

일반논문 (Regular Paper)

방송공학회논문지 제29권 제4호, 2024년 7월 (JBE Vol.29, No.4, July 2024)

<https://doi.org/10.5909/JBE.2024.29.4.443>

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

## VCM을 위한 관심 영역 스케일링 압축 기법

이 예 지<sup>a)</sup>, 윤 경 로<sup>b)†</sup>

### RoI Scaling-based Coding for VCM

Yegi Lee<sup>a)</sup> and Kyoungro Yoon<sup>b)†</sup>

#### 요 약

MPEG(Moving Picture Experts Group)은 비디오와 오디오를 포함한 멀티미디어 기술의 국제 표준화를 선도하는 기구이다. 최근 다양한 산업분야에서 인공지능 기술의 적용이 확산되면서, 지능형 영상분석에 대한 수요가 증가하고 있다. 이에 따라 MPEG은 2019년 VCM(Video Coding for Machines)이라는 AhG(Ad hoc Group)을 설립했으며<sup>[1]</sup>, 머신비전을 위한 비디오 압축 기술에 대해 지속적으로 논의하고 있다. 감시나 자율주행 자동차와 같은 환경에서는 사람이나 자동차와 같은 객체를 탐지 및 추적하는 것이 주요 목표이다. 따라서 VCM에서는 프레임 내 관심 영역(RoI, Region of Interest) 정보를 기반으로 영상 압축을 하는 기술들이 제안되고 있다. 본 논문에서는 이러한 관심 영역 정보를 바탕으로 머신비전 성능에 영향을 주지 않는 스케일 요소(scale factor)를 찾아 해당 관심 영역을 다운스케일(down-scaling)을 수행하는 관심 영역 스케일링 방법에 대해 소개한다. 실험 결과, 제안하는 방법이 VCMRS(VCM Reference Software) v0.7과 v0.8과 비교했을 때 각각 BD-rate가 -6.87%, -4.02%로 압축 효율이 우수한 것을 확인하였다.

#### Abstract

MPEG(Moving Picture Experts Group) operates under ISO/IEC JTC1 SC29 and leads in the international standardization of multimedia technologies, including video and audio coding. In response to the increasing demand for intelligent video analysis due to the wide application of artificial intelligence technologies across various industries, MPEG established the VCM(Video Coding for Machine) AhG(Ad hoc Group) in 2019<sup>[1]</sup>, and discussions on video coding technology for machine visions have been ongoing. In applications like surveillance and autonomous driving, where detecting and tracking objects like people or vehicles is a primary goal, several contributions have been proposed based on RoI(Region of Interest) information within frames. This paper, proposes a RoI scaling method that performs downscaling on the specified RoI based on the extracted interest area information from each frame without affecting machine performance by finding a scale factor. Experimental results show that the proposed method achieves better compression efficiency with a BD-rate of -6.87% compared to VCMRS(VCM Reference Software) v0.7, and a BD-rate of -4.02% compared to VCMRS v0.8.

Keyword : MPEG, Video Coding for Machine, Machine Vision, Video Compression, RoI

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

압축이 되지 않은 원본 비디오 데이터는 막대한 저장 공간을 차지하며, 이는 효율적인 전송을 방해하는 주요 걸림돌이 될 수 있다. 따라서 코덱은 이러한 데이터를 더 적은 용량으로 압축하여 관리와 전송을 쉽게 하는 중요한 역할을 한다. 이 과정에서 가능한 한 데이터 압축을 최대화하면서도 비디오의 품질 손실을 최소화하면서 소비자에게 최상의 시각적 경험을 제공하기 위한 목적을 가지고 있다. 디지털 멀티미디어 기술의 국제적 표준화를 선도하는 기구인 MPEG은 ISO/IEC JTC1 SC29 산하에서 활동하며, 비디오 및 오디오를 포함한 멀티미디어 기술 발전에 크게 기여하고 있다. MPEG에서는 고품질의 비디오 데이터를 가능한 적은 데이터로 압축하여 효율적으로 저장, 전송, 재생하기 위해 MPEG-4 파트 10 AVC(H.264, Advanced Video Coding)<sup>[2]</sup>, MPEG-H 파트 2 HEVC(H.265, High Efficiency Video Coding)<sup>[3]</sup>, MPEG-I 파트 3 VVC(H.266, Versatile Video Coding)<sup>[4]</sup>와 같은 다양한 영상 압축 표준을 개발해왔다. AVC는 HD 비디오 스트리밍의 보급을 주도했으며, HEVC는 그 성능을 더욱 개선하여 4K 울트라 HD 콘텐츠의 효율적인 전송을 가능하게 하였다. VVC는 다가오는 8K 비디오와 같은 더 높은 해상도와 가상현실 콘텐츠를 포함하여 다양한 멀티미디어 응용 분야에 걸쳐 한층 더 개선된 압축 효율을 제공한다.

