

# 基于机器视觉的公共空间微观行为计算方法

李特<sup>1</sup>, 金衍孜<sup>2</sup>, 姚佳欣<sup>3</sup>, 闫超<sup>4</sup>

**摘要:** 社交行为是公共空间研究的核心内容, 但是其量化表征一直是难点。随着机器视觉技术的发展, 对空间行为的自动化追踪和量化采集已得以实现, 然而, 尚缺乏对于微观行为数据的计算分析及量化描述方法。本研究提出一种基于位姿估计算法的机器视觉行为计算方法, 探索公共空间场景中微观行为的量化分析 workflow, 展示基于视频监控的微观行为计算工具的可行性。本研究针对微观步态计算为例, 一方面通过行为序列重建和实测实验, 获取最佳平均拍摄角度、距离, 提升数据采集和行为解析准确度; 另一方面, 通过开发 Grasshopper 模块进行行为序列计算, 基于 17 关节点数据对微观行为的步态指标进行可视化分析, 并使用动作捕捉实验验证机器视觉行为重建工作方法的准确性。最后, 本研究以社交距离实验为例, 初步佐证了该方法可以提升复杂公共空间中的微观行为分析精度, 进而可以有效表征社交距离等公共空间指标。

**关键词:** 行为分析; 机器视觉; 人体位姿估计; 行为计算; 微观行为

**Keywords:** Behavioural Analysis; Machine Vision; Human Pose Estimation; Behavioural Calculations; Micro-Behaviour

**资助项目情况:** 中央高校基本科研业务费专项资金资助 22120240290

## 引言

社交行为研究是行为性能化研究的一个重要方向, 表征了特定建成环境中空间使用者之间的交互关系, 进而推及日常体验。常见的社交行为研究可根据空间评估和优化的目的不同分为两类: 一类社交行为研究以空间的品质提升和活力促进为目的, 另一类以空间效率提升和风险控制为目的。前者关注的社交行为包括交往行为<sup>[1]</sup>、游憩行为<sup>[2-3]</sup>、消费行为<sup>[4]</sup>等; 后者关注人流交汇和避让行为<sup>[5]</sup>、密接行为<sup>[6]</sup>等。

目前的研究方法已经提高了传统行为采集数据中使用的人工观察统计和问卷访谈调研方法的信息采集效率, 使用 WiFi 信号、手机信令、GPS 等位置信息追踪人的时空轨迹, 扩展了空间中人员流动规律的研究规模, 并减少了观测过程中对研究对象的干扰。<sup>[7-8]</sup>然而上述方法仅能获取人的位置信息, 提高了信息获取的效率和广度, 但无法取得更深层次的微观行为信息, 精细化行为研究存在量化表征困难。依托以计算机视觉 (Computer Vision) 为主导的观测和计算技术进步, 自动化微观行为计算研究成为可能。

<sup>1</sup> 李特, 同济大学建筑与城市规划学院, 硕士研究生

<sup>2</sup> 金衍孜, 同济大学建筑与城市规划学院, 硕士研究生

<sup>3</sup> 姚佳欣, 同济大学建筑与城市规划学院, 硕士研究生

<sup>4</sup> 闫超(✉), 同济大学建筑与城市规划学院, 助理教授, yanchao@tongji.edu.cn

# 1 基于机器视觉的社交行为研究

计算机视觉方法可以自动化处理视频记录，如目标检测（Object Detection）和目标追踪（Object Tracking）算法可以识别行人和计数。<sup>[9]</sup>类似于前文提到的位置信息分析方法，基于目标检测和跟踪的方法使得人体行为数据得以大规模、无接触、自动化地记录分析，避免霍桑效应（Hawthorne Effect）造成的实验误差，并极大地提高计算效率。此外，基于视频的方法计算过程更加可视化；响应速度较快，可以做到人流数据实时监测。<sup>[10]</sup>该方法可以作为进一步获取微观行为数据的基础。

另外一类研究使用深度相机（RGB-D）采集行为数据，相关研究将深度相机穿戴在实验者身上，采集个体行为活动和人际交互信息。<sup>[11-12]</sup>该方法可以解析多元行为信息，获取空间深度信息，结合深度相机开发工具中常见的人体位姿估计（Human Pose Estimation）算法，可以获取交互对象的具体行为。随着深度学习技术的发展，基于普通单目相机的人体位姿估计算法的速度和准确度也不断提高，并可以估算的关键点的三维坐标。

