

일반논문 (Regular Paper)

방송공학회논문지 제30권 제6호, 2025년 11월 (JBE Vol.30, No.6, November 2025)

<https://doi.org/10.5909/JBE.2025.30.6.1065>

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

## 사람 자세 코드북을 이용한 사용자 신체 메쉬 복원

성 강 윤<sup>a)</sup>, 김 원 준<sup>a)†</sup>

## Ego-body Mesh Reconstruction via Human Pose Codebook

Kangyun Sung<sup>a)</sup> and Wonjun Kim<sup>a)†</sup>

### 요 약

최근 웨어러블 기기의 사용이 증가함에 따라 다양한 분야에서의 활용을 위해 사용자의 자세를 정확하게 파악하는 기술이 필수적인 요소로 부상하고 있다. 하지만 머리 움직임 정보만으로 몸 전체의 자세를 예측하는 것은 매우 어려운 일이다. 이를 해결하고자 대부분의 사용자 신체 메쉬 복원(Ego-body Mesh Reconstruction) 연구들은 보이지 않는 몸의 부분을 생성하기 위해 확산 모델(Diffusion Model) 기반의 구조를 선택하였지만, 반복적인 추론 과정으로 인해 실시간 적용에 한계가 있었다. 본 논문은 이러한 한계를 극복하기 위해, 확산 모델과 달리 단 한 번의 스텝으로 사용자 신체 메쉬를 복원하는 코드북(Codebook) 기반의 새로운 방법을 제안한다. 기준 데이터셋(Benchmark Dataset)에서의 성능 평가를 통해 제안하는 방법이 빠른 추론 속도와 정확한 자세 예측에 효과적임을 보였다.

### Abstract

The increasing use of wearable devices has made it essential to accurately understand and analyze postures of users for various real-world applications. However, reconstructing full-body poses from head motions alone remains a significant challenge. To address this, several studies have adopted diffusion-based architectures to generate unseen body parts, however, their iterative inference process has limited their real-time applications. In this paper, we propose a codebook-based architecture that reconstructs a 3D mesh in a single step. Experimental results on the benchmark dataset demonstrates that the proposed method achieves both fast inference speed and improved performance of ego-body mesh reconstruction.

Keyword : Ego-body Mesh Reconstruction, Human Pose Codebook

a) 건국대학교 전기전자공학부(Department of Electrical and Electronics Engineering, Konkuk University)

† Corresponding Author : 김원준(Wonjun Kim)

E-mail: wonjkim@konkuk.ac.kr

Tel: +82-2-450-3396

ORCID: <https://orcid.org/0000-0001-5121-5931>

※ 본 논문은 2025년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2018-0-00207, RS-2018-II180207, 이며시브 미디어 전문연구실)

· Manuscript September 19, 2025; Revised October 20, 2025; Accepted October 20, 2025.

