특집논문 (Special Paper)
방송공학회논문지 제30권 제6호, 2025년 11월 (JBE Vol.30, No.6, November 2025)
https://doi.org/10.5909/JBE.2025.30.6.1055
ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

멀티해상도 헥스평면 융합 가우시안 스플래팅 기반 다중 인체 상호작용 복원

이 희 경[a)‡], 김 서 연[b)], 임 대 규[b)], 이 주 호[b)], 정 원 식[a)]

Multi-Human Interaction Reconstruction Based on Multi-Resolution HexPlane Integrated Gaussian Splatting

HeeKyung Lee[a)‡], Seoyeon Kim[b)], Dae-Gyu Lim[b)], Joo Ho Lee[b)], and Won-Sik Jeong[a)]

요 약

본 논문은 가우시안 스플래팅(GS)에 멀티해상도 헥스평면 인코딩을 결합한 다중 인체 상호작용 복원 프레임워크를 제안한다. 제안 구조는 정점 법선과 자세 신호를 활용해 전역 형상과 세밀한 외관 표현을 통합적으로 반영한 헥스 피처를 구성하며, 외관 변화 모듈을 통해 기하와 색상 변화를 안정적으로 추정한다. 학습 단계에서는 GeoAvatar의 기하 정합 기반 손실에 더해 MLP 오프셋 제어 손실과 헥스평면 Grid-based Total-Variation 손실을 도입하여 손실 구조를 확장하였다. 이를 통해 과도한 형태 변형을 억제하고, 헥스평면 피처의 시공간적 부드러움과 일관성을 강화하여 전역‑국소 형상의 균형적 재구성을 달성한다. Hi4D 데이터셋 실험에서는 제안 기법이 Multi-GART[1] 및 GeoAvatar[2] 대비 PSNR, LPIPS, P2S 등 주요 지표에서 일관된 성능 향상을 보였으며, 빠른 동작과 빈번한 신체 접촉이 발생하는 복잡한 상호작용 장면에서도 안정적인 복원 성능을 확인하였다. 이를 통해 본 연구는 희소한 8개 시점 입력만으로도 경계 선명도와 세부 표현을 유지하며 복잡한 근접 상호작용을 고정밀로 복원할 수 있음을 입증한다.

Abstract

This paper proposes a multi-human interaction reconstruction framework that integrates multi-resolution HexPlane encoding with Gaussian Splatting (GS). The proposed architecture constructs HexPlane features that jointly capture global geometry and fine-grained appearance by leveraging vertex normals and pose signals, while the appearance variation module reliably estimates geometric and color variations. During training, in addition to the geometry-alignment loss used in GeoAvatar, we introduce an MLP offset regularization loss and a HexPlane grid‑based total variation loss, thereby expanding the overall loss structure. These losses suppress excessive shape distortion and enhance the spatiotemporal smoothness and consistency of the HexPlane features, achieving a balanced reconstruction of global and local geometry. Experiments on the Hi4D dataset demonstrate that the proposed method consistently outperforms Multi-GART [1] and GeoAvatar [2] in key metrics such as PSNR, LPIPS, and P2S, and maintains stable reconstruction performance even in highly dynamic scenes involving rapid motion and frequent physical interactions. This verifies that our approach can faithfully reconstruct complex close-range human interactions with high precision while preserving boundary sharpness and fine details, even when using only eight sparse input views.

Keyword : Gaussian Splatting, Multi-Resolution HexPlane, Dual-Branch Offsets, Spatio-Temporal Losses, Multi-Human Reconstruction

Ⅰ. 서 론

메타버스 아바타, 영화·방송 VFX, 가상 피팅, 스포츠 분석 등에서 4차원 인체 복원은 모션, 형상, 외관을 디지털 공간에 정밀하게 재현하는 핵심 기반 기술이다. 4D 인체 복원 기술들은 여러 시점에서 촬영한 색상 비디오들로부터 프레임별로 사람의 자세를 추정한 후, 이를 바탕으로 대표 자세에 대한 3차원 외관을 복원한다. 이때, 정밀한 복원을 위해 많은 시점의 비디오 촬영이 요구된다. 이는 촬영 시스템의 비용 증가로 이어지며 해당 기술 보편화의 걸림돌로 작용한다. 따라서 적은 수의 카메라로 고품질의 인체 복원을 하는 것은 도전적인 과제이다.

최근 인체 외관에 대한 표현 방식으로 가우시안 스플래팅 방식이 깊이 연구되고 있다. 3D 가우시안 스플래팅은 수많은 3차원 가우시안을 3차원 가상 세계에 분포시킨 후 각 카메라 시점의 관찰된 영상정보와 가우시안들의 밀도 기반 렌더링을 통해 계산된 재구성된 영상정보를 비교하여 각 가우시안의 속성(가우시안의 색상, 밀도, 위치, 모양) 등을 최적화하는 기술이다. 이 기술을 4차원 인체 복원 기술에 적용하기 위해 각 프레임의 추정된 자세를 기준으로 인체 주변 공간을 인체 형상에 따라 왜곡하여 대표 자세로 매핑하여 각 프레임의 외관 단서를 대표 자세를 기준으로 모은다. 하지만, 인체 인근 공간을 거리에 따라서만 왜곡하기에 각 프레임에서의 옷의 외관을 대표 공간에 정확히 정렬할 수 없으며, 이는 흐려진 외관 복원 결과로 이어지게 된다.

a) 한국전자통신연구원(Electronics and Telecommunications Research Institute)
b) 서강대학교 시각컴퓨팅연구실(Visual Computing Lab, Sogang University)
‡ Corresponding Author : 이희경(HeeKyung Lee)
E-mail: lhk95@etri.re.kr
Tel: +82-42-860-6615
ORCID: https://orcid.org/0000-0002-1502-561X
※ 이 논문의 결과 중 일부는 한국방송·미디어공학회 2025년 하계학술대회에서 발표한 바 있음
※This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2018-0-00207, Immersive Media Research Laboratory)
· Manuscript November 4, 2025; Revised November 21, 2025; Accepted November 24, 2025.

