基于运动相机阵列系统的动作捕捉与模型重建

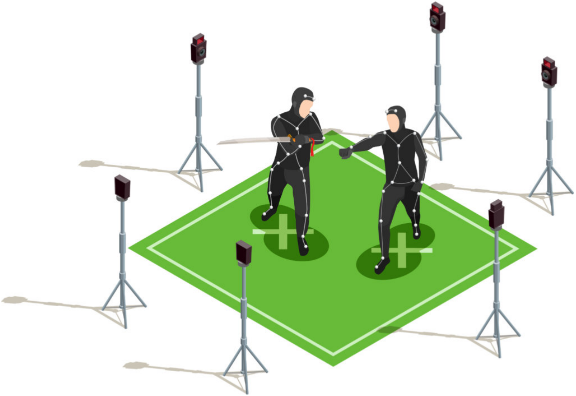
答辩人: 赵志翔

指导老师: 郝祁

南方科技大学

研究背景

动作捕捉 (Motion Capture) 是利用外部设备获取并记录人体结构的位移从而进行姿态还原的技术。



- 电影制作
- 虚拟现实
- 游戏开发
- 人体工程学研究
- 模拟训练
- 医疗保健

图: 来源: https://skywell.software/blog/fast-growing-motion-capture-industry-predictions-and-analysis/

相关工作

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

相关工作

- 机械式:[1]
- 电磁式:[2]
- 惯性导航式:[3]、[4]
- 光标式:[5]
- 无标志式:[6]、[7]、[8]、[9]

设计需求

1. 多场景
2. 成本
3. 侵入性
4. 实时性
5. 准确率

无人机动捕系统

图: 户外基于多无人机的无标记式动捕系统[9]

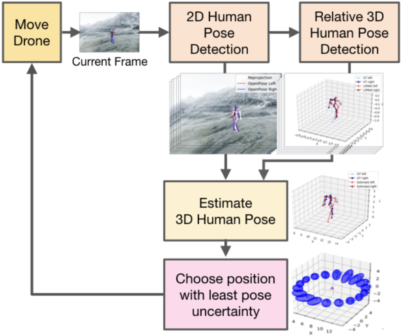


图: ActiveMoCap: 主动式优化视角点选择的单无人机动捕系统[10]

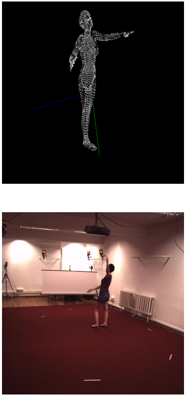
系统对比

动作捕捉系统对比

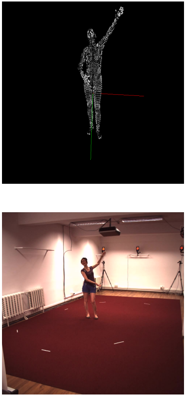
方法	场景	成本	侵入性	实时性	准确率
[4]	多场景	高	$\surd$	30fps	$\approx 3cm$
[7]	多场景	低	$\times$	no info	$<$10cm
[8]	多场景	低	$\times$	no info	5cm
[5]	多场景	中	$\surd$	no info	$\approx 10cm$
[9]	室外	高	$\times$	offline	$\approx 10cm$
本文方案	多场景	低	$\times$	1.5fps	$\approx 5cm$

表: 各类动作捕捉系统对比

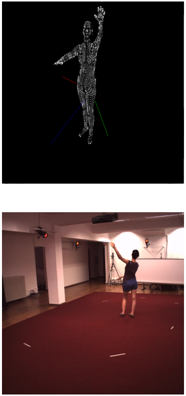
视角点选择问题



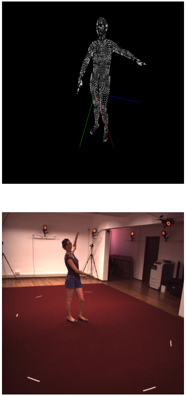
(e) 机位 1



(f) 机位 2



(g) 机位 3



(h) 机位 4

图: 在不同机位对同一姿势的三维重建比较 (Human3.6M 数据集 S1-pose2): 重点观察人物右手，被摄者右手收紧在腹部，只有机位 2 能够正确识别，而其他机位识别的右手手臂都为向外张开。

