CVPR Nashville JUNE 11-15, 2025

Session Order: 159, ExHall D Poster Session3, June 14

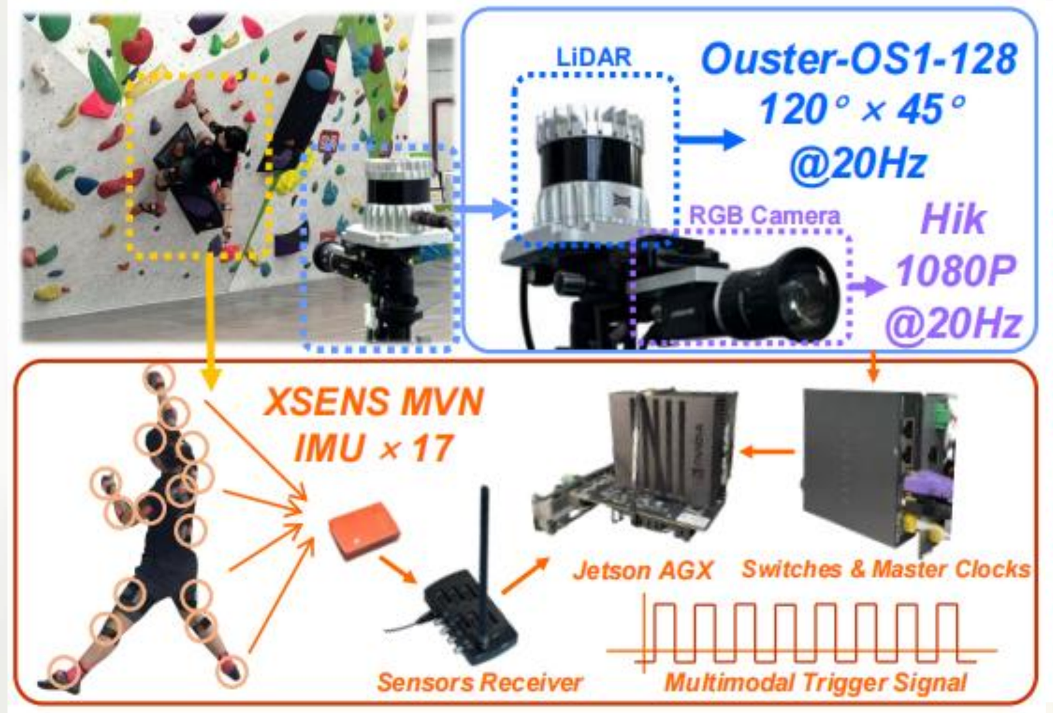
Music City Center, Nashville, TENNESSEE USA

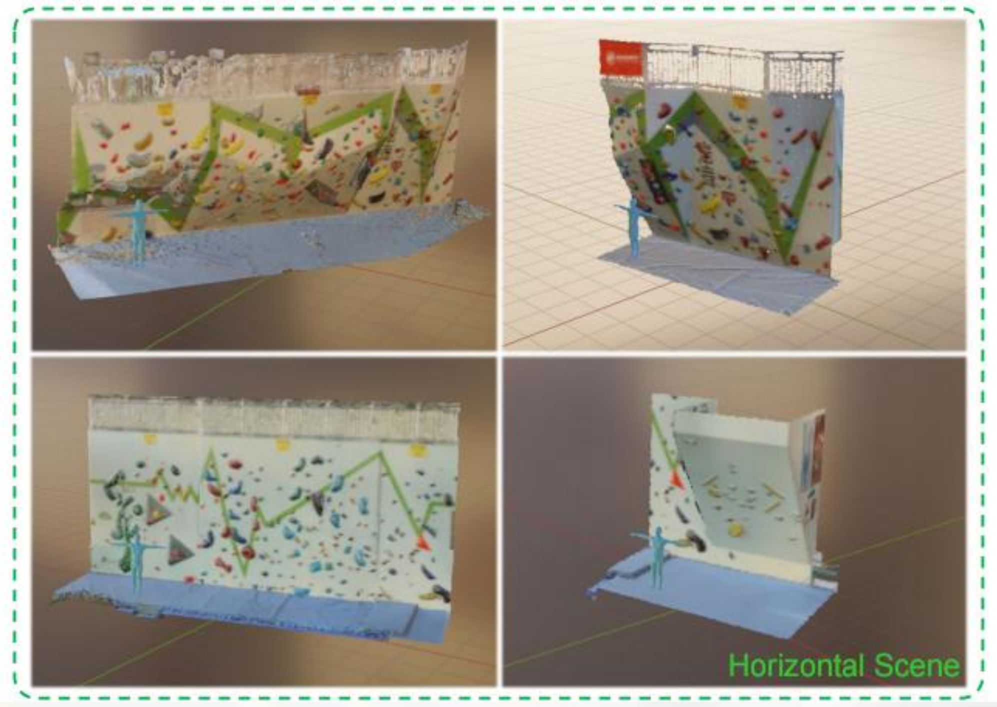
ClimbingCap: Multi-Modal Dataset and Method for Rock Climbing in World Coordinate

