

DOI: 10.3969/j.issn.1671-0673. 2026. 01. 007

# 一种融合全局人脸—行人特征的在线跨镜 多目标跟踪方法

胡 钦, 周炜林, 唐 俊, 孔红山

(信息工程大学, 河南 郑州 450001)

**摘要:** 针对目标跟踪单镜视野受限和特征融合不足问题, 提出一种融合全局人脸—行人特征的在线跨镜多目标跟踪方法。首先, 通过构建全局人脸—行人特征池, 采用全局长时身份管理机制, 在单镜跟踪同时实现跨镜信息实时融合与身份关联; 其次, 通过集成多类别特征提取、阈值自适应等模块, 并引入时空—表观联合约束模型, 提升跨镜关联准确性与鲁棒性; 最后, 为验证方法, 构建涵盖不同复杂度的跨镜行人跟踪数据集。实验结果表明: 单镜消失重现中, 身份切换次数(IDS)较对比方法平均降低 20 次以上, 身份识别  $F1$  分数(IDF1)最高提升约 30%; 跨镜跟踪任务中, 多人场景下 IDS 从对比方法的 132 次降至 126 次; 消融实验验证了各模块有效性; 实时性测试显示该方法帧率(FPS)可达 25.15。

**关键词:** 跨镜多目标跟踪; 人脸—行人特征融合; 在线跟踪; 数据集构建

中图分类号: TP391.41

文献标识码: A

文章编号: 1671-0673(2026)01-0048-08

## An Online Multi-Target Multi-Camera Tracking Method with Global Face-Pedestrian Feature Fusion

HU Qin, ZHOU Weilin, TANG Jun, KONG Hongshan

(Information Engineering University, Zhengzhou 450001, China)

**Abstract:** To address the problem of limited single-camera views and insufficient feature fusion of target tracking, an online multi-target multi-camera tracking method with global face-pedestrian feature fusion is proposed. Firstly, by constructing a global face-pedestrian feature pool and adopting a global long-term identity management mechanism, real-time multi-camera information fusion and identity association are realized during single-camera tracking. Secondly, multi-category feature extraction and adaptive threshold adjustment are integrated, and a spatio-temporal-appearance joint constraint model is introduced to enhance the accuracy and robustness of multi-camera correlation. Finally, in order to verify the method, a multi-camera pedestrian tracking dataset covering different complexities is constructed. Experimental results show that compared with those of the comparison methods, the average number of ID switches (IDS) is reduced by more than 20 times, and the maximum number of ID  $F1$  score (IDF1) is increased by about 30%. In the multi-camera tracking task, IDS in the multi-pedestrian scene is reduced from 132 to 126. Ablation experiments confirm the effectiveness of each module of the proposed method, and the real-time test shows that frames per second (FPS) of this method can reach 25.15.

**Key words:** multi-target multi-camera tracking; face-pedestrian feature fusion; online tracking; dataset construction

收稿日期: 2025-09-29; 修回日期: 2025-10-26

作者简介: 胡 钦(1993—), 男, 硕士生, 主要研究方向为计算机视觉。E-mail: nongmo\_hu@outlook.com

随着智能监控网络广泛部署,在线跨镜多目标跟踪在公共安全与智慧城市中应用需求日益增长。然而,目标与背景因光照、移动和遮挡产生的随机变化增加了跟踪难度<sup>[1-2]</sup>。针对单镜视野受限、目标身份切换等问题,OCMCTrack<sup>[3]</sup>(Online Multi-Target Multi-Camera Tracking with Corrective Matching Cascade)引入“校正匹配级联”机制,减少遮挡导致的误匹配。ReidTrack<sup>[4]</sup>(Reid-only Multi-Target Multi-Camera Tracking)利用外观特征聚类实现跨镜关联,但在高遮挡场景下表现不足。LMGP<sup>[5]</sup>(Lifted Multicut Meets Geometry Projections)结合升维多割模型与几何投影,在Duke MTMC数据集实现高精度关联,但属于离线处理,实时性差。MVDeTr<sup>[6]</sup>(Multiview Detection with Shadow Transformer)利用Transformer进行多视角融合,在WILDTRACK数据集上准确率达91.5%,但实时性受限。DyGLIP<sup>[7]</sup>(Dynamic Graph Model with Link Prediction)通过动态图链接预测处理重叠与非重叠场景,在NLPR\_MCT数据集上准确率达89.4%,但多模态融合不充分。TRACTA<sup>[8]</sup>(Tracklet-to-Target)采用轨迹—目标分配与非负矩阵分解进行关联,对高密度人群适应性有限。UMMT<sup>[9]</sup>(Unified Multi-View Multi-Person Tracking)多视角多人跟踪框架,在WILDTRACK数据集

上多目标检测准确率(Multiple Object Detection Accuracy, MODA)达92.3%,但计算开销大。MVT<sup>[10]</sup>(Multi-View Target Transformation)通过脚部特征定位解决空间断裂问题,在MultiviewX数据集上多目标检测精确率(Multiple Object Detection Precision, MODP)达95.0%,但仅适用于离线场景。