## I. 서 론

최근 헤드 마운트 장치 또는 스마트 글래스 등 웨어러블 기기에 대한 수요가 증가하면서 자기중심(Egocentric) 센서로부터 3차원 공간을 이해하는 것<sup>[1,2,3]</sup>이 중요한 문제가 되었다. 특히, 사용자의 자세나 움직임<sup>[4,5,6,7]</sup>을 파악하는 것은 다양한 분야에서 활용할 수 있음에 따라 많은 연구자에게 큰 관심을 받고 있다. 이에 따라 자기중심 센서로부터 사용자의 3차원 사람 메쉬를 복원하는 연구가 활발하게 진행되고 있다. 가장 핵심이 되는 부분은 적은 입력 신호로부터 보이지 않는 몸의 메쉬까지 복원해야 하는 것이다. 이러한 어려움을 해결하기 위해 사람 모션에 대한 강력한 사전 정보를 활용하는 다양한 방법들이 제안되었다. 구체적으로, Jiang<sup>[4]</sup> 등은 두 손과 머리 모션 정보를 인코딩하고 이를 전신 모션으로 디코딩하는 간단한 신경망을 대규모 모션 데이터를 통해 학습시키는 방법을 제안하였다. 하지만, 이 방식은 여전히 보이지 않는 하반신 예측에 취약하다는 한계가 있었다. 이를 해결하기 위해, Castillo<sup>[5]</sup> 등은 손과 머리 모션을 입력으로 하는 조건부 확산 모델<sup>[8]</sup>(Conditional Diffusion Model)을 통해 상반신뿐만 아니라 더 그럴듯한 하반신 자세를 생성하여 정확도를 향상시켰다. 가장 최근, ProjectAria<sup>[9]</sup>와 같은 스마트 글래스의 등장으로 머리 움직임 단독 입력으로부터 정확한 전신 자세 예측에 대한 연구 또한 활발하게 되고 있다. 예를 들어, Li<sup>[6]</sup> 등은 자기중심 비디오 영상으로부터 머리 움직임을 옵티컬 플로우(Optical Flow) 기반의 신경망과 위치 측정 및 동시 지도화<sup>[10]</sup>(Simultaneous Localization and Mapping) 알고리즘을 통해 예측하고 이를 전신 움직임을 예측하기 위한 확산 모델의 입력으로 사용하였다. 이에 따라 다양한 고품질의 전신 모션을 생성할 수 있었지만, 학습 데이터셋의 부족한 다양성으로 일반화 성능에 문제가 있었다. 구체적으로, 현재 존재하는 학습 데이터셋(예, AMASS<sup>[11]</sup>)의 경우 물리적으로 제한된 공간에서 얻어졌기 때문에 긴 궤적의 움직임 등에 취약하였다. 이를 해결하기 위해 Yi<sup>[7]</sup> 등은 입력으로 사용되는 머리의 움직임을 시공간 불변의 표현으로 나타내는 방법을 제안하고 이를 조건부 확산 모델의 입력으로 사용하여 일반화 성능 향상을 이뤄내기도 하였다. 위처럼 다양한 방법들이 주목할 만한 성능을 보였지만, 확산 모델의 반

복적인 복원 절차는 시간을 많이 소요하는 근본적인 문제는 실제 활용에 큰 제약이 되고 있다.

본 논문에서는 머리 움직임으로부터 3차원 사람 메쉬를 복원하는 코드북<sup>[12]</sup>(Codebook) 기반의 네트워크를 제안한다. 제안하는 방법의 핵심은 사람의 자세에 대해 학습된 코드북에서 주어진 머리 움직임에 적합한 자세 토큰을 찾는 것이다. 이러한 절차는 한 번의 단계만으로도 정확한 자세의 사용자 신체 메쉬를 복원할 수 있게 하며 기존의 확산 모델 기반의 방법에 비해 빠른 추론 속도를 가질 수 있게 한다. 기준 데이터셋(Benchmark Dataset)에 대한 성능 평가를 통해 기존 방법 대비 제안하는 방법이 기존 확산 모델 기반 방법들 대비 더 빠른 추론 속도와 정확한 자세 예측에 효과적임을 보였다.

본 논문의 구성은 다음과 같다. 2장에서는 제안하는 구조에 대해 자세히 설명하며, 3장에서는 다양한 실험을 통해 제안하는 방법이 기존 방법들보다 효과적임을 검증한다. 마지막으로 4장에서는 본 논문의 결론을 서술한다.

## II. 제안하는 방법

그림 1은 제안하는 신경망의 구조를 보여준다. 제안하는 방법은 두 개의 단계로 구성되어 있다. 첫 번째 단계에서는 사람의 전신 자세 정보를 학습하기 위한 코드북을 학습한다. 두 번째 단계에서는 이전 단계에서 학습된 코드북을 활용하여 머리 움직임으로부터 3차원 사람 메쉬를 복원한다.