4D 가우시안 스플래팅(4DGS[3], Yang et al. ICLR 2024)은 공간 3차원 및 시간 1차원의 4차원 비디오 자료를 4차원 가우시안들을 분포시켜 재구성하는 기술이다. 4DGS는 하나의 공간 좌표계에서 외관을 재구성하여 별도의 자세 추정을 필요로 하지 않는다. 하지만, 인체 중심의 3차원 공간에 외관을 재구성하지 않았기에, 인체 외관 복원 이후 인체에 대한 편집이 불가하다. 또한, 프레임별로 다른 가우시안들이 외관을 표하기에 외관의 일관성 및 디테일이 다소 떨어진다.

이에 본 연구는 최근 제안된 멀티해상도 헥스평면 인코딩(Wu et al. CVPR 2024)[4]을 통해 시공간별 외관 변화를 표현한다. 이후, 제안한 자세 및 법선 기반 오프셋 계산 모듈을 통해 가우시안의 위치 및 색상 변화를 계산한다. 이를 통해, 미세한 자세 차이 및 법선 차이를 인지하여 대표 자세 공간과 프레임별 공간 간의 외관 변화를 정확히 표현하여 적은 카메라 관찰로부터의 4차원 고품질 인체 영상을 복원한다. 본 논문의 주요 기여는 다음과 같다.

멀티해상도 헥스평면 피처 및 자세, 법선 정보 통합: 다중 해상도 헥스평면 피처 표현 방식을 통해 전역 형상과 세밀한 국소 디테일을 학습하고 자세 및 법선 피처를 통합하여 유사한 자세일수록, 유사한 방향일수록 비슷한 변화가 추정되도록 한다.

자세 및 법선 선택적 가우시안 변화 계산 모듈 학습: 가우시안의 각 속성 또는 속성 변화를 물리적 연관성에 따라 자세 또는 법선 정보에 영향을 받도록 추정 모듈의 신경망을 학습한다. 이를 통해 프레임별로 외관 변화를 정확히 계산하여 경계 왜곡과 디테일 손실을 완화한다. 이를 실험으로 입증하였다.

이 통합적 설계를 통해, 희소한 8개 시점 입력으로 자세에 따라 변화하는 객체별 경계, 세밀한 지역 디테일, 안정적인 시공간 표현을 달성한다.

Ⅱ. 관련 연구

3차원 인체 복원은 크게 내재 함수(SDF), 광선 적분(NeRF), 가우시안 분포(GS) 기반으로 나뉜다. 이들 방법은 단일 인체 복원에서는 높은 성과를 보였지만, 다중 인체가

표 1. 3D 인체 복원 접근법 비교
Table 1. Comparison of 3D Human Reconstruction Approaches

Methods	Key Idea & Advantages	Representative Studies	Limitations in Multi-object Reconstruction
Implicit Function (SDF)	Models surfaces as a continuous Signed Distance Function (SDF) from coordinates → provides smooth surfaces and accurate normals	PIFu (2019)[5], PaMIR (2022)[6], MultiPly (2024)[7]	Boundaries between multiple humans often become entangled due to the continuous representation, necessitating separate learning and inference for each object
Ray Integration (NeRF)	Predicts density and color given position + view, then integrates along rays → enables realistic lighting and material representation	Mip-NeRF (2021)[8], HumanNeRF (2022)[9]	Memory consumption and training cost increase exponentially with the number of humans; real-time multi-object processing remains difficult
Gaussian Splatting (GS)	Approximates the scene with many 3D Gaussians on GPU → supports real-time, high-resolution rendering without meshes; provides smooth visual quality	GaussianAvatar (2024)[10], GPS-Gaussian (2024)[11], GART (2024)[1], GeoAvatar (2025)[2]	Overlapping objects cause blur and boundary distortions; the number of Gaussians grows rapidly, making optimization harder; limited ability to simultaneously recover both global shapes and fine local details

근접 상호작용하는 환경으로 확장될 경우 각기 다른 제약이 존재한다. 본 논문은 이를 표 1에서 비교 정리하였다.

표 1에서 보듯이, 내재 함수(SDF)는 경계 혼재 문제를 피하기 어렵고, NeRF 계열은 객체 수가 늘어날수록 연산 비용과 메모리 사용량이 급격히 증가한다. GS는 실시간성과 해상도 측면에서 가장 유망하지만, 다객체가 근접하거나 중첩될 경우 경계 왜곡, 디테일 손실, 시간적 불안정성이 발생한다. 따라서 단일 인체에서 입증된 강점을 다중 인체 상호작용으로 확장하기 위해서는 이러한 제약을 보완할 새로운 설계가 요구된다.

Wu et al. (CVPR 2024)은 기존 GS의 시간축 불연속성과 효율성 문제를 완화하기 위해 헥스플레인 기반 3D 가우시안 스플래팅을 제안하였다[4]. 이 방법은 4차원 입력 (x, y, z, t)을 여섯 개의 2차원 평면 (x,y), (x,z), (y,z), (x,t), (y,t), (z,t)으로 분해하는 헥스플레인 인코딩을 도입하였으며, 멀티해상도 구성을 통해 전역 형상과 국소 디테일을 함께 포착할 수 있도록 설계되었다. 본 논문은 해당 기법을 기반으로 가우시안의 자세별 및 법선별 외관 모듈을 학습하여 프레임별 외관을 정확히 복원하고자 한다.

