특집논문 (Special Paper)
방송공학회논문지 제29권 제6호, 2024년 11월 (JBE Vol.29, No.6, November 2024)
https://doi.org/10.5909/JBE.2024.29.6.842
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

VCM 관심 영역 기반 압축 기술 성능 비교

이 예 지[a], 윤 경 로[b]‡

Comparison of RoI-based Video Coding Technologies in VCM

Yegi Lee[a] and Kyoungro Yoon[b]‡

요 약

최근 AI(Artificial Intelligence) 기술의 발전으로 감시 시스템, 자율주행 자동차, 교통 관리 시스템, 의료 등 다양한 산업 분야에서 딥러닝 기반 기술이 적용되고 있으며, 막대한 양의 미디어 데이터가 AI 학습 및 분석에 활용되고 있다. 이에 따라 멀티미디어 기술의 국제 표준화 단체인 MPEG(Moving Picture Experts Group)에서는 이러한 변화에 대응하기 위해 2019년 VCM(Video Coding for Machines) AhG(Adhoc Group)을 설립하고, 머신 비전을 위한 비디오 코딩 기술의 표준화를 진행하고 있다. 현재 VCM에서는 객체 탐지 및 추적 등에 뛰어난 머신비전 성능을 보여주면서 낮은 비트율을 달성하기 위해 공간적 리샘플링, 시간적 리샘플링, 관심 영역 기반 기술 등 다양한 기술들이 제안되고 있다. 그중에서도 관심 영역 기반 기술은 객체 탐지나 추적 등 객체 중심의 머신 작업에서 매우 높은 성능을 보여주고 있으며, 매 회의에서 다양한 기술들이 제안되고 있다. 본 논문에서는 현재까지 제안된 다양한 VCM 기술 중 관심 영역 기반 기술에 대해 주로 논의된 기술들을 살펴보고 성능을 비교 분석하고자 한다.

Abstract

Recently, the rapid advancement of AI(Artificial Intelligence) technologies has led to the application of deep learning across various fields, including surveillance, autonomous vehicles, traffic systems, and the medical sectors. Consequently, a substantial amount of multimedia data is being utilized for AI training and analysis. In response to these changes, the MPEG(Moving Picture Experts Group) recognized the need for new video coding technology and established the VCM(Video Coding for Machines) AhG(Adhoc Group) in 2019 to standardize coding technologies for machine vision. Various contributions have been proposed, suggesting technologies such as temporal resampling, spatial resampling, and RoI(Region of Interest) based coding, which achieve high machine vision performance in object detection and tracking at low bitrates. Among these, RoI-based coding technologies have shown exceptional performance in object-centric machine tasks. This paper reviews and compares the performance of RoI-based coding technologies proposed within the VCM group.

Keyword : MPEG, Video Coding for Machines, Machine Vision, RoI-based Coding

a) 건국대학교 컴퓨터공학과(Dept. of Computer Science and Engineering, Konkuk University)
b) 건국대학교 스마트ICT융합공학과(Dept. of Smart ICT Convergence, Konkuk University)
‡ Corresponding Author : 윤경로(Kyoungro Yoon)
E-mail: yoonk@konkuk.ac.kr
Tel: +82-2-450-4129
ORCID: https://orcid.org/0000-0002-1153-4038
※이 논문의 연구 결과 중 일부는 한국방송·미디어공학회 2024년 하계학술대회에서 발표한 바 있음.
※본 논문은 2024년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임. (No. 2020-0-00011, (전문연구실)기계를 위한 영상부호화 기술)
· Manuscript September 3, 2024; Revised October 16, 2024; Accepted October 17, 2024.

Ⅰ. 서 론

ISO/IEC JTC1 SC29 산하에서 활동하는 MPEG은 비디오와 오디오를 포함한 멀티미디어 표준을 개발하는 국제 표준화 기구로, 1988년에 설립되었다. MPEG은 지난 수십 년간 사용자의 요구사항과 하드웨어 발전에 발맞추어 AVC(H.264, Advanced Video Coding)[1], HEVC(H.265, High Efficiency Video Coding)[2], VVC(H.266, Versatile Video Coding)[3] 등과 같은 다양한 영상 압축 표준을 개발해 왔다. 이를 통해 MPEG은 고품질의 비디오 데이터를 효율적으로 압축, 저장, 전송 및 재생할 수 있도록 하였으며, 디지털 미디어 발전에 핵심적인 역할을 수행해 왔다. 이러한 표준화 기술은 오늘날 비디오 스트리밍, 방송, 저장 매체, 통신 등 다양한 산업 분야에서 필수적인 기술로 자리 잡았다.

