

## **HUMAN4D: A Human-Centric Multimodal Dataset for Motions and Immersive Media**

유럽(그리스) 메타버스 프로젝트 논문/ VRTogether H2020 프로젝트 일환

### **\*\*\*\*논문 흐름**

1. 우리가 이 데이터셋 만든이유
2. overview 기존 데이터셋들과의 차별 우리가 뭐 더했는지 기술
3. 데이터셋 디테일 묘사 데이터셋 통계적으로 정리함
4. benchmark 2d 3d pose 실험
5. 3d codec benchmark
6. discussion impact of dataset
7. 결론이랑 future work(앞으로 할 것들)

### **\*\*\*\*데이터셋 설명**

시공간과 RGBD 멀티뷰 가 일관된 정렬된 데이터셋 공개함

각각 2명의 남자 여자 Mocap전문배우 전신 움직임 촬영함

여러대의 RGBD촬영 기기와 카메라를 동기화함 (동시에 촬영)

- 2D pose estimation bench mark(성능 분석) 논문에서 sota(state of the art)라고 소개를 했지만 sota는 아님

Openpose bottom up / Alphapose top down

- 3D pose estimation bench mark

Learnable Triangulation

- 3D 압축 bench mark

Draco, Corto and CWIPC

### **\*\*\*\*기존 연구들**

- MHAD : 2013년도 버클리 데이터셋 모션캡처랑, RGBD제공함 spatio-temporally(시공간 일치하는)  
12RGB, 2Kinect v2 RGBD 6wearable inertial sensors 4MIC// 12명의 실험자에 의한 659 시퀀스 데

이더 11개의 인간 행동을 제공함. But HW-SYNCed 안됨

- Human3.6M : 3.6 백만개의 포즈 뭉치 포함 5여자 6남자. Human4D데이터셋과 유사함 일상생활 행동 데이터셋 포함됨. 4개의 rgb 동기화된 카메라와 1개의 depth sensor제공 computer vision분야에서 많이 쓰임. 한 사람에 대해 포커스 맞춤. But 컬러카메라만 동기화 되어있고 depth 해상도 낮음

- CMUPanoptic : 사람 view에서 가장 큰 데이터셋임 최대 8명의 상호작용도 캡처하였음 rgb(Kinect v2)카메라로 구성된 대규모 다중뷰. But human4d보다 동기화 정밀도 떨어짐, 마커 없음

- HUMBI

➔데이터셋 받아봐야함 승인 지연됨

Smpl 기반 3d데이터 mesh제공

- HUMAND4d

위의 부족한 부분들을 채우는 방향으로 제작됨

가림, 시야에서 안나오는 것들 그리고 여러대의 동기화된 depth값

24대의 모션캡처 카메라와 4대의 rgbd

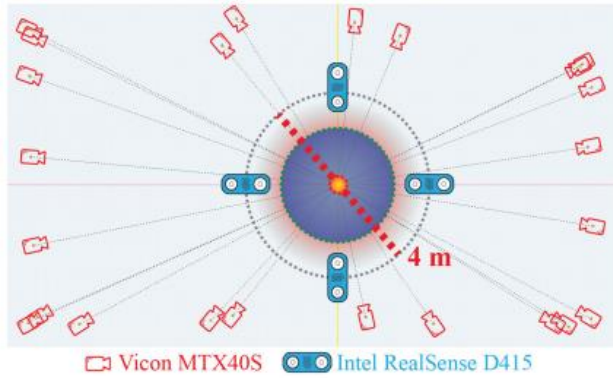
상세 자료는 appendix참조

**\*\*\*\*어떻게 실험 수행했는지 기술**

**\*\*\*\*Appendix**

용어 설명

HW-SYNCed :하드웨어 동기화



**FIGURE 2.** Capturing space floor-plan showing the poses of 24 Vicor MTX40S cameras and 4 Intel RealSense D415 sensors.

- 1인 14개, 2인 5개(총 19개)의 복합 데이터(매일 신체 운동 포함) 및 사회 활동, 각각 총 56개의 1인 및 10개의 2인 시퀀스.
- 앵커 기반 보정 방법을 사용하여 검색된 투영 행렬 및 외부 보정 카메라 매개변수는 쌍별 누적 오류를 줄여 다양한 카메라 보기에 대한 4D 데이터의 2D 투영을 가능하게 하고 그 반대로도 가능함.
- 배우가 특정 스크립트 및 시나리오에 따라 말하고 행동해야 하는 일부 활동에 대한 30개의 오디오 큐 타임스탬프 데이터를 제공하여 양식 간의 동기화.
- 전문 배우 중 한 명의 스캔 및 조작된 3D 모델 1개.

**TABLE 2.** Details with respect to HUMAN4D physical, daily and social activities.

	activity	# frames	audio	type
Single-person	running	2,050	✗	physical
	jumping_jack	1,974	✗	physical
	bending	2,156	✗	physical
	punching_n_kicking	2,079	✗	physical
	basketball_dribbling	2,124	✗	physical
	laying_down	4,082	✗	physical
	sitting_down	3,288	✗	daily
	sitting_on_a_chair	2,797	✗	daily
	talking	2,377	✓	daily
	object_dropping_n_picking	1,768	✗	daily
	stretching_n_talking	2,787	✓	physical
	talking_n_walking	2,889	✓	daily
	watching_scary_movie	2,194	✓	daily
	in-flight_safety_announcement	6,192	✓	daily
Multi-person	watching_football_together	1,760	✓	social
	dancing_together	1,356	✓	social
	physical_examination	2,328	✓	social
	whispering	3,045	✓	social
	card_trick	3,060	✓	social
		50,306		

**TABLE 1.** Summary of state-of-the-art datasets and HUMAN4D with respect to the available features and modalities.

	MHAD <sub>(2013)</sub> [37]	Human3.6M <sub>(2014)</sub> [38]	CMUPanoptic <sub>(2015)</sub> [39]	HUMBI <sub>(2018)</sub> [40]	HUMAN4D <sub>(2020)</sub>
<i>Body Pose</i>	✓	✓	✓	✓	✓
<i>Marker-based MoCap</i>	✓	✓	✗	✗	✓
<i>Body Part Segments</i>	✗	✗	✗	✓	✗
<i>Multi-view RGB</i>	✓	✓	✓	✓	✓
<i>Multi-view Depth</i>	✓	✗	✓	✗	✓
<i>3D Meshes</i>	✗	✗	✗	✓	✓
<i>Point-clouds</i>	✗	✗	✓	✗	✓
<i>Audio Cues</i>	✓	✗	✗	✗	✓
<i>Gaze Features</i>	✗	✗	✗	✓	✗
<i>Hand Features</i>	✗	✗	✓	✓	✗
<i>Facial Features</i>	✗	✗	✓	✓	✗
<i>Rigged Characters</i>	✗	✓	✗	✗	✓
<i>Multi-person</i>	✗	✗	✓	✗	✓