### 1. 사람 자세 코드북

제안하는 방법은 VQ-VAE<sup>[13]</sup> 구조를 활용하여 3차원 자세를 이산적 표현으로 인코딩하는 방법을 학습한다. 전체 구조는 그림 1의 위쪽 부분에서 볼 수 있다. 입력은 Skinnd Multi-person Linear(SMPL)<sup>[14]</sup>의 자세 파라미터  $\theta \in R^{126}$ 이고 인코더를 통해 자세 토큰  $z \in R^{M \times d_c}$ 이 계산된다. 여기서  $M$ 과  $d_c$ 는 전체 자세 토큰의 수와 각 토큰의 차원이다.  $z$ 는 코드북  $C \in R^{N \times d_c}$  각 요소와의 유clidean 거리(Euclidean Distance)를 통해 가장 유사한 항목을 찾아 양자화된다. 여기서  $N$ 은 코드북의 항목 수이다. 이후 양자화

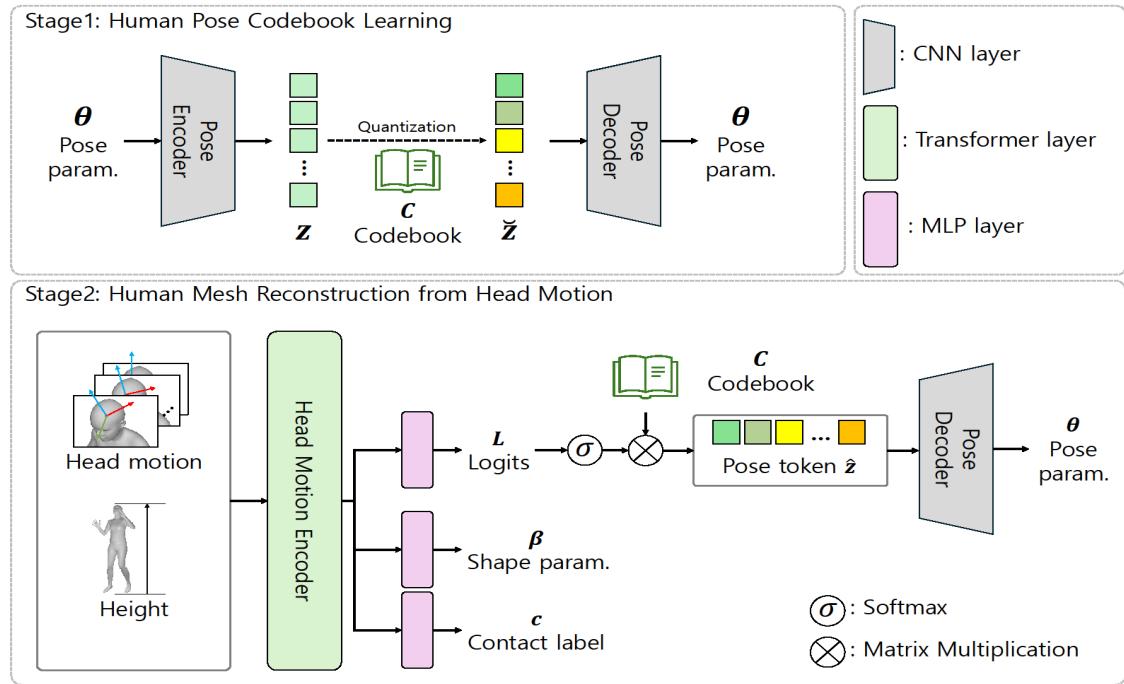


그림 1. 제안하는 방법의 전체 구조

Fig. 1. Overall architecture of the proposed method

된 자세 토큰  $\tilde{z} \in R^{M \times d_c}$ 은 디코더를 통해 SMPL 자세 파라미터로 복원된다. 사용된 인코더와 디코더는 합성곱 신경망<sup>[13]</sup>(Convolutional Neural Network)으로 구성되어 있다.