计算机视觉的技术进步使得微观行为的数据采集能力已足够精细，然而目前从数据中计算分析得到有意义的信息的方法尚不成熟。量化和解析行为采集数据中蕴含的信息需要关注下列问题：微观行为研究关注哪些行为指标？在理论和方法上如何表征和计算相应的行为指标？如何表达行为指标与空间（和时间）的耦合关系？本文将构建一套微观行为计算 workflow，并回应上述问题。

## 2 基于机器视觉的微观行为计算

### 2.1 工作框架

在目标检测和追踪的基础上，结合人体位姿估计算法获得人体关键点坐标，是提高行为采集信息分辨率，进而开展微观行为分析与计算的关键基础。例如，相关研究使用监控摄像机结合 Openpose 二维人体位姿估计算法研究建筑中人的时空行为分布，该研究中主要采集的信息仍然是室内空间中人的位置信息，以与人相邻的 POI 的类别区分人的行为<sup>[13]</sup>。然而，当前此类研究没有充分挖掘人体位姿估计方法收集到的人体关键点信息，即没有深入到行为的微观层面。

本文提出基于机器视觉（单目相机）的微观行为计算工作框架分为两个部分。第一部分是行为采集和预处理：在目标检测和目标追踪的基础上使用三维人体位姿估计（3D Human Pose Estimation）算法得到人体关键点的空间位置信息，使得大规模行为信息采集的精细化程度提高；第二部分是行为指标计算：基于关节点三维坐标和人体模型重建，解析计算有意义的微观行为指标，得到深层次的微观行为信息。

行为采集和预处理部分通过组合若干计算机视觉算法实现从单目相机图像输入到人体三维关键点坐标输出的技术路径。具体而言，首先通过 YOLOX 目标检测算法和 ByteTrack 目标追踪算法实现视频序列中的行人识别，并保持连续视频帧中行人检测的一致性。<sup>[14-15]</sup>在此基础上使用 HRNet 获取二维人体关键点位置。<sup>[16]</sup>然后使用 GAST-Net 回归生成 3D 人体关键点位置。<sup>[17]</sup>

行为指标计算部分基于 Rhino 8 Grasshopper 平台将关节点三维坐标重建成空间点，并按照一定次序相互连接，得到可视化的人体微观时空行为模型。在此基础上，根据研究目的，确定行为研究指标。利用 Grasshopper 平台及其内置的代码编辑器进行数据处理，得到相应指标的量化

表征，并在重建模型的基础上数据可视化。下文以步态分析为例说明行为指标计算的实现方法，以近身空间（PPS）研究实验为例验证机器视觉微观行为计算方法的可行性。

## 2.2 微观行为数据采集与校准

本研究采用 17 个关键点表征人体姿态（图 1）。为了使结果可视化，便于研究人体时空行为与空间的交互关系，进而指导空间设计，使用 Grasshopper 平台进行人体时空姿态重建。在 Grasshopper 中输入记录每帧视频画面解析出的关键点坐标的逗号分隔值（csv）文件；通过数据清洗和处理，整理出 17 个关节点的三维坐标；将坐标组装成点；按照图 1 的顺序连接点，形成人体姿态的骨架模型。如同使用连续曝光照片研究行为的早期实验那样，本方法得到了连续的人体姿态三维模型。为了方便在稠密的模型中看清人体姿态，可以设置抽帧显示，也可以进一步将每帧模型间拉开一定的距离。