최근 인공지능 기술의 급속한 발달로 스마트 시티, 자율주행 자동차, 감시, 모니터링 등 다양한 분야에서 딥러닝 기반 기술이 다양하게 활용되고 있다. 특히 지능형 영상 분석에 대한 수요가 급증하고 있으며, 영상 콘텐츠의 주된 소

비 주체가 사람에서 기계로 전환될 것이라는 전망이 제기되고 있다. 기존의 영상 압축 표준은 인간의 인지 특성에 맞추어 개발되었기 때문에, 기계에 의한 영상 분석이 주를 이루는 환경에서는 이러한 방식이 적합하지 않을 수 있다. 따라서 머신 비전을 위한 영상 압축 기술을 개발하는 것은 새롭게 연구가 필요한 분야이다. 이러한 시장의 요구와 기술적 변화를 반영하여, MPEG에서는 지능형 영상 분석과 관련된 요구사항을 충족시키기 위해 2019년 VCM AhG을 설립하였다. 약 2년동안 제안된 다양한 후보 기술은 크게 비디오 압축 방식과 특징맵 압축 방식으로 나누어졌다. 136차 회의에서는 이를 두 개의 트랙으로 분리하여 머신비전을 위한 비디오 압축 방법과 관련된 기술은 VCM으로 특징맵 압축 방법과 관련된 기술은 FCM(Feature Coding for Machine)으로 나누어 회의를 진행하기로 결정하였으며, 머신비전이 비디오 데이터를 보다 효율적으로 처리하고 이해할 수 있도록 지원하는 새로운 영상 및 특징맵 압축 기술에 대해 논의가 지속해서 이루어지고 있다.

감시 시스템이나 자율주행 자동차와 같은 환경에서는 객체의 탐지 및 추적, 예를 들어 사람이나 차량과 같은 대상을 식별하는 것이 중요한 목적 중 하나이다. 따라서 VCM에서는 관심 영역 정보를 기반으로 영상을 압축하는 방식에 대하여 제안되고 있다. 본 논문에서는 머신을 위한 영상 압축 기술로서 관심 영역 기반 스케일링 압축 기법을 소개한다. 제안하는 방법은 입력된 영상의 각 프레임을 원본 해상도 대비 75%, 또는 50%로 줄여 압축하는 방식을 적용하여, 이 과정에서 다운스케일된 프레임 내에서 관심 영역이 여전히 효과적으로 탐지될 수 있는지를 확인한다. 이를 통해 머신의 성능에 영향을 주지 않으면서도 최적의 스케일 요소를 도출 해내고, 해당 스케일 요소를 적용하여 관심 영역을 다운스케일링하고 부호화함으로써 압축 효율을 향상하는 것이 본 기법의 핵심이다.

본 논문은 다음과 같이 구성되어 있다. 2장은 CFP 대응에 대하여 논의했던 140차 회의부터 145차 회의까지의 VCM 표준화 동향에 대해 채택된 기술 및 주요 논의 사항에 대하여 서술한다. 3장은 실험 환경 및 본 논문에서 제안하는 머신을 위한 관심 영역 기반 스케일링 압축 방법에 대해 논한다. 4장은 실험 결과와 그 분석을 제시하고, 5장에서는 본 논문의 결론을 도출한다.

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>

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

· Manuscript April 5, 2024; Revised June 12, 2024; Accepted June 12, 2024.

## II. VCM 표준화 동향

140차 MPEG 회의에서 CFP(Call for Proposal)에 대한 응답 기술에 대해 논의한 이후 WG2(Working Group 2)인 Requirement 그룹에서 WG4 Video 그룹으로 이전<sup>[6]</sup>되었다. CFP 대응 과정에서 총 17개의 기술 제안<sup>[7]</sup>이 이루어졌으며 이 중 2건은 end-to-end 신경망 기반 코덱을 사용하는 기술이었고, 13건은 기존에 표준화된 비디오 코덱을 확장하는 방식이며, 복호화의 수정을 필요로 하는 제안이었다. 또한, 기존 표준화된 비디오 코덱을 사용하면서 복호화의 변경 없이 성능을 개선하는 기술이 1개 제안되었고, 레이어드 디자인을 기반으로 한 압축 기술도 1개 포함되었다. 140차 회의에서는 이러한 기술 제안을 포괄하고 확장할 수 있는 참조 소프트웨어 구조에 대하여 논의를 하였으며, 이를 기반으로 VCMRS v0.3가 배포 되었다<sup>[8]</sup>.

