

특집논문 (Special Paper)

방송공학회논문지 제30권 제2호, 2025년 3월 (JBE Vol.30, No.2, March 2025)

<https://doi.org/10.5909/JBE.2025.30.2.112>

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

관심 영역 기반 영상 압축에서 배경 영역이 머신 비전 성능에 미치는 영향

이 예 지^{a)b)}, 윤 경 로^{a)†}, 이 진 영^{b)}, 정 순 흥^{b)}

Impact of Background Regions on Machine Vision Performance in ROI-based Video Compression

Yegi Lee^{a)b)}, Kyoungro Yoon^{a)†}, Jin Young Lee^{b)}, and Soon-heung Jung^{b)}

요 약

최근 지능형 비디오 분석 및 활용이 증가함에 따라 MPEG(Moving Picture Experts Group)에서는 VCM(Video Coding for Machines) 그룹을 구성하여 사람 중심이 아닌 기계 중심의 영상 부호화 표준 기술을 개발하고 있다. 현재 VCM은 VCMRS(VCM Reference Software) 개발과 함께 표준 초안 작업이 진행중이며, 관심 영역 리타겟팅, 공간·시간적 리샘플링, 루마 채널 비트 절단 등 다양한 기술이 제안되었다. 이 중 관심 영역 리타겟팅 기술은 각 프레임에서 관심 영역 정보를 추출한 뒤 해당 영역 외의 배경을 제외하여 원본보다 낮은 해상도로 부호화하는 방식으로, 객체 탐지 및 객체 추적과 같은 신경망에서 높은 압축 효율을 제공한다. 하지만 이러한 방식은 전체 프레임 정보가 요구되는 깊이 추정이나 의미적 분할 작업에서는 한계가 있을 수 있다. 본 논문에서는 146차 MPEG 회의에서 제안된 Pandaset 데이터셋을 활용하여 의미적 분할 환경에서 VCMRS v0.10의 결과와 본 논문에서 제안하는 배경 영역 전송 방식을 비교하고, 배경 정보가 기계 비전 성능에 미치는 영향을 분석하고자 한다.

Abstract

As the demand for intelligent video analysis continues to grow, MPEG(Moving Picture Experts Group) has initiated the development of machine-centric video coding standards through the VCM(Video Coding for Machines) AhG. Along with the development of the VCMRS(VCM Reference Software), the group has begun drafting the standard, incorporating technologies such as ROI(Region of Interest)-based retargeting, spatio-temporal resampling, and luma channel bit truncation. Among them, ROI retargeting technology extracts ROI information from each frame and encodes it at a reduced resolution by excluding background areas. This approach achieves high compression efficiency in neural networks for object detection and tracking tasks. However, it may not be suitable for scenarios requiring full-frame information, such as depth estimation or semantic segmentation. This paper leverages the Pandaset, proposed at the 146th MPEG meeting, to compare the performance of VCMRS v0.10 with the proposed method of transmitting background regions. The study evaluates how background information affects machine vision performance in semantic segmentation tasks and demonstrates the applicability of the proposed approach in such environments.

Keyword : MPEG, VCM, ROI, Machine Vision, Semantic Segmentation

Copyright © 2025 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."

I. 서 론

MPEG은 비디오와 오디오를 포함한 멀티미디어 표준을 개발하는 국제 표준화 기구로 하드웨어의 발전과 여러 요구사항에 따라 다양한 동영상 압축 표준 기술을 개발하고 있다. 기존의 영상 압축 표준 기술은 주로 사람의 소비를 목적으로 비디오 데이터를 효율적으로 저장하고 전송하기 위해 개발되었으며, 원본 동영상 데이터를 더 적은 용량으로 압축하여 관리, 전송, 재생, 저장 등을 용이하게 하는 데 있어 중요한 역할을 하였다. 이러한 기술은 사람의 시각 인지 체계와 동영상의 특징을 기반으로 높은 시각적 품질을 유지하면서 시간적, 공간적, 통계적 중복성을 줄여 높은 압축 성능을 보여준다.

최근 기계가 데이터를 직접 처리하고 분석하는 역할이 점점 중요해지고 있다. 특히, 자율주행, 스마트 시티, 보안 시스템 등 다양한 응용 분야에서 기계 학습 기반 비전 시스템의 활용이 확대되면서, 기계 중심의 영상 압축 기술이 요구되고 있다. 이러한 변화는 기존의 압축 기술이 아닌, 기계가 효율적으로 데이터를 처리할 수 있도록 설계된 새로운 영상 부호화 방식으로의 전환을 필요로 하고 있다. 이러한 요구를 충족하기 위해 MPEG에서는 VCM 표준화 그룹을 구성하여 기계 학습 기반 비전 시스템에서 활용 가능한 영상 압축 기술 개발을 목표로 표준화를 진행중에 있다. 140 차까지 VCM은 WG(Working Group) 2인 Requirement 그룹에서 기계를 위한 영상 부호화 기술 개발을 위한 유즈케이스, 요구사항, 평가 방법 및 데이터셋과 관련하여 논의를 하였으며, 표준 기술 개발과 관련한 여러 후보 기술들이 제안되었다. 140차 회의 이후에 VCM은 WG 4인 비디오 그

룹으로 넘어가 본격적인 표준화 초안 작업 및 참조 소프트웨어 개발을 진행중이다. 현재 VCM 표준화 단계는 WD(Working Draft) 단계에 있으며, 표준화 과정에서 채택된 대표적인 기술로는 관심 영역 리타겟팅^[1], 공간적·시간적 리샘플링^{[2]-[3]}, 그리고 밝기 채널에 대한 비트 절단 기술^[4] 등이 있다.

