基于视觉的目标检测与跟踪综述

尹宏鹏 [1, 2] 陈 波 [2] 柴 毅 [2] 刘兆栋 [2]

摘 要 基于视觉的目标检测与跟踪是图像处理、计算机视觉、模式识别等众多学科的交叉研究课题, 在视频监控、虚拟现实、人机交互、自主导航等领域, 具有重要的理论研究意义和实际应用价值. 本文对目标检测与跟踪的发展历史、研究现状以及典型方法给出了较为全面的梳理和总结. 首先, 根据所处理的数据对象的不同, 将目标检测分为基于背景建模和基于前景建模的方法, 并分别对背景建模与特征表达方法进行了归纳总结. 其次, 根据跟踪过程有无目标检测的参与, 将跟踪方法分为生成式与判别式, 对基于统计的表观建模方法进行了归纳总结. 然后, 对典型算法的优缺点进行了梳理与分析, 并给出了其在标准数据集上的性能对比. 最后, 总结了该领域待解决的难点问题, 对其未来的发展趋势进行了展望.

关键词 计算机视觉, 目标检测, 目标跟踪, 背景建模, 表观建模

引用格式 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

DOI 10.16383/j.aas.2016.c150823

Vision-based Object Detection and Tracking: A Review

YIN Hong-Peng[1, 2] CHEN Bo[2] CHAI Yi[2] LIU Zhao-Dong[2]

Abstract Vision-based object detection and tracking is an active research topic in image processing, computer vision, pattern recognition, etc. It finds wide applications in video surveillance, virtual reality, human-computer interaction, autonomous navigation, etc. This survey gives a detail overview of the history, the state-of-the-art, and typical methods in this domain. Firstly, object detection is divided into background-modeling-based methods and foreground-modeling-based methods according to the different data objects processed. Background modeling and feature representation are further summarized respectively. Then, object tracking is divided into generative and discriminative methods according to whether the detection process is involved. Statistical based appearance modeling is presented. Besides, discussions are presented on the advantages and drawbacks of typical algorithms. The performances of different algorithms on benchmark datasets are given. Finally, the outstanding issues are summarized. The future trends of this field are discussed.

Key words Computer vision, object detection, object tracking, background modeling, appearance modeling

Citation 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

随着信息技术的发展, 基于视觉的运动目标检测与跟踪, 已逐渐渗透到人们生活的方方面面, 其重要性日益突出, 吸引了越来越多的海内外学者及研究机构参与到此领域的研究. 目前, 基于视觉的运动目标检测与跟踪已广泛应用于视频监控、虚拟现实、人机交互、行星探测、行为理解等领域, 实现了公共安全监控与管理, 意外事件防范、检测及处理, 应急推演, 老幼病残监护以及自主导航等功能, 其具体应用分类详见表 1.

早在 60 年代, 国外就已开展了对运动目标检测与跟踪的研究[1]. Papageorgiou 等[2] 提出了静态图像中用于目标检测的一般框架, 该框架直接从样本中学习特征, 不需要任何的先验知识、模型或者运动分割. Viola 等[3] 将积分图用于图像特征表达, 采用级联分类器实现了对目标的鲁棒实时检测. Lowe[4] 通过获取图像关键点附近的梯度信息来描述运动目标, 提出了尺度不变特征 (Scale invariant feature transform, SIFT). Dalal 等[5] 提出了梯度直方图特征 (Histogram of oriented gradient, HOG), 将其用于解决静态图像的行人检测问题. Felzenszwalb 等[6] 将 HOG 与支持向量机 (Support vector mach-

收稿日期 2015-12-14 录用日期 2016-05-16
Manuscript received December 14, 2015; accepted May 16, 2016
国家自然科学基金 (61203321), 重庆市基础科学与前沿研究技术专项重点项目 (cstc2015jcyjB0569), 中央高校基本科研业务专项基金 (106112016CDJZR175511, 106112015CDJXY170003), 重庆市研究生科研创新项目 (CYB14023) 资助
Supported by National Natural Science Foundation of China (61203321), Chongqing Nature Science Foundation of Fundamental Science and Frontier Technologies (cstc2015jcyjB0569), China Central Universities Foundation (106112016CDJZR175511, 106112015CDJXY170003), Chongqing Graduate Student Research Innovation Project (CYB14023)
本文责任编委 刘跃虎
Recommended by Associate Editor LIU Yue-Hu
1. 信息物理社会可信服务计算教育部重点实验室 (重庆大学) 重庆 400030 2. 重庆大学自动化学院 重庆 400044
1. Key Laboratory of Dependable Service Computing in Cyber Physical Society (Chongqing University), Ministry of Education, Chongqing 400030 2. College of Automation, Chongqing University, Chongqing 400044

表 1 基于视觉的目标检测与跟踪应用领域

Table 1 Applications of vision-based object detection and tracking

应用领域	具体应用
智能监控	公共安全监控 (犯罪预防、人流密度检测)、停车场、超市、百货公司、自动售货机、ATM、小区 (外来人员访问控制)、交通场景、家庭环境 (老幼看护) 等
虚拟现实	交互式虚拟世界、游戏控制、虚拟工作室、角色动画、远程会议等
高级人机交互	手语翻译、基于手势的控制、高噪声环境 (机场、工厂等) 下的信息传递等
动作分析	基于内容的运动视频检索, 高尔夫、网球等的个性化训练, 舞蹈等的编排, 骨科患者的临床研究等
自主导航	车辆导航、机器人导航、太空探测器的导航等
机器人视觉	工业机器人、家庭服务机器人、餐厅服务机器人、太空探测器等

ine, SVM) 相结合, 提出了可变形部件模型 (Deformable part model, DPM), 逐渐成为近年来最受欢迎的目标检测模型之一, 该工作在 2010 年被授予了模式分析, 统计建模, 计算学习视觉目标分类 (Pattern Analysis, Statistical Modelling and Computational Learning Visual Object Classes, PASCAL VOC) 挑战赛[7] "终身成就奖".

国内从 2002 年到 2012 年, 相继召开了 3 届全国智能视觉监控学术会议, 其内容主要涉及了智能视频监控、背景建模、图像分割、目标检测与跟踪、行为理解与分析、数据压缩与传输、系统构建与集成等内容. 另外, 从 2011 年到 2016 年相继召开了 6 届视觉与学习青年研讨会 (Vision and Learning Seminar, VALSE) 也涉及了计算机视觉、图像处理、模式识别与机器学习等众多领域, 重点讨论了图像分类、目标检测、视觉跟踪等相关技术的进展.

随着对基于视觉的目标检测与跟踪研究的深入, 大量的相关成果不断涌现. 为此, 众多学者逐渐展开了对目标检测与跟踪的综述工作, 对该领域进行了归纳与总结, 典型的综述文献及主要内容如表 2 所示. 在已有的综述文献中, 部分综述性工作只是面向某一类具体应用进行展开, 如文献 [8] 对手势识别中的目标跟踪进行了深入讨论, 文献 [9] 针对行星地形分类问题对目标检测方法展开了综述; 部分文献只是讨论了目标检测与跟踪的一个较小的组成部分, 如文献 [10−12] 讨论了目标表观建模方面的一些研究进展, 文献 [13−14] 讨论了目标检测中的背景建模方法, 文献 [15] 对多种跟踪算法进行了实验对比与分析; 部分文献在总结时, 对目标检测与跟踪部分只进行了简要概述, 如文献 [16−17] 分别综述了行为理解和智能监控的研究进展, 在其底层部分简要地介绍了目标的检测与跟踪, 没有给出针对该领域较全面的整理; 当然, 也有一些学者对运动目标检测与跟踪进行了较为全面的阐述、分析与整理, 如文献 [18−22], 这部分工作极大地推动并促进了该领域的研究与发展. 然而, 随着时代的发展, 目标检测与跟踪技术也在突飞猛进, 近几年来该领域的研究成果可谓日新月异, 各种优秀算法也是层出不穷, 特别是计算机视觉三大会议 (CVPR (Computer Vision and Pattern Recognition), ICCV (International Conference on Computer Vision), ECCV (European Conference on Computer Vision)) 每年均有许多相关报道, 对这些工作及成果进行梳理与总结, 将会有效地推动和促进计算机视觉领域的发展. 运动目标的检测和跟踪主要用于获取运动目标的位置、姿态、轨迹等基本运动信息, 是理解服务对象或对目标进行控制的前提和基础. 本文试图对运动目标检测与跟踪, 在不同层面上的典型算法进行较为全面的梳理与总结, 将其整体结构归纳为如图 1 所示的框架. 其中, 目标检测可看作是目标跟踪的组成部分, 主要用于对目标状态的初始化, 目标跟踪则是在目标检测的基础上, 对目标的状态进行连续估计的过程.

本文分别对运动目标检测及其跟踪两方面进行了梳理与总结. 第 1 节按算法所处理的数据对象的不同, 将目标检测分为基于背景建模与基于目标建模的检测方法, 并分别阐述了目标检测技术的常用方法, 重点讨论了背景建模方法以及表观建模中的特征表达方法; 第 2 节将目标跟踪按其与目标检测的关系的不同, 分为生成式与判别式跟踪方法, 重点讨论了表观建模中的统计建模方法; 第 3 节综合目标检测、目标跟踪的研究现状, 对该领域未来的研究趋势进行了展望.

1 目标检测

目标检测的目的是从不同复杂程度的背景中辨识出运动目标, 并分离背景, 从而完成跟踪、识别等后续任务. 因此, 目标检测是高层理解与应用的基础任务, 其性能的好坏将直接影响后续的目标跟踪、动作识别以及行为理解等中高层任务的性能.

表 2 目标检测与跟踪相关综述文献

Table 2 Related surveys about object detection and tracking

文献	题目	主要内容	讨论主题	发表年限	不足之处
[8]	Vision based hand gesture recognition for human computer interaction: a survey	从检测、跟踪与识别三方面对手势识别的发展现状进行了梳理与总结	检测、跟踪、识别	2015	只进行了某些具体应用方向上的梳理
[9]	A survey on recent object detection techniques useful for monocular vision-based planetary terrain classification	对行星地形分类中的目标检测技术进行了总结	目标检测	2014	
[10]	Sparse coding based visual tracking: review and experimental comparison	对基于稀疏编码的目标跟踪进行了全面的梳理与总结, 给出了实验对比与分析	表观建模	2013	只讨论了目标检测与跟踪的组成部分
[11]	A survey of appearance models in visual object tracking	从全局与局部信息描述的角度探讨了目标跟踪中的视觉表达问题	表观建模	2013	
[12]	面向目标检测的稀疏表示方法研究进展	综述了稀疏表示方法在目标检测领域中的国内外重要研究进展	表观建模	2015	
[13]	Background subtraction techniques: a review	对几种常用的背景减除方法进行了总结	背景建模	2004	
[14]	Traditional and recent approaches in background modeling for foreground detection: an overview	对目标检测中背景建模方法进行了详细讨论	背景建模	2014	
[15]	Visual tracking: an experimental survey	对 19 种先进的跟踪器在 315 段视频序列上进行了对比实验与性能评估	目标跟踪	2014	
[16]	Automated human behavior analysis from surveillance videos: a survey	在人体行为理解的底层处理部分, 对目标检测、分类及其跟踪进行了详细阐述	人体行为理解	2014	没有展开讨论检测跟踪问题
[17]	智能视频监控技术综述	在智能视频监控的底层部分, 对目标检测与跟踪进行了讨论	智能监控	2015	
[18]	Object tracking: a survey	对目标跟踪中的目标表达、特征或运动模型选取等问题进行了分类归纳	目标跟踪	2006	发表年限比较久远, 不断更新的理论和方法亟需梳理总结
[19]	视觉跟踪技术综述	分类归纳了视觉跟踪, 并论述了其在视频监控、图像压缩和三维重构等的应用	目标跟踪	2006	
[20]	运动目标检测算法的探讨	对 2007 年以前的主流运动目标检测方法进行了分类讨论	目标检测	2006	
[21]	运动目标跟踪算法研究综述	将运动目标跟踪问题分为运动检测与目标跟踪, 并对跟踪算法进行了综述工作	目标跟踪	2009	
[22]	微弱运动目标的检测与跟踪识别算法研究	对强噪声背景下的微弱运动目标检测与跟踪算法进行了探讨	目标检测与跟踪	2010	

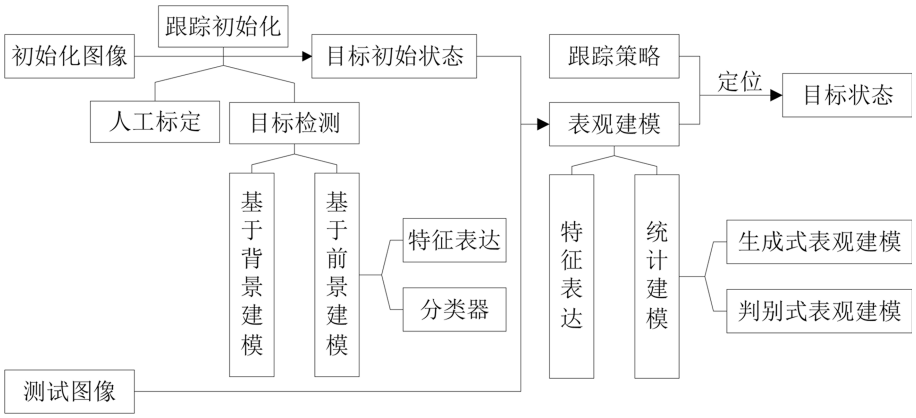

图 1 基于视觉的目标检测与跟踪框架

Fig. 1 General framework of vision-based object detection and tracking

目标检测的任务是要分割“我们不关心”的背景从而获取“我们所关心”的前景目标. 因此, 按算法处理对象的不同, 目标检测方法可以分为基于背景建模的目标检测方法和基于前景建模的目标检测方法. 其中, 基于背景建模的方法通过对背景进行估计, 建立起背景模型与时间的关联关系, 将当前帧与所建背景模型进行对比作差, 间接地分离出运动前景, 最后经过前景分割得到跟踪目标; 基于前景目标建模的方法则是采用灰度、颜色、纹理等同质特征, 建立起跟踪目标的表观模型, 并设计适当的分类器对其进行分类与检测.

