



일반논문 (Regular Paper)

방송공학회논문지 제30권 제5호, 2025년 9월 (JBE Vol.30, No.5, September 2025)

<https://doi.org/10.5909/JBE.2025.30.5.743>

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

다중 시점 기반 적응형 신뢰도 임계값 설정을 통한 3D 자세 복원

박정탁^{a)}, 이학범^{a)}, 서영호^{a)b)†}

3D Human Pose Reconstruction Using Multi-View Adaptive Confidence Thresholding

Jung-Tak Park^{a)}, Hak-Bum Lee^{a)}, and Young-Ho Seo^{a)b)†}

요약

본 논문은 다중 시점 기반 3차원 인간 자세 추정에서 연산 효율성과 이상치 제거의 문제를 동시에 해결하고자, 비학습 기반의 경량화된 필터링 기법을 제안한다. 기존 방법들은 모든 시점의 예측 결과를 동일하게 취급하거나, 고정된 임계값을 기준으로 관절을 필터링함으로써, 신뢰도가 낮거나 중복되는 정보에 의해 재구성 품질이 저하되는 문제가 존재하였다. 이에 본 논문에서는 (1) 관절별 신뢰도 분포에 기반한 소프트 가중 기반 적응형 임계값 설정(Soft-Weighted Thresholding), (2) 시점 간 코사인 유사도 분석을 활용한 비정상 임계값 방지를 위한 하한 설정(Redundancy-Aware View Filtering), (3) 일정 기간 동일 시점을 반복 활용하는 시간에 따른 시점 간 일관성 유지(Temporal View Consistency) 기법을 결합하여, 제한된 시점 수와 실제 환경(wild condition)에서도 정확한 3D 복원이 가능하도록 제안한다. 제안된 방법은 대수적 삼각 측량에 기반하여 구현되었으며, CMU Panoptic Studio 데이터셋을 활용한 실험 결과로 기존 baseline 대비 최대 17%의 MPJPE 감소를 달성하였다.

Abstract

This paper proposes a non-learning-based lightweight filtering method to simultaneously address the challenges of computational efficiency and outlier removal in multi-view-based 3D human pose estimation. Existing approaches often treat the predictions from all viewpoints equally or filter joints using fixed thresholds, leading to a degradation in reconstruction quality due to unreliable or redundant information. To overcome these limitations, this study introduces (1) Soft-Weighted Thresholding, an adaptive thresholding method based on the confidence distribution of each joint; (2) Redundancy-Aware View Filtering, which sets a lower bound to prevent abnormal thresholding by analyzing cosine similarity between views; and (3) Temporal View Consistency, which ensures inter-view consistency over time by reusing identical viewpoints for a certain duration. The proposed method, implemented using algebraic triangulation, demonstrates its effectiveness under limited viewpoints and real-world (wild condition) scenarios. Experimental results using the CMU Panoptic Studio dataset show up to a 17% reduction in MPJPE compared to existing baselines.

Keyword : 3D Human Pose Estimation, Multi-View Vision, Real-Time Reconstruction

I. 서론

인간 포즈 추정(Human Pose Estimation, HPE)은 이미지나 비디오 같은 시각 데이터를 기반으로 인간 신체의 관절 위치를 추정하는 문제로, 컴퓨터 비전 분야에서 오랜 기간 활발히 논쟁되어 왔다^{[1][2][3]}. 이 기술은 인간-컴퓨터 상호작용(Human-Computer Interaction, HCI), 행동 분석, 증강 및 가상현실(AR/VR), 스포츠 동작 평가 등 다양한 실세계 응용 분야에 필수적인 기반 기술로 자리 잡았다. 특히 3차원(3D) 포즈 추정은 신체의 구조적 이해를 가능하게 하여 보다 정밀한 분석과 자연스러운 상호작용을 제공한다^{[4][11]}. 3D 자세 추정 기술은 의료 재활, 로봇 제어, 감정 및 행동 인식, 군집 행동 분석, 교통 감시 및 보행자 행동 예측 등에서도 활용되고 있으며, 최근에는 메타버스, 모션 캡처 기반 콘텐츠 제작, 무인 이동체의 사람-환경 상호작용 논문 등으로 그 활용 범위가 지속적으로 확장되고 있다^[6]. 이러한 확장성으로 인해 3D 자세 추정은 다양한 산업군에서 핵심 기술로 주목받고 있다. 초기 논문은 수작업으로 설계된 특징과 단순한 스틱 피규어 모델(Stick Figure Model)에 의존하여 복잡한 장면 처리에 한계가 있었다. 이후 합성곱 신경망(CNN)의 등장으로 2D 포즈 추정의 성능이 크게 향상되었고, DeepPose^[1], HRNet^[2] 등 다양한 모델이 개발되었다. 이러한 발전은 단일 인물에서 다중 인물 포즈 추정으로 확장되었으며, 하향식 및 상향식 방식의 병행 발전으로 복잡한 장면에서도 정확한 추정이 가능해졌다^{[3][7][8]}.

2D 포즈 추정의 성공은 자연스럽게 3D 포즈 추정 논문으로 이어졌다. 단일 시점 영상으로부터 3D 포즈를 예측하는 2단계 접근법이 대표적이었으나, 이 방식은 깊이 정보의 소실로 인해 여전히 본질적으로 불안정한(ill-posed) 문제로 남아 있다^{[9][10]}. 또한 정확한 3D 주석 데이터 수집에는 값비

싼 모션 캡처 시스템과 환경적 제약이 요구된다^[5]. 특히 단일 시점 기반 3D 복원은 시점 종속성(view-dependency), 깊이 모호성(depth ambiguity), 그리고 occlusion 문제에 취약하며, 실제 장면에서는 다양한 조명 조건, 피사체 회전, 빠른 자세 변화 등의 영향으로 추정 정확도가 저하될 수 있다^{[11][12]}. 역문제(inverse problem)인 3D 복원이 본질적으로 과소결정(underdetermined)된 문제임을 반영하며, 소량의 2D 오차가 전체 3D 복원 결과에 증폭되어 반영될 가능성이 존재한다. 따라서 최근 논문은 다중 시점 기반 방법이나 관성 센서(Inertial Measurement Unit, IMU), 심도 센서, 라이다(Light Detection And Ranging, LiDAR) 등과의 융합을 통해 3D 포즈 추정의 정밀성과 강건성을 높이려는 방향으로 진행되고 있다^{[13][14]}.

2D 포즈 추정의 성능 향상은 3D 포즈 추정 논문으로 자연스럽게 이어졌으며, 다양한 접근법이 시도되어 왔다. 일반적으로 단일 시점 기반 접근법과 다중 시점 기반 접근법으로 구분되며, 각 방식은 장단점이 공존한다. 단일 시점 기반 3D 포즈 추정은 2D 관절 좌표를 기반으로 3D 좌표를 복원하는 회귀 모델을 활용한다^{[9][15]}. 이 방식은 시스템 구성이 간단하고 장비 부담이 비교적 적다는 장점이 있어 널리 활용되고 있다. 그러나 시점이 단일할 경우 깊이 정보가 제한되어 있으며, 가림(occlusion), 관절의 좌우 혼동(ambiguity), 카메라-피사체 상대 위치 변화 등 다양한 요인에 의해 추정 성능이 영향을 받을 수 있다. 다중 시점 기반 접근법은 이러한 단일 시점 기반의 한계를 보완하고자, 여러 시점에서 취득한 2D 정보를 융합하여 3D 복원을 수행하는 방식을 활용한다^{[5][16][17]}. 대수 기반 삼각 측량(Algebraic Triangulation)과 볼류메트릭 통합(Volumetric Fusion) 기법 등이 대표적으로 사용되고 있으며^{[13][16][18]}, 최근에는 학습 기반 신뢰도 추정이나 확률 기반 최적화 기법을 병합한 다양한 시도가 이루어지고 있다^{[19][20][21]}. 이러한 발전에도 불구하고, 다중 시점 기반 방식에서도 여전히 몇 가지 기술적 과제가 존재한다. 예를 들어, 볼류메트릭 접근법은 높은 해상도의 복셀 공간을 다룰수록 연산량이 급격히 증가하여, 실시간 응용이나 경량화가 요구되는 시스템에서는 적용에 제약이 있을 수 있다^{[16][23]}. 대수 기반 삼각 측량 방식은 비교적 연산량이 적은 장점이 있으나, 관측된 2D 키포인트가 이상치를 포함하는 경우 전체 3D 복원 결

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>

※ 본 연구는 2023년도 중소벤처기업부의 창업성장기술개발사업 지원에 의한 연구임 [RS-2023-00264741]

· Manuscript July 23, 2025; Revised August 21, 2025; Accepted August, 22, 2025.