针对上述问题,提出一种融合全局人脸—行人特征的在线跨镜多目标跟踪方法,在单镜跟踪同时通过全局管理机制实时融合跨镜信息,实现身份在线判别与跨镜关联。首先,提出全局长时身份管理机制,实现人脸、行人特征深度融合与在线更新,提供鲁棒身份表示;其次,设计时空—表观联合约束模型,将摄像头拓扑先验转化为时空匹配得分,并与外观相似度动态融合以抑制误匹配;最后,构建涵盖不同人数与遮挡复杂度的跨镜多人跟踪数据集,弥补公开数据缺失,为模型评估提供支持。

## 1 跨镜跟踪框架构建

### 1.1 单镜人脸—行人联合跟踪框架

本文提出的人脸—行人联合跟踪框架基于“类内独立跟踪+跨类别动态绑定”思想,整体结构如图1所示。

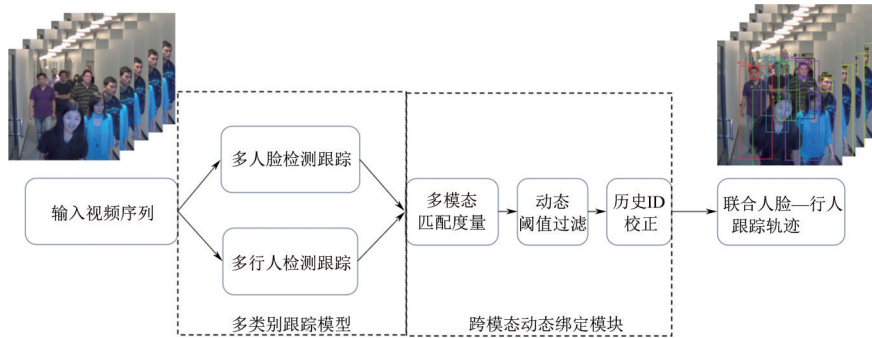


图1 人脸—行人联合跟踪框架

在改进的FairMOT<sup>[11]</sup>(on the Fairness of Detection and Re-Identification in Multiple Object Tracking)基础上,扩展检测头与特征通道,将单类别跟踪模型扩展为可独立跟踪人脸与行人的多类别模型,再通过联合匹配实现二者动态绑定。框架包含以下两个阶段。

1)类内独立跟踪。人脸跟踪分支:通过DLA-34<sup>[12]</sup>主干网络提取人脸特征,输出人脸框 $B_f \in \mathbb{R}^{N_f \times 4}$  ( $N_f$ 为人脸数)及人脸特征 $F_f \in \mathbb{R}^{128}$ ,利用联合检测与嵌入(Joint Detection and Embedding, JDE)跟踪器生成人脸轨迹 $T_f = \{(I_f, B_f, F_f)\}$ ,  $I_f$ 为人脸ID。行

人跟踪分支:使用共享的DLA-34主干网络提取行人特征,输出行人框 $B_p \in \mathbb{R}^{N_p \times 4}$  ( $N_p$ 为行人数)及行人重识别特征 $F_p \in \mathbb{R}^{128}$ ,利用JDE跟踪器生成行人轨迹 $T_p = \{(I_p, B_p, F_p)\}$ ,  $I_p$ 为行人ID。

2)跨类别ID绑定。基于时空—表观联合度量(详见第2.3节),动态关联 $T_f$ 和 $T_p$ ,输出绑定后联合轨迹 $T_j = \{(I_j, B_f, B_p, F_f, F_p)\}$ ,  $I_j = (I_f, I_p)$ 。

### 1.2 在线跨镜跟踪框架

通过扩展单镜人脸—行人联合跟踪框架,融合全局目标特征池与身份管理机制,构建在线跨镜跟踪框架,如图2所示。

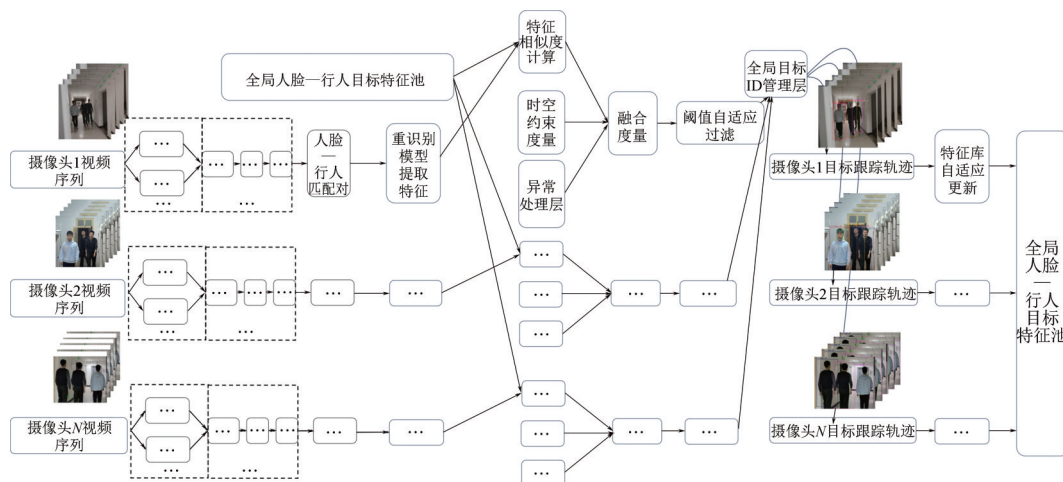


图2 基于全局人脸-行人联合特征的跨镜多目标跟踪框架

该框架的核心创新为引入双类别特征联合匹配机制与时空约束,以获取时空合理且特征相似匹配结果。

系统采用分层处理结构,包含以下模块:1)多类别特征提取:用于并行获取人脸特征与行人外观特征;2)动态匹配决策:融合时空约束与特征相似度计算;3)特征自适应更新:实现特征库时序感知维护;4)阈值自适应调整:基于场景特征的自适应阈值机制。