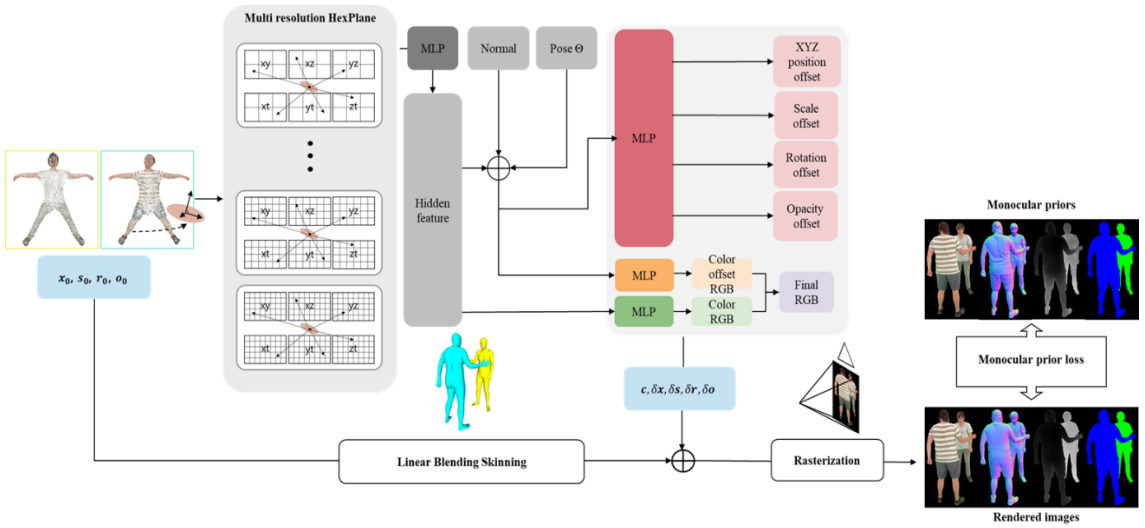

그림 1. 멀티해상도 헥스평면 융합 GS 네트워크 파이프라인
Fig. 1. Multi-Resolution HexPlane Integrated GS Pipeline for Multi-Human Interaction Reconstruction

Ⅲ. 멀티해상도 헥스평면 융합 GS 기반 다중 인체 상호작용 복원

본 절에서는 그림 1에 제시된 멀티해상도 헥스평면 융합 가우시안 스플래팅(GS) 프레임워크의 핵심 구성과 학습 설계를 설명한다. 제안 프레임워크는 제한된 다시점 입력(8개 시점)만으로도 복수 인체의 근접 상호작용이 포함된 복잡한 장면을 정밀하게 복원하도록 설계되었으며, 데이터 준비/처리 흐름(Ⅳ절)과 구분해 네트워크 구조와 손실 함수에 초점을 맞춘다.

제안 프레임워크는 기존 GS 기반 파이프라인(GeoAvatar[2], CVPR et al. 2025)을 토대로, 색상·깊이·법선·인스턴스 맵을 사전 단서로 활용하고 SMPL-X 기반 T-pose를 대표 자세 공간에 초기화한 뒤, 가우시안 단위 변형과 래스터화를 거쳐 멀티뷰 합성 영상을 생성한다.

본 연구의 차별성은 이러한 표준 GS 기반 절차에 새로운 표현 구조와 학습 전략을 도입한 데 있다. 이후 절에서는 (1) 멀티해상도 헥스평면 기반의 헥스 피처 형성, (2) 자세 및 법선 기반 가우시안 변화 계산, (3) 학습 목적 함수를 차례로 설명한다.

1. 헥스 피처 형성: 멀티해상도 인코딩과 자세/법선 정보 통합

본 연구의 Multi-resolution 헥스평면은 전역 형상과 지역 디테일을 동시 학습하도록 설계되었다. 입력 4차원 좌표 (x, y, z, t)는 여섯 개의 이차원 평면 (x,y), (x,z), (y,z), (x,t), (y,t), (z,t)으로 분해되어 헥스평면 인코딩을 거친다. 본 연구는 각 평면에 기본 해상도(base resolution, 32)를 설정하고 다중 해상도 계수 [1, 2, 4, 8]을 적용하였다. 이는 물리적 격자 해상도를 높이는 방식이 아니라, 동일한 4D 좌표에 대해 서로 다른 스케일(정규화 간격)을 갖는 피처 평면을 구성함으로써 피처 공간 상에서 전역적(저주파) 구조와 지역적(고주파) 세부 정보를 동시에 학습하는 의미적 다중 해상도 표현(multi-scale semantic representation)을 구현한 것이다. 낮은 해상도의 평면은 인체의 전체 형상, 포즈, 실루엣 등 전역 구조(저주파 성분)를 주로 반영하고, 높은 해상도의 평면은 의복 주름, 머리카락, 얼굴 윤곽 등 지역적 디테일(고주파 성분)을 보완한다. 이렇게 생성된 다중 해상도 피처는 head MLP에서 융합(fusion)되어, 서로 다른 스케일 간의 상호작용을 학습함으로써 단순한 병합(concatenation)을 넘어 전역 형상 복원과 세부 디테일 표현 간의 균형을 이룬다. 모든 해상도 피처는 동일한 학습 체계 내에서 공동 최적화되며, 별도의 손실 분리나 독립 학습 경로는 존재하지 않는다. 최종적으로 융합된 피처는 은닉 피처로 투영된 뒤, 정점별 법선과 자세 정보를 결합하여 헥스 피처를 형성한다. 이 헥스 피처는 기하 정보와 자세 정보를 포함하는 확장 표현으로, 이후 외관 변화 계산 모듈의 입력으로 사용된다. 그림 1에서도 여섯 개의 평면 구조를 유지하되, 각 평면 내부에 [1, 2, 4, 8]의 다중 해상도 계층이 존재하며, 이를 통해 공간적·주파수적 정보가 계층적으로 융합되는 과정을 시각적으로 제시하였다.