## 2. 머리 움직임으로부터 3차원 사람 메쉬 복원

두 번째 단계에서는 머리 움직임을 기반으로 코드북에서 적절한 코드를 찾아 3차원 사람 메쉬를 복원한다. 전체 구조는 그림 1의 아래쪽에서 볼 수 있다. 머리 움직임은 셀프 어텐션(Self-attention) 연산 기반의 트랜스포머로 구성된 인코더를 지나 형태와 접촉 확률, 그리고 자세 토큰에 대한 로짓(Logit)  $L \in R^{T \times M \times N}$ 을 예측한다. 여기서  $T$ 는 입력 시퀀스의 수이다. 양자화된 자세 토큰의 예측에 있어 미분 불가능한 지점이 있기 때문에, 곧바로 코드의 인덱스를 예측하기보다 로짓 기반의 접근 방법<sup>[13]</sup>을 사용하고 대량의 양자화된 자세 토큰  $\hat{z} \in R^{T \times M \times d_c}$ 은 수식 (1)과 같이 계산된다.

$$\hat{z} \simeq \sigma(L) \times C, \quad (1)$$

여기서  $\sigma$ 는 소프트 맥스 함수를 의미한다. 이를 통해 얻은 대량의 양자화된 자세 토큰은 이전 단계에서 학습된 디코더를 통과하여 최종적인 자세를 생성한다.

## 3. 제안하는 신경망 훈련을 위한 손실 함수

본 논문에서는 EgoAllo<sup>[7]</sup>에서 제안된 자세( $L_{pose}$ ), 형태( $L_{shape}$ ), 관절( $L_{kp3d}$ ), 속도( $L_{vel}$ ), 접촉( $L_{contact}$ ), 그리고 미끄러짐( $L_{sliding}$ ) 손실 함수를 통해 제안하는 신경망을 학습한다. 먼저,  $L_{pose}$ ,  $L_{shape}$ ,  $L_{kp3d}$ , 그리고  $L_{contact}$ 는 예측과 정답 값 간의 차이를  $L_1$  노름(Norm)을 이용하여 계산한다.  $L_{contact}$ 는 각 관절이 지형지물과 닿아 있는지에 대한 확률값을 학습하기 위함으로 이진 분류 손실 함수로 구성되어 있다.  $L_{sliding}$ 는 월드 좌표계에서 빌어 미끄러지는 현상을 최소화하기 위해 수식 (2)와 같이 계산한다.

$$L_{sliding} = \sum_{t=0}^{T-1} \| p^t \cdot v_f^t \|^2, \quad (2)$$

여기서  $p^t$  와  $v_f^t$ 는  $t$  번째 프레임에서 관절에 대한 접촉 확률과 관절의 속도 값이다. 최종적인 손실 함수는 수식 (3) 와 같다.

$$L_{total} = \lambda_{pose} L_{pose} + \lambda_{shape} L_{shape} + \lambda_{kp3d} L_{kp3d} + \lambda_{vel} L_{vel} + \lambda_{contact} L_{contact} + \lambda_{sliding} L_{sliding}, \quad (3)$$

여기서  $\lambda_{pose}$ ,  $\lambda_{shape}$ ,  $\lambda_{kp3d}$ ,  $\lambda_{vel}$ ,  $\lambda_{contact}$  그리고  $\lambda_{sliding}$  는 각각 1, 0.01, 1, 0.1, 1 그리고 1로 설정되었다.

