

# 뉴스 레터

## 회원 소개

### 박상효 교수

경북대학교 IT대학 컴퓨터학부



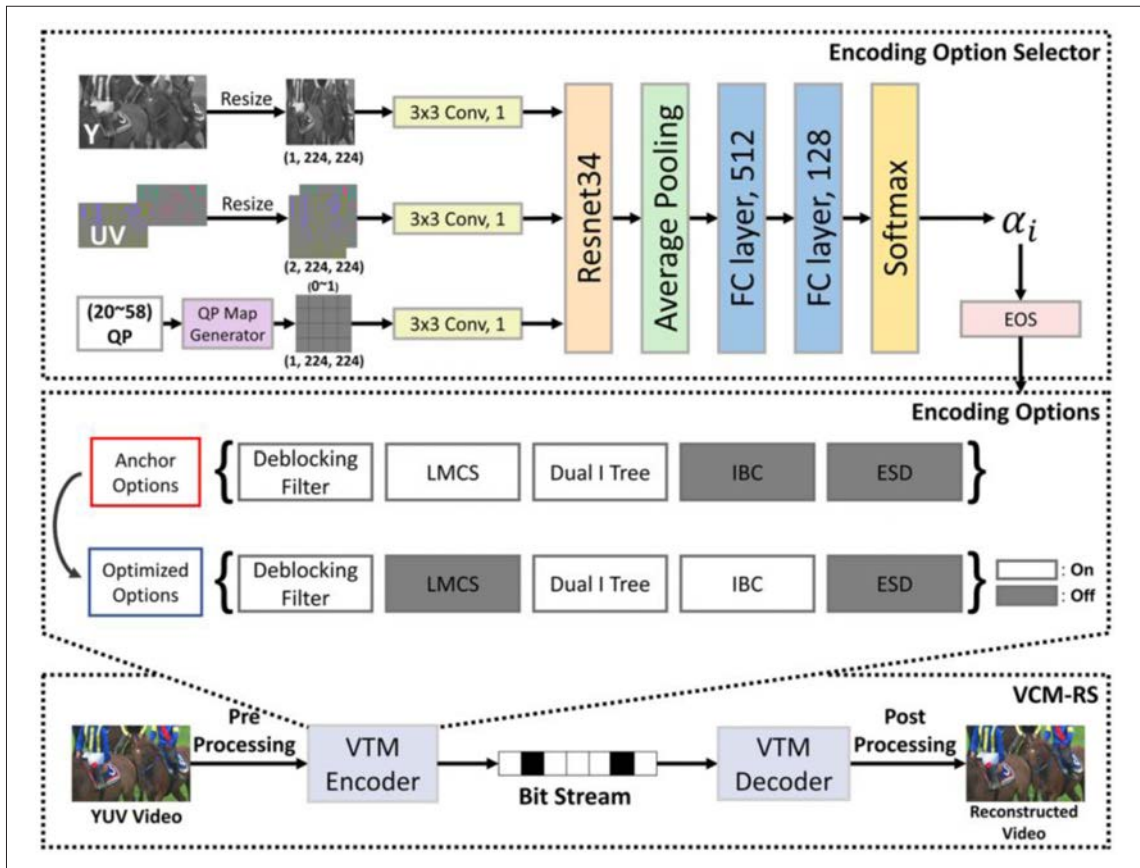
## I. 자기 소개

박상효 교수는 현재 경북대학교 IT대학 컴퓨터학부의 부교수 직을 수행하고 있으며, 인공지능컴퓨팅 전공주임을 담당하고 있다. 또한, 현재 동영상지능연구실(Video Intelligence Lab)을 운영하고 있다. 대표적으로 Data compression, Multi-modal learning, 2D/3D reconstruction 등을 연구하고 있으며, 미디어/AI 분야 최상위 저널 및 국제학회에 논문을 게재 및 발표하였다. 또한, 국제표준화기구 MPEG에 다년간 한국대표단으로 참여하며, Systems/Video/3D 그룹에 활발히 표준화 기고를 발표하고 있다.