VCM의 평가 데이터셋 및 방법에 대하여 정의되어 있는 CTC(Common Test Condition)^[5]에서는 SFU 데이터셋^[6]과 TVD-video 데이터셋^[7]을 필수 실험 데이터셋으로 지정하고 있으며, 각각 객체 탐지와 객체 추적 성능을 평가하기 위해 설계된 신경망을 평가 네트워크로 사용하고 있다. VCM 내 기술 채택은 이들 데이터셋을 활용하여 머신 성능 대비 비트율 감소율(BD-rate, Bjøntegaard Delta Rate)을 기반으로 이루어진다. 이러한 접근은 객체 중심 작업 환경에서 기술의 성능을 효과적으로 검증할 수 있으나, 기술 적용 범위를 제한적으로 평가한다는 한계가 있다. 그 중에서도 관심 영역 리타겟팅 기술은 부호화 과정에서 객체 탐지 신경망을 사용하여 각 프레임 내 중요한 관심 영역만을 추출하고, 배경 정보를 제외한 상태로 부호화하는 방식을 따른다. 이 방식은 전송 비트율을 크게 줄이는 동시에 객체 탐지 및 추적과 같은 태스크에서 높은 성능을 보인다. 특히, 관심 영역 정보만을 전송하여 부호화 효율성을 극대화할 수 있어, VCM 표준화 과정에서 핵심 기술로 채택되었다. 그러나 해당 기술은 주로 객체 중심의 작업 환경에 최적화되어 있어, 전체 프레임 정보가 필요한 작업, 예를 들어 의미적 분할과 같은 환경에서는 한계가 발생할 가능성이 있다.

머신 비전 작업은 단순히 객체 탐지와 추적에 국한되지 않으며, 의미적 분할과 같은 태스크 또한 중요한 비중을 차지한다. 의미적 분할 작업에서는 전체 프레임 내의 모든 영역이 분석 대상이 되며, 배경 정보를 포함한 프레임 전체 정보가 필요하다. 그러나 기존 관심 영역 리타겟팅 기술은 배경 영역을 제외하여 부호화 하기 때문에, 이러한 작업 환경에서 기술의 적합성을 평가하는 데 한계가 있다. 이에 따라, 배경 정보가 포함되지 않은 상태에서의 기술 적용이 적절한지에 대한 심도 있는 검토가 필요하다. 146차 MPEG 회의에서는 이러한 한계를 보완하기 위해 Nokia에서 Pandaset^[8]을 제안하였다. Pandaset은 자율주행 연구를 목적으로 설계된 73개의 영상 데이터셋으로, 기존 CTC 테이

a) 건국대학교(Konkuk University)

b) 한국전자통신연구원(Electronics and Telecommunications Research Institute)

‡ Corresponding Author : 윤경로(Kyoungro Yoon)
E-mail: yoonk@konkuk.ac.kr

Tel: +82-2-450-4129

ORCID: <https://orcid.org/0000-0002-1153-4038>

※ 본 논문은 2024년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2020-0-00011, (전문연구실)기계를 위한 영상부호화 기술).

※ 이 논문의 연구 결과 중 일부는 한국방송·미디어공학회 2024년 추계학술대회에서 발표한 바 있음.

· Manuscript January 9, 2025; Revised February 25, 2025; Accepted February 26, 2025.

터셋과는 달리 의미적 분할을 포함한 복잡한 머신 비전 태스크를 평가하기 위한 네트워크를 포함하고 있다. 이는 객체 중심 작업 외에도 전체 프레임 정보가 요구되는 환경에서 기술의 성능을 평가할 수 있는 기회를 제공하며, 기존 평가 방식의 범위를 확장하는 데 기여한다. 그러나 Pandaset은 현재 CTC 내 필수(mandatory) 데이터셋으로 포함되지 않고 선택적(optional) 데이터셋으로 정의되어 있다. 이에 따라, 기술 성능 평가에서 Pandaset을 활용한 분석은 필수적으로 이루어지지 않는 상황이다.

본 논문에서는 Pandaset을 활용하여 현재 VCMRS(VCM Reference Software)에 구현된 관심 영역 리타겟팅 기술이 의미적 분할 환경에서 머신 비전 성능에 미치는 영향을 평가하고자 한다. 또한, 배경 영역을 함께 전송했을 때의 성능 변화를 비교하여 기존 기술의 한계를 보완하고, 보다 효율적이고 유연한 영상 부호화 기술의 방향성을 제시하고자 한다. 이를 통해, 객체 중심 작업뿐만 아니라 전체 프레임 정보를 필요로 하는 다양한 머신 비전 태스크에서도 적용 가능한 영상 부호화 기술의 가능성을 탐구한다.

본 논문의 구성은 다음과 같다. 2장에서는 본 논문에서 제안하는 배경 영역을 추가로 부호화하는 방법을 서술한다. 3장에서는 실험 환경과 결과를 제시하며, 제안된 방법과 기존 기술의 효과성을 비교 분석한다. 마지막으로, 4장에서는

실험 결과를 바탕으로 본 연구의 결론을 도출하고 향후 연구 방향을 논의한다.