### III. 실험 결과 및 분석

본 논문에서는 제안하는 방법의 학습과 평가를 위해 AMASS<sup>[11]</sup> 데이터셋을 활용하였다. AMASS은 45시간 분량의 다양한 사람 움직임의 포함하고 있는 대용량 합성 데이터셋이다. 제안하는 방법은 Yi<sup>[7]</sup> 등을 따라 학습과 평가 데이터셋을 나누었다. AMASS는 비디오 영상 및 자기중심적 데이터가 없어서 머리 자세와 미간-중심의 좌표계를 통해 머리 움직임을 생성한다. 제안하는 방법은 Pytorch<sup>[15]</sup>를 기반으로 구현되었다. 코드북의 구조는 TokenHMR<sup>[13]</sup>에서 제안하였던 방법을 사용하고 AMASS 데이터셋을 통해 사전 학습된 가중치를 사용하였다. 제안하는 신경망 (그림 1 의 아래부분)은 머리 움직임을 효과적으로 인코딩 하기 위해 트랜스포머를 활용한다. 구체적으로, 각 프레임별 머리 움직임과 키 정보는 다층 퍼셉트론(Multi Layer Perceptron)을 통해 토큰화되고 각 토큰의 위치 정보는 Rotary Position Embedding<sup>[16]</sup> 기법을 통해 더해진다. 이 토큰들은 셀프 어텐션 기반의 트랜스포머 인코더 3개를 통해 프레임 간의 관계를 학습하고, 각각의 목적에 맞는 서로 다른 3개의 다층 퍼셉트론을 거쳐 디코딩된다. 제안하는 신경망 최적화하기 위한 알고리즘은 Adam<sup>[17]</sup>을 사용하였다. 학습 속도(Learning Rate)  $1 \times 10^{-5}$  으로 설정하고 배치 크기 8로 20 에포크(Epoch) 동안 학습을 진행하였다. 학습과 성능 평

가에는 Intel(R) Xeon(R) w5-3433 CPU와 NVIDIA RTX 4090 GPU 1대가 이용되었다.

제안하는 방법의 성능 개선 효과를 검증하기 위해 4개의 베이스 라인 모델과의 비교한다. VAE+Opt는 기존 SLAHMR<sup>[18]</sup> 프레임워크에 머리 움직임 정보를 입력으로 하는 모델이다. EgoEgo<sup>[6]</sup>는 머리 움직임을 입력으로 하는 조건부 학산 모델을 기반으로 하고 있는 모델이고, 이와 유사하게 EgoAllo<sup>[7]</sup>의 경우 기존 머리 움직임을 정규화한 뒤 입력한다. NoShape은 EgoAllo 모델에서 몸의 형태 정보를 예측하지 않는 모델이다. 제안하는 방법의 정확도를 평가하기 위해 3차원 사람 메쉬 복원 분야에서 주로 사용하는 두 가지 평가 지표로 제안하는 방법의 성능을 측정하였다. 구체적으로, 평균 관절별 위치 오차(Mean Per-joint Position Error)와 정답 자세와 Procrustes 정렬 이후 평균 관절별 위치 오차를 활용하여 정량적으로 성능을 비교하였으며, 해당 결과를 표 1에 나타내었다. 표에서 볼 수 있듯, 제안하는 방법은 기존의 대표적인 방법인 EgoEgo<sup>[6]</sup>와 EgoAllo<sup>[7]</sup> 보다 좋은 성능뿐만 아니라 빠른 추론 속도로 결과를 생성한다.

표 1. 기준 데이터셋에서 정량적 평가 비교

Table 1. Performance comparison on the benchmark dataset

Methods	MPJPE(mm)	PA-MPJPE (mm)	Inference Time(sec)
VAE+Opt <sup>[7]</sup>	199.5	191.4	-
EgoEgo <sup>[6]</sup>	184.0	158.6	9.0
NoShape <sup>[7]</sup>	138.1	118.8	-
EgoAllo <sup>[7]</sup>	129.8	109.8	3.3
Ours	122.6	109.7	0.06

제안하는 방법의 3차원 메쉬 복원 정확도를 확인하기 위해 기존 방법인 EgoAllo<sup>[8]</sup>과의 비교 결과를 그림 2에서 확인할 수 있다. 첫 번째부터 다섯 번째 행의 결과에서 볼 수 있듯, 기존 방법보다 더 목표로 자세에 가까운 예측을 하는 것을 볼 수 있다. 구체적으로, 네 번째와 다섯 번째 행의 결과를 보면, 그럴듯하지만 정답 메쉬와 차이가 큰 것을 확인할 수 있다. 이는 기존 방법이 학산 모델을 채용하여 예측을 하기에 발생하는 문제로 좋은 품질의 샘플을 생성하는데는 뛰어나지만 입력에 따른 유일한 정답 메쉬를 복원하

