

일반논문 (Regular Paper)

방송공학회논문지 제29권 제3호, 2024년 5월 (JBE Vol.29, No.3, May 2024)

<https://doi.org/10.5909/JBE.2024.29.3.363>

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

MPEG FCM 테스트 모델에 대한 시간적 재 표본화 적용 및 성능 분석

장 승 환^{a)}, 정 혜 원^{a)}, 임 달 홍^{a)}, 김 휘 용^{a)†}

Analyzing Performance about MPEG FCM Test Model with Temporal Resampling

Seunghwan Jang^{a)}, Hyewon Jeong^{a)}, Dalhong Lim^{a)}, and Hui Yong Kim^{a)†}

요 약

최근 딥러닝 기술의 발전으로, 기계 시각 작업을 위한 기계의 영상 소비량이 큰 폭으로 증가했다. 이러한 기계의 영상소비량 증가에 따라 기계에 최적화된 압축 기술 연구에 대한 필요성이 커졌는데, MPEG의 WG04 에서 커지는 필요성에 발맞춰 기계를 위한 압축 기술 연구 그룹인 FCM(Feature Coding for Machines)을 창설했다. FCM은 2024년 1월에 있던 MPEG 145th 회의에서 우수한 성능을 보인 기고를 바탕으로 새로운 표준 테스트 모델인 FCTM-v2를 만들었다. 이 논문은 FCTM-v2에 시간적 재 표본화 기술을 적용하여 성능을 분석한다. 시간적 재 표본화 기술은 압축 대상이 되는 특징맵 비디오의 프레임들을 일정 주기로 표본화하여 발생 비트량을 저감하는 기술이다. 실험을 통해 시간적 재 표본화 기술이 CTTC에 정의된 모든 데이터셋에 대해서 성능향상이 있고, 인코딩과 디코딩 과정에서 연산 복잡도를 크게 줄임을 확인했다.

Abstract

Due to recent advancements in deep learning technology, there has been a significant increase in the volume of video consumed by machines for computer vision tasks. This shift has necessitated the development of compression technologies optimized for machines. In response to this need, the MPEG's WG04 established the Feature Coding for Machines (FCM) research group. Based on the outstanding contributions at the MPEG 145th meeting in January 2024, FCM developed a new common test model, FCTM-v2. This paper applies temporal resampling techniques to FCTM-v2 and analyzes their performance. Temporal resampling involves sampling the frames of feature map videos at regular intervals to reduce the bitrate. The experiments confirmed that temporal resampling improves performance across all datasets defined by CTTC, and significantly reduces computational complexity during the encoding and decoding processes.

Keyword : Feature Coding for Machines, NN-based coding, Compression, Video coding

a) 경희대학교 컴퓨터공학부(Kyung Hee University Department of Computer Science and Engineering)

† Corresponding Author : 김휘용(Hui Yong Kim)

E-mail: hykim.v@gmail.com

Tel: +82-31-201-3760

ORCID: <https://orcid.org/0000-0001-7308-133x>

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

※ 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 2023년도 SW중심대학사업의 결과로 수행되었음(2023-0-00042)

· Manuscript April 22, 2024; Revised May 13, 2024; Accepted May 13, 2024.

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

컴퓨터 비전 분야에서 딥러닝 기술이 널리 쓰임에 따라, 기계 시각 작업을 위한 영상의 소비량이 크게 증가했다. 이에 따라 기계의 시각 작업을 위한 영상 압축 기술의 필요성 커졌다. 하지만, 기계가 영상을 소비하는 형태는 사람이 영상을 소비하는 형태와 차이가 존재한다. 기계는 사람과 달리 원본 영상에서 중요한 특징을 추출한 특징맵을 사용하여 시각 작업을 수행하는데, 이러한 영상의 소비 방법의 차이로 기계 시각 작업을 최적화하기 위해서 새로운 압축 방법을 모색해야 한다. 이에 따라, 기계 시각 작업에 최적화된 영상 압축 기술에 대한 연구의 필요성을 느낀 ISO(International Standard Organization)에서 기계 시각 작업에 최적화된 특징맵 압축 기술을 연구하기 위해서 MPEG(ISO/IEC JTC1/SC 29)에서 VCM(Video Coding for Machines) 그룹을 창설했다^[1].

VCM 그룹은 상용 영상압축 코덱인 HEVC(High Efficiency Video Coding)^[2]나 VVC(Versatile Video Coding)^[3] 대비 우수한 특징맵 압축 효율을 갖는 압축 기술을 연구하기 위해 2023년 4월 CfP(Call for Proposal) 문서를 발행했고^[1], 23년 10월 144th 회의에서 CE(Core Experiment)1에서 한밭대학교와 China Telecom이 제안한 기고^[4,5]를 바탕으로 표준 테스트 모델인 FCTM-v1을, 145th 회의에서 경희대학교가 제안한 기고^[6]를 바탕으로 FCTM-v2를 만들었다.

기계 시각 작업을 위한 특징맵을 압축하는 작업과 사람이 소비하기 위한 영상을 압축하는 작업에는 두 가지 큰 차이가 있다. 먼저, 사람이 소비하는 영상은 밝기를 표현하는 단일 채널이나 색 정보를 포함하는 3채널로 표현되는 저차원 영상이지만, 기계 인지 작업을 위한 특징맵은 대체로 100채널이 이상의 다채널 영상을 압축해야 한다^[1]. 또한, 사람이 소비하는 영상은 한 프레임이 한 장의 영상으로 구성되지만, 기계를 시각 작업을 위한 특징맵은 한 프레임이 여러 레이어의 특징맵으로 이루어진 다계층 구조를 가질 수 있어 다 계층의 특징맵을 효율적으로 압축할 수 있어야 한다^[1].

