용 시 비트율 증가가 빈번하게 발생하였다. Class A에서는 BD-rate가 최대 43.68%까지 증가하였으며, BD-mAP 또한 -3.12 수

표 2. RoI-Based Processing 단독 활성화 조건에서의 성능 결과
 Table 2. Performance Results under Single Tool-On Condition with RoI-Based Processing

	Object Detection	End-to-End BD-Rate [%] mAP	End-to-End BD-mAP
Random Access	Class A	43.68%	-3.12
	Class B	-48.71%	#####
	Class C	18.51%	-0.39
	Class D	-4.89%	0.23
	Average	-4.00%	#####
Low Delay	Class A	7.24%	-1.12
	Class B	-35.59%	2.40
	Class C	21.69%	-1.05
	Class D	#####	-5.84
	Average	#####	-1.79
All Intra	Class A	#####	-5.40
	Class B	#####	#####
	Class C	4.38%	0.50
	Class D	6.39%	-0.78
	Average	#####	#####

준으로 감소하였다. 이는 고해상도 시퀀스에서 RoI 확장, 누적, 리타겟팅 과정에서 발생하는 부가 비용이 배경 축소에 따른 이득을 상쇄하기 때문으로 해석할 수 있다. Class C와 Class D에서는 RoI-based processing 단독 적용 시 환경 및 구성에 따라 상반된 결과가 나타났다. 일부 조건에서는 제한적인 비트율 절감 효과가 관측되었으나, 다른 조건에서는 BD-rate 증가 또는 평가 지표 연산 불가 문제가 발생하였다. 특히 일부 LD 조건과 All Intra 구성에서는 실험 기준과 비교 대상 간 RD 곡선의 비교 가능한 성능 구간이 충분히 형성되지 않아 BD-rate가 정상적으로 산출되지 않는 사례가 확인되었다.

3. Spatial Resampling과 RoI-based processing 간 상호작용에 따른 성능 및 평가 불안정성 분석

해당 절에서는 Spatial Resampling과 RoI-based processing을 동시에 활성화한 조건에서의 실험 결과를 분석하였다. 표 3은 동 조건에서의 실험 결과를 나타낸다.

표 3. Spatial Resampling과 RoI-based processing 간 상호작용에 따른 성능 결과
 Table 3. Performance Results under Joint Activation of Spatial Resampling and RoI-Based Processing

	Object Detection	End-to-End BD-Rate [%] mAP	End-to-End BD-mAP
Random Access	Class A	42.64%	-3.12
	Class B	-47.70%	#####
	Class C	22.77%	-0.88
	Class D	30.78%	-0.73
	Average	9.48%	#####
Low Delay	Class A	37.10%	-3.45
	Class B	-36.61%	#####
	Class C	22.44%	-1.23
	Class D	#####	-5.55
	Average	#####	#####
All Intra	Class A	#####	-5.30
	Class B	#####	#NUM!
	Class C	-2.18%	#####
	Class D	-6.13%	-0.21
	Average	#####	#####

실험 결과, 두 도구를 동시에 적용한 조건에서는 단일 도구 적용 대비 성능이 더욱 복잡하게 나타났으며, 전반적으로 성능 및 평가 지표의 불안정성이 가장 크게 관측되었다. Class B에서는 결합 조건에서도 비교적 큰 비트율 절감 효

