특집논문 (Special Paper)
방송공학회논문지 제30권 제6호, 2025년 11월 (JBE Vol.30, No.6, November 2025)
https://doi.org/10.5909/JBE.2025.30.6.997
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

다시점 기반의 사실적 3D 얼굴 합성 시스템

이 학 범[a], 서 영 호[a)b)‡]

Multi-camera-based 3D Facial Reconstruction System

Hak-Bum Lee[a] and Young-Ho Seo[a)b)‡]

요 약

본 연구는 다중 카메라 시점에서 획득한 동기화된 고해상도 이미지를 기반으로 사실적인 3D 얼굴을 복원하는 시스템을 제안한다. 제안 시스템은 (1) 멀티카메라 하드웨어 구성 및 동기화, (2) 촬영 품질을 담보하는 운영 체크리스트, (3) 포토그래메트리 기반 3D 재구성 및 좌표계 정렬, (4) 텍스처 합성과 재투영 기반 품질 평가로 구성된다. 우리는 전면·측면·측후면으로 배치된 약 60대 카메라 리그와 균일 조명, 마커 기반 정합을 통해 입력 영상과의 광학적 일관성(photometric consistency)을 유지하는 고해상도 얼굴 표현을 얻는다. 품질 평가는 마스크 기반 PSNR/SSIM과 재투영 차이 비교로 수행하며, 결과는 실제 촬영 영상과 유사한 시각적 자연스러움을 보인다. 실제 촬영 영상과 유사한 수준의 광학적 특성을 지닌 3D 얼굴 표현을 가능케 하는 기술적 기반을 마련하는 데 목적이 있다.

Abstract

This study proposes a system for reconstructing realistic 3D facial models from synchronized high-resolution images captured by a multi-camera setup. The proposed framework consists of four main components: (1) multi-camera hardware configuration and synchronization, (2) an operational checklist to ensure capture quality, (3) photogrammetry-based 3D reconstruction with coordinate alignment, and (4) texture synthesis combined with reprojection-based quality evaluation. Using a rig of approximately 60 cameras arranged in frontal, lateral, and oblique positions, together with uniform illumination and marker-based alignment, we achieve high-resolution facial representations that maintain photometric consistency with the input images. The reconstruction quality is evaluated through mask-based PSNR/SSIM metrics and reprojection error analysis, demonstrating visual fidelity comparable to real captured footage. The proposed system provides a technological foundation for generating 3D facial representations with optical characteristics closely resembling real video recordings.

Keyword : Multi-camera system, 3D Facial reconstruction, Photogrammetry, Digital Human

a) 광운대학교 전자재료공학과(Department of Electronic Materials Engineering, Kwangwoon University)
b) 오모션 주식회사(Omotion Inc.)
‡ Corresponding Author : 서영호(Young-Ho Seo)
E-mail: yhseo@kw.ac.kr
Tel: +82-2-300-0263
ORCID: https://orcid.org/0000-0003-1046-395X

※이 논문의 결과 중 일부는 한국방송·미디어공학회 2025년 하계학술대회에서 발표한 바 있음
※이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. RS-2025-16071490) This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. RS-2025-16071490)
· Manuscript September 16, 2025; Revised November 5, 2025; Accepted November 5, 2025.

Ⅰ. 서 론

메타버스(Metaverse), 가상 제작(Virtual Production), 디지털 휴먼(Digital Human) 기술의 산업적·학문적 활용 범위의 확대는 사실적 얼굴 표현의 중요성을 더욱 부각시키고 있다. 얼굴은 감정과 발화를 표현하는 핵심 채널로서, 사용자 몰입감과 내러티브 전달력을 결정짓는 요소이다. 산업 현장에서는 게임·영화·방송·광고 등에서 얼굴 애니메이션과 립싱크(Lip-Sync)의 부자연스러움이 출시 지연이나 추가 후처리 비용의 주요 원인으로 지적되고 있다. 특히 국내 제작 환경에서는 페이셜 캡처(Facial Capture) 전문 인력과 고성능 장비의 부족으로 인해 한국어 및 다국어 대사의 자연스러운 표현이 제약되고 있다.

기존의 단일 시점 또는 소수 시점 기반 3D 복원은 심도 모호성(Depth Ambiguity)과 가려짐(Occlusion), 피부 반사로 인한 이상치(Specular Reflection) 때문에 눈꺼풀, 구강 내부, 턱선과 같은 불안정 영역을 정밀하게 재현하기 어렵다. 또한 언어별 입술 움직임(Lip Motion)을 정교하게 표현하려면 시점별 시간 차이를 최소화한 멀티뷰(Multi-View) 데이터가 요구되지만, 다수의 카메라를 완전하게 동기화하고 일관된 노출·색 보정을 유지하는 것은 여전한 실무적 난제이다.

또한 최근 딥러닝 기반 3D 얼굴 복원(예: 3D Morphable Model 회귀, NeRF/3DGS 기반 생성 모델) 기법이 활발히 연구되고 있으나, 이들은 대규모 학습 데이터셋에 의존하고 특정 도메인에 과적합(overfitting)되기 쉽다[1][2]. 특히 법과학·디지털 휴먼 제작과 같이 높은 재현성(reproducibility)과 정량 검증(quantitative QA)이 요구되는 응용에서는, 블랙박스 모델의 불투명한 추론 과정이 결과의 신뢰성을 떨어뜨릴 수 있다[2][3]. 이에 비해 포토그래메트리 기반 접근은 입력 영상과의 광학적 일관성(photometric consistency)을 유지하고, 좌표계 정렬 및 품질 평가 과정을 명확히 정의할 수 있어 실험 반복이 용이하다[3][4].

실제 산업 현장에서도 Sony, Artec, 3dMD 등 주요 캡처 시스템은 여전히 다수의 카메라를 활용한 멀티뷰 포토그래메트리 방식을 채택하고 있으며, 의학 및 법과학 분야에서도 신뢰성 확보를 위해 동일한 접근법을 활용한다[5][6]. 본 연구는 이러한 실무적·산업적 요구를 고려하여, 학습 기반 생성 모델이 아닌 실측 기반 baseline 파이프라인을 제시함으로써 후속 연구와 데이터셋 구축의 기준선을 제공하고자 한다.