기존에도 기계 시각 작업을 위한 다 계층 구조의 특징맵을 압축하기 위해 여러 가지 연구들이 있었다^[4,5,6,7,8,9,10,11,12]. 먼저 특징맵을 효과적으로 압축하기 위해서, 특징맵을 공간적으로 보다 작은 크기로 줄일 필요가 있다. 특징맵의 공간적 크기를 줄이는 방법으로, 주성분 분석과 같은 알고리즘을 사용하는 방법^[10,11,12]과 인공신경망을 사용하는 방법^[4,5,6,9]으로 나눌 수 있다. 이 중에서 인공신경망을 사용하는 연구들은 다계층 특징맵을 병렬적으로 또는 순차적으로 융합하여 단일 계층의 특징맵을 생성하고, 이 단일 계층의 특징맵을 오토인코더 구조를 통해 공간적으로 보다 작은 크기의 특징맵으로 압축한다^[4,5,6,7,8,9]. 앞선 과정을 통해 압축된 특징맵을 송신하기 위해 비트열로 변환하는 방법으로, 단순 피쳐맵 정보를 송신하는 방법^[7]과 인공신경망 기반의

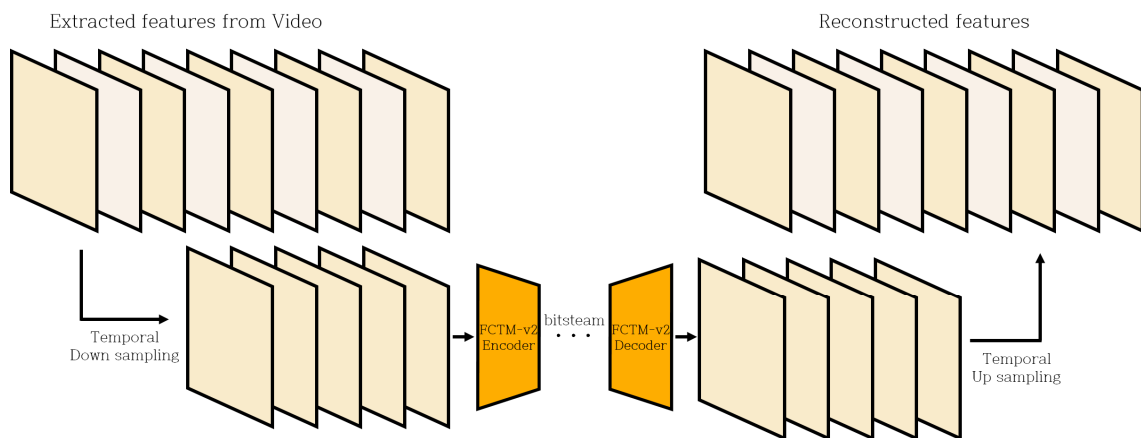


그림 1. FCTM-v2에서 시간적 재 표본화가 작동하는 구조
Fig. 1. Structure of temporal resampling in FCTM-v2

엔트로피 모델을 이용하는 방법^[9], 상용 코덱을 이용하는 방법이 있다^[4,5,6,8]. MPEG FCM CE1은 상용 코덱을 이용하여 비트열로 변환하는 방법을 연구하고 있는데, MPEG 145th 회의에서 경희대학교가 제안한 기고^[6]가 FCM CE1에 채택되어 표준 테스트 모델인 FCTM-v2^[14]가 만들어졌다.

압축된 특징맵을 비트스트림으로 변환하는 과정에서 인공신경망 기반의 엔트로피 모델을 이용하는 방법^[9]과, VVC^[3]와같은 상용 코덱을 이용하는 방법^[4,5,6]은 동영상 압축에 대응하는 방법에서 큰 차이를 갖는다. 효율적인 동영상 압축을 위해서 프레임 간 중복성을 이용해야 하는데, 인공신경망 기반의 엔트로피 모델은 프레임 간 중복성을 이용하기 위해서 별도의 신경망 모듈이 필요하다. 하지만, VVC^[3]와 같은 상용 코덱은 이미 잘 만들어진 동영상 압축 기술을 구비하고 있기에 별도의 모듈 없이 효과적으로 동영상을 압축할 수 있다^[4,5,6].

MPEG 145th 회의에서 표준 테스트 모델로 정해진 FCTM-v2는 인공신경망을 이용한 효과적인 다계층 특징맵 융합 및 복원 기술과 상용 코덱의 뛰어난 동영상 압축 기술 덕에 여러 기계 시각 작업 특징맵 대하여 우수한 압축 성능을 달성했다.

하지만, FCTM-v2은 느린 인코딩과 디코딩 속도를 갖는다는 한계가 있기에, 이러한 한계를 극복하기 위해서 인코딩과 디코딩 속도를 높일 수 있는 기술이 필요하다. 반면에, MPEG 145th 회의에서 기고된 시간적 재 표본화 기술^[11]은 압축 대상이 되는 프레임의 수를 절반으로 줄여, 높은 수준

의 인코딩과 디코딩 속도의 향상을 보였다. 하지만, 시간적 재 표본화 기술^[11]이 적용된 기고^[16,17]에서는 FCTM-v2가 아닌, 다른 압축 모델을 사용하기에 FCTM-v2에서의 성능이 검증되지 못했다는 한계가 존재한다.

본 논문에서는 높은 압축 성능을 보여준 FCTM-v2의 압축 파이프라인에, [그림 1]과 같이 MPEG 145th 회의에서 기고된 시간적 재 표본화^[15] 기술을 추가하여 인코딩과 디코딩 시간을 획기적으로 줄이고 추가적인 압축성능의 향상을 달성했다. 시간적 재 표본화 기술은 기존 압축 파이프라인의 전처리 및 후처리로 시간적 표본화와 시간적 보간을 수행한다. 먼저 시간적 표본화 절차에서는 인코더 단에서, 입력 다계층 특징맵의 프레임을 일정 주기로 표본화하는데, 본 논문의 경우 표본화 주기를 2로 설정하여 두 프레임 당 한 프레임을 표본화 하였다. 인코더 직전 단에서 표본화가 진행된 후, 기존 압축 파이프라인을 거쳐 디코더 단에서 표본화된 상태의 복원된 특징맵 동영상을 획득하게 된다. 이렇게 획득한 특징맵 동영상에 시간적 보간을 적용하여 표본화 과정에서 버려진 특징맵 영상을 복원한다.

