

# 评论：用于通用多摄像头人员跟踪的分层聚类 and 细化技术

Zeal Liang  
z5325156

## 1. 引言

在本文中，作者提出了一种新的多目标多摄像头跟踪 (MTMCT) 框架。该框架在监控场景中具有巨大潜力。然而，由于视角不同、遮挡严重和光照变化等因素，这项任务极具挑战性。作者的主要目标是应对这些挑战，以提高 MTMCT 的准确性和鲁棒性。MTMCT 领域存在许多问题。首先，随着虚拟监控场景的发展，将 MTMCT 应用于多样化的监控场景越来越值得探索。其次，单个摄像头内的目标跟踪只能捕捉短期轨迹，鲁棒性不够，无法将目标与单个轨迹联系起来，并且容易受到视角、光照和图像分辨率变化的影响。在公共场所，同一身份往往会在同一摄像头下多次出现和消失，从而导致同一身份在同一摄像头下出现多个轨迹。因此，作者提出了一个分层聚类 and 细化框架，希望通过逐步聚类和细化来减少错误匹配。

完善策略[1]。

解决这些问题有几个重要方面。首先，在现实世界的监控应用中，确保不同摄像头追踪人员的准确性和一致性至关重要，这可以提高公共安全系统的效率。其次，随着虚拟监控场景的兴起，将 MTMCT 扩展到虚拟环境可以为现实和虚拟世界的监控提供综合解决方案。通过引入分层聚类 and 细化，作者希望克服现有方法中存在的问题，从而为 MTMCT 领域的发展带来积极影响。

如果这些问题得到有效解决，将会带来多重益处。监控系统操作员可以获得更准确、更一致的人员跟踪信息，从而改善应急响应和安全状况。在实际应用中，它还能能为犯罪侦查、事件分析等提供更丰富的信息。此外，该方法的成功实施将改变目前 MTMCT 领域的实践方法，并为未来的研究和应用提供新的思路 and 方向。

## 2. 方法

作者提出的框架包含四个主要部分：人物检测、人物再识别、单摄像头跟踪和多摄像头跟踪。整个过程可归纳如下：(1) 获取一个人形图像

使用人物检测器从每个摄像头视图中提取边界框；

(2) 使用预先训练好的再识别模型，从每个边界框中提取行人再识别特征；(3) 使用单摄像头跟踪模型，为每个摄像头生成单摄像头轨迹片段；(4) 对轨迹片段特征进行聚类，以关联摄像头内的轨迹片段；(5) 应用聚类方法关联摄像头间的轨迹片段；(6) 使用外观、时空和面部特征[1]，完善摄像头间的轨迹片段。

人员检测是多摄像机人员跟踪的第一步，因此选择一个可靠的检测器至关重要。作者选择使用 YOLOv5 作为检测器[1]。这是一种高精度、高效率的目标检测算法，适用于各种应用场景。但是，论文没有讨论可能存在的缺点，例如 YOLOv5 可能不擅长检测小尺寸目标，这在某些情况下可能会影响跟踪精度。因此，我建议作者应该尝试几种不同的人员检测算法，并进行比较，以确保选择了最合适的方法。

人员再识别在多摄像头人员跟踪中起着至关重要的作用。作者选择了基于 TransReID-SSL 的人员再识别模型，该模型在 LUperson 数据集上经过预训练，能够提取稳健且领域不变的 ReID 特征。通过在多个数据集上对模型进行微调，作者进一步提高了模型的性能。不过，这种方法在重新识别过程中也可能受到不匹配的影响，特别是在跨摄像头跟踪中，可能会出现 ID 切换问题。此外，论文也没有明确说明这些数据集之间的领域差异，以及如何解决领域偏差问题。

作者使用 ByteTrack 作为单摄像头跟踪算法，该算法能够通过考虑运动信息和视觉相似性，将每个检测帧与唯一身份关联起来。在多摄像头跟踪中，作者使用聚类方法关联摄像头内的跟踪小块，并利用外观、时空和人脸约束来完善跨摄像头的跟踪小块[1]。不过，我注意到这篇论文没有讨论 ByteTrack 算法是否能在密集场景中保持其效率和准确性，因此可能需要更多的验证和分析。

在多摄像头人物关联方面，作者使用 K-means 聚类算法对汇总的轨迹片段进行分组。然而，K 均值聚类对噪声和异常值非常敏感，可能需要根据数据分布调整参数。作者提到了一种加权策略来解决某些问题，但没有提