기존의 SfM/MVS 기반 포토그래메트리 파이프라인은 일반적으로 정합점 추출 - 번들 조정 - 메싱 - 텍스처 합성의 절차로 구성되어, 다양한 응용 분야에서 널리 활용되어 왔다. 다만, 촬영 표준화나 좌표계 정렬 절차가 파이프라인 외부에서 개별적으로 수행되는 경우가 많아, 세션 간 결과의 일관성(reproducibility) 확보나 품질 평가(QA)의 정량화에는 다소 제약이 있었다. 본 연구에서는 이러한 부분을 보완하고 절차 간 통합성을 높이기 위해, 마커 기반 좌표계 정렬을 도입하여 세션 간 정합성을 향상시키고, 조명·노출의 균질화 절차를 추가하여 재현성을 강화하였으며, 재투영 기반 정량 품질 평가(QA)를 통합함으로써 전체 워크플로우의 신뢰성을 개선하였다.

이러한 한계를 극복하고 반복 가능성을 확보하기 위해, 멀티카메라 포토그래메트리(Photogrammetry)를 활용한 다시점 기반 사실적 3D 얼굴 합성 기준 시스템을 제안한다. 제안 시스템은 (1) 다수의 카메라로 구성된 원주형 리그와 하드웨어·소프트웨어 트리거를 조합한 준실시간 동기화, (2) 촬영 품질 보장을 위한 체크리스트와 피사체 준비 가이드, (3) 포토그래메트리 기반 3차원 재구성과 좌표계 정렬, (4) 시점 간 색상·노출 보정을 고려한 텍스처 블렌딩과 재투영(Reprojection) 기반 품질 평가로 구성된다.

본 연구의 주요 기여는 다음과 같다.

- 멀티카메라 리그 및 데이터 파이프라인 설계: 수십 대 규모의 카메라를 활용하여 다양한 시점에서 고해상도 영상을 획득한다.
- 촬영 및 운영 표준화: 마커 배치, 조명 환경, 피사체 준비를 포함한 촬영 체크리스트를 제시하여 데이터 수집의 반복 가능성과 품질 일관성을 확보하였다.
- 안정적 기하 복원: 포토그래메트리 기반 재구성과 마커 기반 좌표계 정렬을 통합하여 재현성 높은 3차원 기하를 복원하였다.
- 객관적 품질 평가: 마스크 기반 PSNR, SSIM 지표와 재투영 차이를 활용해 시각적·광학적 일관성을 정량화하였다.

본 연구는 결과적으로 다국어 립싱크, 몰입형 콘텐츠, 디지털 휴먼 제작 파이프라인 등에서 활용 가능한 고품질 3D

얼굴 데이터를 구축할 수 있는 실용적 기반을 제공한다.

Ⅱ. 관련 연구

최근 얼굴 3D 복원 기술은 크게 학습 기반(learning-based) 방법과 비학습(non-learning based photogrammetric) 방법으로 양분된다. 학습 기반 접근은 NeRF/3DGS 기반 표현 학습, 대규모 다중 시점 데이터셋으로 훈련된 feed-forward MVS 네트워크, 그리고 대형 Foundation 모델을 이용한 implicit surface 회귀로 발전해 왔다[1][6]. 이러한 방법들은 빠른 추론과 novel view 합성에서 높은 표현력을 보이지만, 학습 데이터셋 편향과 추론 과정의 불투명성 때문에 디지털 휴먼·제조·의료·법과학 등 결과의 재현성과 신뢰성이 중요한 응용에서는 적용에 제약이 따른다. 이에 따라 촬영, 동기화, 캘리브레이션 절차의 표준화부터 기하 복원, 텍스처 합성, 재투영 기반 정량 평가까지 전 과정을 통제할 수 있는 비학습 포토그래메트리 접근법이 여전히 중요하게 다뤄지고 있다.

다시점 기반 얼굴 재구성에서 포토그래메트리적 접근은 하드웨어 구성, 동기화·보정, 표면 복원과 텍스처링, 재투영 기반 품질 보증으로 이어지는 전형적 파이프라인을 중심으로 발전해 왔다[1].

다시점 촬영 시스템 설계 측면에서, Giuliani et al.(2024)는 법과학 응용을 위한 다중 시점 얼굴 포토그래메트리 시스템을 제안하고, 저비용 카메라 배열·표준화된 촬영 절차·정밀 캘리브레이션으로 고세부 얼굴 메시에 도달하는 구성을 보고하였다[1]. 동기화 측면에서 Zhou et al.(2024)는 하드웨어 트리거 없이 글로벌 기준 비디오와의 시간 캘리브레이션을 통해 프레임 이하(subframe) 정밀도의 정렬을 달성하는 방법을 제시하였다[2]. 텍스처 합성 분야에서는 Liu et al.(2025)의 TRSP 연구가 다시점 이미지와 재구성 표면을 이용하여 사전(prior) 기반 텍스처 재구성을 수행하고, 기하·관측의 일치성을 높이는 절차를 보고하였다[3]. 정량적 품질 평가 연구에서는 Ruiu et al.(2025)이 재투영 오차, 카메라 포즈, 점군(point cloud) 지표 등 전통적 SfM/MVS 평가 지표의 체계적 활용을 재정리하였다[4]. 또한, 모바일 기기로 구축한 다차원 얼굴 데이터셋 Uniss-MDF(Ruiu et al., 2025)가 공개되면서, 전통 파이프라인의 현장 적합성과 일반화 성능을 검증할 수 있는 실증 자료가 확장되었다[5].

종합하면, 최근의 비학습 기반 연구는 다중 시점 포토그래메트리 시스템의 현장 적용성 강화, 무트리거 정렬을 포함한 정밀 동기화, 관측 일관성 기반 텍스처 재구성, 재투영 중심의 표준화된 정량 QA로 발전하고 있다.

Ⅲ. 제안 방법

본 장에서는 앞서 논의한 문제 정의와 관련 연구를 바탕

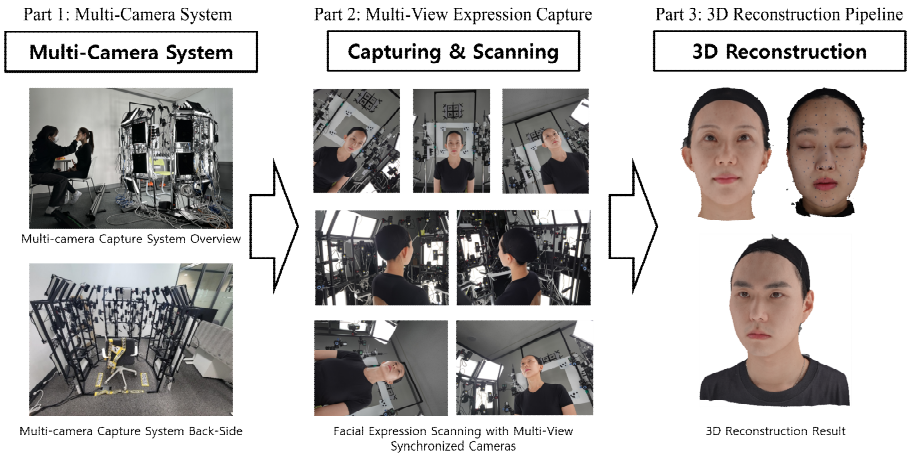

그림 1. 제안한 3차원 재구성 프레임워크의 전체 처리 절차 개요도.
Fig. 1. Overview of the proposed 3D Reconstruction Framework

으로 멀티카메라 포토그래메트리 기반의 사실적 3D 얼굴 합성 시스템을 제안한다. 제안 시스템은 다중 시점 캡처, 정밀 동기화 및 보정, 포토그래메트리 기반 3차원 재구성, 3D 품질 평가로 구성된 파이프라인을 따른다.