框架流程可形式化表示为

$$T(F_f, F_p, C) = \arg\max \left[ (1 - \alpha) \cdot S_f(F_f, F_p) + \alpha \cdot S_{st}(C) \right] \quad (1)$$

式中:  $T(F_f, F_p, C)$  为目标跟踪结果;  $F_f$  为人脸特征,  $F_p$  为行人特征;  $C$  为时空约束集合;  $\alpha \in [0, 1]$  为平衡系数,用于控制融合特征与时空约束贡献比例;  $S_f(F_f, F_p)$  为融合特征得分;  $S_{st}(C)$  为时空约束得分。

## 2 在线跨镜多目标跟踪方法

### 2.1 跨镜多目标跟踪数据集构建

由于目前尚无公开数据集适用于基于人脸-行人特征的在线跨镜跟踪方法,本文自行拍摄多组跨镜视频并标注,构建专用数据集,录制示意图如图3所示。

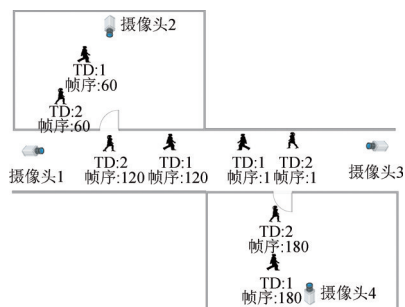


图3 跨镜监控录制场景示意图

数据集由4个摄像头同步采集,每次录制生成4段视频,构成1组跨镜序列。按场景人数分别录制1~4人视频组,共4组16段视频。数据集基本信息如表1所示。

表1 自建跨镜多目标跟踪数据集基本信息

| 数据集 | 视频数 | 行人数 | 人脸框   | 人脸密度 | 行人框  | 行人密度 | 帧数/(s <sup>-1</sup> ) |
|-----|-----|-----|-------|------|------|------|-----------------------|
| 第1组 | 4   | 1   | 338   | 1.00 | 717  | 1.00 | 1 820                 |
| 第2组 | 4   | 2   | 1 038 | 1.63 | 1874 | 1.62 | 2 784                 |
| 第3组 | 4   | 3   | 1 603 | 2.19 | 2730 | 2.11 | 2 941                 |
| 第4组 | 4   | 4   | 2 151 | 2.46 | 3436 | 2.36 | 3 058                 |
| 合计  | 16  | 4   | 5 130 | 1.97 | 8757 | 1.89 | 10 603                |

### 2.2 多类别特征建模

1)人脸-行人特征建模。定义目标特征为人脸特征与行人外观特征的联合表示:

$$F = \{F_f, F_p\} \quad (2)$$

式中:  $F_f \in \mathbb{R}^{2048}$ ;  $F_p \in \mathbb{R}^{2048}$ 。两者分别基于 ResNet50 训练的人脸、行人重识别模型进行特征提取。

2)人脸特征质量评估。人脸特征可信度评估公式为

$$Q_f = 0.5 \min \left( \frac{A_b}{2000}, 1 \right) + 0.5 S_d \quad (3)$$

式中:  $A_b$  为人脸检测框面积; 2 000 为基于实验与统计确定的基准面积;  $S_d$  为人脸检测置信度; 权重 0.5 根据初步实验确定,用于平衡清晰度与检测置信度。

### 2.3 动态匹配策略

#### 2.3.1 特征融合模型

动态调整特征权重,并对融合特征进行归一化处理,实现鲁棒匹配:

$$S_{\text{fuse}} = \frac{w_p S_p + w_f S_f}{w_p + w_f} \quad (4)$$

$$\begin{cases} w_p = 0.7(1 + 0.2(1 - Q_i)); \\ w_f = 0.3Q_i. \end{cases} \quad (5)$$

式中:  $S_p$  为行人相似度得分;  $S_f$  为目标人脸相似度得分。经网格搜索, 设定行人特征基础权重  $w_p$  为 0.7, 人脸特征基础权重  $w_f$  为 0.3。根据人脸特征可信度, 动态调整行人和人脸特征基础权重。

### 2.3.2 时空约束模型

为提升跨镜匹配时空合理性, 本文构建时空联合约束模型。通过融合摄像头拓扑约束与动态时间容差机制, 实现跨镜匹配的物理可解释性与场景自适应性。对多摄像头拓扑结构作如下定义: 建立摄像头网络  $G = (V, E)$ 。其中: 节点  $v_i \in V$ , 表示摄像头  $c_i$ ; 边  $e_{ij} \in E$ , 表示转移时间  $t_{ij}$ 。

