

基于踪片 Tracklet 关联的视觉目标跟踪: 现状与展望

刘雅婷^{1,2,3} 王坤峰^{1,3} 王飞跃^{1,4}

摘要 近年来, 由于计算机视觉技术的发展和计算机硬件性能的提高, 基于视觉的目标跟踪方法得到了飞速的发展. 其中, 基于踪片 (Tracklet) 关联的目标跟踪方法因为具有对目标遮挡的强鲁棒性、算法运行的快速性等优点得到了广泛关注, 本文对这类方法的最新研究进展进行了综述. 首先, 简明地介绍了视觉目标跟踪的基本知识、研究意义和研究现状. 然后, 通过感兴趣目标检测、跟踪特征提取、踪片生成、踪片关联与补全四个步骤, 系统详尽地介绍了基于踪片关联的目标跟踪方法, 分析了近年来提出的一些踪片关联方法的优缺点. 最后, 本文指出了该研究问题的发展方向, 一方面要提出更先进的目标跟踪模型, 另一方面要采用平行视觉方法进行虚实互动的模型学习与评估.

关键词 视觉目标跟踪, 踪片关联, 网络流, 马尔科夫随机场, 平行视觉

引用格式 刘雅婷, 王坤峰, 王飞跃. 基于踪片 Tracklet 关联的视觉目标跟踪: 现状与展望. 自动化学报, 2017, 43(11): 1869–1885

DOI 10.16383/j.aas.2017.c170117

Tracklet Association-based Visual Object Tracking: The State of the Art and Beyond

LIU Ya-Ting^{1,2,3} WANG Kun-Feng^{1,3} WANG Fei-Yue^{1,4}

Abstract In the past decade, benefitting from the progress in computer vision theories and computing resources, there has been a rapid development in visual object tracking. Among all the methods, the tracklet-based object tracking method has gained its popularity due to its robustness in occlusion scenarios and high computational efficiency. This paper present a comprehensive survey of research methods related to tracklet-based object tracking. First, the basic concepts, research significance and research status of visual object tracking are introduced briefly. Then, the tracklet-based tracking approach is described from four aspects, including object detection, feature extraction, tracklet generation, and tracklet association and completion. Afterwards, we propose a detailed review and analyze the characteristics of state-of-the-art tracklet-based tracking methods. Finally, potential challenges and research fields are discussed. In our opinion, more advanced object tracking models should be proposed and the parallel vision approach should be adopted to learn and evaluate tracking models in a virtual-real interactive way.

Key words Visual object tracking, tracklet association, network flow, Markov random field, parallel vision

Citation Liu Ya-Ting, Wang Kun-Feng, Wang Fei-Yue. Tracklet association-based visual object tracking: the state of the art and beyond. *Acta Automatica Sinica*, 2017, 43(11): 1869–1885

视觉目标跟踪是指利用目标的颜色、纹理等视觉信息以及运动信息, 确定视频数据中感兴趣目标

的位置、速度等信息, 并将相邻图像帧的相同目标进行关联, 实现对目标的位置预测和持续追踪, 以便完成更高级的任务. 视觉目标跟踪不仅可以获得目标的运动状态和运动轨迹, 也为运动分析、场景理解、行为或事件监测提供先验知识. 它融合了模式识别、人工智能、图像处理等多个学科, 在智能监控、人机交互、视觉导航、军事指导以及医疗诊断等领域有着广泛的应用^[1–3].

由于视觉跟踪技术具有广阔的市场前景和理论价值, 国内外很多大学和科研机构都开展了相关理论研究. 国外研究启动相对较早, 牛津大学动态视觉研究组针对视觉目标跟踪展开了大量研究, 包括灵活目标跟踪、对抗伪装, 并应用到交通监控、安保等领域; 加利福尼亚大学视觉研究实验室 (VRL) 展开了摄像机网络中的行人跟踪研究^[4];

收稿日期 2017-03-04 录用日期 2017-08-18
Manuscript received March 4, 2017; accepted August 18, 2017
国家自然科学基金 (61533019, 71232006, 91520301) 资助
Supported by National Natural Science Foundation of China (61533019, 71232006, 91520301)

本文责任编辑 张军平
Recommended by Associate Editor ZHANG Jun-Ping
1. 中国科学院自动化研究所复杂系统管理与控制国家重点实验室 北京 100190 2. 中国科学院大学 北京 100049 3. 青岛智能产业技术研究院 青岛 266000 4. 国防科技大学军事计算实验与平行系统技术研究中心 长沙 410073
1. The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190 2. University of Chinese Academy of Sciences, Beijing 100049 3. Qingdao Academy of Intelligent Industries, Qingdao 266000 4. Research Center for Computational Experiments and Parallel Systems Technology, National University of Defense Technology, Changsha 410073

诺丁汉大学计算机视觉实验室 (CVL) 展开使用视频中语义信息进行人或人群跟踪研究; 瑞士联邦理工学院 (ETH, Zurich) 计算机视觉实验室开展了动态场景自动驾驶中的目标跟踪研究^[5], 将视觉跟踪技术与机器人技术相结合; 南加州大学计算机视觉实验室研究无约束环境中的视觉跟踪问题, 并提出基于语境的跟踪方法^[6]; NEC 实验室研究视觉监控场景中的多人跟踪问题, 以期满足实时性要求^[7]; 卡耐基梅隆大学机器人研究所计算机视觉小组则针对机器人可能遇到的环境约束中的视觉跟踪问题进行了大量研究^[8]. 美国国防部高级研究计划局 (Defense Advanced Research Projects Agency, DARPA) 开展了重大视频监控项目 VSAM (Visual Surveillance and Monitoring) 并产生了先进的成果^[9-10], 在国内, 视觉跟踪的研究也逐渐取得一系列成果, 许多高校和科研单位在视觉跟踪理论方面进行了深入研究. 早在 2001 年, 清华大学运用相关目标识别和跟踪技术开发了一套适用于野外环境的视觉侦查系统; 中科院自动化所在行人视觉分析, 交通场景与行为事件理解、视觉监控等领域也取得了科研成果. 近年来, 以深度学习为代表的机器学习热潮再次掀起, 激发越来越多的企业与科研机构投入到视觉目标跟踪领域.

随着目标检测算法的成熟、检测准确度的提高, 越来越多的研究者^[1, 11-13] 采用 Tracking-by-detection 思路进行研究, 通过提取感兴趣目标的目标区域, 再运用生成模型或者训练分类器得到跟踪轨迹. 进一步地利用该轨迹来解决目标之间、目标与背景遮挡的问题, 并且融入先验知识提升跟踪精度. 本文综述了一类 Tracking-by-detection 方法——基于踪片 (Tracklet) 关联的目标跟踪方法. 需要指出, 目前对英文专业术语 Tracklet 的中文翻译不统一, 常见的译名有轨迹片段、短轨迹、踪迹片段等, 都不够简洁达意; 本文将 Tracklet 翻译为“踪片”, 言简意赅. 基于踪片关联的目标跟踪方法依据目标检测的结果, 找到目标能被稳定检测的视频帧, 将其置信度较高的位置进行关联形成踪片, 再将不同的踪片进一步关联, 形成最终的完整轨迹. 当目标发生遮挡、重叠时, 通过填补成功关联踪片之间的空缺能够得到完整的轨迹集合; 当目标再次进入视野时, 提取其特征与之前的轨迹进行匹配, 实现目标的稳定关联, 提高跟踪的鲁棒性.

除上述优点外, 基于踪片关联的目标跟踪还可用在不同的跟踪情形中. 目标跟踪情形可按如下方式分类: 按照跟踪目标的个数可划分为单目标跟踪和多目标跟踪; 按照跟踪目标的类型分为刚体跟踪

与非刚体跟踪; 按照摄像头数量分为单摄像头与多摄像头跟踪; 按照摄像头的运动状态分为静止摄像头与运动摄像头跟踪; 按照应用场景分为单场景和多场景目标跟踪^[18]. 在以上不同的情况中, 基于踪片关联的跟踪都能根据相应的目标检测结果进行关联形成对应的踪片, 并填补空缺形成完整轨迹, 因此有较强的普适性.

本文其他部分内容安排如下: 第 1 节介绍基于踪片关联的视觉目标跟踪算法流程; 第 2 节总结目标跟踪常用的公开数据集; 第 3 节详细阐述基于踪片关联的视觉目标跟踪研究进展, 包括算法介绍、算法的优缺点分析、在公共数据集上的测试结果比较; 第 4 节分析了现有的目标跟踪方法的优点和局限性, 并对该研究领域的发展趋势做出展望; 第 5 节总结全文.

1 基于踪片关联的视觉目标跟踪算法流程

基于踪片关联的目标跟踪算法主要包括感兴趣目标检测、跟踪特征提取、踪片生成、踪片关联四个步骤, 基本流程如图 1 所示. 输入视频序列, 首先通过目标检测获得感兴趣目标的位置等特征, 并将相关检测结果进行分析, 提取出恰当的跟踪特征后关联形成踪片, 通过图论等数学方法将踪片进一步关联形成长轨迹, 通过轨迹补全、轨迹校正等后处理方法填补轨迹空缺, 进行轨迹平滑, 校正轨迹关联错误, 从而得到最终的输出轨迹.

1.1 感兴趣目标检测

基于踪片关联的目标跟踪方法以目标检测的结果为前提, 继而进行关联直到形成最终的跟踪轨迹, 因此目标检测是该方法的基础. 目标检测是指依据一定的算法和先验信息把图像中的前景目标从背景中提取出来. 在关联过程中, 只对需要关注的目标而不是所有目标进行实时检测的特定目标检测方法被称为感兴趣目标检测. 然而物体运动复杂多变, 感兴趣目标在视频序列中可能会出现暂时离开场景或者被遮挡的情况; 目标与背景外观等特征较为相似, 前景和背景区分不准确; 受天气、光照等外界条件以及背景自身内部因素影响, 图像中的背景具有复杂性和动态变化性, 这些都增加了提取感兴趣目标的难度.

近年来, 机器学习特别是深度学习技术在目标检测领域获得了广泛应用, 越来越多的研究者使用人工神经网络^[10, 19] (CNN、RCNN、Fast RCNN、GAN 等)、支持向量机^[20]、Adaboost^[21] 等方法训练分类器实现前景与背景的分隔. 这些方法通常首先选定样本 (包括正、负样本), 将所有样本分

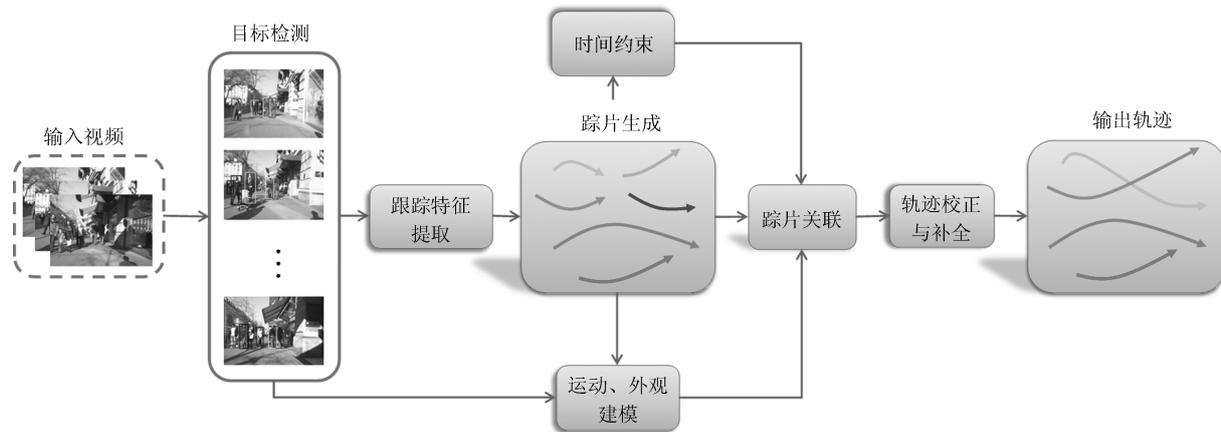


图 1 基于踪片关联的视觉目标跟踪方法流程图

Fig. 1 Flowchart of visual object tracking based on tracklet association

成训练集和测试集两部分. 在训练集上运用机器学习算法训练分类器, 生成分类模型; 在测试集上运用该模型生成预测结果; 最后用相应的评估指标评估分类器性能. 机器学习方法能够克服背景扰动, 处理目标运动复杂的场景, 抗干扰性强. 通过以上方法实现了感兴趣目标的检测和定位, 为踪片关联打下基础.