视角点选择问题

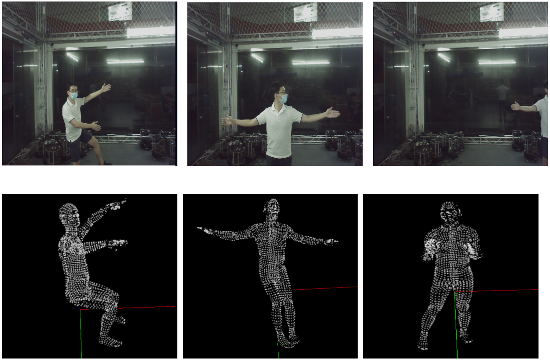


图: 画面仅捕捉到的部分人体骨架，HMR 模型[11]。

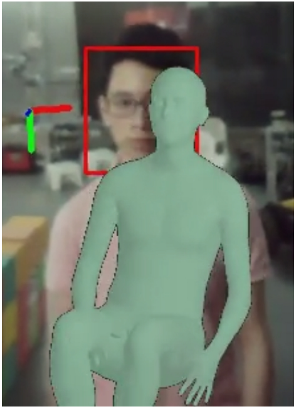
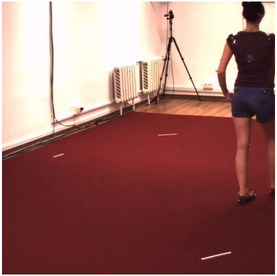


图: VIBE 模型[12]

本项目针对的场景与问题定义

(a) 第一类视角点

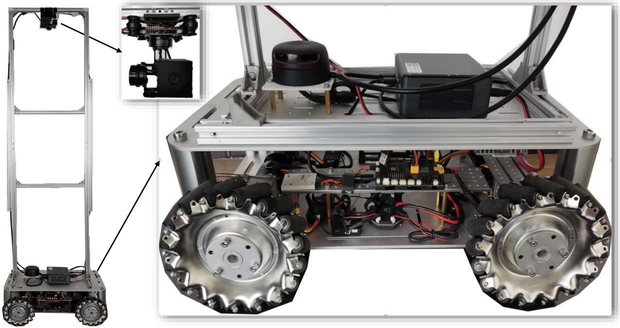


(b) 第二类视角点

图: 第一类视角点：画面中人体的骨架完整。第二类视角点：相机不能捕捉到完整的人体骨架信息

使用场景：足够宽的平整场地，在场地上只有运动相机机器人和被摄者，例如动捕工作室场景，室内的大型运动场地，或者是室外宽阔的平整地面。

移动相机机器人



系统模块

南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

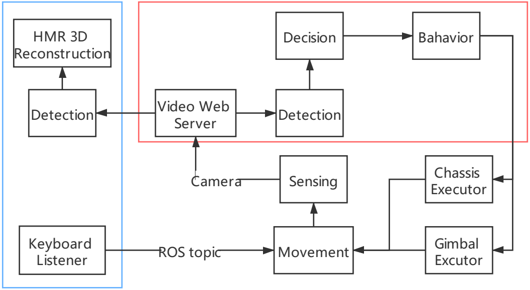
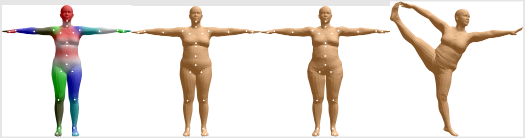


图: 系统运行流程：蓝色框为数据处理模块，橙红框为运动控制模块。

- 目标检测：HOG + SVM
- 决策：行为树
- 模型推理：HMR

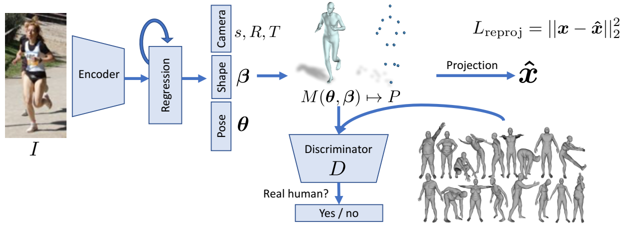
南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

