

芭蕾呪法

Phase 2: 3D 포즈 비교 실험 리포트

2026-02-16 | /sc:duo 자동 생성

1 작업 요약

Phase 1에서 확인된 카메라 앵글 민감성 문제를 해결하기 위해, MediaPipe의 3D world_landmarks를 활용한 비교 모드를 구현하고, 2D vs 3D 비교 실험을 수행하였다.

핵심 가설: 3D world_landmarks는 골반 중심 기준 미터 단위 좌표이므로, 카메라 각도에 의한 2D 투영 왜곡이 제거되어 더 정확한 포즈 비교가 가능하다.

2 구현 내용

2.1 수정된 모듈

파일	변경 내용	담당
src/comparator.py	use_world 파라미터 추가, calculate_l2_distance_3d() 함수 구현, 3D 모드에서 정규화 자동 비활성화	Gemini
src/pose_extractor.py	Phase 2-1에서 이미 world_landmarks 출력 지원 완료 (변경 없음)	기존
tests/test_phase2.py	6개 테스트 케이스 신규 작성	Gemini

2.2 핵심 설계 결정

- 3D 거리 함수:** $\sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}$ (3축 유clidean 거리)
- 정규화 불필요:** world_landmarks는 이미 골반 중심 기준 + 미터 단위이므로 별도 정규화 없이 직접 비교
- 하위 호환성:** use_world=False(기본값)이면 기존 2D 파이프라인과 동일하게 동작

3 테스트 결과

#	테스트 항목	결과	유형
1	test_calculate_l2_distance_3d_basic	PASS	단위
2	test_calculate_l2_distance_3d_same_point	PASS	단위
3	test_compare_poses_use_world_metadata	PASS	통합
4	test_compare_poses_2d_backward_compatible	PASS	통합
5	test_3d_loss_lower_than_2d	PASS	통합
6	test_compare_poses_3d_no_normalization	PASS	통합

전체 테스트 스위트: 18 passed, 3 skipped (Phase 0 + Phase 2 + PM Agent 통합)

4 실험: 2D vs 3D 비교

4.1 실험 설정

동일한 그랑 바뜨망(Grand Battement) 영상 쌍을 2D 정규화 비교와 3D world_landmarks 비교로 각각 처리하여 결과를 비교하였다.

- 참조 영상: reference_grand_battement.mp4 (301 프레임, 24fps)
- 비교 영상: compare_grand_battement.mp4 (301 프레임, 24fps)
- 비교 프레임 수: 301 프레임 (동일)

4.2 통계 비교

지표	2D	3D	변화율
평균 손실 (mean)	4.6161	0.2916	-93.7%
최대 손실 (max)	11.4594	0.5682	-95.0%
최소 손실 (min)	0.0000	0.0000	-
표준 편차 (std)	0.9537	0.1411	-85.2%
최대 손실 프레임	49	40	-
최소 손실 프레임	12	12	동일

핵심 발견: 3D 비교의 평균 손실이 2D 대비 **약 15.8배 감소** ($4.62 \rightarrow 0.29$). 이는 2D 비교에서의 높은 손실이 상당 부분 카메라 앵글 차이에 기인했음을 실증적으로 확인한 것이다.