그림 1은 제안 시스템의 전체 파이프라인을 나타내며, (1) 다수의 카메라와 조명으로 구성된 캡처 리그를 이용한 동시 촬영, (2) 동기화된 멀티뷰 데이터 획득 및 피사체 표정 스캔, (3) 포토그래메트리를 통한 정밀 3차원 재구성 단계를 포함한다. 이를 통해 디지털 휴먼을 정밀하게 복원할 수 있다.

1. 전체 시스템 구성

본 시스템은 피사체를 둘러싼 약 60대의 디지털 카메라(예: Sony RX0 II)를 원주형으로 배치하고, 이더넷 기반 트리거 및 전원 장치(예: Sony CCB‑WD1)와 1 GbE 스위치 허브를 통해 동시에 촬영하는 방식을 취한다. 하드웨어 트리거를 사용하여 각 카메라의 셔터 지연과 스큐를 10ms 미만으로 억제하고, 이러한 구성을 표준 촬영 리그로 정의한다. 전송 및 저장 설계는 H.264 스트리밍을 기준으로 2K@30fps에서 약 5‑10Mbps, 4K@30fps에서 약 15‑25Mbps의 대역폭을 가정하며, MJPEG 코덱 사용 시 해상도에 따라 프레임당 데이터량이 크게 증가하므로 코덱과 해상도를 품질과 처리량 간의 균형에 맞게 선택한다.

하드웨어의 안정적인 운용을 위해 본 시스템은 트리거-동기화 모듈, 전원 분배 모듈, 데이터 수집 서버를 분리하여 설계하였다. 트리거-동기화 모듈은 PoE(Power over Ethernet)를 기반으로 전체 카메라에 동시 신호를 분배하여 촬영 타이밍을 맞추고, 전원 모듈은 각 카메라의 과전압·과전류를 방지하는 보호 회로를 포함한다. 데이터 수집 서버는 다중 NIC(Network Interface Card)를 탑재하여 병렬 스트림을 수신하며, RAID 스토리지를 사용해 실시간으로 데이터를 저장함으로써 대규모 촬영에서도 데이터 유실을 최소화한다. 모든 모듈은 1U/2U 랙마운트 규격으로 구성해 이동형 촬영 환경에서도 손쉽게 설치·해체할 수 있도록 하였다.

촬영 품질을 결정짓는 조명 시스템은 5600K 기준의 소프트박스 LED를 원주형으로 설치하여 균일한 확산광을 제공한다. 피사체 상·하부에는 추가 보조 조명을 배치해 그림자와 하이라이트를 조절하며, 조도 편차는 ±5% 이내로 유지하도록 조명 출력을 미세 조절한다. 피부 표면의 반사로 인한 이상치를 줄이기 위해 편광 필터를 사용할 수 있으며, 조명 장치 전반에는 Flicker-free 기술을 적용하였다. 이러한 조명 설계는 포토그래메트리 단계에서의 광학적 일관성을 높이고, 텍스처 재구성 시 음영 오류를 최소화하는 데 기여한다.

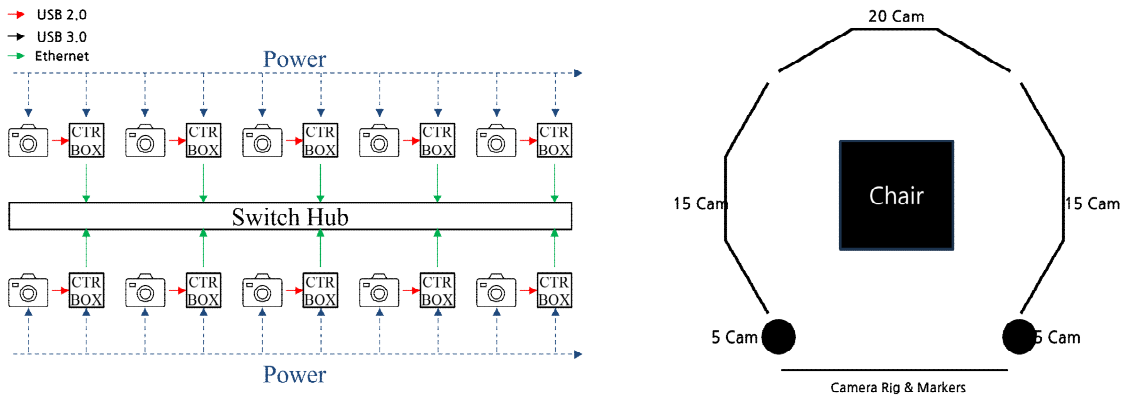

그림 2. 하드웨어 시스템 구성 및 카메라/마커 배치 (a) 카메라·컨트롤러 박스·스위치 허브 간의 전원 및 데이터 연결 구성, (b) 피사체를 중심으로 한 다중 시점 카메라 리그 및 마커 배치 예시

Fig. 2. Hardware system configuration and camera/marker layout, (a) Power and data connection between cameras, controller boxes, and switch hub, (b) Multi-view camera rig and marker placement around the subject

캘리브레이션과 운영 표준화 또한 중요한 구성 요소이다. 카메라 내·외부 파라미터는 fisheye 모델을 포함한 다양한 왜곡 모델을 고려해 정기적으로 추정하며, ArUco/ChArUco 마커를 이용해 글로벌 좌표계를 정의한다. 시스템을 분해·재설치한 후에도 동일한 품질을 유지할 수 있도록 표준화된 캘리브레이션 프로토콜을 마련하였다. 더불어, 촬영 세션 간 품질의 일관성을 확보하기 위해 피사체 위치·자세·표정 제어, 트리거 타임스탬프 기록, 노출과 화이트밸런스 로그 저장 등 운영 체크리스트를 작성하고, 촬영 데이터와 메타데이터를 세션별로 통합 관리하여 후처리 단계에서 자동 정렬 및 QA(품질 평가)에 활용하도록 하였다.