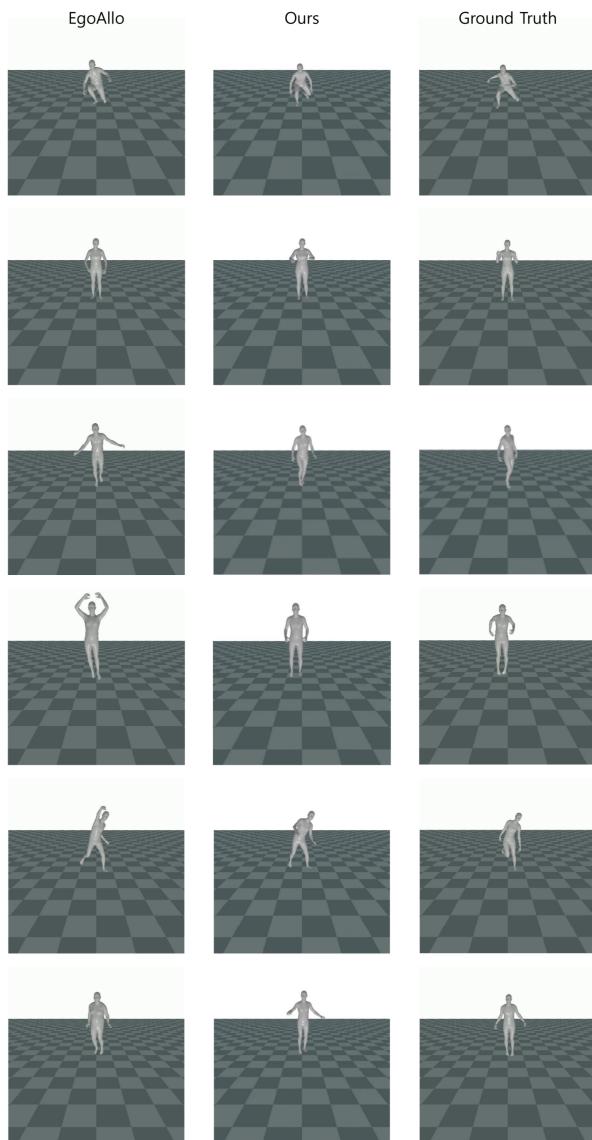


그림 2. AMASS<sup>[11]</sup> 데이터셋에 대한 에고-바디 메쉬 복원 결과, 원쪽부터 오른쪽으로: EgoAllo<sup>[7]</sup>의 결과, 제안하는 방법의 결과, 그리고 정답 메쉬  
 Fig. 2. Results of Ego-body mesh reconstruction on the AMASS<sup>[11]</sup> dataset. From left to right: EgoAllo<sup>[7]</sup> results, ours, and ground truth

는 과정에는 불확실성이 개입되어 정확도에 부정적인 영향을 미치기 때문이다. 자세뿐만 아니라 사람의 형태 예측에서도 이러한 경향을 보인다는 것을 마지막 행의 결과를 통해 알 수 있었다. 반면, 제안하는 방법은 주어진 머리 움직임과 사용자의 키 정보에 적절한 자세 토큰 찾도록 유도하-

기에 확률적인 요소 없이 정답에 가까운 메쉬를 생성할 수 있다는 장점이 있다.

제안하는 신경망의 구조를 검증하기 위한 하이퍼파라미터 변동에 따른 성능의 차이를 표 2에 나타내었다. 제안하는 신경망은 간단한 트랜스포머 구조를 가지고 있어 신경망의 깊이 또는 특징의 차원 수에 따른 비교를 하였다. 표 2에서 볼 수 있듯, 3개의 층과 256 차원에서 큰 성능 향상이 있었음을 확인할 수 있었다.