II. 배경 프레임 부호화 방법

VCM에 채택된 관심 영역 리타겟팅 기술은 부호화 과정에서 각 프레임별 객체 탐지 신경망을 사용하여 관심 영역을 추출하고, 해당 영역 외부의 배경 정보를 제거하거나 단순화하여 부호화하는 방식이다. 이 기술은 전송 비트율을 효과적으로 감소시키는 동시에, 객체 탐지 및 추적과 같은 태스크에서 높은 효율성을 제공한다. 관심 영역 리타겟팅 과정에서는 중요도가 높은 관심 영역은 화질 손상을 최소화하며 유지하고, 관심 영역 외부는 저역통과 필터를 적용하여 단순화함으로써 부호화 효율을 극대화한다. 최종적으로, 관심 영역 외곽 부분은 제외되거나 원본 해상도보다 작은 해상도로 부호화되어 전송된다. 이러한 방식은 객체 중심 작업에 적합하지만, 의미적 분할과 같이 전체 프레임 정보가 필요한 작업에서는 성능 저하를 초래할 수 있는 한계를 가진다.

본 연구에서는 기존 관심 영역 리타겟팅 기술의 한계를 극복하기 위해, 배경 영역 전송 기능을 추가로 구현하였다. 그림 1은 VCMRS v0.10 기반의 기존 ROI 부호화 컴포넌트

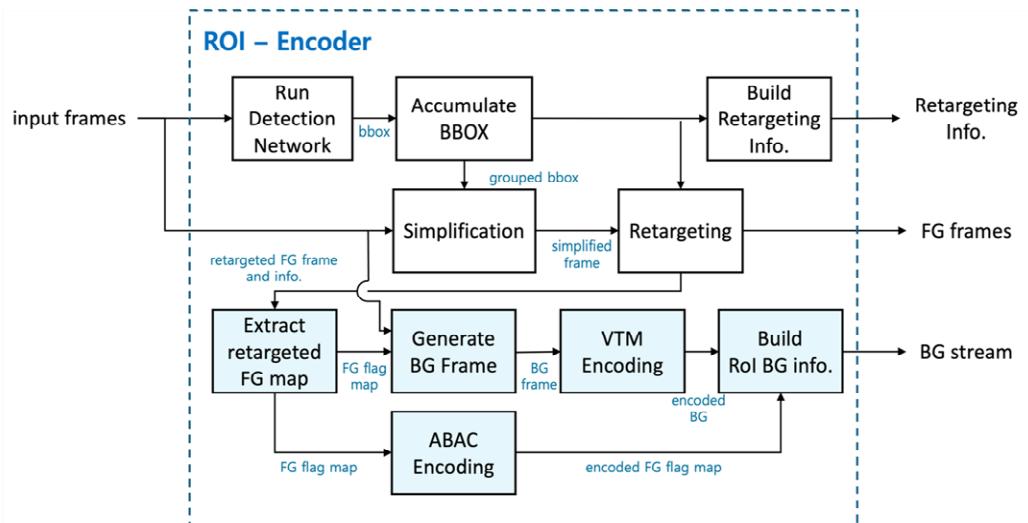


그림 1. ROI 부호화 컴포넌트 내 배경 영역 추출 및 부호화 방법
Fig. 1. Background Extraction and Encoding Method in ROI Encoding Component

구조에 배경 프레임을 추출하고 부호화 과정을 추가한 새로운 인코딩 구조를 나타내며, 파란색 박스로 표시된 부분이 본 연구에서 구현한 배경 영역 전송 과정을 나타낸다. 본 연구에서는 기존 관심 영역 리타겟팅 구조를 확장하여 전경과 배경 영역 모두를 효율적으로 처리할 수 있도록 설계하였다. 부호화 과정에서는 기존 ROI 컴포넌트 과정에 따라 입력 프레임이 객체 탐지 신경망을 통해 분석되어 관심 영역 정보가 추출된다. 추출된 관심 영역 정보는 부호화 모드에 따라 설정된 누적 정보를 기반으로 마진을 적용하여 부호화할 관심 영역이 존재하는 전경 프레임을 생성한다. 이후 단순화 과정에서는 관심 영역 내부는 기존 화질을 유지하고, 관심 영역 외부에는 저역 통과 필터를 적용하여 데이터의 복잡도를 감소시킨다. 리타겟팅 과정에서는 관심 영역 외곽 부분을 제거하고, 원본 해상도보다 낮은 해상도로 리샘플링하여 부호화를 수행한다.

본 논문에서 제안하는 배경 영역 전송은 리타겟팅 이후 전경 프레임과 배경 프레임을 분리하는 과정에서 시작된다. 이 과정에서 생성된 전경 플래그 맵(FG Flag Map)은 리타겟팅된 전경 프레임에서 각 64×64 블록이 관심 영역에 속한지 아닌지를 나타낸다. 플래그 맵은 '1' 또는 '0'의 값을 할당하며, 이를 순차 주사 방식(Progressive scan)을 통해 1 차원 스트림으로 변환하고 적응형 이진 산술 코딩(ABAC, Adaptive Binary Arithmetic Coding)을 통해 부호화한다. 이와 동시에, FG 플래그 맵을 사용하여 배경 프레임이 추

출되며, 추출된 배경 프레임은 부호화 전에 가로와 세로 해상도를 각각 25%로 다운샘플링하여 데이터 크기를 줄인다. 이후 VCMRS의 Inner Codec을 사용하여 AI(All Intra) 모드에서 입력 QP(Quantization Parameter)에 따라 부호화된다. 배경 정보는 기존 ROI 서브 스트림이 아닌, 새로운 ROI_BG 서브 스트림 데이터로 분리 전송된다. ROI_BG 서브 스트림은 전경 플래그 맵과 배경 스트림을 함께 전송하며, 기존 ROI 문법에 영향을 미치지 않고 독립적으로 배경 정보를 처리할 수 있도록 설계되었다.