2. 하드웨어 시스템 구성카메라/조명/마커 배치

카메라 배치는 피사체 주변을 원주형으로 감싸며 다양한 시점을 확보하는 것이 핵심이다. 본 연구는 전면 20대, 좌·우 측면 각 15대, 측후면 각 5대의 카메라를 배치하여 360°에 걸쳐 균일한 촬영 시점을 확보한다. 하악부(턱선) 디테일을 개선하기 위해 전면 시점을 24대로 증설하는 개선안을 함께 제시하며, 측면·후면 시점에서는 필요 이상의 중복을 최소화해 정보 효율을 높인다. 모든 카메라는 동일한 렌즈 스펙과 화각을 사용하되, 피사체와의 거리 및 높이를 미세 조정해 시점별 겹침 영역을 균일화하였다.

조명은 좌·우·정면에 수직 확산광 패널을 설치하여 얼굴 표면 전체에 균일한 조도를 제공한다. 사용된 LED는 약 55W급 고출력 모델로, 1m 거리에서 약 3000lux 이상의 밝기를 유지하도록 설계하였다. 디퓨저를 장착해 빛을 부드럽게 확산시키고, 약 130°의 넓은 빔 각을 확보하여 광원이 카메라의 시야에 직접 노출되지 않도록 하였다. 이러한 조명 배치는 피사체 피부의 미세한 질감과 색상을 정확히 포착하고, 다중 시점 간 색상·노출의 일관성을 유지하는 데 필수적이다.

글로벌 좌표계를 정의하고 스케일을 고정하기 위해 마커 시스템을 도입하였다. 6×6 아루코(ArUco) 마커와 원형 마커를 병용하여 변 길이 0.40m의 정사각형 꼭짓점에 배치함으로써, 각 카메라의 외부 파라미터를 정확히 추정할 수 있도록 하였다. 마커의 배열과 무게중심은 월드 좌표계의 원점으로 설정하고, 마커 판독에 영향을 미칠 수 있는 배경이나 조명 반사도는 사전에 제거하였다. 또한 마커와 카메라 간의 거리와 각도를 일정하게 유지해 캘리브레이션 오차를 최소화했다.

추가적으로, 카메라 간 동기화와 촬영 편의를 위해 케이블 관리 구조와 배선 공간을 설계하였다. 각 카메라의 트리거 라인을 동일 길이로 유지해 신호 지연을 균등화하며, 전원 케이블과 데이터 케이블을 분리된 채널에 배치해 EMI(전자기 간섭)를 줄인다. 카메라 부착 지그는 높이 조절이 가능한 구조로 설계되어 피사체 신장과 자세에 맞춰 조정할 수 있으며, 기계적 떨림을 억제하는 댐퍼를 적용해 촬영 중 진동을 최소화하였다. 이러한 세부 설계는 전체 시스템의 안정성을 높이고 반복 촬영 시 동일한 세팅을 유지하는 데 기여한다.

3. 촬영 운영 가이드

촬영 준비는 환경 정비부터 시작한다. 모든 카메라를 정면 기준으로 정확히 얼라인하고, 조명 위치와 차광 장치를 점검하며 외부인의 출입을 통제해 광원 변화와 잡음을 최소화한다. 촬영 전에 카메라 렌즈를 청소하고 메모리와 배터리 상태를 확인하며, 하드웨어 트리거와 네트워크 연결이 정상적으로 동작하는지 테스트 샷을 통해 검증한다. 이러한 사전 준비는 촬영 품질을 좌우하는 중요한 단계이다.

피사체 준비 과정에서는 피부 윤광을 줄이고 표정 주름을 안정적으로 추적하기 위한 조치를 취한다. 구체적으로는 파우더를 사용해 피부의 유분을 제거하고 광택을 줄이며, 헤어넷과 스프레이를 통해 이마와 옆머리를 고정해 얼굴 윤곽이 잘 드러나도록 한다. 얼굴의 움직임 추적을 위해 마커 페인팅을 실시할 때에는 피부 자극이 적은 아이라이너를 사용하고, 눈썹·눈두덩·입술·턱선 등 표정 변화가 큰 영역에 일정한 간격으로 점을 찍어 변형 추적의 안정성을 높인다. 이때 귀와 목 주변 마커의 누락이 없는지 재확인하여 후처리 과정에서의 오류를 줄인다.

촬영 직전에는 각 카메라의 메모리를 초기화하고 시간 동기화를 다시 한 번 확인한다. 촬영 체크리스트에 따라 중앙 카메라와 피사체 정면을 일치시키고, 테스트 샷을 통해 조명과 포커스를 최종 점검한다. 필요 시 노출과 화이트밸런스를 미세 조정해 모든 카메라가 동일한 밝기와 색온도

를 유지하도록 한다. 촬영 세션 중에는 트리거 타임스탬프, 카메라 온도, 렌즈 정보 등 운영 로그를 실시간으로 기록하며, 예상치 못한 이상 현상이 발생할 경우 즉시 대응할 수 있도록 시스템 모니터링을 수행한다.

마지막으로, 촬영이 끝난 후에는 모든 데이터를 정리하고 세션별 폴더 구조와 메타데이터 파일(JSON/XML)을 생성해 저장한다. 각 세션의 환경 변수(촬영일, 카메라 구성, 조명 설정 등)와 피사체 정보(성별, 언어, 표정 종류 등)를 기록하여 이후 실험에서 반복 가능성을 확보한다. 또한, 촬영 중 발생한 문제점과 개선 사항을 운영 일지로 남겨 다음 촬영 시 반영함으로써 시스템 운영의 효율성을 지속적으로 향상시킨다.

4. 3D 재구성 파이프라인

3차원 재구성 파이프라인은 정확한 카메라 파라미터 추정과 영역 설정에서 시작한다. 먼저 번들 조정(Bundle Adjustment, BA)을 포함한 카메라 캘리브레이션을 수행하여 왜곡을 고려한 내부·외부 파라미터를 추정하고, 이 정보에 기반하여 얼굴과 목을 중심으로 재구성할 영역을 마스크로 정의해 배경과 잡음 영역을 배제한다. 이후 구조에서 모션(Structure from Motion, SfM) 알고리즘을 통해 희소한 특징점과 카메라 포즈를 복원한 뒤, 다중 시점 스테레오(Multi View Stereo, MVS)를 적용하여 조밀 점군과 각 시점의 깊이 지도를 계산한다. 이러한 단계는 포토그래메트리 기반 접근 방식으로 눈꺼풀, 입술, 턱선 등 복잡한 부분의 기하를 안정적으로 추정하는 데 핵심적이다.