1) 时空匹配得分。给定目标  $A$  (摄像头  $c_a$ , 最后出现时间  $t_a$ ) 与目标  $B$  (摄像头  $c_b$ , 最后出现时间  $t_b$ ), 时空得分计算为

$$S_{st} = \begin{cases} \max\left(0, 1 - \frac{1}{\delta_t} |t_b - t_a - t_{ab}|\right), & c_b \in N_{c_a}; \\ 0, & \text{其他}. \end{cases} \quad (6)$$

式中:  $t_{ab}$  为摄像头  $c_a$  到  $c_b$  的转移时间;  $\delta_t$  为时间容差参数;  $N_{c_a}$  为  $c_a$  的邻接摄像头集合。

2) 联合匹配得分。最终匹配得分由融合特征得分  $S_{fuse}$  和时空得分  $S_{st}$  线性组合:

$$S_{total} = (1 - \alpha)S_{fuse} + \alpha S_{st} \quad (7)$$

式中,  $\alpha \in [0, 1]$ 。经在验证集上的网格搜索最终确定, 跨镜匹配时, 为强化时空约束, 缓解视角差异导致的外观特征失真问题, 取  $\alpha = 0.3$ ; 同镜匹配时, 则侧重外观相似性, 避免短时重复匹配的时空误判, 取  $\alpha = 0.1$ 。通过参数  $\alpha$ , 平衡外观特征与时空约束的贡献度。

## 2.4 特征自适应更新机制

### 2.4.1 特征时序衰减策略

为解决目标外观动态变化问题, 设计基于指数衰减的特征更新机制, 通过时间衰减函数增强近期特征表示权重, 平滑处理因光照和姿态变化引起的外观波动, 保留长期特征中稳定成分, 提升特征匹配鲁棒性。

设当前时刻为  $t$ , 特征采集时刻为  $t_i$ , 衰减因子为  $\lambda$ , 则特征权重计算为

$$w_i = e^{-\lambda \cdot (t - t_i)} \quad (8)$$

为适应行人外观变化等情况, 针对行人特征使用较快衰减, 经初步实验确定, 取  $\lambda_p = 0.1$ ; 为使人脸特征具有更长记忆周期, 针对人脸特征使用较慢衰减, 取  $\lambda_f = 0.05$ 。

### 2.4.2 特征更新机制

行人特征采用加权平均更新:

$$\mathbf{F}_p^{\text{new}} = \sum_{i=1}^n w_p^i \mathbf{F}_p^i / \sum w_p^i \quad (9)$$

人脸特征采用质量感知更新策略:

$$\mathbf{F}_f^{\text{new}} = \begin{cases} \mathbf{F}_f^{\text{current}}, & Q_f \geq 0.4; \\ \sum_{i=1}^n w_f^i \mathbf{F}_f^i / \sum w_f^i, & \text{其他}. \end{cases} \quad (10)$$

行人特征  $\mathbf{F}_p^{\text{new}}$  始终使用加权平均进行更新, 结合时序衰减权重, 强化近期特征。人脸特征采用高质量时覆盖更新, 低质量时衰减加权平均方式进行更新, 阈值 0.4 根据初步实验确定, 尽可能保留历史可靠特征。

## 2.5 基于场景特征的自适应阈值机制

针对传统固定阈值方法在复杂场景中鲁棒性不足问题, 本文提出动态阈值调整策略。通过分析场景特征分布特性, 建立阈值自适应计算模型。

1) 相似度窗口。设  $W_c$  为跨镜相似度滑动窗口,  $W_s$  为同摄像头相似度滑动窗口, 窗口长度分别设置为  $L_c = 100$  和  $L_s = 30$ , 用于记录历史相似度数据。

2) 动态阈值计算。给定基础阈值  $T_b$ , 当前窗口标准差  $\sigma$ , 最大经验标准差  $\sigma_{\max}$ , 则动态阈值  $T_d$  计算为

$$T_d = T_b + k \cdot \left(1 - \frac{\sigma}{\sigma_{\max}}\right) \quad (11)$$

式中, 调节系数  $k \in [0.05, 0.1]$ , 通过实验确定最优值。该公式的物理意义在于: 当特征分布离散度即当前窗口标准差  $\sigma$  增大时, 适当降低阈值以保持匹配灵敏度; 当特征集中时, 提高阈值以减少误匹配。

本文方法相比传统固定阈值方案具有以下优势: 1) 自适应不同摄像头视角的特征分布差异; 2) 动态平衡召回率与准确率指标; 3) 降低人工调参成本。

## 3 实验与分析

为验证方法有效性, 本文在自建跨镜多目标跟踪数据集上进行 3 项实验: 跨镜多目标跟踪性能测试、消融实验以及实时性测试。

前两项实验核心评价指标包括: 身份识别 F1 分数 (ID F1 Score, IDF1)、多目标跟踪准确率 (Multiple Object Tracking Accuracy, MOTA)、多目标跟踪精确率 (Multiple Object Tracking Precision, MOTP)、