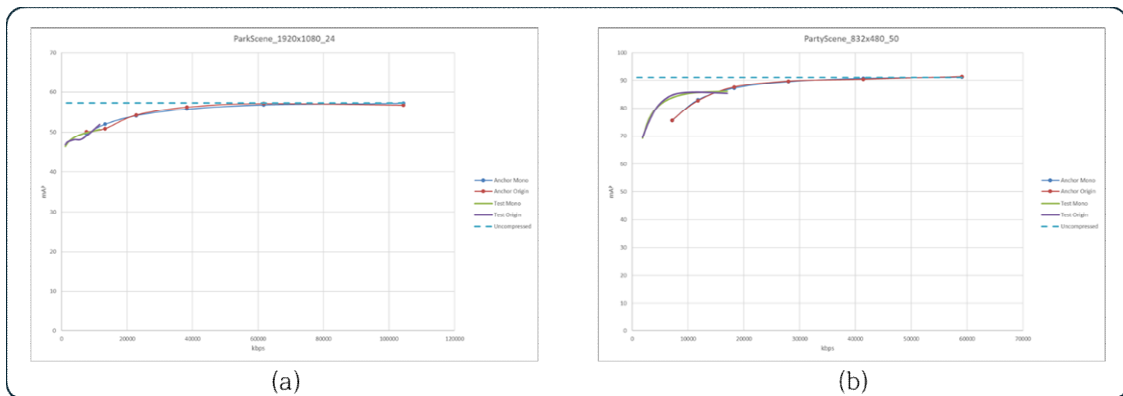


그림 3. All Intra 환경에서의 대표적인 BD-rate 연산 문제의 예시
 Fig. 3. Representative Examples of BD-Rate Calculation Failure in the All Intra Configuration

과가 유지되었다. 일부 조건에서는 BD-rate가 -40% 이하로 유지되어, RoI-based processing 단독 활성화 조건과 유사한 수준의 성능이 관측되었다. 이는 객체가 명확한 시퀀스에서는 RoI-based processing의 객체 보존 효과가 Spatial Resampling의 프레임 해상도 축소로 인한 영향을 일부 상쇄할 수 있음을 의미한다. 반면, Class A, Class C, Class D에서는 Spatial Resampling과 RoI-based processing의 결합으로 인해 성능 저하와 평가 불안정성이 현저히 증폭되었다. Class A에서는 BD-rate가 최대 42.64%까지 증가하였으며, BD-mAP 또한 -3.12 수준으로 감소하였다. Class C와 Class D에서는 BD-rate 증가와 함께 다수의 조건에서 BD-rate가 정상적으로 도출되지 않는 문제가 발생하였다. 그림 3은 AI 환경에서 Spatial Resampling과 RoI-based processing이 동시에 활성화된 조건에서의 대표적인 RD-curve 사례를 나타낸다. 그림 3 (a)는 Class B의 ParkScene_1920x1080_24.yuv 시퀀스, 그림 3 (b)는 Class C의 PartyScene_832x480_50.yuv 시퀀스의 실험결과다. 두 시퀀스는 해당 환경에서 BD-rate 연산이 정상적으로 수행되지 않은 대표적인 사례이다.

ParkScene_1920x1080_24.yuv 시퀀스의 경우, 실험 기준과 비교 대상 간 mAP 범위가 충분히 겹치지 않는다. 이로 인해 동일한 비트율에서 성능 차이를 비교할 수 있는 구간이 형성되지 않으며, 결과적으로 BD-rate가 정의되지 않는다. 반면 PartyScene_832x480_50.yuv 시퀀스에서는 비교 대상의 곡선이 실험 기준의 곡선과 교차하는 현상이 관측된다. 저비트율 영역과 중간 비트율 영역에서 두 곡선의 상대적 우열 관계가 역전되면서, RD-curve 전 구간에서 일관된 비교 관계가 유지되지 않는다. 이러한 곡선 교차는 BD-rate 계산이 전제하는 비교 조건을 위반하여 연산 불안정성을 유발한다. 두 사례는 Spatial Resampling에 의한 프레임 해상도 축소와 RoI-based processing의 객체 중심 리타겟팅이 동시에 적용될 경우, 객체 크기 축소 효과가 누적되면서 RD-curve의 형태 자체가 왜곡될 수 있음을 보여준다.

종합하면, Spatial Resampling과 RoI-based processing의 결합 조건에서는 성능이 양호한 조건에서는 과도하게 강화되고, 취약한 조건에서는 더욱 악화되는 양극화 현상과 함께 BD-rate 연산 불가 문제와 같은 평가 지표 불안정성이 빈번히 발생하였다. 이는 도구 간 상호작용으로 인해 실험

결과 해석과 신뢰성 확보가 어려워지는 구조적 문제로 해석할 수 있다.