4.3 프레임별 손실 비교

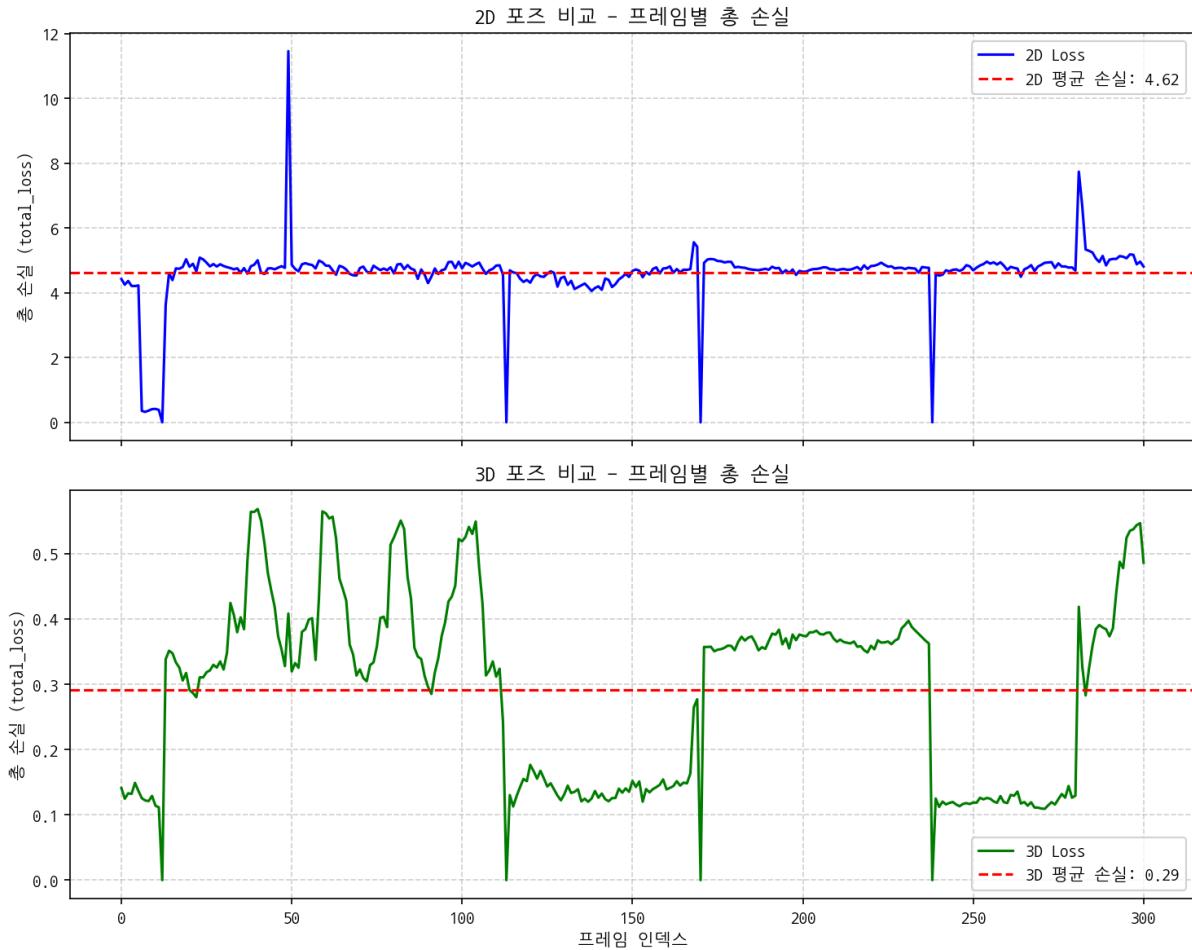


Figure 1: 2D vs 3D 프레임별 총 손실. 3D(하단)에서 손실 범위가 0.057로 대폭 축소됨.

4.4 2D/3D 손실 중첩 그래프

2D 및 3D 포즈 비교 - 프레임별 총 손실 중첩

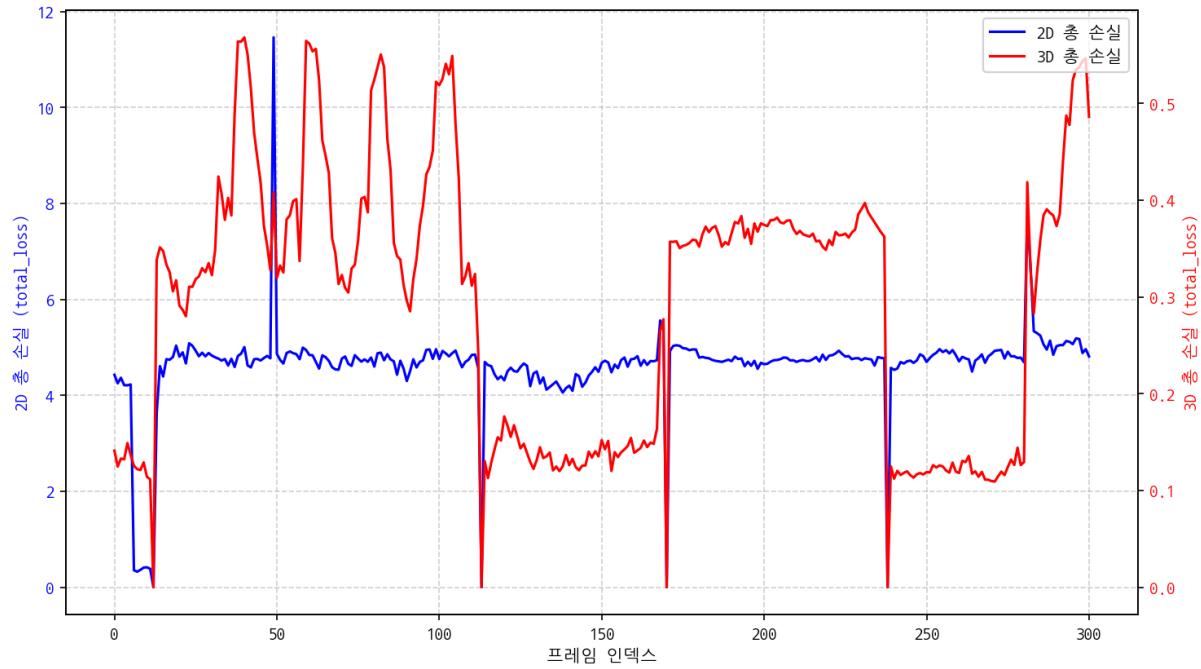


Figure 2: 2D(파란색)와 3D(빨간색) 손실을 이중 Y축으로 중첩. 3D가 더 세밀한 동작 차이를 포착함.