身份切换次数(ID Switches, IDS)。第3项实验测试单镜跟踪与跨镜关联的运行时间及平均帧率(Frames Per Second, FPS)。

### 3.1 跨镜多目标跟踪方法性能测试

#### 3.1.1 单镜行人消失重现跟踪实验

在跨镜跟踪任务中,单镜行人消失重现场景易引起ID切换,影响轨迹连续性。

为评估所提方法在该场景下跟踪能力,选取自建数据集中存在消失重现情形单镜视频,与FairMOT(单阶段联合检测跟踪模型)、ByteTrack(分段式检测跟踪模型)及PP-Tracking(百度飞桨开源目标跟踪模型)进行对比。

各方法跟踪效果对比如图4所示。图4中,相同颜色框表示同一ID,以边界框颜色区分各自跟踪过程中的目标轨迹ID。从图4可见,本文方法在目标重现后仍能保持ID一致,跟踪连续性明显优于对比方法。FairMOT与ByteTrack作为单镜跟踪

模型,对消失重现目标几乎无法维持ID;PP-Tracking采用“单镜跟踪—跨镜聚类”的离线模式,在单镜内仍将重现目标视为新ID;而本文方法通过在线身份管理实现了重现目标的稳定关联,反映出本文方法处理单镜情况下目标消失重现的优势。

各方法跟踪测试结果如表2所示。由表2可知,本文方法在身份一致性指标上显著优于对比方法;IDF1达78.17%,身份精确率(ID Precision, IDP)为80.68%,身份召回率(ID Recall, IDR)为75.93%,IDS仅为16次,远低于FairMOT(42次)、ByteTrack(46次)和PP-Tracking(36次),说明人脸—行人联合特征与全局身份管理有效抑制了ID混淆;整体跟踪精度方面,MOTA为91.45%,MOTP为5.82%,均优于对比方法;误检数(False Positive, FP)为27,漏检数(False Negative, FN)为169,轨迹中断数(Fragmentation, FM)为29,均为最低,进一步体现其在复杂场景下的高召回与稳定性。

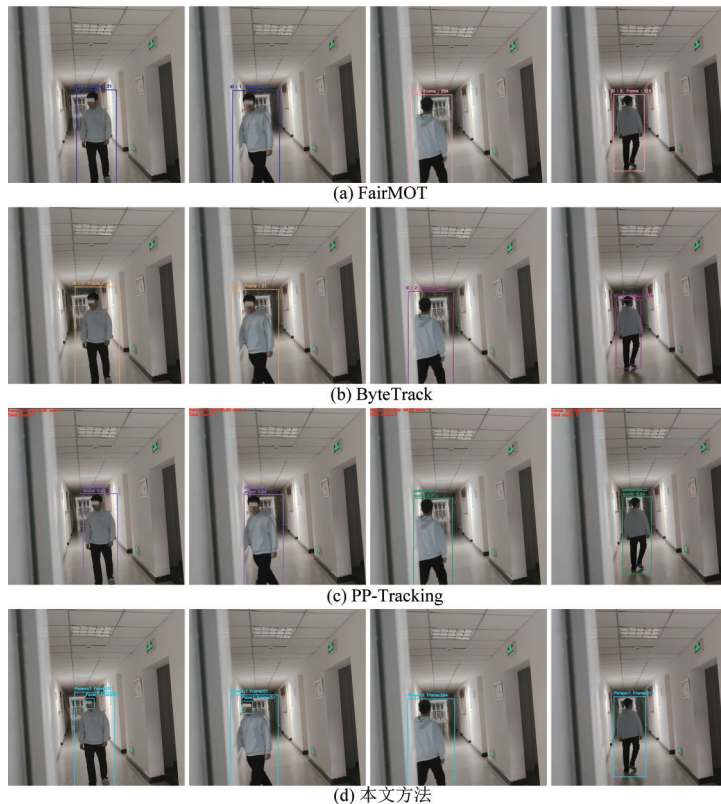


图4 各方法对单镜行人消失重现跟踪效果对比图

表2 各方法对单镜行人消失重现跟踪测试结果

| 方法          | IDF1/% | IDP/% | IDR/% | FP | FN  | IDS | FM | MOTA/% | MOTP/% |
|-------------|--------|-------|-------|----|-----|-----|----|--------|--------|
| FairMOT     | 50.62  | 52.23 | 49.11 | 32 | 180 | 42  | 41 | 89.76  | 6.37   |
| ByteTrack   | 48.65  | 50.35 | 47.06 | 29 | 191 | 46  | 39 | 89.27  | 6.29   |
| PP-Tracking | 50.79  | 52.35 | 49.31 | 31 | 175 | 36  | 33 | 90.24  | 6.31   |
| 本文方法        | 78.17  | 80.68 | 75.93 | 27 | 169 | 16  | 29 | 91.45  | 5.82   |

