

2026년 한국방송·미디어공학회 하계학술대회 후기

류은석 교수(조직위원장, 성균관대학교), 박상효 교수(프로그램위원장, 경북대학교), 방 건 책임(프로그램위원장, ETRI)



준비위원회

- 대회장 : 김규현 회장(한국방송·미디어공학회, 경희대학교)
- 조직위원장 : 류은석 교수(성균관대학교)
- 프로그램위원장 : 박상효 교수(경북대학교), 방 건 책임(ETRI)
- 프로그램위원

강정원 실장(ETRI)	강제원 교수(이화여자대학교)	고준호 수석(ITA)
공경보 교수(부산대학교)	김원준 교수(건국대학교)	김재광 교수(성균관대학교)
김진필 교수(성균관대학교)	라상중 책임(ETRI)	백승권 책임(ETRI)
서영우 단장(KBS)	서영호 교수(광운대학교)	서정일 교수(동아대학교)
양성빈 팀장(IITP)	유 성 소장(SBS)	이상민 교수(고려대학교)
이영한 센터장(KETI)	이윤재 팀장(KBS)	이장원 교수(성균관대학교)
이재협 교수(경북대학교)	이종민 교수(중앙대학교)	진강옥 팀장(RAPA)
정든솔 교수(경북대학교)	정병희 수석(KBS)	정종범 교수(국립군산대학교)
정진우 책임(KETI)	정현민 교수(서울과학기술대학교)	조병철 교수(동아방송예술대학교)
진경환 교수(고려대학교)	최기호 교수(경희대학교)	최종원 교수(중앙대학교)
최해철 교수(국립한밭대학교)	추현곤 실장(ETRI)	한성희 부장(KBS)
홍기훈 실장(ETRI)	임영권 수석(삼성전자)	



하계학술대회 프로그램 구성 및 내용

2026년도 한국방송·미디어공학학회 하계학술대회는 6월 18일 목요일부터 6월 20일 토요일까지 제주특별자치도 서귀포시 롯데호텔 제주에서 개최되었다.

최근 방송미디어 및 관련 학문 분야는 생성형 인공지능, 몰입형 미디어, 3차원 및 4차원 콘텐츠, 초실감 영상, 지능형 방송 서비스 등 새로운 기술 패러다임을 중심으로 빠르게 확장되고 있다. 이러한 변화 속에서 연구자와 산업계 전문가들이 최신 성과를 공유하고, 미래 미디어 기술의 발전 방향을 함께 논의하는 일은 그 어느 때보다 중요해지고 있다. 따라서, 올해의 하계학술대회는 방송미디어, 영상처리, 인공지능, 실감미디어, 차세대 콘텐츠 및 미디어 융합 기술 분야의 최신 연구 성과를 공유하고, 산·학·연 전문가들이 함께 미래 기술의 방향을 논의하는 뜻깊은 교류의 장으로 준비되었다.

올해 학술대회에서는 IITP, TTA, RAPA 등의 기관과 ETRI, KETI, 삼성, KBS 및 학회 내의 연구회들의 총 26개의 특별세션이 구성되어, 빠르게 변화하는 방송미디어 기술의 다양한 연구 주제와 산업적 이슈를 폭넓게 다루었다. 또한 각종 기업이 참여하는 9건의 취·창업 세션을 통해 회사 소개, 기술 소개, 채용 및 취업 안내가 함께 이루어져, 학생과 신진 연구자들에게 실질적인 진로 탐색과 네트워킹의 기회를 제공하였다. 더불어 6명의 신진연구자 발표세션을 통해 방송미디어 분야의 새로운 연구 역량과 참신한 시각을 공유하는 자리도 마련하였다.

무엇보다 올해의 학술대회에서 가장 주목할 프로그램은 최우수 국제학술대회 논문 발표 세션이다. 세계적 수준의 국제학술대회에서 우수한 성과를 거둔 19건의 최우수 논문을 해당 연구를 이끈 교수들이 직접 발표하는 자리로, 참여 연구자들이 최신 국제 연구 흐름과 핵심 기술 동향을 깊이 있게 접할 수 있는 매우 의미 있는 세션이었다.

초청 강연 및 특별 프로그램은 학술대회의 양적·질적 수준을 한층 높였다. 초청 강연에서는 조병철 교수(동아방송예술대)가 좌장을 맡은 가운데 두 건의 발표가 진행되었다. KBS 김민중 본부장이 ‘KBS 방송기술분야 비전 및 로드맵’을 통해 미디어 환경 변화에 따른 공영방송의 기술적 지향점을 제시하였고, Fraunhofer HHI의 Corneliuss Hellge 박사가 ‘Gaussian Splat Unleashed: Compression for Real-World Deployment’를 주제로 실세계 적용을 위한 3DGS 압축 기술의 최신 동향을 발표하였다.

국제 최우수컨퍼런스 논문 PI 초청 발표 세션은 세계 최고 권위의 학회에서 성과를 거둔 우수 논문 약 20건을 선정하여 총 5개 세션으로 확대 운영되었다. 전병우 교수(성균관대), 김규현 교수(경희대), 이영렬 교수(세종대), 박종일 교수(한양대), 김문철 교수(KAIST)가 각 세션의 좌장을 맡았으며, 김은우 교수

(중앙대), 김정옥 교수(경희대), 김학구 교수(중앙대), 박은병 교수(연세대), 김재광 교수(성균관대) 등 국내외 주요 연구팀이 참여하여 글로벌 탐티어 연구 결과를 공유하였다.

신진연구자 세션은 이장원 교수(성균관대)의 진행 아래 최근 학계에 부임한 젊은 연구자 6명을 초청하여 Physical AI, 역렌더링, 3D 파운데이션 모델 등 차세대 핵심 미디어 연구 계획을 소개하고 산·학·연 협력의 기반을 다졌으며, 본 세션에는 정든솔 교수(경북대), 이경호 교수(성균관대), 이종민 교수(중앙대), 유재민 교수(서울대), 박준형 교수(한국외대), 이재홍 교수(한국외대)가 참여하였다.