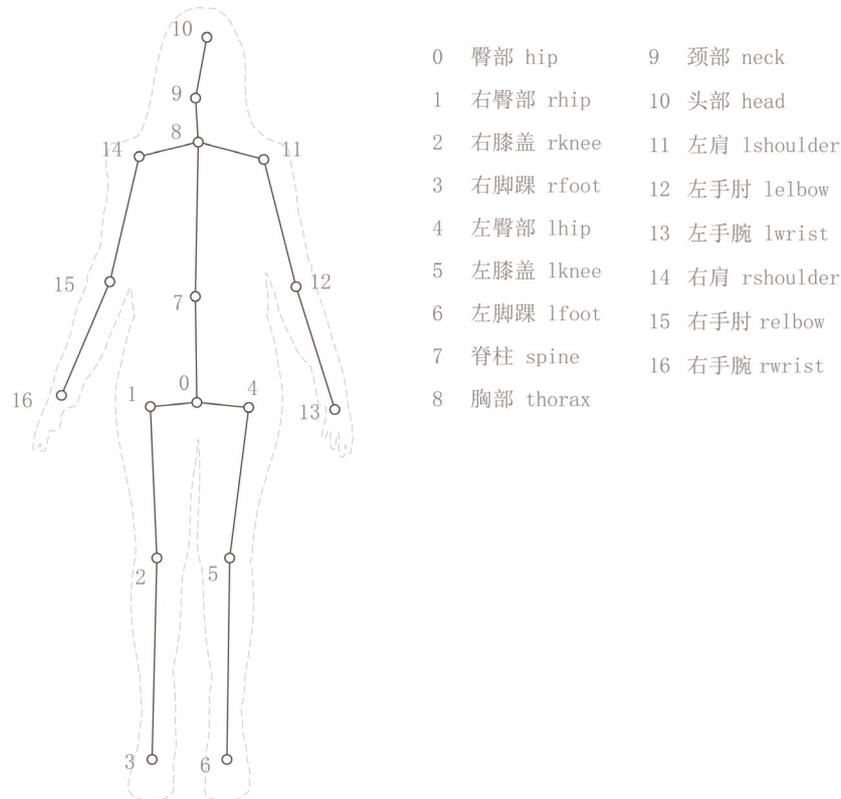


图 1 本研究采用的人体骨骼模型关节编号和连接方式（图片来源：作者自绘）

通过预实验发现，相机拍摄方式影响人体关键点信息采集的准确度。单目相机的视频估计方法具有相当大的灵活性，对于相机位置和相关配置的要求很低，但是受制于行为采集和预处理过程中算法的性能，相机与实验目标的位置关系仍然影响结果的准确性。本研究通过比较视频录制时不同距离和相机角度得到的重建结果，评估数据采集过程中提高重建准确度的最佳拍摄方法。受试者 8 号关节位于相机画面中央，设置相机与水平方向的俯角  $\alpha$  为 15 度、30 度、45 度、60 度和 75 度，相机与人的直线距离  $D$  为 3 米、5 米、7 米、9 米、11 米（图 2）。使用上述角度和距离进行实验时，排除相机高度  $H$  超过 5 米的情况，得到 12 组实验，相关参数如表 1 所示。实验中 45 度 3 米、60 度 3 米、75 度 3 米条件下的 3 组无法识别人体姿态，余下的 9 组实验得到重建的模型。测量受试者腿部关节 4-5 和关节 5-6 的真实长度，并与重建结果对比，结果表明 15 度 5 米条件下得到了均方误差（RSME）最小的重建结果。在所有俯角小于等于 30 度，距离小于

等于 9 米的测试中，关节 4-5 的重建长度均方误差均小于 13 厘米，关节 5-6 的重建长度均方误差均小于 9 厘米。关节长度重建误差在部分已知受试者身高的实验中可以通过整体缩放重建模型的高度而进一步减小。