2020년도에 작성된 Cisco 보고서[4]에 따르면 비디오 감시, 의료 모니터링, 운송, 패키지 또는 자산 추적 등 점점 더 많은 산업에서 M2M(Machine to Machine) 연결이 증가할 것으로 예상하고 있으며, 2023년 이후 전 세계 장치 및 연결의 절반 이상이 M2M으로 연결된 장치가 될 것으로 예상된다. 이는 기계 간의 통신의 중요성이 점점 더 부각되고 있음을 시사하며, 기존의 인간 중심의 멀티미디어 소비 환경이 기계 중심으로 변화하고 있음을 나타낸다. 이에 더해, Grand View Research[5]는 컴퓨터 비전 기술의 발전이 교육, 의료, 로봇공학, 소비자 전자, 소매, 제조, 보안 및 감시 등 다양한 산업 분야에서 컴퓨터 비전 시스템의 활용 범위를 넓힐 것으로 전망한다. 이러한 기술적 발전은 지능형 영상 분석의 수요를 증가시키고 있으며, 이는 영상 소비 주체가 사람에서 기계로 변화되고 있음을 시사한다.

MPEG에서는 기계를 위한 영상 압축 기술의 새로운 표준 개발 필요성을 인식하여 2019년에 VCM이라는 AhG를 설립하였다. VCM은 초반에 지능형 비디오 분석 및 활용을 목표로 논의되었으며, 기계가 비디오 데이터를 더 적은 비트로 효율적으로 이해하고 분석할 수 있도록 하는 새로운 영상 및 특징맵 압축 기술에 대해 논의를 하였다. 하지만 영상 압축 방식과 특징맵 압축 방식을 분리하여 논의할 필요성이 대두되었고, 136차 회의에서는 이를 두 개의 트랙으로 분리하여, 머신비전을 위한 비디오 압축 관련 기술은 VCM에서, 특징맵 압축 관련 기술은 FCM(Feature Coding for Machines)에서 다루기로 결정하였다. VCM은 140차 회의 이후 WG4(Working Group 4)인 비디오 그룹으로 이전되었으며, 이후 VCM 참조 소프트웨어가 배포되었다. 이 과정에서 시/공간적 리샘플링, 루마 샘플에 대한 비트 절단 기술, 관심 영역 기반 기술 등 다양한 기술이 채택되었다. 특히, 관심 영역 기반 기술의 경우 객체 탐지나 추적과 같은 신경망을 사용하는 응용에서는 객체 주변의 특징맵이 활성화되므로, 이러한 영역이 머신 비전 성능에 큰 영향을 미친다. 이에 따라 매 회의에서는 관심 영역을 효과적으로 추출하고 압축하는 기술, 즉 관심 영역 기반 압축 방법에 대한 다양한 제안들이 활발히 이루어지고 있다. 본 논문에서는 142차 및 145차 회의에서 채택된 관심 영역 기반 기술들과 145차 회의에서 논의된 관심 영역 기반 스케일링 기술을 소개하고, 이들의 성능을 비교 분석하고자 한다.

본 논문의 구성은 다음과 같다. 제2장에서는 VCM에서 논의된 주요 관심 영역 기반 압축 기술들을 설명하고, 제3장에서는 실험 환경과 결과를 제시하며, 이를 바탕으로 각 기술의 효과성을 비교 분석한다. 마지막으로, 제4장에서는 실험 결과를 바탕으로 결론을 도출한다.

Ⅱ. VCM 관심 영역 기반 압축 기술

1. 관심 영역 스태킹 기술

140차 회의 이후, CfP(Call for Proposal)에서 제안된 기술 중 관심 영역 기반 기술과 관련된 후보 기술들은 CE1(Core Experiment 1)에서 논의되었다. CE1은 143차 회의까지 진행되었으며, 총 6개의 관심 영역 기반 기술 중 최종적으로 명지대학교와 한국전자통신연구원에서 제안한 관심 영역 스태킹 기술[6]이 채택되었고, 그 구조는 그림 1과 같다.