4.5 관절별 손실 분석 (3D)

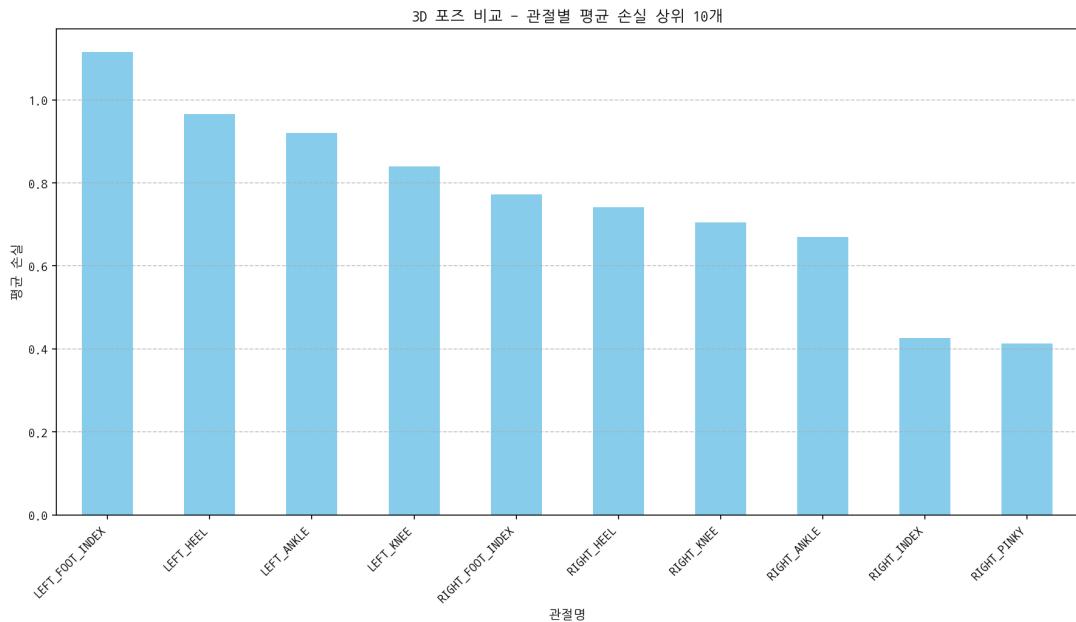


Figure 3: 3D 비교에서의 관절별 평균 손실 상위 10개. 하체 관절이 여전히 최상위이나 절대값이 대폭 감소.

4.6 통계 요약 비교

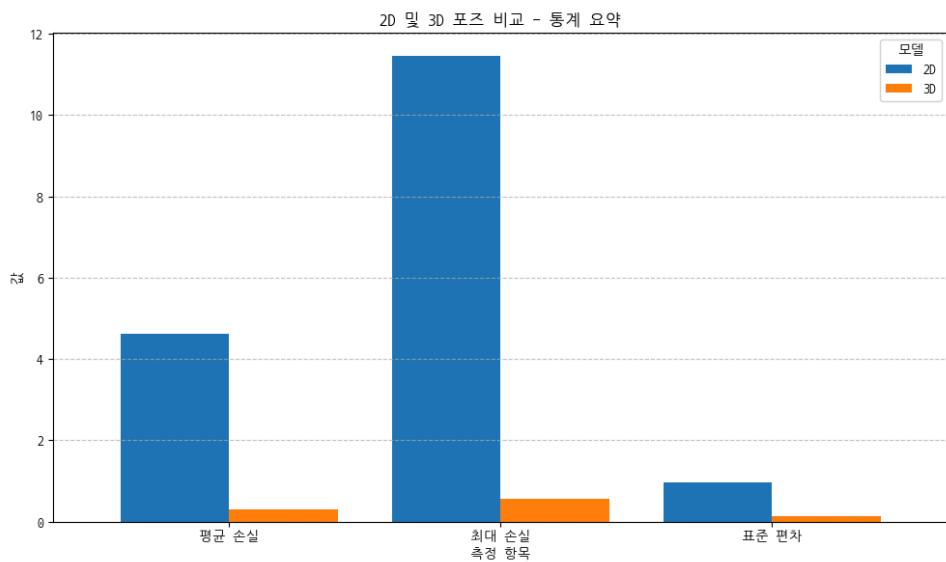


Figure 4: 2D와 3D의 주요 통계 비교. 모든 지표에서 3D가 현저히 낮은 값을 보임.