본 논문에서는 시간적 재 표본화 기술^[15]을 표준 테스트 모델인 FCTM-v2에 적용 및 성능을 측정하여, 시간적 재 표본화 기술이 기계 인지 작업 성능 감소 대비 높은 비트율 저감으로 전체적인 압축 성능 향상이 있음을 보인다. SFU 데이터셋에서는 발생 비트량에 따른 객체 탐지 성능의 BD-rate가, 클래스 A/B 에서 -18.40%, 클래스 C에서 -12.61%, 클래스 D에서 -24.56%의 BD-rate 이득이 있었고, TVD와 HiEve 데이터셋에서 발생 비트레이트에 따른

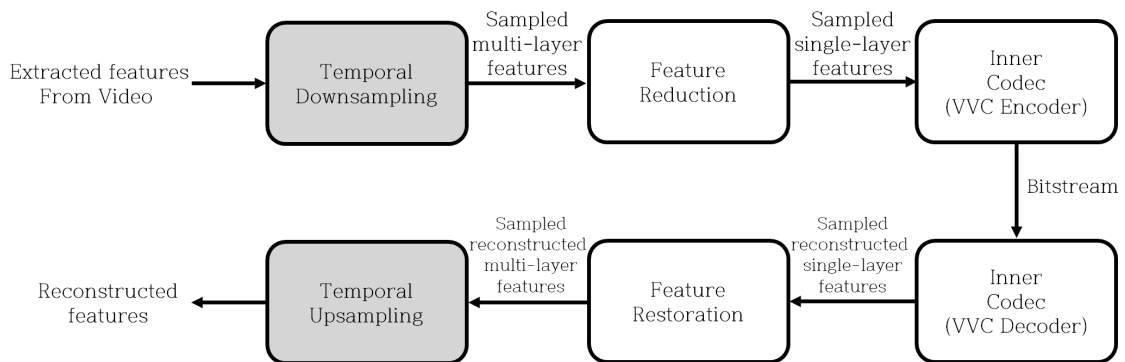


그림 2. 시간적 재 표본화를 적용한 FCTM-v2의 작동 파이프라인
 Fig. 2. Pipeline of FCTM-v2 with temporal resampling applied

객체 추적 성능의 BD-rate가, TVD에서 -12.42%, HiEve (1080p)에서 -21.64%, HiEve (720p)에서 -5.08%의 BD-rate 이득이 있었다. 연산복잡도 측면에서는 인코딩 과정에서, SFU는 -47.75%에서 -49.95%, TVD는 -43.53%에서 -49.23%, HiEve에서는 -50.27%에서 -58.29%의 절반에 가까운 압축 시간 감소를 보였다. 디코딩 과정에서는 SFU가 -45.63%에서 -47.42%, TVD에서 -25.94%에서 -29.57%, HiEve에서 -31.90%에서 -31.95%로 큰 폭의 연산량 감소가 있었다.

II. 관련 연구

1. FCM Common Test Model version 2

기계 시각 작업에 효과적인 압축 코덱을 만들기 위해서 다양한 기술들이 연구되어 왔다^[6,7,15,18,19,20]. 이런 다양한 특징맵 압축 기술들의 성능을 검증하기 위해서 표준 테스트 모델이 필요하다. FCM에서도 이러한 표준 테스트 모델의 필요성을 만족하기 위해서, 2024년 1월에 있었던 MPEG 145th 회의에서 경희대학교가 제안한 기고^[6]를 바탕으로 기존 FCTM을 개량한 FCTM-v2를 만들었다. 경희대학교가 제안한 기고^[6]는 L-MSFC^[12]를 발전시킨 L-MSFC-v2^[9]를 기반으로 하여, 특징맵을 비트열로 변환하는 방법으로 사용 코덱인 VVC^[3]를 접목한 하이브리드 구조를 갖는다. L-MSFC는 인공지능망을 이용한 특징맵 추출 및 복원을 수행하며, 엔트로피 모델을 사용하여 특징맵을 비트열로 변환하기에 종단 간 학습을 통하여 압축 성능을 최적화 할 수 있다는 장점이 있다. L-MSFC-v2는 L-MSFC의 신경망

구조를 ELIC^[21]과 유사한 구조로 개량한 것으로, L-MSFC 대비 높은 압축 성능을 보인다.

2. 시간적 재 표본화 (Temporal resampling)

MPEG 145th 회의에서 시간적 재 표본화 기술^[15] 및 이를 이용한 모델이 기고되었다^[22,23]. 이 기고서^[22,23]에는 시간적 재 표본화 기술을 적용한 하이브리드 구조의 모델과 VVC로 인코딩 및 디코딩할 때 참조하는 환경변수의 모음인 구성(config) 파일이 정의되어 있으며, 이 구성 파일에서 이미지 프레임 그룹인 GOP(Group of Pictures) 크기를 기존 크기 대비 절반으로 정의한다. 이러한 기술들을 통해, 하이브리드 구조의 모델에 시간적 재 표본화 기술을 적용하면 비트량 대비 기계 시각 작업 성능이 향상됨을 보였다.

3. Common Test and Training Condition

MPEG 144th 회의에서 FCM에 대한 CTTC(Common Test and Training Condition)^[13]가 정해졌다. CTTC의 테스트 환경은 [표 1]과 같이 5개의 데이터셋에 대하여 정의되어 있다. 각각의 데이터셋에 대하여 기계 시각 작업이 정의되어 있고, 이 기계 시각 작업에 사용될 네트워크와 압축될 대상 특징맵의 네트워크 내 위치를 나타내는 분할점이 정의되어 있다. 또한, 기계 시각 작업 성능 대비 비트 발생률을 계산하기 위해, 비트 발생률 측정 지표와 기계 시각 작업 성능 측정 지표가 정의되어 있다. 다섯 개의 데이터셋 중 비디오 데이터셋으로 SFU, TVD, HiEve가 있으며, SFU는 FPN^[24] 기반, TVD와 HiEve는 DarkNet-53 기반^[25]의 특징맵 추출 과정을 거친다.

표 1. CTTC에 정의된 데이터셋에 대한, 작업과 특징맵 추출 네트워크와 추출 분할 포인트 및 비트율 측정지표와 작업 성능 측정 지표
Table 1. Tasks and feature map extraction Network with Split points defined in CTTC datasets, along with bitrate measurement metrics and task performance evaluation metrics

Dataset	Task	Network	Split point	Rate measure	Task measure
OpenImagesV6	Object detection	FasterRCNN-X101-FPN	P-layer (P2-P5)	BPP	mAP @ 0.5
OpenImagesV6	Instance segmentation	MaskRCNN-X101-FPN	P-layer (P2-P5)	BPP	mAP @ 0.5
TVD	Object tracking	JDE-1088x608	Darknet-53	Kbps	MOTA
HiEve	Object tracking	JDE-1088x608	Layers 75, 90, and 105 ("ALT1")	Kbps	MOTA
SFU	Object detection	FasterRCNN-X101-FPN	P-layer (P2-P5)	Kbps	mAP @ 0.5-0.95