1.1　基于背景建模的目标检测

基于背景建模的方法将当前帧与背景参考模型进行比对, 再通过阈值法来判断各个像素是否属于运动前景, 最终对检测出的运动前景进行分割得到跟踪目标. 基于背景建模的目标检测方法一般包含背景模型的初始化、模型维护及前景检测与分割等步骤, 其处理的一般流程如图 2 所示, N 表示用于背景模型初始化的视频帧数.

背景初始化 (Background initialization)[23] 指对背景模型的初始化工作. 其中, 最简单的初始化方法是得到一帧不包含任何运动目标的背景图像. 通常的初始化模型可以从一段较短且不包含前景目标的训练序列中获取. 然而, 实际场景却较难满足不包含前景目标的纯背景特性, 这就要求我们使用包含前景目标的一组序列去获取背景初始模型. Wang 等[24] 对初始化背景模型 *Median* 进行改进, 提出了一种能容纳超过 50 % 前景目标或噪声的鲁棒初始化模型. Colombari 等[25] 提出了基于块的背景初始化方法, 能够处理前景目标在场景中静止一段时间的复杂情况.

实际场景中背景因受光照变化、场景中目标的进入或退出等因素的影响而时刻发生变化, 准确的背景模型变得难以获取. 因此如何构造鲁棒、自适应的背景模型是基于背景建模的运动目标检测算法的关键. 目前, 对于背景建模已有大量的研究工作. Lee 等[26] 将均值背景模型用于视频监控, McFarlane 等[27] 在对小猪的跟踪过程中采用了中值滤波模型. 另外, 还有帧间差分、最大最小值滤波等其他早期常被使用的基本模型. 随着对背景建模研究工作的推进, 又出现了统计模型[28−29]、聚类模型[30−31]、神经网络模型[32−33]、估计模型[34−35]、模糊模型[36]、鲁棒子空间模型[37]、稀疏模型[38]、变换域模型[39−40] 等背景模型. Bouwmans 等分别在 2010 年[28]、2011 年[29] 和 2014 年[14] 对背景建模进行了综述工作, 对背景建模方法进行了详细的分类与总结, 有兴趣的读者可以参考相关文献.

背景模型中, 高斯模型[41] 和支持向量模型[42] 通常用于处理动态背景; 子空间学习模型[37] 多用于处理光照变化问题; 模糊模型[36] 能同时处理光照变化和动态背景; 鲁棒子空间模型[37] 特别是鲁棒独立主成分分析 (Robust principal components analysis, RPCA) 能利用视频帧间的关联关系, 较好地处理时空约束, 在前背景的分离上潜力巨大, 但该类模型通常不能较好地满足实时性能, 需要更多的研究提升算法的时效性; 神经网络模型[33] 在精确性能和计算成本上寻求了一个折中, 该类模型中的空间相干性自组织背景减除法 (Spatially coherent self-organizing background subtraction, SC-SOBS)[43] 与 3D 自组织背景减除改进算法 (Enhanced 3D self-organizing background subtraction, 3dSOBS+)[44] 在 ChangeDetection.Net 数据集[45] 上取得领先的检测结果. 此外, 在该数据集上检测结果较好的算法还有统计模型中的视觉背景提取器 (Visual background extractor, ViBe)[46]、像素自适应分割 (Pixel-based adaptive segmenter, PBAS)[47] 算法等, 它们都使用了鲁棒更新模型, 能同时处理光照变化和背景动态变化等情况. Sobral 等[48] 基于 OpenCV 搭建了背景减除方法的通用框架

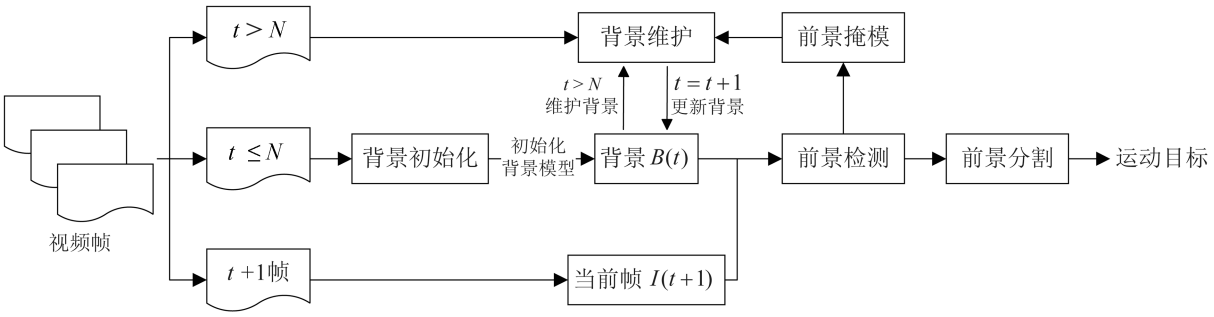

图 2　基于背景建模的目标检测流程图

Fig. 2　Flow chart of object detection based on background modeling

(Background subtraction library, BGSLibrary), 提供了 37 种背景建模算法的实现[1], 极大地促进了学术研究与工程实践.

1.2 基于前景目标建模的目标检测

基于目标建模的目标检测方法通常分为离线训练与在线检测两个阶段. 离线训练阶段对训练样本中的前景目标与背景分别进行特征表达, 建立起目标或背景表观模型, 再进行分类器训练得到分类器模型. 在线检测阶段对测试样本在多个尺度上进行滑动窗口扫描后, 采用同样的特征表达方法建立起表观模型, 然后再用离线阶段训练得到的分类器模型对其进行分类, 从而判断各个窗口是否为前景目标. 基于前景目标建模的目标检测方法的一般处理流程如图 3 所示. 与基于背景建模的方法不同, 该类方法不受场景限制, 应用范围相对较广泛, 且检测结果不需要进行再度分割.

事实上, 上述检测方法用到了目标检测、目标识别以及行为识别等任务的一个通用框架, 即 "特征表达" + "分类器" 的框架. 因此, 如何高效准确地进行特征表达以及构造适当的分类器是此类方法的核心所在. 如果所选取的表达特征能够有效地反映检测目标的本质, 那么目标检测与跟踪任务便会取得事半功倍的效果.

1.2.1 特征表达

图像的特征表达就是将原始图像像素映射到一个可区分维度空间数据的过程, 它是打破底层像素与高层语义之间的语义鸿沟至关重要的一步. 图像特征按其能否通过自学习得到, 可以分为基于人工设计的特征与基于学习的特征. 其中, 人工的设计特征是一种利用人类先验知识与智慧, 并将这类知识应用于目标检测、识别等任务的很好的方式. 这类方法实现相对容易、计算也比较简单, 但其极大地依赖了人类知识、经验的总结, 且不能做到对图像或目标模型最本质的刻画. 基于学习的特征表达主要是通过无监督学习的方式, 让机器自动地从样本中学习到表征这些样本更加本质的特征.

1) 基于人工设计的特征表达

基于人工设计的特征是人们通过手动设计提取得到的特征, 即存在一个对输入信号 "显式" 的预处理过程. 目前, 人工设计的特征按视觉特性与特征计算的不同可以大致分为四类: 梯度特征、模式特征、形状特征以及颜色特征.

梯度特征通过计算空间区域上的梯度强度和方向等的分布来描述目标. 其中, 使用最为广泛的梯度特征是 Lowe[4] 提出的尺度不变特征 (SIFT), 其通过获取特定关键点附近的梯度信息来描述目标, 通常关键点在空间上分布相对较稀疏, 该类特征具有非常卓越的尺度、旋转不变特性, 其改进特征主要有 PCA-SIFT[49]、梯度位置方向直方图 (Gradient location-orientation histograms, GLOH)[50]、加速鲁棒特征 (Speed-up robust features, SURF)[51]、雏菊花特征 (DAISY)[52] 等. 另外, Dalal 等[5] 为解决静态图像中的行人检测问题而提出的梯度直方图特征 (HOG), 也逐渐成为近年来最具影响的特征之一, 其改进特征主要有变尺寸梯度直方图 (HOG with variable size, v-HOG)[53]、共生梯度直方图 (Co-occurrence histogram of oriented gradients, CoHOG)[54]、GIST[55] 等.

模式特征是通过分析图像局部区域的相对差异而得出的一种特征描述, 通常被用于对图像纹理信息的表示. 1997 年, Jain 等[56] 将 Gabor 滤波器用于目标检测中, 取得了较好的检测结果. Ahonen 等[57] 将局部二值模式 (Local binary patterns, LBP) 用于人脸特征描述, 实现了对人脸的识别. 其中, LBP 的改进特征主要有中心对称局部二值模式

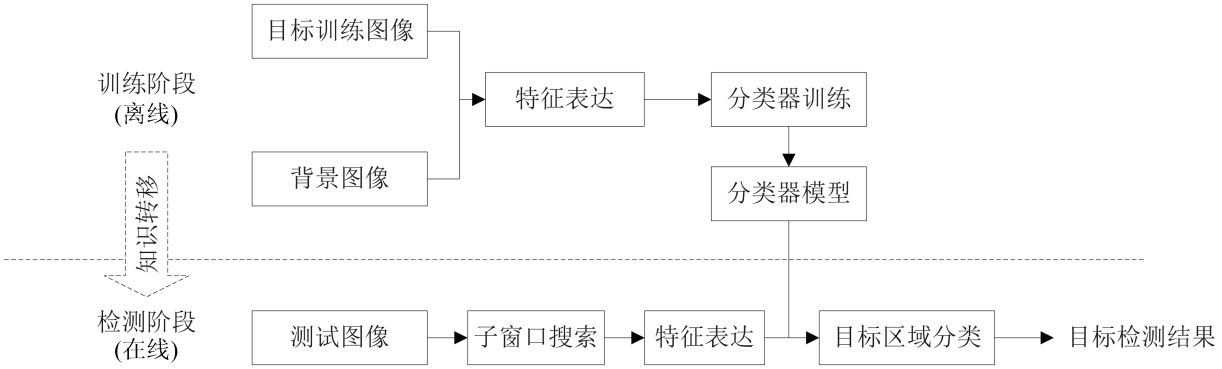

图 3 基于目标建模的目标检测流程图

Fig. 3 Flow chart of object detection based on object modeling

[1] https://github.com/andrewssobral/bgslibrary

(Center symmetric LBP, CS-LBP)[58]、无冗余局部二值模式 (Non redundant LBP, NR-LBP)[59] 等. 此外, 比较常用的模式特征还有 Haar-like[60]、统计变换直方图 (Census transform histogram, CENTRIST)[61]、姿态描述子 (Poselets)[62]、稀疏表示目标描述 (Sparselets)[63]、中层共享部分描述 (Shared mid-level parts, Shufflets)[64]、局部区域描述 (Regionlets)[65]、局部组合二值 (Locally assembled binary, LAB)[66] 等. 相对于梯度特征而言, 模式特征的数据维度较高, 因此, 在特征选择与分类学习时计算负担相对较大.