산출된 점군은 삼각화 또는 스크린드 포아송(Screened Poisson) 기반 메싱 알고리즘을 사용해 연속적인 표면(mesh)으로 변환한다. 메싱 과정에서 노이즈 제거와 홀 필링을 수행해 위상 결함을 교정하고, 생성된 메시를 마커 정보를 이용해 월드 좌표계로 스케일·회전·이동을 정렬한다. 이 단계에서 마커 기반 좌표계 정렬을 통해 각 세션 간 결과를 동일한 기준으로 비교할 수 있으며, 실제 환경과의 일치성도 향상된다. 필요한 경우 메시 스무딩과 준위(Subdivision) 과정을 추가해 곡면의 자연스러운 연속성을 확보한다.

텍스처 합성 단계는 멀티뷰 영상에서 추출한 색상과 기하 정보를 표면에 투영해 솔기 없는 일관된 텍스처를 생성하는 과정이다. 본 연구에서는 가시성, 입사각, 노출을 반영한 가중치를 적용하고, 블렌딩 과정에서 멀티밴드 및 그래

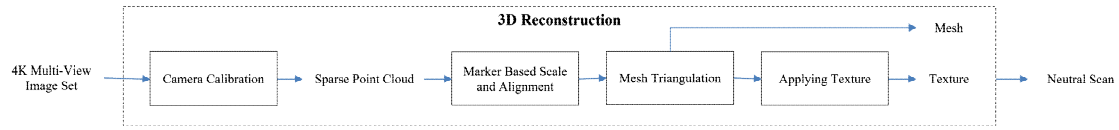

그림 3. 제안한 3차원 재구성 파이프라인 개요
Fig. 3. Overview of the proposed 3D reconstruction pipeline

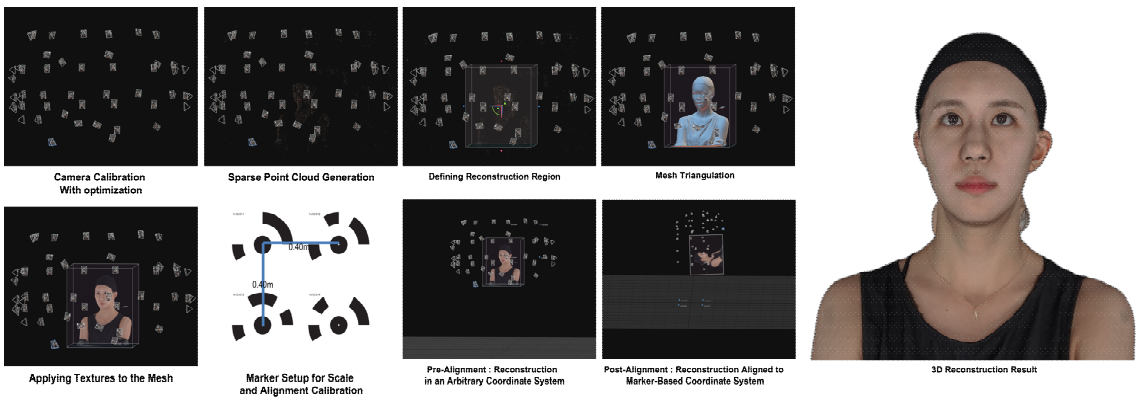

그림 4. 3D 재구성 파이프라인 단계별 결과
Fig. 4. Step-by-step results of the 3D reconstruction pipeline

디언트 도메인 기법을 활용하여 시점 경계에서 발생하는 색상 불연속을 최소화하였다. 텍스처 베이킹(Texture Baking)에서는 UV 전개가 완료된 메시를 기준으로 디퓨즈(Diffuse), 노멀(Normal), 디스플레이스먼트(Displacement) 맵을 생성한다. 디퓨즈 맵은 최대 16K 해상도로 생성해 피부의 미세 질감을 보존하고, 노멀 맵은 8K까지 생성하여 세밀한 표면 변화를 표현한다. 디스플레이스먼트 맵은 조밀 깊이 데이터와 메시 간의 잔차를 32비트 부동소수점으로 인코딩하여 서브디비전 시 미세 형상 복원이 가능하도록 한다.

최종적으로 생성된 메시는 언리얼 엔진(Unreal Engine), 블렌더(Blender) 등 실시간 렌더러에 바로 적용할 수 있도록 표준 규격(예: sRGB/Linear, MikkTSpace)을 준수한다. 또한 다양한 언어의 립모션과 표정을 포함한 데이터셋을 구축하여 후속 연구나 산업 현장에서 활용 가능하도록 하며, 시스템 동작 흐름과 재구성 파이프라인을 반자동 스크립트화하여 연구자와 제작자가 반복 수행하기 쉽게 도왔다. 이처럼 포토그래메트리 기반 재구성, 좌표계 정렬, 텍스처 베이킹을 통합한 파이프라인은 학습 기반 모델에 의존하지 않고도 고품질 3D 얼굴 합성을 실현한다는 데 의의가 있다.

5. 3D 재투영 기반 품질 평가

제안된 시스템의 품질 평가는 3D 재투영(reprojection)에 기반하여 수행된다. 우선 내부·외부 파라미터로 보정된 입력 카메라의 동일한 위치와 방향에서 복원된 메시에 대한 렌더링을 진행하고, 그 결과를 원본 영상과 정합하여 차이 영상을 생성한다. 이 때 모든 비교는 렌즈 왜곡 보정(undistortion)과 감마 보정을 적용한 이후에 진행하며, 배경과 머리카락, 강한 반사 하이라이트와 같은 포화 픽셀은 마스크로 제외하여 얼굴 표면의 광학적 일관성만을 평가하도록 한다.

정량적 평가 지표는 피크 신호 대 잡음비(PSNR)와 구조적 유사도 지수(SSIM)를 사용하여 각 시점별 복원 품질을 수치화한다. PSNR은 원본과 재투영 영상 간의 픽셀 단위 차이를 로그 스케일로 표현해 높은 값일수록 노이즈가 적음을 의미하며, SSIM은 밝기, 대비, 구조 특성을 종합적으로 비교해 1에 가까울수록 구조적 유사성이 높음을 나타낸다. 각 카메라 시점별로 PSNR과 SSIM 값을 계산한 후, 평균과 표준편차를 함께 제시하여 전체 시스템의 일관성과 변동성을 분석한다.