2. 자세 및 법선 선택적 가우시안 변화 계산

형성된 헥스 피처는 자세 정보와 법선 정보와 함께 외관 변화 계산 모듈에 입력으로 들어간다. 이를 통해 자세별로 달라지는 가우시안의 모양, 위치, 색상을 계산한다. 입력으로 법선 정보를 같이 넣어줌으로써, 다른 법선 정보를 갖는 가우시안에 대해 계산 결과가 다르도록 유도하였다. 상세히 말하자면, 외관 계산 모듈에서는 포즈 변화에 따른 위치·스케일·회전 변형을 학습적으로 추정하며, 포즈 변환 후의 가우시안 법선과 포즈 정보를 입력으로 사용하여 대표 자세 공간에서 위치 및 스케일 변화를 정밀하게 보정한다. 색상의 경우, 색상(RGB_{base})과 포즈 변화에 따른 색상 편차(RGB_{offset})를 분리하여 학습한다. RGB_{base}는 헥스평면 인코딩 피처를 입력으로 하여 포즈와 무관한 고유 색상 성분을 모델링하고, RGB_{offset}은 포즈 변환 후의 법선 및 포즈 정보를 추가 입력으로 활용해 동적 색상 변화를 정교하게 반영한다. 본 연구에서는 색상 변화 계산 모듈과 기하 변화 계산 모듈을 달리두어 학습하였고, 분리 시 개선되었음을 확인하였다.

3. 학습 목적 함수

본 연구의 손실 함수는 기존 GeoAvatar[2]의 기본 손실 구

성(L_{base})에 MLP 변화 손실(L_{offset})과 헥스평면 Grid-based Total-Variation 손실(L_{tv})을 추가하여 확장된 형태로 설계하였다.

각 손실의 목적과 역할은 다음과 같다.

- GeoAvatar Loss (L_{base})
  기존 GeoAvatar에서 사용한 기하 정합 기반 손실로, 입력 영상의 기하 및 색상 정보를 정확히 복원하도록 유도한다. L_{base}는 광도 손실(L_c), 표면 정렬 손실(L_{so}), 단안 기하 손실 (L_d, L_n), 그리고 법선·깊이·프레임 단위 일관성 항(L_m, L_{rd}, L_{reg})으로 구성된다. 특히 경계 모호화와 침투 현상은 헥스평면과는 별도로, GeoAvatar의 기하 정합 학습과 표면 정렬 손실 (surface alignment loss)을 통해 완화된다.

$$L_{base}= \lambda_c L_c+\lambda_{so} L_{so}+\lambda_d L_d+\lambda_n L_n+\lambda_{rn} L_{rn}+\lambda_{rd} L_{rd}+\lambda_{reg} L_{reg}$$

- MLP Offset Loss (L_{offset})
  MLP가 출력하는 offset 값의 크기를 제어하여 불필요한 변형을 방지하고, 대표 자세 공간의 기하 구조를 안정적으로 유지한다.

$$\lambda_{offset} L_{offset}= ||\delta_\mu, \delta_r, \delta_s, \delta_o||^2$$

- 헥스평면 Grid-based Total-Variation Loss (L_{tv})
  Wu et al.의 Total-Variation 손실 함수로, 헥스평면 피처 격자에서 인접 격자 간의 과도한 값 변화를 억제하여 피처 공간의 불연속성과 노이즈를 줄이고 시공간적 부드러움을 유도한다. 이를 통해 비슷한 외관이 비슷한 피처로 표현되도록 하여 피처 공간과 외관 공간의 유사성이 높아지도록 한다.
  본 손실은 다음 세부 항으로 구성된다.
  · L_{plane}: 평면 내 인접 픽셀 간의 2차 차분을 최소화해 공간적 노이즈를 억제하고 피처 필드의 부드러움을 유지한다.
  · L_{time}: 연속된 프레임 간 시간축 방향의 2차 차분을 최소화해 시공간적 일관성을 보장한다.
  · L_{l1time}: 시간 평면 전체 값을 1에 근접하게 유지해 과도한 시간적 변조나 누적 오프셋을 방지한다.

이상의 세 손실 항은 각각 시공간적 안정성, 표면 연속성, 그리고 학습 안정성 확보에 기여하며, 가중치 (λ)는 경험적 탐색을 통해 균형 있게 설정하였다. 최종 손실 함수는 다음과 같다.

$$L=L_{base}+\lambda_{offset} L_{offset}+\lambda_{tv} L_{tv}$$

Ⅳ. 데이터 처리 흐름

본 절은 앞서 제안한 네트워크 설계(Ⅲ절)와 달리, 실제 데이터 준비 - 전처리 - 표현 모델링 - 렌더링 - 메시 생성으로 이어지는 상위 수준의 전체 시스템 파이프라인을 설명한다. 즉, 다중 카메라 입력으로부터 최종 3차원 메시를 생성하기까지의 절차를 단계적으로 요약하며, 전체 프레임워크의 데이터 흐름을 정리한다. 데이터 처리 흐름은 그림 2와 같이 데이터 정렬 및 포즈 추정 - 표현 모델링 및 렌더링 - 학습 및 메시 생성의 세 단계로 구성된다.