形状特征来源于基于模型的目标检测[67], 一般用于对目标轮廓的描述. 文献 [68] 对形状特征提取进行了综述工作, 常见形状特征描述有形状上下文 (Shape context)[69] 、角矩阵[70]、k 近邻分割 (k adjacent segments, kAS)[71]、图形基元[72]、线描述子 (Bunch of lines descriptor, BOLD)[73]、尺度不变形状特征[74] 等. 形状特征具有优良的尺度、旋转、平移不变特性, 可用于描述形变表观, 但其优良性能很大程度上依赖了如边界检测及分割等预处理步骤[75]. 另外, 许多不同类型的目标也可能会拥有相似的形状, 因此基于形状特征的检测方法在应用层面上受到了一定的限制. 此外, 形状特征表达忽略了纹理和颜色等有用信息, 这在一定程度上降低了其目标检测的可靠性.

颜色特征是通过计算局部的图像属性 (如灰度、颜色等) 的概率分布而得到一种特征描述, 该类特征主要包括了颜色 SIFT 特征[76]、颜色属性 (Color names)[77]、颜色共生矩阵 (Color co-occurrence matrix, CCM)[78]、颜色自相似 (Color self-similarity, CSS)[79]、局部自相似 (Local self-similarity, LSS)[80]、全局自相似 (Global self-similarity, GSS)[81]、协方差特征 (Covariance feature)[82]、C-SIFT[83]、HSV 颜色空间 SIFT 特征 (HSV-SIFT)[84]、色调直方图 SIFT 特征 (Hue-SIFT)[85] 等. 颜色特征能得到对光照不敏感的颜色表达, 近年来广泛地应用于目标检测[86]、目标跟踪[87] 等任务, 取得了很好的应用效果. 其中, 基于熵的显著性特征[88] 通过熵值的计算来获取感兴趣区域; CCM[78] 通过计算颜色的联合概率分布来获取对目标的局部表观描述; 颜色属性[77] 通过将图像像素映射至相应的属性获取对目标的颜色描述, 一般需要融合梯度特征一起来对目标进行表观建模. 颜色特征同时描述了视觉目标的亮度信息与彩色信息, 能有效地描述颜色通道类内一致的目标, 极大地提高了视觉目标的检测精度. 然而, 颜色信息及其不稳定性在一定程度上限制了颜色特征的应用, 如颜色特征不适用于对红外图像的处理. 此外, 颜色特征的融合也会增加特征维数及其计算成本. 总之, 颜色特征能很好地用于目标检测与跟踪任务中, 但在其使用的过程中需要适当地抑制其不稳定性.

基于人工设计的特征表达方法具有设计简单、实现容易、易于理解等优点, 受到科研人员的广泛研究与使用. 近年来, 众多海内外学者以及研究机构在相关方面, 已经相继开展了许多研究工作, 取得了一系列显著的研究成果, 表 3 对典型的基于人工设计

表 3 基于人工设计的特征表达方法

Table 3 Human-engineering-based feature representation methods

序号	文献	典型算法	主要思想	提出年限	方法类别
1	[4]	SIFT	通过获取特定关键点附近的梯度信息来描述运动目标, 具有旋转、尺度不变等优良特性, 其改进特征主要有 PCA-SIFT[49]、GLOH[50]、SURF[51]、DAISY[52] 等	2004	梯度特征
2	[5]	HOG	通过计算空间分布区域的梯度强度及其方向信息来描述运动目标, 其改进特征主要有 v-HOG[53]、CoHOG[54]、GIST[55] 等	2005	
3	[56]	Gabor	利用 Gabor 滤波器对图像卷积得到, 在一定程度上模拟了人类视觉的细胞感受野机制	1997	模式特征
4	[57]	LBP	通过计算像素点与周围像素的对比信息, 获得的一种对光照不变的局部描述, 其改进特征主要有 CS-LBP[58]、NR-LBP[59] 等	2004	
5	[60]	Haar-like	通过计算相邻矩形区域的像素和之差来描述线性、边缘、中心点以及对角线特征, 其改进特征主要有 LAB[66] 等	2001	
6	[6]	DPM	其实质是一种弹性形状模型, 是通过将梯度直方图 (HOG) 特征与 Latent SVM 相结合而训练得到的一种目标形状描述模型	2010	形状特征
7	[69]	Shape context	通过获取形状上某一参考点与其余点的距离分布来描述目标轮廓	2002	
8	[71]	kAS	使用一组近似线性的线段对目标形状进行描述, 具有平移、尺度等不变特性	2008	
9	[77]	Color names	通过将图像像素值映射至相应的语义属性来对目标进行描述, 该特征通常包含 11 种语义属性, 一般需要结合梯度特征一起使用	2009	颜色特征
10	[88]	基于熵的显著性特征	通过计算图像像素的灰度概率分布来获取目标的感兴趣区域	2004	

的特征表达方法进行了归纳与总结.

基于人工设计的特征表达充分利用了人类知识与智慧, 这类特征已经能够很好地应用于目标检测、识别等任务, 但它们对于目标的描述存在着一个 "显式" 的处理过程, 对目标的刻画不够本质. 根据神经科学关于哺乳动物的信息表达的研究[89−90] 表明, 哺乳动物大脑中关于执行识别等任务的大脑皮层并没有一个对信号进行 "显式" 预处理的过程, 而是将输入信号在一个大脑的复杂的层次结构中传播, 通过每一层次对输入信号进行重新提取和表达, 最终让哺乳动物感知世界. 而基于学习的特征表达就是通过无监督学习的方式让机器自动地, 从样本中学习到表征样本更加本质的特征, 从而使计算机模拟人脑感知视觉信号的机制, 实现目标检测与跟踪等视觉功能.

2) 基于学习的特征表达

基于学习的特征表达主要是采用无监督学习的方法, 使机器自动学习刻画样本更加本质的特征. 其中, 该类方法中最常用的是基于深度学习的特征表达方法, 其通过逐层地构建一个多层网络, 使机器自动地学习隐含在数据内部的关系. 基于深度学习的特征表达一直是目标检测的研究热点之一, 为了促进该领域的研究与发展, 相继出现了一系列的深度学习开源平台, 如 2014 年 Jia 等[91] 使用 C++ 搭建了用于快速特征提取的深度学习框架 (Caffe), 并封装了 Python 和 Matlab 接口, 广泛用于计算机视觉、语义理解等应用中; Google 在第一代深度学习系统 (DistBelief)[92] 的基础上对网络架构等进行优化, 推出了第二代深度学习系统 (TensorFlow)[93] 使其学习速度更快、精度更高, 并在 2015 年 11 月将该系统宣布开源, 支持卷积神经网络 (Convolutional neural network, CNN)、递归神经网络 (Recurrent neural network, RNN) 以及长短期记忆单元 (Long short-term memory, LSTM) 等算法. 此外, 使用较多的工具还有 Torch7[94]、CudaConvNet[95]、MatConvNet[96]、Pylearn2[97]、Theano[98] 等.

基于深度学习的特征表达按其构成单元的不同, 一般可以分为基于限制玻尔兹曼机 (Restricted Boltzmann machine, RBM)[99]、基于自编码机 (Auto encoder, AE)[100] 和基于卷积神经网络 (Convolutional neural network, CNN)[101] 的特征表达方法. 其中, RBM 是深度置信网络 (Deep belief nets, DBN) 的基本单元, 基于玻尔兹曼机的特征表达通过使学习到的模型产生符合条件的样本的概率最大进行特征表达. 基于自编码机的特征表达将输入信号进行编码得到表达特征, 将该特征输入解码器后得到的重构信号与原始信号满足最小残差的约束. 基于卷积神经网络的特征表达通过应用不同的卷积核, 提取不同的观测特征, 同时引入了子采样过程进行特征降维.

a) 基于限制玻尔兹曼机的特征表达

RBM 是一个双层的无向图模型, 它是玻尔兹曼机的简化模型. RBM 的示意图如图 4 所示, 其中 v 是可见层单元, h 是隐层单元, 其层内单元间没有连接关系, 层间单元呈全连接关系, 这种层间独立的条件使得 RBM 的训练显得十分高效[99].

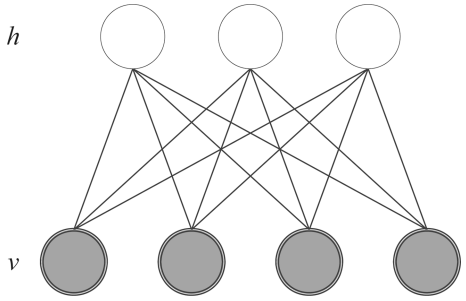

图 4 限制玻尔兹曼机

Fig. 4 Restricted Boltzmann machine

将 RBM 逐层叠加, 就构成了 DBN, 底层的输出特征作为上一层的输入信号, 对每层分别使用对比散度的方法单独进行训练. 2006 年, Hinton 等[99] 提出了 DBN 的高效训练方法. Lee 等[102] 将卷积概念引入 RBM, 提出了卷积深度置信网络 (Convolutional deep belief network, CDBN) 用于分层的特征表达, 得到很好的高层视觉特征. 受 Lee 等[102] 的启发, Nair 等[103] 将生成式梯度与判别式梯度相结合, 提出一种 3 阶 RBM 的高层模型, 用于对 3D 目标的识别中. Eslami 等[104] 将深度玻尔兹曼机 (Deep Boltzmann machine, DBM)[105] 引入对目标形状的描述, 提出形状玻尔兹曼机 (Shape Boltzmann machine, SBM), 消除了背景中与目标形状相似的物体的干扰.

b) 基于自编码机的特征表达

基于自编码机的特征表达方法通过对输入信号的自动编码, 能有效地去除冗余信息, 获取输入信号的主要信息, 使输入信号得以更加简洁的表达. 自编码机的主要结构如图 5 所示, 其中 v 是可见的输入层单元, h 是特征隐层单元. 将自编码机逐层堆叠, 就构成了深度自编码机 (Stacked auto-encoder). 与 DBN 一样, 深度自编码机的底层输出特征作为上一层的输入信号, 每一层单独地进行优化.

文献 [106] 将 K 均值聚类、稀疏编码、主成分分析也理解为一个自编码机. Xiong 等[107] 使用多堆叠自编码机 (Multiple stack auto-encoders, MSAE) 来模拟人类视觉, 从不同角度对识别目标进行深层特征表达. Yin 等[108] 采用单层的稀疏自编码

机 (Sparse auto-encoder, SAE) 提取场景特征, 并结合 SVM 对场景进行分类. Bai 等[109] 对彩色与深度图像分别采用 SAE 特征表达, 并利用递归神经网络 (Recurrent neural network, RNN) 对学习特征进行降维, 最终学习到鲁棒的分层表达特征, 用于 RGB-D 图像的目标检测. Su 等[110] 将稀疏自编码用于深度图像中的人体检测, 学习到了能表征人体内在结构的特征.

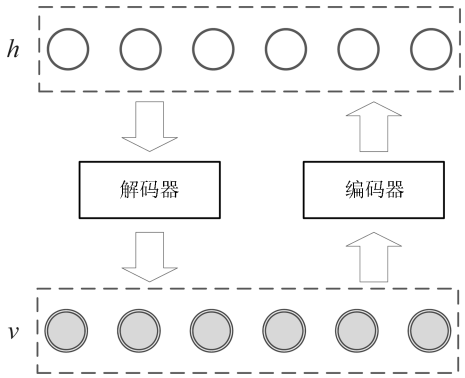

图 5 基于自编码机的特征表达

Fig. 5 Feature representation based on auto-encoder

c) 基于卷积神经网络的特征表达

单层卷积神经网络包含了卷积与子采样 2 个过程, 其实现过程如图 6 所示. 其中, 卷积过程通过引入不同的卷积核提取信号的不同特征, 实现对输入信号特定模式的观测; 子采样过程主要用于对特征图的降维, 通常采用平均池化或最大值池化操作, 该过程虽然降低了特征图的分辨率, 但能较好地保持高分辨率特征图的特征描述.

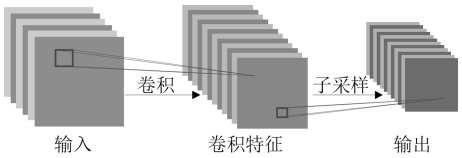

图 6 单层卷积神经网络

Fig. 6 Single layer convolutional neural network

为方便与前两种基本构成单元相对比, 将上述过程等价为图 7. 其中, 每两个节点之间的连线, 表示从输入节点经历卷积、子采样变为输出节点的过程. 根据 Hubel 等[111] 对猫视觉皮层的研究表明, 视觉系统对外界的认知是局部感知的. 受此生物学研究启发, CNN 也采用了局部连接, 这与层间全连接的 RBM 有所不同. CNN 采用这种局部连接与权值共享的思想, 极大地减少了网络参数, 加快了训练速度, 使其对网络的训练变得切实可行.