과에 영향을 받을 수 있으며, 이상치 처리를 위해 적절한 필터링이 요구된다^{[19][21]}. 이를 보완하기 위해 기존 논문들에서는 신뢰도 기반 필터링을 적용하는 경우가 있으나, 대부분 관절별 또는 시점별 관측 분포의 다양성을 충분히 반영하기 어려운 고정 임계값(fixed threshold)을 사용하는 경우가 많아, 일부 관절에서 이상치가 완전히 제거되지 않거나 유효한 정보가 과도하게 제거되는 가능성이 있다^{[20][21][23]}. 또한, 모든 시점을 동시에 사용하는 경우 시점 간 유사도가 높은 중복된 정보가 포함될 수 있으며, 이러한 경우 연산량은 증가하는 반면 추가적인 성능 향상에는 한계가 발생할 수 있다^[23]. 한편, 대부분의 기존 논문에서는 프레임 단위의 시점 선택이 독립적으로 이루어지는 경우가 많아, 연속적인 자세 변화가 포함된 환경에서는 시점 구성이 자주 변화하게 되어 시간적 일관성이 저하될 가능성도 존재한다^{[6][23]}.

본 논문에서는 이러한 기존 연구 결과들을 보완할 수 있는 방안으로서, 관절별 신뢰도 분포 기반 적응형 임계값 설정과 시점 간 유사도 분석을 결합한 비학습 기반 경량 필터링 프레임워크를 제안한다. 본 논문에서는 다중 시점 기반 3차원 인간 자세 추정 과정에서 발생할 수 있는 이상치, 신뢰도 편차, 카메라 중복 및 시간적 불안정성 문제를 완화하기 위해, 학습 기반 모델에 의존하지 않는 경량화된 임계값 설정 기반 처리 프레임워크를 사용한다. 시간적 연속성을 고려하지 않은 시점 선택 방식 역시 프레임 간 결과의 안정성 저하를 초래할 수 있는 한계를 개선하기 위한 방법도 논의한다. 결과적으로 본 논문의 제안 기법은 대규모 학습 데이터나 복잡한 신경망 아키텍처에 의존하지 않고도 적용 가능하며, 제한된 시점 수와 실환경 조건에서도 안정적인 3D 자세 복원이 가능하도록 설계되었다. 이러한 경량화된 프레임워크는 실제 모바일 시스템, 실시간 응용, 그리고 센서 수 제한이 존재하는 다양한 환경에서의 활용 가능성을 높인다.

본 논문은 2장에서는 3D 인간 자세 추정과 관련된 기존 논문들을 2D 포즈 추정, 단일 시점 기반 3D 복원, 다중 시점 기반 접근으로 구분하여 고찰한다. 제3장에서는 제안하는 3D 자세 복원 프레임워크의 전체 구성과 세부 알고리즘을 소개한다. 제4장에서는 CMU Panoptic Studio 데이터셋을 활용한 실험 결과를 제시하고, 제안 기법의 성능을 기존

방식과 비교 분석한다. 마지막으로 제5장에서는 본 논문의 결론과 향후 논문 방향을 제시한다.

II. 이전 연구의 분석

본 장에서는 3차원 인간 자세 추정과 관련된 기존 접근을 2D 포즈 추정, 단일 시점 기반 3D 복원, 다중 시점 기반 접근의 세 가지 범주로 나누어 고찰한다. 2D 인간 자세 추정은 딥러닝 기술의 급속한 발전으로 정확도가 향상되어 3D 복원의 기반 기술로 자리 잡았다. 단일 시점 기반 3D 추정은 구조가 단순하고 실시간 처리에 적합하지만 깊이 정보 부족과 가림 현상에 취약하며, 다중 시점 기반 방식은 이를 보완하나 높은 연산량과 이상치 문제, 시점 중복 및 시간 일관성 부족과 같은 한계가 존재한다.

1. 2D 인간 자세 추정

2D HPE는 이미지로부터 인체의 주요 관절 위치를 추정하는 기술로, DeepPose^[1], OpenPose^[3], HRNet^[2], HigherHRNet^[7], AlphaPose^[8] 등 다양한 모델이 존재한다. 최근에는 semi-/self-supervised 학습 기법을 통해 다양한 환경에 대한 일반화 성능이 향상되었고, 예측 결과는 관절 좌표와 신뢰도(confidence score)로 구성되어 3D 복원에 입력으로 활용된다. 실제 환경에서는 가림이나 화질 저하로 인해 예측 정확도에 편차가 생기며, 특히 다중 시점 기반 3D 복원에서는 이 오류가 직접 반영되므로 이상치 제거 기술이 중요하다. 2단계 접근법은 2D 관절 좌표를 기반으로 회귀를 통해 3D 위치를 복원하며, 모델화가 용이하다. Martinez et al.^[9]과 Zhou et al.^[24]은 간단한 신경망 구조로도 뛰어난 성능을 입증하였고, Shan et al.^[20]은 시계열 정보를 활용한 P-STMO로 시간 일관성을 반영하였다. 이 접근법은 전체 파이프라인 최적화가 가능하지만, 2D 추정 오류의 누적이라는 근본적 한계를 가진다. 단일 단계 접근법은 이미지에서 직접 3D 좌표를 회귀하여 중간 2D 단계 없이 중단 간 학습이 가능하다. X. Xu et al.^[5]은 coarse-to-fine 구조의 볼류메트릭 회귀 방식으로 정확도를 높였으며, 이는 이후 관련 연구에 큰 영향을 끼쳤다. 다만 고해상도 복셀

처리로 인한 연산량과 메모리 부담이 크며, 학습 데이터 부족 시 성능 저하가 발생할 수 있다. Hossain and Little^[11]은 LSTM 기반 시계열 모델로 시간적 일관성을 확보하는 방법을 제안하였다.

2. 다중 시점 기반 3D 인간 자세 추정

다중 시점 기반 방식은 깊이 추정의 정확성을 높이고 가림에 강건하지만, 복잡한 연산 구조와 이상치 처리가 주요 과제다. Tu et al.^[16]의 불류메트릭 통합은 2D 히트맵을 3차원 복셀로 역투영한 후 누적하여 3D 관절 분포를 예측하며, 강건한 성능을 보인다. 그러나 해상도가 높아질수록 연산량이 세제곱으로 증가하여 실시간 응용에 부적합하다. 대수 기반 삼각 측량은 상대적으로 효율적이며, Iskakov et al.^[13], Tome et al.^[12], Bartol et al.^[19]은 시점별 신뢰도 가중치를 학습하여 성능을 향상시켰다. Zhang et al.^[21], Wang et al.^[23]은 관절별 신뢰도 분포를 고려한 필터링 및 시점 선택 기법을 제안하였고, Chen et al.은 유효 시점 선정을 통해 중복을 줄이려 하였다. 하지만 프레임 간 시점 변동이 잦을 경우 시간 일관성이 저하되며, 일부 접근은 처리 지연과 시스템 안정성에 부담을 준다.