III. 시간적 재 표본화

1. 전체 파이프라인 (Overall Pipeline)

시간적 재 표본화 과정^[11]이 추가된 FCTM-v2^[14]의 실행 파이프라인은 [그림 2]와 같은 구조를 갖는다. 원본 이미지에서 특징맵 추출 과정을 통해 추출된 특징맵을 특징맵 압축 파이프라인에 입력한다. 입력된 특징맵은 시간적 저해상화를 수행하는데, 이 과정에서 입력된 표본화 주기에 따라 입력 특징맵의 시간적 저해상화 수준이 결정된다. 본 논문에서는 표본화 주기를 2로 하여, 특징맵 동영상 두 프레임 당 한 프레임을 표본화하게 된다. 표본화된 프레임들은 기존 FCTM-v2의 압축 파이프라인과 동일한 절차를 걸쳐 압축되고 복원된다. 복원된 프레임들은 표본화가 수행된 상태이므로, 기계 시각 작업에 입력하기 위해 시간적 고해상화를 통해서 프레임 보간을 수행한다.

2. 시간적 재 표본화 (Temporal Resampling)

시간적 재 표본화는 크게, 시간적 저해상화와 시간적 고해상화로 구성된다. 시간적 저해상화는 [그림 3(b)]와 같이 일정 간격으로 프레임을 표본화 하는데, 이 논문에서는 프레임이 표본화되는 간격을 2로 하여 두 프레임마다 한 프레임을 표본화한다. 즉, 영상의 번호를 나타내는 POC에 대해서, 2 POC마다 한번 표본화가 수행된다. 이때, 최종 프레임은 복원과정에서 이후 프레임을 참조할 수 없기 때문에 표본화를 적용하지 않으며, 화면 내 예측을 수행하는 I 프레임에는 표본화를 적용하지 않고 화면 간 예측을 수행하는 B 프레임만 표본화가 적용된다. 시간적 저해상화를 수행하게 되면, 압축 대상 프레임이 대략 절반으로 줄어들어 압축 파이프라인에서 발생 비트량을 감소시키고, 연산량을 저감하는 효과가 있다.

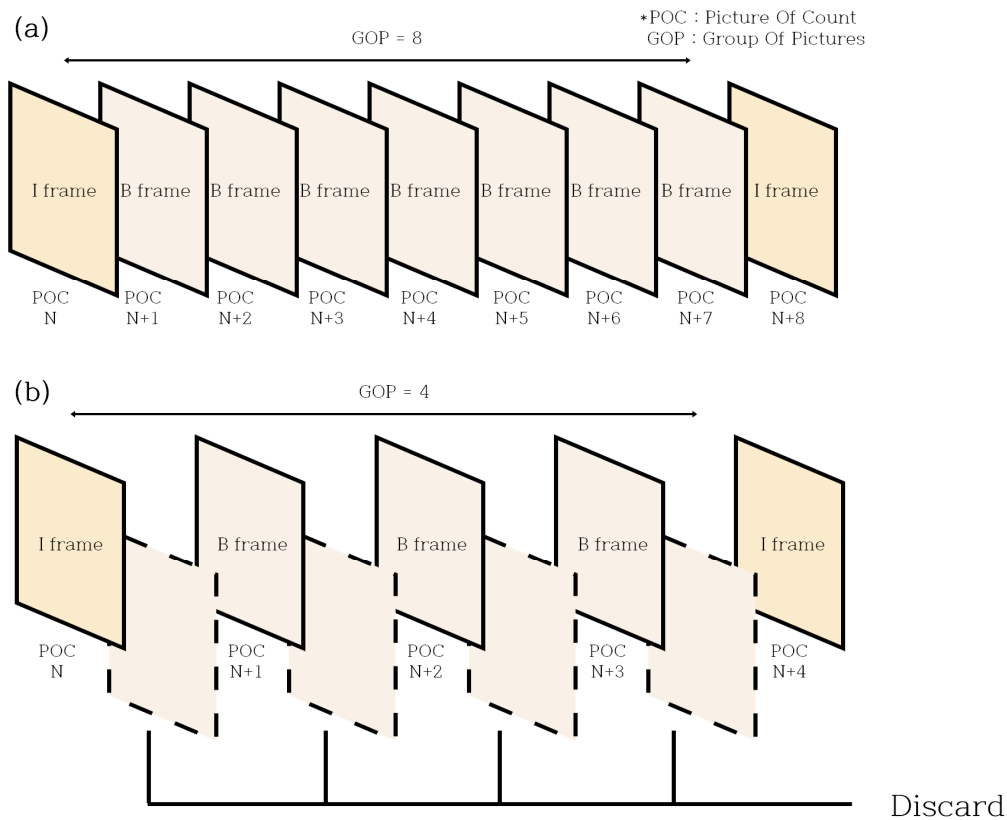


그림 3. (a) 시간적 재 표본화를 적용하지 않은 GOP 구조와 (b) 적용한 GOP 구조비교
 Fig. 3. Comparison of GOP structure (a) without temporal resampling applied and (b) with temporal resampling applied

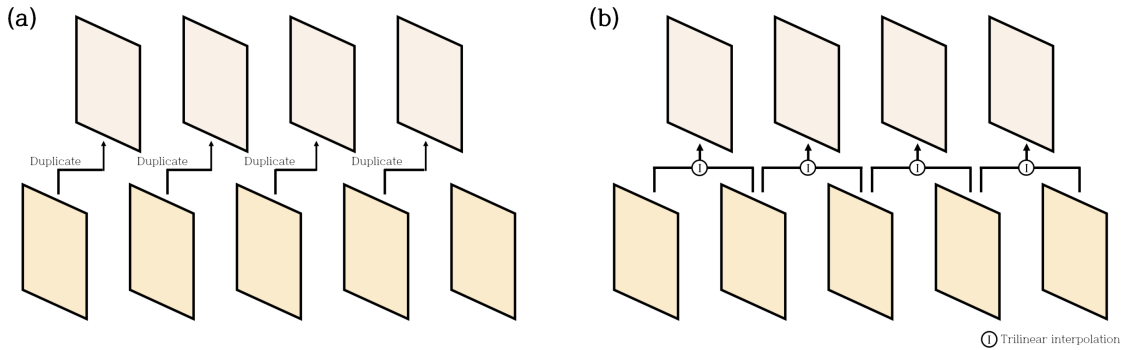


그림 4. 시간적 고해상화를 위해 (a) 이전 프레임을 복사하는 기술과 (b) 앞뒤 인접 프레임으로 보간하는 기술
 Fig. 4. Techniques for temporal super-resolution enhancement (a) duplicating previous frames and (b) interpolating between adjacent frames in both directions