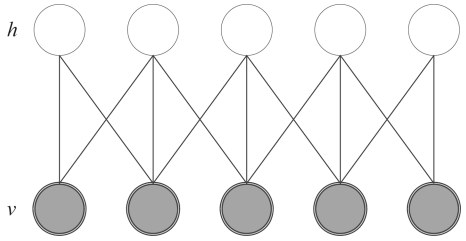

图 7 基于单层卷积神经网络的特征表达

Fig. 7 Feature representation based on single layer CNN

将单层卷积神经网络进行逐层堆叠, 就构成了 CNN, 底层的输出特征作为上一层的输入信号. Donahue 等[112] 提出了深度卷积激活特征 (Deep convolutional activation feature, DeCAF) 用于通用的视觉识别. Girshick 等[113] 将大容量的卷积神经网络 (CNN) 应用于自下而上的区域方法, 提出了基于区域的卷积神经网络 (Regions with CNN features, R-CNN), 并基于 Caffe 平台实现了对目标的精确检测与语义分割. Girshick 等将空间金字塔池化网络 (Spatial pyramid pooling based neural network, SPPNet)[114] 用于 R-CNN, 对其进行加速提出了 Fast R-CNN[115], 在计算速度和准确度上均有所提高, 而后在此基础上又提出了 Faster R-CNN[116]. Zhu 等[117] 将上下文信息引入深度卷积神经网络中提出了 segDeepM 模型, 在 PASCAL VOC 2010 数据集[7] 上检测精度比 R-CNN 高 4.1 %. Han 等[118] 使用深度卷积神经网络提取特征成功用于 MatchNet 中.

基于学习的特征表达方法受到了广泛的关注和研究, 表 4 对该类特征表达方法进行了归纳总结. 与人工设计的特征相比, 由于深度学习特征是通过构建深层的网络结构, 直接从原始图像像素中提取得到, 故其将特征设计问题转换为了网络架构问题. 这种自学习的方法极大地减少了不必要的特征设计细节, 同时深度神经网络的高层特征映射也显示出了一定的语义属性, 在 PASCAL VOC[119]、ImageNet 大规模视觉识别挑战赛 (ImageNet Large Scale Visual Recognition Challenge, ILSVRC)[120] 等相关国际赛事中, 基于深度学习的模型取得了最好的效果, 这都体现了深度学习的强大学习能力. 虽然深度学习特征表达具有更本质的特征表现, 但由于学习深层神经网络涉及了大量的参数, 网络的训练需要大量的数据, 因此计算过程比较繁重, 需要进一步优化.

1.2.2 分类器

文献 [131] 对现有的分类器进行了详细的阐述, 其中, 支持向量机 (SVM) 是使用最为广泛的分类

表 4 基于学习的特征表达方法

Table 4 Learning-based feature representation methods

类别	方法名称
基于深度学习的特征表达	CDBN[102], SBM[104], DeCAF[112], R-CNN[113], SPPNet[114], Fast R-CNN[115], Faster R-CNN[116], segDeepM[117], MatchNet[118], OverFe-at[121], NIN[122], GoogLeNet[123], VGGNet[124], DeepID-Net[125], Vox-Net[126], SuperCNN[127], MDNet[128], DeepSRDCF[129], SODLT[130]

器之一. 基于对本文篇幅的考量, 这里仅对 SVM 分类器进行简要的阐述, 对其他分类器将不作赘述, 有兴趣的读者可以参考相关文献. 目前, SVM 在数据分类任务上取得了很好的效果, 尤其是引入了核方法的 SVM. 其中, 合理的核函数选取将有效地改善分类器性能, 常用的核函数主要有 Linear、Sigmoid、RBF、GussianRBF 等[132].

随着时代的发展, 更多的核方法相继被提出, 如 Lu 等[133] 将空间失配核 (Spatial mismatch kernels, SMK) 引入 SVM 对图像进行分类, Lazebnik 等[134] 提出了空间金字塔匹配核 (Spatial pyramid matching, SPM) 并将其用于自然场景分类问题. Yang 等[135] 通过学习过完备稀疏特征, 将稀疏编码 (Sparse coding, SC) 与 SPM 相结合提出了 ScSPM 方法, 使用线性 SVM 达到了当时最好的分类效果. 然而, 稀疏编码对特征变化非常敏感[136], 且稀疏编码过程忽略了局部特征之间的相关性. 针对上述问题, Gao 等[137] 通过引入正则化项, 保证了相似的局部特征之间稀疏编码的一致性, 提出了 LScSPM 方法, 使分类精度得以提升. 核方法的使用使 SVM 分类性能得以极大提升, 然而, 其最大的优点也是其最致命的弱点, 分类器的分类性能也会极大地依赖核的选取. 因此, 如何根据实际需求选取合适的核方法仍需进一步探索.

尽管基于核的 SVM 已广泛地应用于分类任务中, 但随着对图像分类的研究的进展, 目前所使用的视觉单词大小越来越大, 相应的图像表达数据维度与日俱增, 传统的单个分类器已经难以满足高维数据的分类要求. 普遍的做法是将多个分类器集成在一起[138], 得到一个分类性能更优良的强分类器, 常用的方法主要有 Bagging、Boosting 以及随机森林等. 然而, 集成分类器方法也存在一些问题尚未有定论, 如子分类器如何选取与组合、如何在不降低分类性能的情况下使分类器数量尽可能少等. 因此, 研究子分类器模型的产生、调整以及整合, 将有助于适应高维数据的分类任务.

1.3 小结

目标检测的目的是从不同复杂程度的环境背景中分离出运动目标, 本小节按算法所处理的数据对象的不同, 将其分为基于背景建模和基于目标建模的检测方法. 其中, 基于背景建模的检测方法一般具有实现简单、运算效率较高等优点, 但其适用范围比较狭小, 通常只能用在固定摄像机拍摄的场景下, 且固定场景也存在诸多干扰因素, 如光照变化、阴影、局部动态背景 (摇晃的树枝、波动的水面) 等, 这些因素都将极大地影响算法的性能, 给前景提取带来巨大的挑战.

基于目标建模的检测方法克服了基于背景建模的方法固定场景的缺点, 可以用于动态环境 (如车载摄像头等) 的目标检测, 且一般检测出的运动区域不需再度进行分割, 扩展了目标检测的应用范围. 但其在应用过程中也存在诸多的挑战, 如较大的遮挡与光照变化, 较小的类间差与较大的类内差, 较大的目标形变与尺度变化, 较低的图像分辨率等. 另外, 特征学习需要依赖大量的学习样本, 且不同场景需要训练不同的分类器, 在实际应用中一般较难满足实时要求. 研究者仍然需要从特征表达与分类器设计上进行思考, 提高算法的精确度、鲁棒性, 同时也不能忽略算法的时效性能. 目前, 在目标检测领域已经公开发布了许多可供算法评测的数据集, 表 5 对典型的数据集及其特点进行了简要的归纳.

2 目标跟踪

运动目标跟踪问题可以等价为在连续的图像帧之间, 构建基于目标位置、速度、形状、纹理、色彩等有关特征的对应匹配问题. 其一般处理流程如图 8 所示, 由目标状态初始化、表观建模、运动估计及目标定位 4 部分组成, 其中 N 表示用于跟踪初始化的视频帧数.

目标状态的初始化一般采用人工标定或目标检测的方法进行实现. 表观建模主要包括了对目标的视觉特征 (颜色、纹理、边缘等) 的描述, 以及如何度量视觉特征之间的相似性, 它是实现鲁棒跟踪的关键所在. 运动估计则是采用某种运动假设来预估目标可能出现的位置, 常用的运动估计方法主要有线性回归[146]、均值漂移[147]、隐马尔科夫模型[148]、卡尔曼滤波[149] 以及粒子滤波[150] 等. 最后, 在表观建模与运动估计的基础上, 采用某种最优化策略获取目标最可能的位置, 实现对跟踪目标的定位.

表 5 目标检测典型数据集

Table 5 Typical data sets for object detection

序号	参考文献	数据集名字	数据规模	是否标注	特点及描述	主页链接	发布时间
1	[139]	MIT CBCL Pedestrian Database	共 924 张图片, 64 × 128, PPM 格式	否	人体目标处于图像正中间, 且图像视角限定为正向或背向	http://cbcl.mit.edu/software-datasets/PedestrianData.html	2000
2	[140−141]	USC Pedestrian Detection Test Set	共 359 张图片, 816 个人	是	包含单视角下无遮挡、部分遮挡以及多视角下无遮挡的行人检测数据	http://iris.usc.edu/Vision-Users/OldUsers/bowu/DatasetWebpage/dataset.html	2005 / 2007
3	[5]	INRIA Person Dataset	共 1 805 张图片, 64 × 128	是	包含了各种各样的应用背景, 对行人的姿势没有特别的要求	http://pascal.inrialpes.fr/data/human/	2005
4	[45, 142]	ChangeDetection.Net	共 51 段视频, 约 140 000 帧图片	是	包含了动态背景、目标运动、夜晚及阴影影响等多种挑战	http://changedetection.net/	2012 / 2014
5	[143]	Caltech Pedestrian Dataset	10 小时视频, 640 × 480	是	视频为城市交通环境下驱车拍摄所得, 行人之间存在一定的遮挡	http://www.vision.caltech.edu/Image_Datasets/CaltechPedestrians/	2009
6	[144]	CVC Datasets	共 9 个数据集	部分标注	提供了多种应用场景, 如城市、红外等场景, 行人间存在部分遮挡	http://www.cvc.uab.es/adas/site/?q=node/7	2007 / 2010 / 2013 ∼ 2015
7	[119]	PASCAL VOC Datasets	共 11 540 张图, 含 20 个类	是	该比赛包括分类、检测、分割、动作分类以及人体布局检测等任务	http://host.robots.ox.ac.uk/pascal/VOC/	2005 ∼ 2012
8	[120]	ImageNet	共 14 197 122 张图片	是	大规模目标识别比赛, 包括目标检测、定位以及场景分类等任务	http://image-net.org/	2010 ∼ 2015
9	[145]	Microsoft COCO	约 328 000 张图片, 含 91 个类	是	自然场景下的图像分类、检测、场景理解等, 不仅标注了不同的类别, 还对类中个例进行了标注	http://mscoco.org/	2014

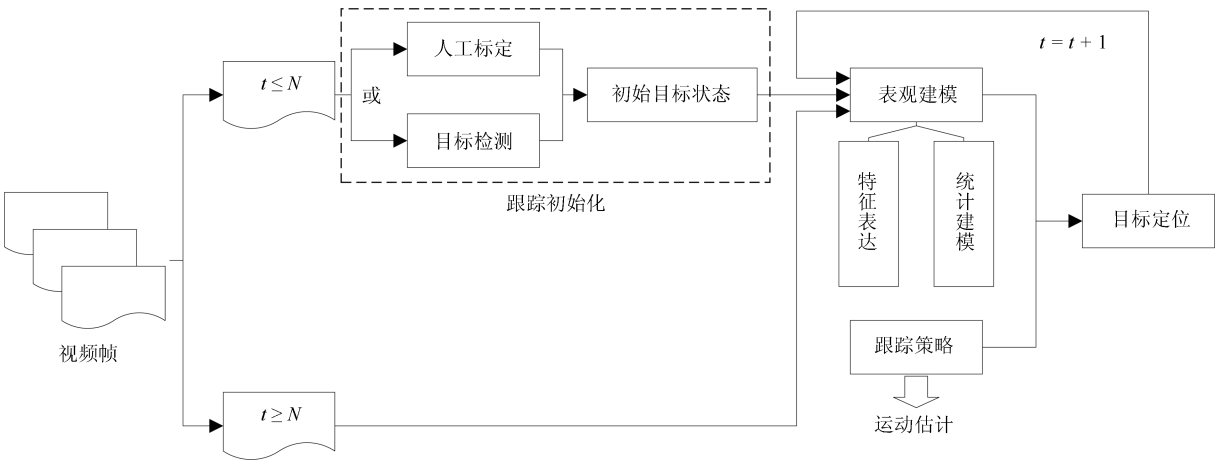

图 8 运动目标跟踪一般流程

Fig. 8 Flow chart of moving object tracking

如图 8 所示, 表观建模可以分为特征表达与统计建模, 关于特征表达在上一小节中已经作了详细的阐述, 这里将不再进行赘述. 一般地, 目标跟踪按有无检测过程的参与, 可以分为生成式跟踪与判别式跟踪. 其中, 生成式跟踪方法是在目标检测的基础上, 对前景目标进行表观建模后, 按照一定的跟踪策略估计跟踪目标的最优位置. 判别式跟踪方法则是通过对每一帧图像进行目标检测来获取跟踪目标状态, 因此这类方法也常被称为基于检测的跟踪方法.