그림 2는 VCMRS v0.10 기반 기존 ROI 복호화 컴포넌트 구조를 확장하여 배경 프레임 복호화와 관심 영역 프레임 및 배경 프레임을 하나의 통합된 프레임으로 재구성하기 위한 복호화 과정을 나타낸다. 복호화 과정에서는 ROI_BG 서브 스트림으로부터 배경 스트림과 전경 플래그 맵이 분리된다. 배경 스트림은 Inner Codec을 통해 복호화되고, 전경 플래그 맵은 적응형 이진 산술 코딩 복호화기를 통해 복원된다. 복호화된 전경 플래그 맵은 각 픽셀의 전경 및 배경 여부를 식별하며, 이를 기반으로 역리타겟팅된 전경 프레임과 복호화된 배경 프레임이 결합되어 최종 재구성된 프레임이 생성된다. 이 재구성 과정은 기존 리타겟팅 기술에서 배경 정보가 배제됨으로써 발생할 수 있는 한계를 보완하며, 의미적 분할과 같이 전체 프레임 정보가 필요한 작업에서 성능 개선을 기대할 수 있는 기반을 제공한다.

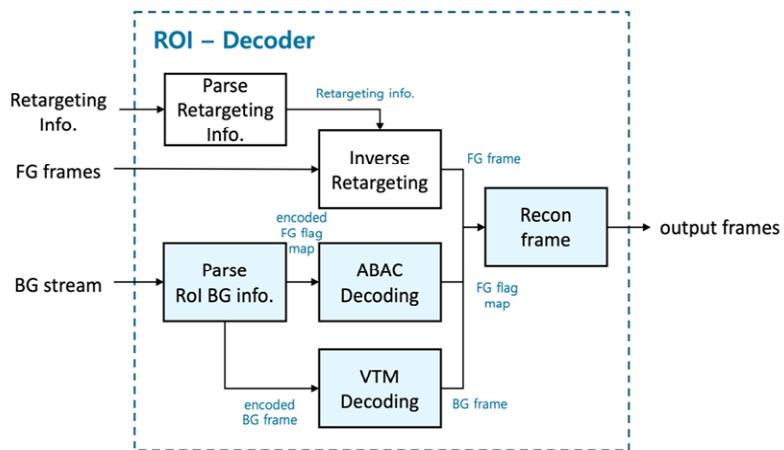


그림 2. ROI 복호화 컴포넌트 내 배경 영역 재구성 방법
 Fig. 2. Background Reconstruction Method in ROI Decoding Component

III. 성능 비교 및 분석

본 연구에서는 Pandaset을 활용하여 실험을 진행하였으며, VCMRS v0.10을 앵커로 설정하고 본 논문에서 제안하는 배경 영역 전송 방식을 적용하여 압축 성능을 비교하였다. 관심 영역 이외의 다른 설정들은 모두 VCM CTC의 기본 조건에 따라 실험이 진행되었다. Pandaset은 1920x1080 해상도의 106개 시퀀스로 구성된 비디오 데이터셋으로, 자율주행 연구를 목적으로 설계되었다. 이 중 73개의 시퀀스가 CTC에 포함되었으며, 기술 제안 시 필수로 제출해야 할 필수 데이터셋이 아닌 선택적 데이터셋으로 정의되어 있다. Pandaset은 데이터셋의 특성과 조건에 따라 A1, B1, C1, A2, B2, C2의 6개 그룹으로 나뉘며, 각 그룹은 A1, B1, C1이 각각 7개, 17개, 12개의 시퀀스로 구성되어 있다. A2, B2, C2는 각각 A1, B1, C1과 유사한 특성을 가지며, 각각

7개, 18개, 12개의 시퀀스로 구성되어 있다. 의미적 분할 성능 평가는 Detectron2의 Panoptic FPN R-101-FPN^[10] 모델을 사용하며, mIoU(Mean Intersection over Union)^[11] 평가 지표를 통해 머신 비전 성능을 계산하였다. 그림 3은 A1, B1, C1 그룹에 속하는 영상 샘플을 보여준다.

표 1은 VCMRS v0.10을 앵커로 설정한 경우, 배경 영역 전송 기법의 BD-mAP 결과를 나타낸다. 일반적으로 비디오 압축 성능 평가는 BD-rate를 주로 사용하지만, 앵커와 테스트 결과 간의 머신 비전 성능이 겹치지 않아 정량적 비교가 어려운 BD-mIoU를 사용하여 성능 비교를 진행하였다.