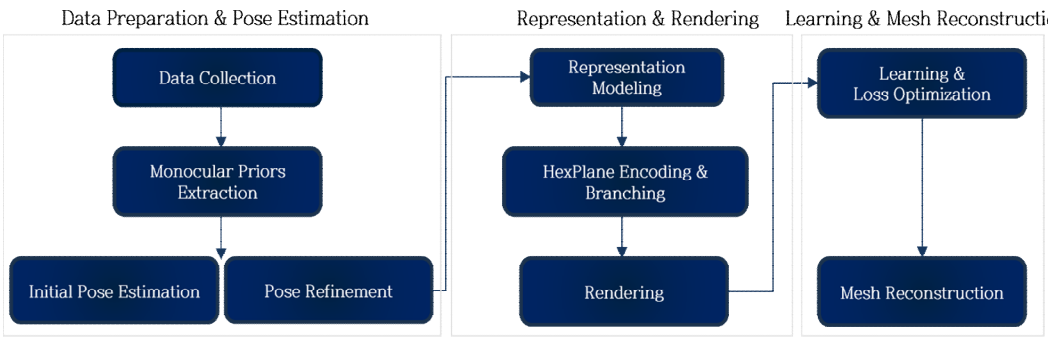

그림 2. 멀티해상도 헥스평면 기반 GS 다중 인체 상호작용 복원 데이터 처리 흐름도
Fig. 2. Data Processing Flow for Multi-Human Interaction Reconstruction Based on Multi-Resolution HexPlane

- 데이터 정렬 및 포즈 추정: 다중 카메라의 캘리브레이션 정보와 타임스탬프 동기화를 통해 시퀀스를 정렬한다. 이후 단일 뷰 네트워크를 이용해 깊이·법선·인스턴스 맵을 추출하고, 이를 기반으로 멀티뷰 키포인트를 3D로 역추정하여 SMPL-X 템플릿을 초기화한다. 초기화된 모델은 반복 정제 과정을 거쳐 프레임 단위 자세 정확도가 향상된다.
- 표현 모델링 및 렌더링: 대표 자세 공간에서 T-pose 기반 평면 가우시안 표현을 구성하고, 멀티해상도 헥스평면 인코딩을 적용한다. 이어서 정점별 법선과 자세 신호를 결합하여 헥스 피처를 형성하고, 자세 및 법선 선택적 가우시안 변화 계산 모듈을 통해 기하와 색상 표현을 보정한다. 렌더링 단계에서는 깊이 정렬 기반 단일 공간에서 모든 객체를 처리하는 볼륨 렌더링을 통해 경계 선명도와 파이프라인 효율성을 모두 확보한다.
- 학습 및 메시 생성: 광도·깊이·법선·인스턴스 손실과 프레임 단위 정규항을 통해 표면 기하의 일관성과 안정성을 확보한다. 여기에 Ⅲ절에서 제안한 MLP Offset 손실(Loffset)을 도입하여 전역 - 국소 형상 균형, 시공간 연속성, 다객체 간 물리적 타당성을 강화한다. 마지막으로 인체 중심 구면 샘플링을 적용하여 가려진 영역을 보완함으로써, 희소한 입력 환경에서도 완전한 3D 인체 메시를 안정적으로 복원한다. 학습 시, 프레임별 자세 또한 같이 최적화한다.

이와 같은 처리 흐름을 통해 희소한 8개 시점 입력만으로도 안정적이고 완전한 다중 인체 상호작용 복원이 가능함을 확인하였다.

Ⅴ. 실험 결과

본 연구의 성능 평가는 ETH Zurich에서 구축하여 CVPR 2023에 공개된 Hi4D 데이터셋[12]을 활용하였다. Hi4D는 두 인체가 실제로 접촉하며 움직이는 고난도 상호작용 장면을 8개 시점(940×1280 RGB)으로 제공하며, 뷰별 RGB 프레임, 인스턴스 마스크, 카메라 파라미터뿐 아니라 프레임 단위 SMPL-X 파라미터와 메시를 포함한다. 특히 근접 접촉, 폐

표 2. 정량 비교
Table 2. Quantitative Comparison

Method	Sequence	Visual Quality			Geometric Accuracy	
		PSNR↑	SSIM↑	LPIPS↓	P2S↓	CD↓
Multi-GART[1]	Hug01	28.6832	0.9532	0.0647	1.6600	2.0785
	Football21	25.5343	0.9411	0.0733	1.3762	2.0795
	Highfive13	25.4589	0.9358	0.0777	1.3804	1.9556
	Fight21	24.0991	0.9311	0.0873	1.6679	2.0016
	Basketball28	24.7084	0.9351	0.0823	1.6942	2.2254
	Average	25.2766	0.9399	0.0769	1.5556	2.0281
GeoAvatar[2]	Hug01	30.9779	0.9629	0.0532	0.5373	0.6581
	Football21	27.4125	0.9492	0.0633	0.5645	0.5943
	Highfive13	26.8472	0.9444	0.0697	0.7873	0.9118
	Fight21	25.3247	0.9380	0.0771	0.9252	0.9979
	Basketball28	26.4895	0.9454	0.0682	0.7311	0.8347
	Average	27.4109	0.9481	0.0670	0.7090	0.7992
Ours	Hug01	31.7620	0.9670	0.0447	0.4667	0.6627
	Football21	28.3924	0.9536	0.0526	0.4610	0.5143
	Highfive13	28.4416	0.9521	0.0535	0.4452	0.5674
	Fight21	28.2064	0.9504	0.0581	0.4402	0.4935
	Basketball28	27.6560	0.9517	0.0559	0.5498	0.6814
	Average	28.8914	0.9547	0.0529	0.4623	0.5834
Perf. gain (%)	M-G vs Geo	+8.44	+0.87	+12.9	+54.4	+60.6
	M-G vs Ours	+14.3	+1.57	+31.2	+70.3	+71.2
	Geo vs Ours	+5.4	+0.70	+21.1	+34.8	+27.0

색, 경계 혼재가 빈번히 발생하는 장면으로 구성되어 있어, 본 연구가 제안하는 프레임워크의 경계 선명도, 시공간 연속성, 기하 정확도를 종합적으로 검증하기에 적합하다.