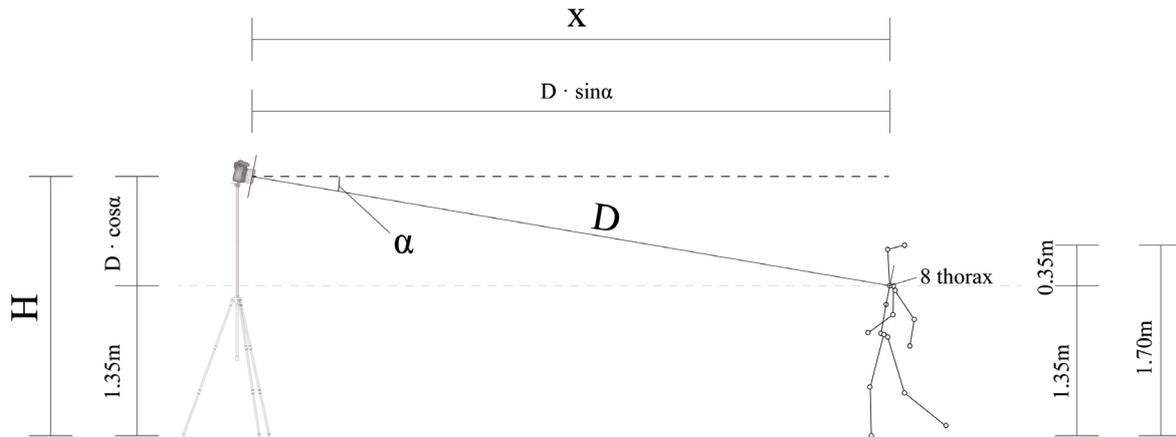


图 2 准确度测量实验示意图（图片来源：作者自绘）

表 1 准确度测量实验参数（图片来源：作者自绘）

$[x(m), H(m)]$	$D=3m$	$D=5m$	$D=7m$	$D=9m$	$D=11m$
$\alpha=15^\circ$	[2.90, 2.13]	[4.83, 2.64]	[6.76, 3.16]	[8.69, 3.68]	[10.63, 4.20]
$\alpha=30^\circ$	[2.60, 2.85]	[4.33, 3.85]	[6.86, 4.85]	[/, 5.85]	[/, 7.85]
$\alpha=45^\circ$	[2.12, 3.47]	[3.54, 4.89]	[/, 6.30]	[/, 7.71]	[/, 8.13]
$\alpha=60^\circ$	[1.50, 3.95]	[/, 5.68]	[/, 7.41]	[/, 9.14]	[/, 10.88]
$\alpha=75^\circ$	[0.18, 4.25]	[/, 6.18]	[/, 8.11]	[/, 10.04]	[/, 11.98]

## 2.3 微观行为指标计算和可视化

在软件界面中，重建的人体骨骼模型可以显示人的微观行为时空序列，通过对时空坐标点的进一步行为计算，可以获取具有分析意义的行为研究指标。以步态分析为例，步态分析相关指标包括步幅长度（Stride Length）、步幅宽度（Stride Width）、步幅速度（Stride Velocity）、双支撑时间（Double Support Time）、步幅变异性（Stride Velocity Variability）等（图 3）。根据脚腕关节 3、关节 6 落地情况进行跨步（stride）的分割，使用落地点坐标进行行为计算，得到上述步态指标的数值。将步态指标数据映射为不同的颜色，赋予每个跨步的骨骼模型，即可将步态指标可视化呈现。

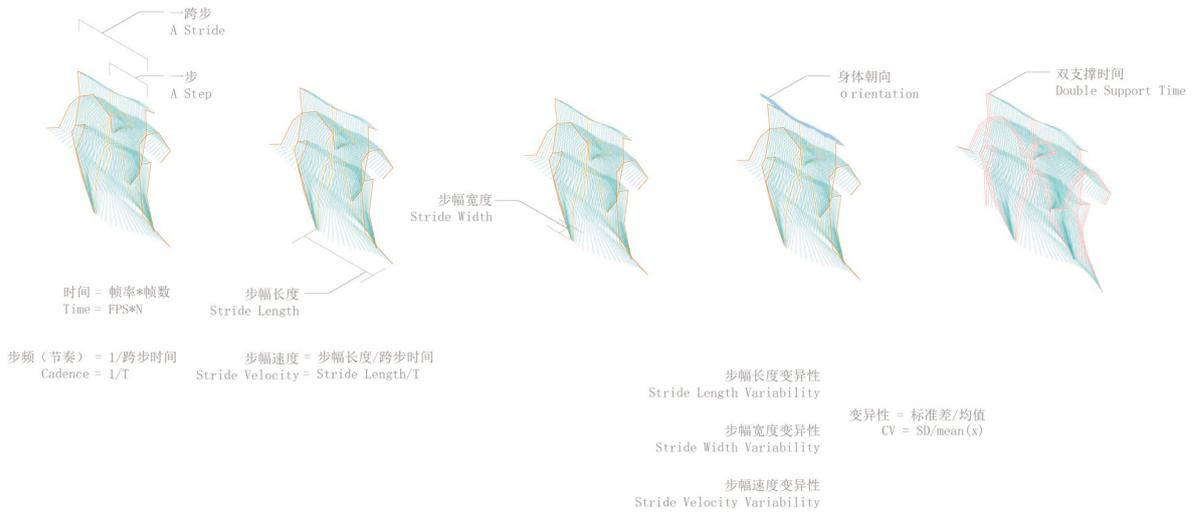


图3 步态相关指标计算（图片来源：作者自绘）

为了确认单目相机视频微观行为计算框架的有效性，本研究比较了动作捕捉（Motion Capture）方法与单目计算方法得到的步态指标差异（图4）。实验中，使用自定义贴点方法将动作捕捉检测反光点贴到与人体位姿估计算法预设的17个关节相同的位置。2名受试者在动作捕捉区域内分别和共同按照十字形路线行走。动作捕捉系统得到了90帧/秒（fps）的17关节坐标数据。实验中同时架设手机录制视频，使用单目相机行为计算方法进行后处理，得到30fps的17关节坐标数据。