AI 모드의 결과를 살펴보면, Inner 모드에서는 BD-mIoU가 7.83, E2E(End-to-End) 모드에서는 7.55로 나타나, 두 경우 모두 VCMRS v0.10 대비 성능이 크게 향상된 것을 확인할 수 있다. 이는 배경 영역을 추가로 전송함으로써 기계 학습 모델이 더 많은 정보를 활용할 수 있게 되어 의미

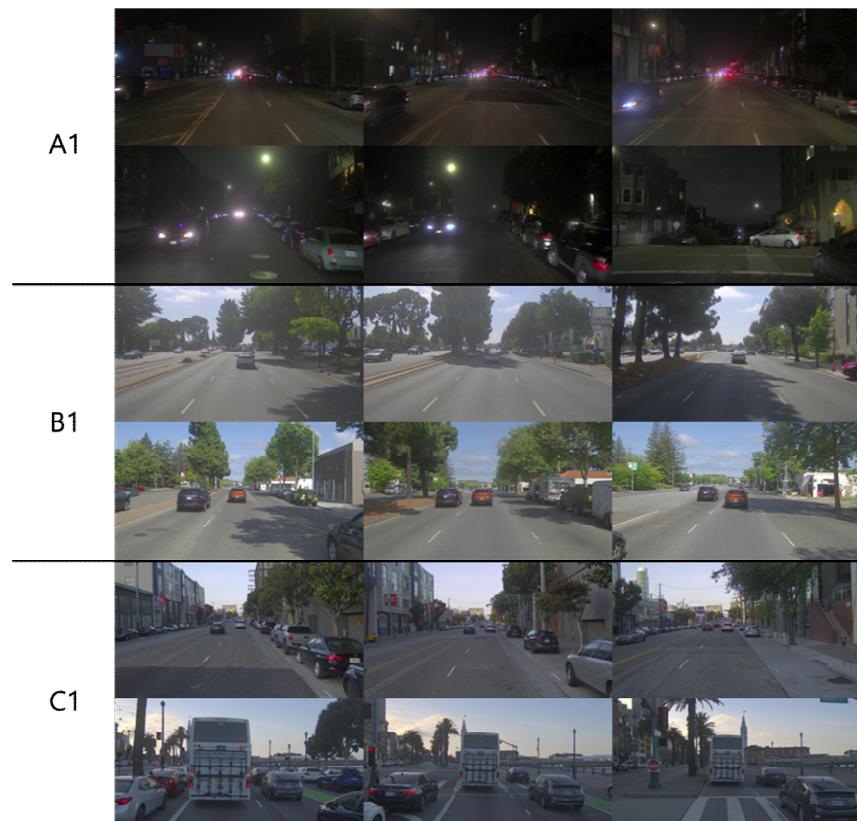


그림 3. Pandaset 영상 샘플
Fig. 3. Sample frames from the Pandaset Dataset

표 1. End-to-End BD-mIoU 결과

Table 1. End-to-End BD-mIoU results

Pandaset	RA_inner	LD_inner	AI_inner	RA_e2e	LD_e2e	AI_e2e
Class A1	-2.93	-4.36	0.99	-2.77	-5.05	3
Class B1	-0.02	-1.47	7.17	-0.91	-3.93	6.49
Class C1	5.91	2.76	11.98	5.52	2.76	11.53
Average (A1,B1,C1)	1.39	-0.62	7.57	0.87	-1.92	7.49
Class A2	-1.97	-3.57	3.66	-3.08	-4.62	5.09
Class B2	0.07	-1.76	7.32	0.02	-4.5	6.27
Class C2	5	1.8	11.81	4.9	2.51	11.07
Average (A2,B2,C2)	1.28	-0.95	8.08	1.01	-2.25	7.6
Average (All)	1.34	-0.79	7.83	0.94	-2.08	7.55