Hi4D 데이터셋은 다양한 다중 인체 상호작용 시나리오로 구성되어 있으나, 시퀀스별로 상호작용 형태와 동작 복잡도가 상이하다. 본 연구에서는 기하적 정합성과 시각적 재현력을 균형 있게 검증하기 위해, Hug01, Football21, Highfive13, Basketball28, Fight21의 총 5개 시퀀스를 평가 대상으로 선정하였다. 이들 시퀀스는 밀접한 신체 접촉(Hug01), 빠른 동작(Football21, Basketball28), 비접촉 상호작용(Highfive13), 격렬한 움직임(Fight21) 등 대표적인 2인 상호작용 유형을 포함한다. 이를 통해 제안 방법의 다양한 동적 상황에서의 복원 성능과 강인성을 효율적으로 검증하였다.

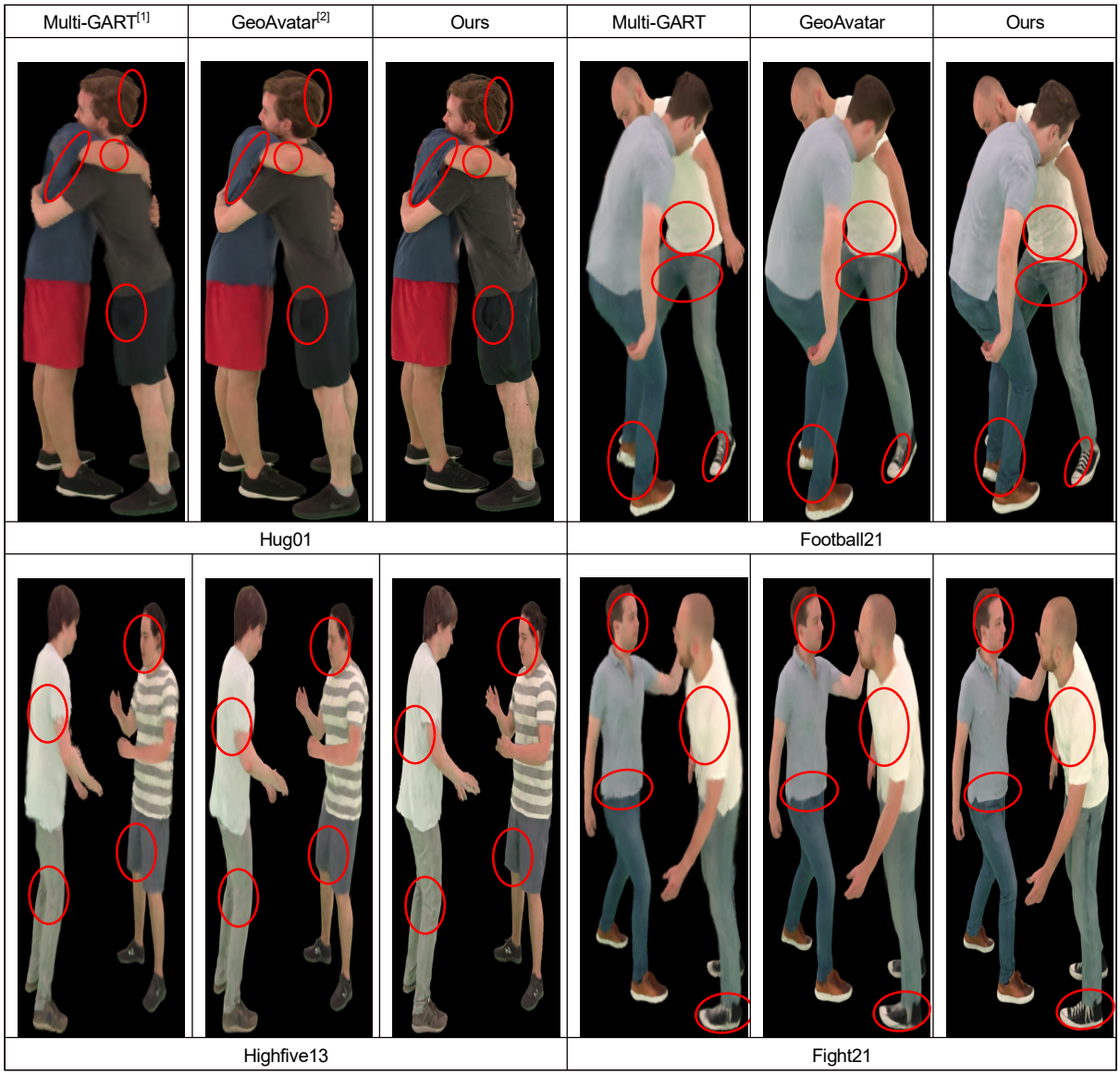

그림 3. Hug01, Football21, Highfive13, Fight21 정성 비교
Fig. 3. Qualitative comparison on Hug01, Football21, Highfive13, and Fight21

표 2는 제안한 방법(Ours)을 기존 기법인 Multi-GART[1] 및 최신 기법인 GeoAvatar[2]와 비교한 정량 평가 결과를 나타낸다. 제안 방법은 Multi-GART[1] 대비 PSNR +3.61dB, SSIM +1.48%, LPIPS −31.2%, P2S −70.3%, CD −71.2%로, 시각 품질(Visual Quality)과 기하 정확도(Geometric Accuracy) 모두에서 현저한 성능 향상을 보였다. 또한 GeoAvatar[2]와 비교했을 때에도 PSNR +5.4%, SSIM +0.70%, LPIPS −21.1%, P2S −34.8% 개선을 보였으며, CD 지표에서는 약 +27.0%의 소폭 열세를 보였으나, 전반적인 정량 평가 지표에서 제안 방법이 우수한 결과를 유지하였다. 특히 빠른 동작이 포함된 Football21 시퀀스에서도 PSNR 28.39dB, P2S 0.461mm를 기록하여, 동적 상호작용 장면에서도 높은 복원 정확도와 강인성을 입증하였다.