3. 한계와 제안 방향

다중 시점 기반 3D 인간 자세 추정은 깊이 정보 복원과 가림 문제의 완화 측면에서 확실한 성능 개선을 보장하지만, 해당 방식들이 실질적으로 갖는 한계는 세 가지 축으로 요약할 수 있다: 계산 복잡성, 이상치 민감도, 그리고 시점 중복 및 시간 불안정성이다. 먼저, 불류메트릭 통합 기법은 고정된 복셀 해상도를 기준으로 전체 관절의 분포를 예측하며 공간적 정밀도를 향상시킨다는 점에서 효과적이지만, 고해상도 복셀 운용 시 연산 복잡도가 공간의 세제곱(cubic complexity)으로 증가하게 되어 실시간 시스템에서는 적용이 어렵다. 특히 모바일 및 임베디드 시스템에서는 메모리 사용량과 처리 지연이 실질적 병목으로 작용한다. 대수 기반 삼각 측량은 계산량 측면에서 상대적으로 효율적이지만, 관측된 2D 관절이 이상치를 포함할 경우 복원된 3D 결과는 왜곡되기 쉬우며, 이로 인해 전체 시스템의 정확도가 급

격히 하락할 수 있다. 특히 고정 임계값 기반 필터링 방식은 관절 및 시점별 신뢰도 분포의 다양성을 고려하지 못해, 지나치게 엄격하거나 관대한 이상치 판별 기준을 야기한다. 이는 실제 관절이 신뢰도 편차가 큰 경우(예: 손목, 발목 등)에 더욱 치명적이다.

또한 시점 수가 많아질수록 발생하는 시점 간 중복은 정보의 다양성보다는 중복(redundant) 정보의 축적을 초래하며, 전체 연산의 비효율성과 오류 누적의 가능성을 높인다. 이러한 시점 구성은 프레임 간 자주 변경될 경우 시간적 일관성 문제로 이어져, 추정된 포즈의 안정성을 저하시킨다. 이는 특히 동작 분석, 인터랙티브 미디어, 스포츠 트래킹 등 시간 연속성이 중요한 응용에서 치명적인 성능 저하 요인이 된다. 기존의 학습 기반 융합 방식들은 이러한 문제를 해결하기 위해 복잡한 네트워크 구조 및 대규모 데이터셋 학습을 요구하며, 결과적으로 시스템의 구현 복잡성과 자원 요구량을 증가시키는 경향이 있다. 학습 기반 모델 없이도 강건한 이상치 제거와 효율적인 시점 선택을 가능케 하는 경량 필터링 프레임워크를 제안한다. 관절별 신뢰도 분포에 기반한 Soft-Weighted Thresholding은 적응형 임계값을 제공하여 정밀한 이상치 제거가 가능하며, 시점 간 cosine similarity를 이용한 Redundancy-Aware View Selection은 중복을 줄이고 정보 다양성을 극대화한다. 또한 Temporal View Consistency는 시점 선택의 시간적 안정성을 보장하여 실시간 시스템에서의 활용성을 높인다. 이러한 접근은 기존의 한계를 구조적으로 보완하며, 정확도와 효율성을 동시에 달성할 수 있는 현실적인 대안을 제시한다.

III. 제안하는 3D 자세 복원

본 장에서는 제한된 시점 수와 실제 환경 조건에서도 강건하게 동작할 수 있도록 설계된 3차원 인간 자세 복원 프레임워크를 제안한다. 전체 절차는 학습 기반 모듈 없이도 동작 가능하도록 구성되었으며, 연산 효율성과 정확도를 동시에 고려하였다.

제안하는 핵심 구성 요소는 다음과 같다. 첫째, 관절별 신뢰도 분포를 기반으로 이상치를 완화하는 관절별 신뢰도

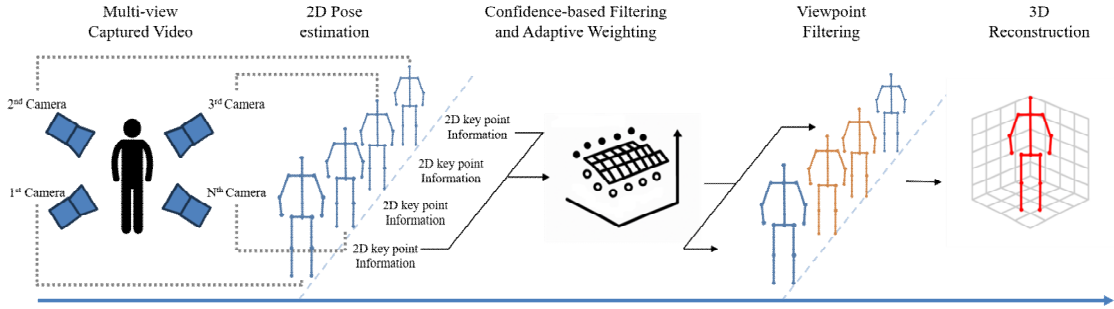


그림 1. 제안한 3차원 자세 복원 프레임워크의 전체 처리 절차 개요도
 Fig. 1. Overview of the overall processing procedure for the proposed 3D pose reconstruction Framework

기본 적응형 임계값 설정 기법, 둘째, 시점 간 코사인 유사도를 활용하여 중복 시점을 제거하는 시점 간 유사도 기반 중복 제거 기법, 셋째, 일정 시간 동안 시점 구성을 고정함으로써 시간적 일관성과 연산 효율을 확보하는 전략이 포함된다.

모든 단계는 대수적 삼각 측량 기반 복원을 중심으로 유기적으로 연결되며, 실시간 적용 가능성과 실제 환경에서의 확장성을 고려하여 설계되었다.

1. 전체 알고리즘의 개요

제안한 3차원 자세 복원 프레임워크는 학습 기반 모듈 없이도 구성할 수 있으며, 간결한 연산 흐름과 명확한 기준에 기반하여 경량화된 구조를 유지한다. 이를 그림 1에 나타내었다. 2차원 자세 추정에는 기존의 범용 관절 검출기를 활용하여 각 관절의 2차원 좌표와 해당 신뢰도 값을 추출하였다. 이 신뢰도 값은 이후 이상치 제거와 시점 선택의 핵심 기준으로 활용된다. 임계값 설정 과정에서는 관절별 신뢰도 분포를 히스토그램 형태로 요약한 뒤, 소프트맥스 기반 가중합을 통해 분포의 중심값을 추정하고 이를 바탕으로 적응형 임계값을 산출하였다. 이 방법은 평균이나 최빈값과 같은 단일 통계값에 의존하지 않고, 분포 전체의 구조를 반영함으로써 이상치의 영향을 효과적으로 억제한다. 산출된 임계값에는 후처리 감산 항을 적용하여 민감도를 조정하였고, 해당 임계값보다 낮은 신뢰도를 가지는 관측값은 삼각 측량 과정에서 제외하였다. 시점 간 중복 제거를 위해 각 시점의 신뢰도 벡터 간 코사인 유사도를 계산하고, 유사

도가 과도하게 높은 시점 쌍 중 하나를 제거하였다. 실험에서는 이 기준값을 0.85에서 0.9 범위로 설정하였으며, 최적의 성능을 보이는 값을 적용하였다. 아울러 선택된 시점 집합은 매 프레임마다 재선택하지 않고 약 1초 간격(30프레임)으로 반복 활용하였다. 이는 짧은 시간 동안 피사체의 구조가 크게 변하지 않는다는 가정 하에 시간적 일관성과 연산 효율을 동시에 확보하기 위한 전략이다.