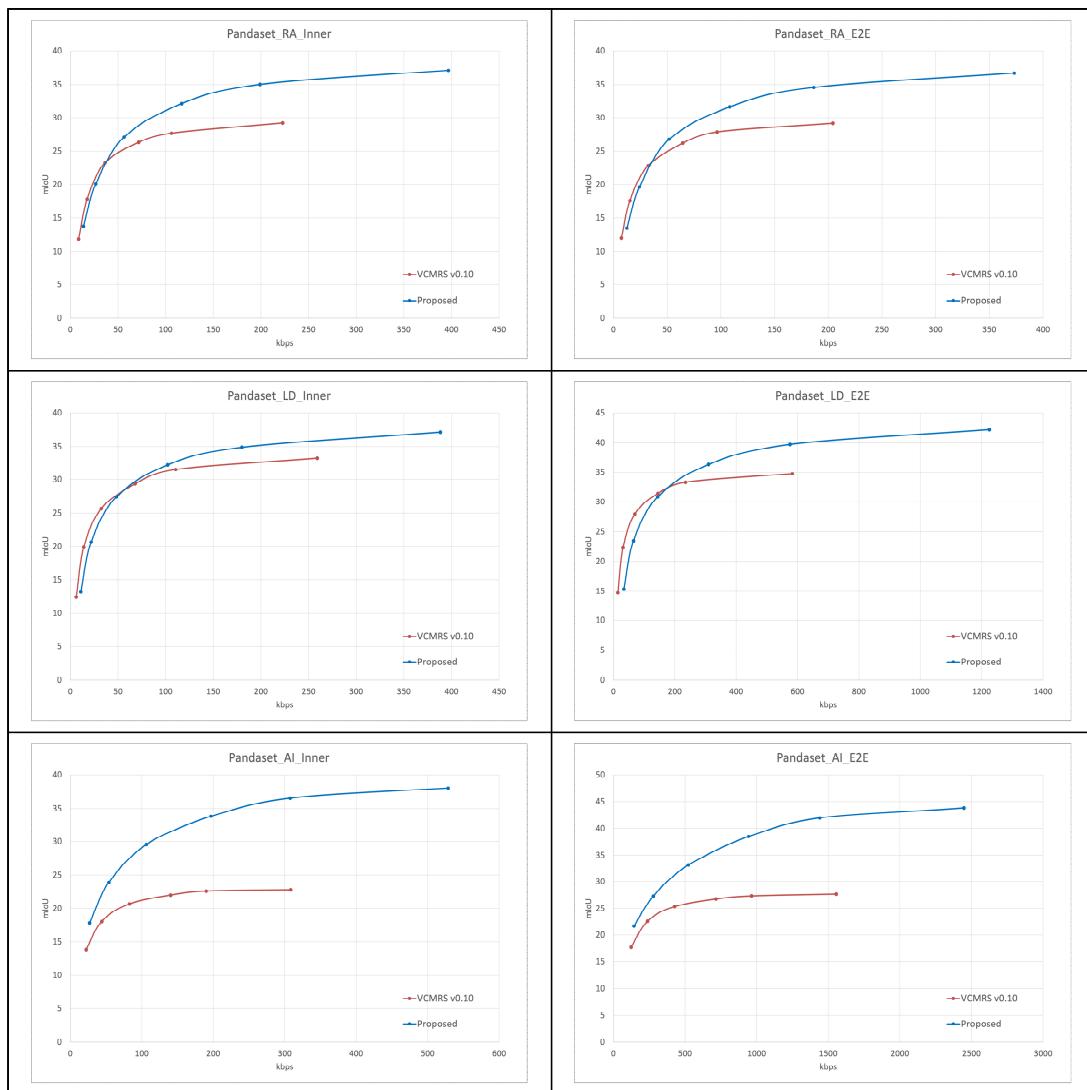


그림 4. 부호화 모드별 Kbps vs mIoU 커브 결과
 Fig. 4. Kbps vs mIoU Curve Results by Encoding Mode

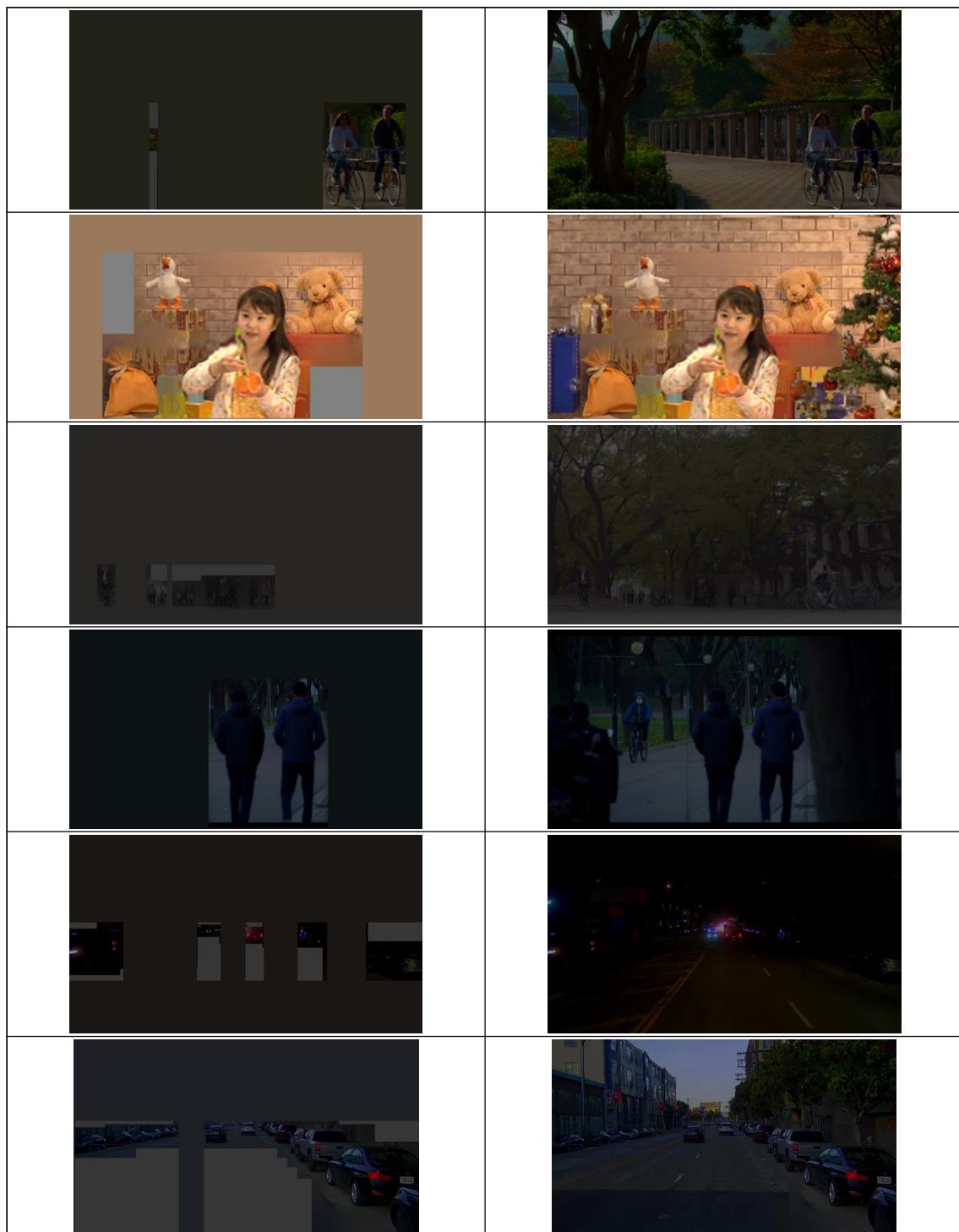


그림 5. VCMRS v0.10 앵커(왼쪽)와 제안된 방법(오른쪽)의 디코딩 샘플

Fig. 5. Decoded Samples of the VCMRS v0.10 anchor(left) and proposed method(right)