SMPL 模型



(a) $\bar{\mathbf{T}}, \mathcal{W}$ (b) $\bar{\mathbf{T}} + B_S(\vec{\beta}), J(\vec{\beta})$ (c) $T_P(\vec{\beta},\vec{\theta}) = \bar{\mathbf{T}} + B_S(\vec{\beta}) + B_P(\vec{\theta})$ (d) $W(T_P(\vec{\beta},\vec{\theta}), J(\vec{\beta}), \vec{\theta}, \mathcal{W})$

HMR 算法



南方科技大学 SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

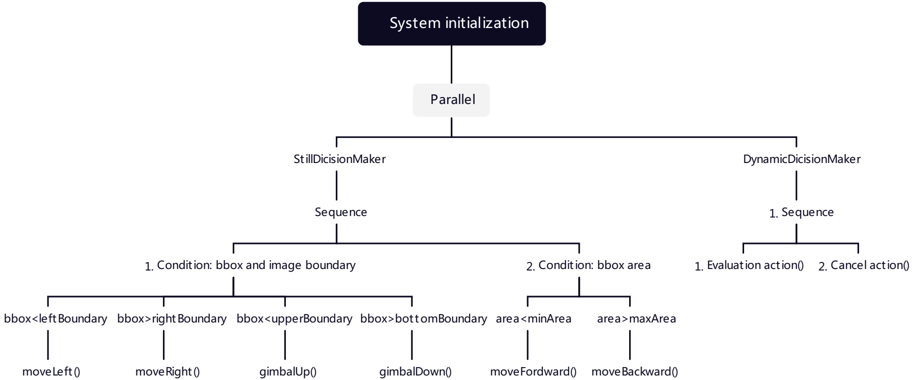
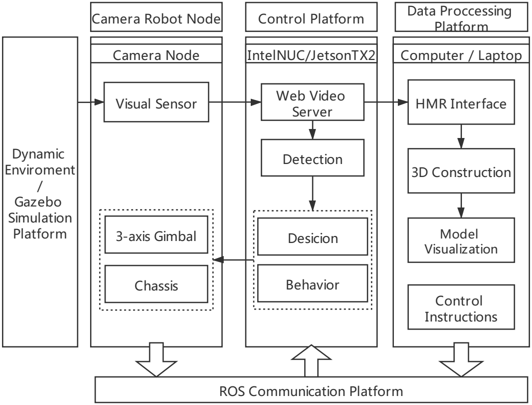
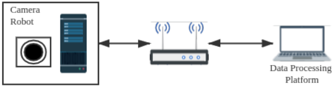


图: 相机的行为树：定义了相机运动相关的行为 (actions) 和相对应的触发条件。

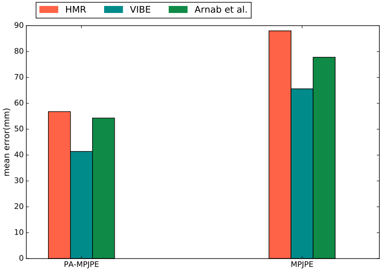
实验系统搭建





- 基于 ROS 的通讯框架
- 控制与计算模块分离
- 物联网

算法精确度分析



方法	PA-MPJPE	MPJPE
HMR	56.8	88
VIBE	41.4	77.8
Arnab et al.	54.3	65.6

表: Human3.6M 基准比较: MPJPE 和 PA-MPJPE, 单位为毫米 (Protocal 2)