VCMRS v0.3 구조는 그림 1과 같다. 크게 전처리 구성요소, 내부 코덱, 후처리 구성요소로 구성된다. 내부 코덱은 VVC를 기반으로 개발이 되어있으며, 전/후처리 요소에는 시간적 리샘플링, 공간적 리샘플링, 관심 영역 기반 프로세싱 및 복호화단 후처리 필터링이 있다. 140차 회의 이후 내부 코덱 기술을 포함하여 전/후 처리 요소들에 대하여 아래와 같이 크게 5가지의 기술로 나누어져 각 기술에 대해 CE(Core Experiment)를 진행하였다.

- CE1: ROI based coding methods
- CE2: Neural network based intra frame coding
- CE3: Spatial resampling
- CE4: Temporal resampling
- CE5: Post filtering

VCMRS v0.3 배포 후, MPEG 회의를 통해 CE와 기술 기고에 기반한 참조 소프트웨어 업데이트가 지속되었다. 현재까지 채택된 기술들은 다음과 같다. 142차 회의에서 CE4 temporal resampling에 대하여 2개의 기술 중 China Telecom, CAS-ICT가 제안한 기고가 채택<sup>[9]</sup>되었으며, VCMRS v0.5에 해당 기술이 통합되었다. 해당 기고는 실시간으로 중간 프레임을 보간할 수 있는 RIFE(Real-time Interpolation Flow Estimation)<sup>[10]</sup> 기반 방법을 제안하였으며, sampling ratio을 4로 설정 하였을 때 VCMRS v0.4 앵커 대비 -44.81%의 Pareto-mAP/MOTA 성능을 보여주었다.

143차 회의에서는 CE1 ROI based coding에 대하여 총 6개의 기술 중 명지대학교와 ETRI에서 제안한 기술이 채택<sup>[11]</sup>되었다. 해당 기고는 부호화 단계 객체 탐지 신경망을 이용하여 관심 영역을 찾은 후 객체 주변 영역을 확장하여 객체만 존재하는 전경(FG, Foreground) 프레임을 만들고, 압축을 수행함으로써 압축 효율을 높였다. 해당 기술은 CE1 평가방법에 기반하였을 때 VCMRS v0.5 대비

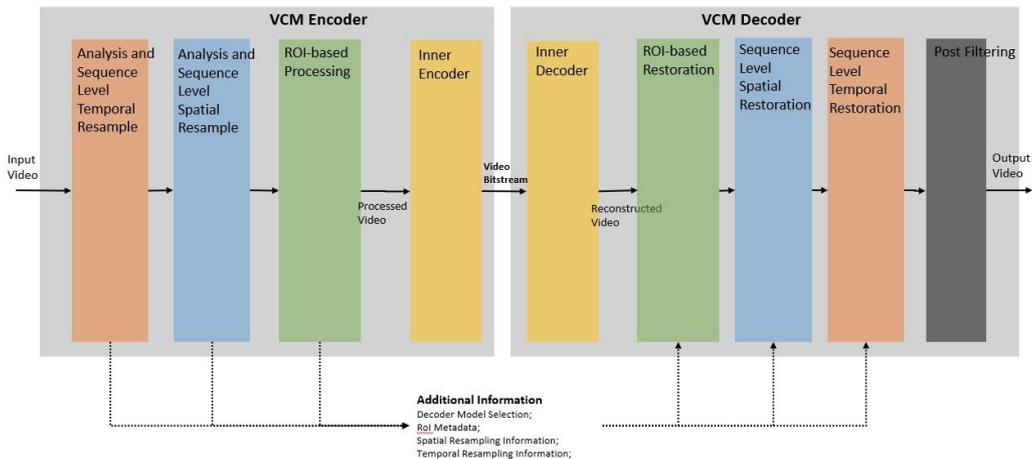


그림 1. VCMRS v0.3 구조<sup>[8]</sup>  
Fig. 1. VCMRS v0.3 architecture<sup>[8]</sup>

- 36.07% BD-rate 성능을 보여주었다.

144차 회의에서는 CE로 진행된 기술이 아닌 Tencent의 bit truncation<sup>[12]</sup> 기술이 채택되었다. 해당 기고는 머신 성능이 bit depth에 덜 민감하다는 특성을 이용하였다. 부호화 측에서 루마 샘플 값을 1비트씩 오른쪽으로 이동시켜 압축 시킨 후 복호화 측에서 조건부에 따라 루마 샘플 값을 왼쪽으로 이동시켜 복원하는 기술이며, VCMRS v0.6 대비 -18.55% Pareto-mAP/MOTA 성능을 보여주었다.