특별 세션은 총 26개 부문으로 세분화되어 특정 목적 지향성 연구와 표준화 동향을 깊이 있게 조망하였다. 주요 세션으로는 송실대 이정진 교수님과 KBS가 참여한 ‘차세대통신 환경과 AI 기반 미디어 제작 활용’, ETRI 백승권 책임이 좌장을 맡고 연세대와 성균관대 등이 발표한 ‘오디오 신호처리 및 응용’, ETRI 배병준 책임 중심의 ‘AI 기반 UI/UX 접근성 기술’, 그리고 국립한밭대와 동아대 등이 ETRI와 공동 연구 성과를 도출한 ‘기계를 위한 미디어 부호화 기술(1, 2)’ 세션이 밀도 있게 운영되었다. 또한, ‘KBS 특별세션: 방송현장, AI와 UHD를 입다’를 통해 방송 실무에 적용된 생성형 AI 파이프라인이 대거 공개되었으며, 건국대, 중앙대, 한양대, 연세대, 성균관대, 한국항공대 등이 참여한 ‘GS 기반 이머시브 미디어’ 및 ‘3DGS 기술과 표준화 전망(1, 2)’ 세션을 통해 최신 공간 미디어 기술의 압축 및 렌더링 성능 최적화 방안이 집중 논의되었다. 이 외에도 ‘우주계층 지능통신망 특화연구실(1, 2)’과 ‘2D 비디오 코덱(1, 2)’ 세션을 통해 국제 R&D 과제와 세부 지표와 코덱 효율 향상 기법이 상세히 공유되었다.

일반 세션은 구두 및 포스터 발표를 통해 풍성한 학술적 검증을 거쳤다. 9개의 구두 세션은 유재민 교수(서울대), 정든솔 교수(경북대), 공경보 교수(부산대), 현명한 교수(경상국립대), 서영호 교수(광운대), 이종민 교수(중앙대), 김규현 교수(경희대), 이재협 교수(경북대), 조병철 교수(동아방송예술대)가 각각 좌장을 맡아 진행하였다. ‘인공지능 기반 신호처리’ 부문이 센서/전송, 신뢰성/탐지, 멀티모달, 3D, 시각지능 등 5개 분과로 심화 운영되었으며, 명지대, 경희대, 성균관대, 경북대, 국민대 등의 연구팀이 최신 인공지능 방법론을 발표하였다. 또한 ‘방송·미디어 영상처리 및 압축(2D/3D)’, ‘미디어 콘텐츠 생성’, ‘방송·미디어 서비스 및 플랫폼’ 세션에서는 세종대, 광운대, 한양대, 국립공주대 등의 다채로운 영상 압축 및 시스템 최적화 연구가 발제되었다. 2개의 포스터 세션 및 캡스톤디자인 경진대회(포스터) 세션은 정대권 교수(한국항공대)와 최해철 교수(국립한밭대)의 심사 하에 진행되었으며, KAIST, 동아대, 서울과학기술대, 경상국립대, 한양대, 한국항공대 등에서 제출한 총 80편의 연구 성과물이 열린 토론 속에 전시되었다.

학문 후속 세대의 성장을 도모하기 위한 대학생 논문 경진대회는 총 3개의 구두 세션으로 편성되었으며, 이상민 교수(고려대)와 정종범 교수(국립군산대) 등이 좌장을 맡아 학부생 연구원들의 참신한 시각과 실증적 연구 결과물들을 객관적으로 평가하였다. 마지막으로 총 2회에 걸쳐 운영된 산업체/취창업 세션 및 전문 기술 소개 세션에서는 정현민 교수(서울과학기술대)와 정진우 책임(KETI)이 좌장을 맡았다. 본



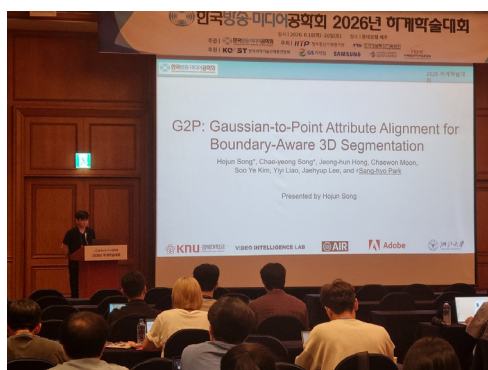
세션에는 (주)픽스트리(김연배 대표), SK텔레콤(전영규, 주현철), 스튜디오메타케이(박정수 대표), KBS(서영우), (주)빅태블릿(문지호 대표), MBC(김성곤), (주)퍼셉트(이은규), 스펙스페이스(이아휘), (주)와이매틱스(방준성) 등 국내 미디어 산업을 선도하는 기관들이 대거 참여하였다. 참여 기관들은 각 기업이 보유한 AI 미디어 실무 파이프라인과 기술 연구 동향을 소개하고 현업에서 요구되는 역량 및 채용 정보를 구체적으로 공유하여, 전문 연구자를 지망하는 대학원생 및 청년 연구자로부터 뜨거운 호응을 얻었다.

일반 세션

일반 세션에서는 총 109편의 논문이 발표되었는데, 4개의 구두 세션인 ‘인공지능 기반 신호처리’, ‘방송미디어 영상처리 및 압축’, ‘미디어 콘텐츠 생성’, ‘방송미디어 서비스 및 플랫폼’에서 41편의 논문과 2개의 포스터 세션에서 68편의 논문이 발표되었다.

‘인공지능 기반 신호처리’ 세션에서는 23편의 논문이 구두 발표되었는데, 성균관대학교(6편), 명지대학교(3편), 경북대학교(3편), 경희대학교(2편), 경북대학교/Adobe Research(2편), 국민대학교(1편), 국립아시아문화전당재단/SK텔레콤(1편), 경북대학교/KETI/Adobe Research/저장대학교(1편), 홍익대학교(1편), 한양대학교(1편)에서 발표하였다. ‘방송미디어 영상처리 및 압축’ 세션에서는 성균관대학교(4편), 세종대학교/ETRI(3편), 광운대학교(1편), 성균관대학교/ETRI(1편)에서 논문을 발표하였다. ‘미디어 콘텐츠 생성’ 세션에서는 ETRI/과학기술연합대학원대학교(1편), KETI(1편), 한양대학교(1편), 광운대학교/ETRI(1편), 광운대학교(1편)에서 발표하였다. ‘방송미디어 서비스 및 플랫폼’ 세션에서는 명지대학교(1편), 경북대학교(1편), 성균관대학교(1편), 국립공주대학교/서울과학기술대학교(1편)에서 발표하였다.