## II. 연구 분야

### 1. Video Coding for Machines

영상 데이터의 처리, 전송, 저장 요구가 증가함에 따라 효율적인 비디오 압축 기술이 필수적이다. 현재 가장 앞선 압축 기술인 Versatile Video Coding(VVC) 표준은 이전 표준인 HEVC 대비 최대 40%, AVC 대비 75%의 비트레이트 절감을 제공한다[1]. 그러나 기존의 인간 시각 최적화 코덱을 기계 비전용 영상에 그대로 적용하는 것은 적절하지 않을 수 있다. 이에 따라 MPEG을 중심으로 Video Coding for Machines(VCM) 프로젝트가 추진되었으며, 이 프로젝트는 기계가 감지한 관심 영역(ROI)의 화질을 높이고, VVC를 기반으로 한 압축 기술을 적용해 압축 효율을



<그림 1> 기계 비전용 동영상 압축을 위한 딥러닝 기반 전환형 적응 네트워크

향상시키는 방식을 제안한다[2]. 이를 통해 VVC 단독 사용 대비 Tencent Video 데이터셋에서 39.73%의 압축 효율 개선이 확인되었다.

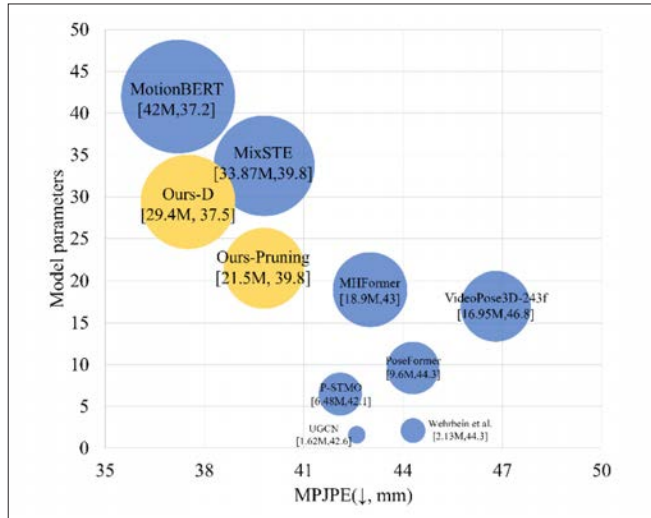
MPEG 및 여러 연구 그룹은 기계 비전 작업을 위한 영상 압축 성능 향상을 위해 ROI 기반 전처리, 시공간 리샘플링 등의 방법을 연구해 왔다. 하지만 기존 VVC 인코딩 옵션은 인간 시각 최적화에 중점을 두고 있어 기계 비전에는 적합하지 않을 수 있다. MPEG VCM 접근법에서는 ROI 외부 영역의 화질을 크게 낮추거나 단색(그레이스케일)으로 변환하는 방식이 사용되지만, 이는 비효율적인 면이 있다. 특히 기계 비전 작업에서는 입력 영상의 특성을 고려한 적응형 인코딩 전략이 필요하며, 모든 영상에 동일한 인코딩 옵션을 일괄 적용하는 것은 최적의 방식이 아닐 수 있다.

이에 본 연구실에서는 VCM 인코딩 옵션이 성능에 미치는 영향을 체계적으로 분석하고, 각 영상 시퀀스에 최적화된 인코딩 도구를 자동으로 선택할 수 있도록 설계된 딥러닝 기반 적응형 전환 네트워크(DASM)를 제안하였다.