选择权重的方法。

总之，虽然本研究提出了 MTMCT 框架，但我认为在方法选择、参数调整和方法局限性方面仍有改进的余地。作者可以进一步探讨方法选择的合理性，提供更多的实验和比较结果，并讨论方法的局限性和可能的改进方向。

### 3. 成果

作者对提议的 MTMCT 框架进行了广泛的实验分析，重点使用 IDF1、IDP 和 IDR 指标来评估提议框架的有效性。文中讨论了实施细节、弱化分析以及与其他模型比较。

用于实验的数据集非常全面，包括来自 130 个摄像头的 1,491 分钟高分辨率视频数据。作者同时使用了真实数据和合成数据，还使用了公开的人物 ReID 数据集进行模型训练，进一步增强了评估的稳健性。

定量分析对分层聚类 and 细化策略进行了弱化研究，证明了它们对整体性能的影响。结果表明，建议的分层聚类策略（相机内和相机间聚类）显著提高了性能。此外，各种细化策略也对结果产生了积极影响，其中外观、时空和轨迹级细化分别将性能提高了 2%、1% 和 2%。这些策略对性能的改善是渐进的，并得到了令人信服的证明。

作者选择 TransReID-SSL 作为骨干网络的决定得到了不同骨干网络弱化分析的支持。经过微调的 TransReID-SSL 在不同骨干网络中都取得了最佳结果，从而提高了论文结论的整体可信度。作者还提供了 MTMCT 最终结果的可视化，使其方法的效果更加清晰。跨摄像机和跨时间的轨迹可视化表明，该模型确实有效地匹配了轨迹，即使在不同视角和遮挡的摄像机之间也是如此。此外，跨集群分析的结果进一步说明了该模型管理身体重叠和遮挡情况的能力，凸显了该模型的稳健性。

方法。

作者将该模型的性能与其他团队进行了比较，结果表明，他们提出的方法在 AICity2023 挑战赛的 Track1 赛道上以 0.921 的 IDF1 分数取得了第五名的好成绩。这一比

较为他们提出的方法的有效性提供了佐证。

总之，本文成功地对提出的 MTMCT 框架进行了稳健的实验分析。通过对多个指标的严格评估、弱化研究以及与其他模型比较，充分证实了所提方法的有效性。为了进一步提高结果的可信度，我建议今后

实验可以探索在各种具有挑战性的条件下测试该框架的性能，如不同的光线、遮挡和人群密度。此外，考虑到保护隐私的人工智能日益受到关注，本文还可以讨论所提出的方法如何处理敏感信息并符合隐私法规。

## 4. 结论

综上所述，本文提出的 MTMCT 框架在许多方面都取得了显著成果。然而，其主要优势和潜在不足仍需在对整个工作进行全面评估时加以综合分析。

首先，本文的框架针对多摄像头跟踪问题提出了一系列创新方法，如分层聚类 and 细化策略，以及引入外观、时空和面部约束等。通过丰富的实验分析表明，这些方法能显著提高多摄像头人物跟踪的准确性和鲁棒性。

不过，在探讨这项工作的潜在隐患时，我注意到有许多方面仍有待进一步探索。例如，论文中提到的 ByteTrack 算法在密集场景中的性能以及不同目标大小的检测问题，都需要更多的实验验证和分析。

在未来的研究中，可以考虑从以下几个方面进一步深化。首先，可以探索更多的实验场景，包括光线变化、遮挡和人群密度等不同情况下的性能表现。此外，由于隐私保护问题日益受到关注，作者还可以考虑如何在提供有效跟踪的同时，确保敏感信息的处理和隐私法规的遵守。

此外，考虑到最新的技术进步，我建议在今后的研究中将新引入的 YOLOv8 模型纳入 MTMCT 框架。YOLOv8 在准确性和速度方面进行了优化，可能会为整个系统带来更好的性能。

总之，作者提出的新 MTMCT 框架确实具有重要价值，为该领域的研究提供了一些启示。不过，在方法选择的合理性、参数调整的探讨以及方法局限性的讨论等方面，仍有进一步改进的空间。

## 参考资料

- [1] 李宗义、王润生、李和、魏伯豪、史宇轩、凌海飞、陈家忠、刘博远、李中阳、郑汉清；《IEEE/CVF 计算机视觉与模式识别（CVPR）研讨会论文集》，2023 年，第 5519-5528 页