이에 본 논문에서는 이러한 평가 불안정성 문제를 완화하기 위한 방안으로, Spatial Resampling의 스케일 결정 과정에 객체 크기 정보를 직접 반영하는 객체 크기 기반 스케일 계수 보정 기법을 새롭게 제안한다. 제안 기법은 객체 크기가 과도하게 축소되는 경우 Spatial Resampling의 축소 계수를 보정함으로써, 프레임 단위 해상도 축소와 RoI-based processing에 의한 객체 중심 리타겟팅이 중첩되어 발생하는 누적 축소 효과를 억제하는 것을 목표로 한다. 이를 통해 Spatial Resampling과 RoI-based processing이 동시에 적용되는 조건에서도 객체 크기 변화가 보다 안정적으로 제어되며, RD-curve의 형태 왜곡을 완화하여 비교 가능한 성능 구간을 확보할 수 있다.

IV. 객체 크기 기반 스케일 계수 보정 기법

Spatial Resampling과 RoI-based processing이 동시에 적용되는 환경에서는 프레임 단위 해상도 축소와 객체 중심 리타겟팅이 중첩되며, 객체 크기 감소 효과가 누적되는 문제가 발생한다. 이러한 현상은 특히 고해상도 영상에서 다수의 소형 객체가 포함된 경우 두드러지며, 객체가 기계 추론 모델의 최소 검출 크기 이하로 축소되는 사례를 빈번히 유발한다. 그 결과 객체 인식 성능 저하와 함께 RD-curve의 비교 가능 조건이 붕괴되어 BD-rate와 같은 성능 지표가 정상적으로 산출되지 않는 평가 불안정성이 발생한다.

본 장에서는 이러한 문제를 해결하기 위해, 객체의 최소 검출 크기를 명시적으로 고려하여 Spatial Resampling의 스케일 계수를 보정하는 객체 크기 기반 스케일 계수 보정 기법을 제안한다. 제안 기법은 RoI-based processing으로 인해 이미 축소된 객체가 Spatial Resampling 단계에서 추가로 과도하게 축소되는 것을 방지함으로써, 두 도구의 누적 축소 효과를 억제하는 것을 목표로 한다.

1. 최소 검출 객체 크기의 정의

일반적인 객체 검출 모델은 합성곱 신경망 기반 구조를

가지며, 입력 영상으로부터 다단계 피쳐 맵을 생성하기 위해 일정한 간격의 스트라이드(stride)를 사용한다. 대표적으로 스트라이드 8, 16, 32를 사용하는 다중 해상도 피쳐 맵 구조가 널리 활용된다. 이러한 구조에서 소형 객체는 주로 가장 작은 스트라이드의 피쳐 맵에서 검출되며, 안정적인 검출을 위해서는 객체가 해당 피쳐 맵 상에서 일정 개수 이상의 피쳐 셀을 차지하는 것이 일반적으로 요구된다.

이를 입력 영상의 공간 해상도로 환산하면, 객체의 짧은 변 길이가 일정 픽셀 이상일 때 검출 성능이 안정적으로 유지되는 경향이 있다. 이에 따라 본 연구에서는 기계 추론 모델이 안정적으로 객체를 인식할 수 있는 최소 객체 크기(임계 크기)를 $\tau = 16$ 픽셀로 정의하였다. 이 임계 크기는 이후 스케일 계수 보정 과정에서 객체 축소 허용 여부를 판단하는 기준으로 사용된다.