표 2. 제안하는 방법에서의 하이퍼파라미터 변동에 따른 성능 분석

Table 2. Performance analysis according to hyperparameter variations in the proposed method

		MPJPE(mm)	PA-MPJPE(mm)
# of layers	1	146.2	131.4
	2	135.9	122.9
	3	122.6	109.7
# of channels	32	144.3	133.5
	64	139.8	120.9
	128	122.6	109.7

#### IV. 결 론

본 논문에서는 머리 움직임으로부터 3차원 사람 메쉬를 복원하는 코드북 기반의 네트워크를 제안하였다. 제안하는 방법의 핵심은 사람의 자세에 대해 학습된 코드북에서 주어진 머리 움직임에 적합한 전신 자세 토큰을 찾아 한 번의 단계로 전신 메쉬를 복원하는 것이다. 제안하는 방법이 간단한 구조를 가졌음에도 기존 방법보다 좋은 성능과 빠른 추론 속도를 가지고 있다는 것을 실험 결과를 통해 확인할 수 있었다.

#### 참 고 문 헌 (References)

- [1] Q. Gu, Z. Lv, D. Frost, S. Green, J. Straub, and C. Sweeney, “Egolifter: Open-world 3d segmentation for egocentric perception,” in *Proc. Eur. Conf. Comput. Vis.*, pp. 382-400, Sep. 2024.  
 doi: [https://doi.org/10.1007/978-3-031-72775-7\\_22](https://doi.org/10.1007/978-3-031-72775-7_22)
- [2] D. Zhang, G. Li, J. Li, and M. Bressieux, “Egogaussian: Dynamic scene understanding from egocentric video with 3d gaussian splatting,”