算法实时性分析

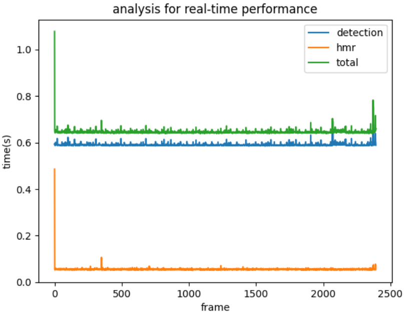


图: 实时性测试数据集:
human3.6m-S1-walking1

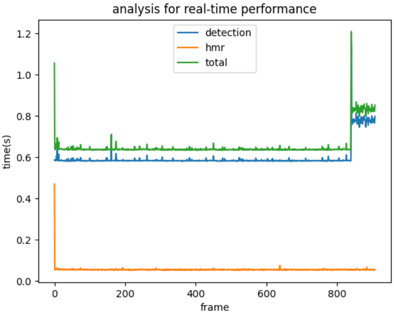


图: 实时性测试数据集:
human3.6m-S1-walkingDog1

现实场景试验（视频）

图: 相机机器人视角

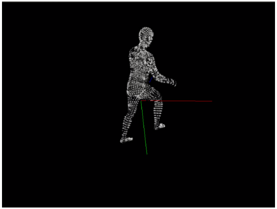


图: 模型重建

图: 旁观者视角

GU X, ZHANG Y, SUN W, et al. Dexmo: An Inexpensive and Lightweight Mechanical Exoskeleton for Motion Capture and Force Feedback in VR[C/OL]//CHI '16: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. San Jose, California, USA: Association for Computing Machinery, 2016: 1991-1995. https://doi.org/10.1145/2858036.2858487. DOI: 10.1145/2858036.2858487.

YABUKAMI S, KIKUCHI H, YAMAGUCHI M, et al. Motion capture system of magnetic markers using three-axial magnetic field sensor[J]. IEEE Transactions on Magnetics, 2000, 36(5): 3646-3648. DOI: 10.1109/20.908928.

ANDREWS S, HUERTA I, KOMURA T, et al. Real-Time Physics-Based Motion Capture with Sparse Sensors[C/OL]//CVMP 2016: Proceedings of the 13th European Conference on Visual Media Production (CVMP 2016). London, United Kingdom: Association for Computing Machinery, 2016. https://doi.org/10.1145/2998559.2998564. DOI: 10.1145/2998559.2998564.

MALLESON C, GILBERT A, TRUMBLE M, et al. Real-Time Full-Body Motion Capture from Video and IMUs[C]//2017 International Conference on 3D Vision (3DV). [S.l. : s.n.], 2017: 449-457. DOI: 10.1109/3DV.2017.00058.

BILESAN A, OWLIA M, BEHZADIPOUR S, et al. Marker-based motion tracking using Microsoft Kinect[J]. IFAC-PapersOnLine, 2018, 51: 399-404. DOI: 10.1016/j.ifacol.2018.11.575.

参考文献 III

- WEI S, RAMAKRISHNA V, KANADE T, et al. Convolutional Pose Machines[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [S.l. : s.n.], 2016: 4724-4732. DOI: 10.1109/CVPR.2016.511.
- DONG J, SHUAI Q, ZHANG Y, et al. Motion Capture from Internet Videos[C]//VEDALDI A, BISCHOF H, BROX T, et al. Computer Vision – ECCV 2020. Cham: Springer International Publishing, 2020: 210-227.
- MARTINEZ J, HOSSAIN R, ROMERO J, et al. A simple yet effective baseline for 3d human pose estimation[J/OL]. CoRR, 2017, abs/1705.03098. arXiv: 1705.03098. http://arxiv.org/abs/1705.03098.

SAINI N, PRICE E, TALLAMRAJU R, et al. Markerless Outdoor Human Motion Capture Using Multiple Autonomous Micro Aerial Vehicles[C]//Proceedings 2019 IEEE/CVF International Conference on Computer Vision (ICCV). [S.l.]: IEEE, 2019: 823-832.

KICIROGLU S, RHODIN H, SINHA S, et al. ActiveMoCap: Optimized Viewpoint Selection for Active Human Motion Capture[C]//CVPR. [S.l. : s.n.], 2020.

KANAZAWA A, BLACK M J, JACOBS D W, et al. End-to-end Recovery of Human Shape and Pose[C]//Computer Vision and Pattern Recognition (CVPR). [S.l. : s.n.], 2018.

参考文献 V

KOCABAS M, ATHANASIOU N, BLACK M J. VIBE: Video Inference for Human Body Pose and Shape Estimation[C]//The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). [S.l. : s.n.], 2020.