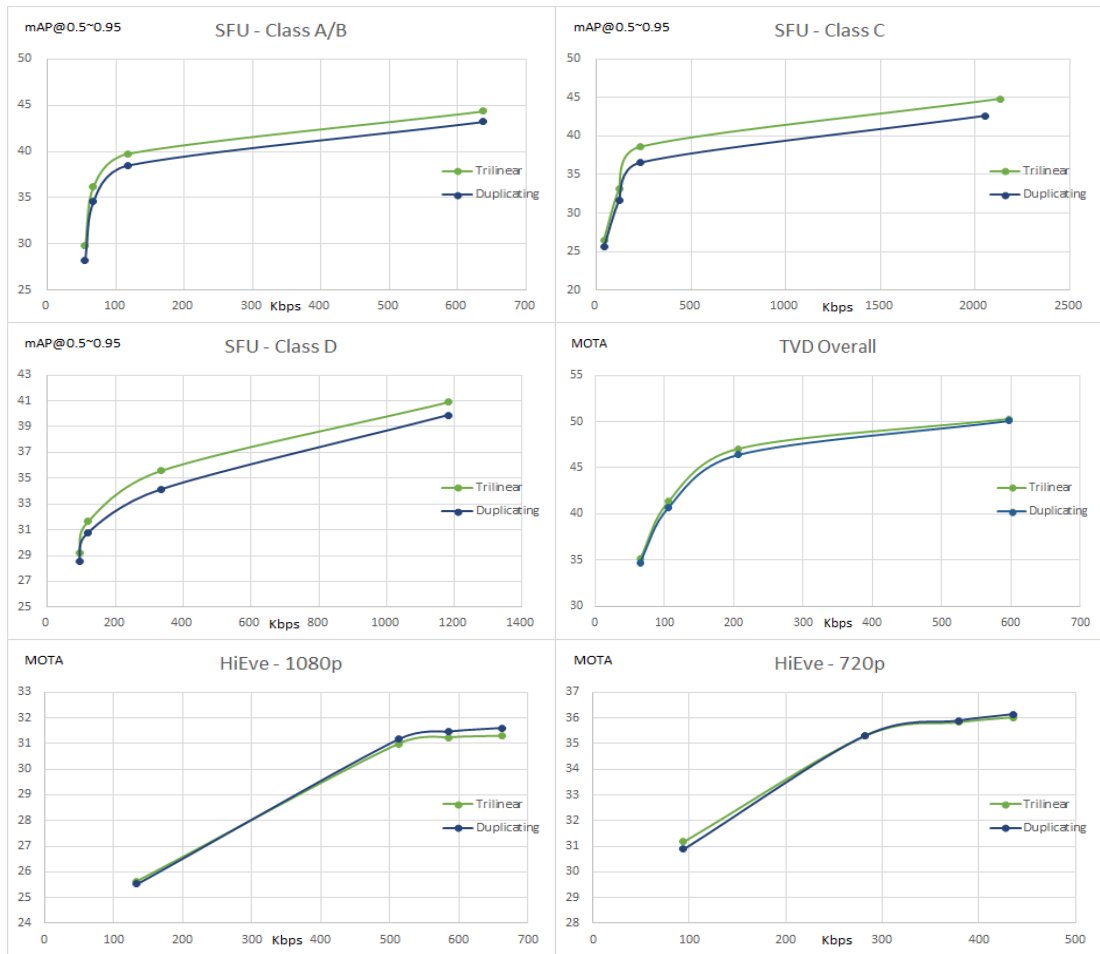


그림 5. 시간적 고해상화 과정에서, 삼선형 보간법과 이전 프레임 복사를 통한 보간법과의 성능 비교 그래프
 Fig. 5. Performance comparison graph between trilinear interpolation and interpolation through copying previous frames in the process of temporal high-resolution enhancement

시간적 고해상화는 [그림 4(b)]와 같이 앞뒤 프레임을 사용하여 중간 프레임을 선형 보간하는 방법을 사용한다. 이때, 입력 데이터가 3차원 정보를 갖기에, 이를 보간 하기 위해서 삼선형 보간법(Trilinear interpolation)을 사용한다^[15].

$$p_t(x, y, z) = \frac{p_{t-1}(x, y, z) + p_{t+1}(x, y, z)}{2} \quad (1)$$

식 (1)은 삼선형 보간법을 이용하여, t 번째 텐서의 x, y, z 좌표에서의 값인 $p_t(x, y, z)$ 를 $t-1$ 와 $t+1$ 번째 텐서를 이용하여 생성하는 방법을 표현한 것으로, x, y, z 좌표에서의 값인 $p_t(x, y, z)$ 는 앞뒤 프레임의 동일 x, y, z 좌표에서의 값에 평균을 취하여 값을 생성한다.

[그림 5]는 이전 프레임을 복제하여 사용하는 방법과 전·후 프레임으로 중간 프레임을 삼선형 보간법으로 보간하는 방법의 성능을 비교한 그래프이다. [그림 5]를 보면 삼선형 보간법을 사용하는 것이 [그림 4(a)]처럼 이전 프레임을 복제하는 것보다 대부분의 경우에서 더 높은 압축 성능을 보인다.

3. VVC config

CTTC^[13]에는 inner codec인 VVC^[3]의 압축 구성 파일이 정의되어 있다. 구성 파일에는 I-frame의 삽입 주기 또한 정의되어 있는데, 시간적 표본화를 적용하면 프레임의 개수가 줄어들면서 기존 구성 파일을 썼을 때와 다른 개수와 위치의 I-frame을 갖게 된다. 동일한 개수와 위치의 I-frame을 유지하기 위해서, 구성 파일에 정의된 I-frame 삽입 주기를

시간적 표본화 수준에 맞춰서 수정할 필요가 있다. 본 논문에서는 [그림 3(b)]와 같이 2 프레임 주기로 표본화가 진행되었기에, I-frame의 삽입 주기를 2로 나누면 시간적 표본화를 적용하지 않은 경우와 동일한 I-frame 개수 및 위치를 얻을 수 있다.

IV. 실험 결과 및 분석

우리는 MPEG 145th 회의에서 정의된 CTTC 환경에서 정의된 동영상 데이터셋, SFU와 TVD, HiEve에 대해 FCTM-v2^[14]에 시간적 재 표본화 기술^[15]을 적용하여 실험했다.

FCTM-v2에 시간적 재 표본화 기술을 적용한 성능을 CTTC 환경에 맞춰, FCTM-v2와 성능 비교를 실시했다. CTTC에 정의된 목표 성능인 PP (Performance Point) 범위^[13]를 가급적 맞추기 위해서, 시간적 재 표본화를 적용한 모델에서 SFU와 HiEve의 VVC 설정 QP를 일부 수정하여 성능을 측정했다, 결과적으로, 동영상 데이터셋인 SFU와 TVD, HiEve에대해서 시간적 재 표본화 기술은 유의미한 BD-rate 성능 향상을 달성했다.

시간적 재 표본화 기술^[15]은 특징맵 압축 과정에서 압축 대상이 되는 프레임의 수를 절반으로 줄여 연산 복잡도를 크게 줄임과 동시에, 압축 성능의 향상을 보였다. [그림 7]은 특징맵 추출 및 복원 시간에 대한 인코딩 및 디코딩 시간의 비를 나타내고, 이를 데이터셋 별로 평균을 취한 결과를 표기한 것이다. 시간적 재 표본화를 적용하면, 적용

표 2. FCTM-v2대비 시간적 재 표본화를 적용한 FCTM-v2의 압축 성능 비교표

Table 2. Comparison chart of compression performance between FCTM-v2 and FCTM-v2 with temporal resampling applied

Proposal vs FCTM-v2		
Task	Dataset	BD-rate
Object detection	SFU (Class A/B)	-18.40%
	SFU (Class C)	-12.61%
	SFU (Class D)	-24.56%
Object Tracking	TVD (OVERALL)	-12.42%
	HiEve (1080p)	-21.64%
	HiEve (720p)	-5.08%

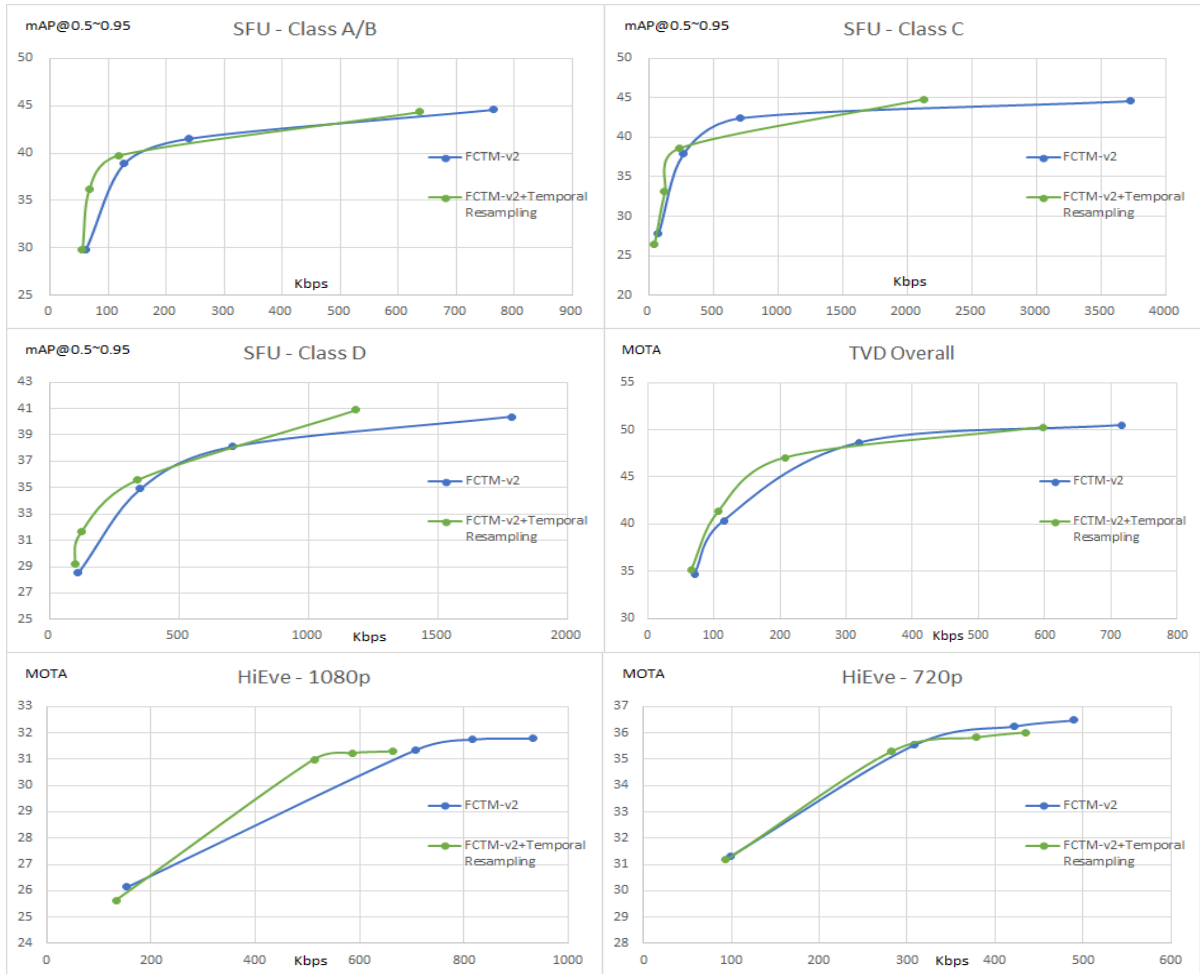


그림 6. 데이터셋 별, FCTM-v2와 시간적 재 표본화를 적용한 FCTM-v2의 압축 성능 비교
 Fig. 6. Comparison of compression performance between FCTM-v2 and FCTM-v2 with temporal resampling, by dataset