- in *Proc. Int. Conf. 3D Vis.*, pp. 1091-1102. Mar. 2025.  
doi: <https://doi.org/10.1109/3DV66043.2025.00105>
- [3] C. Plizzari, S. Goel, T. Perrett, J. Chalk, A. Kanazawa, and D. Damen, “Spatial cognition from egocentric video: Out of sight, not out of mind,” in *Proc. Int. Conf. 3D Vis.*, pp. 1211-1221. Mar. 2025.  
doi: <https://doi.org/10.1109/3DV66043.2025.00115>
- [4] J. Jiang, P. Strelci, H. Qiu, A. Fender, L. Laich, P. Snape, and C. Holz, “Avatarposer: Articulated full-body pose tracking from sparse motion sensing,” in *Proc. Eur. Conf. Comput. Vis.*, pp. 443-460, Oct. 2022.  
doi: [https://doi.org/10.1007/978-3-031-20065-6\\_26](https://doi.org/10.1007/978-3-031-20065-6_26)
- [5] A. Castillo, M. Escobar, G. Jeanneret, and A. Pumarola, “Bodiffusion: Diffusing sparse observations for full-body human motion synthesis,” in *Proc. Int. Conf. Comput. Vis.*, pp. 4221-4231, Oct. 2023.  
doi: <https://doi.org/10.48550/arXiv.2304.11118>
- [6] J. Li, K. Liu, and J. Wu, “Ego-body pose estimation via ego-head pose estimation,” in *Proc. Conf. Comput. Vis. Pattern Recognit.*, pp. 17142-17151, Jul. 2023.  
doi: <https://doi.org/10.48550/arXiv.2212.04636>
- [7] B. Yi, V. Ye, M. Zheng, and Y. Li, “Estimating body and hand motion in an ego-sensed world,” in *Proc. Conf. Comput. Vis. Pattern Recognit.*, pp. 7072-7084, Jul. 2025.  
doi: <https://doi.org/10.48550/arXiv.2410.03665>
- [8] J. Ho, A. Jain, and P. Abbeel, “Denoising diffusion probabilistic models,” in *Proc. Conf. Neural Inf. Process. Syst.*, pp. 6840-6851, Dec. 2020.  
doi: <https://doi.org/10.48550/arXiv.2006.11239>
- [9] X. Pan, N. Charron, Y. Yang, S. Peters, T. Whelan, C. Kong, O. Parkhi, R. Newcombe, and Y. Ren, “Aria digital twin: A new benchmark dataset for egocentric 3d machine perception,” in *Proc. Int. Conf. Comput. Vis.*, pp. 20133-20143. Oct. 2023.  
doi: <https://doi.org/10.1109/ICCV51070.2023.01842>
- [10] Z. Teed and J. Deng, “DROID-SLAM: Deep visual slam for monocular, stereo, and RGB-D cameras,” in *Proc. Conf. Neural Inf. Process. Syst.*, pp. 16558-16569, Dec. 2021  
doi: <https://doi.org/10.48550/arXiv.2108.10869>
- [11] N. Mahmood, N. Ghorbani, N. Troje, G. Pons-Moll, and M. Black, “AMASS: Archive of motion capture as surface shapes,” in *Proc. Int. Conf. Comput. Vis.*, pp. 5442-5451, Oct. 2019.  
doi: <https://doi.org/10.1109/ICCV.2019.00554>
- [12] V. Oord, A. Aaron, and O. Vinyals, “Neural discrete representation learning,” in *Proc. Conf. Neural Inf. Process. Syst.*, pp. 6309 - 6318, Dec. 2017.  
doi: <https://doi.org/10.48550/arXiv.1711.00937>
- [13] S. Dwivedi, Y. Sun, P. Patel, Y. Feng, and M. Black, “Tokenhmhr: Advancing human mesh recovery with a tokenized pose representation,” in *Proc. Conf. Comput. Vis. Pattern Recognit.*, pp. 1323-1333, Jun. 2024.  
doi: <https://doi.org/10.1109/CVPR52733.2024.00132>
- [14] F. Bogo, A. Kanazawa, C. Lassner, P. Gehler, J. Romero, and M. Black, “Keep it SMPL: Automatic estimation of 3d human pose and shape from a single image,” in *Proc. Eur. Conf. Comput. Vis.*, pp. 561-578, Oct. 2016.  
doi: <https://doi.org/10.1145/2816795.2818013>
- [15] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, “PyTorch: An imperative style, high-performance deep learning library,” in *Proc. Conf. Neural Inf. Process. Syst.*, pp. 8024 - 8035, Dec. 2019.  
doi: <https://dl.acm.org/doi/10.5555/3454287.3455008>
- [16] J. Su, M. Ahmed, Y. Lu, S. Pan, W. Bo, and Y. Liu, “Roformer: Enhanced transformer with rotary position embedding,” *Neurocomputing*, vol. 568, Feb 2024.  
doi: <https://doi.org/10.1016/j.neucom.2023.127063>
- [17] D. P. Kingma and J. Ba, “Adam: a method for stochastic optimization,” in *Proc. Int. Conf. Learn. Represent.*, pp. 1-15, May 2015.  
doi: <https://doi.org/10.48550/arXiv.1412.6980>
- [18] V. Ye, G. Pavlakos, J. Malik, and A. Kanazawa, “Decoupling human and camera motion from videos in the wild,” in *Proc. Conf. Comput. Vis. Pattern Recognit.*, pp. 21222-21232, Jul. 2023.  
doi: <https://doi.org/10.1109/CVPR52729.2023.02033>

---

## 저자 소개

---

### 성 강 윤



- 2018년 3월 ~ 2024년 2월 : 건국대학교 전기전자공학부 학사
- 2024년 3월 ~ 현재 : 건국대학교 전기전자공학부 석사
- ORCID : <https://orcid.org/0009-0003-8661-6972>
- 주관심분야 : 컴퓨터 비전, 영상처리, 심층학습

---

저자소개

---



김원준

- 2012년 8월 : 한국과학기술원(KAIST) 박사
- 2012년 9월 ~ 2016년 2월 : 삼성종합기술원 전문 연구원
- 2016년 3월 ~ 2020년 2월 : 건국대학교 전기전자공학부 조교수
- 2020년 3월 ~ 현재 : 건국대학교 전기전자공학부 부교수
- ORCID : <https://orcid.org/0000-0001-5121-5931>
- 주관심분야 : 컴퓨터 비전, 영상처리, 기계학습