2. 객체 크기 기반 스케일 계수 보정 기법 및 안정화 효과

객체 크기 기반 스케일 계수 보정 기법은 **Spatial Resampling** 이후에도 모든 객체가 임계 크기 τ 이상을 유지하도록 스케일 계수를 조정한다. 입력 프레임에 포함된 각 객체를 j 라 할 때, 객체의 짧은 변 길이를 실제 객체 크기 a_j 로 정의한다. 또한 **RoI-based processing**으로 인한 객체 축소 비율을 r_j , **Spatial Resampling** 단계에서 적용되는 스케일 계수를 s 라 하면, **Spatial Resampling** 이후 객체의 실제 크기 a'_j 는 a_j, r_j, s 의 곱인 $a'_j = a_j \times r_j \times s$ 로 표현된다. 객체가 안정적으로 검출되기 위해서는 $a'_j \geq \tau$ 조건을 만족해야 하므로, 이를 스케일 계수 s 에 대해 정리하면 $s \geq \tau / (a_j \times r_j)$ 조건을 얻을 수 있다.

즉, 각 객체 j 에 대해 최소 검출 크기를 만족시키기 위한 최소 스케일 계수가 존재하며, 프레임 내 모든 객체에 대해 해당 조건을 평가할 수 있다. 본 연구에서는 프레임 내 객체 중 가장 보수적인 조건을 만족하도록 하기 위해, 필요한 최소 스케일 계수의 최댓값을 $s_{adjusted} = \max \{ \tau / (a_j \times r_j) \}$ 로 정의한다. 이는 해당 프레임에서 모든 객체가 최소 검출 크기 이상으로 유지되기 위해 요구되는 보정 스케일 계수를 의미한다. 이후 이 값과 **Spatial Resampling**의 사전 분석 단계에서 결정된 초기 스케일 계수 $s_{initial}$ 을 비교하여, 두 값 중

더 큰 값을 최종 스케일 계수 $s_{final} = \max(s_{initial}, s_{adjusted})$ 로 적용한다. 그림 4는 이러한 객체 크기 기반 스케일 계수 보정 알고리즘의 흐름도를 보여준다.

제안 기법은 **Spatial Resampling** 단계에서 객체 크기를 직접 제약함으로써, **RoI-based processing** 이후에도 객체가 최소 검출 크기 이하로 축소되는 상황을 사전에 차단한다. 이를 통해 객체 인식 성능의 급격한 저하를 방지할 뿐만 아니라, **RD-curve**의 형태 왜곡을 완화하여 실험 기준과 비교 대상 간 비교 가능한 성능 구간을 보다 안정적으로 확보할 수 있다.

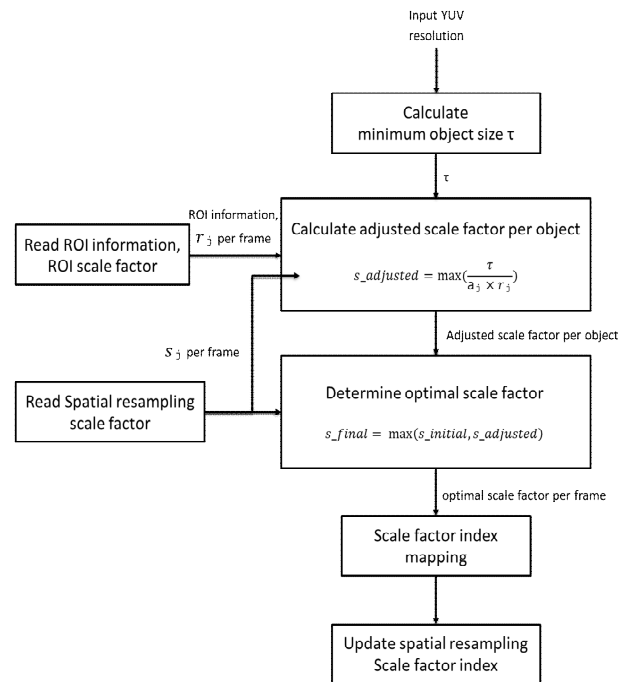


그림 4. 객체 크기 기반 스케일 계수 보정 기법의 흐름도
 Fig. 4. Flowchart of the Object Size-Based Scale Adjustment Method

