

# 실루엣 기반의 카메라 투영을 이용한 논파라메트릭 3D 인체 복원 개선 연구

변지성, 염세경\*

동국대학교 일반대학원 산업시스템공학과

quswltjd12@naver.com, sekyoungyoum@gmail.com

\* Corresponding Author

## A Study on Improving Non-parametric 3D Human Reconstruction Using Silhouette-based Camera Projection

Byeon Ji Seong, Youm Se Kyoung  
Dongguk Univ.

### 요약

본 논문은 단일 RGB 이미지로부터 3D 인체 형상을 복원하는 과정에서 발생하는 2D-3D 투영 정합 오류를 해결하기 위해 실루엣 기반의 카메라 투영 기법을 제안한다. 기존의 PIFu 계열 모델은 학습 및 추론 시 스케일, 이동, 회전 등의 정답 카메라 파라미터가 필수적으로 요구되거나 데이터셋에 고정된 투영 가정을 전제하는 한계가 있다. 그러나 공개 데이터셋에서는 이러한 파라미터가 제공되지 않거나 값이 불안정하여 실제 환경 적용 시 투영 오차가 커지며 전신 비례의 왜곡을 야기한다. 이에 본 연구는 이미지에서 추출한 실루엣의 기하학적 정보만을 활용하여 투영 스케일과 오프셋을 최적화함으로써 픽셀 정합 내재 함수가 정확한 특징을 참조하도록 유도하였고, 결과적으로 복원된 3D 인체 모델의 구조적 안정성과 품질을 개선하였다.

### I. 서론

최근 헬스케어와 메타버스 산업의 성장으로 단일 이미지 기반 3D 인체 복원 기술의 중요성이 크게 부각되고 있다 [1, 2]. 특히 PIFu[3]와 같은 논파라메트릭(non-parametric) 방식은 고정된 템플릿 제약 없이 개인 고유의 세밀한 체형을 복원할 수 있다는 장점이 있으나, 3D 포인트를 2D 이미지 평면으로 투영하는 정합 과정의 정확도에 매우 민감하다. 정답 카메라 파라미터가 제공되지 않거나 불안정한 실제 환경에서는 투영 오차로 인해 픽셀 특징 참조 위치가 어긋나고, 그 결과 전신 비례 왜곡이나 표면 아티팩트가 발생할 수 있다[4]. 아래의 그림 1은 이러한 투영 정합 오류로 인해 2D-3D 정렬이 잘못된 상태에서 학습이 진행될 경우, 추론 단계에서 왜곡된 3D 복원 결과가 나타날 수 있음을 보여준다.

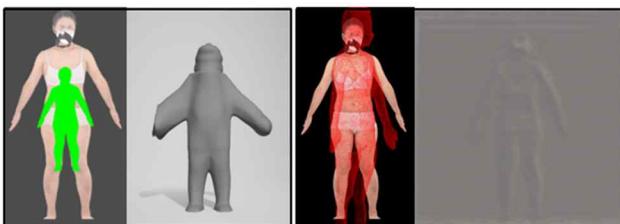


그림 1 3D-2D 투영 오류로 인해 잘못 학습된 경우의 복원 결과 예시

따라서 본 연구에서는 정답 카메라 파라미터에 대한 의존을 배제하고, 이미지에서 추출한 실루엣(silhouette) 정보만을 기반으로 투영 스케일과 오프셋을 정렬하여 2D-3D 투영 정합을 안정화하는 실루엣 기반 카메라 투영 기법을 제안한다.

### II. 본론

기존의 3D 인체 복원 연구들 중 PIFu 계열 모델은 3D 포인트를 2D 이미지 평면에 정확히 투영하기 위해 학습 단계에서 스케일(scale), 이동(translation), 회전(rotation) 등의 카메라 투영 파라미터가 필수적으로 요구된다[3, 4]. 일부 데이터셋은 이러한 카메라 파라미터의 정답값을 제공하지만, 데이터 구매 비용이 발생하는 고가의 데이터셋에 의존하는 경우가 많다. 반면, 다수의 공개 데이터셋에서는 카메라 파라미터가 제공되지 않거나 값이 불안정하여, 데이터셋이 달라지거나 파라미터 정보가 없는 환경에서 투영 정합의 불안정성이 커지고 일반화에 한계가 발생한다. 따라서 GT 카메라 파라미터가 없는 상황에서도 다양한 데이터셋에 robust하게 적용 가능한 2D-3D 투영 기법이 필요하다.

본 연구에서 제안하는 기법은 정답 파라미터가 없는 상황에서도 이미지에서 추출된 2D 실루엣 마스크를 직접적인 기준점으로 활용하여 3D 메쉬를 정합시킴으로써, 투영 정합을 안정화한다. 먼저, 제안

기법이 전체 추론 파이프라인에서 적용되는 위치와 흐름을 그림 2에 제시한다. 이후 실루엣 바운딩 박스를 이용한 스케일, 오프셋 계산과 회전 각도 최적화 과정을 구체적으로 설명한다.