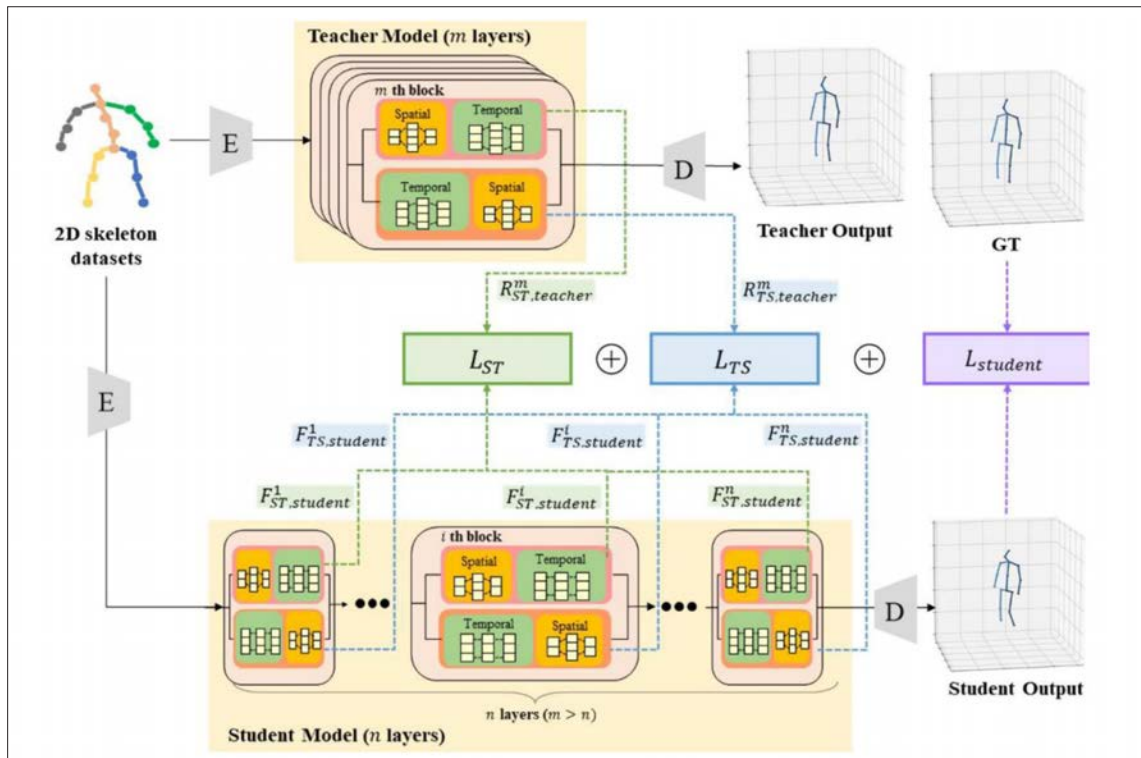
## 2. AI Model Compression

자세 추정 모델은 단안(monocular) 데이터나 다중 시점 2D 데이터를 활용해 인간의 활동을 탐지하지만, 높은 정

확도를 위해 많은 연산 자원과 전송/처리 비용이 요구된다. 특히 3차원 자세 추정용 딥러닝 모델은 일반적으로 성능과 크기가 비례하며(<그림 2> 참조), 과도한 파라미터는 과적합 문제를 일으킬 수 있다.



<그림 2> 3차원 자세 추정용 AI 모델들의 정확도 관련 성능(MPJPE) 대비 모델의 파라미터(크기) 비교 : 정확도가 가장 좋은 MotionBERT의 경우, 4천 2백만 개의 모델 파라미터가 필요함



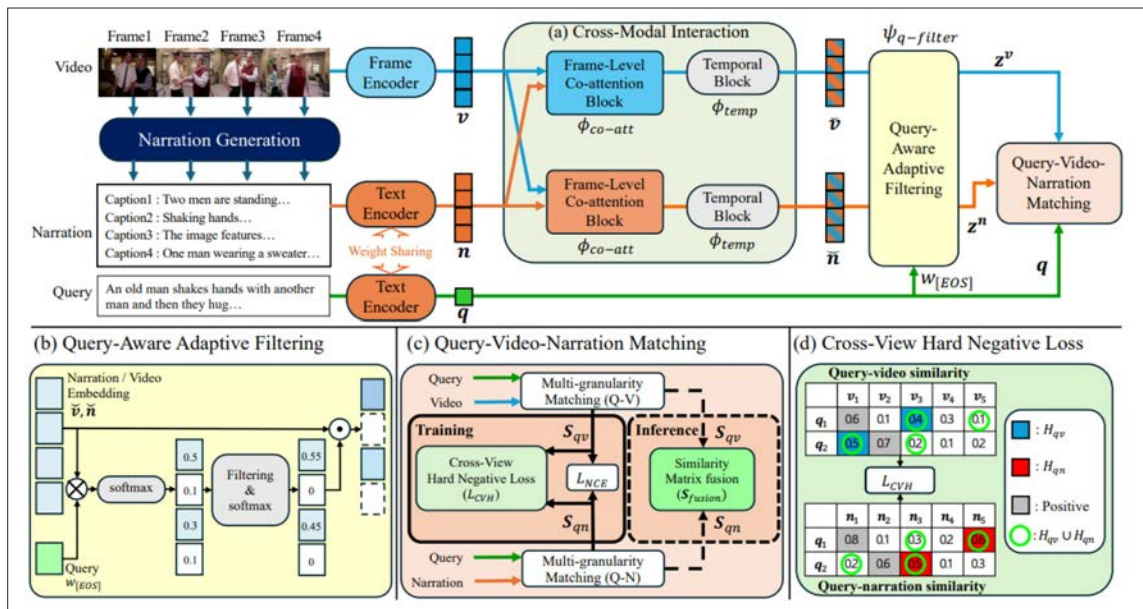
<그림 3> 듀얼 스트림 트랜스포머 기반 3차원 자세 추정을 위한 가지치기 유도 기반의 response-to-feature distillation 프레임워크

이에 본 연구실에서는 기존 3D 포즈 추정 모델 중 가장 복잡한 구조를 대상으로 가지치기(pruning)와 지식 증류(KD)를 결합한 경량화 전략을 제안(<그림 3> 참조)하여, 초기 최적의 학습 가능한 파라미터 수를 결정하고 모델 크기를 줄이면서도 실시간 반응성을 확보하고자 한다. 또한, 제안된 방법은 듀얼 스트림 트랜스포머 아키텍처에 적합하도록 수정된 ‘response-to-feature distillation(RFD)’ 기법을 도입하여 손실 함수를 업데이트하고, 미세 조정을 통해 성능 저하를 최소화하였다. Human3.6M 데이터셋에 대한 실험 결과, 제안된 방법은 모델 크기를 현저히 줄이면서도 MPJPE 기준으로 높은 정확도를 유지함을 입증하였으며, 본 연구의 주요 기여는 가지치기와 KD 결합 전략, 새로운 RFD 기법 도입, 그리고 모델 크기 감소와 높은 정확도의 균형 확보에 있다.

### 3. Multi-modal Learning: Video Retrieval

여러 비디오 스트리밍 플랫폼의 인기와 함께 텍스트-비디오 검색이 학계와 산업계에서 주목받고 있다. 텍스트와 비디오 간 이질적 데이터로 인한 다중모달리티(multi-modality) 차이와 비디오 내 시간 변화라는 두 가지 어려움이 존재한다. 이를 해결하기 위해 CLIP, BLIP 등의 비전-언어 사전 학습 모델과 트랜스포머 인코더, 3D 컨볼루션 등 시간적 모듈이 활용되고 있다. 또한, 기존의 비디오 수준 캡션 대신 프레임 단위 캡션을 사용해 영상의 다양한 객체와 속성을 포착하려는 시도도 진행 중이다.

이에 본 연구실에서는 프레임 단위 캡션을 전체 텍스트-비디오 검색 과정에서 종합적으로 활용하는 NarVid 프레임워크를 제안한다. NarVid 프레임워크는 프레임 단위 캡션(내레이션)을 다양한 방식으로 활용하여 검색 성능을 향상시킨다. 비디오 프레임과 생성된 캡션 간의 상호 주의를 통해 특징을 추출하고, 쿼리와 유사도를 기반으로 관련 없는 프레임과 부정확한 캡션을 필터링하여 정제한다. 이후 텍스트-비디오와 텍스트-내레이션 유사도를 융합해 매칭 과정을 개선하며,



<그림 4> 텍스트-동영상 검색을 위한 내레이션 기반의 NarVid 프레임워크

쿼리-비디오-내레이션 관계를 활용한 새로운 hard negative loss를 도입해 구별력 있는 특징 학습을 지원한다. 이러한 접근법은 다중모달 간의 차이를 효과적으로 줄여 최첨단 텍스트-비디오 검색 결과를 이끌어 낸다. 해당 연구는 2025년 6월에 열리는 시분야 최상위 학술대회인 CVPR에 Poster로 발표 예정이다[3].