V. 객체 크기 기반 스케일 계수 보정 적용에 따른 실험 결과

본 장에서는 **Spatial Resampling**과 **RoI-based processing**이 동시에 적용될 때, 객체 크기 기반 스케일 계수 보정 기법 적용 전·후의 실험 결과를 정량적으로 비교·분석한다. 모든 실험은 **VTM**을 앵커로 설정하였으며, **SFU** 데이터셋을 대상으로 수행되었다. 본 장의 목적은 성능의 절대적 개

선 여부를 논의하는 것이 아니라, 결합 조건에서 반복적으로 발생하던 BD-rate 계산 불가 및 성능 불안정성 문제가 실제로 완화되었는지를 수치적으로 확인하는 데 있다.

보정 기법 적용 이전의 실험 결과에서는 다수의 클래스와 환경에서 평가 지표가 정상적으로 산출되지 않는 문제가 발생하였다. RA 환경에서 Class A는 BD-rate 42.64%, BD-mAP -3.12로 큰 성능 저하가 관측되었으며, Class B는 BD-rate -47.70%가 계산되었으나 BD-mAP는 산출되지 않았다. Class C와 Class D 역시 각각 22.77%, 30.78%의 BD-rate 증가와 함께 BD-mAP 감소가 관측되었고, SFU 평균 BD-rate는 9.48%로 계산되었으나 평균 BD-mAP는 계산 불가 상태였다. LD 환경에서는 Class A와 Class C에서 각각 37.10%, 22.44%의 BD-rate 증가가 발생하였고, Class D에서는 BD-rate 자체가 산출되지 않아 평균 지표 계산이 불가능하였다. AI 환경에서는 문제 양상이 더욱 심각하여, Class A와 Class B에서 BD-rate가 정의되지 않았고, 평균 BD-rate와 BD-mAP 모두 연산 불가 상태로 남았다. 이러한 결과는 Spatial Resampling과 RoI-based processing의 누적 축소 효과로 인해 RD-curve 비교 조건이 붕괴되었음을 수치적으로 보여준다.

이에 비해 객체 크기 기반 스케일 계수 보정 기법을 적용한 이후에는 모든 환경과 모든 클래스에 대해 BD-rate와 BD-mAP가 예외 없이 정상적으로 산출되었다. 표 4는 이러한 객체 크기 기반 스케일 계수 보정 기법 적용 후의 실험 결과를 보여준다.

RA 환경에서 Class A의 BD-rate는 기존 42.64%에서 12.68%로 감소하였고, BD-mAP 역시 -3.12에서 -0.96으로 완화되었다. Class C 또한 BD-rate가 22.77%에서 6.03%로 크게 감소하였다. 기존에 BD-mAP가 산출되지 않았던 Class B는 -22.74%의 BD-rate와 +0.80의 BD-mAP가 동시에 도출되어, 결합 조건에서도 성능 비교가 가능해졌다. SFU 평균 기준으로는 BD-rate가 9.48%에서 4.18%로 감소하였으며, 평균 BD-mAP 역시 연산 불가 상태에서 -0.23으로 안정적으로 계산되었다. LD 환경에서도 동일한 경향이 확인되었다. 적용 이전에는 평균 지표가 전혀 산출되지 않았으나, 적용 이후에는 평균 BD-rate -5.46%, 평균 BD-mAP +0.20으로 계산되었다. 특히 기존에 BD-rate가 연산 불가였던 Class D는 BD-rate -6.99%로 정상적인 수치가 도출되

표 4. 객체 크기 기반 스케일 계수 보정 기법 적용 후의 실험 결과
Table 4. Performance Results after Applying Object Size-Based Scale Adjustment

	Object Detection	End-to-End BD-Rate [%] mAP	End-to-End BD-mAP
Random Access	Class A	12.68%	-0.96
	Class B	-22.74%	0.80
	Class C	6.03%	-0.37
	Class D	20.41%	-0.67
	Average	4.18%	-0.23
Low Delay	Class A	7.60%	-0.93
	Class B	-27.82%	1.50
	Class C	9.58%	-0.44
	Class D	-6.99%	0.14
	Average	-5.46%	0.20
All Intra	Class A	4.31%	-0.19
	Class B	-20.32%	1.09
	Class C	-3.97%	0.40
	Class D	-5.31%	0.20
	Average	-7.81%	0.46