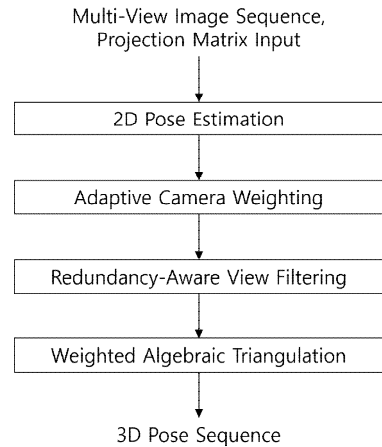


그림 2. 제안한 전체 알고리즘
 Fig. 2. The proposed algorithm

2. 관절별 신뢰도 기반 적응형 필터링

시점 수가 제한된 환경에서 이상치 관절을 효과적으로 제거하기 위해, 관절별 신뢰도 분포에 기반한 소프트 가중 임계값 설정 기법을 제안한다. 본 방법은 각 관절에 대해

시점별 신뢰도 값 중 일정 수준 이상(최소 신뢰도 기준값)을 갖는 값을 선별하고, 그 중 상위 일정 개수를 추출하여 임계값 계산에 활용한다. 이후 선택된 신뢰도 값을 히스토그램으로 변환하고, 온도 조절이 적용된 소프트맥스 함수를 통해 구간별 가중치를 산출한 뒤, 이를 기반으로 기대값을 계산하여 최종 임계값으로 설정한다. 이 방법은 평균, 최댓값 등 단일 대표값에 의존하지 않고 분포 전체의 특성을 고려함으로써 이상치의 영향을 완화하고, 신뢰도 중심의 적응형 이상치 제거를 가능하게 한다. 구체적으로, 신뢰도 분포의 히스토그램에 대해 식 (1)과 같이 소프트맥스 가중 평균을 계산하였다.

$$\theta_j = \sum_{i=1}^B \text{Softmax}\left(\frac{h_i}{\tau}\right) \cdot c_i - \delta \quad (1)$$

여기서 θ_j 는 관절 j 에 대한 임계값이고, c_i 는 구간 i 의 중심값에 해당한다. h_i, τ 및 δ 는 각각 히스토그램 구간의 높이, 소프트맥스 조절 계수, 그리고 후처리 감산 상수이다. 다음으로 본 논문에서 사용한 소프트맥스는 식 (2)와 같이 정의된다.

$$\text{Softmax}(x_i) = \frac{e^{x_i}}{\sum_{j=1}^k e^{x_j}} \quad (2)$$

이 함수는 신뢰도 분포 내에서 상대적으로 높은 빈도를 갖는 구간의 영향을 강조하여 분포 중심에 대한 민감도를 높이며, 조절 계수 τ 값을 조정함으로써 임계값의 민감도(Sharpness)를 제어할 수 있다. 본 논문에서는 이러한 소프트맥스 조절 계수(τ)와 후처리 감산 상수(δ)의 값을 CMU Panoptic Studio 데이터셋을 활용한 반복 실험을 통해 최적화하였으며, 그 결과 가장 낮은 MPJPE를 기록하는 조합을 최종 파라미터로 선정하였다.

추출된 신뢰도 값들의 분산이 매우 낮은 경우, 해당 분포는 의미 있는 정보로 간주되지 않으며, 이때는 가장 낮은 신뢰도 값에 일정 값을 감산한 값을 임시 임계값으로 설정한다. 이는 식 (1)의 조건으로 처리된다.

$$\text{if } \max(\text{conf}) - \min(\text{conf}) < \epsilon \Rightarrow \theta_j \leftarrow \min(\text{conf}) - 0.1 \quad (3)$$

또한, 신뢰도 분포가 좁거나 전반적으로 낮은 경우에는 앞서 제시한 소프트 가중 방식으로 계산된 임계값이 0 또는 음수로 출력되어, 신뢰도 값이 0인 관절이 이상치 제거 단계를 통과하는 문제가 발생할 수 있다. 이러한 현상을 방지하기 위해 관절별 임계값의 하한선을 식 (4)와 같이 설정하였다.

$$\theta_j \leftarrow \max(\theta_j, 0.05) \quad (4)$$

이를 통해 최소한의 신뢰도를 갖추지 못한 관절이 이상치 제거 단계를 우회하여 재구성 과정에 포함되는 것을 효과적으로 방지할 수 있다.

그림 3에서 제시된 알고리즘은 관절별 2D 신뢰도 정보를 기반으로 적응형 필터링 임계값을 산출하는 절차를 단계적으로 설명하고 있다. 먼저, 각 관절에 대해 여러 시점에서 계산된 2D 신뢰도 값들을 입력으로 받아, 전체 신뢰도 중에서 가장 낮은 값을 가지는 시점들을 우선적으로 제거하는 최소 신뢰도 필터링(Min-confidence filtering) 단계를 수행한다. 이후 남은 시점들 중 상위 k 개의 높은 신뢰도 값을 선택하는 Top-k confidence selection을 통해, 신뢰도가 비교적 안정적인 시점들만을 기준으로 삼는다. 그다음 단계에서는 선택된 신뢰도 분포에서 특정 백분위(percentile) 값을 기준으로 적응형 임계값(adaptive threshold)을 설정하며, 이 값은 고정된 기준이 아닌 데이터 분포에 따라 동적으

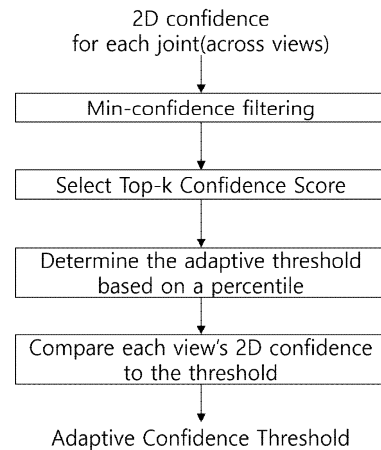


그림 3. 소프트 가중 기반 적응형 임계값 설정 방법 절차도
Fig. 3. Soft-weighted based adaptive threshold setting method procedure diagram

로 변화한다는 특징이 있다. 마지막으로, 각 시점에서의 2D 신뢰도 값과 계산된 임계값을 비교하여, 임계값 이하인 관절을 이상치로 판단하고 제거하는 필터링을 수행한다. 이와 같은 방식은 전체 시점의 품질 편차를 반영하면서도, 강건하고 데이터 기반적인 임계값 설정을 통해 신뢰성 있는 관절 선택을 가능하게 하며, 3D 재구성의 정확도를 높이는 데 핵심적인 역할을 한다.

3. 시점 간 유사도 기반 중복 제거

모든 시점을 동일하게 사용하는 대신, 시점 간 정보 중복을 줄이기 위해 각 시점의 신뢰도 벡터 간 코사인 유사도를 계산하였다. 계산된 유사도를 기반으로, 유사도가 가장 낮은 시점 중에서 대표성과 다양성을 확보할 수 있도록 일정 개수의 시점을 선택하여 재구성에 활용하였다. 이러한 과정은 연산량을 효과적으로 감소시키는 동시에, 학습 기반 모델 없이도 시점 선택의 타당성을 확보할 수 있다. 카메라 시점 수가 많을수록 3차원 복원 정확도는 향상될 수 있으나, 유사도가 매우 높은 시점을 중복으로 사용할 경우 추가적인 정보 이득은 제한적인 반면, 불필요한 연산량은 증가하게 된다. 본 논문에서는 시점별 신뢰도 벡터 $v_c = [conf_{(c,1)}, conf_{(c,2)}, \dots, conf_{(c,n)}]$ 에 대해 식 (5)와 같이 정의한다.

$$CosSim(v_c, v_{c'}) = (v_c \cdot v_{c'}) / (\|v_c\| \times \|v_{c'}\|) \quad (5)$$

이를 계산하고, 특정 임계값(예: 0.9)을 초과하는 시점 쌍 (c, c') 중 하나를 제거함으로써 중복된 시점을 걸러낸다. 본

연구에서는 CMU Panoptic Studio 데이터셋을 활용하여 다양한 임계값 구간(0.85~0.95)을 비교하였으며, 그 중 MPJPE가 최소화된 0.9를 최종값으로 채택하였다. 이를 통해 소수의 다양한 시점을 선택하여 알고리즘의 연산 효율성을 향상시킨다.