145차 회의에서는 CE3 spatial resampling과 관련하여 총 3개의 기술 중 연세대학교와 ETRI가 제안한 MI-RPR (Machine Inference- based Reference Picture Resampling) 기술<sup>[13]</sup>이 채택이 되었다. MI-RPR은 YOLOv7<sup>[15]</sup>을 사용하여 관심 영역을 추출한 후 해당 정보를 이용하여 프레임 별 spatial resampling 분석결과를 구한 후 VVC 내부에서 현재 영상과 다른 해상도를 가지는 영상을 참조할 수 있는 RPR 기술을 적용하였다. VCMRS v0.7 대비 -3.45% BD-rate 성능을 보여주었다. 또한 관심영역 기반 기술과 관련하여 포즈난대학교에서 제안한 RoI retargeting 기술<sup>[14]</sup>이 채택되었다. 포즈난대학교의 RoI retargeting의 경우 기 채택된 RoI 기술에 관심 영역을 기반으로 프레임 별 관심 영역을 모두 포괄하는 outline을 구성한 후, 해당 정보를 다시 모아 가장 큰 outline 영역을 계산하고, 그 외 영역을 잘라 해상도를 줄인다. 그 후, 관심 영역 그룹에 대해 각 프레임 별 줄어든 해상도에 대하여 영역을 확장하는 기술이다. 또한 관심 영역 사이의 거리가 좁으면 배경(background, BG) 영역의 픽셀 값을 127로 설정하는 것이 아니라 블러링을 수행함으로써 인접한 영역에서 픽셀 값의 변화가 크게 변동되는 것을 막았다. 포즈난대학교 제안의 경우 VCMRS v0.7 대비 -20.50% BD-rate 결과를 얻었다. 또한 143차에 채택된 temporal resampling 기술과 관련하여 SSIM 계산 및 후속 조건 선택을 생략하여 복호화 과정에서 공간적 보간 절차를 단순화한 ITRI의 제안<sup>[16]</sup>도 채택이 되었다. 해당 기고는 약 -1.45% BD-rate 정도로 다른 기술들에 비해 성능 개선이 작지만, 복호화단 연산을 단순화할 수 있는 제안으로 채택이 되었다.

추가적으로 142차 MPEG 회의까지, VCMRS에서는 VVC 기반과 Hybrid 기반의 두 가지 내부 코덱이 앵커로 사용되었지만, Hybrid 기반 기술이 아직 CE2에서 세부 기

술 채택을 기다리고 있었기 때문에, 143차 회의 이후로는 VVC 기반 내부 코덱만이 채택되어 사용되고 있다. 또한 이전에는 VVC 기반 내부 코덱은 VTM(VVC Test Model) 12.0 버전을 사용하였으나 145차 회의 이후로는 버전이 20.0으로 업데이트 되었다. 평가 방법론 측면에서는, BD-rate가 주된 평가 요소로 사용되었으나, 머신 성능이 모노토닉하지 않았을 때 에러 발생의 문제가 있어, 144차까지 CE에 따라 Pareto-mAP/MOTA를 사용하거나 CE만의 평가 방법을 만드는 등 기준을 다르게 하였다. 하지만 145차 이후 Curve Fitting 기고를 채택<sup>[17]</sup>되어 머신 성능을 모노토닉하게 만들어 BD-rate가 주요 평가요소로 사용되고 있다. 또한 VCMRS가 새롭게 배포 될 때마다 테스트 시퀀스 별, 인코딩 모드 별 QP값이 다르게 세팅되어 있었으며, QP값을 세팅하는 기준에 대하여 합의가 이루어지지 않았다. 따라서 VCMRS가 업데이트가 계속되고 이전 버전과의 결과 비교에 어려움이 있었지만 144차 회의에서 CTC 내 앵커 QP값에 대한 기고가 채택되어<sup>[18]</sup> VCMRS v0.7부터 시퀀스 별 고정된 QP값을 사용하게 되었다.

### III. 관심 영역 기반 스케일링 압축 방법

본 논문에서 제안하는 관심 영역 기반 스케일링 압축 방법은 머신 성능에 영향을 주지 않는 스케일 요소를 찾아 해당 관심 영역에 다운스케일링을 적용해 부호화하고, 복호화 과정에서 다운스케일링한 관심 영역을 다시 원본 해상도로 복원하는 방법이다. 연구는 VCMRS v0.7을 기준으로 시작되었으며, 제안하는 압축 방식의 설명도 이 버전을 바탕으로 진행된다. 하지만 145차 회의에서 채택된 RoI retargeting 기술이 포함된 VCMRS v0.8을 기반으로 한 실험도 수행되었으며 이와 관련된 변경 사항 역시 본 논문에서 다룬다.