Ming Yan*[1] Xincheng Lin*[1] Yuhua Luo[1] Shuqi Fan[1] Yudi Dai[4] Qinxin Zhong[2] Lincai Zhong[2] Yuexin Ma[3] Lan Xu[3] Chenglu Wen[1] Siqi Shen†[1] Cheng Wang[1]

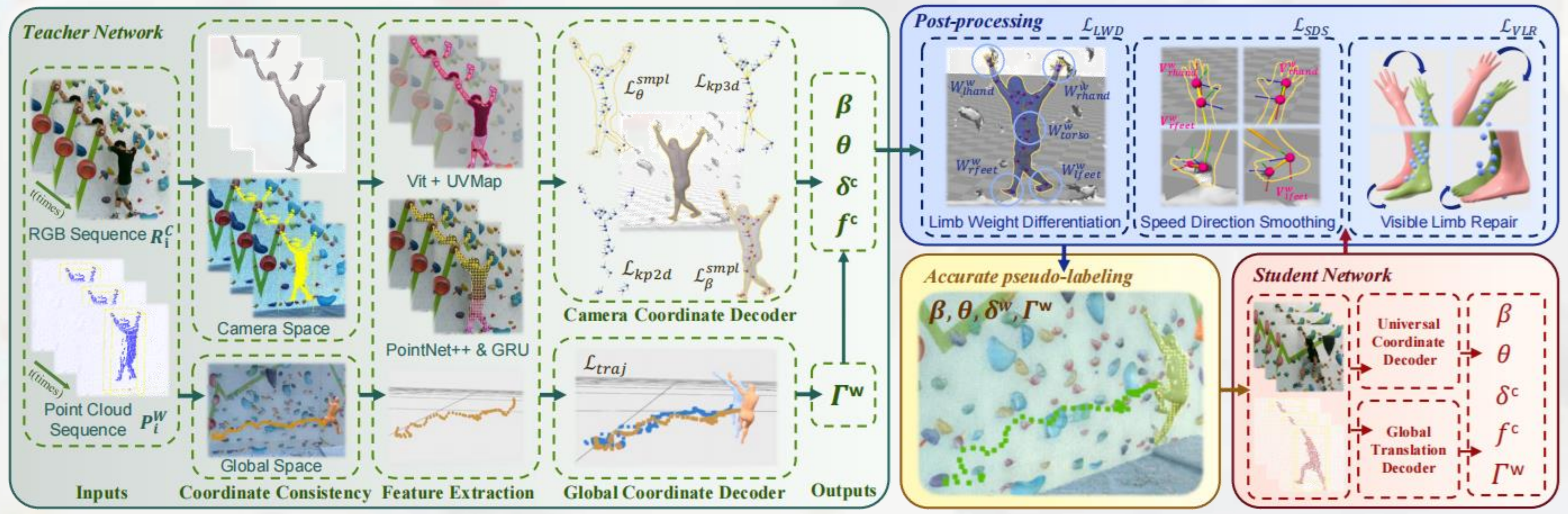
[1]Xiamen University [2]National Climbing Team of China [3]ShanghaiTech University [4]ETH Zürich

Global Human Motion Recovery (HMR) struggles with complex human poses and dynamic interactions, especially in climbing—an off-ground sport requiring hand-foot coordination to ascend walls. As an Olympic event, climbing lacks sufficient research due to limited datasets (e.g., SPEED21: 2D, CIMI4D: small-scale 3D with trivial motions). Capturing climbing motion is challenging, needing precise pose estimation and global coordinate consistency, yet existing HMR methods focus on ground-based motions and fail to address climbing's unique demands.