추가적으로, 일반화 성능을 점검하기 위해 기존 시점뿐 아니라 새로운 시점(novel view)에서의 재투영 비교를 수행한다. 복원된 3D 얼굴 모델을 임의의 가상 카메라 위치에서 렌더링하고, 해당 시점에서 촬영한 실제 영상이 존재할 경우 비교하여 오차를 측정한다. 또한 고질적인 오차나 이상치(outlier)가 발생하는 프레임에 대해서는 사전에 정의한 임계 기준에 따라 별도로 분석하고, 원인이 촬영 단계의 조명 편차인지, 캘리브레이션 오차인지, MVS 알고리즘의 한계인지 등을 파악하여 개선 방안을 제시한다.

정량 지표와 함께 정성적 평가를 병행하는 것도 중요하다. 전문가 평가자들이 복원된 모델과 원본 영상을 비교하여 눈꺼풀, 구강 내부, 턱선 등 민감 영역의 재현성을 확인하고, 다국어 립모션에서의 발음 동기화와 표정 표현이 자연스러운지 평가한다. 이러한 정성적 피드백은 시스템의 실무 적용 가능성을 판단하는 데 중요한 기준이 되며, 정량 지표만으로 포착하기 어려운 섬세한 품질 차이를 반영한다. 이처럼 정량·정성 평가를 결합한 3D 재투영 기반 품질 평가는 본 연구가 제시한 시스템의 신뢰성과 개선 방향을 종합적으로 제시하는 데 필수적이다.

Ⅳ. 실험 결과

1. 데이터셋 및 절차

본 연구는 N명의 피실험자를 대상으로 전면 및 측면 시점에서 다양한 표정(중립, 미소, 입 벌림, 눈 감김 등)을 촬영한 자체 캡처 데이터셋을 구축하였다. 촬영은 3장에서 기술한 멀티카메라 리그와 조명 환경을 그대로 적용하여 표정별 다중 시점 데이터를 확보하였다. 각 세션은 촬영 직전 캘리브레이션을 수행하여 카메라 내부·외부 파라미터를 갱신하고, 피사체 위치와 자세를 통제하여 데이터의 반복 가능성을 보장하였다.

또한, 데이터셋에는 언어 발화가 포함된 세트도 별도로 수집하여 립모션 표현의 정합성도 함께 평가할 수 있도록 하였다. 캡처 후에는 모든 시점에 대해 렌즈 왜곡 보정 및

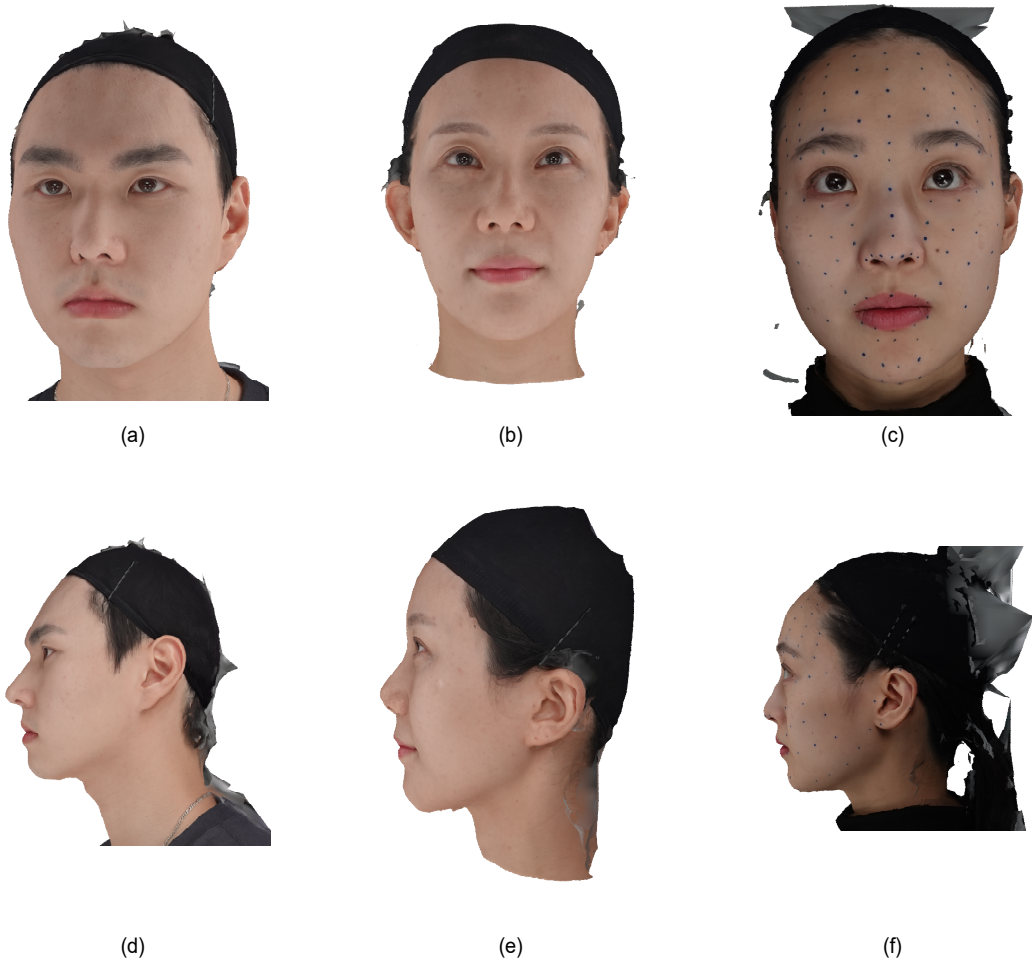

그림 5. 포토그래메트리 기반 3D 얼굴 재구성 결과, (a) 남성 피사체 전면, (b) 여성 피사체 전면, (c) 마커 부착 여성 피사체 전면, (d) 남성 피사체 측면, (e) 여성 피사체 측면, (f) 마커 부착 여성 피사체 측면

Fig. 5. Photogrammetry-based 3D face reconstruction results, (a) Male subject (front view), (b) Female subject (front view), (c) Female subject with facial markers (front view), (d) Male subject (side view), (e) Female subject (side view), (f) Female subject with facial markers (side view)

색상·노출 정규화를 적용하였고, 촬영 메타데이터(트리거 타임스탬프, 카메라 설정 값)를 JSON 형식으로 저장해 후처리 단계에서 재사용할 수 있도록 하였다.