### 3.1.2 跨镜行人跟踪实验

为验证所提方法在跨镜行人跟踪任务中的性能,在自建数据集上设置单人及多人两种场景(均

使用4个摄像头),并与PP-Tracking进行对比。PP-Tracking与本文方法的跟踪效果对比如图5所示,跟踪测试结果如表3所示。

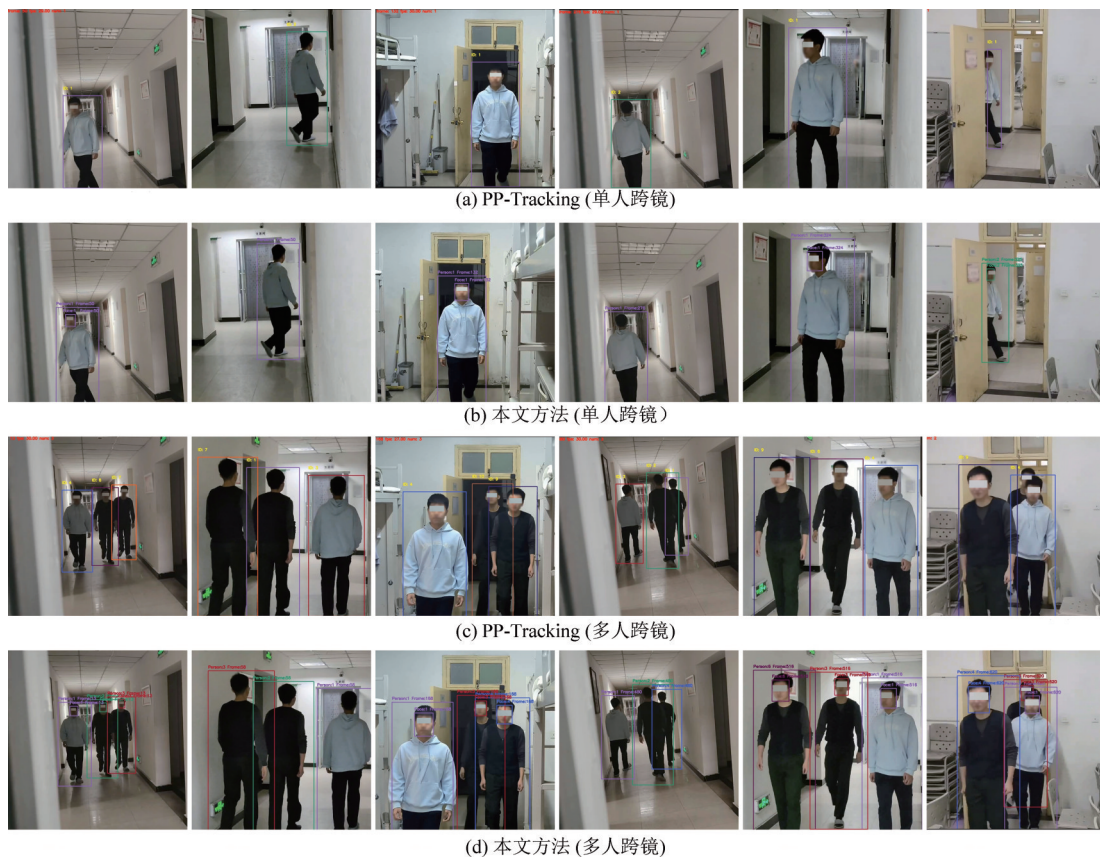


图5 跨镜行人跟踪效果对比图

表3 跨镜行人跟踪测试结果

| 实验场景 | 方法          | IDF1/% | IDP/% | IDR/% | FP  | FN  | IDS | FM | MOTA/% | MOTP/% |
|------|-------------|--------|-------|-------|-----|-----|-----|----|--------|--------|
| 单人跨镜 | PP-Tracking | 92.12  | 96.08 | 88.84 | 15  | 66  | 12  | 10 | 87.03  | 5.59   |
|      | 本文方法        | 95.05  | 98.37 | 92.33 | 4   | 45  | 4   | 6  | 92.61  | 5.50   |
| 多人跨镜 | PP-Tracking | 67.98  | 69.51 | 67.06 | 147 | 198 | 132 | 36 | 74.53  | 6.50   |
|      | 本文方法        | 69.17  | 70.72 | 68.23 | 131 | 182 | 126 | 34 | 76.56  | 6.44   |

在单人跨镜场景下,本文方法在多数指标上优于PP-Tracking:IDF1达95.05%(PP-Tracking为92.12%),IDS仅为4次(PP-Tracking为12次),MOTA提升至92.61%,FP与FN进一步降低。从图5(a)和图5(b)可见,两类方法基本实现了跨镜连续跟踪,但PP-Tracking在图5(a)第3张和第5张出现ID切换,本文方法仅在图5(b)第6张发生1次,稳定性更强。图5(a)第1张和第4张与图5(c)第1张和第4张均显示,PP-Tracking在单镜目标消失重现时仍无法保持ID连续,与第3.1.1节结论一致,进一步体现本文方法在简单场景下的良好身份保持能力。

在多人跨镜场景中,两类方法性能均有所下

降,但本文方法仍更具鲁棒性:IDF1由67.98%提升至69.17%,IDS从132次降至126次,表明身份一致性维护更好;MOTA由74.53%提高至76.56%。从图5(d)可见,在3人跨镜跟踪时,本文方法ID切换明显少于PP-Tracking,跟踪更稳定。