生成式跟踪方法采用一定的跟踪策略, 估计下一帧中跟踪目标的状态, 其跟踪过程与检测过程是相互独立的, 二者有一定的时间先后顺序. 判别式

跟踪方法将跟踪问题看作前景与背景的二分类问题, 通过学习分类器, 在当前帧中搜索与背景区分最大的前景区域, 其跟踪过程与检测过程彼此联系, 二者是同时进行的.

2.1 生成式跟踪方法

生成式跟踪方法假设跟踪目标可以由某种生成过程所描述, 其目标是从众多候选目标中寻找最优的候选目标. 生成式跟踪方法的关键在于如何精确地对跟踪目标进行重构表达, 采用在线学习方法对跟踪目标进行表观建模以适应目标表观的变化, 实现对目标的跟踪. 目前, 生成式表观模型的建立可以分为基于核的方法、基于子空间的方法以及基于稀疏表示的方法三类.

2.1.1 基于核的方法

基于核的方法通常采用核密度估计的方式构建表观模型, 并使用 Mean shift 方法[147] 对运动目标位置进行估计. Comaniciu 等[151] 使用颜色直方图建立表观模型, 采用各向同性的核函数施加空间平滑, 通过 Mean shift 寻找局部极值对目标进行定位. Rahmati 等[152] 使用颜色直方图建立跟踪目标表观模型的同时, 对基于核的 Mean shift 方法进行改进, 实现了对婴儿四肢的跟踪以辅助其运动. Melzer 等[153] 对典型相关分析 (Canonical correlation analysis, CCA) 进行非线性扩展, 提出了基于核的典型相关分析 (Kernel-CCA) 方法用于表观模型的构建, 并将其应用于目标姿态估计与运动跟踪.

早期的基于核的方法虽然考虑了跟踪目标的颜色及灰度等信息, 但其忽略了如梯度、形状等其他重要的视觉信息, 在复杂的场景、目标的部分遮挡、快速运动以及尺度变化等情况下容易出现漂移问题. 为了解决目标尺度的自适应问题, 研究者提出了一系列方法, 如 Yilmaz[154] 将非对称核引入 Mean shift 方法, 实现了对跟踪目标的尺度自适应以及方向的选择. Hu 等[155] 通过计算主成分的协方差矩阵来更新跟踪目标的方向, 并使用相关特征值检测目标的尺度变化, 实现了对跟踪目标尺度及方向的自适应估计.

2.1.2 基于子空间的方法

基于子空间的方法的关键在于如何构建相关的基以及它们所张成的子空间, 对目标表观进行表示. Levey 等[156] 与 Brand[157] 采用增量奇异值分解 (Singular value decomposition, SVD) 的方法获取子空间学习的解, 将其应用于计算机视觉处理以及音频特征提取中. De 等[158] 提出了一种鲁棒子空间学习 (Robust subspace learning, RSL) 的通用框架, 该框架适用于各类线性学习问题, 如特征分析、运动结构获取等. Li[159] 结合子空间学习提出了一种增量 PCA 方法, 并将其推广到了鲁棒 PCA 方法, 极大地提高了算法效率. Skocaj 等[160] 将加权增量学习用于子空间学习中, 分别对人脸表观与动态背景进行建模, 均取得了较好的识别效果.

上述基于增量 PCA 的方法在子空间学习过程中, 样本均值不能及时地得到在线更新. 针对该问题, Ross 等[161] 考虑了样本均值的在线更新, 其提出的跟踪算法能够增量地学习低维子空间, 在线地自适应目标表观的变化. Wang 等[162] 使用偏最小二乘 (Partial least squares, PLS) 分析来学习低维可区分的特征子空间, 通过表观模型的在线更新, 减轻了跟踪漂移问题. Li 等[163] 提出了一种高效的在线张量子空间学习算法, 其通过增量地学习一个低阶的张量特征子空间建立表观模型, 考虑了样本均值和特征基的自适应更新. Wen 等[164] 针对传统的张量表示方法易受光照变化影响的问题, 给出了一种加权张量子空间 (Weighted tensor subspace, WTS) 方法, 增量地学习光照变化, 适应了跟踪过程中目标表观的变化.

相对于训练数据处于同一线性子空间的子空间模型, 研究者还尝试了利用非线性子空间的方法对目标表观进行建模. Khan 等[165] 在 Grassmann 流形上对跟踪目标进行非线性动态建模, 解决了跟踪目标在部分遮挡情况下的表观模型更新问题. Chin 等[166] 使用核独立主成分分析 (Kernel principal component analysis, KPCA) 构建非线性子空间模型, 并在此基础上提出了增量计算方法, 解决了标准 KPCA 方法不能进行在线处理的问题.

2.1.3 基于稀疏表示的方法

基于稀疏表示的方法通常假设跟踪目标在一个由目标模板所构成的子空间内, 其跟踪结果是通过寻求与模板重构误差最小而得到的最佳候选目标. Mei 等[167] 通过对重构系数引入稀疏约束, 获取跟踪目标的表观模型, 实现了对目标的跟踪. Li 等[168] 为提高 Mei 等[167] 的方法的时效性, 将压缩感知理论引入到跟踪目标的表观模型建立中, 极大地提高了算法速度, 达到了实时跟踪的性能要求.

在 Mei 等[167] 的工作基础上, Jia 等[169] 将跟踪目标进行局部分块处理, 采用局部稀疏表示与对齐池化对目标表观建模, 其跟踪结果对局部遮挡和光照变化具有较好的鲁棒性, 极大地提高了跟踪精确. Dong 等[170] 将联合稀疏表示引入图像的多特征融合, 建立起对目标的多特征表观描述, 并在粒子滤波框架下进行视觉跟踪. Hu 等[171] 在全局模板集中引入稀疏权重约束动态选取相关模板, 将多特征联合稀疏表示用于遮挡情况下的多目标跟踪.

Zhang 等[172] 认为大多数基于稀疏表示的跟踪

方法, 仅考虑了多特征融合或局部表观建模, 而忽略了候选目标的内在结构. 为此他们提出了结构稀疏跟踪器 (Structural sparse tracking, SST), 充分地利用了候选目标的内在结构及其局部分块间的空间布局信息, 极大地提高了跟踪精度. Zhong 等[173] 提出了基于稀疏表示混合模型的跟踪方法, 综合利用了全局模板和局部表达, 能高效地处理目标表观变化, 解决跟踪漂移问题. Bai 等[174] 采用块正交匹配追踪算法 (Block orthogonal matching pursuit, BOMP) 对结构稀疏表观模型进行求解, 降低了计算成本.

Zhang 等[175] 将在线学习引入稀疏表示, 分别处理跟踪过程中目标与背景的可区分性和目标表观变化的鲁棒性, 也取得了鲁棒的跟踪结果. 另外, 基于字典学习的方法[176−177] 也被广泛地运用于目标跟踪中. 文献 [10] 对 2013 年以前的基于稀疏表示的目标跟踪方法进行了综述工作, 有兴趣的读者可以参考相关文献.

生成式跟踪方法使用了丰富的图像表示, 能精确地拟合目标的表观模型. 然而, 由于实际应用中跟踪目标通常没有特定的表观形式, 因此对此类方法的正确性的验证显得极其困难. 同时, 该类方法忽略了背景信息, 当场景中出现与目标表观相似的物体时, 跟踪算法极易受到干扰, 出现跟踪失败. 为能充分地利用背景信息, 克服生成式跟踪方法的不足, 通常采用判别式跟踪方法.

2.2 判别式跟踪方法

判别式跟踪方法将视觉目标跟踪视为一个二分类问题, 其基本思路是寻求跟踪目标与背景间的决策边界. 判别式跟踪方法通常采用在线增量学习的方法, 获取前景目标与背景的分界面, 降低计算成本, 提升计算效率. 由于该方法通常是对每一帧图像进行目标检测来获取目标状态, 因此这类方法也常被称为基于检测的跟踪方法. 目前, 判别式跟踪方法可以分为基于在线 Boosting 的方法、基于支持向量机的方法、基于随机学习的方法以及基于判别分析的方法 4 类.

2.2.1 基于在线 Boosting 的方法

基于在线 Boosting 的方法[178] 来源于 Valiant 提出的 PAC 学习模型[179], 其基本思路是通过对弱分类器进行重新整合来提升分类性能. 由于该类方法具有较强的判别学习能力, 因此其已广泛地应用于目标跟踪任务中[180]. 一般地, 此类算法通过自适应地选择区分性较强的特征, 根据目标的变化, 自适应地改变分类器完成跟踪任务.

Liu 等[181] 通过初始化一个弱分类器集合, 将梯度特征选择整合到了在线 Boosting 的学习框架下, 用于构建判别式表观模型, 极大地提升了算法的效率. 然而, 该类方法没有考虑特征之间的相关性, 容易造成其所选特征具有极大的冗余性, 且该类方法不能较好地利用不同特征之间的互补性. 解决此类问题的常见做法是为候选特征引入一个加权的策略, 对其特征进行不同程度的加权.

Avidan[182] 通过对弱分类器的特征加权, 重新整合弱分类器对像素进行分类, 其最大的不足在于该方法需要对特征池中的所有特征进行计算和存储, 因此其具有较高的计算复杂度. Parag 等[183] 对弱分类器进行改进, 使其能自适应环境的变化, 提出了一种新的特征加权算法, 在实验中取得了较好的跟踪效果. 然而, 这类特征加权的方法通常需要固定弱分类器的个数, 故其应用过程不够灵活.

实际应用中通常需要对弱分类器数量进行动态调整, 以适应表观变化下的自适应跟踪任务. Visentini 等[184] 将弱分类器进行动态整合, 打破了传统方法不能自适应表观变化的局限. 随着研究的推进, 一些研究者开始将粒子滤波引入特征选择中, 以提高目标跟踪算法效率. Okuma 等[185] 将混合粒子滤波与 Adaboost 相结合, 提出了一种级联粒子滤波器用于多目标检测与跟踪. Wang 等[186] 通过在特征选择过程中嵌入粒子滤波, 同时引入 Fisher 判别准则在线地选取区分性强的特征, 对跟踪目标进行表观建模.

2.2.2 基于支持向量机的方法

基于 SVM 的方法通过引入最大化分类间隔约束, 学习到具有较强分类性能的 SVM 分类器, 对目标与非目标进行划分, 最终实现对运动目标的跟踪. Avidan[187] 将 SVM 分类器与基于光流的跟踪方法相结合, 提出了支持向量跟踪器 (Support vector tracking, SVT) 对车辆目标进行跟踪, 并在分类阶段采用了由粗到精的方法, 解决了跟踪目标发生较大运动的问题. Williams 等[188] 将稀疏贝叶斯学习与基于核的 SVM 相结合, 提出一种概率表观模型用于目标定位. Tian 等[189] 通过对多个 SVM 分类器进行加权, 整合了多个线性 SVM 对目标表观进行建模, 实现了复杂场景下的目标跟踪. 尽管该方法十分简单, 但由于其能高效地更新线性分类器, 且能较好地利用历史信息, 故其对较大的表观变化也具有良好的鲁棒性.

上述表观模型的更新, 通常是在上次跟踪结果附近, 启发式地获取正负样本数据进行训练得到的, 其训练样本的选取策略存在着极大的不确定性、随机性、不可靠性. 为了尽可能地提高样本选取的可靠性, 研究者提出了一系列相应的策略, 可以分为基于结构化输出的策略[190] 和基于 Ranking SVM 的

策略[191]. Hare 等[192] 结合核 SVM 提出了一种基于结构化输出预测的框架 (Struck), 该框架能够非常容易地整合其他附加特征以及不同的核最终提升跟踪性能. Yao 等[190] 针对目标形变与部分遮挡问题, 使用潜变量对未知部分进行建模, 避免了较复杂的初始化过程. Bai 等[191] 将跟踪问题看作弱分类器排序问题, 提出了在线拉普拉斯 Ranking SVM 跟踪器, 实现了对运动目标的鲁棒跟踪.

Tang 等[193] 提出了一种在线半监督学习框架, 使用协同训练方法进行新数据的分类以及分类器的更新, 充分地利用了无标记数据. Zhang 等[194] 为解决前背景分界线模糊问题, 使用了混合支持向量机 (Hybrid SVMs) 进行表观建模, 有效地避免了漂移问题. Zhang 等[195] 结合压缩感知理论与增量 LS-SVM, 有效地获取了上下文信息, 提高了跟踪精度.