그림 3은 Hi4D 데이터셋의 Hug01, Football21, Highfive13, Fight21 시퀀스에 대해 제안 프레임워크를 적용한 정성 비교 결과를 보여준다. 다양한 포즈와 접촉이 포함된 장면에서도 인체 간의 세밀한 형상과 질감을 사실적으로 재현하며, 빈번한 접촉 상황에서도 뚜렷한 형상 경계와 안정적인 시공간 표현을 유지하는 것을 확인할 수 있다.

표 3은 각 모듈에 대한 소거 실험 결과를 나타낸다. 단일 해상도의 헥스평면 피처만 사용 시 기하 및 렌더링 결과가 다소 하락함을 확인하였다. 외관 변화 모듈을 제거 시 재구성 품질이 급격히 떨어진다. 이는 프레임에 따라 자세가 달라지면서 기하 표면이 위치에 따라 다르게 변화하는데 이를 고려하지 못하기 때문이다. TV 손실 함수의 경우, 제거 시 과적합으로 인해 색상/기하 품질이 소폭 상승함을 확인

표 3. 모듈별 소거 실험 정량 비교
Table 3. Quantitative Comparison of Module-wise Ablation Study

	RENDER			MESH		NORMAL	
	PSNR↑	SSIM↑	LPIPS↓	P2S↓	CHAMF↓	COS↓	L2↓
w/o multires	27.5374	0.9472	0.0602	0.535643	0.702941	0.0147	0.0419
w/o offset	23.4771	0.9217	0.0720	1.021527	1.115230	0.0238	0.0566
w/o tv	28.3712	0.9514	0.0545	0.454328	0.566688	0.0135	0.401
ours	28.3232	0.9513	0.0538	0.455929	0.575290	0.0135	0.401

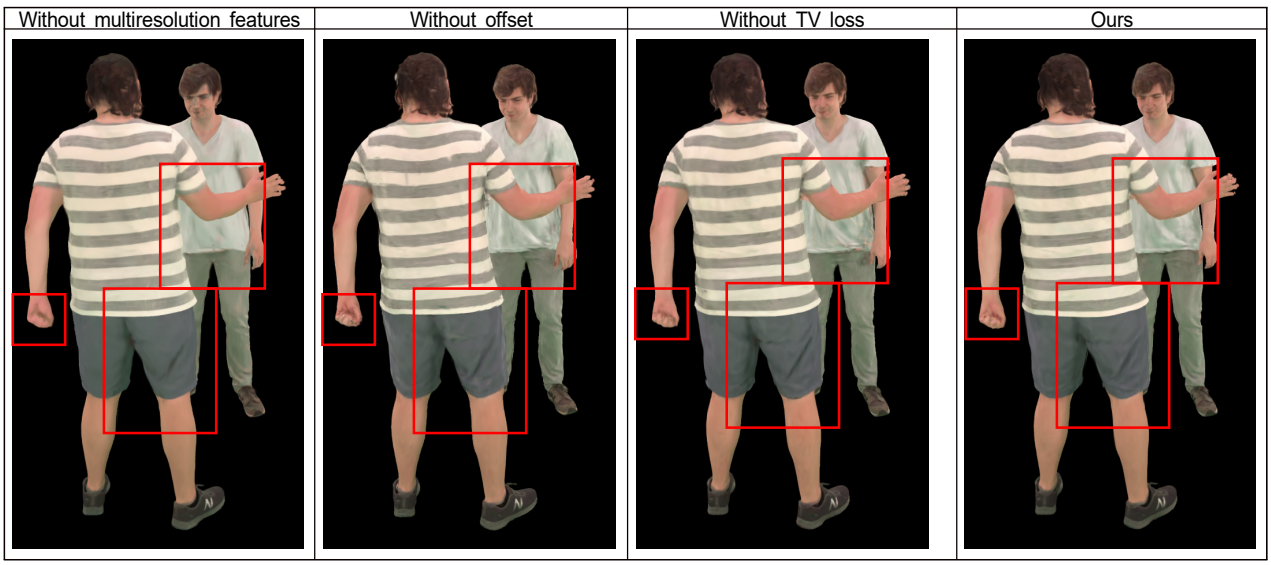

그림 4. 모듈별 소거 실험 질적 비교
Fig. 4. Qualitative Results of Module-wise Ablation Study

하였다. 하지만, 피처의 연속성을 지향하지 않기에 연속된 프레임에서 피처의 차이가 클 경우 중간 프레임 복원 시 가우시안 외관 변화 모듈이 학습하지 않은 입력 피처가 생성될 수 있어 안정성이 떨어진다.

그림 4는 소거 실험의 질적 결과를 나타낸다. 단일 해상도의 피처만 사용 시 손, 머리카락 등의 디테일이 복원되지 않는다. 외관 변화를 고려하지 않을 시(without offset) 프레임이 달라질 때 가우시안들이 적절히 조건화 되지 않아 외관 변화를 표현하지 못해 품질이 하락했다. TV 손실 함수를 배제 시(without TV loss) 정량적으로는 좋은 수치가 나왔으나 질적으로 보았을 때 가우시안 특유의 모양이 잘 드러남을 확인했다. 우리 방법(ours)은 상대적으로 부드러운 외관을 복원함을 확인하였다.

제안 프레임워크의 연산 효율성을 정량적으로 분석한 결과는 다음과 같다. NVIDIA RTX A6000 환경에서 학습에는 약 5–6시간이 소요되었으며, 학습 완료 후 모델 크기는 약 750MB, GPU 메모리 사용량은 약 7GB 수준이었다. 추론(inference) 단계에서는 약 11–12FPS의 렌더링 속도를 보였으며, 이는 MLP 기반 아바타 모델링 기법의 평균적인 실시간 처리 수준과 유사하다. 참고로, 유사한 구조를 사용하는 Animatable Gaussian에서도 약 10FPS가 보고된 바 있다. 따라서 제안 프레임워크는 기존 GS 계열 연구와 동등한 수준의 실시간 렌더링 성능을 유지하면서도, 다중 인체 상호작용 상황에서의 기하 정합성과 시공간 일관성을 동시에 확보하였다.