实验表明,本文方法在不同复杂度场景下均具有良好的适应性与稳定性,尤其在抑制身份切换、提升跨镜关联可靠性方面优势明显,可为实际跨镜多目标跟踪系统提供有效解决方案。

### 3.2 消融实验

为评估各核心模块有效性,在基于行人特征的基线模型(Baseline)上逐步增加以下模块进行消融实验:人脸—行人联合特征模块(Joint Face-Pe-

destrian Feature Module, JFPFM)、时空约束模型(Spatio-Temporal Constraint Model, STCM)、特征自适应更新机制(Feature Adaptive Update Mechanism, FAUM)以及自适应阈值机制(Adaptive Threshold Mechanism, ATM)。实验在自建跨镜多行人数据集上进行,结果如表4所示。

表4 跨镜多目标跟踪方法消融实验结果

| 方法                  | IDF1/% | IDP/% | IDR/% | FP  | FN  | IDS | FM | MOTA/% | MOTP/% |
|---------------------|--------|-------|-------|-----|-----|-----|----|--------|--------|
| 基线模型                | 41.86  | 42.70 | 41.12 | 132 | 224 | 243 | 40 | 76.87  | 7.27   |
| JFPFM               | 44.85  | 45.80 | 44.05 | 139 | 231 | 209 | 45 | 77.64  | 9.05   |
| JFPFM+STCM          | 51.18  | 52.12 | 50.27 | 134 | 226 | 200 | 40 | 78.38  | 6.41   |
| JFPFM+STCM+FAUM     | 56.33  | 57.41 | 55.33 | 132 | 224 | 190 | 39 | 78.92  | 6.52   |
| JFPFM+STCM+FAUM+ATM | 68.20  | 69.88 | 66.99 | 135 | 227 | 131 | 42 | 80.97  | 6.97   |

加入JFPFM后,IDF1提升2.99个百分点(41.86%至44.85%),IDS减少34次,表明人脸特征增强了对相似行人的区分能力。但同时FP、FN各增加7次,MOTP从7.27%退化至9.05%,说明低质量人脸特征可能引入噪声,影响定位精度。

引入STCM后,IDF1进一步提高6.33个百分点(44.85%至51.18%),IDS再降9次,MOTA从77.64%提升至78.38%,MOTP优化至6.41%。结果显示时空约束能有效排除不合理关联,提升轨迹合理性与定位准确性。

增加FAUM后,IDF1继续提升5.15个百分点(51.18%至56.33%),IDS减少10次,MOTA同步上升。该机制通过动态更新特征适应外观变化,提升长期跟踪的鲁棒性,且未引入明显匹配噪声。

加入ATM带来最大性能增益,IDF1进一步增长11.87个百分点(56.33%到68.20%),IDS大幅减少59次,说明动态阈值能更好平衡精确率与召回率。但MOTP从6.52%升至6.97%,推测因匹配更

多定位不精确目标导致。

完整模型相比基线模型IDF1累计提升26.34个百分点,MOTA提高4.1个百分点,MOTP基本持平,说明方法主要优化身份关联而非基础跟踪精度,符合设计预期。实验揭示了特征丰富性与噪声(如JFPFM)、关联鲁棒性与定位精度(如ATM)之间的权衡,提示后续可开发更鲁棒的特征融合策略与更精细的自适应阈值机制。

3.3 实时性能测试

为评估所提方法实时性能,实验采用多线程模拟分布式架构:各摄像头分配独立线程进行局部跟踪与特征提取,全局线程负责跨镜关联与身份管理。以仅使用行人重识别特征的基线模型为对比,在包含1~4人、4路摄像头的测试集上分别评估两类方法单镜跟踪与跨镜关联的平均帧率。仿真实验在统一硬件平台(CPU i7-14700KF, GPU RTX 4060Ti 16 GB, RAM 48 GB)上进行,每组实验重复3次取均值。实验结果如表5所示。

表5 跨镜多目标跟踪方法实时性测试结果

| 方法   | 场景    | 总帧数   | 单镜<br>平均帧数 | 单镜<br>平均耗时/s | 单镜平均<br>FPS/(s <sup>-1</sup> ) | 全局单镜<br>平均FPS/(s <sup>-1</sup> ) | 跨镜关联<br>平均耗时/s | 跨镜关联<br>平均FPS/(s <sup>-1</sup> ) | 全局跨镜关联<br>平均FPS/(s <sup>-1</sup> ) |
|------|-------|-------|------------|--------------|--------------------------------|----------------------------------|----------------|----------------------------------|------------------------------------|
| 本文方法 | 1人跨4镜 | 1 820 | 455        | 16.15        | 28.17                          | 25.15                            | 3.84           | 473.96                           | 257.09                             |
|      | 2人跨4镜 | 2 784 | 696        | 27.48        | 25.33                          |                                  | 11.05          | 252.06                           |                                    |
|      | 3人跨4镜 | 2 941 | 735        | 30.32        | 24.25                          |                                  | 17.11          | 171.89                           |                                    |
|      | 4人跨4镜 | 3 058 | 765        | 33.46        | 22.85                          |                                  | 23.44          | 130.46                           |                                    |
| 基线模型 | 1人跨4镜 | 1 820 | 455        | 14.21        | 32.02                          | 29.64                            | 1.87           | 973.26                           | 477.39                             |
|      | 2人跨4镜 | 2 784 | 696        | 22.35        | 31.14                          |                                  | 6.19           | 449.76                           |                                    |
|      | 3人跨4镜 | 2 941 | 735        | 25.59        | 28.73                          |                                  | 10.36          | 283.88                           |                                    |
|      | 4人跨4镜 | 3 058 | 765        | 28.67        | 26.67                          |                                  | 15.09          | 202.65                           |                                    |