평가 프로토콜은 입력 시점에서의 재투영 품질뿐만 아니라 신규 시점(Novel View)에서의 렌더 품질까지 점검할 수 있도록 설계되었다. 이를 위해 캡처 세션 중 일부 카메라를 학습/재구성 과정에서 제외하고, 해당 카메라를 “검증용 시점”으로 활용하여 일반화 성능을 측정하였다. 비교 시에는 배경, 투명(알파) 영역, 포화 하이라이트 영역을 마스크로 제거해 얼굴 표면 영역에만 집중하였다. 이러한 절차는 불필요한 배경 잡음이 정량 지표에 미치는 영향을 최소화하고, 모델의 실제 얼굴 재현 성능을 공정하게 비교하는 데 기여한다.

2. 정량 평가 지표

정량 평가는 광학적 일관성을 수치화할 수 있는 대표 지표인 피크 신호 대 잡음비(PSNR)와 구조적 유사도(SSIM)를 사용하였다. PSNR은 재투영 영상과 원본 영상의 평균 제곱 오차(MSE)를 로그 스케일로 변환한 값으로, 값이 높

그림 6. 재투영 기반 정량 평가 예시, (a) 원본 이미지 (전면), (b) 재투영 결과와 원본을 겹쳐 놓은 이미지 (전면), (c) 원본과 재투영 결과의 차이맵 (전면), (d) 원본 이미지 (측면), (e) 재투영 결과와 원본을 겹쳐 놓은 이미지 (측면), (f) 원본과 재투영 결과의 차이맵 (측면)

Fig. 6. Example of reprojection-based quality assessment, (a) Original image (front view), (b) Overlap of original and reprojected result (front view), (c) Difference map between original and reprojected result (front view), (d) Original image (side view), (e) Overlap of original and reprojected result (side view), (f) Difference map between original and reprojected result (side view)

을수록 재구성 결과가 원본과 근접함을 의미한다. SSIM은 밝기, 대비, 구조 특성을 종합적으로 비교하는 지표로, 1에 가까울수록 두 영상이 구조적으로 유사함을 나타낸다. 본 연구에서는 시점별 PSNR/SSIM을 산출하고, 평균과 표준편차를 함께 보고하여 전반적인 품질 수준과 시점 간 편차를 동시에 평가하였다.

추가적으로, 시각적 지각 품질을 평가하기 위해 LPIPS (Learned Perceptual Image Patch Similarity)를 산출하였다. LPIPS는 신경망 특성 공간에서 두 영상의 차이를 측정하는 지표로, 사람의 지각적 선호와 높은 상관성을 보이는 것으로 알려져 있다[9]. 이를 통해 단순한 픽셀 일치뿐 아니라 인물의 정체성과 감정 표현이 제대로 유지되는지를 종합적으로 평가하였다.

3. 재투영 결과 및 정량 분석

그림 6은 실제 촬영 영상과 동일 카메라 파라미터로 재투영한 결과를 비교한 예시를 보여준다. 반투명 중첩(α=0.5) 방식으로 비교한 결과, 전·측면 모두에서 안면 윤곽선과 주요 특징점(눈썹, 콧날, 입술)이 정밀하게 정합됨을 확인하

(a) (b) (c)

그림 7. VGGT 복원 결과의 시점별 및 재투영 이미지 비교 예시, (a) 정면 7시점 기반 복원 결과, (b) 단일 정면 시점 복원 결과, (c) VGGT 재투영 이미지
Fig. 7. Comparison of VGGT reconstruction results by view type and reprojection, (a) Reconstruction from seven frontal views, (b) Reconstruction from a single frontal view, (c) VGGT reprojection image

였다. 이 예시 샘플에서 측정된 PSNR은 38.2dB, SSIM은 0.964로, 매우 높은 구조적 유사성을 보였다.

정량 평가에서는 총 50개 시점을 샘플링하여 평균 PSNR과 SSIM을 산출하고, 배경 포함 여부에 따른 차이를 별도로 분석하였다. 배경을 포함한 경우 평균 PSNR은 34.33dB였으나, 배경을 마스크로 제거하면 38.23dB로 상승하였다. SSIM은 마스크 적용 여부에 관계없이 0.960~0.964 범위를 유지하여 높은 구조적 정합성을 입증하였다. 이러한 결과는 제안 시스템이 다중 시점에서 일관성 있는 3D 재구성을 달성했음을 시사하며, 후속 단계에서의 애니메이션 리타게팅 및 디지털 휴먼 제작에도 적용 가능성을 높인다.

추가적으로, 시점 변화에 따른 복원 품질의 일관성을 확인하기 위해 정면(Frontal)과 측면(Lateral) 시점으로 구분하여 품질을 비교하였다. 표 2는 두 시점별 재투영 정량 평가 결과를 나타내며, 정면 시점에서 PSNR 38.03dB, 측면 시점에서 32.58dB로 약 0.5dB의 차이에 불과하였다. 두 시점 모두에서 SSIM $\geq$ 0.96, LPIPS $\leq$ 0.02로 높은 시각적 일관성을 보였으며, 이는 제안 시스템이 시점 변화에 강건한 복원 품질을 유지함을 의미한다.

표 1. 재투영 기반 정량 평가 결과 (최고/최소 사례)
Table 1. Quantitative results of reprojection-based quality assessment (best and worst cases)

Metric	PSNR (Including Background) [dB]	SSIM	Lpips
Highest	38.032	0.994	0.05747
Smallest	28.438	0.9678	0.009476

표 2. 시점별 복원 품질 비교 결과 (정면 vs 측면)
Table 2. Quantitative results according to camera view (frontal vs lateral)

View Type	PSNR (Including Background) [dB]	SSIM	Lpips
Frontal	38.032	0.994	0.05747
Lateral	32.58	0.968	0.0203

4. 딥러닝 기반 방법과의 비교

최근 발표된 Wang et al.(2025)의 Visual Geometry Grounded Transformer(VGGT)[10]는 단일 또는 소수 시점의 입력 영상으로부터 End-to-End 형태의 3D 얼굴 복원을 수행하는 대표적인 학습 기반 모델이다. 이 접근법은 별도의 구조 재구성 단계 없이 직접 3D 표현을 예측할 수 있다는 장점이 있으나, 대규모 학습 데이터와 GPU 연산 자원에

대한 의존도가 높고, 입력 시점 간의 기하적 제약이 명시적으로 주어지지 않는다는 특성을 가진다[10].

그림 7은 동일 인물의 정면 7시점 영상과 단일 정면 영상에 VGGT를 적용하여 복원한 결과를 비교한 예시이다. 7시점 입력의 경우 단일 시점보다 더 풍부한 표면 정보를 재현하였으나, 학습 기반 모델의 특성상 카메라 외부 파라미터(extrinsic)가 명시적으로 추정되지 않아 시점 간 정합이 완전하지 않았다. 이러한 이유로 재투영 과정에서 일부 얼굴 영역(특히 광택 반사 및 명암 경계부)에서 불균일한 정합 패턴이 관찰되었다.