结果表明动作捕捉和视频分析方法在步数划分方面结果一致。经过投影变换后，轨迹形状基本一致。据此得到描述同一次行走（选取了连续的4步）的两个版本步态指标数据。在连续的4步中，步幅长度的均方误差约为19厘米。视频分析方法可以在行为轨迹中划分出较为符合实际的落脚点，可供展开与空间耦合的微观行为研究，如近身空间研究等；视频分析方法可以对齐共时行为的时间帧，结合微观行为朝向和停留时间等信息，可供展开时空耦合的微观行为研究，如人群密接研究等。

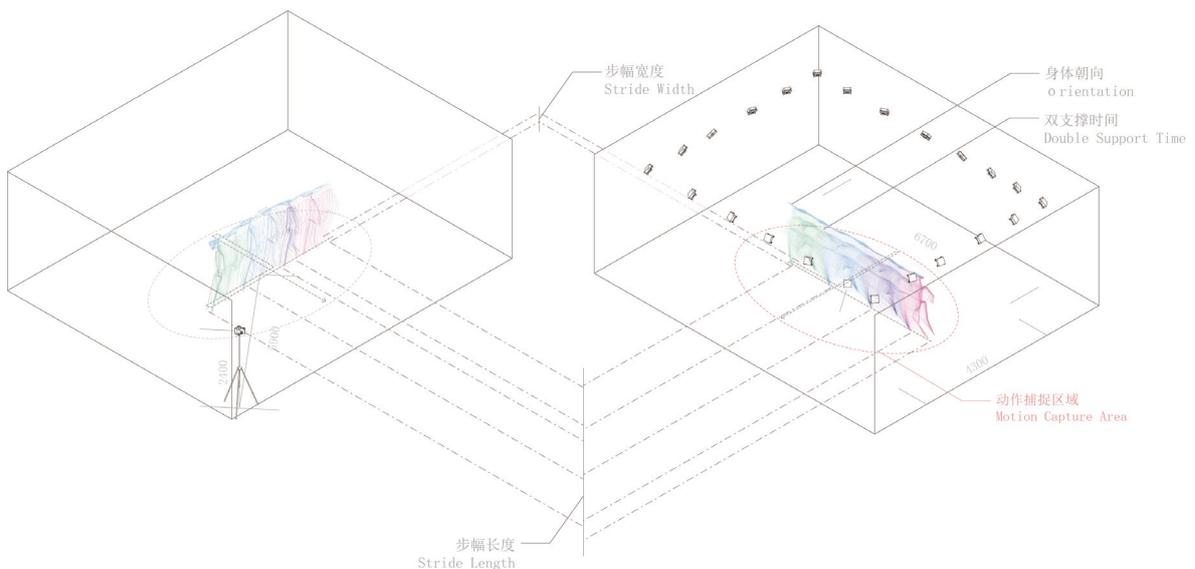


图4 动捕实验示意图（图片来源：作者自绘）

### 3 微观行为计算方法验证

#### 3.1 社交行为与近身空间实验

为了验证本文提出的行为计算方法的有效性，在一项社交行为研究实验中进行了初步应用。该实验探究人的近身空间（Peripersonal Space, PPS）与环境的关系。在某教学楼三楼电梯口空间放置障碍物和障碍人，观测走出电梯的受试者绕过障碍的方式。为了测试本文提出方法的优势，实验分别使用轨迹转折点、最近距离两个宏观指标和步态转折点这一微观指标表征 PPS 半径，探索人在绕行障碍物时的避让距离规律（图 5）。轨迹转折点由受试者的重心轨迹与不放置障碍物（人）时行人的平均轨迹相比较得出；最近距离是障碍物到人体重心运动轨迹的最短距离；步态转折点由落脚点相互连接得到的包含微观信息的步态几何，辅以每一步的向量差模长顺序信息综合判断得出。图 6 说明了以步态转折点表征近身空间的方法：左右两边较粗的折线表示左右脚的实际步幅，颜色代表步幅与前一步的向量差模长的大小，红色代表向量差模长大，蓝色代表向量差模长小；中间较细的折线表示每一小步（step）中的左右脚距离，红色代表距离短，蓝色代表距离长；通过综合分析，向量差模长较大且左右脚距离较短的落脚点为转折点，物体/人与步态转折点连接，连线长度为转折点到障碍物的距离，用以表征受试者在当前环境下对近身空间的判断。