었고, Class A와 Class C 역시 각각 7.60%, 9.58%로 BD-rate 증가 폭이 현저히 완화되었다. AI 환경에서는 적용 이전 평균 BD-rate와 BD-mAP가 모두 정의되지 않았던 반면, 적용 이후에는 평균 BD-rate -7.81%, 평균 BD-mAP +0.46으로 안정적인 성능 지표가 확보되었다. Class A와 Class B 역시 기존에는 비교가 불가능하였으나, 적용 이후 각각 BD-rate 4.31%, -20.32%로 명확한 수치 비교가 가능하다.

이러한 수치 비교 결과는 객체 크기 기반 스케일 계수 보정 기법이 Spatial Resampling과 RoI-based processing 결합 조건에서 발생하던 구조적인 평가 불안정성 문제를 실질적으로 완화하였음을 보여준다. SFU뿐만 아니라 TVD 데이터셋에서도 동일한 지표 정상화 및 RD-curve 안정화 경향이 확인되었으며, 이는 제안 기법이 특정 데이터셋에 국한되지 않고 도구 간 상호작용에 기인한 구조적 문제를 일반적으로 완화할 수 있음을 시사한다. 적용 이전에는 특정 클래스나 환경에서 BD-rate가 과도하게 증가하거나 지표 자체가 정의되지 않는 문제가 빈번히 발생하였으나, 적용 이후에는 모든 조건에서 RD-curve 비교 가능성이 확보되었고, 평균 성능 지표 역시 의미를 갖는 형태로 회복되었다. 일부 클래스에서는 여전히 BD-rate 증가가 관측되지만, 이는 평가 지표가 정상적으로 정의된 상태에서 도출된 결과로서, 결합 환경에서의 성능 해석이 가능하다는 점에서 의의가 있다.

VI. 결론

본 논문에서는 Spatial Resampling과 RoI-based processing이 동시에 적용될 때 발생하는 부호화 도구 간 상호작용 문제를 분석하고, 이로 인해 나타나는 평가 지표 불안정성 문제를 해결하기 위한 객체 크기 기반 스케일 계수 보정 기법을 제안하였다. 기존 CEO 실험에서는 개별 도구의 성능 기여도를 분석하기 위해 단독 비활성화 조건이 주로 사용되어 왔으나, 이러한 방식은 실제 파이프라인 환경에서 여러 도구가 동시에 활성화될 때 발생하는 상호작용 효과를 충분히 설명하지 못하는 한계를 가진다.

본 연구에서는 Spatial Resampling과 RoI-based processing을 각각 단독으로 활성화한 단독 활성화 실험과, 두 도구를 동시에 적용한 결합 실험을 수행하여 성능 결과를 분석하였다. 실험 결과, 두 도구를 동시에 적용할 경우 일부 클래스에서는 비트를 절감 효과가 유지되는 반면, 다른 클래스에서는 성능 저하가 과도하게 증폭되거나 BD-rate 및 BD-mAP가 정상적으로 산출되지 않는 문제가 빈번히 발생함을 확인하였다. 이는 Spatial Resampling에 의한 프레임 해상도 축소와 RoI-based processing의 객체 중심 리타겟팅이 중첩되면서 객체 크기 축소 효과가 누적되고, 그 결과 RD-curve 비교 조건이 붕괴되는 구조적 문제로 해석할 수 있다. 이러한 문제를 해결하기 위해 객체의 최소 검출 크기를 고려하여 Spatial Resampling의 스케일 계수를 보정하는 객체 크기 기반 스케일 계수 보정 알고리즘을 제안하였다. 제안 기법은 프레임 내 모든 객체가 탐지기가 안정적으로 인식할 수 있는 최소 크기 이상을 유지하도록 스케일 계수를 조정함으로써, 소형 객체의 과도한 축소를 방지하고 Spatial Resampling과 RoI-based processing의 누적 축소 효과를 제어한다.

실험 결과, 제안 기법을 적용한 이후에는 RA, LD, AI 환경의 모든 클래스에서 BD-rate와 BD-mAP가 예외 없이 정상적으로 산출되었으며, 기존 결합 실험에서 반복적으로 발생하던 평가 지표 연산 불가 문제가 효과적으로 해소되었다. 또한 과도하게 증가하던 BD-rate가 완화되고, 클래스 및 환경 간 성능 분산이 감소하여 평균 성능 지표가 의미를 갖는 상태로 회복되었음을 확인하였다. 이는 제안 기법이 단순한 성능 개선을 목표로 하는 것이 아니라, 결합 도구 환경에서의 평가 안정성과 실험 결과 해석 가능성을 확보하는 데 핵심적인 역할을 수행함을 의미한다.

참고 문헌 (References)

- [1] J. Zhang and D. Tao, "Empowering things with intelligence: A survey of the progress, challenges, and opportunities in artificial intelligence of things," *IEEE Internet of Things Journal*, Vol. 8, No. 10, pp. 7789 - 7807, May 2021.
- [2] Information technology - Coding of audio-visual objects - Part 10: Advanced Video Coding, ISO/IEC 14496-10:2023, 2023.
- [3] Information technology - High efficiency coding and media delivery in heterogeneous environments - Part 2: High Efficiency Video Coding, ISO/IEC 23008-2:2023, 2023.
- [4] Information technology - Coded representation of immersive media - Part 3: Versatile Video Coding, ISO/IEC 23090-3:2023, 2023.
- [5] ISO/IEC JTC 1/SC 29/WG 2, "Call for Proposals on Video Coding for Machines," ISO/IEC, Output Document N00246, October 2022.
- [6] ISO/IEC JTC 1/SC 29/WG 2, "Call for Evidence on Video Coding for Machines," ISO/IEC, Output Document N00215, July 2022.
- [7] ISO/IEC JTC 1/SC 29/WG 4, "Common test conditions for video coding for machines," ISO/IEC, Output Document N00592, November 2024.
- [8] ISO/IEC JTC 1/SC 29/WG 4, "Algorithm description of tools in vcm reference software," ISO/IEC, Output Document N00593, November 2024.
- [9] ISO/IEC JTC 1/SC 29/WG 2, "Use cases and requirements for video coding for machines," ISO/IEC, Output Document N00190, April 2022.
- [10] H. Kwon, S. Cheong, J. Choi, T. Lee, and J. Seo, "Standardization trends in video coding for machines," *Electronics and Telecommunications Trends*, Vol. 35, No. 5, pp. 102 - 111, October 2020.
- [11] W. Gao, S. Liu, X. Xu, M. Rafie, Y. Zhang, and I. Curcio, "Recent standard development activities on video coding for machines," *arXiv preprint*, arXiv:2105.12653, May 2021. Available: <https://arxiv.org/abs/2105.12653>
- [12] D. Lee, S. Jeon, Y. Jeong, J. Kim, and J. Seo, "Exploring the Video Coding for Machines Standard: Current Status and Future Directions," *Journal of Broadcast Engineering*, Vol. 28, No. 7, pp. 888 - 903, 2023.
- [13] ISO/IEC JTC 1/SC 29/WG 4, "VCM-CE 0: Tool experiments," ISO/IEC, Output Document N00732, October 2025.
- [14] N. Barman, M. Martini, and Y. Reznik, "BjØntegaard Delta (BD): A Tutorial Overview of the Metric, Evolution, Challenges, and Recommendations," *arXiv preprint*, arXiv:2401.04039, 2024. Available: <https://arxiv.org/abs/2401.04039>
- [15] A. Kim, E. An, K. Seo, S. Jung, W. Cheong, J. Lee, and H. Choo, "[VCM] Adaptive spatial resampling based on MI RPR for VCM," Yonsei University and ETRI, Geneva, Switzerland, ISO/IEC JTC 1/SC 29/WG 4 Input Document m64124, July 2023.
- [16] E. An, A. Kim, S. Jung, S. Kwak, J. Lee, W. Cheong, H. Choo, and K. Seo, "Adaptive spatial down-sampling method based on object occupancy distribution for video coding for machines," *EURASIP Journal on Image and Video Processing*, Vol. 2024, No. 36, 2024. doi: <https://doi.org/10.1186/s13640-024-00647-y>
- [17] S. Kim, M. Jeong, J. Lee, H. Lee, H. Choo, and S. Jung, "[VCM] Cfp