1.2 跟踪特征提取

对视频图像进行逐帧检测并得到感兴趣目标的检测结果后, 该方法需要从结果中提取恰当的特征从而形成可靠稳定的踪片, 以便实现踪片的准确关联, 提高目标跟踪精度. 目标跟踪中常用到的目标特性表达主要包括视觉特征、统计特征、代数特征等. 视觉特征包括图像边缘、轮廓、区域、纹理; 统计特征包括直方图、矩; 代数特征如图像矩阵的奇异值分解.

在视觉特征中, 沿着边缘方向移动, 像素变化较缓慢, 而垂直于边缘方向移动, 像素变化很剧烈为边缘特征, 可以采用梯度、Sobel、Roberts^[22]等梯度算子以及卷积神经网络^[23]提取. 将检测到的像素不连续的部分连接成完整边界就形成了轮廓. 轮廓跟踪^[24]利用封闭的曲线轮廓表示运动目标, 并且能够实时更新轮廓位置, 去除了背景像素, 对非刚体与其他轨迹复杂的运动有良好的跟踪效果. 区域特征不仅包括运动目标, 也包括部分背景区域, 通常用矩形或者椭圆形框表示. 区域特征跟踪^[25]对无遮挡目标的跟踪精度高, 但其计算复杂度高, 对目标有遮挡时跟踪效果差. 区域特征提取的方法包括区域生长法^[26]、区域分裂与聚合^[27]、阈值法^[28]. 图像纹理^[29]通过图像的颜色、光强信息描述, 提取方法有结构建模法和统计数据法, 研究者需根据场景不同来选择适合的纹理识别方式.

在统计特征中, 用直方图^[30]描述图像的灰度、HOG、HOF等特征信息, 帮助分析图片中曝光水平, 粗略描绘出目标区域颜色分布, 计算效率高. 针对一幅图像, 若用二维随机变量表示像素位置, 则可以用二维灰度密度函数表示灰度图像, 或用矩描述灰度图像的特征^[31-34].

代数特征^[35]是将图像看作矩阵, 运用代数方法得到空间表征能力强的特征向量作为图像特征. 相关代数方法包括奇异值分解^[36]、主成分分析^[37]、独立成分分析^[38]等.

除了上述特征提取方法外, 近年来 SIFT 算子^[39]、卷积神经网络^[40-42]等特征提取方法得到了广泛的应用, 获得了较好的特征提取效果. 另外, 还可以通过融合多种特征代替单一特征的方式来提高特征提取的鲁棒性和精确性.

1.3 踪片生成

提取到感兴趣目标的特征后, 进一步将检测结果进行关联生成踪片. 该过程通常需要牺牲连接片段的长度来生成置信度较高的踪片确保已关联的片段准确, 称为初级关联. 目标在连续两帧图像中变化缓慢, 相邻帧之间目标的尺寸、运动状态、外表形态等特征变化不大, 逐帧进行关联具有高的可靠性. 具体地, 对相邻帧的检测结果可以提取位置、速度、外观属性计算相似性, 并设置阈值来进行匹配, 相似度高高于阈值则认为属于同一目标. 形成初级关联的踪片初步关联的相似性可以表示为:

$$P_{associate}(f_1, f_2) = A_{position}(f_1, f_2) \times A_{appearance}(f_1, f_2) A_{velocity}(f_1, f_2) \quad (1)$$

其中, f_1 、 f_2 表示待比较的两帧, $P_{associate}(\cdot)$ 表示相邻帧中待关联目标的相似性, $A_{appearance}(\cdot)$ 表示外观相似性, $A_{velocity}(\cdot)$ 表示速度相似性. 其中外观

相似性可以通过面积和颜色的相似性衡量, 由于相邻两帧之间目标移动缓慢, 研究者通常将该过程看做线性运动, 在保证关联准确的情况下尽可能简化计算. 位置和面积相似性可以采用高斯核函数的方式来计算; 颜色的相似性计算方式如下: 首先计算每帧目标的颜色直方图, 接着计算相邻帧之间的距离 (Hellinger 距离^[43]、巴氏距离^[44] 等) 作为相似性的衡量标准. 通过以上所述的方法构建不同帧中待关联检测的相似性, 将踪片生成的过程进行量化表示.

决定踪片生成数量和长度的因素是阈值的选取, 以实现漏检率与错误率的折中. 阈值选取较大值时, 生成的踪片数量较少, 关联精确度高, 但也增加了漏检的可能性; 阈值取值较小时, 生成的踪片数量较多, 但也可能将不属于同一目标的片段错误关联起来. 由于在初级关联时要保证生成的踪片足够精确, 所以此时往往选较大阈值.

将所有检测结果按照以上准则计算相似度并进行关联, 最终得到待跟踪视频初级关联结果.

1.4 踪片关联

形成踪片后, 对其进行高层关联形成长轨迹. 该过程也可视为踪片间的匹配问题, 即如何进行踪片匹配使得关联后的轨迹具有更高的可靠性、稳定性以及鲁棒性. 获得踪片之间相似性需要综合考虑时间关系、外观以及运动等特征, 从而保证踪片关联的精度和完整性. 该步骤是基于踪片跟踪的关键, 有效的踪片关联算法能够大幅提高跟踪精度. 这里总结了踪片关联的一般方法.

轨迹关联在时间上满足以下两种约束:

1) 同一目标在同一时间不可能出现在多于一条运动轨迹;

2) 同一运动轨迹不可能同时属于多个目标.

由上面的结论可知, 时间上重叠的踪片一定不属于同一目标, 可公式化表示为:

$$P_t(T_i, T_j) = \begin{cases} 1, & \text{if } f_{j_s} - f_{i_e} > 0 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

其中 P_t 表示 T_i 与 T_j 关联的可能性, T_i 与 T_j 表示两条待匹配片段, 且 T_i 出现时间早于 T_j . f_{j_s} 为 T_j 的初始帧号, f_{i_e} 为 T_i 的结尾帧号.

与踪片形成阶段相似, 踪片关联也要综合考虑片段的颜色、纹理、面积等因素, 从而判断待关联的踪片是否属于同一目标. 外观相似性模型按照式 (3) 建立, 其中 $P_{app}(\cdot)$ 表示相似度, A_{T_i} 和 A_{T_j} 表示轨迹 T_i 和 T_j 的外观 (颜色、面积、纹理等) 约束.

$$P_{app}(T_i, T_j) = \text{corr}(A_{T_i}, A_{T_j}) \quad (3)$$

对运动特征, 基于目标运动轨迹连续的原则, 时

间差与目标移动距离之间有着相关关系. 通过对前一个踪片的尾部帧和当前踪片的起始帧进行运动相似性匹配确定关联情况. 运动模型可以按照如下方程建立:

$$P_{mo}(T_i, T_j) = \frac{\text{corr}(P_{T_i}^e + V_{T_i}^e \Delta t, P_{T_j}^s)}{\text{corr}(P_{T_j}^s - V_{T_j}^s \Delta t, P_{T_i}^e)} \quad (4)$$

其中, $P_{T_i}^e$ 表示踪片 T_i 的结束位置, $P_{T_j}^s$ 表示踪片 T_j 起始位置, $V_{T_i}^e$ 表示 T_i 结束时刻的速度, $V_{T_j}^s$ 表示 T_j 起始时刻速度, Δt 为 T_i 结束时刻与 T_j 起始时刻的时间差. 由于候选关联踪片之间时间间隔较短, 目标速度在该短时间内可看作是恒定的. 因此该过程可以按照如下步骤进行: 提取前一条踪片结束时刻的位置与速度, 通过线性预测方法预测其经过 Δt 时间间隔后的位置, 并与后一条轨迹的起始位置进行比较, 计算相关性; 另外, 将后一轨迹中的起始时刻按照同样的方法进行倒推, 得到 Δt 时刻之前的状态, 并与前一条轨迹的结束位置进行比较, 得到位置的相关性. 如图 2 所示.

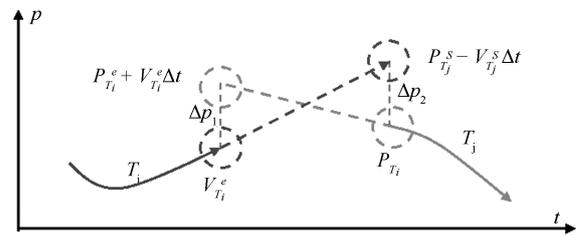


图 2 位置相关性示意图

Fig. 2 Sketch map of position relations

最后将时间、外观和运动相似性模型结合起来计算两个踪片的关联概率表示为:

$$P_{ass}(T_i, T_j) = P_t(T_i, T_j) P_{app}(T_i, T_j) P_{mo}(T_i, T_j) \quad (5)$$

判断关联过程中轨迹是否生成或终止可以采用如下判定依据:

1) 当前帧与前一帧进行匹配计算相似度小于阈值, 认为当前帧出现了新目标, 生成新轨迹;

2) 当前帧与后一帧进行匹配计算相似度小于阈值, 认为当前帧的目标轨迹已终止.

根据以上方法可以得到踪片之间的关联概率. 对比不同踪片的关联概率, 并将概率值最大的踪片关联起来, 可以获得目标相对较长的运动轨迹. 最后, 由于交叉重叠等因素影响, 获得的长轨迹还需要通过插值法^[45] 等进一步连接, 从而形成平滑完整的轨迹, 最终实现目标轨迹跟踪.

2 目标跟踪的公共数据集

为了方便研究者进行目标跟踪实验以及评估实

验结果, 促进目标跟踪领域的发展, 学术界建立了部分开放的公共数据集. 这些数据集由不同的场景、光照、天气、视角、采集而来, 包含行人、车辆等各种要素以及不同要素相互遮挡、轨迹重叠、离开以及重回视野等复杂的运动模式. 将算法在这些数据集上运行, 对跟踪结果与已有的基准进行比较, 能够全面地反映算法的性能, 客观地评价算法的优缺点. 本文将数据集划分为实际数据集和虚拟数据集两种类型分别介绍. 常见的数据集名称及其特点如下表 1 所示.

实际数据集由实际场景中采集到并通过人工方式被标记, 传统数据集一般都属于实际数据集, 如表 1 所示. 但是这种获取数据的方法不仅成本昂贵, 而且在复杂天气条件或是低照度情况下人工标注准确率也难以得到保证. 此外, 受到实际条件约束, 实际数据集无法模拟如极端恶劣天气、目标复杂运动等不常见的情景, 获取的数据集规模也受制约, 这些因素都促使了研究者开展人工场景研究. 近年来, 游戏引擎、虚拟现实技术的发展也进一步推动了虚拟数据集的建立, 表 1 中所示的 Virtual KITTI 和 SYNTHIA 已成为常见的虚拟数据集. 这些数据集利用计算机图形学等综合性生成复杂多样的、动态的、可自动标注的虚拟场景, 从而实现逼真地模拟各种复杂挑战的实际场景. 相关实验^[46]已经表明: 经过真实环境训练的跟踪方法在虚拟数据集和真实数据集上有相同的表现程度, 并且在虚拟数据上进行

预训练能够提高目标跟踪性能.

3 基于踪片关联的视觉目标跟踪进展

近年来, 基于踪片关联的跟踪引起了研究者的广泛关注, 取得了一定的研究进展. 解决踪片跟踪问题的关键是对生成的踪片进行准确关联, 从而形成可靠完整的轨迹. 本文将基于踪片关联的跟踪方法分为图论方法和其他方法, 并具体介绍部分代表性成果.

3.1 图论方法

得到目标检测结果并进行初级关联形成踪片后, 可以利用图论知识建立匹配模型. 概率图可以具体的图论方法有: 贝叶斯网络 (Bayesian network)、条件随机场 (Condition random field, CRF)、马尔科夫随机场 (Markov random field, MRF) 等概率图模型以及网络流 (Network flow, NF)、二分图匹配 (Bipartite graph match) 等模型.

3.1.1 概率图模型

1) 贝叶斯网络

Huang 等^[55]首次提出了基于检测的三层次关联方法, 以解决单摄像机、嘈杂环境下的多目标跟踪问题. 在低层次的关联中, 通过极大化连接相似性的约束产生可靠的轨迹, 该阶段只连接相邻帧的检测结果, 并且用双阈值的方式抑制错误的连接; 在中层次的关联中, 从低层次获得的踪片被迭代地输入, 通

表 1 多目标跟踪常见的公共数据集

Table 1 Frequently used public datasets for multi-target tracking research

| 数据集 | 建立时间 | 描述 | 规模 | 类型 |
|-------------------------------|--------|--|------------------------------|-------|
| PETS ^[47] | 2009 年 | 拥挤的公共区域多传感器跟踪和事件识别 | 3 个不同环境视频序列 8 个视角 | 实际数据集 |
| MOT challenge ^[48] | 2015 年 | 不仅标记了行人, 车辆、静态的人、遮挡物体等都被标注 | 22 个视频序列, 共 11 286 帧图像 | 实际数据集 |
| CAVIAR ^[49] | 2003 年 | 行人会面、购物, 穿越拥挤人群及在公共场所遗失行李等复杂场景 | 28 段视频 | 实际数据集 |
| i-LIDS ^[50] | 2006 年 | 多摄像机配置, 可以选择多视角的数据进行实验 | 10 小时视频 | 实际数据集 |
| UA-DETRAC ^[51] | 2015 年 | 多个数据采集地; 涉及汽车、公共汽车、货车等多种车辆; 包含多云、夜晚、晴天和下雨等天气条件 | 10 小时视频 | 实际数据集 |
| 文献 [52] 中的数据集 | 2014 年 | 从拥挤繁忙的火车站采集 | 42 million 的轨迹 | 实际数据集 |
| KITTI ^[53] | 2012 年 | 每幅图像多达 15 辆车和 30 个行人; 包含三维立体, 光流, 视觉光度法, 3D 物体检测和 3D 跟踪 | 50 个视频序列 | 实际数据集 |
| Virtual KITTI ^[46] | 2016 年 | 数据从不同的成像和天气条件下的五个虚拟世界生成. 有准确, 完整的 2D 和 3D 多对象跟踪注释, 并有像素级别和实例级别标签, 以及深度标签 | 50 个高分辨率单目视频, 共 21 260 帧 | 虚拟数据集 |
| SYNTHIA ^[54] | 2016 年 | 多样化的场景; 多种动态物体种类; 多季节; 不同的照明条件和天气情况; 多传感器多视角 | 2 分 23 秒雪景及 1 分 48 秒傍晚车载视频序列 | 虚拟数据集 |

过复杂的相似性测量方法将上述的踪片进行关联形成长轨迹, 关联过程被看作最大后验概率^[56-57]问题, 其不仅考虑轨迹片的初始化, 终止和 ID 转换, 还考虑踪片的误报警等. 在高层次的关联中, 文章基于前一级得到的踪片估计出一个新的场景结构模型, 它能有效地建模目标进入、退出和场景遮挡问题. 借助于基于场景知识的推理执行长轨迹关联, 以减少轨迹分割并防止 ID 转换. 该方法通过有效地将踪片与不准确的检测响应和长时间遮挡相结合, 显著地改善了跟踪性能. 文章提出的这种分层框架是一种通用的方法, 其他相似性度量或优化方法可以很容易地集成到这个框架中.

2) 条件随机场模型

在现有的主流目标跟踪工作中, 外观模型是预先定义的或通过在线学习的方式得到. 虽然多数情况下这种方法能够有效区分目标, 但当目标具有相似外观并且在空间上接近时该方法将会失效. 运动模型, 线性运动模型目前也被广泛使用. 轨迹之间的关联概率通常基于满足线性运动假设的程度, 即假定目标以恒速度沿着原方向运动. 然而, 如果目标不遵循线性运动模型或是相机运动造成视角变化, 利用线性假设估计踪片之间的关联性会出现很大偏差. 在线学习条件随机场 (CRF) 模型能够在相机运动下可靠地跟踪目标, 并提高不同目标的区分度, 特别是在空间上接近并且具有相似外观的难区分目标. 因此条件随机场模型也广泛用在踪片关联中.

Yang 等^[58] 提出了一种条件随机场模型在线学习方法. 该方法主要分为 CRF 创建、一元项学习、二元项学习、最小化能量函数得到踪片关联四步: 首先寻找首尾间隔满足一定阈值条件的踪片对, 作为 CRF 节点. 然后基于运动模型以及外观辨别模型定义了一元项和二元项能量函数, 分别用于区分踪片之间的关联程度以及邻近的踪片对之间的关联程度. 其中运动模型的一元项由踪片线性运动模型所得的估计位置之间的差别定义, 二元项则由踪片对的尾部位置相近 (Tail-close) 或者头部位置相近 (Head-close) 的关系得到, 如图 3 所示; 外观模型的一元项与二元项则通过选取颜色、纹理、形状等特征, 采用在线学习外观区分模型 (OLDAMs) 以得到正负学习样本, 最后使用 RealBoost^[59] 算法学习得到最终外观模型. 通过最小化总的能量函数即可得到踪片关联. 该方法利用 CRF 一阶和二阶能量项提高了算法的鲁棒性, 时间复杂度为指数级别. 该算法保证了良好的快速性, 并在多个公共数据集实验结果中的多个性能指标中表现良好.

3) 马尔科夫随机场

Wu 等^[60] 将人脸聚类 and 跟踪结合起来, 用以同时提升人脸识别与轨迹跟踪问题. 该方法通过两个

问题相互提供有用的信息以及约束条件, 提高彼此的性能. 文章通过隐马尔科夫随机场模型将人脸聚类标签和人脸轨迹跟踪结合起来, 转化为贝叶斯推理问题, 提出了有效的坐标下降解法. 输入一个视频序列, 利用 Viola-Jones 脸部检测方法^[61] 来产生可靠的检测结果, 通过外观、边框位置、尺度等将检测结果关联起来形成踪片. 为避免身份转换, 作者对匹配分数设置了阈值. 文章基于隐马尔科夫随机场模型表示聚类标签和轨迹连接关系的联合依赖, 提出了同时聚类和关联长视频序列中不同人类的面孔. 该方法不仅减少了由于关联不同聚类标签踪片而导致的错误, 而且在同一目标的长追踪轨迹中进行聚类能够极大增强聚类准确性.

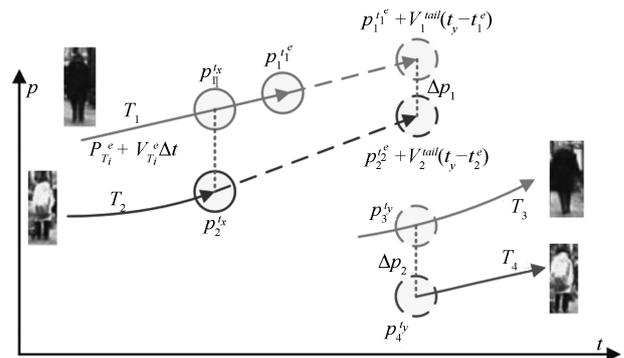


图 3 踪片对运动相似性估计^[58]

Fig. 3 Estimation of motion similarity between a pair of tracklets^[58]

Leung 等^[62] 尝试使用马尔科夫逻辑网络解决目标长时间遮挡问题. 首先利用常见的跟踪方法得到跟踪轨迹, 再检查其中错误关联的部分并断开形成踪片, 最后通过马尔科夫逻辑网络将这些踪片重新关联形成正确的轨迹. 对踪片之间关系建立如下的三个查询谓词来描述: 相同目标 (sameObject)、连接 (join) 以及聚类 (isGroup). 利用踪片的外观相似性和时空一致性构造马尔科夫逻辑网络, 其中踪片的外观采用该轨迹特征颜色直方图的均值和标准差建模, 相似性依据它们均方差归一化直方图交集的大小衡量, 时空一致性则应用踪片之间的时间差和空间位置差计算. 通过最优化该网络获得每个踪片或者踪片对之间的三个查询谓词的赋值情况, 进而可以形成稳定的跟踪轨迹. 例如, 对于如图 4 形式的轨迹, isGroup (tracklet3) 的赋值结果为 1, sameObject (tracklet1, tracklet4) 的赋值结果为 1, join (tracklet1, tracklet3) 的赋值结果为 1, 而 sameObject (tracklet1, tracklet5) 为 0, isGroup (tracklet1) 为 0, join (tracklet1, tracklet5) 为 0. 该算法适用于较拥挤的场景和有长期遮挡的情况, 对于无遮挡或短期遮挡情况, 该算法的复杂度较高并

且性能提升较小.

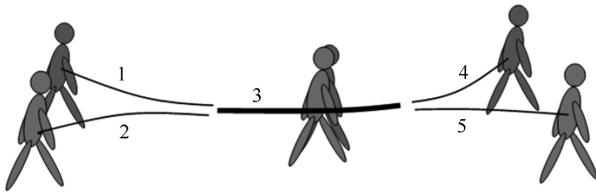


图 4 有遮挡的目标跟踪轨迹^[62]

Fig. 4 Tracklet association for occluded objects^[62]

3.1.2 网络流模型

网络流目标跟踪方法就是将关联问题转化为最大后验概率问题, 将相关轨迹转化为网络流的路径, 将踪片之间的相似性转化为网络流中的费用, 通过最小费用流算法找到全局最优的轨迹关联方式.

Wang 等^[63] 提出了基于外观的在线特定目标的度量模型以及运动模型, 与网络流算法相结合来实现目标跟踪. 文章通过学习距离函数和修正踪片两个步骤构造网络流图中稳定的踪片节点. 首先将包含了较多帧而非仅相邻帧的检测结果关联起来形成初始踪片, 通过学习距离函数来度量两个踪片的相似性. 该初始的踪片只用到了时空信息并包含了较多帧, 稳定性和可靠性较低, 需要利用外观特征将长的踪片切分成小而稳定的踪片从而实现对踪片的修正. 为了计算网络流边的成本即踪片间的相似度, 文章基于运动、时空以及目标离开视野条件建立约束. 关联形成特定目标轨迹的方法为: 给定该目标的起始节点和终止节点, 寻找一条花费最小的路径即为目标跟踪轨迹. 具体流程如下图 5 所示. 该算法在实现了稳定跟踪的同时也减少了目标 ID 转换, 然

而, 该算法鲁棒性还有待提升, 例如目标先离开后又进入视野时, 该算法将得到一条从源节点到目标节点的完整路径, 不符合该目标轨迹不连续的事实, 这主要是由于最小费用流算法只考虑了从源节点到目标节点的全局信息而没有考虑到局部信息.

Wu 等^[64] 基于网络流算法从局部连接、全局连接以及多摄像头三个方面处理轨迹遮挡问题. 文章提出建立一个跟踪图, 分为混合 (Merge) 和分开 (Split) 两部分. Merge 部分是指跟踪关联时从独立轨迹到重叠轨迹, split 是指从重叠轨迹分开成两条或多条独立轨迹. 在构建图时, 每个节点代表形成的踪片. 局部连接时先用网络流算法得到初始可行路径, 之后仅考虑邻接距离为 2 以内的节点直接连接关系, 迭代获得总花费最小的一组路径确定为踪片连接方式. 在全局连接的过程中, 遍历所有从源节点到目标节点可能的路径, 找到使总成本最低的一组路径即为最终的匹配方式. 当场景中有多个视角时, 可以将不同的视角的轨迹按局部或全局连接的方式计算, 并匹配不同视角中同一轨迹, 实现目标在多视角情况下的准确跟踪. 该文章主要对跟踪过程中的遮挡提出解决思路, 采用局部和全局的连接解决短期遮挡问题, 并利用多摄像头数据解决长期遮挡问题.

Shitrit 等^[65] 利用多商品网络流设计踪片关联算法, 以加速多目标轨迹跟踪过程. 文章将定义在栅格占有概率图 (POM) 上的轨迹跟踪问题转化为在有向无环图 (DAG) 中的整数规划问题, 并将其松弛为一个线性规划问题, 以提高求解速度. 该算法首先使用检测算法获得视频序列中的目标,

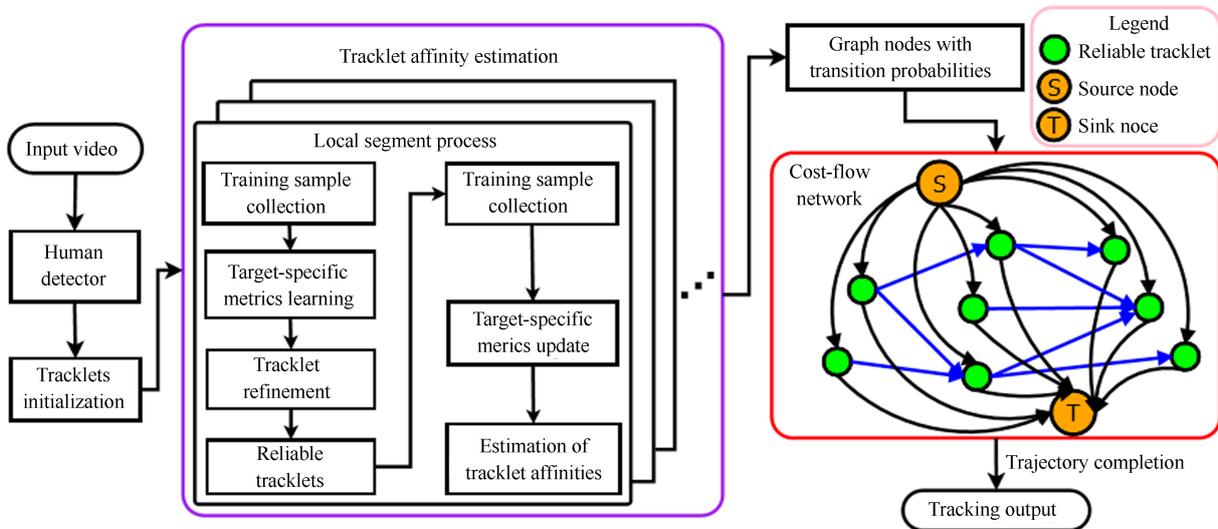


图 5 特定目标度量的踪片关联框架图^[63]

Fig. 5 Framework of tracklet association through target-specific metric learning^[63]

之后根据检测结果将每帧图像分割成相同数目的网格. 将一段视频序列构造成一个层级网络, 其中层级数为视频帧数, 每层节点数为网格数, 层级之间连接只考虑邻域网络的连接结果, 通过线性规划方法来求解网络流问题从而找到相似性最大的长轨迹. 文章提高算法速度及鲁棒性的方法包括: 在构建图的过程中通过剪枝的方法减小图的复杂度; 在连接过程中, 将得到的长轨迹进行分割形成踪片方法为: 检查一条轨迹中每个网格的邻域是否与另一个轨迹中网格的邻域有重合, 若有则将该点断开, 形成踪片, 再利用网络流算法关联成最终轨迹. 该算法能够鲁棒性地减少 ID 转换, 并且适用于计算时间间隔较长的外观相似性. 不足之处是若两物体在相当长的时间内都距离较近时该算法表现较差.

3.1.3 二分图匹配模型

Song 等^[66] 提出了衍生的随机图模型解决视频的长期跟踪的问题. 文章利用粒子滤波来产生稳定可靠的踪片, 利用运动检测方法来自动检测运动的物体. 当无法继续跟踪时, 算法将终止跟踪形成踪片. 为防止 ID 转换, 文章设置了相近目标对之间的距离阈值, 当距离小于阈值时则断开跟踪, 形成踪片. 另外, 文章基于 MCMC (Markov chain Monte Carlo)^[67] 采样方法获得最优特征空间的相似性度量模型, 并利用预测特征来进行相似性测量. 作者定义了一个踪片关联图模型: 用节点代表踪片, 边的权重代表踪片之间的相似性, 将每一个踪片的起始节点和终止节点放在两个子集中, 轨迹连接问题被转化为带权重的二分图最大匹配问题. 如果二分图的相似性分数 (边缘权重) 精确已知并且假设独立, 则可以通过上述方法利用匈牙利算法求最优解. 另外, 作者设计了踪片关联代价 (Track-

let association cost, TAC) 函数, 提出了一种基于 Metropolis-Hastings^[56] 抽样算法的适应方案, 以确保沿着踪片关联后生成的路径上目标特征分布相似, 约束相似性计算的误差, 校正错误踪片, 最终形成长轨迹.

Zamir 等^[68] 利用全局最小基团图算法 (Generalized minimum clique graphs, GMCG) 进行踪片关联. 在踪片的生成阶段, 文章将视频序列按照相同的帧数进行分割, 然后在关联检测目标时比较这一组帧中所有检测结果的相似性, 从而得到踪片. 该关联过程主要依据检测目标的外观相似性与运动相似性, 如图 6 所示. 形成踪片之后, 通过在整段踪片上计算与预设模型的偏差定义全局运动损失模型, 从而使用优化的方法对轨迹进行修正. 另外, 由于按照一组帧进行关联, 当目标发生遮挡时在某些帧中将无法得到该目标的检测结果. 作者巧妙地设计了隐含节点来解决该问题, 当目标在当前组中没有相似的检测与之关联时将其与隐含节点进行关联. 隐含节点的外观特征通过该组中所有非隐含节点特征的均值计算. 在设置隐含节点的位置特征时, 若其为非起始帧, 则用该时刻之前的位置来线性预测此时的位置, 反之则用该时刻以后的位置倒推得到此时的位置. 该方法能够解决遮挡造成的踪片零散化问题, 并且以组为单位进行搜索也能大大提高外观模型的辨识能力.

3.2 其他方法

除了采用传统的图论方法进行关联, 卷积神经网络等深度学习的方法也被引入踪片跟踪中; 社会学等学科相关知识的运用也提高了踪片关联的正确性; 建立成本矩阵以及能量函数, 寻找使得能量最小的踪片关联方法也被广泛使用.

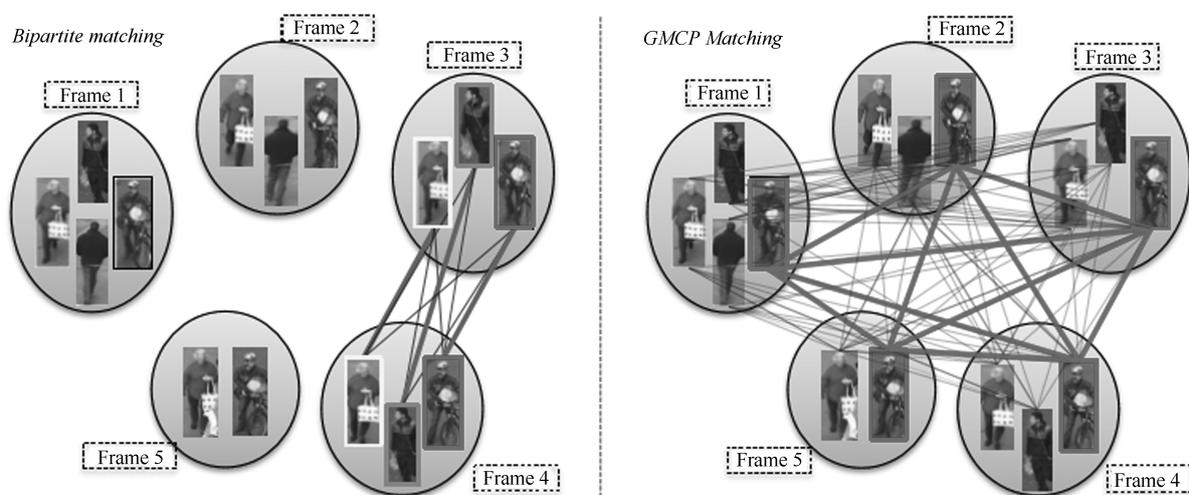


图 6 二分图算法和 GMCP 算法比较^[68]

Fig. 6 The comparison of bipartite and GMCP matching^[68]

Wang 等^[69] 将卷积神经网络 (CNN) 应用到踪片关联中, 作者设计了一个孪生卷积神经网络结构来学习踪片的外观相似性, 并提出一种引入时域约束多任务学习机制进行端到端的学习, 具体流程如图 7 所示. 此外, 作者还创建了 40 个全标记的视频序列进行踪片关联结果的评估. 给定视频输入, 得到每帧中的目标检测结果后, 采用双阈值策略生成可靠的踪片, 并在离线辅助数据上预训练孪生 CNN. 该孪生网络由两个共享参数 (即权重和偏差) 的子卷积神经网络组成, 由于不同时刻目标外观差异较大, 作者考虑到视频段的局部信息在损失函数中引入时域约束, 使得孪生 CNN 能学习到更鲁棒、分辨力更强的外观特征, 并得到更加精准的外观相似性. 之后根据踪片之间的空间位置信息计算运动相似性, 结合外观与运动相似度, 作者将踪片关联问题转化为广义线性分配问题, 所得踪片被表示在一张广义线性分配图 (Generalized linear assignment, GLA), 并采用软分配算法^[70] 求解, 最后基于线性运动模型进行轨迹插值后处理得到最终完整的目标运动轨迹.

Qin 等^[71] 提出了一个在任意基本关联模型中嵌入社交组行为的通用最优化框架, 并将其运用到踪片关联模型中, 以提高目标跟踪的准确性. 通常基本的踪片关联算法仅仅利用了视觉信息, 如外观和运动信息. 本文作者利用了社会学和现有的计算机视觉研究结果, 指出通常彼此靠近的行人具有相似的速度和轨迹, 并且倾向于分组行走. 将视觉信息与社会分组行为信息结合能进一步提高踪片关联鲁棒性和准确性. 文章建立了基本的踪片关联模型: 利用时间约束、外观约束以及运动约束建立踪片之

间的相似性, 也融合了社交行为模型: 加入踪片与踪片组之间的距离约束作为正则项, 如图 8 所示. 作者将最优化框架构造成一个在轨迹空间连接和轨迹分组空间中的最大化一致性问题, 并借鉴最大期望 (Expectation maximization, EM) 算法优化思想设计了一个两步交替迭代优化算法来求解该拉格朗日对偶问题. 利用分组信息进行踪片连接的方式能够大大减少踪片被错误连接的概率, 处理场景繁忙、遮挡严重的情况.

Sun 等^[72] 提出了一种基于场景自适应模型、线性增量外观判别模型和非线性运动模型的多目标跟踪框架. 该方法通过建立场景自适应模型, 关联踪片和检测结果产生局部连接的轨迹; 利用线性增量判别分析方法将踪片进行关联产生全局连接的轨迹; 采用非线性运动模型来填充有空缺的全连接轨迹来保证目标轨迹的完整性和平滑性. 算法首先进行局部关联: 基于鲁棒的特征建立特征空间来描述检测结果和踪片之间的相关性, 利用检测结果更新踪片集合, 选择不可靠的踪片建立分层的特征空间, 再通过迭代实现局部连接. 在全局连接过程中运用线性增量外观判别模型来减少计算复杂性以及提高信息传递的持久性. 为了确保同一目标之间的踪片具有高相似性以及不同目标之间轨迹具有良好的可区分性, 文章提出了基于类间散度最大原则的生成矩阵, 通过特征值分解来获取类间散度和类内散度的主成分, 最后采用非线性运动模型来填充不完整轨迹. 这种策略有效地解决了有重大形变、外观改变的目标识别问题和长时间遮挡问题.

Xing 等^[73] 使用多视角多部分行人检测器

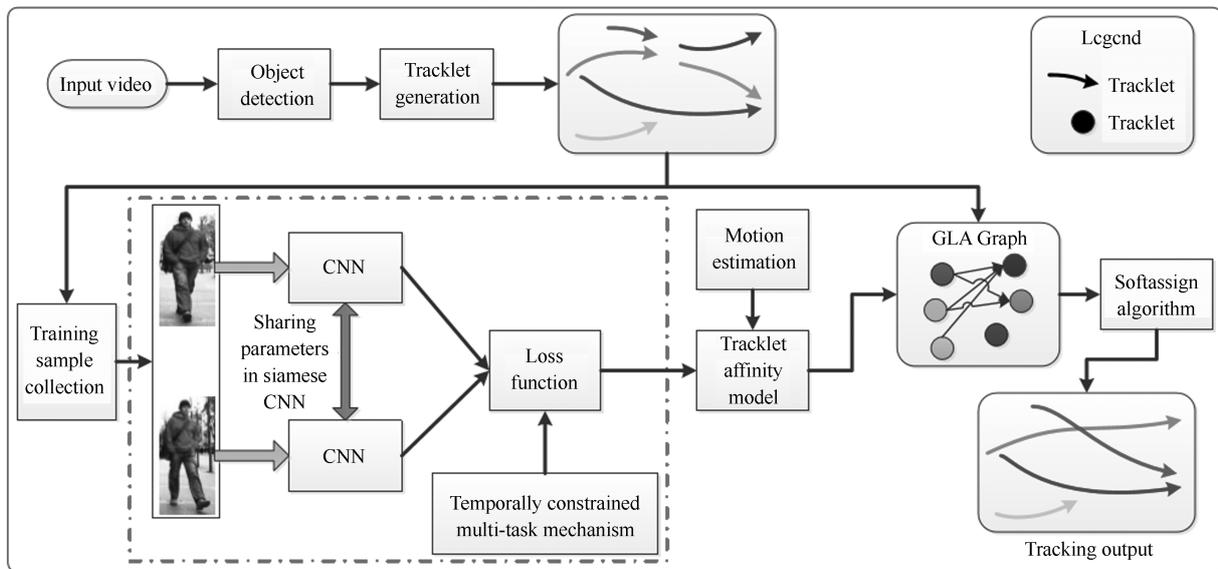
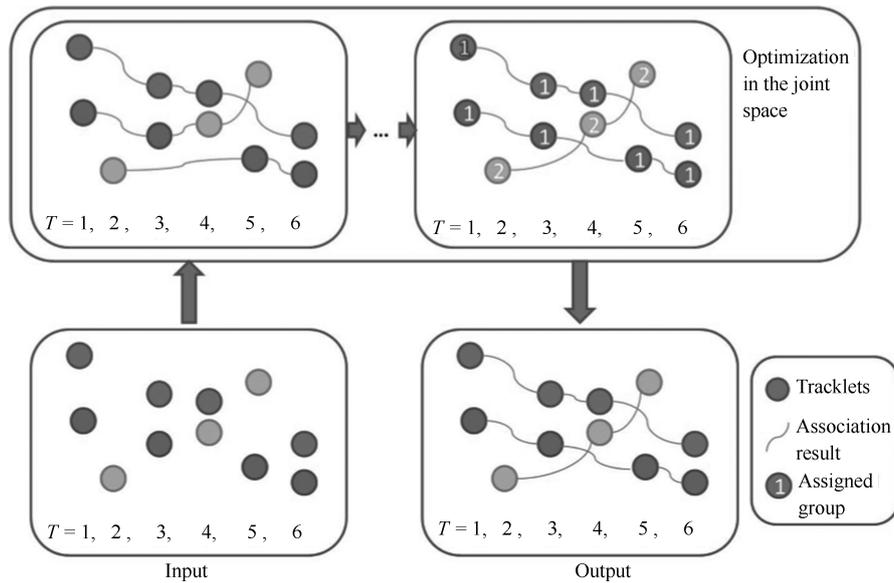


图 7 基于卷积神经网络和时空约束的踪片关联示意图^[69]

Fig. 7 Illustration of tracklet association based on convolutional neural networks and spatio-temporal constraint^[69]

图 8 基本关联与社交组关联结合^[71]Fig. 8 Illustration of the combination of tracklet association and social grouping^[71]

处理行人跟踪问题. 作者将人体分为了头肩 (Head-shoulder, HS)、头躯干 (Head-torso, HT) 以及完整身体 (Full-body, FB) 三个部分分别进行检测与跟踪. 当目标完全没有被遮挡时, 能得到以上三部分的检测结果. 如若目标部分遮挡, 则一定可以检测到 HS 或者 HT, 仍能够得到该目标的轨迹, 从而提高轨迹遮挡的跟踪效果. 文章将相似性高的踪片称作可靠踪片, 将相似度低的踪片称作可能踪片. 通过粒子滤波器关联检测结果生成可靠踪片集合, 然后通过断点使用滑动窗口关联其中的踪片, 并根据观测量优化, 最终生成可能踪片集合. 在形成长轨迹后, 利用可能踪片对其进行校正, 提高跟踪的准确性. 关联踪片时将可靠踪片和可能踪片在外观、尺寸、运动模型进行匹配, 建立成本矩阵并利用匈牙利算法求解得到最佳关联策略. 该方法显著地提高了跟踪的准确性和跟踪的快速性, 同时能够鲁棒地处理目标间的遮挡问题.

Bae 等^[74] 提出基于置信度的踪片连接方法, 包括可检测性与连续性. 踪片的置信度是指构造的踪片与真实轨迹的匹配情况. 文章通过踪片的长度、遮挡情况以及踪片与关联检测之间的相关性来确定踪片的置信度. 具有高置信度的踪片与在线提供的检测局部关联, 从而逐渐生长成局部最佳的踪片. 具有较低置信度的踪片则与其他踪片和检测进行全局连接. 文章还提出一种在线学习方法用于区分目标外观, 从而建立踪片与检测结果的关联关系. 基于外观、形状、运动特征建立踪片局部连接和全局连接过程中的相似性模型, 再使用匈牙利算法计算得到踪片关联情况. 关联形成长轨迹后, 文章运用增量线

性判别分析方法 (Incremental linear discriminant analysis, ILDA) 将踪片的特征投影到另一特征空间中学习高区分度特征, 从而更准确、更鲁棒地聚类同一目标的轨迹. 该方法能够提高踪片关联精度, 具有长时间记忆判别信息的能力, 且能够精确地识别姿势和外观显著变化或长期处于遮挡情况下的目标.

Yang 等^[75] 提出了基于非线性运动模型和鲁棒外观特征学习方法用于多目标跟踪. 文章在文献 [76] 方法的基础上, 使用多层轨迹关联方法生成踪片. 作者将置信度高的踪片进行筛选从而进行非线性拟合: 若该踪片去掉起始几帧和末尾几帧后剩余轨迹仍旧满足非线性的约束, 则踪片可以被用来进行轨迹的非线性估计. 若某个非线性踪片与两个待连接踪片之间距离较近, 则可以按照该非线性模式连接补全, 通过计算最后帧的预测位置和实际位置之间的误差得到非线性运动模型. 在外观模型方面, 该文章提出了多实例增量学习的方法, 正样本从非线性运动误差较低的踪片中获取, 负样本从有重叠或交叉的踪片中获取, 学习得到一个二分类器作为外观模型. 得到长轨迹后, 作者定义了判定踪片是否同组的约束条件. 对于一段时间内足够接近的轨迹, 可认为属于同一个运动群组, 基于此估计该组中消失轨迹位置的期望值, 从而补全由于误检或遮挡形成的短轨迹或间断轨迹. 该方法能够较为精确地估计不符合线性运动的目标轨迹, 同时也在一定程度上解决了遮挡问题.

为方便比较以上不同方法的效果, 文章将算法在公共数据集上的结果进行了汇总并对有代表性的指标进行了比较, 如表 2 所示. 表中各项指标的含义

都可在所涉及论文查到, 在此不再赘述.

表 2 踪片关联跟踪方法在公共数据集上的测试情况表
Table 2 Testing results of tracklet association-based tracking methods on public datasets

| 方法 | 数据集 | 评价指标 | | | | | |
|---------|---------------|--------|--------|------|-------|--------|-------|
| | | MT (%) | ML (%) | FRAG | IDS | MOTA | MOTP |
| 文献 [58] | PETS09 | 89.9 | 0.0 | 13 | 0 | | |
| | TUD | 70.0 | 0.0 | 1 | 0 | | |
| 文献 [63] | TUD | 100 | 0.0 | 3 | 0 | | |
| 文献 [66] | CAVIAR | 84.0 | 4.0 | 6 | 8 | | |
| 文献 [68] | PETS09-view1 | | | 8 | 90.3% | 69.02% | |
| 文献 [69] | PETS09 | 94.7 | 0.0 | 8 | 4 | 95.8% | 86.4% |
| | MOT Challenge | 11.2 | 44.8 | 943 | 712 | 29.6% | 71.8% |
| 文献 [71] | CAVIAR | 88.0 | 2.6 | 5 | 6 | | |
| 文献 [73] | CAVIAR | 84.3 | 3.6 | | 14 | | |
| 文献 [75] | PETS09 | 89.5 | 0.0 | 21 | 15 | | |
| | CAVIAR | 89.1 | 0.7 | 11 | 5 | | |

4 思考与展望

目前, 基于踪片关联的跟踪方法已经取得了进展. 该方法在智能交通、无人驾驶、智能监控、人机交互等领域逐渐得到应用. 在 PAMI、IJCV、ICCV、CVPR、ECCV 等计算机视觉和模式识别领域的期刊和会议中, 涉及该课题的论文数量也逐年上升. 该方法鲁棒性强、计算迅捷、准确性高, 受到了研究者越来越多的关注. 虽然发展迅猛, 但基于踪片关联的跟踪方法的研究与应用仍面临诸多挑战, 下文将对此展开详细讨论, 并提出可能的发展方向.

4.1 对踪片关联跟踪方法的思考

基于踪片关联的跟踪方法主要分为目标检测、跟踪特征提取、踪片生成、踪片关联四个步骤. 目标检测与特征提取的局限性不是本文重点讨论内容, 具体可见文献 [77–80]. 本文主要讨论踪片生成与踪片关联过程中的局限性与发展方向.

在踪片生成过程中, 多数研究者将检测结果按照一定的高阈值进行关联, 此过程会产生孤立的节点, 如何将这些节点进行有效关联也是非常重要的. 通常采用目标的运动特征、外观特征、以及时间约束在相邻帧之间进行关联, 以获得高稳定性和高可靠性. 其中, 运动特征通过建立线性或者非线性模型实现对目标位置的预测, 从而给待匹配踪片赋不同的概率值. 与模型预测位置相近的踪片被赋予较高的概率, 而反之踪片被赋予较低的概率值. 常见的方法有线性高斯预测, 卡尔曼滤波 (Kalman filter,

KF)、KF 计算效率较高, 但只能处理线性的、服从高斯分布的运动模型. 若只采用线性模型估计运动状况, 目标运动复杂时并不能准确跟踪, 很可能将错误的轨迹片段进行关联. 因此在建立运动模型时也需要适当引入非线性模型, 扩展卡尔曼滤波 (Extended Kalman filter, EKF)、粒子滤波 (Particle filter, PF) 等方法也经常引入该过程中进行预测; 缺点在于: EKF 虽然适用于非线性运动模型, 但其依赖于对非线性运动模型的泰勒展开进行一阶线性化截断, 对状态方程的阶次、多模态以及初始分布较敏感; PF 虽然不受模型线性和高斯假设的约束, 但其准确度依赖于在状态空间进行搜索的粒子数目, 高精度状态估计伴随着高昂的计算代价, 其精度与收敛性还受到重采样过程中的随机性、粒子退化、权值退化影响. 另外, 在火车站、机场等拥挤的场所, 目标的运动不仅受到自我行为习惯的影响, 也同样受到其附近行人的影响; 目标可移动的空间范围也有限制. 将这些信息加入目标运动模型中能够在提高踪片关联的准确性的同时也降低了空间搜索的复杂性. 对于外观特征相似性估计, 大多数算法都基于手工设计得到外观特征, 如提取踪片的面积、纹理、颜色直方图等, 这种方法的优点是思路直接, 便于理解, 无需训练过程, 算法运行速度较快, 缺陷在于区分效果不佳. 若使用机器学习的方法, 或者采用 CNN 避免直接提取特征而建模这种相似性, 能够大幅提高特征的区分性, 获得更有效的外观相似性度量. 通过引入时间约束, 去掉时间上有冲突以及相隔较远的片段以缩小搜索空间, 可减少搜索复杂度, 从而提高搜索效率以及连接的准确性. 另外, 将多帧检测结果聚合分组进行比较能够得到更长的踪片, 更便捷地提取出关联信息, 避免众多小踪片的产生, 对于处理遮挡严重的视觉跟踪具有较强的鲁棒性. 然而, 该方法会增加计算复杂度, 增大算法运行的时间, 较难满足跟踪实时性.

在踪片关联过程中, 马尔科夫随机场和网络流算法被广泛使用. 网络流算法搜索从源节点到目标节点的一条路径, 使其成本最小. 该算法的计算复杂度较低, 但是为了解决目标离开后又重新进入视野等问题, 需手动巧妙设计相应的损失函数, 比较依赖于人工配置经验. 另外, 网络流算法的节点和边权值的定义也有很大的可操作空间. 每一个点可以代表单个或者多个踪片, 边权值则代表着连接两个节点所需的费用. 此外, 搜索时需要搜索大量路径, 会有很大的冗余, 需要手动添加限制条件来实现剪枝等过程. 马尔科夫随机场与网络流图相似, 也是通过构造一个能量函数, 计算并找出使能量函数值最小的匹配方式实现目标关联. 该方法能够综合考虑不同因素, 但是由于能量函数也是人为设定, 多靠人工的

经验积累或者实验设置,具有一定的偶然性.该算法对出现复杂的运动情况也无法准确建模.

4.2 展望

由于存在以上缺陷,基于踪片关联的跟踪方法仍有很大的研究空间.模型创新性需要提升:现有的大多数方法只考虑了计算机领域相关知识,模型约束较单一,仅通过优化调整模型参数,算法效果难以显著提升.因此,在今后的研究中可以结合其他多学科领域知识,通过多学科交叉的方式建模目标真实运动轨迹,增加模型的创新性和多样性,引入不同学科的知识增加模型的区分度,减少搜索空间从而减小算法的复杂度.例如,引入社会学、微观动力学^[81-82]等理论,建模空间上相邻目标之间的交互行为,可进一步提高跟踪效果.

社会学主要研究分析社会群体间的互动、跟踪源头及发展过程,分析群体活动对各个成员的影响以及社会特征对日常生活所带来的效果.根据社会学理论,社会群体间存在排斥力与吸引力.社会排斥力是指行人之间由于心理、空间等因素互相保持一定距离;社会吸引力是指行人行动也会有意识地受到亲人、朋友等周围人影响,同时,也无意识的追随某个行人以便保证路径畅通,这使得他们在空间中相互吸引.以群组的形式研究目标运动轨迹能够大大增加踪片关联的正确性.当得知某群组中某个物体的运动规律时能够推测出同组中其他物体的位置,从而有效地解决部分遮挡问题.这不仅可以实现跟踪的鲁棒性,而且能够预测组中所有成员的轨迹.当得知目标的邻近节点位置时能够限制该目标的搜索空间,提升搜索的可靠性和快速性.引入该学科的知识对处理拥挤场景的目标跟踪具有更加重要的意义.由于拥挤场景中每个目标所占的像素较少,外观等特征区分度不大,而目标之间空间距离较小,人群为避免碰撞也总是保持与人流方向一致,因此目标

在下一时刻的候选区域将会限定在较小范围内.将这些信息引入跟踪模型中能够大幅减少跟踪的复杂度,防止与该信息不符的目标踪片相关联,从而提高关联准确度.

微观动力学是分子层次阐明和揭示化学反应过程的化学物理的分支学科,又称微观反应动力学.将其应用在目标跟踪领域构建行人以及车辆的运动轨迹,能够更加鲁棒地实现目标跟踪.图9展示了在不同场合下建立了人群的动力学模型.个体寻求畅通无阻的行走方向,但不喜欢过多偏离从起始点到目的地之间的直接路径.因此,这些个体会在避免障碍物和最直接的途径之间尽量减少绕行路线的情况下找到权衡.当目标运动未受到阻挡时,其轨迹是平滑的直线,当运动区域存在障碍时,运动目标为了避开障碍会实时改变其运动方向和速度,群体也自发地形成了固定的行走路线,可以按照动力学模型进行预测.图中所示的场景分别是走廊、单个障碍、多个障碍、逃生出口、散开、麦加朝圣、小路的场景中人群体的运动学模型.通过将运动学模型引入目标跟踪能够提高对目标运动预测的准确性,与实际轨迹进行比较、匹配以及修正,提高运动目标的跟踪效果.

另外,为了设计和评估目标跟踪算法,需要大规模多样性的标注数据集,但是从实际场景中难以获得满意的数据集.我们认为可以采用平行视觉思路进行研究.2004年,王飞跃等^[84]提出了基于人工社会、计算实验、平行执行(Artificial societies, computational experiments, and parallel execution, ACP)有机组合而成的平行系统理论,以解决复杂系统管理与控制领域的科学研究与实践问题.近年来,基于ACP的平行理论得到了多领域研究者的关注,并且在控制^[85-86]、军事^[87]、机器学习^[88-90]、智能驾驶^[91-92]等领域得到了应用.2016

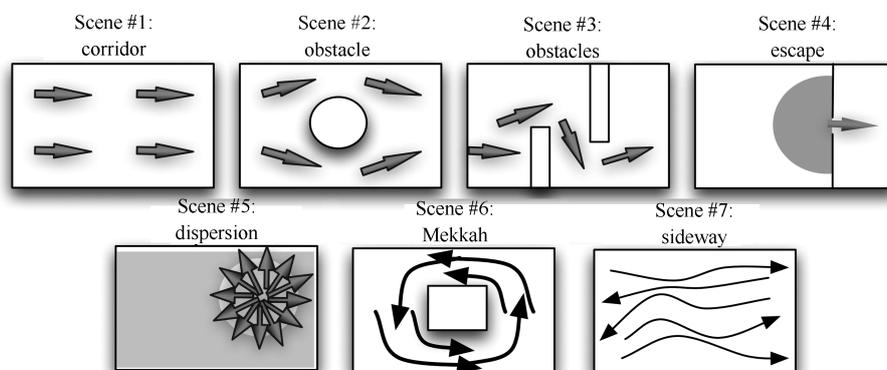


图9 不同场景下的人群动力学模型^[83]

Fig. 9 The crowd dynamic models in different scenes^[83]

年, 王坤峰等^[93-95] 将 ACP 方法推广到计算机视觉领域, 提出平行视觉的概念、框架和关键技术. 如图 10 所示, 在平行视觉中, 通过实际场景与人工场景之间的虚实互动, 衍生出 3 个基本功能模块, 即学习与训练、实验与评估、感知与理解. 具体而言, 平行视觉利用人工场景来模拟和表示复杂挑战的实际场景, 通过计算实验进行各种视觉模型的训练与评估, 最后借助虚实互动的平行执行来在线优化视觉模型, 实现对复杂环境的智能感知与理解. 平行视觉方法使用人工场景生成数据集, 具有数据规模更大、多样性更强、可控可重复、可自动标注等优点. 同时, 人工场景可以是不同光照条件、天气甚至极端环境下的场景, 其多样性也远超过实际数据, 有助于训练出鲁棒性更强的模型. 通过平行视觉方法, 可以产生大量逼真、异质、带标注的虚拟数据, 是视觉算法训练数据来源的很好补充, 可以在工程实践中通过虚实互动进一步优化算法性能.

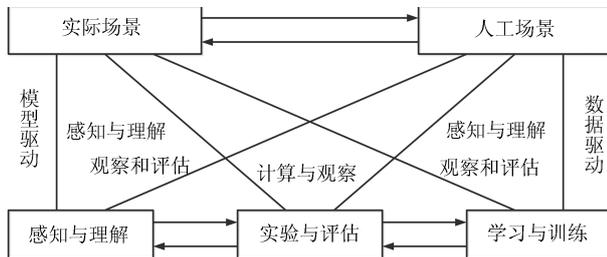


图 10 平行视觉的基本框架与体系结构^[93]
Fig. 10 Basic framework and architecture for parallel vision^[93]

将基于踪片关联的目标跟踪问题与基于 ACP 的平行视觉方法相结合, 可能成为未来一个重要的研究方向. 首先借助三维重建、场景渲染、行为动画建模等方法, 建立高逼真的人工场景, 并渲染模拟得到不同天气、干扰、遮挡等条件下的虚拟场景, 同时自动生成带有数据标注的目标跟踪数据集. 然后, 将人工数据与实际数据相结合, 通过“学习与训练”得到算法模型. 此时, 一种人工数据与数据结合的思路为首先利用人工场景数据集对跟踪模型进行预训练, 再利用实际数据对模型参数进行微调, 以消除数据集偏移. 也可以利用人工数据与实际数据同时训练模型. “实验与评估”则将所得模型用于人工场景和实际场景中, 先在人工场景中进行预测试, 然后在实际场景中进行实地测试来评价算法的准确率、鲁棒性等性能. 此过程中, 应针对不同天气、光照、运动等环境条件下实现对视觉算法全面、充分地测试与评估, 利用统计学等方法来衡量不同算法的性能表现.

5 结论

本文从宏观的角度介绍了目标跟踪的含义、研究意义与现阶段的研究成果, 具体综述了基于踪片关联的目标跟踪方法. 该方法首先利用目标检测方法对需要跟踪的目标对象进行检测, 从每一帧图像中得到感兴趣目标的检测结果, 并用方框或者目标轮廓表示出来. 通过比较相邻帧之间检测结果的外观、运动等相似性将其关联形成踪片. 将得到的踪片用网络流、马尔科夫随机场等方法进一步关联, 得到目标长轨迹. 补全由于遮挡等引起的轨迹中断以形成最终轨迹.

现有的踪片关联准则较为单一, 在随后的研究中可以将社会学、微观动力学等学科与目标跟踪相结合, 建立空间相邻目标之间的交互模型, 并引入机器学习、深度学习等方法学习目标的外观特征, 进一步提高踪片关联的准确性和快速性. 本文还提出将平行视觉理论引入踪片关联过程中, 通过人工场景、计算实验、平行执行三个步骤将虚拟与现实结合起来, 提高跟踪的鲁棒性和准确性, 为基于踪片关联的跟踪提供了一种新的思路.

References

- 1 Kalal Z, Mikolajczyk K, Matas J. Tracking-learning-detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, **34**(7): 1409–1422
- 2 Sun Hong-Guang. The Study of Military Affairs Target Recognition and Tracking Method Based on Wavelet Analysis [Ph.D. dissertation], Changchun University of Science and Technology, China, 2008.
(孙红光. 基于小波分析的军事目标识别及跟踪方法研究 [博士学位论文], 长春理工大学, 中国, 2008.)
- 3 Rautaray S S, Agrawal A. Vision based hand gesture recognition for human computer interaction: a survey. *Artificial Intelligence Review*, 2015, **43**(1): 1–54
- 4 Thakoor N S, An L, Bhanu B, Sunderrajan S, Manjunath B S. People tracking in camera networks: three open questions. *Computer*, 2015, **48**(3): 78–86
- 5 Ess A, Schindler K, Leibe B, Van Gool L. Object detection and tracking for autonomous navigation in dynamic environments. *The International Journal of Robotics Research*, 2010, **29**(14): 1707–1725
- 6 Dinh T B, Vo N, Medioni G. Context tracker: exploring supporters and distracters in unconstrained environments. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Colorado Springs, CO, USA: IEEE, 2011. 1177–1184
- 7 Choi W. Near-online multi-target tracking with aggregated local flow descriptor. In: Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015. 3029–3037
- 8 Alismail H, Browning B, Lucey S. Robust tracking in low light and sudden illumination changes. In: Proceedings of

- the 4th International Conference on 3D Vision (3DV). Stanford, CA, USA: IEEE, 2016. 389–398
- 9 Wang Jiang-Feng. Researches on Object Tracking and Event Detection Based on Tracklet Association [Ph. D. dissertation], National University of Defense Technology, China, 2011.
(王江峰. 基于轨迹片段关联的目标跟踪与事件检测方法研究 [博士学位论文], 国防科学技术大学, 中国, 2011.)
- 10 Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks. *Science*, 2006, **313**(5786): 504–507
- 11 Le N, Heili A, Odobez J M. Long-term time-sensitive costs for CRF-based tracking by detection. In: European Conference on Computer Vision. Amsterdam, The Netherlands: Springer International Publishing, 2016. 43–51
- 12 Lan X S, Xiong Z W, Zhang W, Li S X, Chang H X, Zeng W J. A super-fast online face tracking system for video surveillance. In: Proceedings of the 2016 IEEE International Symposium on Circuits and Systems (ISCAS). Montreal, QC, Canada: IEEE, 2016. 1998–2001
- 13 Huang C H, Allain B, Franco J S, Navab N, Ilic S, Boyer E. Volumetric 3D tracking by detection. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016. 3862–3870
- 14 Yin Hong-Peng, Chen Bo, Chai Yi, Liu Zhao-Dong. Vision-based object detection and tracking: a review. *Acta Automatica Sinica*, 2016, **42**(10): 1466–1489
(尹宏鹏, 陈波, 柴毅, 刘兆栋. 基于视觉的目标检测与跟踪综述. 自动化学报, 2016, **42**(10): 1466–1489)
- 15 Wang X Y, Han T X, Yan S C. An HOG-LBP human detector with partial occlusion handling. In: Proceedings of the 12th International Conference on Computer Vision (ICCV). Kyoto, Japan: IEEE, 2009. 32–39
- 16 Cong Y, Liu W Y, Zhang Y L, Liang H. The research of video tracking based on improved SIFT algorithm. In: Proceedings of the 2016 IEEE International Conference on Mechatronics and Automation (ICMA). Harbin, China: IEEE, 2016. 1703–1707
- 17 Dewan M A A, Granger E, Marcialis G L, Sabourin R, Roli F. Adaptive appearance model tracking for still-to-video face recognition. *Pattern Recognition*, 2016, **49**: 129–151
- 18 Huang Kai-Qi, Chen Xiao-Tang, Kang Yun-Feng, Tan Tie-Niu. *Intelligent visual surveillance: a review*. *Chinese Journal of Computers*, 2015, **38**(6): 1093–1118
(黄凯奇, 陈晓棠, 康运锋, 谭铁牛. 智能视频监控技术. 计算机学报, 2015, **38**(6): 1093–1118)
- 19 Wang Kun-Feng, Gou Chao, Duan Yan-Jie, Lin Yi-Lun, Zheng Xin-Hu, Wang Fei-Yue. Generative adversarial networks: the state of the art and beyond. *Acta Automatica Sinica*, 2017, **43**(3): 321–332
(王坤峰, 苟超, 段艳杰, 林懿伦, 郑心湖, 王飞跃. 生成式对抗网络 GAN 的研究进展与展望. 自动化学报, 2017, **43**(3): 321–332)
- 20 Hua K L, Sari I N, Yeh M C. Human pose tracking using online latent structured support vector machine. In: Proceedings of the 23rd International Conference on Multimedia Modeling. Reykjavik, Iceland: Springer, 2017. 626–637
- 21 Xiang X Z, Bao W L, Tang H W, Li J J, Wei Y M. Vehicle detection and tracking for gas station surveillance based on AdaBoosting and optical flow. In: Proceedings of the 12th World Congress on Intelligent Control and Automation (WCICA). Guilin, China: IEEE, 2016. 818–821
- 22 Miao Yuan. Research of Image Matching Algorithm [Master dissertation], Hefei University of Technology, China, 2013.
(缪源. 图像匹配算法的研究 [硕士学位论文], 合肥工业大学, 中国, 2013.)
- 23 LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, **86**(11): 2278–2324
- 24 陆宗骥. 图象处理领域轮廓跟踪及应用. 中国计算机用户, 1994, (10): 49–52
- 25 Zhang Ji-Ping, Liu Zhi-Fang. Background estimation and moving target detection. *Computing Technology and Automation*, 2004, **23**(4): 51–54
(张继平, 刘直芳. 背景估计与运动目标检测跟踪. 计算技术与自动化, 2004, **23**(4): 51–54)
- 26 Adams R, Bischof L. Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1994, **16**(6): 641–647
- 27 Lin Kai-Yan, Wu Jun-Hui, Xu Li-Hong. A survey on color image segmentation techniques. *Journal of Image and Graphics*, 2005, **10**(1): 1–10
(林开颜, 吴军辉, 徐立鸿. 彩色图像分割方法综述. 中国图象图形学报, 2005, **10**(1): 1–10)
- 28 Han Si-Qi, Wang Lei. A survey of thresholding methods for image segmentation. *Systems Engineering and Electronics*, 2002, **24**(6): 91–94
(韩思奇, 王蕾. 图像分割的阈值法综述. 系统工程与电子技术, 2002, **24**(6): 91–94)
- 29 Wang Hui-Ming, Shi Ping. Methods to extract images texture features. *Journal of Communication University of China Science and Technology*, 2006, **13**(1): 49–52
(王惠明, 史萍. 图像纹理特征的提取方法. 中国传媒大学学报自然科学版, 2006, **13**(1): 49–52)
- 30 Wang Qi-Wei. Study on image histogram feature and application [Ph. D. dissertation], University of Science and Technology of China, China, 2014.
(汪启伟. 图像直方图特征及其应用研究 [博士学位论文], 中国科学技术大学, 中国, 2014.)
- 31 Ding Ming-Yue, Chang Jin-Ling, Peng Jia-Xiong. Research on moment invariants algorithm. *Journal of Data Acquisition & Processing*, 1992, **7**(1): 1–9
(丁明跃, 常金玲, 彭嘉雄. 不变矩算法研究. 数据采集与处理, 1992, **7**(1): 1–9)
- 32 Yan Bo-Jun, Zheng Lian, Wang Ke-Yong. Fast target-detecting algorithm based on invariant moment. *Infrared Technology*, 2001, **23**(6): 8–12
(严柏军, 郑链, 王克勇. 基于不变矩特征匹配的快速目标检测算法. 红外技术, 2001, **23**(6): 8–12)
- 33 Zhang Wei, He Jin-Guo. Construction and generalization of Hu moment invariants. *Journal of Computer Application*, 2010, **30**(9): 2449–2452
(张伟, 何金国. Hu 不变矩的构造与推广. 计算机应用, 2010, **30**(9): 2449–2452)

- 34 Liu Jin, Zhang Tian-Xu. The generalization of moment invariants. *Chinese Journal of Computers*, 2004, **27**(5): 668–674
(刘进, 张天序. 图像不变矩的推广. 计算机学报, 2004, **27**(5): 668–674)
- 35 Hong Zi-Quan, Yang Jing-Yu. Algebraic feature extraction of images for recognition. *Acta Automatica Sinica*, 1992, **18**(2): 233–238
(洪子泉, 杨静宇. 用于图象识别的图象代数特征抽取. 自动化学报, 1992, **18**(2): 233–238)
- 36 Zhao Feng, Huang Qing-Ming, Gao Wen. An image matching algorithm based on singular value decomposition. *Journal of Computer Research and Development*, 2010, **47**(1): 23–32
(赵峰, 黄庆明, 高文. 一种基于奇异值分解的图像匹配算法. 计算机研究与发展, 2010, **47**(1): 23–32)
- 37 Jiang Ming, Zhang Gui-Lin, Hu Ruo-Lan, Chen Zhao-Yang. Research of an image matching method based on principal component analysis. *Infrared and Laser Engineering*, 2000, **29**(4): 17–21
(蒋明, 张桂林, 胡若澜, 陈朝阳. 基于主成分分析的图像匹配方法研究. 红外与激光工程, 2000, **29**(4): 17–21)
- 38 Yang Zhu-Qing, Li Yong, Hu De-Wen. Independent component analysis: a survey. *Acta Automatica Sinica*, 2002, **28**(5): 762–772
(杨竹青, 李勇, 胡德文. 独立成分分析方法综述. 自动化学报, 2002, **28**(5): 762–772)
- 39 Zhang Chun-Mei, Gong Zhi-Hui, Sun Lei. Improved SIFT feature applied in image matching. *Computer Engineering and Applications*, 2008, **44**(2): 95–97
(张春美, 龚志辉, 孙雷. 改进 SIFT 特征在图像匹配中的应用. 计算机工程与应用, 2008, **44**(2): 95–97)
- 40 Nam H, Han B. Learning multi-domain convolutional neural networks for visual tracking. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016. 4293–4302
- 41 Chen Y, Yang X N, Zhong B N, Pan S N, Chen D S, Zhang H Z. CNNTracker: online discriminative object tracking via deep convolutional neural network. *Applied Soft Computing*, 2016, **38**: 1088–1098
- 42 Bertinetto L, Valmadre J, Henriques J F, Vedaldi A, Torr P H S. Fully-convolutional siamese networks for object tracking. In: European Conference on Computer Vision. Amsterdam, The Netherlands: Springer, 2016. 850–865
- 43 Zhao Liang, Liu Jian-Hui, Wang Xing. Hellinger distance based similarity analysis for categorical variables in mixture dataset. *Computer Science*, 2016, **43**(6): 280–282
(赵亮, 刘建辉, 王星. 基于 Hellinger 距离的混合数据集中分类变量相似度分析. 计算机科学, 2016, **43**(6): 280–282)
- 44 Xuan Guo-Rong, Chai Pei-Qi. Feature selection based on Bhattacharyya distance. *PR & AI*, 1996, **9**(4): 324–329
(宣国荣, 柴佩琪. 基于巴氏距离的特征选择. 模式识别与人工智能, 1996, **9**(4): 324–329)
- 45 周蕴时, 何天晓, 常玉堂. 多元插值法. 工数数学, 1985, (1): 12–16
- 46 Lo S C B, Chan H P, Lin J S, Li H, Freedman M T, Mun S K. Artificial convolution neural network for medical image pattern recognition. *Neural Networks*, 1995, **8**(7–8): 1201–1214
- 47 Ferryman J, Shahrokni A. PETS2009: dataset and challenge. In: Proceedings of the 20th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS-Winter). Snowbird, UT, USA: IEEE, 2009. 1–6
- 48 Leal-Taixé L, Milan A, Reid I, Schindler K. MOTChallenge 2015: towards a benchmark for multi-target tracking. arXiv preprint arXiv: 1504.01942, 2015.
- 49 Fisher R B. The PETS04 surveillance ground-truth data sets. In: Proceedings of the 6th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance. New York, USA: IEEE, 2004. 1–5
- 50 Home Office Scientific Development Branch. Imagery library for intelligent detection systems (i-LIDS). In: Proceedings of the 2006 Institution of Engineering and Technology Conference on Crime and Security. London, UK: IET, 2006. 445–448
- 51 Wen L Y, Du D W, Cai Z W, Lei Z, Chang M C, Qi H G, Lim J, Yang M H, Lyu S. UA-DETRAC: a new benchmark and protocol for multi-object detection and tracking. arXiv preprint arXiv: 1511.04136, 2015.
- 52 Alahi A, Ramanathan V, Li F F. Socially-aware large-scale crowd forecasting. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus, OH, USA: IEEE, 2014. 2203–2210
- 53 Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? the KITTI vision benchmark suite. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Providence, RI, USA: IEEE, 2012. 3354–3361
- 54 Ros G, Sellart L, Materzynska J, Vazquez D, Lopez A M. The SYNTHIA dataset: a large collection of synthetic images for semantic segmentation of urban scenes. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016. 3234–3243
- 55 Huang C, Wu B, Nevatia R. Robust object tracking by hierarchical association of detection responses. In: Proceedings of the 10th European Conference on Computer Vision. Marseille, France: Springer, 2008. 788–801
- 56 Richard M D, Lippmann R P. Neural network classifiers estimate Bayesian a posteriori probabilities. *Neural Computation*, 1991, **3**(4): 461–483
- 57 Greig D M, Porteous B T, Seheult A H. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1989, **51**(2): 271–279
- 58 Yang B, Nevatia R. Multi-target tracking by online learning a CRF model of appearance and motion patterns. *International Journal of Computer Vision*, 2014, **107**(2): 203–217
- 59 Overett G, Petersson L, Brewer N, Andersson L, Pettersson N. A new pedestrian dataset for supervised learning. In: Proceedings of the 2008 IEEE Intelligent Vehicles Symposium. Eindhoven, Netherlands: IEEE, 2008. 373–378
- 60 Wu B Y, Lyu S, Hu B G, Ji Q. Simultaneous clustering and tracklet linking for multi-face tracking in videos. In: Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV). Sydney, NSW, Australia: IEEE, 2013. 2856–2863

- 61 Viola P, Jones M J. Robust real-time face detection. *International Journal of Computer Vision*, 2004, **57**(2): 137–154
- 62 Leung V, Herbin S. Flexible tracklet association for complex scenarios using a Markov Logic Network. In: Proceedings of the 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops). Barcelona, Spain: IEEE, 2011. 1870–1875
- 63 Wang B, Wang G, Luk Chan K, Wang L. Tracklet association with online target-specific metric learning. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus, OH, USA: IEEE, 2014. 1234–1241
- 64 Wu Z, Kunz T H, Betke M. Efficient track linking methods for track graphs using network-flow and set-cover techniques. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Colorado Springs, CO, USA: IEEE, 2011. 1185–1192
- 65 Shitrit H B, Berclaz J, Fleuret F, Fua P. Multi-commodity network flow for tracking multiple people. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, **36**(8): 1614–1627
- 66 Song B, Jeng T Y, Staudt E, Roy-Chowdhury A K. A stochastic graph evolution framework for robust multi-target tracking. In: Proceedings of the 11th European Conference on Computer Vision. Heraklion, Crete, Greece: Springer, 2010. 605–619
- 67 Geyer C J. Practical Markov chain monte Carlo. *Statistical Science*, 1992, **7**(4): 473–483
- 68 Zamir A R, Dehghan A, Shah M. GMCP-tracker: global multi-object tracking using generalized minimum clique graphs. *Computer Vision – ECCV 2012*. Berlin, Heidelberg: Springer, 2012. 343–356
- 69 Wang B, Wang L, Shuai B, Zuo Z, Liu T, Chan K L, Wang G. Joint learning of convolutional neural networks and temporally constrained metrics for tracklet association. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Las Vegas, NV, USA: IEEE, 2016. 1–8
- 70 Gold S, Rangarajan A. Softmax to softassign: neural network algorithms for combinatorial optimization. *Journal of Artificial Neural Networks*, 1996, **2**(4): 381–399
- 71 Qin Z, Shelton C R. Improving multi-target tracking via social grouping. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Providence, RI, USA: IEEE, 2012. 1972–1978
- 72 Sun X, Zhu S H, Jin D L, Liang Z W, Xu G Z. Tracklet association for object tracking. In: Proceedings of the 2016 Chinese Control and Decision Conference (CCDC). Yinchuan, China: IEEE, 2016. 107–112
- 73 Xing J L, Ai H Z, Lao S H. Multi-object tracking through occlusions by local tracklets filtering and global tracklets association with detection responses. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Miami, FL, USA: IEEE, 2009. 1200–1207
- 74 Bae S H, Yoon K J. Robust online multi-object tracking based on tracklet confidence and online discriminative appearance learning. In: Proceedings of the 2014 IEEE conference on Computer Vision and Pattern Recognition (CVPR). Columbus, OH, USA: IEEE, 2014. 1218–1225
- 75 Yang B, Nevatia R. Multi-target tracking by online learning of non-linear motion patterns and robust appearance models. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Providence, RI, USA: IEEE, 2012. 1918–1925
- 76 Kuo C H, Nevatia R. How does person identity recognition help multi-person tracking? In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Colorado Springs, CO, USA: IEEE, 2011. 1217–1224
- 77 Kumar G, Bhatia P K. A detailed review of feature extraction in image processing systems. In: Proceedings of the 4th International Conference on Advanced Computing & Communication Technologies (ACCT). Rohtak, India: IEEE, 2014. 5–12
- 78 Kulchandani J S, Dangarwala K J. Moving object detection: review of recent research trends. In: Proceedings of the 2015 International Conference on Pervasive Computing (ICPC). Pune, India: IEEE, 2015. 1–5
- 79 Shukla A P, Saini M. “Moving object tracking of vehicle detection”: a concise review. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 2015, **8**(3): 169–176
- 80 Ding Zhong-Xiao. Survey on moving object detection methods for video surveillance images. *Video Engineering*, 2008, **32**(5): 72–76
(丁忠校. 视频监控图像的运动目标检测方法综述. 电视技术, 2008, **32**(5): 72–76)
- 81 Moussaïd M, Helbing D, Theraulaz G. How simple rules determine pedestrian behavior and crowd disasters. *Proceedings of the National Academy of Sciences of the United States of America*, 2011, **108**(17): 6884–6888
- 82 Helbing D, Farkas I, Vicsek T. Simulating dynamical features of escape panic. *Nature*, 2000, **407**(6803): 487–490
- 83 Courty N, Allain P, Creusot C, Corpetti T. Using the AGO-RASET dataset: assessing for the quality of crowd video analysis methods. *Pattern Recognition Letters*, 2014, **44**: 161–170
- 84 Wang Fei-Yue. Parallel system methods for management and control of complex systems. *Control and Decision*, 2004, **19**(5): 485–489, 514
(王飞跃. 平行系统方法与复杂系统的管理和控制. 控制与决策, 2004, **19**(5): 485–489, 514)
- 85 Wang Fei-Yue. Parallel control: a method for data-driven and computational control. *Acta Automatica Sinica*, 2013, **39**(4): 293–302
(王飞跃. 平行控制: 数据驱动的计算控制方法. 自动化学报, 2013, **39**(4): 293–302)
- 86 Bai Tian-Xiang, Wang Shuai, Shen Zhen, Cao Dong-Pu, Zheng Nan-Ning, Wang Fei-Yue. Parallel robotics and parallel unmanned systems: framework, structures, process, platform and applications. *Acta Automatica Sinica*, 2013,

43(2), 161–175

(白天翔, 王帅, 沈震, 曹东璞, 郑南宁, 王飞跃. 平行机器人与平行无人系统: 框架、结构、过程、平台及其应用. *自动化学报*, 2017, **43**(2): 161–175)

- 87 Bai Tian-Xiang, Wang Shuai, Zhao Xue-Liang, Qin Ji-Rong. Parallel weapons: weapons towards intelligent warfare. *Journal of Command and Control*, 2017, **3**(2): 89–98 (白天翔, 王帅, 赵学亮, 秦继荣. 平行武器: 迈向智能战争的武器. *指挥与控制学报*, 2017, **3**(2): 89–98)
- 88 Li Li, Lin Yi-Lun, Cao Dong-Pu, Zheng Nan-Ning, Wang Fei-Yue. Parallel learning — a new framework for machine learning. *Acta Automatica Sinica*, 2017, **43**(1): 1–8 (李力, 林懿伦, 曹东璞, 郑南宁, 王飞跃. 平行学习 — 机器学习的一个新型理论框架. *自动化学报*, 2017, **43**(1): 1–8)
- 89 Li L, Lin Y L, Zheng N N, Wang F Y. Parallel learning: a perspective and a framework. *IEEE/CAA Journal of Automatica Sinica*, 2017, **4**(3): 389–395
- 90 Liu Xin, Wang Xiao, Zhang Wei-Shan, Wang Jian-Ji, Wang Fei-Yue. Parallel data: from big data to data intelligence. *Pattern Recognition and Artificial Intelligence*, 2017, **30**(8): 673–681 (刘昕, 王晓, 张卫山, 汪建基, 王飞跃. 平行数据: 从大数据到数据智能. *模式识别与人工智能*, 2017, **30**(8): 673–681)
- 91 Wang F Y. Scanning the issue and beyond: parallel driving with software vehicular robots for safety and smartness. *IEEE Transactions on Intelligent Transportation Systems*, 2014, **15**(4): 1381–1387
- 92 Wang F Y, Zheng N N, Cao D P, Martinez C M, Li L, Liu T. Parallel driving in CPSS: a unified approach for transport automation and vehicle intelligence. *IEEE/CAA Journal of Automatica Sinica*, 2017, **4**(4): 577–587
- 93 Wang Kun-Feng, Gou Chao, Wang Fei-Yue. Parallel vision: an ACP-based approach to intelligent vision computing. *Acta Automatica Sinica*, 2016, **42**(10): 1490–1500 (王坤峰, 苟超, 王飞跃. 平行视觉: 基于 ACP 的智能视觉计算方法. *自动化学报*, 2016, **42**(10): 1490–1500)
- 94 Wang K F, Gou C, Zheng N N, Reh J M, Wang F Y. Parallel vision for perception and understanding of complex scenes: methods, framework, and perspectives. *Artificial Intelligence Review*, 2017, **48**(3): 299–329
- 95 Wang Kun-Feng, Lu Yue, Wang Yu-Tong, Xiong Zi-Wei, Wang Fei-Yue. Parallel imaging: a new theoretical framework for image generation. *Pattern Recognition and Artificial Intelligence*, 2017, **30**(7): 577–587 (王坤峰, 鲁越, 王雨桐, 熊子威, 王飞跃. 平行图像: 图像生成的一个新型理论框架. *模式识别与人工智能*, 2017, **30**(7): 577–587)



刘雅婷 中国科学院自动化研究所复杂系统管理与控制国家重点实验室博士研究生. 主要研究方向为视觉目标跟踪, 机器学习.

E-mail: liuyating2015@ia.ac.cn

(LIU Ya-Ting Ph.D. candidate at The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences. Her research interest covers visual object tracking and machine learning.)



王坤峰 中国科学院自动化研究所复杂系统管理与控制国家重点实验室副研究员. 主要研究方向为智能交通系统, 智能视觉计算, 机器学习.

E-mail: kunfeng.wang@ia.ac.cn

(WANG Kun-Feng Associate professor at The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences. His research interest covers intelligent transportation systems, intelligent vision computing, and machine learning.)



王飞跃 中国科学院自动化研究所复杂系统管理与控制国家重点实验室研究员. 国防科技大学军事计算实验与平行系统技术研究中心主任. 主要研究方向为智能系统和复杂系统的建模、分析与控制. 本文通信作者.

E-mail: feiyue.wang@ia.ac.cn

(WANG Fei-Yue Professor at The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences. Director of the Research Center for Computational Experiments and Parallel Systems Technology, National University of Defense Technology. His research interest covers modeling, analysis, and control of intelligent systems and complex systems. Corresponding author of this paper.)