5 3D 스켈레톤 비교 시각화

5.1 최대 손실 프레임 (Frame 40)

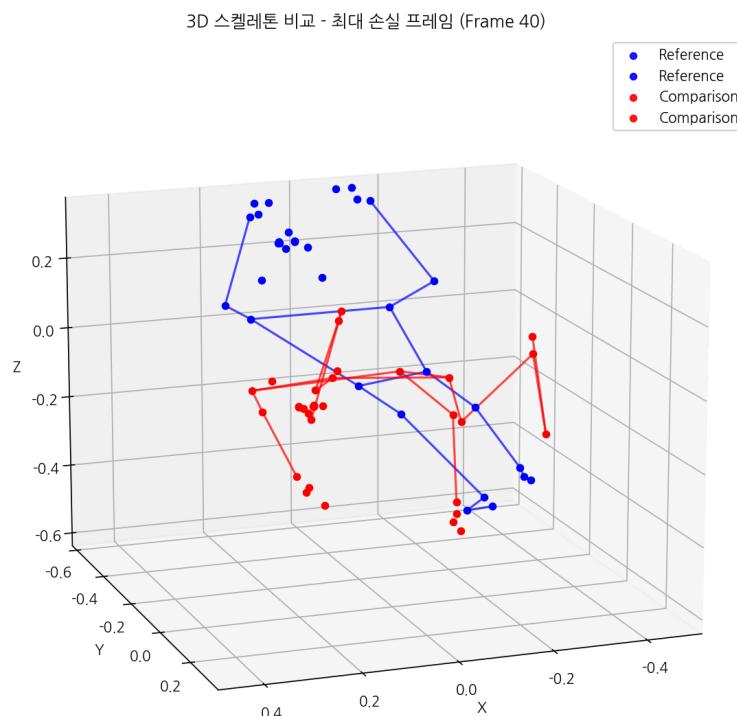


Figure 5: 3D 스켈레톤 오버레이 - 최대 손실 프레임. 참조(파란색)와 비교(빨간색)의 하체 위치 차이가 뚜렷함.

5.2 최소 손실 프레임 (Frame 12)

3D 스켈레톤 비교 - 최소 손실 프레임 (Frame 12)

Comparison
Comparison

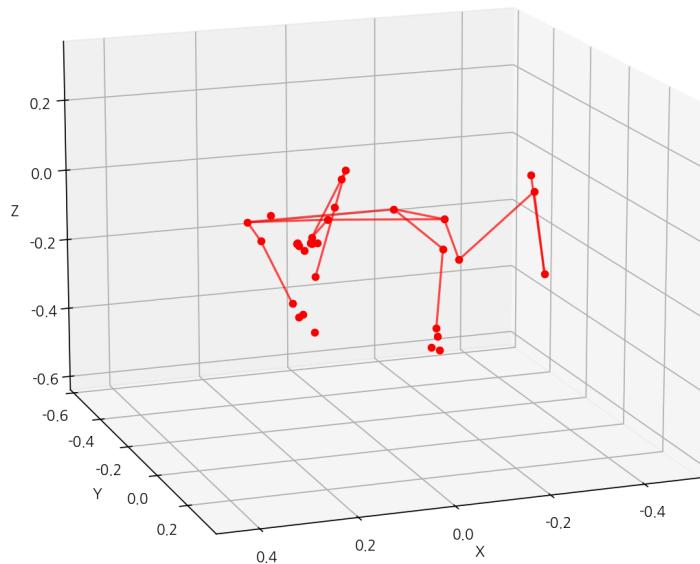


Figure 6: 3D 스켈레톤 오버레이 - 최소 손실 프레임. 두 포즈가 거의 완벽히 일치하여 겹쳐 보임.