response: Region-of-interest based video coding for machine,” ETRI, Mainz, Germany, ISO/IEC JTC 1/SC 29/WG 2 Input Document m60758, October 2022.

- [18] O. Stankiewicz, T. Grajek, S. Mackowiak, J. Stankowski, S. Rozek, M. Lorkiewicz, M. Wawrzyniak, and M. Domanski, “Region-of-Interest-Based Video Coding for Machines,” Proceedings of the IEEE International Conference on Multimedia and Expo Workshops (ICMEW), Los Alamitos, CA, USA, pp. 1 - 6, July 2024.
doi: <https://doi.org/10.1109/ICMEW63481.2024.10645441>
- [19] H. Choi, E. Hosseini, S. Ranjbar Alvar, R. Cohen, I. Bajić, A. Karabutov, Z. Yin, and E. Alshina, “A dataset of labelled objects on

raw video sequences,” Data in Brief, Vol. 34, Art. 106701, 2021.
doi: <https://doi.org/10.1016/j.dib.2020.106701>

- [20] X. Xu, S. Liu, and Z. Li, “Tencent Video Dataset (TVD): A video dataset for learning-based visual data compression and analysis,” arXiv preprint arXiv:2105.05961, May 2021. [Online]. Available: <https://arxiv.org/abs/2105.05961>
- [21] E. An, M. Kim, K. Seo, S. Kwak, A. Kim, S. Jung, W. Cheong, J. Lee, and H. Choo, “[VCM] Report on the non monotonic bitrate issues,” Yonsei University and ETRI, Geneva, Switzerland, ISO/IEC JTC 1/SC 29/WG 4 Input Document m70693, November 2024.

저 자 소 개



김민석

- 2024년 2월 : 연세대학교 컴퓨터정보통신공학부 학사
- 2026년 2월 : 연세대학교 전산학과 석사
- ORCID : <https://orcid.org/0009-0002-5665-3740>
- 주관심분야 : 기계를 위한 비디오 코딩, 몰입형 미디어



신효승

- 2023년 3월 ~ 현재 : 연세대학교 소프트웨어학부 학사과정
- ORCID : <https://orcid.org/0009-0003-1817-1625>
- 주관심분야 : 영상처리, 비디오 코딩



박세희

- 2023년 3월 ~ 현재 : 연세대학교 소프트웨어학부 학사과정
- ORCID : <https://orcid.org/0009-0009-2505-9766>
- 주관심분야 : 비디오 코딩



서광덕

- 1996년 2월 : 한국과학기술원(KAIST) 전기공학과 학사
- 1998년 2월 : 한국과학기술원(KAIST) 전기공학과 석사
- 2002년 8월 : 한국과학기술원(KAIST) 전기공학과 박사
- 2002년 8월 ~ 2005년 2월 : LG전자 선임연구원
- 2012년 9월 ~ 2013년 8월 : 미국 플로리다 대학교 초빙교수
- 2005년 3월 ~ 현재 : 연세대학교 소프트웨어학부 교수
- ORCID : <http://orcid.org/0000-0001-5823-2857>
- 주관심분야 : 비디오 코딩, 시각적 통신, 디지털 방송, 멀티미디어 통신 시스템