

# 기계를 위한 비디오 부호화 성능 개선을 위한 향상된 공간적 재표본화 기술

김민석 / 연세대학교 전산학과 영상통신연구실 (Visual Communications Lab.)

최근 객체 탐지, 추적, 분류와 같은 기계 추론 기술이 자율주행, 스마트 시티, 지능형 감시 등 다양한 응용 분야로 확산되면서, 비디오 데이터는 더 이상 인간의 주관적 시청 품질만을 고려하여 소비되는 대상에 그치지 않고, 기계가 영상으로부터 의미 있는 특징을 안정적으로 추출하고 활용하기 위한 입력 데이터로서의 역할 또한 중요해지고 있다. 이러한 환경에서는 화질의 시각적 우수성보다, 기계 추론에 직접적으로 기여하는 객체의 크기, 형태, 경계와 같은 특징 정보의 보존 여부가 더욱 중요한 요소로 작용한다.

그러나 기존의 비디오 부호화 기술은 인간 시각 특성(Human Visual System, HVS)을 중심으로 설계되어, 시각적으로 덜 민감한 정보의 제거를 통해 압축 효율을 높이는 데 초점을 두고 있다. 이러한 접근은 인간 시청 관점에서는 효과적일 수 있으나, 기계 추론에 중요한 특징 정보까지 함께 손실될 가능성이 존재한다. 이와 같은 배경에서 MPEG에서는 인간 시각 중심 부호화 방식의 한계를 보완하고, 기계 인식 성능을 고려한 새로운 부호화 패러다임을 정립하기 위해 VCM(Video Coding for Machines) 표준화를 추진하고 있다. VCM 표준에는 다양한 도구

들이 정의되어 있으며, 그중 공간적 리샘플링(Spatial Resampling)은 비트율 절감과 기계 인식 성능 유지의 균형을 달성하기 위한 핵심 전처리 기술로 채택되었다.

VCM-RS(VCM Reference Software) 파이프라인에서 공간적 리샘플링은 실제 해상도 축소 이전에 사전분석 단계를 통해 프레임 내 객체 분포를 요약하고, 이를 기반으로 스케일 계수를 결정한다. 구체적으로, 객체 점유 분포(Object Occupancy Distribution, OOD)는 초기 스케일 계수를 산출하는 데 사용되며, 이후 평균 점유 면적비(Average Object Area Ratio, Average OAR)와 입력 해상도 점수를 결합하여 해당 계수를 단계적으로 보정한다. 이렇게 보정된 값이 최종 공간적 리샘플링 스케일 계수로 적용된다. 이 방식은 계산 복잡도가 낮고 다양한 콘텐츠에서 안정적으로 동작한다는 장점이 있으나, 프레임 내 객체 크기 분포의 다양성을 충분히 반영하지 못한다는 한계를 가진다. 특히 소형 객체와 대형 객체가 혼재된 경우 평균화 과정에서 소형 객체의 영향이 희석되며, 결과적으로 해상도 축소 이후 객체 크기가 탐지기의 최소 감지 한계 이하로 감소할 가능성이 존재한다.

이러한 문제는 고해상도 영상이나 관심 영역 기반 처리

(ROI based processing)와 결합될 때 더욱 두드러진다. 공간적 리샘플링은 관심 영역 기반 처리 이전에 수행되므로, 이후 관심 영역 축소로 인한 객체 크기 변화가 스케일 계수 결정 과정에 반영되지 않는다. 그 결과 공간적 리샘플링 단계에서 분석된 객체 크기와 실제 부호화 이후의 객체 크기 사이에 불일치가 발생하며, 이는 기계 추론 성능 저하로 이어질 수 있다. 따라서 기존 평균 기반 스케일 계수 보정 방식만으로는 다양한 해상도 조건과 모듈 간 상호작용을 동시에 고려하는 데 한계가 있다.