포스터 세션 1에서는 총 38편의 논문이 발표되었는데, KAIST(11편), 동아대학교(3편), 한국항공대학교(3편), 경상국립대학교(3편), ETRI/성균관대학교(2편), SK텔레콤(2편), 성균관대학교(2편), ETRI(1편), 고려대학교(1편), 서울과학기술대학교(1편), 한양대학교(1편), 서울대학교(1편), KAIST/스텔라비전(1편), KAIST/중앙대학교(1편), 건양대학교(1편), KETI(1편), 광운대학교(1편), 경북대학교/KETI/Adobe Research(1편), 서울연구원(1편)에서 발표하였다. 포스터 세션 2에서는 총 30편의 논문이 발표되었는데, 성균관대학교(6편), 중앙대학교(4편), (주)와이메틱스(2편), 광운대학교(2편), 국립한국해양대학교(2편), 서울연구원(2편), 한양대학교(2편), 홍익대학교(2편), (주)피아이이(1편), KAIST(1편), SK텔레콤(1편), 경북대학교/KETI(1편), 세종대학교(1편), 엘이지 디펜스앤에어로스페이스(1편), 조선대학교/한국에너지공과대학교(1편), 홍익대학교/주PAL문화유산센터(1편)에서 발표하였다.



국제 최우수컨퍼런스 논문 PI 초청 발표

멀티미디어 분야의 국제 최고 수준 연구 성과를 공유하기 위해, 최근 top-tier 국제학술대회에 발표된 국내 우수 논문을 초청하여 연구 내용과 주요 결과를 소개하는 자리를 마련하였다. 본 세션은 연구자 간 최신 기술 동향과 성과를 공유하고, 후속 연구에 대한 관심과 연구 의욕을 높이기 위한 목적으로 운영하였으며, 발표자 및 내용은 아래와 같다.

[CVPR 2026] Stake the Points: Structure-Faithful Instance Unlearning

/ 홍기성, 신정규, 김은우 (중앙대학교)

머신러닝은 지정된 학습 데이터의 영향을 제거하면서도 나머지 데이터에서 얻은 지식을 유지해야 하지만, 기존 방법은 삭제 과정에서 보존 대상 샘플 사이의 의미적 관계까지 무너뜨리는 구조적 붕괴를 일으킬 수 있다. 본 연구는 언어로 기술한 속성 정보를 CLIP과 같은 의미 인코더로 변환해 ‘stake’라 부르는 의미적 기준점을 만들고, 언러닝 전후의 지식 구조가 이 기준점을 중심으로 유지되도록 설계하였다. 구조 인식 정렬은 보존 데이터의 상대적 배치를 맞추고, 정규화 항은 구조 보존에 중요한 파라미터가 과도하게 변하는 것을 제한한다. 이미지 분류·검색·얼굴 인식 실험에서 삭제 성능과 보존 성능의 균형을 개선했으며, 보지 못한 데이터에 대한 일반화도 함께 향상시켰다.

[CVPR 2026] Learning from Oblivion: Predicting Knowledge-Overflowed Weights via Retrodiction of Forgetting

/ 장진혁, 김재홍, *김정욱 (ETRI, *경희대학교)

이 연구는 더 많은 데이터를 수집하지 않고도 기존 데이터보다 풍부한 지식을 담은 사전학습 가중치를 만들 수 있는지를 다룬다. 저자들은 학습 데이터를 단계적으로 줄이며 순차 미세조정하면 가중치가 무작위로 망가지는 것이 아니라 일정한 방향의 ‘구조적 망각 궤적’을 보인다는 점에 주목하였다. 이를 바탕으로 데이터 규모가 줄어드는 과정에서 발생한 가중치 전이들을 학습 자료로 구성하고, 가중치의 변화 법칙을 배우는 메타 모델 KNOWN을 학습한다. 추론 시에는 이 망각 궤적을 역방향으로 예측하여 더 큰 데이터로 학습한 것과 유사한 KNOW 가중치를 생성한다. 다양한 데이터셋과 네트워크에서 단순 미세조정 및 일반적인 가중치 예측보다 높은 전이 성능과 일반화 성능을 보였다.

[ICCV 2025] MemoryTalker: Personalized Speech-Driven 3D Facial Animation via Audio-Guided Stylization

/ 김형규, *이상민, 김학구 (중앙대학교, *고려대학교)

음성 기반 3차원 얼굴 애니메이션은 입 모양뿐 아니라 화자마다 다른 말하기 습관과 표정 스타일을 재현해야 한다. 기존 방법은 화자 ID, 스타일 라벨 또는 추가 얼굴 메시를 추론 단계에서 요구해 실제 활용성이 제한되었다. Memo-

ryTalker는 먼저 여러 화자의 일반적인 얼굴 움직임을 모션 메모리에 저장하고 검색하는 Memorizing 단계와, 입력 음성에서 추출한 말하기 스타일 특징으로 검색된 모션을 개인화하는 Animating 단계로 구성된다. 모델은 특정 음성에서 어떤 얼굴 움직임을 더 강조해야 하는지를 학습하므로, 추론 시에는 음성만으로 화자 특유의 움직임을 생성할 수 있다. 정량 평가와 사용자 평가에서 기존 방법보다 자연스럽게 개인화된 얼굴 애니메이션을 생성함을 확인하였다.

[CVPR 2026] Multi-view Pyramid Transformer: Look Coarser to See Broader

/ 강경진, *양승권, *남승태, 이영근, *김정우, *박은병 (성균관대학교, *연세대학교)

다수의 입력 영상으로 큰 3차원 장면을 한 번의 순전파로 복원하려면, 모든 뷰와 모든 고해상도 토큰을 직접 연결할 때 발생하는 막대한 계산량을 줄여야 한다. MVP는 뷰 사이에서는 가까운 뷰에서 시작해 뷰 그룹과 전체 장면으로 범위를 넓히는 local-to-global 계층을 사용하고, 각 뷰 내부에서는 세밀한 표현을 점차 압축된 토큰으로 바꾸는 fine-to-coarse 계층을 사용한다. 두 계층을 함께 적용하면 세부 구조를 유지하면서도 장면 전체의 관계를 효율적으로 통합할 수 있다. 3D Gaussian Splatting을 출력 표현으로 결합해 수십 장부터 수백 장의 영상까지 확장 가능하며, 다양한 뷰 수와 대규모 장면에서 높은 일반화 복원 품질과 빠른 처리 성능을 보였다.