적 분할 태스크에서 성능 개선 효과를 가져온 것으로 해석된다. RA(Random Access) 모드의 결과에서는 Inner 모드에서 1.34, E2E 모드에서 0.94로, AI 모드만큼 큰 향상은 아니지만 여전히 VCMRS v0.10 대비 성능이 개선된 것을 확인할 수 있다. 이는 RA 모드에서도 배경 영역 전송이 머신 비전 성능에 긍정적인 영향을 미쳤음을 시사한다. 반면, LD(Low Delay) 모드에서는 Inner와 E2E 모드에서 VCMRS v0.10 대비 성능이 다소 낮게 나타났다. 이는 LD 모드에서 관심 영역 정보를 누적하는 파라미터가 상대적으로 높게 설정되어, 전경 프레임이 차지하는 비율이 크기 때문에 상대적으로 배경 영역 전송의 효과가 제한되었기 때문으로 분석된다.

그림 4는 Pandaset에 대해 각 부호화 모드 별 비트레이트와 mIoU 커브 결과를 보여준다. LD(Low Delay) 모드에서는 낮은 비트레이트 구간에서 앵커 결과가 더 우수한 성능을 보였으나, 높은 비트레이트 구간에서는 배경 영역 전송 방식을 적용한 결과가 더 높은 성능을 나타내는 성능 교차 지점이 확인되었다. 이는 비트레이트가 증가할수록 배경 영역 전송의 성능 향상 효과가 두드러지게 나타남을 시사한다. 앵커의 최고 mIoU 값이 약 35%였던 반면, 배경 영역을 전송하였을 때 42%를 초과하는 성능을 기록하였다. 이는 배경 정보의 전송이 의미적 분할 성능 향상에 긍정적으로 기여하고 있음을 보여준다. 압축 효율이 가장 높게 나왔던 AI 모드에서는 다른 부호화 모드와 달리 성능 교차 지점이 없이 앵커보다 항상 높은 머신 성능을 보여주는 것을 확인할 수 있다. 이는 AI 모드에서 관심 영역 누적 계수가 1로 설정되어 다른 프레임의 관심 영역 정보가 누적되지 않았기 때문이다. 이러한 설정은 전경 프레임의 영역을 줄이고, 배경 영역이 차지하는 비율을 상대적으로 증가시키며, 결과적으로 배경 영역 전송의 중요성을 더 크게 만들었다. 이로 인해 AI 모드에서 특히 높은 성능 향상을 보인 것으로 해석된다.

그림 5는 VCMRS v0.10 기반 디코딩 결과와 본 논문에서 제안하는 배경 영역 전송 결과에 대하여 SFU, TVD-video, Pandaset 디코딩 결과 샘플을 보여준다. VCMRS v0.10의 경우, 디코딩된 프레임에서 관심 영역 정보만 표시되지만, 배경 영역을 전송한 방식에서는 관심 영역 외의 배경 정보도 함께 디코딩되었다. 비록 배경 영역이 낮은 해상도로 압축되었음에도, 주관적으로 보았을 때 배경 정보가 원

본에 가깝게 잘 보존되었음을 확인할 수 있다. 또한, 기존 관심 영역 내 단순화 과정으로 인해 관심 영역과 배경 영역 간 경계에서 아티팩트가 발생했으나, 전반적으로 양호한 화질을 유지하며 관심 영역과 배경 영역이 잘 복원된 것을 확인할 수 있다. 이러한 결과는 배경 정보가 포함됨으로써 디코딩 결과에서 추가적인 시각적 정보를 제공하며, 의미적 분할과 같은 머신 비전 태스크에서 활용 가능성을 높일 수 있음을 시사한다.

IV. 결 론

본 논문에서는 의미적 분할 태스크에서 배경 영역 전송이 머신 비전 성능에 미치는 영향을 분석하기 위해, VCMRS v0.10과 본 연구에서 제안한 배경 영역 전송 기법의 성능을 비교하였다. 현재 VCM 표준에서는 관심 영역만을 추출하고 배경 영역은 전송하지 않는 방식이 채택되어 있다. 이러한 방식은 객체 탐지, 추적과 같은 객체 중심 태스크에서는 적은 비트로 중요한 정보를 효율적으로 전송할 수 있으나, 의미적 분할이나 깊이 추정과 같이 전체 영상의 정보가 필요한 태스크에서는 한계가 있을 수 있다. 본 연구에서는 이러한 한계를 극복하기 위해 배경 영역을 선택적으로 전송하는 방식을 제안하였으며, 실험 결과 AI 모드에서 배경 영역 전송이 기존 VCMRS v0.10 대비 높은 BD-mIoU 성능을 나타냈다. 이는 전경 정보뿐만 아니라 중요한 배경 정보를 함께 활용함으로써 의미적 분할 태스크에서 성능이 개선된 결과로 해석할 수 있다. 또한, RA(Random Access) 모드에서도 일정 수준 이상의 성능 향상이 확인되었으며, LD(Low Delay) 모드에서는 낮은 비트레이트 구간에서 성능 교차 지점이 발생하여 평균 성능이 낮아졌으나, 높은 비트레이트 구간에서는 VCMRS v0.10 대비 더 우수한 머신 비전 성능을 보였다. 본 연구 결과는 배경 정보 전송이 의미적 분할과 같은 전체 영상 정보를 필요로 하는 머신 비전 태스크에서 중요한 역할을 한다는 점을 확인하였다. 본 논문에서는 머신 비전 중심의 영상부호화 기술이 발전하여, 현재 주로 고려되는 객체 중심 태스크뿐만 아니라 다양한 머신 비전 태스크에서도 VCM 기술이 효과적으로 활용될 수 있기를 기대한다.