본 논문에서는 이러한 구조적 한계를 해결하기 위해 입력 영상의 해상도 조건에 따라 서로 다른 스케일 계수 보정 방식을 선택적으로 적용하는 하이브리드 스케일 계수 보정 기법을 제안한다. 제안 기법은 저해상도 환경과 고해상도 환경을 구분하여 동작한다. 저해상도 환경에서는 기존 Average OAR 기반 보정 방식을 유지하여 전체적인 부호화 효율과 안정성을 확보하고, 고해상도 환경에서는 객체 크기 정보를 직접 고려하는 객체 크기 기반 보정 방식을 적용한다. 객체 크기 기반 보정 방식은 공간적 리샘플링과 관심 영역 기반 처리 이후에도 각 객체의 크기가 기계 추론 모델이 안정적으로 탐지할 수 있는 최소 임계 크기 이상을 유지하도록 스케일 계수를 조정한다. 이를 통해 소형 객체의 과도한 축소를 방지하고, 고해상도 입력 영상에서도 기계 인식 성능의 급격한 저하를 완화한다.

본 연구의 제안기법은 VCM-RS에 구현되었으며, 실험은 VCM CTC(Common Test Conditions)에 따라 AI(All Intra)와 RA(Random Access) 부호화 환경에서 수행되었다. 성능 평가는 BD(Bjontegaard delta)-rate를 기준으로 압축 효율을 비교하였으며, 이때의 BD-rate는 기계 임무 정확도 기반의 BD-rate를 사용하였다. 실험 결과, 제안하

는 하이브리드 스케일 계수 보정 기법은 기존 VCM-RS 대비 SFU 데이터셋을 이용한 객체 검출(Object Detection) 실험에서 RA(Random Access) e2e 환경 기준 평균 -0.65%의 BD-rate 감소, LD(Low Delay) e2e 환경에서는 평균 -3.99%의 BD-rate 감소, 그리고 AI(All Intra) e2e 환경에서도 평균 -0.98%의 BD-rate 감소를 달성하였다. 특히 SFU Class A 및 Class B와 같이 고해상도 영상이 포함된 조건에서 BD-rate 감소 폭이 상대적으로 크게 나타나, 제안 기법이 고해상도 검출 환경에서 더욱 효과적으로 동작함을 확인하였다. 한편, TVD 데이터셋을 이용한 객체 추적(Object Tracking) 실험에서는 RA e2e 환경에서 평균 -6.08%, AI e2e 환경에서 평균 -7.37%의 BD-rate 감소가 관찰되어, 추적 과제에서도 제안 기법이 압축 효율을 유의미하게 개선함을 확인하였다. 이러한 결과는 제안하는 하이브리드 스케일 계수 보정 기법이 부호화 구성(RA, LD, AI)과 기계 임무 유형(Detection, Tracking)에 관계없이 안정적으로 동작하며, 특히 고해상도 영상과 관심 영역 기반 처리가 결합된 환경에서 효과적인 성능 향상을 제공함을 보여준다.

본 연구는 VCM 표준 환경에서 공간적 리샘플링의 동작 특성과 한계를 분석하고, 해상도 조건과 객체 크기 특성을 동시에 고려하는 개선 방안을 제시하였다. 제안된 하이브리드 스케일 계수 보정 기법은 기존 평균 기반 접근법의 안정성을 유지하면서도, 고해상도 및 복합 처리 환경에서 발생하는 성능 저하 문제를 완화할 수 있는 구조를 제공한다. 해당 기법은 VCM-RS에 채택되어 v1.2부터 반영되어 있으며, 이를 통해 공간적 리샘플링이 기계 중심 비디오 부호화 환경에서 보다 효과적으로 활용될 수 있는 기반을 마련하였다.



## 김민석

- 2024년 2월 : 연세대학교 소프트웨어학부 학사
- 2026년 2월 : 연세대학교 전산학과 석사졸업(예정)
- 주관심분야 : 기계를 위한 비디오 부호화, 몰입형 미디어