5.3 다중 시점 비교 (Frame 40)

3D 스켈레톤 다중 시점 - 최대 손실 프레임 (Frame 40)

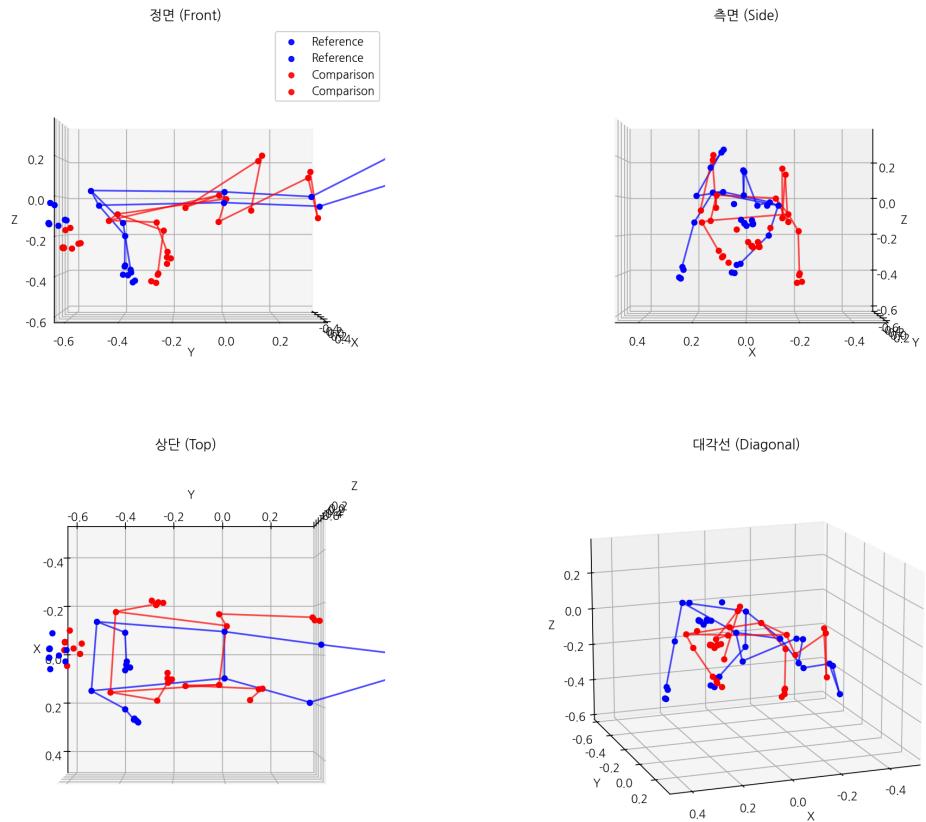


Figure 7: 최대 손실 프레임을 정면/측면/상단/대각선 4개 시점에서 관찰. 측면 뷰에서 깊이 차이가 가장 명확.

5.4 2D vs 3D 스켈레톤 직접 비교 (Frame 40)

2D vs 3D 스켈레톤 비교 - Frame 40

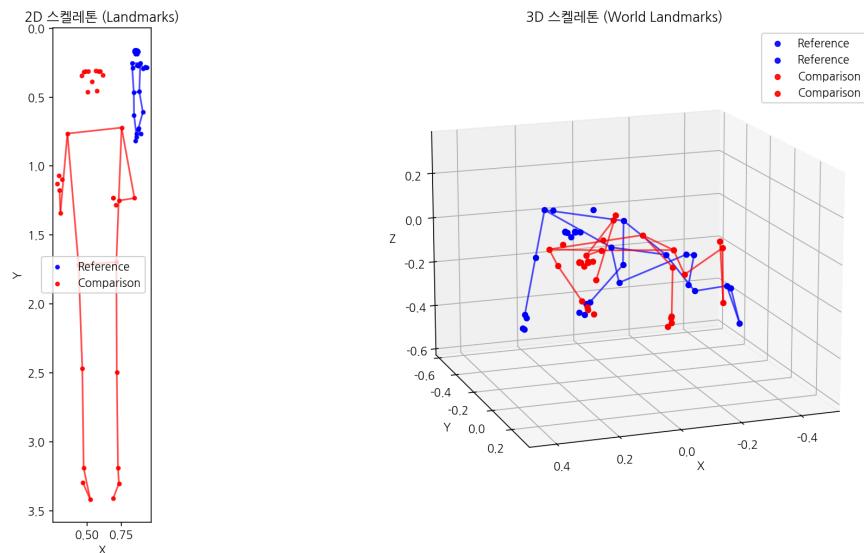


Figure 8: 같은 프레임의 2D(좌)와 3D(우) 스켈레톤 비교. 2D에서는 카메라 투영으로 인해 포즈가 유사해 보이지만, 3D에서는 실제 공간상의 차이를 정확히 보여줌.