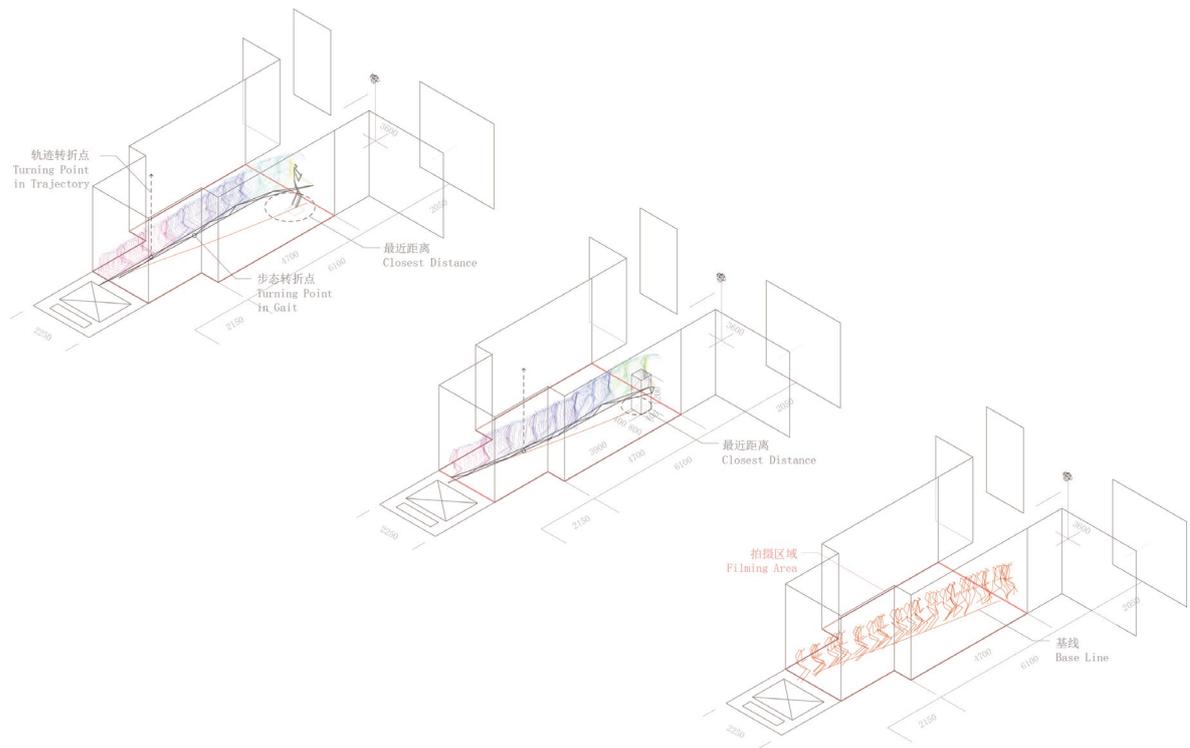


图 5 PPS 实验示意图（图片来源：作者自绘）

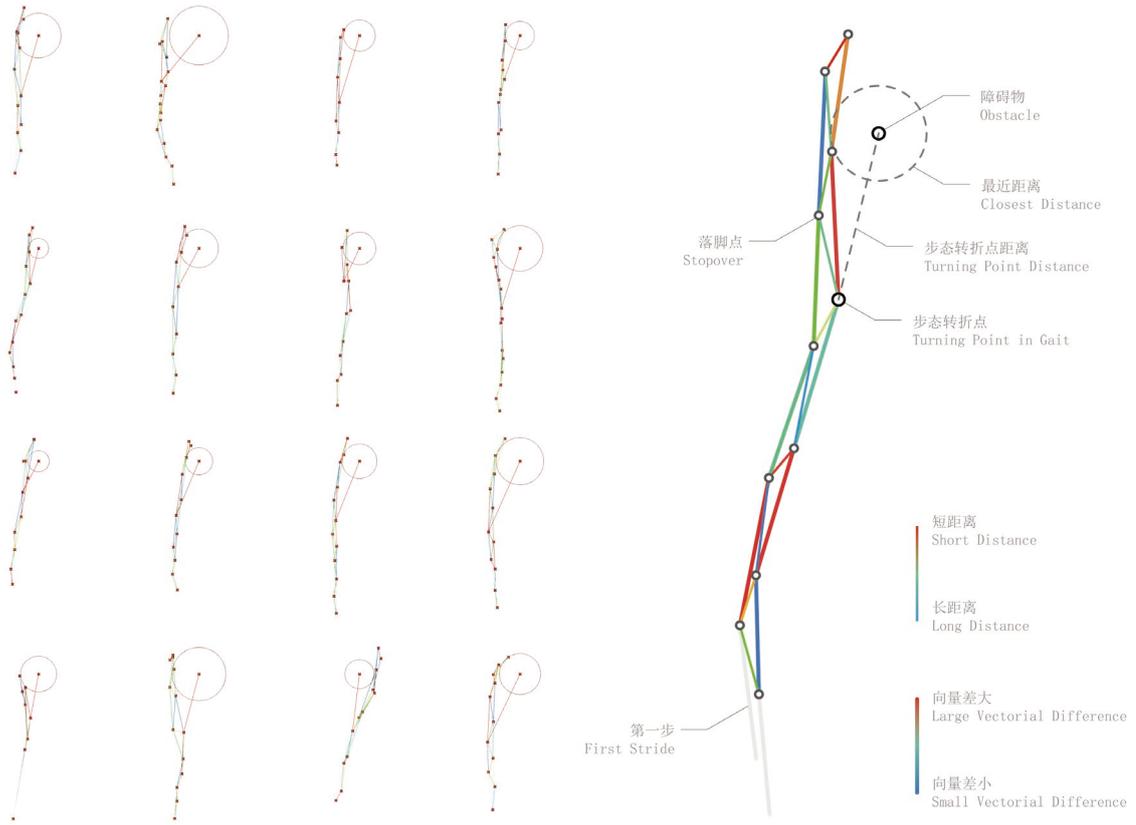


图6 步态几何：向量差可视化和转折点位置（图片来源：作者自绘）

### 3.2 基于微观步态数据的近身空间（PPS）分析

轨迹转折点、最近距离、步态转折点三种方法表征 PPS 半径的结果略有不同。取得轨迹转折点需要与不放置障碍的对照组平均轨迹进行对比：将实验者的轨迹用八次多项式回归拟合，并与对照组平均轨迹的八次多项式回归结果相减，差值函数的二阶导数取得最大值时，将原轨迹上的相应点作为轨迹转折点。实验中部分实验者走出电梯时的初始方向与对照组轨迹相差较远，此时计算的轨迹转折点往往落在在电梯口附近，这对于 PPS 的表征不准确。最近距离是人体重心与障碍物的最近水平距离，不需要与基准对照，但每个人受到障碍物的影响不同，使用最近距离表征 PPS 会受到个体差异的影响。PPS 指示人显著意识到障碍物存在时的距离，相较于另外两种表征方式，步态转折点这一微观指标可以排除部分干扰因素，能够更客观、准确地表征 PPS 半径。根据实验录像评估三种表征方法，微观步态分析得到的结果更接近真实情况。