선택된 시점 구성은 약 30프레임(약 1초) 동안 유지되며, 매 프레임마다 시점을 재선택하지 않는다. 이는 짧은 시간 동안 피사체의 구조가 크게 변하지 않는다는 가정에 기반하며, 시간적으로 일관된 시점 구성을 반복적으로 활용함으로써 3차원 자세 시퀀스의 연속성을 확보하고 전체 연산 효율성을 향상시킨다.

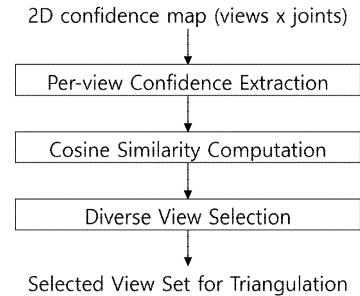


그림 4. 시점 간 신뢰도 유사도 비교를 통한 중복 최소화 기반 시점 필터링 기법
 Fig. 4. View filtering technique based on redundancy minimization through inter-view confidence similarity comparison

IV. 실험 결과

본 절에서는 제안한 기법의 정확도 및 연산 효율성을 검

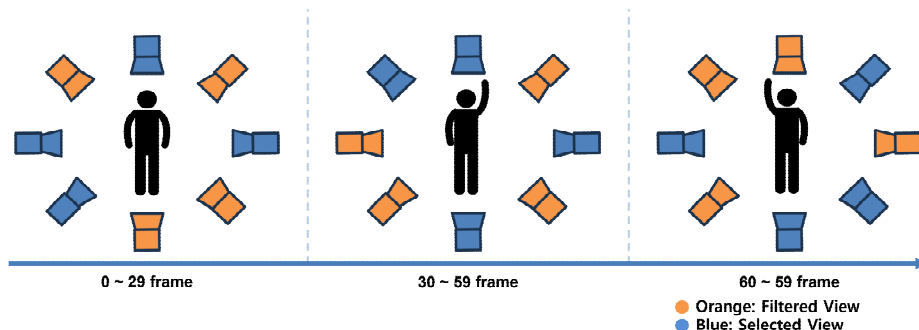


그림 5. 시간 일관성을 고려한 시점 선택 전략 개념도
 Fig. 5. Conceptual Diagram of View Selection Strategy Considering Temporal Consistency

증하기 위해, 다중 시점 기반 3차원 자세 복원 분야에서 일반적으로 사용되는 벤치마크 데이터셋인 CMU Panoptic Studio를 활용하여 정량적 및 정성적 분석을 수행하였다. 비교 대상은 기존 대수적 삼각 측량 기반 복원 기법이며, 본 논문에서 제안한 적응형 임계값 설정 기법의 효과를 다양한 관점에서 평가하였다.

1. 실험 환경 및 평가 지표

실험에 사용된 CMU Panoptic 데이터셋은 총 31개의 동기화된 RGB 카메라를 통해 다중 시점 영상을 제공하며, 각 프레임에는 정확한 3차원 관절 위치 주석이 포함되어 있다. 모든 프레임은 초당 30프레임, 해상도 1920×1080 조건에서 촬영되었다. 2차원 자세 추정에는 기존의 범용 관절 검출기를 활용하여 수행하였으며, 추정된 관절 위치와 신뢰도 값을 기반으로 이상치 제거와 삼각 측량을 적용하였다. 성능 평가는 MPJPE(Mean Per Joint Position Error), PA-MPJPE(Procrustes aligned MPJPE) 그리고 NMPJPE(Normalized MPJPE)의 정량적 지표를 기준으로 수행하였다. MPJPE(Mean Per Joint Position Error)는 예측된 3D 관절 위치와 ground truth 간의 평균 유클리디안 거리(단위: cm)를 측정하는 것이고, PA-MPJPE(Procrustes aligned MPJPE)는 정렬(회전, 이동, 스케일 보정)을 적용한 후의 MPJPE에 해당한다. 마지막으로, NMPJPE(Normalized MPJPE)는 스케일만 정규화한 MPJPE이다.

2. 적응 임계값 설정 기법의 성능 분석

제안한 관절별 신뢰도 기반 소프트 가중 임계값 설정 기법의 이상치 제거 과정과 그 효과를 시각적으로 분석하였다. 기존 고정 임계값 방식은 모든 관측값에 동일한 임계값을 적용하여, 관절별 신뢰도 분포의 차이를 충분히 반영하지 못하고 기준이 과도하거나 부족하게 설정될 위험이 있었다. 반면, 본 논문에서 제안한 소프트 가중 임계값 설정 방식은 관절별 분포 특성을 고려하여 임계값을 적응적으로 결정함으로써 이상치 제거 안정성과 효과를 동시에 향상시킬 수 있었다. 그림 6은 관절 번호 2번과 9번에서 얻어진 신뢰도 값의 분포를 히스토그램으로 나타내고, 평균, 최빈값, 그리고 제안한 소프트 가중 임계값을 함께 표시하였다. 제안 기법은 평균의 대표성과 최빈값의 강건성 간 균형을 유지하도록 설계되었으며, 관측값 분포의 구조적 특성을 일부 반영한다.

그림 7은 임계값 설정 결과를 관절 번호와 시점 번호를 기준으로 시각화한 그래프이다. 초록색 점은 소프트 가중 임계값 이상으로 판단되어 재구성에 활용된 관측값을, 빨간색 점은 임계값 이하로 판단되어 제거된 관측값을 나타낸다. 검은색 선은 각 관절에 대해 설정된 소프트 가중 임계값을 의미한다. 대부분의 관절에서 낮은 신뢰도의 관측값이 효과적으로 제거되고, 높은 신뢰도의 주요 관측값은 유지되는 경향이 확인되었다. 특히 특정 관절(예: 15번 관절)에서 신뢰도 편차가 크고 이상치가 존재하는 경우, 이상치

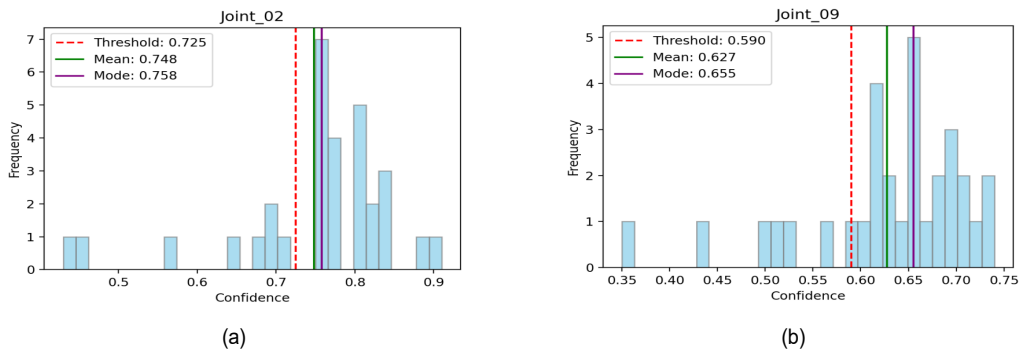


그림 6. 관절별 신뢰도 분포 및 임계값 시각화 결과. 분포 내에서 평균(초록색), 최빈값(보라색), 소프트 가중 임계값(빨간색). (a) 2번 조인트의 실험 결과 (b) 9번 조인트의 실험 결과

Fig. 6. Joint-wise confidence distribution and threshold visualization results. The mean (green), mode (purple), and soft-weighted threshold (red) are displayed together within the distribution. Experimental result of (a) the second joint and (b) the ninth joint

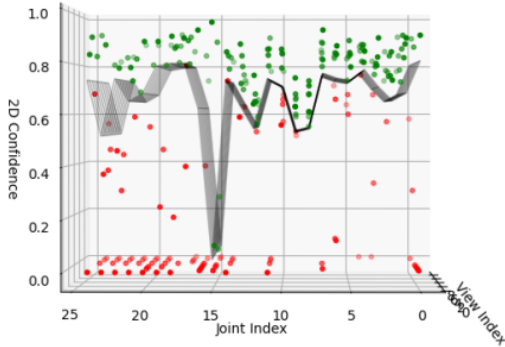
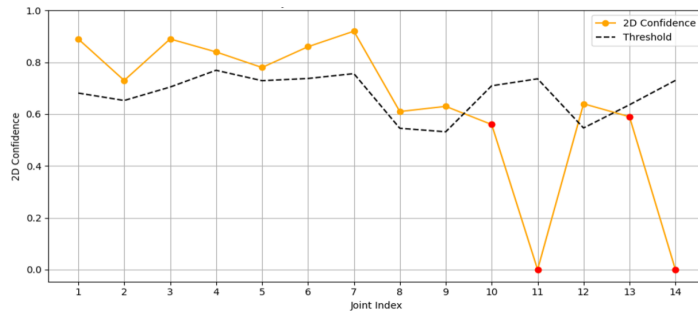


그림 7. 관절별 시점에 대한 2D 신뢰도 분포 및 임계값 설정 결과. 관절 번호와 시점 번호 평면에서 소프트 가중 임계값 이상인 관측값(초록색 점)은 재구성에 활용되었으며, 임계값 이하인 관측값(빨간색 점)은 제거되었다. 검은색 선은 각 관절별로 설정된 소프트 가중 임계값을 나타낸다.

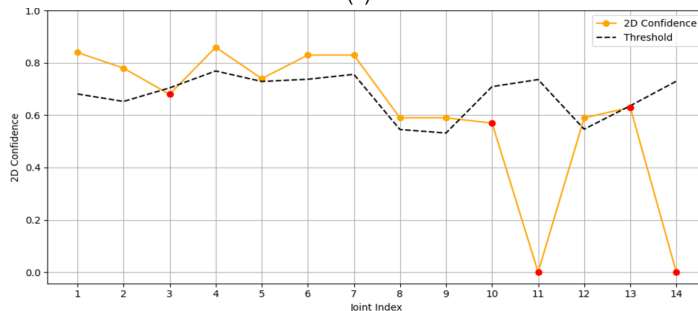
Fig. 7. 2D confidence distribution and threshold setting results for joint-wise Viewpoints. In the joint number and viewpoint number plane, observations above the soft-weighted threshold (green dot) were utilized for reconstruction, while observations below the threshold (red dot) were removed. The black line represents the soft-weighted threshold set for each joint.

제거 과정을 통해 이러한 값들의 영향이 효과적으로 억제되었다. 이러한 결과는 이상치 제거 과정이 관절별 신뢰도 분포에 기반하여 설정된 적응형 임계값에 따라 선택적으로 작동함을 시각적으로 검증하는 사례이다. 또한 소프트 가중 임계값 설정 기법이 평균, 최빈값 등 단일 기준값 기반 이상치 제거 방식에 비해 이상치에 덜 민감하게 동작할 가능성을 시사한다. 다만, 이상치 제거 임계값은 각 관절의 분포 특성과 관측 노이즈 수준에 따라 추가적인 보완 및 조건별 조정이 필요할 수 있다.

그림 8은 2D 관절 신뢰도와 적응형 임계값의 관계를 시각화하여, 특정 시점에서의 관절 신뢰도 기반 이상치 제거 전략의 효과를 보여준다. 그림 8 (a)는 0번 시점의 결과로, 전반적으로 높은 신뢰도를 보이는 관절들이 다수 존재하지만, 9번 이후의 관절들에서 신뢰도가 급격히 하락하며 임계값을 하회하고 있으며, 특히 11번과 14번 관절은 2D 신뢰도가 0에 가까워 명백한 이상치로 판단하고 제거하였다. 그



(a)



(b)

그림 8. 전체 31개 시점 중 일부 시점의 2차원 신뢰도 및 적응형 임계값 시각화. 임계값 이상인 관절(노란색)과 이상치로 제거된 관절(빨간색) (a) 0번 시점 (b) 7번 시점

Fig. 8. 2D confidence and adaptive threshold visualization for some viewpoints among the total 31 viewpoints. joints above the threshold (yellow) and joints removed as outliers (red) were distinguished and displayed (a) the first viewpoint (b) the eighth viewpoint

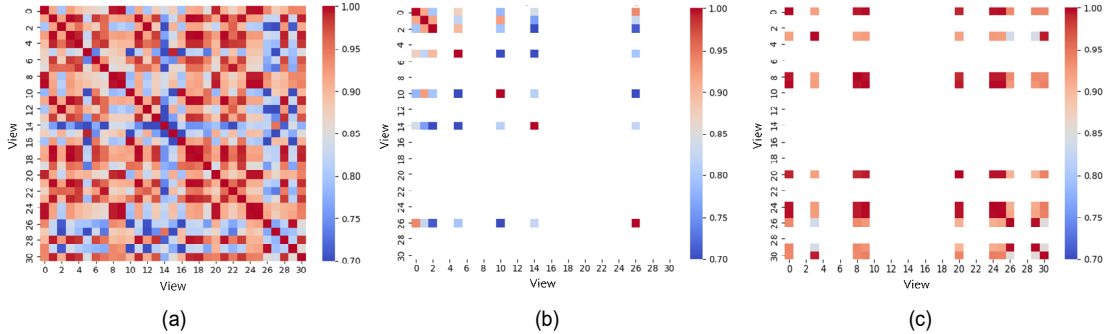


그림 9. 코사인 유사도 분석 결과. (a) 전체 시점 간 유사도 (b) 유사 시점 선택 결과 (c) 분산된 시점 선택 결과
 Fig. 9. Cosine similarity analysis results. (a) overall Inter-view similarity (b) similar view selection results (c) distributed view selection results

림 8 (b)의 7번 시점 역시 유사한 경향을 보이며, 신뢰도가 낮은 관절들이 임계값보다 아래에 위치하여 제거되었고, 특히 11번과 14번은 두 시점 모두에서 공통적으로 제거된 것으로 나타나 특정 관절이 복잡한 시각 조건이나 피사체의 자세에 따라 안정적인 검출이 어려울 수 있다. 이처럼 제안된 적응형 임계값 기반 필터링은 시점별, 관절별 신뢰도 편차를 반영하여 이상치를 효과적으로 제거하며, 이는 3D 재구성 단계에서의 오류 누적을 줄이고 전체적인 정밀도를 높일 수 있다.

표 1은 관측 시점 수와 시점 선택 방식(similar vs. diverse)에 따른 3D 포즈 추정 성능(MPJPE)을 비교한 것으로, 다양한 시점을 선택하는 것이 유사한 시점 선택보다 현저히 낮은 오차를 기록함을 보여준다. 관측 시점이 3개일 때, 유사한 시점 선택은 19.6025의 매우 높은 MPJPE를 보였지만, 다양한 시점 선택은 이를 3.3225로 크게 감소시켰으며, 이는 약 83%의 오류 감소에 해당한다. 관측 시점 수가 증가할수록 두 방법 모두 오차가 감소하는 경향을 보였지만, diverse 방식은 항상 similar 방식보다 뛰어난 성능을 유지하였다. 예를 들어, 6개의 시점을 사용할 경우에도 similar은 10.9116, diverse는 2.2157로 5배 이상의 차이를 나타낸다. 이러한 결과는 시점의 수만큼이나 시점 간 다양성의 확보가 포즈 추정 정확도에 결정적인 요소라는 것을 보여준다.

또한, 표 1은 제안 기법을 최신 학습 기반 방법인 AdaFuse와 비교한 결과도 함께 보여준다. AdaFuse 대비 제안 방법은 모든 시점 수에서 일관되게 낮은 MPJPE를 기

록하며, 특히 분산 시점 선택(diverse) 조건에서 현저한 성능 우위를 보인다. 이는 제안 기법이 비학습 기반의 경량 프레임워크임에도 불구하고 최신 학습 기반 접근법보다도 높은 정확도를 확보하였음을 의미하며, 적은 수의 시점에서 강건한 성능을 유지할 수 있음을 입증한다.

표 1. 유사 시점, 분산 시점 선택 및 AdaFuse와의 3차원 자세 복원 성능 비교

Table 1. Comparison of 3D pose reconstruction performance according to similar view, distributed view selection, and AdaFuse

# of view	Method	MPJPE
3	AdaFuse	19.5
3	similar	19.6025
3	diverse	3.3225
4	AdaFuse	14.0
4	similar	15.1402
4	diverse	2.5078
5	AdaFuse	13.8
5	similar	13.0029
5	diverse	2.3976
6	AdaFuse	13.5
6	similar	10.9116
6	diverse	2.2157

표 2에서 Baseline은 전체 31개 시점을 사용하는 기존 방법의 결과이며, 이는 최대 성능(reference upper bound)으로 설정하였다. 3~6개 시점에서는 동일한 시점 선택 조건(분산 시점 집합)에서 필터링 적용 유무에 따른 성능 및 처리 속도를 비교하였다. 관측 시점 수에 따른 3D 포즈 추정 성

표 2. 임계값 설정 기법 적용 유무에 따른 성능 비교

Table 2. Performance comparison according to the application of threshold setting technique

# of view	Method	MPJPE	PA-MPJPE	NMPJPE
31	Baseline (full-view)	1.4920	0.0077	1.4985
3	Cos-sim only	3.9812	0.0248	3.9821
3	Ours	3.3225	0.0195	3.3367
4	Cos-sim only	2.5288	0.0154	2.5206
4	Ours	2.5078	0.0142	2.4950
5	Cos-sim only	2.4707	0.0146	2.4730
5	Ours	2.3976	0.0132	2.3935
6	Cos-sim only	2.3082	0.0134	2.3131
6	Ours	2.2157	0.0118	2.2055

능을 비교한 것으로, MPJPE, PA-MPJPE, NMPJPE 세 지표 모두에서 제안 기법(Ours)이 Cos-sim only 방식보다 일관되게 낮은 오차를 기록하며 우수한 성능을 보였다. 특히 관측 시점 수가 3개일 때 제안 기법은 MPJPE 기준 약 16.5%의 오차 감소를 보이며 큰 개선 효과를 나타냈고, 관측 시점이 증가함에 따라 오차는 점진적으로 감소하는 경향을 보였다. 6개의 시점을 사용할 경우, 제안 기법은 Full-view(31 views) 대비 약간의 정확도 손실만을 보이면서도 훨씬 적은 관측 수로 효율적인 포즈 추정이 가능함을 입증하였다. 이러한 결과는 제안 기법이 제한된 시점에서도 견고한 성능을 발휘할 수 있는 효과적인 방식임을 보였다.

그림 10은 제안한 방식(빨강), 기존 방식(파랑), 그리고 실제 정답(Ground Truth, 검정)의 3D 포즈 재구성 결과를 비교한 시각적 예시로, 제안한 방식의 정밀도와 구조적 일관성을 직관적으로 보여준다. 전체 신체 실루엣부터 팔, 다리, 무릎, 팔꿈치 등 주요 관절을 확대하여 비교한 결과, 빨간 점과 선으로 표시된 제안한 방식의 관절 위치는 대부분의 부위에서

검정색 GT와 거의 정확히 일치하며, 파란색 기존 방식보다 훨씬 정렬이 잘 되어 있음을 알 수 있다. 특히, 팔과 다리의 관절 연결 선이 GT의 방향성과 각도를 더 충실히 따르고 있으며, 이는 단순 위치 오차뿐 아니라 인체 형태 복원 능력 측면에서 제안 방식이 우수하다는 것을 보여주었다. 또한 확대된 부분에서는 기존 방식에서 관절이 비정상적으로 휘거나 방향이 틀어진 경우가 보이는 반면, 제안 방식은 GT와 거의 평행하게 재구성되어 있어서 보다 자연스럽게 일관된 자세 추정 결과를 제공함을 확인할 수 있다.

V. 결론

본 논문에서는 다중 시점 기반 3차원 인간 자세 추정 과정에서 발생하는 이상치, 신뢰도 편차, 시점 중복, 시간적 불안정성과 같은 문제를 해결하기 위한 경량화된 필터링 프레임워크를 제안하였다. 관절별 신뢰도 분포에 기반한

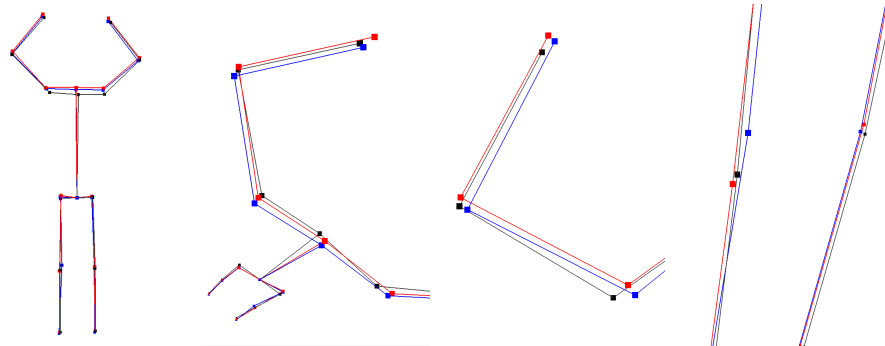


그림 10. 제안 방식에 따른 3D 스켈레톤 재구성 결과 비교

Fig. 10. Comparison of 3D Skeleton Reconstruction Results According to the Proposed Method

적응형 임계값 설정을 통해 이상치를 효과적으로 제거하고, 시점 간 코사인 유사도에 기반한 중복 시점 제거 기법을 도입하여 정보의 중복을 최소화하였다. 또한, 일정 시간 동안 동일한 시점 구성을 유지하는 전략을 적용함으로써 시간적 일관성과 연산 효율을 동시에 확보하였다. 제안한 전략은 별도의 학습 과정 없이도 적용 가능하며, CMU Panoptic Dataset을 활용한 실험을 통해 기존 방법 대비 향상된 정확도와 계산 효율을 입증하였다. 특히, 제안된 기법은 제한된 수의 시점 및 실환경 조건에서도 안정적인 성능을 유지할 수 있어, 실시간 처리나 모바일 환경에서의 응용 가능성을 높인다. 본 기법은 특히 사용 가능한 시점 수가 3개에서 5개 이하로 제한된 상황에서 기존의 고정 임계값 방식보다 더 높은 정밀도를 달성하였다. 그러나 시점 수가 충분하여 정보가 과잉 제공되는 경우에는 분포 기반 임계값 설정의 상대적 이점이 감소하는 경향이 있으며, 일부 관절에서는 필터링 효과가 제한적으로 나타날 수 있다. 또한, 제안한 시점 고정 전략은 단기적인 구조적 안정성을 전제로 하므로, 빠른 자세 변화가 빈번하게 발생하는 시나리오에서는 시점 갱신 조건에 대한 추가적인 조정이 필요할 수 있다. 향후 연구에서는 임계값 설정 계산의 동적 조건 최적화, 관절별 이력 기반의 시점 유지 전략 등 보다 일반화된 적용을 위한 확장 방안을 모색할 수 있을 것이다.

참 고 문 헌 (References)

- [1] A. Toshev and C. Szegedy, "DeepPose: Human pose estimation via deep neural networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, Jun. 2014, pp. 1653-1660.
doi: <https://doi.org/10.1109/CVPR.2014.214>
- [2] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, Jun. 2019, pp. 5693-5703.
doi: <https://doi.org/10.1109/CVPR.2019.00584>
- [3] Z. Cao, G. Hidalgo, T. Simon, S. Wei, and Y. Sheikh, "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 1, pp. 172-186, Jan. 2021.
doi: <https://doi.org/10.1109/TPAMI.2019.2929257>
- [4] D. C. Luvizon, D. Picard, and H. Tabia, "2D/3D pose estimation and action recognition using multitask deep learning," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, Jun. 2018, pp. 5137-5146.
doi: <https://doi.org/10.1109/CVPR.2018.00539>
- [5] X. Xu, H. Chen, F. Moreno-Noguer, L. A. Jeni, and F. De la Torre, "3D human shape and pose from a single low-resolution image with self-supervised learning," in Proceedings of the European Conference on Computer Vision (ECCV), Glasgow, UK, Aug. 2020, pp. 284-300.
doi: https://doi.org/10.1007/978-3-030-58545-7_17
- [6] Q. Shuai, C. Geng, Q. Fang, S. Peng, W. Shen, X. Zhou, and H. Bao, "Novel view synthesis of human interactions from sparse multi-view videos," in Proceedings of the ACM SIGGRAPH Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '22), Vancouver, BC, Canada, Jul. 2022, pp. 1-10.
doi: <https://doi.org/10.1145/3528233.3530704>
- [7] B. Cheng, B. Xiao, J. Wang, H. Shi, T. S. Huang, and L. Zhang, "HigherHRNet: Scale-aware representation learning for bottom-up human pose estimation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, Jun. 2020, pp. 5386-5395.
doi: <https://doi.org/10.1109/CVPR42600.2020.00543>
- [8] H.-S. Fang, J. Li, H. Tang, C. Xu, H. Zhu, Y. Xiu, Y.-L. Li, and C. Lu, "AlphaPose: Whole-body regional multi-person pose estimation and tracking in real-time," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 6, pp. 7157-7173, Jun. 2023.
doi: <https://doi.org/10.1109/TPAMI.2022.3222784>
- [9] J. Martinez, R. Hossain, J. Romero, and J. J. Little, "A simple yet effective baseline for 3D human pose estimation," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, Oct. 2017, pp. 2640-2649.
doi: <https://doi.org/10.1109/ICCV.2017.288>
- [10] G. Moon and K. M. Lee, "I2L-MeshNet: Image-to-lixel prediction network for accurate 3D human pose and mesh estimation from a single RGB image," in Proceedings of the European Conference on Computer Vision (ECCV), Glasgow, UK, Aug. 2020, pp. 752-768.
doi: https://doi.org/10.1007/978-3-030-58571-6_44
- [11] M. R. I. Hossain and J. J. Little, "Exploiting temporal information for 3D human pose estimation," in Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, Sep. 2018, pp. 68-84.
doi: https://doi.org/10.1007/978-3-030-01249-6_5
- [12] D. Tome, C. Russell, and L. Agapito, "Lifting from the deep: Convolutional 3D pose estimation from a single image," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, Jul. 2017, pp. 2500-2509.
doi: <https://doi.org/10.1109/CVPR.2017.603>
- [13] A. Kadkhodamohammadi, A. Gangi, M. de Mathelin, and N. Padov, "A multi-view RGB-D approach for human pose estimation in operating rooms," in Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), Santa Rosa, CA, USA, Mar. 2017, pp. 303-311.
doi: <https://doi.org/10.1109/WACV.2017.47>

- [14] Z. Dong, J. Song, X. Chen, C. Guo, and O. Hilliges, "Shape-aware multi-person pose estimation from multi-view images," in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, Oct. 2021, pp. 11158-11167. doi: <https://doi.org/10.1109/ICCV48922.2021.01097>
- [15] D. Mehta, H. Rhodin, D. Casas, P. Fua, O. Sotnychenko, W. Xu, and C. Theobalt, "VNect: Real-time 3D human pose estimation with a single RGB camera," ACM Transactions on Graphics, vol. 36, no. 4, pp. 44:1-44:14, Jul. 2017. doi: <https://doi.org/10.1145/3072959.3073596>
- [16] H. Tu, C. Wang, and W. Zeng, "VoxelPose: Towards multi-camera 3D human pose estimation in wild environment," in Proceedings of the European Conference on Computer Vision (ECCV), Glasgow, UK, Aug. 2020, pp. 197-212. doi: https://doi.org/10.1007/978-3-030-58452-8_12
- [17] G. Pavlakos, X. Zhou, K. G. Derpanis, and K. Daniilidis, "Coarse-to-fine volumetric prediction for single-image 3D human pose," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, Jul. 2017, pp. 7025-7034. doi: <https://doi.org/10.1109/CVPR.2017.139>
- [18] K. Isakov, E. Burkov, V. Lempitsky, and Y. Malkov, "Learnable triangulation of human pose," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), Seoul, South Korea, Oct. 2019, pp. 7718-7727. doi: <https://doi.org/10.1109/ICCV.2019.00781>
- [19] K. Bartol, D. Bojanić, T. Petković, and T. Pribanić, "Generalizable human pose triangulation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, Jun. 2022, pp. 11028-11037. doi: <https://doi.org/10.1109/CVPR52688.2022.01075>
- [20] W. Shan, Z. Lu, Q. Wang, A. Bovik, and L. Zhang, "P-STMO: Pre-trained spatial temporal many-to-one model for 3D human pose estimation," in Proceedings of the European Conference on Computer Vision (ECCV), Tel Aviv, Israel, Oct. 2022, pp. 461-478. doi: https://doi.org/10.1007/978-3-031-20065-6_27
- [21] Z. Zhang, C. Wang, W. Qin, and W. Zeng, "AdaFuse: Adaptive multiview fusion for accurate human pose estimation in the wild," International Journal of Computer Vision, vol. 129, no. 3, pp. 703-718, Mar. 2021. doi: <https://doi.org/10.1007/s11263-020-01398-9>
- [22] Z. Wang, X. Nie, X. Qu, Y. Chen, and S. Liu, "Distribution-aware single-stage models for multi-person 3D pose estimation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, Jun. 2022, pp. 11038-11047. doi: <https://doi.org/10.1109/CVPR52688.2022.01275>
- [23] D. Pavlo, C. Feichtenhofer, D. Grangier, and M. Auli, "3D human pose estimation in video with temporal convolutions and semi-supervised training," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, Jun. 2019, pp. 7753-7762. doi: <https://doi.org/10.1109/CVPR.2019.00794>
- [24] K. Zhou, X. Han, W. Wang, and J. Lu, "HEMlets pose: Learning part-centric heatmap triplets for accurate 3D human pose estimation," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), Seoul, South Korea, Oct. 2019, pp. 2344-2353. doi: <https://doi.org/10.1109/ICCV.2019.00243>

저 자 소 개



박 정 탁

- 2022년 2월 : 광운대학교 전자재료공학과학과 졸업(공학사)
- 2025년 8월 : 광운대학교 일반대학원 졸업(공학석사)
- ORCID : <https://orcid.org/0000-0002-3825-2428>
- 주관심분야 : 멀티뷰 카메라 캘리브레이션, 딥러닝 기반 3D 포즈 추정, 3D 그래픽스



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

저 자 소 개



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