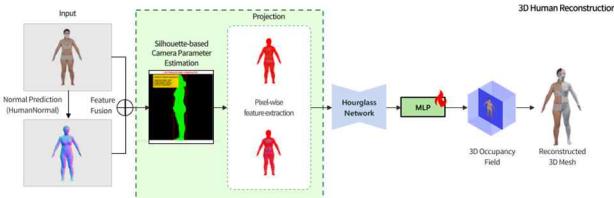


그림 2 제안하는 실루엣 기반의 카메라 파라미터 투영 기법이 반영된 프레임워크

구체적으로는 2D 이미지의 실루엣 바운딩 박스의 크기  $(w, h)$ 와 정규화된 메쉬의 크기  $(w_m, h_m)$ 를 비교하여 투영 스케일  $s = \min(\frac{w}{w_m}, \frac{h}{h_m})$ 을 산출한다. 이후 실루엣 중심 좌표  $(c_x, c_y)$ 와 메쉬 중심의 투영 위치가 일치하도록 오프셋  $t_x = c_x - (s \cdot X_c)$  및  $t_y = c_y - (s \cdot Y_c)$ 를 계산한다. 또한 정면 시점에서 미세하게 벗어난 경우를 대비하여 실루엣 정합도가 최대가 되는 최적의 Y축 회전 각도  $\alpha^* = \arg \max_{\alpha} \text{IOU}(\text{Silhouette}, \text{ProjMesh}(\alpha))$ 를 탐색하여 투영의 안정성을 극대화한다. 이러한 방식은 카메라의 물리적 파라미터를 수치적으로 회귀 예측하는 것이 아니라, 실루엣과 메쉬가 이미지 평면에서 기하학적으로 어긋나지 않도록 투영 프로세스 자체를 안정화하는 데 목적이 있다. 결과적으로 3D 포인트는  $u = (s \cdot X') + t_x$ ,  $v = (s \cdot Y') + t_y$ 로 투영되며, 이를 통해 각 3D 포인트에 대응하는 픽셀 위치에서 정확한 특징(feature)을 안정적으로 추출할 수 있다. 그 결과 그림 3과 같이 투영 정합 오류로 인한 전신 비례 왜곡과 표면 아티팩트를 완화하여 보다 정밀한 3D 복원을 가능하게 한다.

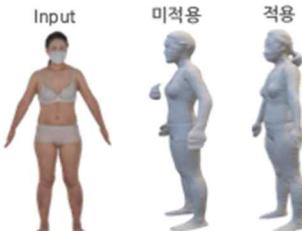


그림 3 실루엣 기반 카메라 투영 적용 전후 3D 복원 결과 비교

제안한 기법을 적용한 결과, 표 1에 제시된 바와 같이 주요 성능 지표에서 개선된 성능이 확인되었다.

표 1 실루엣 기반 카메라 투영 기법 적용에 따른 3D 인체 복원 성능 비교

모델	CD	P2S	MIoU	SSIM	PSNR	NORMAL CONSISTENCY
PIFu	<b>0.53</b>	<b>0.20</b>	0.27	0.20	6.53	0.16
Ours	0.56	0.22	<b>0.50</b>	<b>0.74</b>	<b>15.21</b>	<b>0.27</b>

### III. 결론

본 논문에서는 실루엣 정보를 활용한 카메라 투영 정렬을 통해 논파라메트릭 3D 인체 복원의 2D-3D 투영 정합 안정성을 크게 향상시켰다. 제안한 방법은

이미지에서 추출되는 실루엣(silhouette)이라는 가벼운 기하학적 정보만으로 투영 스케일과 오프셋을 정렬하여, 픽셀 특징 참조 위치의 어긋남을 완화한다. 실험 결과 제안하는 방식은 기존 모델 대비 mIoU 및 SSIM 등 주요 성능 지표에서 우수한 결과를 나타냈으며, 특히 전신 부피와 표면 방향의 일관성을 보다 안정적으로 복원함을 확인하였다. 이는 투영 단계의 정렬 최적화만으로도 픽셀 기반 내재 함수의 학습 및 추론 안정성을 높일 수 있음을 시사한다. 향후 정밀한 신체 치수 측정 및 모바일 기반 디지털 트윈 시스템 구축에 있어 중요한 기술적 토대가 될 것으로 기대된다.

### ACKNOWLEDGMENT

본 연구는 한국연구재단(NRF)의 지원을 받아 수행되었습니다. (과제번호: NRF 2023R1A2C2004252)

### 참 고 문 헌

- [1] Gage, A. D., Knight, M. A., Bintz, C., Aldridge, R. W., Angelino, O., Dieleman, J. L., … and Haakenstad, A., “Disparities in telemedicine use and payment policies in the United States between 2019 and 2023”, Communications Medicine, 5(1), pp.52, 2025.
- [2] Katsoulakis, E., Wang, Q., Wu, H., Shahriyari, L., Fletcher, R., Liu, J., … and Deng, J., “Digital twins for health: a scoping review”, NPJ Digital Medicine, 7(1), pp. 77, 2024.
- [3] Saito, S., Huang, Z., Natsume, R., Morishima, S., Kanazawa, A., and Li, H., “PIFu: Pixel-Aligned Implicit Function for High-Resolution Clothed Human Digitization”, In: Proc. IEEE/CVF International Conference on Computer Vision (ICCV), 2019.
- [4] Saito, S., Simon, T., Saragih, J., and Joo, H., “PIFuHD: Multi-Level Pixel-Aligned Implicit Function for High-Resolution 3D Human Digitization”, In: Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.