6 관절별 고손실 순위 비교

순위	2D 관절	2D 손실	3D 관절	3D 손실
1	LEFT_FOOT_INDEX	10.68	LEFT_FOOT_INDEX	1.12
2	LEFT_HEEL	10.65	LEFT_HEEL	0.97
3	RIGHT_FOOT_INDEX	10.54	LEFT_ANKLE	0.92
4	LEFT_ANKLE	10.45	LEFT_KNEE	0.84
5	RIGHT_HEEL	10.41	RIGHT_FOOT_INDEX	0.77

관찰: 2D와 3D 모두 하체 관절(발, 발목, 무릎)이 최상위 손실을 보이지만, 3D에서의 절대값은 약 10배 감소. 이는 그랑 바뜨망 동작의 실제 차이를 더 정확하게 반영한다.

7 프레임별 손실 상세 분석 (3D)

7.1 프레임별 3D 손실 그래프

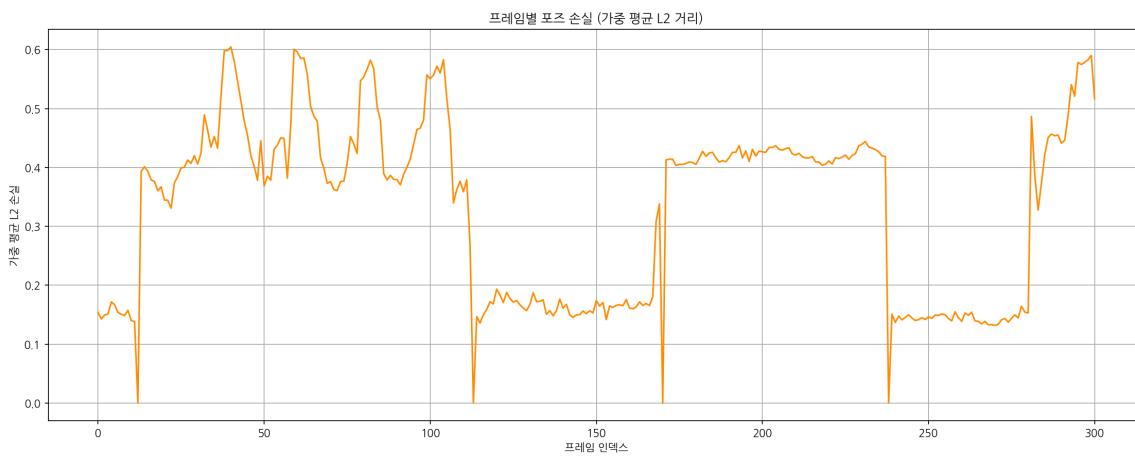


Figure 9: 3D world_landmarks 기반 프레임별 손실. 동작 전환 구간에서 손실 피크가 발생함.

7.2 고오차 구간 하이라이트

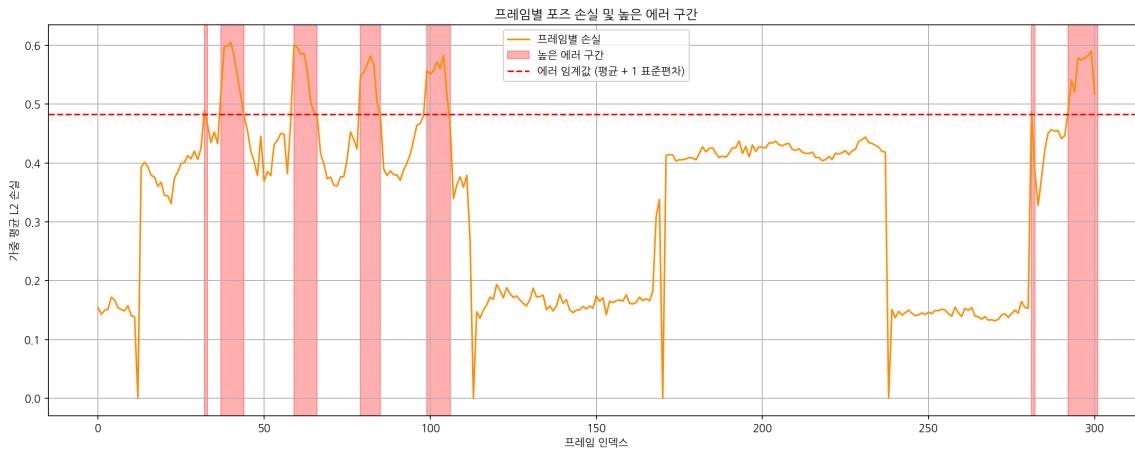


Figure 10: 평균 + 1 표준편차를 초과하는 고오차 구간(빨간 영역)이 자동 표시됨. 집중 교정이 필요한 구간을 시각적으로 식별 가능.

8 SMPL 메쉬 기반 3D 포즈 시각화

포인트+라인 기반 스켈레톤의 한계를 극복하기 위해, SMPL(Skinned Multi-Person Linear) 바디 모델을 활용하여 사실적인 인체 메쉬 시각화를 구현하였다.