이 기술은 부호화 과정에서 프레임별 관심 영역을 추출한 후, 해당 관심 영역으로만 이루어진 전경(FG, Foreground) 부분을 추출하여 전송하는 기술이며, 구체적인 부호화 과정은 다음과 같다. 먼저, 객체 탐지 신경망을 사용하여 각 프레임별 관심 영역을 추출한다. 객체 탐지 신경망은 테스트 시퀀스에 따라 다르게 동작하며, SFU 데이터셋[7]에

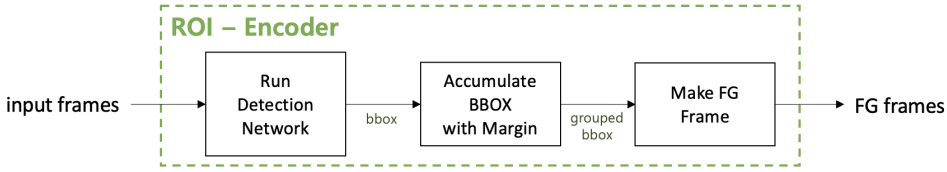

그림 1. 관심 영역 스태킹 기술 구조
Fig. 1. RoI Stacking Technology Architecture

서는 detectron2의 Faster R-CNN X101-FPN[8]을, TVD 비디오 데이터셋[9]에서는 Yolov3[10]을 사용하여 관심 영역을 추출하였다. 그 후, 추출된 관심 영역은 복호화 후 더 높은 머신 성능을 위해 상하좌우 방향으로 일정 픽셀만큼 확장하여 전경 프레임을 생성한다. 이때, RA(Random Access)와 같은 인코딩 모드에서는 효율적인 화면 간 부호화를 위해 일정 프레임 간격으로 추출된 관심 영역을 누적하여 최종 전경 프레임을 생성하였다. 전경 프레임 생성 시, 관심 영역 외의 모든 영역은 회색으로 처리되며, 배경 영역은 전송되지 않도록 설계되었다. 이 기술은 비디오 콘텐츠에서 중요한 정보만을 전송함으로써 비트율을 크게 감소시키는 효과가 있다.

해당 기술은 전경 프레임 내에서 객체 주변의 특징 맵이 신경망의 컨볼루션 수행 시 활성화되는 현상을 고려하여, 수용 영역을 일정 픽셀만큼 확장함으로써 객체 탐지 및 분석 성능을 유지할 수 있었다. 이로 인해 VCM-RS v0.5 앵커와 비교했을 때 BD-rate가 -36.07%로 나타나, 총 6개의 CE1 후보 기술 중 가장 뛰어난 성능을 보였다.

2. 관심 영역 리타켓팅 기술

145차 회의에서는 기존에 채택된 관심 영역 스태킹 기술을 확장한 포즈난대학교의 관심 영역 리타겟팅 기술[11]이 제안되었다. 이 기술은 각 프레임에서 추출된 관심 영역 정보를 바탕으로 리타겟팅을 수행하여, 관심 영역이 포함되지 않은 외곽 부분을 제외함으로써 원본 해상도보다 작은 해상도로 부호화를 수행하는 기술이며, 부/복호화 과정은 그림 2와 같다.

관심 영역 리타겟팅 기술의 부호화 과정은 기존의 스태킹 기술 구조를 확장하여 단순화 과정과 리타겟팅 과정을 추가한 구조로 이루어져 있다. 단순화 과정에서는 중요도가 높은 관심 영역은 기존 화질을 유지하면서, 관심 영역 외의 부분은 저역 통과 필터를 적용하여 단순화한다. 이후, 리타겟팅 과정에서는 관심 영역이 포함되지 않은 외곽 부분을 제외하여 원본 해상도보다 작은 해상도로 부호화를 수행한다. 리타겟팅 과정에서 압축할 해상도를 결정하는 방법은 두 가지로 나뉜다. 첫 번째 방법은 전체 프레임을