2.2.3 基于随机学习的方法

基于随机学习的方法[196] 通过对随机特征与输入的选取建立跟踪目标的表观模型, 典型的方法主要有在线随机森林[197]、MIForests[198] 与随机朴素贝叶斯[199] 等.

Wang 等[200] 将随机森林同时用于在线学习分类与视觉目标跟踪中, 在 UCI 数据集[2]上进行实验取得了鲁棒的实验结果. 与随机森林相比, 随机朴素贝叶斯在训练阶段实时性能更高. Godec 等[199] 利用随机朴素贝叶斯具有较低时间与空间复杂度的优点, 提出了基于随机朴素贝叶斯的目标跟踪方法, 提升了视频处理的时效性能. Leistner 等[198] 结合多示例学习 (Multiple instance learning, MIL) 分类器的优点, 提出了基于随机树的多示例学习算法 (MIForests) 用于表观建模.

由于随机学习的方法通常可以使用 GPU 实现并行加速计算, 故相比于基于在线 Boosting 和基于 SVM 的方法, 基于随机学习的方法处理速度更快、效率更高, 且易扩展到对多分类问题的处理. 但由于该类方法的特征选取比较随机, 故在不同的应用环境下, 该类方法的跟踪性能不够稳定.

2.2.4 基于判别分析的方法

基于判别分析的方法的基本思路是通过寻找一个具有高类间差异的低维子空间对跟踪目标表观进行建模, 这类方法主要有线性判别分析、基于度量学习的判别分析以及基于图的判别分析方法.

线性判别分析是较简单、使用也较广泛的一种方法, Lin 等[201] 将目标与背景分别看作高斯分布, 提出了一种基于增量 Fisher 的线性判别分析的表观建模方法, 能较好地自适应目标外观及背景的变化. Nguyen 等[202] 对前景与背景分别进行局部纹理特征提取, 并采用线性判别分析 (Linear discriminant analysis, LDA) 对其进行判别跟踪, 实现了视角及光照变化下的鲁棒跟踪. Li 等[203] 将二维线性判别分析引入了跟踪目标的表观建模中, 并采用了矩阵形式进行计算, 极大地提高了跟踪效率.

基于度量学习的判别分析方法的基本思想是在满足样本类内距离最小、类间距离最大的约束下, 将图像原始特征空间映射至另一个可度量的空间来完成表观模型的构建. Wang 等[204] 将表观建模与视觉匹配看作是视觉跟踪的一个单目标优化问题, 基于度量学习提出了一种判别式表观模型用于视觉跟踪中. Tsagkatakis 等[205] 将在线距离度量学习 (Distance metric learning, DML) 与最近邻分类器相结合, 在缺少目标表观先验信息下, 实现了对运动目标的稳定跟踪. 然而, 上述方法在目标或背景呈现多模态分布时, 不能很好地适用于目标跟踪任务, 解决此类问题的常用思路是首先对数据进行聚类分析, 然后在每个类上分别使用判别式分析方法进行表观模型构建. Xu 等[206] 采用最近邻聚类方法将数据进行了分类, 提出了自适应子类判别分析方法 (Subclass discriminant analysis, SDA) 解决了前背景的多模态分布问题.

基于图的判别学习方法可以分为基于图嵌入以及图直推学习的方法. 其中, 基于图嵌入的方法通过将高维样本嵌入到一个具有判别能力的低维空间, 实现对运动目标的判别与跟踪. Zhang 等[207] 假设目标类样本近似于高斯分布, 背景类样本服从多模态分布, 使用基于图嵌入的判别分析方法来构造目标表观模型, 并结合增量学习自适应了目标表观以及光照的变化. 基于图直推学习的方法通过估计候选样本属于目标类的似然概率对目标或背景进行分类. 查宇飞等[208] 将跟踪问题看作是一个基于图的直推学习问题, 并以正样本和候选样本为顶点建立了一张图, 同时学习目标所在的流形以及样本的聚类结构, 提出了一种基于图直推模型的跟踪方法, 对姿态、表情及光照的变化、部分遮挡等具有良好的鲁棒性.

2.3 算法评测

目前, 能用于目标跟踪评测的公开视频序列比较多[3,4,5,6], 表 6 对比较常用的典型数据集进行了

[2] https://archive.ics.uci.edu/ml/datasets.html
[3] http://www.cvpapers.com/datasets.html
[4] http://homepages.inf.ed.ac.uk/cgi/rbf/CVONLINE/entries.pl?TAG363
[5] http://www.computervisiononline.com/datasets
[6] http://riemenschneider.hayko.at/vision/dataset/

简要归纳, 并给出了相关数据集的下载链接. 文献 [217] 在数据集 VOT2015 上对典型的跟踪算法进行了评测, 其评测结果如表 7 所列出 (性能前三的指标分别用粗体、粗体加斜体与花体显示), 所选取的评价指标共有 4 个, 即跟踪精度、平均失败数、重叠率以及跟踪速度. 其中, 为了保证算法在不同实现平台上的速度等效, 跟踪速度指标采用文献 [227] 所给出的等效滤波运算 (Equivalent filter operations, EFO) 进行衡量.

从评测结果可以明显看到, MDNet[128]、DeepSRDCF[129] 跟踪器在跟踪准确度、平均失败数以及重叠指标三个跟踪精度指标上, 分别取得了第一和第二的成绩, SODLT[130] 跟踪器在跟踪准确度上与 DeepSRDCF[129] 跟踪器并居第二, 这三个跟踪器都采用了基于 CNN 的特征表达方法来对跟踪目标进行表观建模, 这说明基于学习的特征表达方法确实获得了刻画目标更加本质的特征. 然而, 尽管它们个别使用了 GPU 进行加速, 其跟踪速度仍然是跟踪

表 6 目标跟踪典型数据集

Table 6 Typical data sets for object tracking

序号	参考文献	数据集	数据规模	是否标注	特点及描述	主页链接	发布时间
1	[209−210]	Visual Tracker Benchmark	100 段序列	是	来源于现有文献, 包括了光照及尺度变化、遮挡、形变等 9 种挑战	http://www.visual-tracking.net	2013
2	[211]	VIVID	9 段序列	是	主要任务为航拍视角下的车辆目标跟踪, 具有表观微小、相似等特点	http://vision.cse.psu.edu/data/vividEval/datasets/datasets.html	2005
3	[212]	CAVIAR	28 段序列	是	主要用于人体目标跟踪, 视频内容包含行走、会面、进出场景等行为	http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/	2003 / 2004
4	[213]	BIWI Walking Pedestrians Dataset	1 段序列	是	主要任务为鸟瞰视角下的行人跟踪, 可用于评测多目标跟踪算法	http://www.vision.ee.ethz.ch/datasets/	2009
5	[214]	“Central” Pedestrian Crossing Sequences	3 段序列	是	行人过街序列, 每 4 帧标定一次	http://www.vision.ee.ethz.ch/datasets/	2007
6	[215]	MOT16	14 段序列	是	无约束环境的多目标跟踪, 有不同视角、相机运动、天气影响等挑战	http://motchallenge.net/	2016
7	[216]	PETS2015	7 段序列	是	关于停车场中车辆旁边不同活动序列, 可用于目标检测与跟踪、动作识别、场景分析等	http://www.pets2015.net/	2015
8	[217]	VOT Challenge	60 段序列 (2015 年)	是	主要用于短视频跟踪算法的评测, 该比赛从 2013 年开始举办	http://votchallenge.net/	2013 ∼ 2015

表 7 典型跟踪算法的性能对比

Table 7 Performance comparison of typical tracking algorithms

序号	参考文献	跟踪器	准确度	平均失败数	平均覆盖率	速度 (EFO)	时间	方法类别
1	[128]	MDNet	**0.60**	**0.69**	**0.38**	0.87	2015	CNN
2	[129]	DeepSRDCF	***0.56***	***1.05***	***0.32***	0.38	2015	
3	[130]	SODLT	***0.56***	1.78	0.23	0.83	2015	
4	[218]	SumShift	*0.52*	1.68	0.23	***16.78***	2011	核学习
5	[219]	ASMS	0.51	1.85	0.21	**115.09**	2013	
6	[217]	S3Tracker	*0.52*	1.77	0.24	*14.27*	2015	
7	[161]	IVT	0.44	4.33	0.12	8.38	2008	子空间学习
8	[220]	CT	0.39	4.09	0.11	12.90	2012	
9	[221]	L1APG	0.47	4.65	0.13	1.51	2012	稀疏表示
10	[222]	OAB	0.45	4.19	0.13	8.00	2014	Online Boosting
11	[223]	MCT	0.47	1.76	0.22	2.77	2011	
12	[224]	CMIL	0.43	2.47	0.19	5.14	2010	
13	[225]	Struck	0.47	1.61	*0.25*	2.44	2014	SVM
14	[217]	RobStruck	0.48	*1.47*	0.22	1.89	2015	
15	[226]	MIL	0.42	3.11	0.17	5.99	2011	随机学习

方法中最慢的一类, 其根本原因在于它们的网络架构的设计需要计算大量复杂的参数. 相比而言, 基于核的跟踪方法在跟踪速度上具有较大的优势, 但其跟踪准确度稍逊于基于学习的方法.

2.4 小结

目标跟踪是在目标检测的基础上对运动目标的状态进行连续估计的过程. 本小节将目标跟踪按其与检测的关系的不同分为了生成式与判别式跟踪方法, 分别对其典型方法进行了归纳与总结, 并给出了常用的评测数据集与典型算法的性能对比. 其中, 生成式跟踪方法使用了丰富的图像表示, 在复杂环境中通常会得到更加精确的拟合结果. 但其在参数估计上容易受到局部极值的影响, 且该类方法忽略了背景信息, 易受背景干扰, 场景中出现与目标相似的背景时容易出现跟踪漂移; 判别式跟踪方法克服了生成式跟踪方法的缺陷, 考虑了背景信息, 对较大的遮挡和变化具有更强的鲁棒性. 然而, 判别式跟踪方法对训练样本的依赖程度相对较高, 样本的选取会极大地影响这类方法的性能表现.

3 结束语

运动目标的检测和跟踪主要用于获取运动目标的位置、姿态、轨迹等基本运动信息, 是理解服务对象或对目标实施控制的前提和基础. 在场景较固定、环境较简单时, 通常采用基于背景建模的方法, 就能够很好地进行运动目标的检测与跟踪任务. 在动态背景下, 环境较复杂时, 一般需要对运动目标进行表观建模, 实现其检测与跟踪任务.

运动目标检测与跟踪的准确性与鲁棒性, 很大程度上依赖于精确的表观建模. 传统的特征表达通常是依靠人类智慧、先验知识, 通过人工的总结设计用于视觉任务的特征, 这类特征能够较好地用于检测与跟踪任务, 且实现比较容易, 实时性能较强. 近年来, 众多研究者针对人工特征的设计难度大、表现不够本质等特性, 纷纷提出了一系列基于自学习的特征, 集中体现为基于深度学习的特征. 这类方法通过对样本的自学习, 获取更加本质的特征表达, 极大地提升了检测与跟踪的精度.

然而, 现有算法大多还只是停留在对特定场景下的运动目标的分析与建模, 较难实现对复杂自然环境下运动目标的精确检测与跟踪. 同时, 其在算法实时性能上也较难保证. 因此, 复杂自然场景下的运动目标检测与跟踪及其实时性能仍然是需要努力的目标. 笔者在这里根据自己的理解总结一下目标检测与跟踪的研究热点以及发展趋势:

1) 场景信息与目标状态的融合

场景信息包含了丰富的环境上下文信息, 对场景信息进行分析及充分利用, 能够有效地获取场景的先验知识, 降低复杂的背景环境以及场景中与目标相似的物体的干扰; 同样地, 对目标的准确描述有助于提升检测与跟踪算法的准确性与鲁棒性. 总之, 尝试研究结合背景信息和前景目标信息的分析方法, 融合场景信息与目标状态, 将有助于提高算法的实用性能.

2) 多维度、多层级信息融合

为了提高对运动目标表观描述的准确度与可信性, 现有的检测与跟踪算法通常对时域、空域、频域等不同特征信息进行融合, 综合利用各种冗余、互补信息提升算法的精确性与鲁棒性. 然而, 目前大多算法还只是对单一时间、单一空间的多尺度信息进行融合, 研究者可以考虑从时间、推理等不同维度, 对特征、决策等不同层级的多源互补信息进行融合, 提升检测与跟踪的准确性.