8.1 구현 방법

- **SMPL 모델:** 6,890 vertices, 13,776 faces, 24 관절의 파라메트릭 인체 모델

- **GMM Pose Prior:** VIBE/SMPLify에서 검증된 8-Gaussian Mixture Model로 자연스러운 포즈만 생성
- **2단계 최적화:** (1) 글로벌 회전+이동 최적화 → (2) 전체 포즈+체형 최적화 (GMM prior 포함)
- **좌표 변환:** MediaPipe Y-down → SMPL Y-up (Y축 반전)
- **관절 매팅:** MediaPipe 33개 중 16개를 SMPL 22개 관절에 매팅

8.2 참조 vs 비교 포즈 오버레이 (Frame 40)

Frame 40: Reference (Blue) vs Comparison (Red)



Figure 11: SMPL 메쉬 오버레이 - 최대 손실 프레임. 참조(파란색)와 비교(빨간색)의 하체 포즈 차이가 사실적인 인체 형태로 명확히 드러남.

8.3 참조 포즈 다중 시점 (Frame 40)

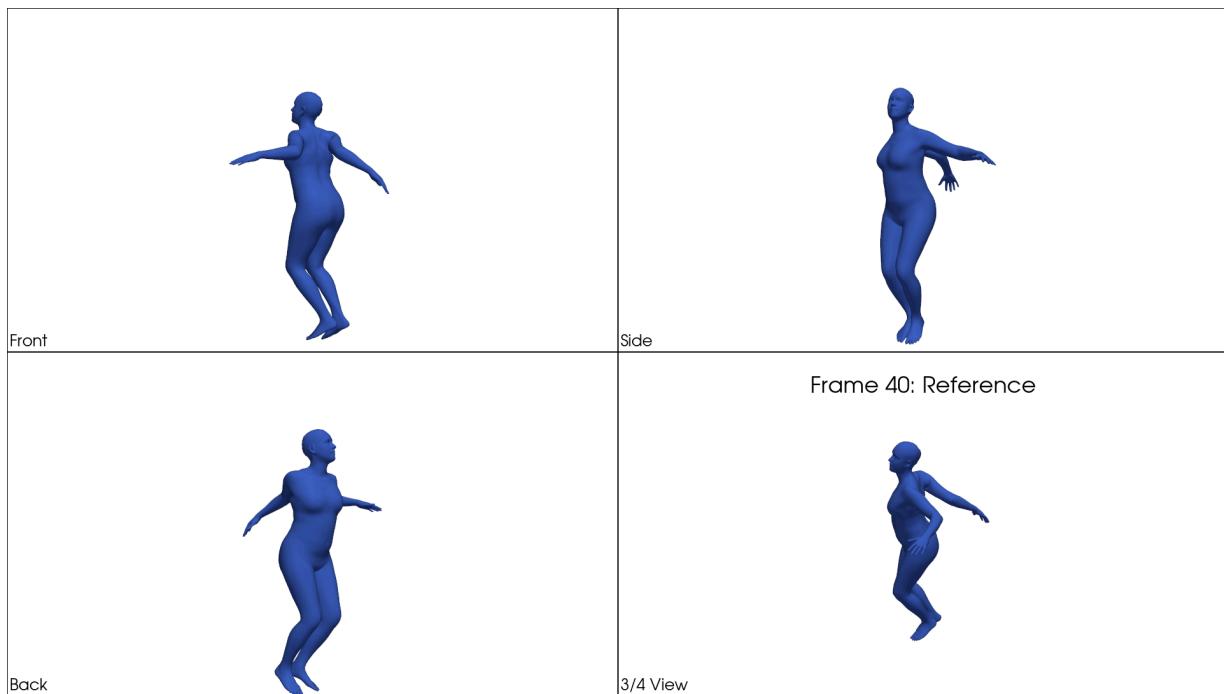


Figure 12: 참조 포즈를 정면/측면/후면/3-4 시점에서 관찰. Y축이 올바르게 정렬되어 서 있는 포즈로 표시됨.

8.4 비교 포즈 다중 시점 (Frame 40)

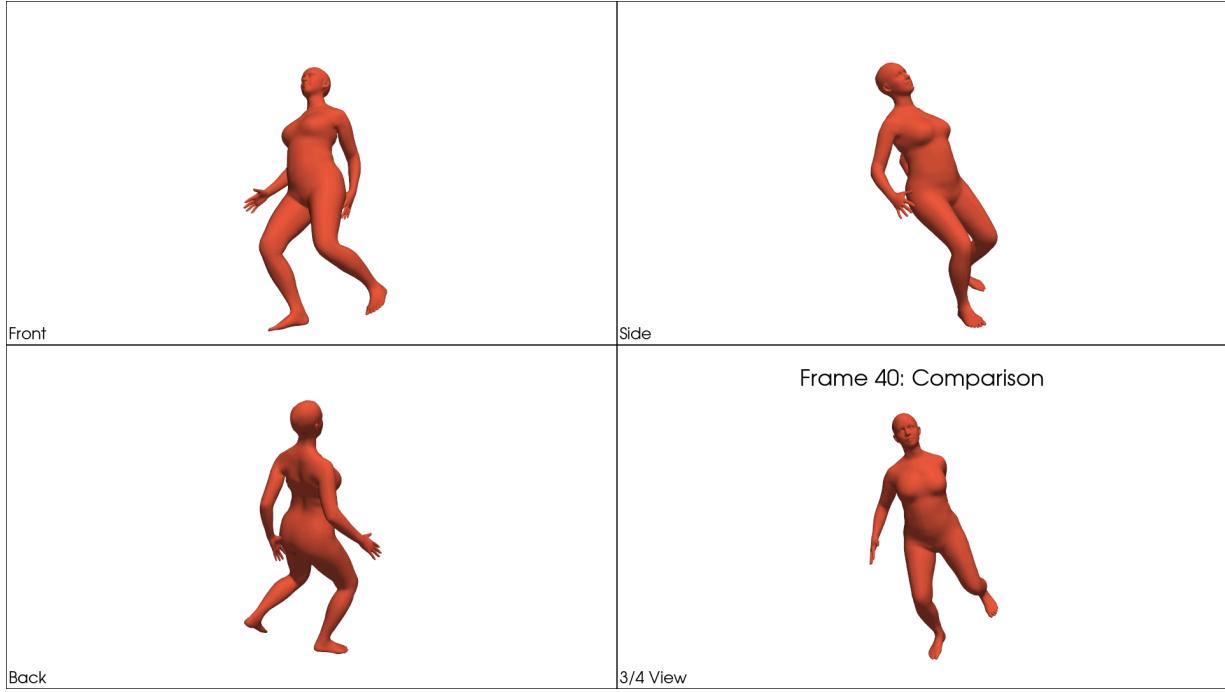


Figure 13: 비교 포즈의 4개 시점 뷰. 참조 대비 다리와 팔의 각도 차이가 시각적으로 확인 가능.

8.5 스켈레톤 → 메쉬 전환 효과

항목	포인트+라인 스켈레톤	SMPL 메쉬
형태 인식	추상적 점/선	사실적 인체
포즈 직관성	전문가만 해석 가능	누구나 포즈 차이 파악
깊이 정보	점의 크기로만 추정	메쉬 표면으로 명확
좌표 안정성	Y축 뒤집힘 문제	SMPL 표준 좌표계 사용

9 Gemini CLI 실행 기록

#	작업	재시도	결과
1	comparator.py 3D 모드 추가	0회	성공
2	2D vs 3D 통계 시각화 (4개 PNG)	0회	성공
3	Phase 2 테스트 작성 (6개)	0회	성공
4	3D 스켈레톤 시각화 (5개 PNG) + 프레임별 분석	0회	성공
5	SMPL 메쉬 시각화 (smpl_visualizer.py)	Gemini 3회 실패	[Claude 직접 수정]

SMPL 시각화 실패 원인 및 해결: Gemini의 자체 IK 구현은 Pose Prior 부재로 메쉬 왜곡 발생. ./sc:research로 joints2smpl 프로젝트를 발견하여 GMM Pose Prior + 2단계 최적화 방식으로 전환 후 성공.

10 결론 및 다음 단계

10.1 결론

1. 가설 검증 성공: 3D world_landmarks 비교가 카메라 앵글 영향을 대폭 감소시킴 (평균 손실 93.7% 감소)
2. 하위 호환성 유지: 기존 2D 파이프라인 완전 보존 (use_world=False 기본값)
3. 실용적 의미: 촬영 각도가 다른 영상 쌍에서도 포즈 차이를 의미 있게 비교 가능

4. **SMPL 메쉬 시각화**: GMM Pose Prior 기반 피팅으로 사실적인 3D 인체 메쉬 비교 달성

10.2 다음 단계

- **Phase 3**: 음악 BPM 동기화 (librosa 기반 비트 추출 + 비트별 손실 집계)
- **Phase 4**: DTW(Dynamic Time Warping)를 통한 자동 시간축 정렬
- **개선 사항**: SMPL 메쉬 애니메이션 (프레임 시퀀스 → GIF/영상 출력)