## 4 讨论

本研究通过方法建构和预实验，初步验证了微观行为计算在解析复杂空间行为规律时的优势。然而，当前的技术方法仍需进一步优化和修正，以提高分析的精准性。例如，在分析步态数据时，跨步的分割是通过设置落脚高度阈值判断 3 号、6 号关节落地情况，进而划分步数。本研究通过 Grasshopper 平台内置的 Galapagos 运算器中的模拟退火算法，以“分割的步数最多”为目标对高度阈值的取值进行优化。这一策略大多数情况下是有效的，但在少数具体案例中仍存在误差，因此需进一步优化跨步的判定规则。

另外, 本研究所提出的技术方法需要进一步考虑实验应用时的操作可行性、便捷性问题。例如, 本研究中行为采集和预处理方法采用自顶至底 (Top Down) 的架构, 即首先进行目标检测, 然后在每个检测框中计算三维位姿估计值。在这个步骤之后, 为了确定人与人之间的相对位置关系, 需要将含有透视的视频图像进行逆透视变形, 得到空间的正投影关系。为解决这一问题, 目前采用的方法是获取视频中图像坐标系的四个点坐标以及现实中世界坐标系相应的四个点坐标, 得到一组透视映射矩阵, 并应用于全部图像。然而, 由于视频拍摄情况不同, 上述过程在实际操作中容易出现较大的人为误差。另一个替代方法是实验之前在场地中标注一个确定边长的矩形区域; 固定相机位置后, 受试者先沿着该区域的边缘行走一圈, 获得人体运动轨迹的基准参照; 在处理视频时, 首先得到未加逆透视变形的人体位姿估计的坐标, 然后根据确定的矩形区域端点位置, 得到估计的坐标与真实位置的映射关系矩阵, 并最终将其应用到所有目标。这种方法可以有效减少人为误差的干扰, 提高了数据准确度, 但是同样也存在让实验步骤变得复杂的问题。