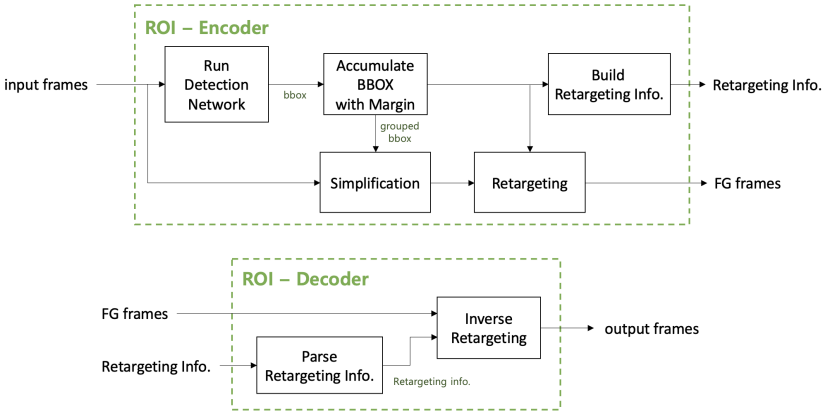

그림 2. 관심 영역 리타켓팅 기술 구조
Fig. 2. RoI Retargeting Technology Architecture

스캔하여 가장 큰 관심 영역의 크기로 해상도를 결정하는 방식이고, 두 번째 방법은 첫 번째 프레임에서 추출된 관심 영역 그룹의 크기를 기준으로 압축할 해상도를 선택하는 방식이다. 압축할 해상도가 결정되면, 각 프레임별로 관심 영역 그룹 외의 부분을 잘라내고, 결정된 해상도에 맞춰 원본 프레임의 관심 영역 그룹 크기를 확장 또는 축소하여 부호화를 진행한다. 또한, 기존 관심 영역 스태킹 기술에서 누적 관심 영역 프레임 수가 64로 고정되었던 것과 달리, 관심 영역 리타겟팅 기술은 RA 모드에서는 Intra Period 단위로, AI(All Intra) 모드에서는 1로 인코딩함으로써 추가적인 압축 효율을 달성하였다. 복호화 과정에서는 잘라낸 부분을 원본 해상도로 재구성하기 위해 역리타겟팅 과정이 수행되며, 이 과정에서 필요한 정보는 부호화 과정에서 전송되고 복호화 과정에서 파싱된다.

관심 영역 리타겟팅 기술은 VCM-RS v0.7 대비 -20.50%의 BD-rate를 보이며 높은 성능을 보여주었지만, 몇 가지 한계점이 존재한다. 전체 시퀀스를 기반으로 해상도를 결정하는 방식은 스트리밍과 같은 실시간 인코딩 모드에서는 적용이 어려운 한계가 있다. 이러한 경우, 시퀀스의 길이가 매우 길다면, 시퀀스의 끝까지 처리한 후에야 해상도를 결정할 수 있어 실시간 처리에 부적합하다. 또한, 첫 번째 프레임의 관심 영역 정보를 바탕으로 해상도를 결정하는 방식은 전체 시퀀스의 해상도를 첫 번째 프레임에만 의존하게 만드는 단점이 있다. 이는 첫 번째 프레임이 전체 시퀀스에서 비중 있는 관심 영역을 포함하지 않거나 그 크기가 실제 중요한 프레임과 상이할 경우, 전체 시퀀스의 해상도 결정이 왜곡될 가능성이 있다. 이러한 단점은 특정 시퀀스나 환경에서 기술의 효율성을 저하시킬 수 있다. 그럼에도 불구하고, 테스트 시퀀스에서 높은 성능을 보여주었기에, 해당 기술이 최종적으로 채택되었다.

3. 관심 영역 스케일링 기술

145차 회의에서는 관심 영역 기반 스케일링 기술[12]도 제안되었다. 이 기술은 프레임별로 추출된 각 관심 영역의 크기를 줄여 부호화함으로써, 머신 비전 성능을 유지하면서 압축 효율을 높이는 방식이다. 제안된 방법은 100%, 75%, 50%의 세 가지 스케일 요소를 도입하여, 각기 다른 스케일 요소로 관심 영역을 부호화하는 기술이며, 기술 구조는 그림 3과 같다.

관심 영역 스케일링 기술의 부호화 과정은 관심 영역별로 최적의 스케일링 요소를 할당하기 위해 기존의 관심 영역 스태킹 기술을 기반으로 관심 영역 추적 과정, 관심 영역 스케일링 할당 과정, 관심 영역 다운스케일링 과정이 추가된 구조를 가진다. 먼저, 기존 관심 영역 스태킹 구조와 동일하게 객체 탐지 신경망을 사용하여 관심 영역 정보를 획득한 후, 프레임별로 관심 영역 정보를 저장한다. 기존 구조에서는 관심 영역을 추출하는 과정에서 관심 영역 박스 좌표만을 획득했지만, 인접 프레임에 대해 동일한 관심 영역에 동일한 스케일 요소를 할당하기 위해, 클래스 정보도 함