표 3의 재투영 기반 정량 평가 결과, PSNR은 27.66 – 33.19dB, SSIM은 0.14 – 0.76 범위로 나타났으며, 이는 VGGT 복원이 입력 조건(시점 수, 조명, 포즈)에 따라 품질 변동성을 보인다는 특징을 반영한다.

표 3. 딥러닝 기반 모델의 재투영 기반 정량 평가 결과 (최고/최소 사례)
Table 3. Quantitative results of reprojection-based quality assessment for learning-based model (VGGT) (best and worst cases)

Metric	PSNR (Including Background) [dB]	SSIM
Highest	33.19	0.7555
Smallest	27.66	0.1406

반면, 본 연구의 포토그래메트리 기반 접근은 실제 촬영된 다중 시점 영상과 정밀한 카메라 파라미터를 이용하여 모든 시점에서 재투영 정합을 수행하므로, 실험 전반에서 일관된 정량 품질을 확보하였다(표 2). 이러한 비교는 두 접근법이 지향하는 목표의 차이를 보여주며, 제안 시스템이 실측 기반 정합 절차의 재현성과 표준화 가능성 측면에서 실용적 이점을 제공함을 확인할 수 있다.

V. 결 론

본 연구에서는 멀티뷰 카메라 기반의 포토그래메트리 파이프라인을 설계하고, 카메라 캘리브레이션·동기화·마커 기반 좌표계 정렬·텍스처 재투영·재투영 기반 정량 품질 평가까지 아우르는 통합 워크플로우를 구현하였다. 실험 결과, 배경을 제외한 조건에서 최대 PSNR 38.2dB, SSIM 0.988, LPIPS 0.0193을 달성하여 원본 이미지와 높은 광학적 일관성을 보였다. 전·측면 재투영 비교에서도 윤곽선 및 주요 표정 부위가 정확히 정합됨을 확인하였으며, 표정 마커가 부착된 데이터에서도 안정적인 복원 성능을 확보하였다.

이러한 결과는 제안 파이프라인이 디지털 휴먼 제작, 립싱크 애니메이션, 법과학 감식 등 정밀 3D 얼굴 복원이 필요한 응용에 직접 적용 가능함을 시사한다. 향후 연구에서는 머리카락·투명 부위까지 복원 가능한 볼류메트릭 확장, 다양한 조명·피부 반사 환경에 대한 강건성 개선, 실시간 처리 속도를 위한 파이프라인 최적화 등을 진행할 계획이다. 이를 통해 보다 범용적이고 확장 가능한 3D 얼굴 디지털 트윈 제작 환경을 구축하고자 한다.

참 고 문 헌 (References)

[1] S. Giuliani, F. Tosti, P. Lopes, C. Ciampini, and C. Nardinocchi, “Design of a multi-view photogrammetric system based on low cost cameras for 3D forensic face recognition,” Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., Vol. XLVIII-2/W8, pp. 177 – 183, May 2024.
doi: https://doi.org/10.5194/isprs-archives-XLVIII-2-W8-2024-177-2024

[2] X. Zhou et al., “Subframe-level synchronization in multi-camera system using time-calibrated video,” Sensors, Vol. 24, No. 21, Art. 6975, pp. 1 – 20, Oct. 2024.
doi: https://doi.org/10.3390/s24216975

[3] Z. Liu et al., “TRSP: Texture reconstruction algorithm driven by prior knowledge of ground object types,” ISPRS J. Photogramm. Remote Sens., Vol. 223, pp. 221 – 243, May 2025.
doi: https://doi.org/10.1016/j.isprsjprs.2025.03.015

[4] P. Ruiu et al., “Uniss-MDF: A multidimensional face dataset for assessing face analysis on the move,” Comput. Vis. Image Underst., Vol. 258, Art. 104384, July 2025.
doi: https://doi.org/10.1016/j.cviu.2025.104384

[5] P. Li et al., “A comprehensive Gaussian splatting evaluation system: from geometric consistency to novel view synthesis,” Proc. SPIE, Vol. 13557, Paper 135570P, pp. 1 – 12, Apr. 2025.
doi: https://doi.org/10.1117/12.3062498

[6] S. Feng, X. Wu, and J. Cao, “A survey of multi-view stereo 3D reconstruction algorithms based on deep learning,” Digit. Signal Process., Vol. 165, No. C, pp. 1 – 23, Oct. 2025.
doi: https://doi.org/10.1016/j.dsp.2025.1052

[7] P. Pérez, M. Gangnet, and A. Blake, “Poisson image editing,” ACM Trans. Graph., Vol. 22, No. 3, pp. 313 – 318, 2003.
doi: https://doi.org/10.1145/882262.88226

[8] R. Zhang et al., “The unreasonable effectiveness of deep features as a perceptual metric,” Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 586 – 595, 2018.
doi: https://doi.org/10.1109/CVPR.2018.00068

[9] J. Deng et al., “ArcFace: Additive angular margin loss for deep face recognition,” Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 4690 – 4699, 2019.
doi: https://doi.org/CVPR.2019.00482

[10] J. Wang et al., “VGGT: Visual Geometry Grounded Transformer,” Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 3452 – 3461, 2025.
doi: https://doi.org/10.48550/arXiv.2503.11651

저 자 소 개

이 학 범

- 2024년 2월 : 광운대학교 전자재료공학과학과 졸업(공학사)
- 2024년 3월 ~ 현재 : 광운대학교 전자재료공학과 일반대학원(석사과정)
- ORCID : https://orcid.org/0000-0003-0721-4944
- 주관심분야 : 멀티뷰 카메라 캘리브레이션, 3D 인체 복원, 컴퓨터 비전, 딥러닝

서 영 호

- 1999년 2월 : 광운대학교 전자재료공학과 졸업(공학사)
- 2001년 2월 : 광운대학교 일반대학원 졸업(공학석사)
- 2004년 8월 : 광운대학교 일반대학원 졸업(공학박사)
- 2004년 9월 ~ 2005년 8월 : 한국전기연구원 연구원
- 2005년 9월 ~ 2008년 2월 : 한성대학교 조교수
- 2008년 3월 ~ 현재 : 광운대학교 전자재료공학과 교수
- 2021년 1월 ~ 현재 : 오모션 주식회사 CTO
- 2023년 1월 ~ 현재 : 오모션 아메리카 CEO
- ORCID : https://orcid.org/0000-0003-1046-395X
- 주관심분야 : 컴퓨터 그래픽스, 디지털 휴먼, 홀로그램 압축, 비디오코덱, 시스템반도체설계