### III. 연구 실적

- D. Kim, S. Lee, J. Bae, S. Cho, B. Bae, J. Lee, and **S. Park**, "Human-oriented Video Retargeting via Object Detection and Patch Decision," *Multimedia Tools and Applications*, Vol. 84, pp. 5565-5582, Mar. 2025.
- J. Y. Lee and **S. Park**, "Fast Depth Intra Mode Decision Using Intra Prediction Cost and Probability in 3D-HEVC," *Multimedia Tools and Applications*, Vol. 83, Oct. 2024.
- A. Kim, S. Woo, M. Park, D. Kim, H. Lim, S. Jung, S. Kwak and **S. Park**, "Deep learning-guided video compression for machine vision tasks," *EURASIP Journal on Image and Video Processing*, 2024, 32, Sep. 2024.
- D. Kim, D. Lee, A. Kim, J. Jeong, J. T. Lee, S. Kim, and **S. Park**, "Pruning-guided feature distillation for an efficient transformer-based pose estimation model," *IET Computer Vision*, vol. 18, Iss. 6, Sep. 2024.
- J. Bae and **S. Park**, "Revisiting video super-resolution: you only look outstanding frames," *Journal of Electronic Imaging* 32(2), 023012 (16 March 2023). <https://doi.org/10.1117/1.JEI.32.2.023012>
- J. Lee and **S. Park**, "Knowledge Distillation for Optical Flow-Based Video Super-resolution," *Journal of Computing Science and Engineering* 17, no. 1 (2023): 13-19.
- K. A. Bhosale and **S. Park**, "Feeding Longer Frames for Efficient Video Denoising Model," *Journal of Computing Science and Engineering* 16, no. 4 (2022): 185-193.
- J. Y. Lee and **S. Park**, "Adaptive Fractional Motion and Disparity Estimation Skipping in MV-HEVC," *Journal of Visual Communication and Image Representation*, vol. 79, 2021
- **S. Park** and J. Kang, "Fast Multi-type Tree Partitioning for Versatile Video Coding Using a Lightweight Neural Network," in *IEEE Transactions on Multimedia*, Nov. 2021. doi: 10.1109/TMM.2020.3042062
- **S. Park** and J. Kang, "Context-Based Ternary Tree Decision Method in Versatile Video Coding for Fast Intra Coding," in *IEEE Access*, vol. 7, pp. 172597-172605, 2019. doi: 10.1109/ACCESS.2019.2956196
- **S. Park** and J. Kang, "Fast Affine Motion Estimation for Versatile Video Coding (VVC) Encoding," in *IEEE Access*, vol. 7, pp. 158075-158084, 2019. doi: 10.1109/ACCESS.2019.2950388
- **S. Park**, "A sub-pixel motion estimation skipping method for fast HEVC encoding," *ICT Express*, Volume 5, Issue 2, 2019, Pages 136-140, ISSN 2405-9595, <https://doi.org/10.1016/j.ict.2018.08.003>.
- **S. Park** and E. S. Jang, "Comments on "Fast Motion Estimation Based on Content Property for Low-Complexity H.265/HEVC Encoder"," in *IEEE Transactions on Broadcasting*, vol. 63, no. 4, pp.

740-742, Dec. 2017. doi: 10.1109/TBC.2017.2711146

- **S. Park**, S. Lee, E. S. Jang, D. Jun, and J. -W. Kang “Efficient biprediction decision scheme for fast high efficiency video coding encoding,” *Journal of Electronic Imaging* 25(6), 063007 (25 November 2016). <https://doi.org/10.1117/1.JEI.25.6.063007>
- R. Wang, T. Huang, **S. Park**, J. -G. Kim, E. S. Jang, C. Reader, and W. Gao, “The MPEG Internet Video-Coding Standard [Standards in a Nutshell],” in *IEEE Signal Processing Magazine*, vol. 33, no. 5, pp. 164-172, Sept. 2016. doi: 10.1109/MSP.2016.2571440
- **S. Park**, K. Choi, E. S. Jang, “Zero coefficient-aware fast butterfly-based inverse discrete cosine transform algorithm,” *IET Image Processing*, 10, 2, pp. 89-100, 2016

## 참 고 문 헌

- [1] B. Bross, Y.-K. Wang, Y. Ye, S. Liu, J. Chen, G.J. Sullivan, J.-R. Ohm, Overview of the versatile video coding (VVC) standard and its applications. *IEEE Trans. Circuits Syst. Video Technol.* 31(10), 3736-3764 (2021)
- [2] S. Liu, C. Rosewarne, Working draft 2 of video coding for machines, document ISO/IEC JTC1/SC29/WG 04, N00465, (2024)
- [3] Hong JH, Lee DH, Kang D, Myeong S, Park SH, Park H, “Narrating the Video: Boosting Text-Video Retrieval via Comprehensive Utilization of Frame-Level Captions.” *arXiv preprint arXiv:2503.05186*, Mar.2025

## 저 자 소 개



### 박 상 호

- 2017년 : 한양대학교 컴퓨터소프트웨어학과 공학박사
- 2017년 ~ 2018년 : 한국전자기술연구원 지능형영상처리센터 Post-Doc
- 2018년 ~ 2018년 : 연세대학교 바른ICT연구소 연구원
- 2019년 ~ 2020년 : 이화여자대학교 전자전기공학전공 Post-Doc
- 2020년 ~ 현재 : 경북대학교 컴퓨터학부 부교수
- 주관심분야 : 동영상 압축, 모델 경량화, 멀티모달리티, 생성모델