实验结果显示,在单镜跟踪阶段,本文方法的FPS为22.85~28.17,低于基线模型的26.67~32.02,说明引入功能模块带来一定计算开销。但在4人跨4镜复杂场景中,本文方法的FPS仍可达22.85,接近实时性下限25<sup>[13]</sup>,具备实际可行性。跨镜关联

阶段,两类模型的FPS均远高于25,表明关联非系统瓶颈。

综合分析表明,尽管功能模块增加了计算负载,但其显著提升了关联准确性,在低复杂度场景中可实现精度与效率的良好平衡。进一步分析发

现,系统瓶颈随场景复杂度动态变化:低负载下以单镜跟踪为主,高负载下跨镜关联逐步成为制约因素。当两阶段FPS接近时,系统达到最佳平衡点,此时可通过特征池容量控制、分层检索或近似最近邻搜索等方法进一步提升效率。本研究为在线跨镜跟踪系统的实际部署提供了模型选择与优化依据,比较适用于公共安全等高精度关联场景。

## 4 结束语

针对单镜视野受限与跨镜在线鲁棒跟踪需求,提出一种融合全局人脸—行人特征的在线跨镜多目标跟踪方法,缓解了复杂场景下身份一致性保持与跨镜关联难题。本文提出融合人脸与行人特征的联合建模方法,通过构建全局特征池和全局长时身份管理机制,结合分层处理架构集成多类别特征提取、动态匹配决策、阈值调整等模块,并引入时空—表观联合约束模型以提升跨镜匹配判别能力。

为验证性能,构建了涵盖不同人数与遮挡场景的跨镜行人跟踪数据集。实验表明,本文方法在多项指标上优于主流对比算法:在单镜消失重现测试中,IDS降至16次,低于FairMOT(42次)、ByteTrack(46次)和PP-Tracking(36次),IDF1达78.17%;在跨镜跟踪中,单人场景IDF1为95.05%、IDS仅4次,多人场景IDS由132次降至126次,显示出良好的身份一致性保持能力。消融实验验证了各关键模块的有效性。实时测试FPS达25.15,测试还表明低复杂度场景下系统瓶颈在单镜跟踪阶段,高复杂度下跨镜关联逐步成为制约因素,为实际部署参数配置提供了依据。未来工作将扩展数据集规模与场景多样性,引入低光照、重度遮挡等极端场景,并验证方法在跨域环境下的泛化能力。

### 参考文献:

- [1] 朱伟杰,朱洪军,伍祥,等.基于相关滤波技术的目标跟踪方法综述[J].信息工程大学学报,2019(6):684-688.
- [2] 李豪. Deep Snake 协助相关滤波目标跟踪[J].信息工程大学学报,2024,25(5):538-544.
- [3] SPECKER A. OCMCTrack: online multi-target multi-camera tracking with corrective matching cascade [C]//Proceedings of the 2024 IEEE/CVF Conference on Computer

Vision and Pattern Recognition Workshops. Piscataway, USA: IEEE, 2024:7236-7244.

- [4] SPECKER A, BEYERER J. ReidTrack: Reid-only multi-target multi-camera tracking [C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway, USA: IEEE, 2023: 5442-5452.
- [5] NGUYEN D M H, HENSCHER R, ROSENHAHN B, et al. LMGP: lifted multicut meets geometry projections for multi-camera multi-object tracking [C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2022:8856-8865.
- [6] HOU Y Z, ZHENG L. Multiview detection with shadow transformer (and view-coherent data augmentation) [C]//Proceedings of the 29th ACM International Conference on Multimedia. New York, USA: ACM, 2021:1673-1682.
- [7] QUACH K G, NGUYEN P, LE H, et al. DyGLIP: a dynamic graph model with link prediction for accurate multi-camera multiple object tracking [C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2021:13779-13788.
- [8] HE Y H, WEI X, HONG X P, et al. Multi-target multi-camera tracking by tracklet-to-target assignment [J]. IEEE Transactions on Image Processing, 2020,29:5191-5205.
- [9] YANG F, ODASHIMA S, YAMAO S, et al. A unified multi-view multi-person tracking framework [J]. Computational Visual Media, 2024,10(1):137-160.
- [10] LEE W Y, JOVANOVIĆ L, PHILIPS W. Multi-view target transformation for pedestrian detection [C]//Proceedings of the 2023 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops. Piscataway, USA: IEEE, 2023:90-99.
- [11] ZHANG Y F, WANG C Y, WANG X G, et al. FairMOT: on the fairness of detection and re-identification in multiple object tracking [J]. International Journal of Computer Vision, 2021,129(11):3069-3087.
- [12] YU F, WANG D Q, SHELHAMER E, et al. Deep layer aggregation [C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2018:2403-2412.
- [13] 国家市场监督管理总局,中国国家标准化管理委员会. 安防监控视频实时智能分析设备技术要求:GB/T 30147—2013[S].北京:中国标准出版社,2014:88.

(编辑:李志豪)