VCMRS v0.7에서 사용되는 관심 영역 기반 압축 기술은 각 입력 시퀀스에 대하여 프레임별 객체 탐지 네트워크를 수행하여 박스 정보를 추출하고, 해당 박스 결과를 각 프레임으로 저장한다. 이후, roi\_accumulate\_period에 따라 프레임 그룹화하여 박스 영역을 누적함으로써 최종 전경 프레임이 생성되며, 내부코덱을 통해 압축이 수행된다. 본 논

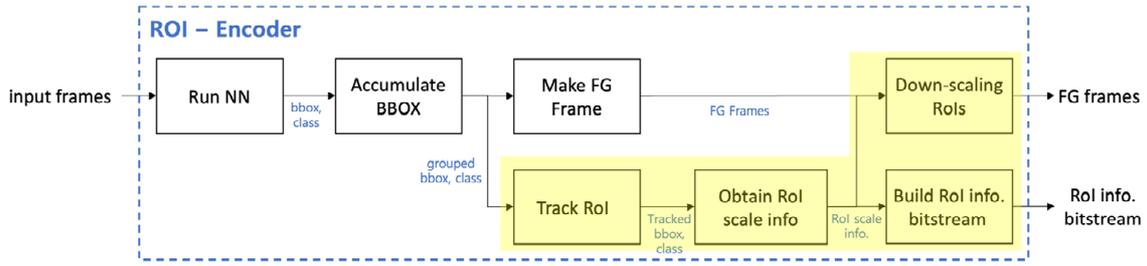


그림 2. VCMRS v0.7 기반 RoI 스케일링 방식의 부호화 과정  
 Fig. 2. Encoding process of the RoI scaling method based on VCMRS v0.7

문에서 제안하는 RoI 스케일링 방법은 기존 VCMRS 기술에 네 가지 추가 단계가 포함되어 있다. 그림 2는 VCMRS v0.7에 내장된 RoI 기반 압축 기술에 본 논문의 제안 방법을 통합한 부호화 과정을 보여준다. 이 그림에서 노란색으로 표시된 부분은 본 논문에서 새롭게 추가된 단계를 나타낸다.

먼저, 기존 RoI 기술 구조와 동일하게 객체 탐지 신경망을 이용하여 관심 영역 정보를 획득한 후 프레임 별 관심 영역 정보를 저장한다. 단, 기존 RoI 기술에 경우 박스(x1, y1, x2, y2) 정보만 추출하지만, RA 및 LD 모드에서 인접

프레임에 대해 동일한 RoI에 동일한 스케일을 할당하기 위해서 본 논문에서는 RoI 추적을 위한 클래스 ID를 박스 정보와 함께 저장한다. 또한 기존 RoI 기술에서는 `roi_accumulation_period`가 코딩 모드에 관계없이 64로 동일하게 설정되었지만, 제안하는 방법에서는 인코딩 모드에 따라 아래와 같이 변경하였으며, 인코딩 모드 별 관심 영역 누적 및 추적 방법은 그림 3과 같다.

AI(All Intra) 모드에서는 `roi_accumulation_period`는 1로 설정되며, RoI 정보는 프레임마다 별도로 추출되므로 RoI 정보가 누적되지 않는다. LD(Low Delay) 모드에서는

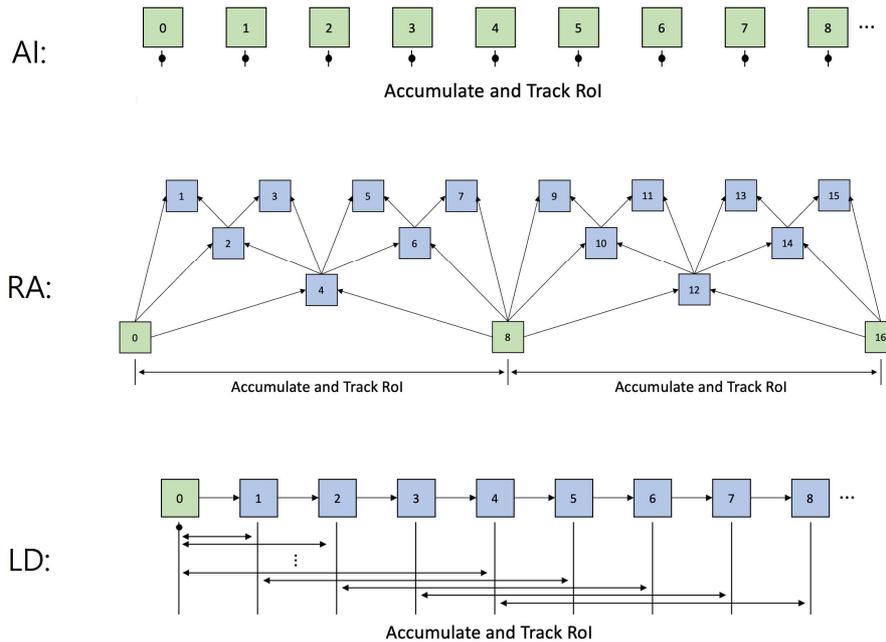


그림 3. 인코딩 모드 별 관심 영역 누적 및 추적 방법  
 Fig. 3. RoI accumulation and tracking method by encoding mode

roi\_accumulation\_period는 현재 프레임 인덱스 번호에 프레임 레이트(frame rate)를 뺀 값으로 설정되며, 만약 roi\_accumulation\_period가 음수인 경우 0으로 설정된다. RA에서는 roi\_accumulation\_period는 IntraPeriod로 설정되며, IntraPeriod 단위로 RoI 정보가 누적된다.

그 후 기존 RoI 기술과 동일하게 전경이미지를 생성한다. 동시에 Track RoI에서는 roi\_accumulation\_period를 기반으로 관심 영역 정보를 검색하고 관심 영역을 추적한다. 추적을 위한 신경망을 사용하지 않으며, 본 논문에서는 IoU(Intersaction of Union)와 이전 프레임과의 클래스 ID 정보 간의 관계를 이용하여 추적 ID를 할당한다. Obtain RoI scale info.에서는 각 프레임 별 탐지된 관심 영역에 대하여 최적의 스케일을 할당하기 위한 작업을 수행한다. 본 논문에서는 75%, 50% 두 가지의 스케일 요소에 대하여 원본 해상도를 스케일 요소를 기반으로 다운샘플링하여 VTM으로 압축을 수행한다. 다음으로 객체 탐지 신경망을 다시 수행하여 관심 영역이 다운샘플링 된 프레임에서도 탐지가 되는지 확인하여 최적의 스케일 요소를 찾는다. 최적의 스케일 요소를 찾는 과정은 먼저 50%로 압축하고, 50%에서 물체가 감지되지 않으면 관심영역의 크기를 75% 줄여 물체가 감지가 되는지 확인하고, 그래도 물체가 감지되지 않으면 RoI에 100% 스케일 요소를 할당한다. Down-scaling RoIs 단계에서는 이전 단계에서 얻은 스케일 정보를 기반으로 RoI를 겹치지 않는 RoI로 다운샘플링한다. 서로 다른 스케일 요소를 가지고 있는 관심 영역이 만약 겹쳐져 있는 경우 복호화 과정에서 복원이 잘못 될 수도 있기 때문에 해당 스케일 요소로 다운샘플링을 수행한 후 부호화 측에서 다시 원본 해상도로 업샘플링(up-sampling)을 수행한다. 마지막으로 부호화 과정에서 다운샘플링 된 관심 영역을 복호화에서 원본 해상도로 복원하기 위해 관

심 영역 스케일 요소, RoI 좌표(x1, y1, x2, y2) 등이 Build RoI info. bitstream 단계에서 전송된다.

VCMRS v0.7 내 RoI 기반 압축 기술에서는 복호화 과정을 바이패스한다. 그러나 본 논문에서 제안하는 방식은 다운샘플링된 관심 영역을 복원해야 하므로, RoI 정보를 불러와 업샘플링하는 과정이 필요하며, 이 과정은 그림 4와 같다. Parse RoI info. 단계에서는 부호화 단계에서 전송된 관심 영역 정보를 불러오며, Up-scaling RoI 단계에서는 이 정보를 바탕으로 해당 관심 영역을 업샘플링을 수행한다.

145차 회의에서 포즈난대학교에 RoI retargeting 기술이 채택되어 VCMRS v0.8에 통합되었다. 포즈난대학교 제안의 경우 관심 영역 정보를 활용하여 최종 해상도를 결정하고, 각 프레임 내에서 관심 영역 그룹의 크기를 최종 해상도에 맞춰 확대하여 사용함으로써 높은 QP에서도 향상된 압축 효율을 달성하였다. 그러나 본 논문에서 제안하는 관심 영역 기반 스케일링 방법은 관심 영역을 다운스케일링하여 머신 성능을 유지하면서 비트레이트를 낮추는 접근 방식이기 때문에 포즈난대학교에서 제안하는 모든 기술을 적용할 경우 오히려 성능이 떨어지는 것을 확인하였다. 따라서 본 논문에서는 관심 영역 간 거리가 좁을 때 불러링된 프레임을 생성하고, 관심 영역 정보를 이용해 최종 해상도를 결정하는 방법만을 선택적으로 적용하여 실험을 진행하였다. 추가적으로 MI-RPR 기술에서 채택된 공간적 샘플링을 위한 분석 결과를 스케일 요소 계산에 활용하여 더욱 정밀한 스케일 결정을 가능하게 하였다.

#### IV. 실험 결과 및 분석

실험은 VCM 프로젝트에서 사용되는 SFU 데이터셋을 기

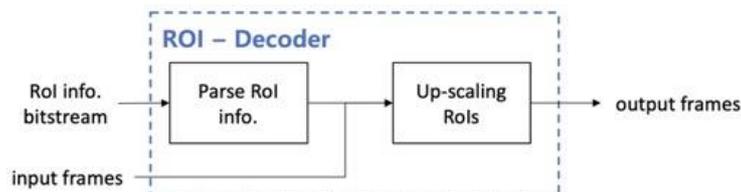


그림 4. RoI 스케일링 방식의 복호화 과정  
Fig. 4. Decoding process of the RoI scaling method

반으로 진행되었으며, 그 외 실험 환경 및 조건은 144차 및 145차 CTC(Common Test Condition)<sup>[19-20]</sup>를 따랐다. 표 1은 VCMRS v0.7 앵커와 VCMRS v0.7 기반의 관심 영역 기반 스케일링 방법을 비교한 BD-rate 결과를 보여준다.

표 1. VCMRS v0.7 기반 관심 영역 스케일링 결과  
Table 1. RoI scaling result based on VCMRS v0.7

Mode	roi_scaling	roi_scaling_poc	roi_scaling_poc_AP
RA	3.92%	0.36%	-0.31%
LD	1.30%	2.26%	0.12%
AI	-5.83%	-5.83%	-16.44%
Avg.	-0.20%	-1.07%	-5.54%

표 1에서의 roi\_scaling\_poc\_AP 결과는 3장에서 설명된 인코딩 모드 별 roi\_accumulation\_period를 조정하고, 최적의 스케일 요소를 찾는 과정에서 LD와 RA 모드를 인코딩할 때 PoC(Picture Order Count) 별 QP(Quantization Parameter) offset을 고려한 결과이다. 그 외의 roi\_scaling 결과는 위의 두 조건을 모두 고려하지 않고 수행한 실험 결과이며, roi\_scaling\_poc 결과는 PoC 별 QP 오프셋만을 적용한 결과를 나타낸다. 실험 결과, 두 조건을 모두 고려했을 때 평균 BD-rate가 -5.54%로 가장 우수한 성능을 보였다.

표 2. 제안하는 방법 대한 LD 추가실험 결과  
Table 2. LD additional experiment results for the proposed method

Mode	roi_scaling_poc_AP	roi_scaling_v2_poc_IP_Ld_noscaling
RA	-0.31%	-0.31%
LD	0.12%	-3.87%
AI	-16.44%	-16.44%
Avg.	-5.54%	-6.87%

표 2의 실험 결과, LD 모드의 경우 BD-rate가 0.12%로 VCMRS v0.7과 비교했을 때 성능이 다소 저하되었다는 점을 확인할 수 있다. 이는 LD 모드의 특성상, 앞선 프레임을 참조하는 과정에서 이전 프레임 내 관심 영역이 스케일링되어 크기가 줄어든 경우, 해당 프레임을 효과적으로 참조하지 못하여 비트레이트가 증가하는 것으로 추정된다. 따라서 본 연구에서는 LD 모드 인코딩 시에만 다른 접근

방식을 도입하였다. 구체적으로, 부호화 과정에서 관심 영역을 해당 스케일 요소로 다운스케일링한 후, 이를 다시 원본 해상도로 업스케일링하는 방법을 적용하였다. 이렇게 함으로써, 디코딩 과정에서 추가적인 업스케일링을 수행할 필요가 없으며, 디코딩을 위한 부가정보는 보내지 않는다. 그 결과, 표 2에 나타난 바와 같이, LD에서 압축 효율이 BD-rate 0.12%에서 -3.87%로 개선되었다는 것을 확인할 수 있었다.

표 3. VCMRS v0.8 기반 관심 영역 스케일링 결과  
Table 3. RoI scaling result based on VCMRS v0.8

Mode	roi_scaling_v0.8 (Faster R-CNN)	roi_scaling_v0.8 (Yolov7)
RA	-5.96%	-4.05%
LD	5.13%	0.80%
AI	-5.41%	-8.82%
Avg.	-2.08%	-4.02%

표 3은 VCMRS v0.8을 기반으로 한 실험 결과를 보여준다. VCMRS v0.8에 통합된 MI-RPR 기술은 YOLOv7을 활용하여 관심 영역을 추출한 다음, 이 정보를 바탕으로 각 프레임에 대한 공간적 스케일 요소를 결정한다. 현재 VCMRS v0.8의 RoI 처리 단계에서는 SFU 데이터셋에 대해 detcon2<sup>[21]</sup>의 Faster R-CNN X-101 FPN을 사용하여 관심 영역 정보를 추출하고 있다. 본 논문에서는 객체 탐지 신경망의 변경이 성능에 미치는 영향을 관찰하고자 했다. 실험 결과, 기존에 Faster R-CNN을 사용했을 때는 BD-rate이 -2.08%였으나, YOLOv7을 사용했을 때는 BD-rate이 -4.02%로, 압축 효율이 향상된 것을 확인할 수 있었다. 이 결과는 신경망 기반의 관심 영역 추출에 YOLOv7을 적용함으로써, RoI 기반 압축 단계 및 공간적 샘플링 단계에서 다른 객체 탐지 신경망을 사용하는 것을 하나로 통일하여 부호화 시 계산 복잡도를 줄이는 동시에 성능을 개선할 수 있음을 보여준다.

## V. 결론

지능형 영상 소비의 증가 추세에 맞추어, MPEG에서는

VCM AhG을 통해 머신용 영상 부호화 기술에 대한 논의를 진행하고 있다. VCM 표준화 작업이 지속됨에 따라, 머신용 비디오 압축 기술에 관한 토론이 계속되고 있으며, 참조 소프트웨어의 성능도 지속해서 개선되고 있다. 본 논문에서는 머신을 위한 영상 압축 기술로 관심 영역 기반 스케일링 압축 방법을 제안하였다. 제안하는 방법은 입력 영상의 각 프레임을 원본 해상도 대비 75%, 50%로 다운스케일링하여 압축을 수행하고, 다운스케일링 상태에서 관심 영역이 탐지되는지를 확인한다. 이후, 머신 성능에 영향을 주지 않는 스케일 요소를 결정하여, 선택된 관심 영역에 대해 다운스케일링을 적용하여 압축 효율을 높였다. 실험 결과, 본 논문에서 제안하는 방법은 VCMRS v0.7과 비교할 때 BD-rate가 최대 -6.87%, VCMRS v0.8과 비교할 때 BD-rate가 최대 -4.02%로, 이는 제안된 압축 기법이 VCM 참조 소프트웨어 기존 앵커보다 우수한 것을 확인하였다. 그러나 VCMRS v0.8 기반 실험에서 LD 모드로 압축을 수행한 결과 앵커에 비해 압축 효율이 다소 낮게 나타나, 해당 부분에 관한 추가 연구가 필요한 것으로 보인다.

### 참 고 문 헌 (References)

- [1] L.-Y. 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.  
doi: <https://doi.org/10.1109/TIP.2020.3016485>
- [2] 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>.
- [3] 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>.
- [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] Cisco, Cisco Annual Internet Report (2018-2023), March 9, 2020.
- [6] WG 04, "Report of 9th meeting," ISO/IEC JTC1/SC29/WG4 output document N0249, October 2022.
- [7] C. Rosewarne, Y. Zhang, C. Hollmann and S. Liu, "BoG report on Video Coding for Machines," ISO/IEC JTC1/SC29/WG4 input document m61401, October 2022.
- [8] H. Zhang, "[VCM] Introduction to the VCM reference software (VCM-RS)," ISO/IEC JTC1/SC29/WG4 input document m62003, January 2023.
- [9] H. Wang, et al., "CE4.2 Temporal Resample coding from China Telecom," ISO/IEC JTC1/SC29/WG4 input document m63086, April 2023.
- [10] Huang, Zhewei, et al. "Real-time intermediate flow estimation for video frame interpolation," *European Conference on Computer Vision (ECCV)*, pp. 624-642, 2022.  
doi: [https://doi.org/10.1007/978-3-031-19781-9\\_36](https://doi.org/10.1007/978-3-031-19781-9_36).
- [11] M. H. Jeong, et al., "[VCM] Report on CE1.4," ISO/IEC JTC1/SC29/WG4 input document m64421, July 2023.
- [12] 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.
- [13] A. Kim, et al., "[VCM] Test results for VCM CE3.2," ISO/IEC JTC1/SC29/WG4 input document m66168, January 2024.
- [14] S. Rózek, et al., "[VCM] Improved RoI preprocessing and retargeting for VCM," ISO/IEC JTC1/SC29/WG4 input document m66523, January 2024.
- [15] 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.
- [16] H. Yang, S. Wang, C. Lin, C. Lin, "[VCM] The simplification of temporal up-sampling method," ISO/IEC JTC1/SC29/WG4 input document m66454, January 2024.
- [17] D. Ding, et al., "[VCM] A Curve Fitting Approach to Transform Non-Monotonic Test Data for BD-Rate Calculation," ISO/IEC JTC1/SC29/WG4 input document m65531, October 2023.
- [18] Y. Lee, et al., "[VCM] Comments on VCM anchor," ISO/IEC JTC1/SC29/WG4 input document m64809, October 2023.
- [19] WG 04, "Common test conditions for video coding for machines," ISO/IEC JTC1/SC29/WG4 output document N0427, October 2023.
- [20] WG 04, "Common test conditions for video coding for machines," ISO/IEC JTC1/SC29/WG4 output document N0467, January 2024.
- [21] detectron2, <https://github.com/facebookresearch/detectron2> (accessed March 25, 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