참 고 문 헌 (References)

- [1] S. Róžek, et al., “[VCM] Improved RoI preprocessing and retargeting for VCM,” ISO/IEC JTC1/SC29/WG4 input document m66523, January 2024.
- [2] H. Wang, et al., “CE4.2 Temporal Resample coding from China Telecom,” ISO/IEC JTC1/SC29/WG4 input document m63086, April 2023.
- [3] A. Kim, et al, “[VCM] Test results for VCM CE3.2,” ISO/IEC JTC1/SC29/WG4 input document m66168, January 2024.
- [4] D. Ding, X. Zhao, Z. Liu and S. Liu, “[VCM] Bitwise Efficiency: Truncating Bit Depth for Machine Video Coding,” ISO/IEC JTC1/SC29/WG4 input document m65525, October 2023.
- [5] WG 04, “Common test conditions for video coding for machines,” ISO/IEC JTC1/SC29/WG4 output document N00543, July 2024.
- [6] H. Choi, E. Hosseini, S. R. Alvar, R. A. Cohen, and I. V. Bajić, “A dataset of labelled objects on raw video sequences,” Data in Brief, 2021, vol. 34, pp. 106701.
- [7] W. Gao, X. Xu, M. Qin, and S. Liu, “An Open Dataset for Video Coding for Machines Standardization,” in 2022 IEEE International Conference on Image Processing (ICIP), 2022, pp. 4008-4012. doi: <http://doi.org/10.1109/ICIP46576.2022.9897525>.
- [8] H. Zhang, et. al., “[VCM] Semantic segmentation for VCM with PandaSet,” ISO/IEC JTC1/SC29/WG4 input document m67677, April 2024.
- [9] VCM-RS v0.10, https://git.mpeg.expert/MPEG/Video/VCM/VCM-RS/-/tree/v0.10?ref_type=tags, (accessed Dec., 22nd, 2024).
- [10] Detectron2, https://github.com/facebookresearch/detectron2/blob/main/MODEL_ZOO.md, (accessed Dec., 22nd, 2024).
- [11] C. L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 4, pp. 834 - 848, Apr. 2018. doi: <http://doi.org/10.1109/TPAMI.2017.2699184>.

저 자 소 개

이 예 지



- 2018년 2월 : 극동대학교 스마트모바일학과 졸업 (학사)
- 2020년 2월 : 건국대학교 스마트ICT융합과 졸업 (석사)
- 2025년 2월 : 건국대학교 컴퓨터공학과 졸업 (박사)
- 2025년 2월 ~ 현재 : 한국전자통신연구원 박사후연수연구원
- ORCID : <https://orcid.org/0000-0002-0292-160X>
- 주관심분야 : 영상처리, 인공지능, 컴퓨터비전

윤 경 로



- 1987년 2월 : 연세대학교 전자전산기공학과 졸업 (학사)
- 1989년 12월 : University of Michigan, Ann Arbor, 전기전산기공학과 졸업 (석사)
- 1999년 5월 : Syracuse University, 전산과학과 졸업 (박사)
- 1999년 6월 ~ 2003년 8월 : LG전자기술원 책임연구원/그룹장
- 2003년 9월 ~ 현재 : 건국대학교 컴퓨터공학부 교수
- 2017년 10월 ~ 현재 : 국립전파연구원 멀티미디어부호화 전문위원회 (JTC1 SC29K) 대표전문위원
- 2019년 10월 ~ 현재 : IEEE SA 2888 WG 의장
- 2024년 10월 ~ 현재 : IEEE SA 디지털콘텐츠기술 표준화 위원회 의장
- ORCID : <https://orcid.org/0000-0002-1153-4038>
- 주관심분야 : 스마트미디어시스템, 멀티미디어검색, 영상처리, 컴퓨터비전, 멀티미디어/메타데이터 처리

저자소개



이진영

- 1998년 5월 : B.S. EECS Michigan State University
- 1999년 12월 : M.S. EECS Michigan State University
- 2008년 12월 : Ph.D. EECS Michigan State University
- 2004년 3월 ~ 현재 : 한국전자통신연구원
- ORCID : 0000-0002-8718-1961
- 주관심분야 : AI기반 영상처리, 멀티미디어 시스템, 메타데이터 처리



정순홍

- 2001년 2월 : 부산대학교 전자공학 (학사)
- 2003년 2월 : KAIST 전기및전자공학 (석사)
- 2016년 2월 : KAIST 전기및전자공학 (박사)
- 2005년 4월 ~ 현재 : 한국전자통신연구원 책임연구원
- ORCID : 0000-0003-2041-5222
- 주관심분야 : 실감미디어, 컴퓨터 비전, 머신러닝, 영상부호화, 영상처리