Ⅵ. 결 론

본 연구는 가우시안 스플래팅 기반 인체 복원 기법을 멀티해상도 헥스평면 인코딩과 법선 정보와 자세 정보와의 헥스 피처 통합하여 자세 및 외관별 외관 변화를 가능케 하였다. 특히 자세 및 법선 선택적 가우시안 변화 계산 모듈을 통해 물리적 연관성이 있는 가우시안 속성과 기하 정보만 선택적 영향을 받도록 하여 보다 일관된 인체 복원이 가능하게 하였다.

Hi4D 데이터셋을 활용한 실험 결과, 제안 프레임워크는 기존 기법 대비 시각 품질과 기하 정확도에서 일관된 향상을 보였으며, 빠른 동작이 포함된 시퀀스에서도 안정적 성능을 유지하였다. 이는 멀티해상도 헥스평면 표현과 확장 손실 설계가 다중 인체 상호작용 복원에 핵심적으로 기여함을 실증적으로 확인한 것이다.

향후 연구에서는 본 프레임워크를 3인 이상 다중 인체 장면과 배경을 포함한 전신 복원으로 확장하여, 보다 복잡한 실세계 응용에 적합한 통합적 솔루션을 제시하고자 한다.

참 고 문 헌 (References)

[1] Lei, Jiahui, et al. “Gart: Gaussian articulated template models,” in Proc. CVPR, pp. 19876–19887, 2024.
doi: https://doi.org/10.1109/CVPR52733.2024.01879

[2] Lee, Soohyun, et al. “GeoAvatar: Geometrically-Consistent Multi-Person Avatar Reconstruction from Sparse Multi-View Videos,” in Proc. CVPR, pp. 21138–21147, 2025.

[3] Yang, Zeyu, et al. “Real-time photorealistic dynamic scene representation and rendering with 4d gaussian splatting,” in Proc. ICLR, 2024.

[4] Wu, Guanjun, et al. “4d gaussian splatting for real-time dynamic scene rendering,” in Proc. CVPR, pp. 20310–20320, 2024.

[5] Saito, Shunsuke, et al. “Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization,” in Proc. ICCV, pp. 2304–2314, 2019.
doi: https://doi.org/10.1109/ICCV.2019.00239

[6] Zheng, Zerong, et al. “Pamir: Parametric model-conditioned implicit representation for image-based human reconstruction,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 6, pp. 3170–3184, 2022.
doi: https://doi.org/10.1109/TPAMI.2021.3050505

[7] Jiang, Zeren, et al. “Multiply: Reconstruction of multiple people from monocular video in the wild,” in Proc. CVPR, pp. 109–118, 2024.

[8] Barron, Jonathan T., et al. “Mip-NeRF: A multiscale representation for anti-aliasing neural radiance fields,” in Proc. ICCV, pp. 5855–5864, 2021.
doi: https://doi.org/10.1109/ICCV48922.2021.00580

[9] Weng, Chung-Yi, et al. “Humannerf: Free-viewpoint rendering of moving people from monocular video,” in Proc. CVPR, pp. 16210–16220, 2022.

[10] Hu, Liangxiao, et al. “Gaussianavatar: Towards realistic human avatar modeling from a single video via animatable 3d gaussians,” in Proc. CVPR, pp. 634–644, 2024.

[11] Zheng, Shunyuan, et al. “Gps-gaussian: Generalizable pixel-wise 3d gaussian splatting for real-time human novel view synthesis,” in Proc. CVPR, pp. 19680–19690, 2024.

[12] Yin, Yifei, et al. “Hi4d: 4d instance segmentation of close human interaction,” in Proc. CVPR, pp. 17016–17027, 2023.
doi: https://doi.org/10.1109/CVPR52729.2023.01632

저 자 소 개

이 희 경

- 1999년 : 영남대학교 컴퓨터공학과(공학사)
- 2002년 : 한국정보통신대학교(ICU) 공학부(공학석사)
- 2002년 ~ 현재 : 한국전자통신연구원 책임연구원
- ORCID : https://orcid.org/0000-0002-1502-561X
- 주관심분야 : 컴퓨터 비전, 기계학습, VCM, 메타버스, 360VR, 시선추적, 메타데이터

김 서 연

- 2024년 : 서강대학교 컴퓨터공학(학사)
- 현재 : 서강대학교 인공지능학 석사과정
- ORCID : https://orcid.org/0009-0001-4977-0810
- 주관심분야 : 컴퓨터비전, 컴퓨터 그래픽스

임 대 규

- 2025년 : 서강대학교 전자공학/컴퓨터공학(학사)
- 현재 : 서강대학교 인공지능학 석사과정
- ORCID : https://orcid.org/0009-0003-8001-716X
- 주관심분야 : 컴퓨터비전, 컴퓨터 그래픽스

이 주 호

- 2012년 : 한국과학기술원 전산학과(학사)
- 2014년 : 한국과학기술원 전산학부(석사)
- 2020년 : 한국과학기술원 전산학부(박사)
- 현재 : 서강대학교 컴퓨터공학과 교수
- ORCID : https://orcid.org/0000-0001-7307-7744
- 주관심분야 : 컴퓨터 그래픽스, 컴퓨터 비전, 시각 컴퓨팅, 3차원 재구성

정 원 식

- 1992년 : 경북대학교 전자공학과(공학사)
- 1994년 : 경북대학교 대학원 전자공학과(공학석사)
- 2000년 : 경북대학교 대학원 전자공학과(공학박사)
- 2000년 ~ 현재 : 한국전자통신연구원 책임연구원
- ORCID : https://orcid.org/0000-0001-5430-2969
- 주관심분야 : 이머시브 미디어 기술, 기계를 위한 영상 부호화, 딥러닝기반 신호처리, 멀티미디어 표준화