[AAAI 2026] Low-Rank Curvature for Zeroth-Order Optimization in LLM Fine-Tuning

/ 승현석, *이재우, **고현석 (위스콘신대학교, *조지아대학교, **한양대학교 ERICA)

역전파 없이 함수값의 변화만으로 기울기를 추정하는 zeroth-order 최적화는 메모리가 제한된 환경에서 대규모 언어 모델을 미세조정할 수 있지만, 무작위 등방성 섭동을 사용하면 추정 분산이 크고 실제 곡률과 맞지 않는 방향을 탐색하기 쉽다. LOREN은 기울기 사전조건화를 ‘어떤 분포에서 섭동을 뽑을 것인가’라는 문제로 다시 정의하고, 자연 진화 전략을 이용해 저랭크 블록 대각 곡률 정보를 반영한 비등방성 섭동 분포를 학습한다. 또한 REINFORCE leave-one-out 추정기를 적용해 분산을 줄인다. 여러 LLM 벤치마크에서 기존 zeroth-order 방법보다 빠르게 수렴하고 높은 정확도를 달성했으며, MeZO-Adam 대비 최대 27.3%의 피크 메모리를 절감하였다.

[WWW 2026] FCRLLM: Aligning LLM with Collaborative Filtering for Long-tailed Sequential Recommendation

/ 허병문, 이남준, 김선아, 김재광 (성균관대학교)

순차 추천에서는 상호작용이 많은 인기 사용자-아이템에 학습 신호가 집중되어, 기록이 짧은 사용자와 비인기 아이템의 추천 성능이 크게 떨어지는 long-tail 문제가 발생한다. FCRLLM은 협업 필터링이 학습한 행동 기반 선호 정보와 LLM이 제공하는 풍부한 의미 정보를 결합하는 Flipped Classroom with LLM 구조를 제안한다. 단순히 두 표현을 동일하게 맞추는 대신, 협업 신호와 언어 의미가 서로 부족한 정보를 보완하도록 정렬하여 희소한 사용자-아이템 표현을 강화한다. 사용자의 시퀀스 길이와 아이템 인기도를 기준으로 여러 그룹을 나누어 평가했으며, 세 개의 실제 데이터셋에서 인기도 구간과 관계없이 순차 추천 성능을 안정적으로 개선하였다.

[NeurIPS 2025] Disentangled Action and Context Concept-Based Explainable Video Action Recognition

/ Jongseo Lee, Wooil Lee, *Gyeong-Moon Park, Seong Tae Kim, Jinwoo Choi (경희대학교, *고려대학교)

영상 행동 인식 모델의 설명은 사람이 어떻게 움직였는지와 주변의 물체·장면이 무엇인지를 구분해 보여줘야 하지만, 기존 saliency 기반 설명은 두 요인을 뒤섞고 언어 설명은 미세한 동작을 정확히 표현하기 어렵다. DANCE는 행동을 모션 다이내믹스, 물체, 장면이라는 세 종류의 개념으로 분해하는 ante-hoc 개념 병목 구조를 사용한다. 모션 개념은 사람의 포즈 시퀀스로 표현하고, 물체와 장면 개념은 LLM을 이용해 자동 구축한 뒤 모델이 반드시 이 개념들을 거쳐 행동을 예측하도록 한다. 네 개의 행동 데이터셋에서 경쟁력 있는 인식 성능을 유지하면서 설명의 명확성을 향상시켰고, 사용자 평가와 모델 디버깅·편집·실패 분석에서도 유용성을 보였다.

[CVPR 2025] SocialGesture: Delving into Multi-person Gesture Understanding

/ Xu Cao, Pranav Virupaksha, Wenqi Jia, *Bolin Lai, *Fiona Ryan, **Sangmin Lee, James M. Rehg (UIUC, *Georgia Tech, **고려대학교)

기존 제스처 연구는 한 사람의 동작을 중심으로 구성되어 실제 대화에서 여러 사람이 서로 반응하며 사용하는 사회적 제스처를 충분히 다루지 못했다. SocialGesture는 자연스러운 다인 상호작용 장면을 대규모로 수집하고, 제스처 존재 여부 판별, 제스처 유형 분류, 시간 구간 탐지 등 여러 과제를 지원하는 데이터셋을 제안한다. 또한 영상 속 사람들의 제스처와 사회적 맥락을 질문·응답 형태로 이해하는 VQA 과제를 도입해 비전-언어모델의 사회적 제스처 이해 능력을 평가한다. 실험 결과 기존 제스처 인식 모델과 VLM이 인물 간 관계, 시간적 경계, 맥락 의존적 의미를 처리하는데 여전히 뚜렷한 한계가 있음을 보여주었다.

[NeurIPS 2025] Adaptive 3D Reconstruction via Diffusion Priors and Forward Curvature-Matching Likelihood Updates

/ 신승혁, 김다빈, 임흥기 (인하대학교)

확산 사전분포와 관측 우도를 결합하는 3차원 복원 방법은 별도 재학습 없이 다양한 측정 조건을 처리할 수 있지만, 우도 갱신의 스텝 크기를 고정하면 수렴이 느리고 복원 품질도 불안정해진다. 본 연구는 확산 샘플링 과정에 Forward Curvature-Matching(FCM) 갱신을 삽입해 현재 관측 오차의 곡률에 맞는 스텝 크기를 매 반복마다 자동 결정한다. 역전파 전체를 계산하지 않고 forward-mode 자동미분과 유한차분 곡률 추정만 사용하므로 효율적으로 최적 스텝을 찾을 수 있다. 동일한 사전 모델에서 관측 연산자만 바꾸어 단일·다중 뷰 및 다양한 입력 모달리티를 처리하며, ShapeNet과 CO3D에서 더 적거나 비슷한 함수 평가 횟수로 F-score를 높이고 CD·EMD를 낮췄다.