3) 基于深度学习的特征表达

基于深度学习的特征表达具有强大的分层自学习能力, 能深度地挖掘隐含在数据内部间的潜在关系. 其中, 基于卷积神经网络的特征表达方法效果尤为突出, 近年来取得了显著的检测效果[126−128]. 同时, 深度学习框架相继开源[92−98], 为思想的碰撞与交融带来了更多可能. 然而, 基于深度学习的特征表达方法也存在一些问题尚未定论, 如深度学习的层数以及隐层节点个数如何确定, 深度学习所学得特征的优劣如何评价等. 因此, 对基于深度学习的特征表达方法的进一步研究可能会产生突破性成果, 最终将促进该领域的发展.

4) 基于核的支持向量机分类方法

支持向量机 (SVM) 因其分类性能优良、操作实现简单等特性, 仍是目前倍受青睐的常用分类方法之一, 尤其是核方法的引入更使其性能得以极大提升. 然而, 其最大的优点也是其最致命的弱点, 其分类性能也会极大地依赖于核的选取, 尽管目前已经开展了相当一部分的工作[133, 135, 137], 但对于不同分类任务下的核方法的选取, 仍然还没有一个普遍通用的定论. 因此, 如何根据实际分类需求, 选取合适的核方法仍需进一步探索.

5) 高维数据的分类方法

随着分类任务研究的发展, 分类中所使用的视觉单词的大小不断地增加, 其相应的图像表达数据维度也与日俱增, 传统的单个分类器已经难以满足高维数据的分类要求. 目前, 普遍的做法是将多个分类器集成在一起, 以获取分类性能更好的强分类器. 然而, 集成分类器方法也存在一些问题尚未定论, 如子分类器如何选取与组合, 如何在不降低分类性能的情况下使分类器数量尽可能少等. 因此, 研究子分

类器模型的产生、调整以及整合, 将有助于适应高维数据的分类任务.

References

1 Harold W A. Aircraft warning system, U. S. Patent 3053932, September 1962

2 Papageorgiou C P, Oren M, Poggio T. A general framework for object detection. In: Proceedings of the 6th IEEE International Conference on Computer Vision. Bombay, India: IEEE, 1998. 555−562

3 Viola P, Jones M J. Robust real-time object detection. *International Journal of Computer Vision*, 2001, **4**: 51−52

4 Lowe D G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004, **60**(2): 91−110

5 Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, CA, USA: IEEE, 2005. 886−893

6 Felzenszwalb P F, Girshick R B, McAllester D, Ramanan D. Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, **32**(9): 1627−1645

7 Everingham M, Van Gool L, Williams C K I, Winn J, Zisserman A. The pascal visual object classes (VOC) challenge. *International Journal of Computer Vision*, 2010, **88**(2): 303−338

8 Rautaray S S, Agrawal A. Vision based hand gesture recognition for human computer interaction: a survey. *Artificial Intelligence Review*, 2015, **43**(1): 1−54

9 Gao Y, Spiteri C, Pham M T, Al-Milli S. A survey on recent object detection techniques useful for monocular vision-based planetary terrain classification. *Robotics and Autonomous Systems*, 2014, **62**(2): 151−167

10 Zhang S P, Yao H X, Sun X, Lu X S. Sparse coding based visual tracking: review and experimental comparison. *Pattern Recognition*, 2013, **46**(7): 1772−1788

11 Li X, Hu W M, Shen C H, Zhang Z F, Dick A, van den Hengel A. A survey of appearance models in visual object tracking. *ACM transactions on Intelligent Systems and Technology (TIST)*, 2013, **4**(4): Article No. 58

12 Gao Shi-Bo, Cheng Yong-Mei, Xiao Li-Ping, Wei Hai-Ping. Recent advances of sparse representation for object detection. *Acta Electronica Sinica*, 2015, **43**(2): 320−332
(高仕博, 程咏梅, 肖利平, 韦海萍. 面向目标检测的稀疏表示方法研究进展. 电子学报, 2015, **43**(2): 320−332)

13 Piccardi M. Background subtraction techniques: a review. In: Proceedings of the 2004 IEEE International Conference on Systems, Man and Cybernetics. The Hague, Holland: IEEE, 2004. 3099−3104

14 Bouwmans T. Traditional and recent approaches in background modeling for foreground detection: an overview. *Computer Science Review*, 2014, **11−12**: 31−66

15 Smeulders A W M, Chu D M, Cucchiara R, Calderara S, Dehghan A, Shah M. Visual tracking: an experimental survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, **36**(7): 1442−1468

16 Gowsikhaa D, Abirami S, Baskaran R. Automated human behavior analysis from surveillance videos: a survey. *Artificial Intelligence Review*, 2014, **42**(4): 747−765

17 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. 智能视频监控技术综述. 计算机学报, 2015, **38**(6): 1093−1118)

18 Yilmaz A, Javed O, Shah M. Object tracking: a survey. *ACM Computing Surveys (CSUR)*, 2006, **38**(4): Article No. 13

19 Hou Zhi-Qiang, Han Chong-Zhao. A survey of visual tracking. *Acta Automatica Sinica*, 2006, **32**(4): 603−617
(侯志强, 韩崇昭. 视觉跟踪技术综述. 自动化学报, 2006, **32**(4): 603−617)

20 Wan Ying, Han Yi, Lu Han-Qing. The methods for moving object detection. *Computer Simulation*, 2006, **23**(10): 221−226
(万缨, 韩毅, 卢汉清. 运动目标检测算法的探讨. 计算机仿真, 2006, **23**(10): 221−226)

21 Zhang Juan, Mao Xiao-Bo, Chen Tie-Jun. Survey of moving object tracking algorithm. *Application Research of Computers*, 2009, **26**(12): 4407−4410
(张娟, 毛晓波, 陈铁军. 运动目标跟踪算法研究综述. 计算机应用研究, 2009, **26**(12): 4407−4410)

22 Niu Xiang-Jie, Huang Yong-Chun. Research on detection and tracking identification algorithm of weak moving target. *Computer Simulation*, 2010, **27**(4): 245−247
(牛芗洁, 黄永春. 微弱运动目标的检测与跟踪识别算法研究. 计算机仿真, 2010, **27**(4): 245−247)

23 Gutchess D, Trajkovics M, Cohen-Solal E, Lyons D, Jain A K. A background model initialization algorithm for video surveillance. In: Proceedings of the 8th IEEE International Conference on Computer Vision. Vancouver, BC, Canada: IEEE, 2001. 733−740

24 Wang H Z, Suter D. A novel robust statistical method for background initialization and visual surveillance. In: Proceedings of the 7th Asian Conference on Computer Vision (ACCV 2006). Hyderabad, India: Springer, 2006. 328−337

25 Colombari A, Fusiello A. Patch-based background initialization in heavily cluttered video. *IEEE Transactions on Image Processing*, 2010, **19**(4): 926−933

26 Lee B, Hedley M. Background estimation for video surveillance. In: Proceedings of the Image and Vision Computing New Zealand. Auckland, New Zealand, 2002. 315−320

27 McFarlane N J B, Schofield C P. Segmentation and tracking of piglets in images. *Machine Vision and Applications*, 1995, **8**(3): 187−193

28 Bouwmans T, El Baf F, Vachon B. Statistical background modeling for foreground detection: a survey. *Handbook of Pattern Recognition and Computer Vision*. Singapore: World Scientific Publishing, 2010. 181−189

29 Bouwmans T. Recent advanced statistical background modeling for foreground detection: a systematic survey. *Recent Patents on Computer Science*, 2011, **4**(3): 147−176

30 Butler D E, Bove V M Jr, Sridharan S. Real-time adaptive foreground/background segmentation. *EURASIP Journal on Advances in Signal Processing*, 2005, **2005**: 2292−2304

31 Kim K, Chalidabhongse T H, Harwood D, Davis L. Background modeling and subtraction by codebook construction. In: Proceedings of the 2004 IEEE International Conference on Image Processing. Singapore: IEEE, 2004. 3061−3064

32 Palomo E J, Domínguez E, Luque R M, Muñoz J. Image hierarchical segmentation based on a GHSOM. In: Proceedings of the 16th International Conference on Neural Information Processing. Bangkok, Thailand: Springer, 2009. 743−750

33 De Gregorio M, Giordano M. Background modeling by weightless neural networks. In: Proceedings of the 2015 Workshops on New Trends in Image Analysis and Processing (ICIAP 2015). Genoa, Italy: Springer, 2015. 493−501

34 Toyama K, Krumm J, Brumitt B, Meyers B. Wallflower: principles and practice of background maintenance. In: Proceedings of the 7th IEEE International Conference on Computer Vision. Kerkyra, Greece: IEEE, 1999. 255−261

35 Ridder C, Munkelt O, Kirchner H. Adaptive background estimation and foreground detection using Kalman-filtering. In: Proceedings of the 1995 International Conference on Recent Advances in Mechatronics. Istanbul, Turkey: Boğaziçi University, 1995. 193−199

36 Kim W, Kim C. Background subtraction for dynamic texture scenes using fuzzy color histograms. *IEEE Signal Processing Letters*, 2012, **19**(3): 127−130

37 Bouwmans T, Zahzah E H. Robust PCA via principal component pursuit: a review for a comparative evaluation in video surveillance. *Computer Vision and Image Understanding*, 2014, **122**: 22−34

38 Cevher V, Sankaranarayanan A, Duarte M F, Reddy D, Baraniuk R G, Chellappa R. Compressive sensing for background subtraction. In: Proceedings of the 10th European Conference on Computer Vision (ECCV 2008). Marseille, France: Springer, 2008. 155−168

39 Wren C R, Porikli F. Waviz: spectral similarity for object detection. In: Proceedings of the 2005 IEEE International Workshop on Performance Evaluation of Tracking and Surveillance. Breckenridge, Colorado, USA: IEEE, 2005. 55−61

40 Baltieri D, Vezzani R, Cucchiara R. Fast background initialization with recursive Hadamard transform. In: Proceedings of the 7th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). Boston, USA: IEEE, 2010. 165−171

41 Bouwmans T, El Baf F, Vachon B. Background modeling using mixture of gaussians for foreground detection — a survey. *Recent Patents on Computer Science*, 2008, **1**(3): 219−237

42 Lin H H, Liu T L, Chuang J H. A probabilistic SVM approach for background scene initialization. In: Proceedings of the 2002 International Conference on Image Processing. Rochester, New York, USA: IEEE, 2002. 893−896

43 Maddalena L, Petrosino A. The SOBS algorithm: what are the limits? In: Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Providence, RI, USA: IEEE, 2012. 21−26

44 Maddalena L, Petrosino A. The 3dSOBS+ algorithm for moving object detection. *Computer Vision and Image Understanding*, 2014, **122**: 65−73

45 Goyette N, Jodoin P M, Porikli F, Konrad J, Ishwar P. Changedetection.net: a new change detection benchmark dataset. In: Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Providence, RI, USA: IEEE, 2012. 1−8

46 Barnich O, Van Droogenbroeck M. ViBe: a powerful random technique to estimate the background in video sequences. In: Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Taipei, China: IEEE, 2009. 945−948

47 Hofmann M, Tiefenbacher P, Rigoll G. Background segmentation with feedback: the pixel-based adaptive segmenter. In: Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Providence, RI, USA: IEEE, 2012. 38−43

48 Sobral A, Bouwmans T. *BGS Library: A Library Framework for Algorithm′s Evaluation in Foreground/Background Segmentation*. London: CRC Press, 2014.

49 Ke Y, Sukthankar R. PCA-SIFT: a more distinctive representation for local image descriptors. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, D. C., USA: IEEE, 2004. II-506−II-513

50 Mikolajczyk K, Schmid C. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, **27**(10): 1615−1630

51 Bay H, Ess A, Tuytelaars T, Van Gool L. Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 2008, **110**(3): 346−359

52 Tola E, Lepetit V, Fua P. Daisy: an efficient dense descriptor applied to wide-baseline stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, **32**(5): 815−830

53 Zhu Q, Yeh M C, Cheng K T, Avidan S. Fast human detection using a cascade of histograms of oriented gradients. In: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE, 2006. 1491−1498

54 Watanabe T, Ito S, Yokoi K. Co-occurrence histograms of oriented gradients for human detection. *Information and Media Technologies*, 2010, **5**(2): 659−667

55 Torralba A, Oliva A, Castelhano M S, Henderson J M. Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological Review*, 2006, **113**(4): 766−786

56 Jain A K, Ratha N K, Lakshmanan S. Object detection using Gabor filters. *Pattern Recognition*, 1997, **30**(2): 295−309

57 Ahonen T, Hadid A, Pietikäinen M. Face recognition with local binary patterns. In: Proceedings of the 8th European Conference on Computer Vision (ECCV 2004). Prague, Czech Republic: Springer, 2004. 469−481

58 Heikkilä M, Pietikäinen M, Schmid C. Description of interest regions with local binary patterns. *Pattern Recognition*, 2009, **42**(3): 425−436

59 Nguyen D T, Ogunbona P O, Li W Q. A novel shape-based non-redundant local binary pattern descriptor for object detection. *Pattern Recognition*, 2013, **46**(5): 1485−1500

60 Viola P, Jones M. Robust Real-time Object Detection, Technical Report CRL-2001-1, Cambridge Research Laboratory, University of Cambridge, United Kingdom, 2001

61 Wu J X, Rehg J M. CENTRIST: a visual descriptor for scene categorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, **33**(8): 1489−1501

62 Bourdev L, Malik J. Poselets: body part detectors trained using 3D human pose annotations. In: Proceedings of the 12th IEEE International Conference on Computer Vision. Kyoto, Japan: IEEE, 2009. 1365−1372

63 Girshick R, Song H O, Darrell T. Discriminatively activated sparselets. In: Proceedings of the 30th International Conference on Machine Learning (ICML-13). Atlanta, GA, USA: ACM, 2013. 196−204

64 Kokkinos I. Shufflets: shared mid-level parts for fast object detection. In: Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV). Sydney, Australia: IEEE, 2013. 1393−1400

65 Wang X Y, Yang M, Zhu S H, Lin Y Q. Regionlets for generic object detection. In: Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV). Sydney, Australia: IEEE, 2013. 17−24

66 Yan S Y, Shan S G, Chen X L, Gao W. Locally assembled binary (LAB) feature with feature-centric cascade for fast and accurate face detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Anchorage, Alaska, USA: IEEE, 2008. 1−7

67 Arman F, Aggarwal J K. Model-based object recognition in dense-range images — a review. *ACM Computing Surveys (CSUR)*, 1993, **25**(1): 5−43

68 Yang M Q, Kpalma K, Ronsin J. A survey of shape feature extraction techniques. *Pattern Recognition*. IN-TECH, 2008. 43−90

69 Belongie S, Malik J, Puzicha J. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(4): 509−522

70 Kontschieder P, Riemenschneider H, Donoser M, Bischof H. Discriminative learning of contour fragments for object detection. In: Proceedings of the 2011 British Machine Vision Conference. Dundee, Scotland: British Machine Vision Association, 2011. 4.1−4.12

71 Ferrari V, Fevrier L, Jurie F, Schmid C. Groups of adjacent contour segments for object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, **30**(1): 36−51

72 Chia A Y S, Rahardja S, Rajan D, Leung M K. Object recognition by discriminative combinations of line segments and ellipses. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). San Francisco, CA, USA: IEEE, 2010. 2225−2232

73 Tombari F, Franchi A, Di L. BOLD features to detect texture-less objects. In: Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV). Sydney, Australia: IEEE, 2013. 1265−1272

74 Jurie F, Schmid C. Scale-invariant shape features for recognition of object categories. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, D. C., USA: IEEE, 2004. II-90−II-96

75 Dhankhar P, Sahu N. A review and research of edge detection techniques for image segmentation. *International Journal of Computer Science and Mobile Computing (IJCSMC)*, 2013, **2**(7): 86−92

76 Rassem T H, Khoo B E. Object class recognition using combination of color SIFT descriptors. In: Proceedings of the 2011 IEEE International Conference on Imaging Systems and Techniques (IST). Penang, Malaysia: IEEE, 2011. 290−295

77 Van De Weijer J, Schmid C, Verbeek J, Larlus D. Learning color names for real-world applications. *IEEE Transactions on Image Processing*, 2009, **18**(7): 1512−1523

78 Vadivel A, Sural S, Majumdar A K. An integrated color and intensity co-occurrence matrix. *Pattern Recognition Letters*, 2007, **28**(8): 974−983

79 Walk S, Majer N, Schindler K, Schiele B. New features and insights for pedestrian detection. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). San Francisco, CA, USA: IEEE, 2010. 1030−1037

80 Shechtman E, Irani M. Matching local self-similarities across images and videos. In: Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Minneapolis, Minnesota, USA: IEEE, 2007. 1−8

81 Deselaers T, Ferrari V. Global and efficient self-similarity for object classification and detection. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). San Francisco, CA, USA: IEEE, 2010. 1633−1640

82 Tuzel O, Porikli F, Meer P. Human detection via classification on Riemannian manifolds. In: Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Minneapolis, Minnesota, USA: IEEE, 2007. 1−8

83 Burghouts G J, Geusebroek J M. Performance evaluation of local colour invariants. *Computer Vision and Image Understanding*, 2009, **113**(1): 48−62

84 Bosch A, Zisserman A, Muñoz X. Scene classification using a hybrid generative/discriminative approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, **30**(4): 712−727

85 Van De Weijer J, Schmid C. Coloring local feature extraction. In: Proceedings of the 9th European Conference on Computer Vision (ECCV 2006). Graz, Austria: Springer, 2006. 334−348

86 Khan F S, Anwer R M, van de Weijer J, Bagdanov A D, Vanrell M, Lopez A M. Color attributes for object detection. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Providence, RI, USA: IEEE, 2012. 3306−3313

87 Danelljan M, Khan F S, Felsberg M, van de Weijer J. Adaptive color attributes for real-time visual tracking. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus, OH, USA: IEEE, 2014. 1090−1097

88 Kadir T, Zisserman A, Brady M. An affine invariant salient region detector. In: Proceedings of the 8th European Conference on Computer Vision (ECCV 2004). Prague, Czech Republic: Springer, 2004. 228−241

89 Lee T S, Mumford D, Romero R, Lamme V A F. The role of the primary visual cortex in higher level vision. *Vision Research*, 1998, **38**(15−16): 2429−2454

90 Lee T S, Mumford D. Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A*, 2003, **20**(7): 1434−1448

91 Jia Y Q, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell T. Caffe: convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM International Conference on Multimedia. Orlando, Florida, USA: ACM, 2014. 675−678

92 Dean J, Corrado G, Monga R, Chen K, Devin M, Mao M, Ranzato M, Senior A, Tucker P, Yang K, Le Q V, Ng A Y. Large scale distributed deep networks. In: Proceedings of the 2012 Advances in Neural Information Processing Systems 25. Lake Tahoe, Nevada, USA: MIT Press, 2012. 1223−1231

93 Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z F, Citro C, Corrado G S, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Jia Y Q, Jozefowicz R, Kaiser L, Kudlur M, Levenberg J, Mane D, Monga R, Moore S, Murray D, Olah C, Schuster M, Shlens J, Steiner B, Sutskever I, Talwar K, Tucker P, Vanhoucke V, Vasudevan V, Viegas F, Vinyals O, Warden P, Wattenberg M, Wicke M, Yu Y, Zheng X Q. TensorFlow: large-scale machine learning on heterogeneous distributed systems. arXiv: 1603.04467, 2016.

94 Collobert R, Kavukcuoglu K, Farabet C. Torch7: a Matlab-like environment for machine learning. In: Proceedings of Annual Conference on Neural Information Processing Systems. Granada, Spain: MIT Press, 2011.

95 Krizhevsky A. CUDA-convnet: high-performance C++/CUDA implementation of convolutional neural networks [Online], available: http://code.google.com/p/cuda-convnet/, August 6, 2016

96 Vedaldi A, Lenc K. MatConvNet-convolutional neural networks for MATLAB. arXiv: 1412.4564, 2014.

97 Goodfellow I J, Warde-Farley D, Lamblin P, Dumoulin V, Mirza M, Pascanu R, Bergstra J, Bastien F, Bengio Y. Pylearn2: a machine learning research library. arXiv: 1308.4214, 2013.

98 The Theano Development Team. Theano: a Python framework for fast computation of mathematical expressions. arXiv: 1605.02688, 2016.

99 Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets. *Neural Computation*, 2006, **18**(7): 1527−1554

100 Hinton G E, Zemel R S. Autoencoders, minimum description length and Helmholtz free energy. In: Proceedings of the 1993 Advances in Neural Information Processing Systems 6. Cambridge, MA: MIT Press, 1993. 3−10

101 Lécun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, **86**(11): 2278−2324

102 Lee H, Grosse R, Ranganath R, Ng A Y. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In: Proceedings of the 26th Annual International Conference on Machine Learning. Montréal, Canada: ACM, 2009. 609−616

103 Nair V, Hinton G E. 3D object recognition with deep belief nets. In: Proceedings of the 2009 Advances in Neural Information Processing Systems 22. Vancouver, B. C., Canada: MIT Press, 2009. 1339−1347

104 Eslami S M A, Heess N, Williams C K I, Winn J. The shape Boltzmann machine: a strong model of object shape. *International Journal of Computer Vision*, 2014, **107**(2): 155−176

105 Salakhutdinov R, Hinton G. Deep Boltzmann machines. In: Proceedings of the 12th International Conference on Artificial Intelligence and Statistics. Clearwater Beach, Florida, USA: ACM, 2009. 448−455

106 Zheng Yin, Chen Quan-Qi, Zhang Yu-Jin. Deep learning and its new progress in object and behavior recognition. *Journal of Image and Graphics*, 2014, **19**(2): 175−184
(郑胤, 陈权崎, 章毓晋. 深度学习及其在目标和行为识别中的新进展. 中国图象图形学报, 2014, **19**(2): 175−184)

107 Xiong M F, Chen J, Wang Z, Liang C, Zheng Q, Han Z, Sun K M. Deep feature representation via multiple stack auto-encoders. In: Proceedings of the 16th Pacific-Rim Conference on Advances in Multimedia Information Processing (PCM 2015). Gwangju, South Korea: Springer, 2015. 275−284

108 Yin H P, Jiao X G, Chai Y, Fang B. Scene classification based on single-layer SAE and SVM. *Expert Systems with Applications*, 2015, **42**(7): 3368−3380

109 Bai J, Wu Y, Zhang J M, Chen F Q. Subset based deep learning for RGB-D object recognition. *Neurocomputing*, 2015, **165**: 280−292

110 Su S Z, Liu Z H, Xu S P, Li S Z, Ji R R. Sparse auto-encoder based feature learning for human body detection in depth image. *Signal Processing*, 2015, **112**: 43−52

111 Hubel D H, Wiesel T N. Receptive fields, binocular interaction and functional architecture in the cat′s visual cortex. *The Journal of Physiology*, 1962, **160**(1): 106−154

112 Donahue J, Jia Y Q, Vinyals O, Hoffman J, Zhang N, Tzeng E, Darrell T. DeCAF: a deep convolutional activation feature for generic visual recognition. arXiv: 1310.1531, 2013.

113 Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus, OH, USA: IEEE, 2014. 580−587

114 He K M, Zhang X Y, Ren S Q, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, **37**(9): 1904−1916

115 Girshick R. Fast R-CNN. arXiv: 1504.08083, 2015.

116 Ren S Q, He K M, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. arXiv: 1506.01497, 2015.

117 Zhu Y K, Urtasun R, Salakhutdinov R, Fidler S. SegDeepM: exploiting segmentation and context in deep neural networks for object detection. arXiv: 1502.04275, 2015.

118 Han X F, Leung T, Jia Y Q, Sukthankar R, Berg A C. MatchNet: unifying feature and metric learning for patch-based matching. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: IEEE, 2015. 3279−3286

119 Everingham M, Eslami S M A, Van Gool L, Williams C K I, Winn J, Zisserman A. The pascal visual object classes challenge: a retrospective. *International Journal of Computer Vision*, 2015, **111**(1): 98−136

120 Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z H, Karpathy A, Khosla A, Bernstein M, Berg A C, Li F F. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, 2015, **115**(3): 211−252

121 Sermanet P, Eigen D, Zhang X, Mathieu M, Fergus R, LeCun Y. OverFeat: integrated recognition, localization and detection using convolutional networks. arXiv: 1312.6229, 2013.

122 Lin M, Chen Q, Yan S C. Network in network. arXiv: 1312.4400, 2013.

123 Szegedy C, Liu W, Jia Y Q, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: IEEE, 2015. 1−9

124 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv: 1409.1556, 2014.

125 Ouyang W L, Luo P, Zeng X Y, Qiu S, Tian Y L, Li H S, Yang S, Wang Z, Xiong Y J, Qian C, Zhu Z Y, Wang R H, Loy C C, Wang X G, Tang X O. DeepID-Net: multi-stage and deformable deep convolutional neural networks for object detection. arXiv: 1409.3505, 2014.

126 Maturana D, Scherer S. VoxNet: a 3D convolutional neural network for real-time object recognition. In: Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Hamburg, