多目标跟踪算法综述

吴慧纯　黄锦涛　高铭鸿　郭文晖　许杰鸿　陈　军*
佛山大学　广东佛山　528000

摘　要：多目标跟踪（MOT）作为计算机视觉的核心任务，需在视频序列中持续定位目标并维护身份一致性。本文系统梳理MOT技术演进脉络，构建基于“运动建模→特征提取→匹配决策→联合建模”的四层演进分类体系，深入剖析各层代表性算法的核心思想。通过对比检测后跟踪（TBD）与联合检测跟踪（JDT）两大范式，揭示其设计哲学、适用场景及固有局限。最后，结合轻量化设计、无监督学习和多模态融合等趋势，展望未来突破方向。

关键词：多目标跟踪；检测－关联范式；运动模型；外观特征；端到端学习；算法对比

一、多目标跟踪技术的评估指标和常用数据集

（一）多目标跟踪技术评估指标和常用数据集合

多目标跟踪（MOT）算法的评估指标随着任务理解的深化和应用需求的提升而不断演化，主要分为经典度量、“CLEAR”度量、ID度量以及HOTA度量四大类。

多目标跟踪数据集从最初的室外行人监控（如MOT16）逐步扩展到无人机、车载、室内外混合、多模态等多种应用场景，呈现出“规模更大、场景更复杂、目标更细致、标注更精准”的发展趋势。以下是主要的多目标跟踪数据集介绍：

1. MOTChallenge系列（MOT16、MOT17、MOT20）

多目标追踪领域的标准评测平台——MOTChallenge。自2015年推出以来持续更新。Mot16和Mot17数据集包含14个视频序列。涵盖行人密集、遮挡严重等复杂场景，提供高质量的手工标注，适用于基准测试。MOT17提供了三种不同的检测结果（DPM、FasterR-CNN和SDP）。以评估跟踪器在不同检测输入下的鲁棒性。MOT20进一步增加了人群密度，极大考验了算法在极端拥挤场景中的表现。

2. UA-DETRAC

由中科院自动化所提出，是用于汽车交通监控的追踪数据集。涵盖多种天气与光照条件。其评估涵盖了检测与跟踪两个任务，共含140段视频、超过1.1万帧手动标注图像。

3. KITTITracking

KITTI是自动驾驶领域的重要数据集，其追踪任务集中在车辆、行人、自行车等街景环境下的目标。KITTI的数据来自于车载摄像头，真实反映自动驾驶场景，广泛用于自主导航领域的算法研究。

4. DanceTrack

这是一个专注于同类目标细粒度区分的新数据集，主要关注跳舞视频中多人的复杂互动。该数据集中的个体差异极小，强调算法对目标外观建模和时序一致性的要求。

5. VisDrone

该数据集由无人机视角采集，包含各种天气、时间、场景条件下的视频。目标包括人、车、自行车等多类对象，适用于评估算法在低视角和多类别场景中的表现。

6. BDD100KTracking

来自伯克利大学的数据集，专为自动驾驶系统设计，覆盖城市、高速公路等多种行车场景。其数据量之场景之多样，为深度学习模型提供了丰富的训练资源。

二、多目标跟踪技术发展概述

多目标跟踪技术的发展呈现出任务整合范式与技术演进路径的双重主线。在任务整合层面，依据检测与跟踪任务的耦合程度，它主要分为两种范式：检测跟踪（TBD）和联合检测跟踪（JDT）。在技术演进层面，方法创新沿“运动建模→特征提取→匹配决策→联合建模”

基金项目：本文系佛山大学学生学术基金一般项目《基于视频的人体姿态捕获和追踪方法研究》（项目编号：xsjj202407zrb02）。

作者简介：
1.吴慧纯（2003--），女，汉族，广东揭阳人，本科，研究方向：数字媒体技术方向。
2.陈军（1986--），男，汉族，河南信阳人，博士，副教授，研究方向：信息与通信工程，为本文的通讯作者。

脉络递进，反映了MOT领域“从局部到全局、从规则到学习”的发展趋势。TBD范式覆盖技术演进的前三层（运动→外观→匹配），体现分步优化的设计哲学；JDT范式对应第四层（端到端架构），标志全局融合的技术跃迁，驱动MOT从手工规则迈向数据驱动。

（一）运动建模层：传统滤波改进

运动建模层聚焦目标物理运动规律（位置、速度）的数学表征，核心是解决线性运动预测与非线性适应的矛盾。最近邻数据关联（NNDA），通过跟踪门（椭圆/矩形）筛选最近量测点，奠定基于位置的距离关联基础。1990年Kalman滤波成为主流运动模型，但无法处理遮挡导致的非线性漂移。

（二）特征提取层：外观特征优化

特征提取层以增强目标判别性为核心任务，通过优化深度特征提取应对形变、遮挡导致的身份混淆问题。相较于运动建模层对物理规律的依赖，本层聚焦语义级特征表达，是应对复杂场景的必要升级。其技术演进始于手工特征时代（如SIFT/HOG），2015年Google提出的TripletLoss推动了监督式ReID特征学习，但检测与特征提取分离的架构导致误差累积。核心挑战在于平衡特征判别力与计算效率，并克服光照变化、极端形变等干扰因素。

（三）匹配决策层：关联策略创新

匹配决策层聚焦多目标关系建模，通过优化关联逻辑解决复杂场景（拥挤、运动模糊）中的匹配歧义问题。相较于前两层的单目标特征处理，本层实现从个体到关系的技术升级，代表多目标跟踪（MOT）范式的技术巅峰。其核心挑战在于当目标外观特征失效或运动轨迹冲突时，如何利用拓扑结构、匹配机制与全局优化策略实现鲁棒关联。传统匈牙利算法等人工规则在密集场景中面临严重性能衰减，需建立更智能的决策体系。

1.典型算法演进

为应对复杂场景下的多目标关联挑战，现代匹配决策研究聚焦三大创新技术路径，显著提升了匹配系统的容错性与泛化能力。

（1）拓扑关系融合

通过图结构建模目标间空间依赖关系：

GSM采用GNN加入拓扑关系，对每个顶点都构建一个有向图，顶点表示外观特征，边表示位置关系特征，随后计算两间每个目标的图相似度进而进行匹配。

SGT提出稀疏图恢复机制，采用GNN聚合邻域信息修复低置信检测，顶点代表检测，边代表两个检测之间的相似度（IoU、拓扑关系、外观三个维度衡量），把两个检测是否属于同一目标，转换成边分类的问题。

（2）匹配流程优化

设计鲁棒关联机制应对检测缺陷：

ByteTrack提出双阈值关联策略，优先匹配高置信检测，再利用低置信检测与未匹配轨迹二次关联，有效避免了因直接丢弃低置信度检测框导致的漏检问题，显著降低了误检数量。

OMC通过记忆分支融合历史外观特征预测漏检目标。将记忆分支生成的高置信度预测结果与检测器输出进行互补性关联，若某预测目标未出现在检测结果中且置信度超过预设阈值，则判定为有效漏检恢复，显著提升了目标关联鲁棒性。

2.关联性与局限

匹配决策层呈现清晰的智能升级路径：决策机制从人工规则（匈牙利算法）演进至图结构建模（如GSM的空间拓扑构建、SGT的稀疏图恢复），最终发展为端到端矩阵学习（如TransMOT的GraphTransformer）；优化维度由单帧匹配（ByteTrack双阈值策略）扩展至时空拓扑融合（SGT邻域聚合），进而实现全局指标优化（DeepMOT可微损失）；问题解决深度从显性冲突处理（OMC漏检预测）深化至隐式关系挖掘（GSM图相似度计算），并向预测性关联演进。其本质在于通过多目标协同推理增强系统容错能力：拓扑融合提升空间鲁棒性（如GSM在MOT20拥挤场景提升IDF15.1%），流程优化补偿检测缺陷（如ByteTrack减少MOT17的FP23.6%），矩阵学习实现全局最优匹配（如TransMOT在DanceTrack提升HOTA12.5%）。

（四）联合建模：统一检测与跟踪

联合建模层通过统一架构整合检测、特征提取与关联流程，实现端到端的跟踪全局优化，标志着多目标跟踪从“分步流水线”向“协同学习”的范式跃迁。其核心突破在于解决传统模块化流程的误差累积问题——检测偏差会逐级传播至关联阶段。2020年Carion提出的DETR为端到端检测提供新范式，而跟踪领域进一步引入时序建模能力形成两大核心技术路径。

1.典型算法演进

（1）基于Transformer的时序建模

TrackFormer创新性地引入了“跟踪查询”的概念，模型通过注意力机制，将这些查询向量（包含历史轨迹和

潜在新检测）与当前帧的图像特征进行交互，同步输出更新的轨迹（包含检测框和轨迹ID）及新产生的检测。

（2）图结构联合优化

GSDT融合图嵌入与半监督学习，将过去轨迹的特征与当前帧的特征图进行图神经网络传播，以同时捕获目标的外观特征和几何位置信息。通过构建并传播目标间的几何关系，辅以稀疏重构与图正则化技术来增强特征的判别力。

Chained-Tracker采用链式注意力结构，以边界框对为回归单元跨帧连接检测，实现轻量级联合推理。

（3）运动引导联合学习

CenterTrack以中心点偏移为跟踪线索，输入双帧中心点热图，端到端预测目标运动关联，直接预测目标在当前帧相对于上一帧中心点的偏移量，显著降低复杂运动匹配歧义，简化了目标关联过程

2.关联性与局限

联合建模层呈现“分步优化→架构统一→全局协同”的深度演进脉络：从模块化流水线（SORT级联结构）演进至Transformer统一架构（如TrackFormer的跟踪查询机制），最终发展为图结构联合优化（GSDT的图嵌入传播）；其信息流由单向误差传递升级为多模态特征跨层交互（TransCenter的运动-外观联合推理）；优化目标从局部最优匹配深化至全局轨迹一致性最大化（Chained-Tracker链式注意力回归）。其本质是通过深度网络的协同推理机制替代人工规则，例如TrackFormer的注意力权重自动学习检测-轨迹关联（MOT17上ID切换降低37%），GSDT的图卷积隐式建模拓扑约束（Market-1501mAP85.3%）。该范式在ID保持与长时跟踪方面展现显著优势。

三、总结与展望

本研究系统梳理了基于视频的多目标跟踪（MOT）技术发展脉络，围绕“检测—关联—优化”的核心框架，深入剖析了从传统运动模型优化（如Kalman滤波及其改进型OCSORT、BoT-SORT）到深度特征提取（如FairMOT、QDTrack的联合学习与对比学习），再到复杂匹配决策（如ByteTrack的双阈值流程、GSM的图结构建模、TransMOT的关联矩阵学习）的技术演进路径，并最终聚焦于端到端联合建模（如TrackFormer、GSDT）这一前沿范式。研究揭示了检测后跟踪（TBD）与联合检测跟踪（JDT）两大范式的本质差异、性能边界及各自适用场景：TBD灵活高效但易受误差传播影响，JDT鲁棒性强却计算成本高昂、依赖标注。通过整合并对比分析经典指标（MT/ML/IDS、MOTA）、ID一致性指标（IDF1）以及新兴的HOTA指标，研究为全面评估算法性能提供了多维视角，并首次系统性地指明了轻量化设计、无监督学习和多模态融合等关键未来方向。

面向未来，多目标跟踪研究需在提升算法鲁棒性、保证实时效率、增强场景泛化能力三者间寻求更优平衡。突破点包括：发展轻量化模型设计技术（如模型蒸馏、神经架构搜索）以降低JDT范式的计算负载，使其更适用于边缘设备；探索无监督或弱监督学习范式（如时序自监督、检测框生成伪轨迹标签），以显著减少对大规模精细标注数据的依赖；融合多模态（如红外、深度）与多传感器（如雷达、LiDAR）信息，以应对夜间、极端遮挡等复杂感知挑战并提升自动驾驶等场景的可靠性；加强长时序关联与记忆机制（如引入神经符号系统、时空图神经网络），有效解决超长遮挡下的身份恢复难题；同时，提升端到端模型的可解释性与对抗鲁棒性，确保算法决策的透明可靠和在极端环境下的稳定表现。通过这些努力，MOT技术有望在智能驾驶、人机交互、视频分析等领域释放更大价值。

参考文献

[1]李天成，谢昱昕，李固冲，等.面向多目标跟踪的数据关联方法研究综述[J].雷达科学与技术，2025，23（01）：10-31.

[2]刘旭东.多目标跟踪算法研究综述[J].吉林省经济管理干部学院学报，2013，27（04）：63-65.

[3]沈锦荣.基于Transformer的跨空间特征关联多目标追踪[J].软件导刊，2024，23（03）：165-171.