[CVPR 2026] Diffusion-Based sRGB Real Noise Generation via Prompt-Driven Noise Representation Learning

/ 고재균, 김동진, 이수민, *Guanghui Wang, 김태현 (한양대학교, *Toronto Metropolitan University)

sRGB 영상의 실제 노이즈는 카메라·촬영 조건·후처리 파이프라인에 따라 복잡하게 달라지지만, 실제 noisy-clean 쌍

은 수집 비용이 높고 기존 생성 방법은 카메라 메타데이터에 크게 의존한다. Prompt-Driven Noise Generation(PNG)은 소량의 실제 노이즈 영상에서 고차원 노이즈 프롬프트를 추출하고, 확산 모델이 이 프롬프트가 나타내는 분포를 따르는 다양한 노이즈를 생성하도록 학습한다. 따라서 센서 모델이나 ISO 같은 명시적 메타데이터 없이도 입력 노이즈의 특성을 재현할 수 있다. 생성된 영상은 시각적·통계적으로 실제 노이즈와 유사했으며, 이를 학습 데이터로 사용한 복원 모델이 여러 실제 노이즈 제거 벤치마크에서 성능을 향상시켰다.

[CVPR 2026] ZOO-Prune: Training-Free Token Pruning via Zeroth-Order Gradient Estimation in Vision-Language Models

/ 김영은, *Youjia Zhang, *Huiling Liu, *정애천, **이선우, *홍성은 (Amazon, *성균관대학교, **인하대학교)

대규모 비전-언어모델은 한 장의 영상에서 수백 개의 시각 토큰을 생성하지만 실제 추론에 필요한 토큰은 일부에 불과하다. 기존 학습 없는 프루닝은 attention 점수의 불안정성이나 특징 다양성만을 기준으로 삼아 중요한 영역을 제거할 수 있다. ZOO-Prune은 가벼운 projection layer에서 각 토큰에 작은 무작위 섭동을 가하고 출력 특징이 얼마나 변하는지를 zeroth-order 방식으로 측정해 토큰 민감도를 계산한다. 이후 민감도와 특징 다양성을 함께 고려해 영향력이 크고 중복이 적은 토큰을 선택한다. 역전파나 추가 학습 없이 여러 VLM에 적용 가능하며, 최대 94.4%의 토큰을 제거하면서 정확도를 유지하고 기준 모델 대비 최대 2.30배의 end-to-end 추론 가속을 달성하였다.

[CVPR 2026] Mark4D: Temporally-Consistent Watermarking for 4D Gaussian Splatting

/ Jaejin Lee, Minjae Jeong, Joonhyuk Park, Yechan Hwang, Seunghun Baek, Won Hwa Kim (POSTECH)

Dynamic 4D Gaussian Splatting (4DGS) 모델에 눈에 보이지 않고 시간적으로 일관된 watermark를 임베딩하는 것은 Gaussian의 연속적인 spatio-temporal deformation과 다양한 motion dynamics로 인해 독특한 어려움을 제기한다. Gaussian splat 내의 파라미터를 직접 fine-tuning하는 기존의 3DGS watermarking 방법들은 동적인 4D 환경으로 확장될 때 geometric fidelity와 temporal coherence를 유지하지 못한다. 이러한 문제를 해결하기 위해, 본 논문에서는 4DGS를 위한 강건하고 감지할 수 없으며 motion-aware한 watermark 임베딩을 달성하는 시간적으로 일관된 watermarking 기술인 Mark4D를 제안한다.

[ICLR 2026] MoE-GS: Mixture of Experts for Dynamic Gaussian Splatting

/ 진인환, 문형주, *김준수, *윤국진, 공경보 (부산대학교, *ETRI)

최근 3D Gaussian Splatting 기반의 dynamic scene reconstruction은 모든 동적인 문제를 단일 접근법으로 해결하지 못하는 한계를 보인다. 이를 극복하기 위해 본 논문은 복수의 특화된 전문가를 Volume-aware Pixel Router로 통합하는 framework인 MoE-GS를 제안한다. LLM의 회소성 지향적 MoE와 달리, MoE-GS는 FLOPs 감소가 아닌 서로 다른 deformation priors를 결합하여 dynamic novel view synthesis 퀄리티를 향상시키도록 설계되었다. 제안하는 router는 미분 가능한 weight splatting을 통해 Gaussian 단위의 가중치를 픽셀 공간으로 투영함으로써 시공간적으로 일관된 렌더링을 보장한다.

[ICLR 2026] FIRE: Frobenius-Isometry Reinitialization for Balancing the Stability-Plasticity Tradeoff

/ 한이삭, 박상연, 오승원, ***김동후, ***이호준, 김경중 (GIST, *KAIST, **Holiday Robotics)

비정상 데이터로 학습하는 심층 신경망은 안정성과 가소성의 균형을 맞추어야 한다. 기존의 가중치 재초기화 방법들은 파라미터 튜닝이 어려워 가소성 회복에 실패하거나 기존 지식을 망각하는 문제를 겪는다. 이를 해결하기 위해 본 논문은 두 요소의 트레이드오프를 명시적으로 조절하는 원리적인 재초기화 방법인 FIRE를 제안한다. FIRE는 과거 가중치와의 근접성을 측정하는 Squared Frobenius Error (SFE)로 안정성을 수량화하고, 가중치의 등방성을 반영하는 Deviation from Isometry로 가소성을 수량화한다. 재초기화 지점은 DfI를 0으로 유지하는 제약 조건 하에 SFE를 최소화하는 최적화 문제를 통해 도출되며, 이는 Newton-Schulz iteration으로 효율적으로 근사된다.