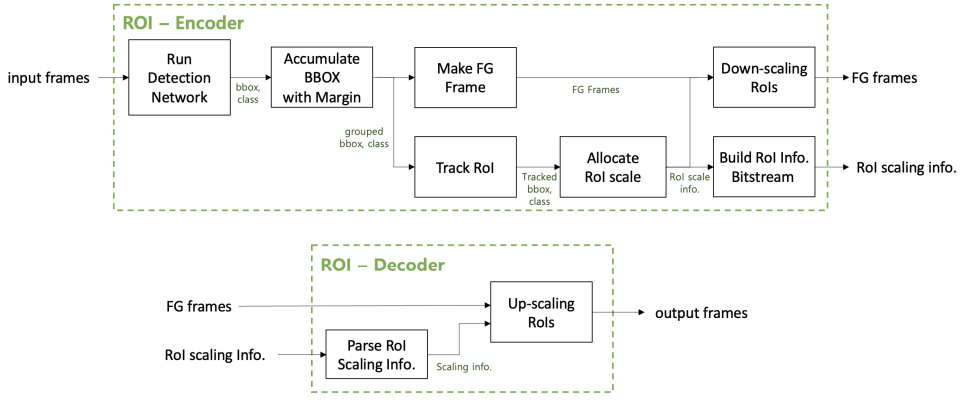

그림 3. 관심 영역 스케일링 기술 구조
Fig. 3. RoI Scaling Technology Architecture

께 저장한다. 또한, 기존 관심 영역 스태킹 기술에서는 누적 관심 영역 프레임 수를 64로 고정하였지만, 관심 영역 스케일링에서는 인코딩 모드에 따라 AI 모드에서는 1로, RA 모드에서는 Intra Period로, LD(Low Delay) 모드에서는 현재 프레임 인덱스 번호에서 프레임 레이트를 뺀 값으로 설정하며, 음수일 경우 0으로 설정하여 인코딩 시 딜레이 발생을 최소화하였다.

관심 영역 추적 단계에서는 동일한 객체에 대해 같은 스케일 요소를 할당하기 위해 인접 프레임 간의 IoU(Intersection over Union)와 클래스 정보를 이용하여 추적 ID를 할당한다. 그 후, 관심 영역 스케일 할당 단계에서는 각 프레임에서 추적된 관심 영역에 대해 최적의 스케일을 할당한다. 스케일 할당 방법은 다음과 같다. 먼저, 75%, 50% 두 가지 스케일 요소를 사용하여 원본 해상도를 다운샘플링한 후 VTM(VVC Test Model)으로 압축을 수행한다. 이후, 객체 탐지 신경망을 다시 수행하여 다운샘플링된 프레임에서 관심 영역이 탐지되는지 확인한 후, 최적의 스케일 요소를 선택한다. 이 과정에서, 먼저 50%로 압축한 뒤 객체가 탐지되지 않으면 75%로 스케일을 조정하고, 여전히 탐지되지 않으면 100% 스케일 요소를 할당한다. 관심 영역 다운스케일링 단계에서는 앞서 얻은 스케일 정보를 기반으로 다운샘플링을 수행한다. 마지막으로, 다운샘플링된 관심 영역을 복호화 과정에서 원본 크기로 복원하기 위해 관심 영역 스케일 요소, 관심 영역 좌표 등이 전송되며, 복호화 단계에서는 해당 정보를 바탕으로 다운샘플링된 관심 영역에 대해 업스케일링을 수행한다. 해당 기술은 VCM-RS v0.7 대비 -10.11% BD-rate 결과를 보였다. 그러나, 해당 기술은 공간적 리샘플링 기술과 함께 사용될 때 성능 저하가 우려되어 추가 연구를 수행하기로 결정하였다.

Ⅲ. 성능 비교 및 분석

본 논문에서는 2장에서 소개된 세 가지 관심 영역 기반 기술에 대한 성능 비교 실험을 수행하였다. 실험은 VCM-RS v0.8[13]을 활용하여 진행되었으며, 관심 영역 관련 기술 외에는 VCM 참조 소프트웨어의 기본 실험 설정을 그대로 유지하였다. 그러나 기존의 CTC[14] 테스트 시퀀스는 길이가 짧고 큰 장면 전환이 없는 특성을 가지므로, 이러한 테스트 환경이 실험 결과에 미치는 영향에 대한 우려가 145차 VCM 회의에서 제기되었다. 이에 따라 본 연구에서는 장

표 1. 실험 영상 정보
Table 1. Test Sequence Configuration

Test Sequence	Resolution	Frame Number	Combined Sequences
SFU_ClassB _1920x1080	1920x1080	356	ParkScene, Cactus, BasketballDrive, BQTerrace
SFU_ClassOB _1920x1080	1920x1080	130	Kimono, BasketballDrive
SFU_ClassC _832x480	832x480	388	BasketballDrill, BQMall, PartyScene, RaceHorses
SFU_ClassD _416x240	416x240	388	BasketballPass, BQSquare, BlowingBubbles, RaceHorses

표 2. 인코딩 모드별 QP 값 구성
Table 2. QP Configuration by Encoding Mode

Encoding Mode	QP	SFU_ClassB _1920x1080	SFU_ClassOB _1920x1080	SFU_ClassC _832x480	SFU_ClassD _416x240
RA, LD	QP0	38	38	27	22
	QP1	42	42	31	26
	QP2	46	46	35	30
	QP3	50	50	39	34
	QP4	54	54	43	38
	QP5	58	58	47	42
AI	QP0	22	22	22	22
	QP1	27	27	27	27
	QP2	32	32	32	32
	QP3	37	37	37	37
	QP4	42	42	42	42
	QP5	47	47	47	47

면 전환이 포함된 더 긴 시퀀스를 사용하여 실험을 수행하기 위해, 기존 SFU 데이터셋에서 해상도가 동일한 여러 영상을 결합하여 새로운 실험 시퀀스를 표 1과 같이 구성하였다. 실험에서 사용된 각 시퀀스와 인코딩 모드별 QP(Quantization Parameter) 값은 표 2에 정리되어 있으며, 모든 테스트 시퀀스에 대해 프레임 레이트를 30fps로 고정하고, RA 인코딩 모드의 경우 Intra Period를 32로 설정하였다.

각 실험 결과는 그림 4와 표 3, 4, 5에 요약되어 있으며,

표 3. AI 모드 실험 결과 (%: BD-rate)
Table 3. Experimental Results for AI mode (%: BD-rate)

Test Sequence	RoI Stacking			RoI Retargeting (Resolution: first)			RoI Scaling		
	All	High 4	Low 4	All	High 4	Low 4	All	High 4	Low 4
SFU_ClassB 1920x1080	14.0%	-13.2%	17.4%	-11.4%	-53.9%	-5.8%	23.1%	-5.6%	24.4%
SFU_ClassOB 1920x1080	-8.6%	-9.6%	-9.0%	56.4%	####	-3.6%	0.3%	22.0%	-2.2%
SFU_ClassC 832x480	13.4%	6.3%	16.0%	-5.9%	-4.9%	-6.9%	22.2%	39.0%	17.2%
SFU_ClassD _416x240	-2.2%	-5.2%	0.9%	0.0%	0.0%	0.0%	-3.5%	-5.4%	-3.3%
Average	4.1%	-5.4%	6.3%	9.8%	####	-4.1%	10.5%	12.5%	9.0%

표 4. RA 모드 실험 결과 (%: BD-rate)
Table 4. Experimental Results for RA mode (%: BD-rate)

Test Sequence	RoI Stacking			RoI Retargeting (Resolution: first)			RoI Scaling		
	All	High 4	Low 4	All	High 4	Low 4	All	High 4	Low 4
SFU_ClassB _1920x1080	4.0%	-3.1%	11.8%	-0.8%	-0.3%	-0.6%	20.5%	15.7%	23.2%
SFU_ClassOB _1920x1080	-25.9%	-28.6%	-24.7%	-16.7%	####	-29.9%	-18.5%	-21.5%	-17.0%
SFU_ClassC _832x480	-1.1%	-17.9%	5.1%	-3.0%	-5.8%	0.0%	5.0%	-1.8%	6.6%
SFU_ClassD _416x240	-6.6%	-0.4%	-12.9%	0.0%	0.0%	0.0%	-5.7%	-6.8%	-9.3%
Average	-7.4%	-12.5%	-5.1%	-5.1%	####	-7.6%	0.3%	-3.6%	0.9%

표 5. LD 모드 실험 결과 (%: BD-rate)
Table 5. Experimental Results for LD mode (%: BD-rate)

Test Sequence	RoI Stacking			RoI Retargeting (Resolution: first)			RoI Scaling		
	All	high 4	low 4	All	high 4	low 4	All	high 4	low 4
SFU_ClassB _1920x1080	4.0%	-3.1%	11.8%	-0.8%	-0.3%	-0.6%	20.5%	15.7%	23.2%
SFU_ClassOB _1920x1080	-25.9%	-28.6%	-24.7%	-16.7%	####	-29.9%	-18.5%	-21.5%	-17.0%
SFU_ClassC _832x480	-1.1%	-17.9%	5.1%	-3.0%	-5.8%	0.0%	5.0%	-1.8%	6.6%
SFU_ClassD _416x240	-6.6%	-0.4%	-12.9%	0.0%	0.0%	0.0%	-5.7%	-6.8%	-9.3%
Average	-7.4%	-12.5%	-5.1%	-5.1%	####	-7.6%	0.3%	-3.6%	0.9%

그림 4. 모든 테스트 시퀀스에 대한 mAP(세로 축) 대 bpp(가로 축) 곡선 결과
Fig. 4. mAP(axis y) vs bpp(axis x) Curve Results for All Test Sequences

기술 간 성능 비교를 위해 BD-rate가 산출되었다. 이때, BD-rate는 6개의 QP 값 전체를 대상으로 한 All, 낮은 QP 값 4개를 사용한 High 4, 그리고 높은 QP 값 4개를 사용한 Low 4로 구분하였다. VCM-RS v0.8의 기본 실험 구성으로 선택된 기술은 관심 영역 리타겟팅 기술이며, 이는 전체 프레임을 기준으로 해상도를 결정하는 방식이다. 본 연구에서는 이 기술을 앵커로 설정하여, 다른 두 기술과의 BD-rate 성능을 비교하였다. 새로운 시퀀스 구성 방식이 기존 데이터셋 기반 실험과 어떻게 다른 결과를 도출하는지를 분석함으로써, 관심 영역 기반 압축 기술의 성능을 보다 정확하게 평가할 수 있었다.