未来研究可以进一步针对多人场景下的行为计算方法开展准确度评估和验证。在多人场景下, 行为计算将扩展到人与人之间的交互方式等更复杂的指标, 这将为真实公共空间的群体行为提供可靠的描述和表征方式。

## 结语

基于机器视觉的行为分析方法在建筑学领域的应用愈发成熟。本文认为, 该技术不仅可以在大规模自动化行为分析的场景中得到应用, 而且可以在提升空间行为分析的精度和深度方面发挥巨大的作用。在此基础上, 本文提出了基于机器视觉的微观行为计算方法, 可以用于评估有效社交距离、微观密接情况等多方面的微观行为指标, 并将微观行为计算结果在空间模型中进行可视化重建。建立微观行为与空间几何的对应关系认知, 为建筑师的设计决策提供基础。

## 参考文献

- [1] 宋晟, 肖敏, 陆熹. 身体、行为、空间——重识公共空间中的“微观尺度”[J]. 国际城市规划, 2024, 39(1): 60-66.
- [2] 杨春侠, 詹鸣, 耿慧志等. 基于行为模拟的户外公共空间研究综述及滨水关键问题探索[J]. 西部人居环境学刊, 2022, 37(5): 89-97.
- [3] 宣晓东, 闫梦辉, 邹俊等. 物理环境介入下游览路径空间品质对游客行为影响探讨——以合肥市三河古镇为例[J]. 南方建筑, 2024(5): 14-28.
- [4] 闫斌, 富志强, 许景鑫. 商业综合体公共空间与行为互动[J]. 城市建筑空间, 2023, 30(3): 105-108.
- [5] 李翔宇, 单镜祎, 崇志国. 城市立体化视角下的地下综合交通枢纽换乘体验提升策略研究——以北京城市副中心站综合交通枢纽为例[J]. 新建筑, 2020(06): 22-26.
- [6] FOUAD A A, EL ANTABLY A. Agent-based modeling and simulation of pandemic propagation in a school environment[J]. International Journal of Architectural Computing, 2023, 21(1): 120-135.
- [7] 林雨铭, 黄蔚欣. 基于 Wi-Fi 定位技术的航站楼旅客时空行为特征分析——以上海虹桥机场 T2 航站楼为例[J]. 新建筑, 2021(02): 24-30.
- [8] 顾宗超, 陆伟, 大佛俊泰. 基于决策层级理论的地下行人系统时空行为研究[J]. 地下空间与工程学报, 2022, 18(S01): 1-13.
- [9] 蒋丹红, 夏林. 视频人流统计技术的应用研究[J]. 现代建筑电气, 2014, 5(1): 1-4.
- [10] 吴韶集, 胡一可. 基于深度学习的商业综合体空间超链接分析——以三里屯太古里和北京 APM 为例[J]. 南方建筑, 2022(1): 61-68.
- [11] ZHANG Z, LONG Y, WANG M, et al. Application of Wearable Cameras in Studying Individual Behaviors in Built Environments. Landscape Architecture Frontiers, 2019, 7(2): 22-38.
- [12] ZHANG N, LIU L, DOU Z, et al. Close contact behaviors of university and school students in 10 indoor environments[J]. Journal of Hazardous Materials, 2023, 458: 132069.
- [13] 杨丽婧, 黄蔚欣, 王月琛. 图像行人识别技术在建筑空间研究中的应用[C]//高等学校建筑学专业教学指导分委员会建筑数字技术教学工作委员会. 数智营造: 2020 年全国建筑院系建筑数字技术教学与研究学术研讨会论文集. 北京: 清华大学建筑学院, 2020: 7.

- [14] GE Z, LIU S, WANG F, et al. Yolox: Exceeding yolo series in 2021[J]. arXiv preprint, 2021, arXiv:2107.08430.
- [15] ZHANG Y, SUN P, JIANG Y, et al. Bytetrack: Multi-object tracking by associating every detection box[C]//European conference on computer vision. Cham: Springer Nature Switzerland, 2022: 1-21.
- [16] SUN K, XIAO B, LIU D, et al. Deep high-resolution representation learning for human pose estimation[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Piscataway: IEEE, 2019: 5693-5703.
- [17] LIU J, GUANG Y, ROJAS J. Gast-net: Graph attention spatio-temporal convolutional networks for 3d human pose estimation in video[J]. arXiv preprint, 2020, arXiv:2003.14179.