[NeurIPS 2025] Temporal Smoothness-Aware Rate-Distortion Optimized 4D Gaussian Splatting

/ 이형민, *백경준 (서울과학기술대학교, *세종대학교)

Dynamic 4D Gaussian Splatting (4DGS)은 대용량 Gaussian 수와 시간적 중복성, 엔트로피 인지 압축 프레임워크의 부제로 인해 대규모 저장 공간을 요구하며, 이는 실제 배포와 전송에 심각한 제약을 준다. 이를 해결하기 위해 본 논문은 Ex4DGS를 베이스라인으로 삼아 유연하고 높은 fidelity의 렌더링을 가능하게 하는 end-to-end Rate-Distortion 최적화 압축 프레임워크를 제안한다. 기존 3DGS 압축 방법을 기반으로 하되, 시공간적 축으로 인해 발생하는 추가적인 문제를 해결하는 데 집중한다. 특히 모션 궤적을 독립적으로 저장하는 대신, 실제 환경의 smoothness 특성을 반영할 수 있는 wavelet transform을 적용하여 저장 효율성을 극대화하고 압축 효율과 렌더링 퀄리티 간의 사용자 제어 균형을 제공한다.

[CVPR 2026] MoRel: Long-Range Flicker-Free 4D Motion Modeling via Anchor Relay-based Bidirectional Blending with Hierarchical Densification

/ Sangwoon Kwak, *Weeyoung Kwon, Jun Young Jeong, *Geonho Kim, Won-Sik Cheong, *Jihyong Oh (ETRI, *중앙대학교)

수천 프레임의 장시간 동적 장면을 4DGS로 한 번에 학습하면 메모리가 계속 증가하고, 영상을 여러 구간으로 나누면 구간 경계에서 외형 변화와 flicker가 발생한다. MoRel은 전체 장면의 기준이 되는 Global Canonical Anchor를 만든 뒤 시간축의 key frame마다 지역 canonical anchor를 배치해 장면을 릴레이 방식으로 연결한다. 이웃한 anchor 사이의 정방향·역방향 변형을 함께 학습하고, 학습 가능한 opacity로 두 결과를 혼합해 구간 경계의 불연속과 가려졌다 나타나는 영역을 안정적으로 처리한다. Feature-variance-guided Hierarchical Densification은 복잡한 고주파 영역에 표현 용량을 집중한다. 새로 구축한 SelfCapLR과 기존 데이터에서 제한된 메모리로 장시간 동작을 재구성하고 시간적 깜빡임을 줄였다.

[ICCV 2025] AJAHR: Amputated Joint Aware 3D Human Mesh Recovery

/ 조현진, 최기윤, 최종원 (중앙대학교)

기존 3차원 인체 메시 복원 모델은 모든 관절과 팔다리가 존재하는 표준 인체 구조를 가정하기 때문에, 절단 장애가 있는 사람에게 적용하면 존재하지 않는 관절을 강제로 추정하거나 잘못된 메시지를 생성할 수 있다. AJAHR는 입력 영상에서 절단된 신체 부위를 판별하는 amputation classifier를 메시 복원 네트워크와 공동 학습하고, 분류 결과에 따라 포즈와 메시 추정을 적응적으로 조정한다. 또한 실제 학습 자료 부족을 해결하기 위해 다양한 절단 위치와 자세를 포함하는 합성 Amputee 3D(A3D) 데이터셋을 구축하였다. 비절단 인체에서는 기존 모델과 경쟁력 있는 성능을 유지하면서, 절단 인체에서는 기존 방법보다 정확한 관절·형상 복원을 달성하였다.

[ICCV 2025, ICLR 2026] SummDiff: Generative Modeling of Video Summarization with Diffusion

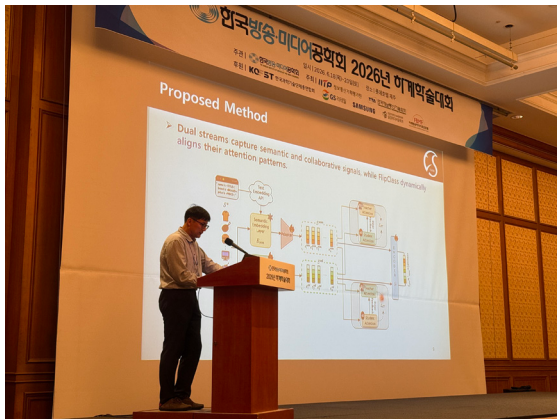
/ Kwanseok Kim, *Jaehoon Hahm, Sumin Kim, **Jinhwan Sul, ***Byunghak Kim, Joonseok Lee (서울대학교, *UIUC, **Georgia Tech, ***Hyundai Card)

비디오 요약은 같은 영상이라도 평가자마다 중요하다고 판단하는 장면이 달라 본질적으로 여러 정답이 존재한다. 그러나 기존 방법은 여러 사람의 프레임 점수를 평균한 하나의 값으로 회귀하여 이러한 주관성을 제거한다. SummDiff는 영상 요약을 입력 영상에 조건화된 생성 문제로 재정의하고, 확산 모델이 ‘좋은 요약’의 분포를 학습해 서로 다른 관점을 반영한 여러 후보 요약을 생성하도록 한다. 영상의 시각적 맥락에 따라 프레임 선택 분포를 동적으로 조정하며, 기존 벤치마크에서 높은 요약 성능과 개별 평가자 선호에 가까운 결과를 보였다. 또한 최종 길이 제한을 적용하는 knapsack 단계의 영향을 분석하고 이를 평가하기 위한 새로운 지표를 제안하였다.

[ICCV 2025] PAN-Crafter: Learning Modality-Consistent Alignment for PAN-Sharpening

/ 도정혁, 김성표, 육근혁, *이재협, 김문철 (KAIST, *경북대학교)

PAN-sharpening은 고해상도 흑백 PAN 영상의 공간 세부와 저해상도 다중분광 MS 영상의 색 정보를 결합하지만, 센서 위치와 촬영 시점·해상도 차이로 두 영상이 픽셀 단위로 어긋나는 경우가 많다. PAN-Crafter의 Modality-Adaptive Reconstruction은 하나의 네트워크가 고해상도 다중분광 영상과 PAN 영상을 함께 복원하도록 하여 PAN의 고주파 구조를 보조 자기지도 신호로 사용한다. Cross-Modality Alignment-Aware Attention(CM3A)은 MS의 텍스처를 PAN 구조에, PAN의 구조를 MS 표현에 양방향으로 맞추며 특징을 보정한다. 여러 위성 데이터셋에서 최신 방법보다 모든 지표에서 우수했고, 50.11배 빠른 추론과 0.63배 메모리로 높은 효율을 보였으며 보지 못한 위성 데이터에도 잘 일반화되었다.



신진연구자 세션

최근 대학으로 부임한 신진 연구자인 경북대학교 정든솔 교수, 성균관대학교 이경호 교수, 중앙대학교 이종민 교수, 서울대학교 유재민 교수, 한국외국어대학교 박준형 교수와 이재홍 교수로부터 연구자 소개, 연구 내용 및 향후 가능한 협업 연구 등에 대한 소개가 있었다. 젊은 연구자로서 한국방송·미디어공학회에서 활발한 학술 활동을 펼칠 수 있도록 학회 구성원들의 지속적인 관심과 지원, 협력이 필요하다.



신진 연구자	소속	학과	연구 분야	연구실 홈페이지
정든솔	경북대학교	컴퓨터학부	장면 그래프 생성 및 인간-객체 상호작용, 신경-상징 학습 및 AI, 다양한 시각 인공지능 관련 연구	https://cv-knu.github.io
이경호	성균관대학교	인공지능융합학과	컴퓨터 그래픽스, 캐릭터 애니메이션	https://sites.google.com/view/moai-lab
이종민	중앙대학교	소프트웨어학부	컴퓨터비전, 기계학습, 시각 기하학	https://bluedream1121.github.io/spatial-intelligence-lab/
유재민	서울대학교	컴퓨터공학부	데이터 중심 인공지능, 테이블 파운데이션 모델, 그래프 신경망, 시계열 분석 및 예측, 추천 시스템, 이상 탐지	https://da.snu.ac.kr
박준형	한국외국어대학교	Language & AI 융합학부	자연어처리, 언어모델, 효율적 인공지능	https://llab.hufs.ac.kr/
이재홍	한국외국어대학교	Language & AI 융합학부	신호처리, 기계학습	https://j-pong.github.io

대학생 논문 경진대회

학부생들의 연구 참여를 활성화하고, 연구 성과 발표 경험이 향후 대학원 진학과 심화 연구로 이어질 수 있는 계기를 마련하고자 대학생 논문 경진대회 구두 세션이 운영되었다. 본 세션에서는 총 21편의 논문이 발표되었으며, 성균관대학교(6편), 국립군산대학교(4편), 홍익대학교(4편), 경북대학교/KETI(3편), 국립한밭대학교(1편), 건양대학교(1편), 세종대학교/성균관대학교(1편), 광운대학교(1편)가 참여하였다. 발표 주제는 3D Gaussian Splatting, 3D 객체 취득 및 재구성, SLAM, 3D 객체 탐지 등 3차원 비전 기술을 비롯하여, 딥페이크 탐지, 눈 깜빡임 검출, 저조도 영상 개선, 영상 기반 객체 탐지 등 다양한 컴퓨터 비전 기술을 포함하였다. 또한 강화학습 기반 최적화, Audio-Visual Question Answering, 공간 컴퓨팅 인터페이스, 옛지 디바이스 기반 LLM 추론, 시각장애인을 위한 오디오 설명 삽입 등 인공지능 기반 미디어 응용 및 소프트웨어 시스템 분야의 다양한 연구 성과가 소개되었다.



캡스톤디자인 경진대회

캡스톤디자인 경진대회에서는 학생들이 전공 지식과 최신 인공지능·미디어 기술을 바탕으로 직접 설계하고 구현한 다양한 프로젝트가 포스터 형식으로 발표되었다. 경성대학교 팀은 Swin-DINO 자기 지도 학습과 YOLO 지도 학습을 연계한 소형 객체 탐지 성능을 분석하고, 옛지 AI 기반 실시간 영상 관계 시스템을 구현하여 객체 탐지 성능을 비교하였다.

경희대학교 팀은 적응적 임계값을 활용한 G-PCC RAHT 화면 간 예측의 압축 성능 향상 기법과 딥러닝 기반 특징점 추출 및 3D Gaussian Splatting을 이용한 3차원 객체 모델링 파이프라인을 제안하였다. 동아대학교와 ETRI 팀은 Depth-wise Convolution 기반의 경량 Attention Block을 설계하고 성능을 평가하였으며, 경북대학교 팀은 Retrieval-based Few-shot 프롬프팅을 활용하여 SQL Agent의 실시간 쿼리 생성 지연을 개선하였다. 세종대학교 팀은 영상 속 이상 상황을 의미 단위로 분해하여 검색 결과의 근거를 제시할 수 있는 설명 가능한 영상 검색 시스템을 구현하였다. 서울과학기술대학교 팀은 전·후면 이미지와 생성형 AI를 활용한 일관성 있는 인물 3D 포인트 클라우드 생성, 메타 퀘스트 3 기반 XR-AI 멀티모달 인터랙션 앱, 단일 360도 이미지 기반 PerRF 공간 재구성 및 가려짐 영역에 따른 성능 한계 분석, 확장현실 기기의 실시간 AI 객체 탐지를 위한 클라우드 시스템을 발표하였다. 한국기술교육대학교 팀은 사용자의 시선 정보를 실시간으로 인식하여 로봇의 동작을 제어하는 시선 기반 로봇 제어 시스템을 구현하였다.



산업체/취창업 세션

산업체/취·창업 세션에서는 인공지능 기술이 방송장비와 미디어 콘텐츠의 제작·전송·송출 과정에 가져오는 변화와 연구성과의 사업화, 디지털 트윈 및 퍼지컬 AI 기술의 산업적 활용 사례가 소개되었다.



대회를 관통한 하나의 흐름: 방송미디어의 지능화와 공간화

2026년 하계학술대회는 단순히 많은 논문과 세션을 모은 행사가 아니었다. 생성형 인공지능이 미디어 제작과 이해 방식을 바꾸고, 3D Gaussian Splatting(3DGS)을 비롯한 공간미디어 기술이 차세대 방송의 표현·압축·전송 구조를 재편하는 시점에서, 국내 방송미디어 연구가 어디로 향하고 있는지를 집약적으로 보여준 자리였다.

프로그램 전반에서는 세 가지 변화가 뚜렷하게 확인되었다. 첫째, 인공지능은 더 이상 방송 제작을 보조하는 개별 도구가 아니라 콘텐츠 생성, 품질 개선, 이용자 접근성, 서비스 운영을 연결하는 핵심 기반 기술로 자리 잡았다. 둘째, 2D 영상 중심의 전통적 방송기술이 3D·4D 공간 표현과 몰입형 미디어로 확장되면서 압축·렌더링·표준화가 하나의 연구 생태계로 결합되고 있다. 셋째, 대학·연구기관·방송사·기업이 각자의 연구 결과를 소개하는 수준을 넘어 실제 현장 적용과 표준화, 후속 공동연구를 함께 논의하는 구조가 강화되었다. 따라서, 전체적으로 프로그램을 네 개의 의미 있는 축으로 다시 보면 다음과 같다.

- AI가 방송 제작과 미디어 서비스를 다시 설계하다

KBS 특별세션, 생성형 미디어, AI 기반 UI/UX 접근성, 미디어 제작보조 및 공간 사전시각화 관련 세션은 인공지능이 방송 현장에 실제로 들어오고 있음을 보여주었다. 특히 생성형 AI는 영상 제작의 일부 단계를 자동화하는 데 그치지 않고, 기획·편집·복원·접근성·유통에 이르는 전체 파이프라인을 재구성하는 방향으로 발전하고 있었다. 이러한 변화는 방송기술 연구의 평가 기준도 바꾸고 있다. 알고리즘 정확도나 압축 효율만으로는 충분하지 않으며, 제작자의 업무 흐름에 얼마나 자연스럽게 결합되는지, 시청자 경험을 어떻게 개선하는지, 신뢰성과 책임성을 어떻게 확보하는지가 함께 논의되어야 한다는 점이 부각되었다.

- 공간미디어와 3DGS가 차세대 핵심 연구축으로 부상하다

GS 기반 이머시브 미디어와 3DGS 기술·표준화 전망 세션은 이번 학술대회의 기술적 중심축 가운데 하

나였다. 여러 세션에서 3차원 장면 복원, 동적 장면 표현, 3D·4D Gaussian의 압축, 렌더링 효율화, 시간적 일관성, 워터마킹과 표준화가 연속적으로 다뤄졌다.

이는 3DGS가 단기적인 유행 기술을 넘어, 실감미디어의 획득·표현·압축·전송·재현을 잇는 공통 플랫폼으로 발전하고 있음을 의미한다. 특히 세계 최고 수준 학술대회에 발표된 4DGS 압축, 장시간 동적 장면 모델링, 다중 전문가 기반 동적 표현 등의 연구가 소개되면서 국내 연구자들이 국제 연구의 최전선과 긴밀하게 연결되어 있음을 확인할 수 있었다.

- 연구성과가 산업 현장과 표준화로 이어지다

이번 대회의 또 다른 특징은 학술 연구, 방송 현장, 산업 기술, 표준화가 서로 분리되지 않았다는 점이다. KBS·MBC·SBS 등 방송사, ETRI·KETI·TTA 등 연구·표준화 기관, 그리고 다양한 미디어 기업이 참여해 현장 적용 사례와 기술 수요를 직접 공유하였다.

초청강연에서 제시된 KBS의 방송기술 비전과 Fraunhofer HHI의 실세계 3DGS 압축 기술은 이러한 연결을 상징적으로 보여주었다. 연구 결과가 논문에 머무르지 않고 실제 배포 가능한 기술이 되기 위해서는 계산 복잡도, 압축 효율, 상호운용성, 표준화와 같은 조건이 함께 해결되어야 한다는 공감대가 형성되었다.

- 세대 간 지식 순환과 연구 생태계를 확장하다

국제 최우수컨퍼런스 논문 PI 초청 발표는 국내 연구자들이 세계 최고 수준의 성과를 학회 구성원들과 직접 공유하는 통로가 되었다. 단순 논문 소개를 넘어 연구 문제의 설정, 핵심 아이디어, 실패와 개선 과정, 후속 연구 가능성을 함께 논의함으로써 최신 연구의 맥락을 전달했다는 데 의미가 있다.

신진연구자 세션은 새롭게 대학에 부임한 연구자들의 연구 방향을 학회 공동체에 소개하고 협업의 접점을 만드는 자리였다. 대학생 논문 및 캡스톤디자인 경진대회는 학부생이 연구 결과를 발표하고 피드백을 받는 경험을 제공했다. 이처럼 석학·중견 연구자·신진 연구자·학생이 하나의 프로그램 안에서 연결된 점은 학회의 지속가능성을 높이는 중요한 성과였다.



결론적으로, 이번 학술대회가 남긴 네 가지 성과는 다음과 같다. (1) 기술 의제의 선명화: 생성형 AI와 공간미디어가 방송미디어 분야의 주변 주제가 아니라 향후 학회의 핵심 의제임을 분명히 했다. (2) 산·학·연 연결의 강화: 학술 발표와 방송 현장, 산업 기술, 표준화 논의가 같은 프로그램 안에서 이어지며 공동 연구와 기술이전의 가능성을 높였다. (3) 국제 연구성과의 확산: top-tier 국제학술대회 성과를 국내 연구자들이 직접 설명함으로써 최신 연구의 접근성을 높이고 후속 연구를 촉진했다. (4) 미래 연구자 육성: 신진연구자와 대학생을 별도의 부가 프로그램이 아니라 학술대회의 핵심 구성원으로 배치해 세대 간 연구 생태계를 확장했다.



맺음말

2026년 한국방송·미디어공학회 하계학술대회는 6월 18일(목)부터 6월 20일(토)까지 롯데호텔 제주에서 성공적으로 개최되었다. 총 314편의 논문이 발표되고 524명의 회원이 등록한 이번 학술대회는 방송 미디어 분야의 최신 기술 동향에 대한 높은 학술적 수요를 객관적으로 입증하였다. 3일간 진행된 행사는 3D Gaussian splatting, 대형 언어 모델, 생성형 인공지능 등 최신 기술 패러다임 변화를 반영하여 특별 세션과 국제 최우수컨퍼런스 논문 PI 초청 발표 세션을 대폭 확대 구성하였다. 이를 통해 참가자들에게 최신 연구 동향을 분석할 수 있는 실효적인 기회를 제공하였다. 안정적인 운영을 가능하게 한 체계적인 기획과 실행력의 준비위원회와 본 행사를 지원하고 참여한 모든 연구자 및 관계자에게 감사를 표한다.