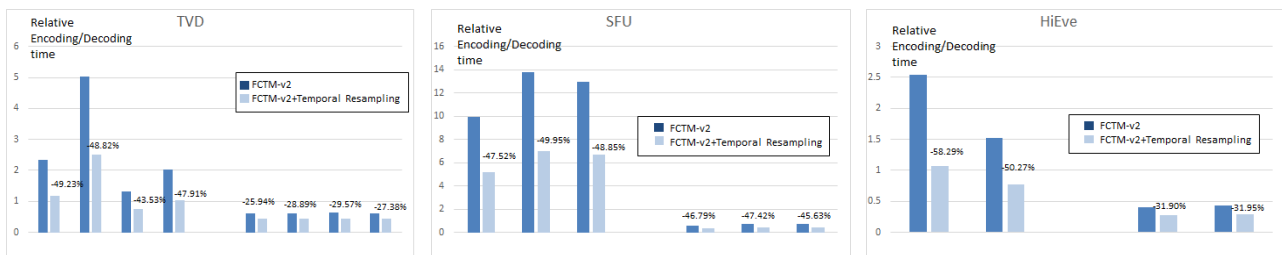


그림 7. 데이터셋 별, FCTM-v2와 시간적 재 표본화를 적용한 FCTM-v2의 인코딩/디코딩 시간 비교 그래프
 Fig. 7. Graph comparing encoding/decoding time between FCTM-v2 and FCTM-v2 with temporal resampling applied, by dataset

하지 않았을 때와 비교해 인코딩 과정 연산시간이 적게는 -43.53%(TVD-03)에서 많게는 -58.29%(HiEve-1080p)까지 감소하는 등 절반 가까운 연산량 절감을 보였다. 디코

딩 때는 적게는 -25.94%(TVD-01)에서 많게는 -47.42%(SFU Class B)까지 감소하는 등 디코딩 과정에서도 유의미한 연산량 절감을 보였다.

표 3. 데이터셋 별, FCTM-v2와 시간적 재 표본화를 적용한 FCTM-v2의 연산 복잡도 비교 표

Table 3. Comparison table of computational complexity between FCTM-v2 and FCTM-v2 with temporal resampling applied, by dataset

Proposal vs FCTM-v2		Complexity reduction	
Dataset	Quality point	Encoding time	Decoding time
SFU Class AB	1	-41.38%	-45.46%
	2	-57.15%	-48.12%
	3	-55.53%	-45.17%
	4	-36.00%	-48.40%
SFU Class C	1	-47.79%	-45.45%
	2	-52.55%	-52.19%
	3	-47.72%	-46.02%
	4	-51.73%	-46.04%
SFU Class D	1	-42.12%	-47.41%
	2	-47.96%	-45.20%
	3	-62.70%	-44.86%
	4	-42.62%	-45.03%
TVD-TOTAL	1	-50.26%	-34.20%
	2	-53.66%	-20.72%
	3	-47.45%	-22.30%
	4	-45.55%	-26.54%
HiEve-1080p	1	-53.68%	-33.64%
	2	-58.41%	-28.59%
	3	-52.76%	-31.41%
	4	-68.31%	-33.97%
HiEve-720p	1	-46.06%	-32.20%
	2	-46.36%	-32.28%
	3	-46.94%	-28.90%
	4	-61.71%	-34.41%

V. 결론

본 논문에서는 현재 FCM의 표준 테스트 모델인 FCTM-v2^[14]에 대해서 시간적 재 표본화 기술^[15]의 유효성을 보였다. 시간적 재 표본화 기술은 기존 압축 파이프라인 전후 단계에서 특징맵을 표본화·보간을 하여, 압축 대상이 되는 특징맵의 프레임 빈도를 조절한다. 시간적 재 표본화 기술은 CTTC에 정의되어 있는 모든 비디오 데이터셋에 대해서 BD-rate 감소 및 압축 파이프라인의 작업 시간이 감소함을 보였다.

본 논문의 시간적 재 표본화 기술은 적은 기계 시각 작업 성능 감소로 큰 비트량 감소를 통해, 압축 성능을 향상시켜 BD-rate를 감소시켰다. 향후 연구 과제로, 추가적 성능향상을 위해 시간적 고해상도 과정에서 삼선형 보간이 아닌 더 효율적인 보간 방법을 탐색할 필요가 있다, 또, 본 논문에서

는 시간적 재 표본화 비율을 2로 설정하여 두 프레임당 한 프레임을 표본화하는데, 더 효과적인 시간적 재 표본화 빈도를 탐색한다면, 추가적 성능 향상을 기대할 수 있을 것이다.

참고 문헌 (References)

- [1] WG2, "Call for Proposals on Feature Compression for Video Coding for Machines," MPEG Technical requirements, vol. ISO/ IEC JTC 1/SC 29/WG 2, no. N282, Apr 2022.
- [2] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," IEEE Trans. CircuitsSyst. Video Technol., vol. 22, no. 12, pp. 1649 - 1668, Dec. 2012. doi: <https://doi.org/10.1109/icip.2015.7351182>
- [3] B. Bross, Y.-K. Wang, Y. Ye, S. Liu, J. Chen, G. J. Sullivan, and J.-R. Ohm, "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/10.17762/turcomat.v10i2.13578>
- [4] H. Han, J. Kang, J. Kim, H. Choi, S. Jung, J. Lee, S. Kwak, W. Cheong, H. Choo, "[FCVCM] Response to CfP : Enhanced Multi-scale Feature Compression for FCVCM", ISO/IEC JTC 1/SC 29/WG 4, m65705, Hannover, October 2023.
- [5] H. Wang, Z. Wu, H. Wang, Y. Zhang, N. Guo, C. Luo, "[FCVCM] Response to the FCVCM CfP from China Telecom", ISO/IEC JTC 1/SC 29/WG 4, m65718, Hannover, October 2023.
- [6] H. Jeong, S. Jang, D. Lim, H. Kim, J. Lee, S. Jeong, Y. Kim, "[FCM] CE 1.1.8. L-MSFC-v2 with fine-tuning", ISO/IEC JTC 1/SC 29/WG 4 m66341, Online, January 2024.
- [7] Zhang, Z., Wang, M., Ma, M., Li, J., & Fan, X, "Msfc: Deep feature compression in multi-task network". In: 2021 IEEE International Conference on Multimedia and Expo (ICME). IEEE, p. 1-6, 2021. doi: <https://doi.org/10.1109/icme51207.2021.9428258>
- [8] Seunghwan Jang, et al. "VVC-based Multi Scale Feature Compression with Learnable Feature Fusion/Reconstruction". Proceedings of the Korean Society for Broadcasting and Media Engineers Conference, 977-980, 2023. link: <https://www.dbpia.co.kr/pdf/pdfView.do?nodeId=NODE11514266>
- [9] H. Jeong, Y. Kim, J. Yu, S. Jang, D. Lim, H. Kim, J. Lee, S. Jeong, Y. Kim, "[FCM] CE 3.2.2 L-MSFC-v2", ISO/IEC JTC 1/SC 29/WG 4 m66342, Online, January 2024.
- [10] C.Rosewarne, R.Nguyen, "[FCM] CE2.3.1: Online PCA", ISO/IEC JTC 1/SC 29/WG 4 m66340, Online, January 2024.
- [11] Y. Kim, S.Jeong, J.Lee, M.Kim, J.Lee "[FCM] CE2.3.6 PreBM PCA +ECS", ISO/IEC JTC 1/SC 29/WG 4 m66464, Online, January 2024.
- [12] F. Racapé, H. Choi, "[FCM] CE2.3.3 and CE2.3.4: Feature reduction based on channel grouping and tensor downscaling", ISO/IEC JTC

- 1/SC 29/WG 4 m66519, Online, January 2024.
- [13] WG 4, MPEG Video Coding, "Common test and training conditions for FCM", ISO/IEC JTC 1/SC 29/WG 4 N00459, Online, January 2024.
- [14] WG 4, MPEG Video Coding, "Algorithm Description of FCTM", ISO/IEC JTC 1/SC 29/WG 4 N00460, Online, January 2024.
- [15] Ding Ding, Yucheng Gong, Zhenzhong Chen, "[FCM] CE 2.1.1 Temporal Resampling", ISO/IEC JTC 1/SC 29/WG 4 m66271, Online, January 2024.
- [16] Ding Ding, Yucheng Gong, Zhenzhong Chen, "[FCM] CE 1.1.5 FENet/DRNet with Compact Channel + Temporal Resampling", ISO/IEC JTC 1/SC 29/WG 4 m66272, Online, January 2024.
- [17] Ding Ding, Yucheng Gong, Zhenzhong Chen, "[FCM] CE 1.1.6 Multi-Scale Down/up Sampler + Temporal Resampling", ISO/IEC JTC 1/SC 29/WG 4 m66273, Online, January 2024.
- [18] KIM, Yeongwoong, et al. "End-to-end learnable multi-scale feature compression for vcm". IEEE Transactions on Circuits and Systems for Video Technology, 2023.
doi: <https://doi.org/10.1109/tcsvt.2023.3302858>
- [19] YOON, Curie, et al. "MEDO: Minimizing Effective Distortions Only for Machine-Oriented Visual Feature Compression". In: 2023 IEEE International Conference on Visual Communications and Image Processing (VCIP). IEEE, 2023. p. 1-5, 2023.
doi: <https://doi.org/10.1109/vcip59821.2023.10402661>
- [20] Hyewon Jeong, et al. "Batch-Merging Training Based Feature Map Compression Method for Multi-Task Support". Korean Institute of Electrical Engineers Conference, 1260-1264, 2023. <https://www.dbpia.co.kr/pdf/pdfView.do?nodeId=NODE11522336>
- [21] HE, Dailan, et al. "Elic: Efficient learned image compression with unevenly grouped space-channel contextual adaptive coding". In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, p. 5718-5727, 2022.
doi: <https://doi.org/10.1109/cvpr52688.2022.00563>
- [22] Ding Ding, Zhenzhong Chen, "[FCM] CE 1.1.5 FENet/DRNet with Compact Channel + Temporal Resampling", ISO/IEC JTC 1/SC 29/WG4 m66272, Online, January 2024.
- [23] Ding Ding, Zhenzhong Chen, "[FCM] CE 1.1.6 Multi-Scale Down/up Sampler + Temporal Resampling", ISO/IEC JTC 1/SC 29/WG4 m66273, Online, January 2024.
- [24] LIN, Tsung-Yi, et al. "Feature pyramid networks for object detection". In: Proceedings of the IEEE conference on computer vision and pattern recognition, p. 2117-2125, 2017.
doi: <https://doi.org/10.1109/access.2021.3100369>
- [25] Z. Wang, L. Zheng, Y. Liu, et al. "Towards real-time multi-object tracking," in European Conference on Computer Vision (ECCV), p. 107-122, 2020.
doi: <https://doi.org/10.1109/mot.2001.937980>

저 자 소 개



장 승 환

- 2019년 3월 ~ 현재 : 경희대학교 컴퓨터공학과 학사
- ORCID : <https://orcid.org/0009-0000-4337-2064>
- 주관심분야 : 영상처리, 비디오 부호화, 컴퓨터 비전, 머신러닝, 딥러닝



정 혜 원

- 2022년 2월 : 경희대학교 소프트웨어융합학과 학사
- 2024년 2월 : 경희대학교 컴퓨터공학과 석사
- 2024년 3월 ~ 현재 : 경희대학교 컴퓨터공학과 박사
- ORCID : <https://orcid.org/0000-0003-1230-870X>
- 주관심분야 : 영상처리, 비디오 부호화, 컴퓨터 비전, 머신러닝, 영상 부호화 국제화 표준

저 자 소 개



임 달 홍

- 2019년 3월 ~ 현재 : 경희대학교 컴퓨터공학과 학사
- ORCID : <https://orcid.org/0009-0000-2733-5036>
- 관심분야 : 영상처리, 비디오 부호화, 컴퓨터 비전, 머신러닝, 영상 부호화 국제화 표준



김 휘 용

- 1994년 8월 : KAIST 전기및전자공학과 공학사
- 1998년 2월 : KAIST 전기및전자공학과 공학석사
- 2004년 2월 : KAIST 전기및전자공학과 공학박사
- 2003년 8월 ~ 2005년 8월 : ㈜에드팩테크놀로지 멀티미디어팀 팀장
- 2005년 11월 ~ 2019년 8월 : 한국전자통신연구원(ETRI) 실감AV연구그룹 그룹장
- 2013년 9월 ~ 2014년 8월 : Univ. of Southern California (USC) Visiting Scholar
- 2019년 9월 ~ 2020년 2월 : 숙명여자대학교 전자공학전공 부교수
- 2020년 3월 ~ 현재 : 경희대학교 컴퓨터공학과 부교수
- ORCID : <https://orcid.org/0000-0001-7308-133X>
- 관심분야 : 비디오 부호화, 딥러닝, 영상처리, 디지털 홀로그램