관심 영역 리타겟팅 기술의 성능은 2장 2절에서 언급한

것과 같이 해상도를 결정하는 방법에 따라 성능 차이가 크게 발생할 수 있음을 확인할 수 있었다. SFU_ClassB_1920x1080 시퀀스는 Kimono와 BasketballDrive 영상을 결합한 시퀀스로, 원본 해상도는 1920x1080이다. Kimono 영상은 단 한 명의 사람만을 포함하고 있는 반면, BasketballDrive 영상은 농구를 하는 여러 사람들과 의자, 공 등 10개 이상의 객체를 포함하고 있다. 관심 영역 리타겟팅 기술에서 인코딩할 해상도를 전체 시퀀스를 기준으로 설정한 경우, 인코딩 해상도는 1600x1024로 설정되지만, 첫 번째 프레임을 기준으로 설정할 경우 인코딩 해상도는 256x512로 설정된다. 이로 인해 AI, RA 모드에서 두 기술 간의 kbps 값이 겹쳐지는 구간이 없어 BD-rate 계산이 불가능해진다. 또한 그림 4의 결과를 보면 첫 번째 프레임을 기준으로 해상도를 결정하는 방식의 결과(파란색 선)가 다른 결과와 다르게 매우 낮은 머신 성능과 낮은 비트율을 가지는 것을 확인할 수 있다.

또한, 관심 영역 리타겟팅 기술에서는 LD 모드에서 64프레임씩 관심 영역을 누적하는 방식이 사용된다. 제 142차 회의에서 채택된 시간적 리샘플링 기술에 따르면, 4개의 프레임 중 1개의 프레임만 인코딩되며, 이로 인해 전체 시퀀스의 길이가 1/4로 줄어든다. 본 논문에서 실험한 영상 중 가장 긴 시퀀스는 388개의 프레임을 가지지만, 시간적 리샘플링에 의해 첫 번째 프레임에 절반 이상의 관심 영역 정보가 누적되기 때문에, 실험 결과에서 앵커와 동일한 결과가 도출되어 모든 BD-rate 값이 0%로 나타났으며, 그림 4에서 그래프가 동일한 것을 확인할 수 있다.

표 3의 AI 실험 결과에서는 일부 시퀀스가 앵커보다 높은 성능을 보였으나, 전체적으로는 전체 시퀀스를 기반으로 해상도를 결정하는 관심 영역 리타겟팅 기술이 더 우수한 성능을 보인다는 것을 확인할 수 있었다. 표 4의 RA 실험 결과에서는 관심 영역 스태킹 기술이 더 높은 성능을 보였으며, 특히 높은 비트율에서 BD-rate가 -12.5%로 앵커 결과보다 더 우수한 성능을 나타냈다. AI 실험에서도 높은 비트율에서 관심 영역 스태킹 기술이 BD-rate -5.4%로 더 높은 성능을 보였다.

전체 시퀀스를 스캔하여 해상도를 결정하는 경우, 시퀀스 내 거의 모든 프레임의 크기가 확장된다. 이로 인해 확장된 영역을 인코딩하면서 kbps 값이 커지지만, 낮은 QP에서는 영상의 열화가 크지 않아 머신 비전 성능이 크게 증가하지 않기 때문에, 관심 영역 스태킹 기술이 더 높은 성능을 나타낸다. 반대로 높은 QP로 압축하는 경우에는 영상의 열화가 크지만, 확장된 영역 덕분에 객체 탐지가 더 잘 이루어져 낮은 비트율에서는 앵커 결과가 더 높은 성능을 보여준다.

표 5의 LD 결과에서는 관심 영역 스케일링 기술이 다른 기술보다 더 높은 성능을 보이는 것으로 확인되었다. 다른 기술들은 LD 모드에서 64프레임씩 관심 영역을 누적하는 반면, 관심 영역 스케일링 기술은 인코딩 딜레이를 최소화하기 위해 이전 프레임의 관심 영역 정보만 일정 프레임 수만큼 누적한다. 이로 인해 인코딩 영역이 줄어들고, 관심 영역별로 최적의 스케일링 요소를 찾아 관심 영역의 크기를 줄여 부호화할 수 있어, 다른 기술들보다 더 높은 성능을 발휘하게 되는 것으로 분석된다.

Ⅳ. 결 론

본 논문에서는 VCM 그룹이 비디오 그룹으로 이전한 이후, 본격적으로 참조 소프트웨어 개발 및 WD(Working Draft) 작업이 진행되면서 가장 많이 기여된 기술 카테고리 중 하나인 관심 영역 기반 기술에 대해 살펴보았다. 또한, VCM 내에서 긴 시퀀스와 장면 전환이 있는 영상의 필요성이 제기됨에 따라, 기존 SFU 데이터셋에서 해상도가 동일한 영상을 결합하여 시퀀스를 통합한 후 실험을 진행하였고, 이를 통해 다양한 기술의 성능을 비교하였다. 실험 결과 RA 모드에서는 관심 영역 스태킹 기술이 더 우수한 성능을 나타냈고, LD 모드에서는 관심 영역 스케일링 기술이 더 높은 압축 효율을 보였다. 전체 프레임을 스캔하여 가장 큰 영역을 기준으로 해상도를 결정하는 관심 영역 리타겟팅 기술은 긴 시퀀스나 스트리밍과 같은 환경에서는 적용하기 어렵다는 한계가 있어, 이에 대한 개선이 필요할 것으로 보인다. 본 연구에서 제안된 기술들과 앞으로 제안될 다양한 기술들을 통해 이러한 한계점들을 개선함으로써, 관심 영역 기반 부호화 기술이 더욱 발전하고, 향후 VCM 표준화에 중요한 기여를 할 수 있기를 기대한다.

참 고 문 헌 (References)

[1] T. Wiegand, G. J. Sullivan, G. Bjontegaard and A. Luthra, “Overview of the H.264/AVC video coding standard,” IEEE Transactions on circuits and systems for video technology, vol. 13, no. 7, pp. 560-576, Jul. 2003.
doi: https://doi.org/10.1109/TCSVT.2003.815165

[2] G. J. Sullivan, J. Ohm, W. Han and T. Wiegand, “Overview of the high efficiency video coding (HEVC) standard,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1649-1668, Dec. 2012.
doi: https://doi.org/10.1109/TCSVT.2012.2221191

[3] 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

[4] Cisco Annual Internet Report (2018-2023) White Paper, https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html, (accessed September 1st, 2024).

[5] Computer Vision Market Size, Share & Growth Report, 2030, https://www.grandviewresearch.com/industry-analysis/computer-vision-market, (accessed September 1st, 2024).

[6] M. H. Jeong, et al., “[VCM] Report on CE1.4,” ISO/IEC JTC1/SC29/WG4 input document m64421, July 2023.

[7] 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.
doi: https://doi.org/10.1016/j.dib.2020.106701

[8] Detectron2, https://github.com/facebookresearch/detectron2, (accessed September 1st, 2024).

[9] 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: https://doi.org/10.1109/ICIP46576.2022.9897525

[10] J. Redmon and A. Farhadi, “Yolov3: An incremental improvement,” arXiv preprint arXiv:1804.02767, 2018.
doi: https://doi.org/10.48550/arXiv.1804.02767

[11] S. Różek, et al., “[VCM] Improved RoI preprocessing and retargeting for VCM,” ISO/IEC JTC1/SC29/WG4 input document m66523, January 2024.

[12] Y. Lee, et al., “[VCM] RoI Scaling for VCM,” ISO/IEC JTC 1/SC 29/WG 4 input document m66297, January 2024.

[13] VCM-RS v0.8, https://git.mpeg.expert/MPEG/Video/VCM/VCM-RS, (accessed September 1st, 2024).

[14] WG 04, “Common test conditions for video coding for machines,” ISO/IEC JTC1/SC29/WG4 output document N00467, January 2024.

저 자 소 개

이 예 지

- 2018년 2월 : 극동대학교 스마트모바일학과 졸업(학사)
- 2020년 2월 : 건국대학교 스마트ICT융합과 졸업(석사)
- 2020년 3월 ~ 현재 : 건국대학교 컴퓨터공학과 박사과정
- 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 컴퓨터공학과/스마트ICT융합공학과 교수
- ORCID : https://orcid.org/0000-0002-1153-4038
- 주관심분야 : 스마트미디어시스템, 멀티미디어검색, 영상처리, 컴퓨터비전, 멀티미디어/메타데이터 처리