We introduce AscendMotion, a pioneering multi-modal dataset designed to advance Human Mesh Recovery (HMR) research for climbing motion in global coordinate systems. Tailored to capture the intricacies of rock climbing, AscendMotion features complex, multi-directional motions on non-planar surfaces, offering a unique and challenging resource for the HMR community. The dataset includes 344 minutes (6 hours) of meticulously labeled data and an additional 441 minutes of unlabeled data, totaling over 12 hours of recordings. It encompasses 22 skilled climbers navigating 12 diverse real-world climbing scenes, showcasing a variety of styles such as bouldering, speed climbing, and lead climbing.

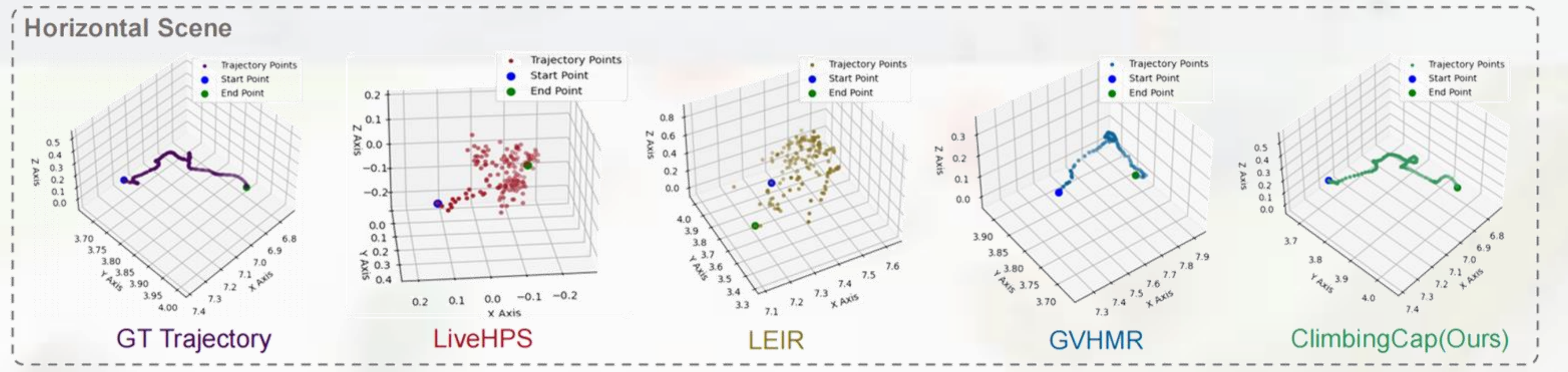


Modality	Method	Camera Coordinate					World Coordinate				
		ACCEL↓	MPJPE↓	PA-MPJPE↓	PVE↓	PCK0.3↑	WA-MPJPE↓	W-MPJPE↓	RTE↓	Jitter↓	T-Error↓
RGB	TRACE [60]	18.68/76.79	875.56/577.60	69.21/85.81	951.52/619.93	0.06/0.09	144.33/**385.71**	254.38/703.35	14.73/26.17	115.96/521.40	2.56/6.62
	SLAHMR [75]	5.46/96.98	232.46/467.88	84.13/285.15	283.24/552.63	0.36/0.12	277.47/447.68	804.85/**613.60**	**3.64**/39.39	4.91/201.33	2.81/6.54
	WHAM [56]	4.59/35.01	110.92/143.17	76.09/**73.36**	124.2/164.91	0.76/0.62	229.42/1125.77	647.70/1499.85	5.16/9.04	**3.58**/40.69	1.77/**2.49**
	GVHMR [55]	**4.50/26.22**	**107.09/124.60**	**60.06**/80.30	**118.89/151.10**	**0.77/0.71**	**105.15**/1002.11	**202.45**/1442.50	4.09/**7.91**	6.85/**32.71**	**1.48**/2.54
LiDAR	LiDARCapV2 [78]	87.99/119.62	244.6/234.52	192.17/156.39	326.45/283.27	0.53/0.50	282.12/1396.42	442.12/1518.29	16.42/10.85	176.95/165.55	1.65/2.89
	LiveHPS [49]	24.06/71.03	138.71/153.33	114.26/132.87	170.84/203.58	0.65/0.68	158.63/201.31	344.54/316.53	13.96/10.37	79.75/155.97	1.60/2.14
LiDAR +RGB	ImmFusion [11]	108.76/74.20	473.18/464.83	254.07/179.51	533.5/529.71	0.17/0.14	324.4/1446.01	487.92/1532.88	16.52/10.86	27.03/**14.49**	1.94/3.99
	FusionPose [14]	112.08/86.44	256.81/315.93	198.55/193.83	306.22/359.68	0.36/0.42	275.02/1445.32	444.47/1532.28	16.29/10.85	92.97/80.79	2.02/6.48
	LEIR [73]	110.18/94.57	297.95/299.62	187.26/150.56	340.61/351.52	0.41/ 0.37	266.82/1313.09	282.31/1435.92	9.78/9.97	73.38/85.03	1.1/**1.20**
	Ours	**5.17/17.25**	**75.45/88.92**	**61.73/74.50**	**94.89/106.42**	**0.91/0.78**	**62.95/85.26**	**78.99/106.95**	**1.57/3.12**	**8.3**/27.75	**1.07**/1.29

We propose ClimbingCap, an innovative Human Mesh Recovery pipeline designed to tackle the unique challenges of capturing climbing motion, which involves extreme limb extensions, full-body exertion, and precise alignment with rock walls in both camera and global coordinate systems. ClimbingCap consists of three core stages: Separate Coordinate Decoding, Post-Processing, and Semi-Supervised Training. In the Separate Coordinate Decoding stage, features from RGB images and LiDAR point clouds, transformed via an extrinsic matrix, are extracted using ViT and PointNet++ to predict SMPL parameters in camera coordinates and global translation parameters in world coordinates. Iterative decoding refines these estimates, guided by losses. The Post-Processing stage refines these outputs using three losses, optimized with the Adam optimizer for global consistency and leveraging LiDAR's 3D data for rigid transformations. Finally, the semi-supervised training stage utilizes AscendMotion's abundant unlabeled data for training. The teacher model generates pseudo labels from unlabeled sequences to train the student model. This cohesive pipeline effectively addresses climbing's off-ground dynamics and intricate human-scene interactions.

Affiliation: Fujian Key Laboratory of Sensing and Computing for Smart Cities, Xiamen University, China.
National Institute for Data Science in Health and Medicine, Xiamen University, China.
Key Laboratory of Multimedia Trusted Perception and Efficient Computing, Xiamen University, China.
Shanghai Engineering Research Center of Intelligent Vision and Imaging, ShanghaiTech University, China
ETH AI Center, ETH Zürich, Switzerland

Project Page: http://www.lidarhumanmotion.net/climbingcap/



We validate the AscendMotion dataset and ClimbingCap through comprehensive experiments, demonstrating the dataset's high-quality annotations and its challenge to existing HMR methods, while showcasing ClimbingCap's superior performance. The annotation quality of AscendMotion is assessed by comparing global optimization outputs with manual annotations across horizontal and vertical rock wall scenes, using metrics like MPJPE, PA-MPJPE, PVE, and acceleration error; confirming the dataset's robustness. ClimbingCap is evaluated against nine state-of-the-art HMR methods—RGB-based, LiDAR-based, and LiDAR+RGB-based—on AscendMotion and CIMI4D datasets, following protocols from WHAM and SLOPER4D for metrics like WA-MPJPE, W-MPJPE, RTE, Jitter, and T-Error. In AscendMotion, ClimbingCap excels in horizontal scenes and significantly outperforms others in vertical scenes, where methods like GVHMR falter due to poor upward motion estimation, highlighting the need for coordinated camera-global alignment. On CIMI4D's less challenging horizontal scenes, ClimbingCap generalizes well, leading in most metrics except WA-MPJPE.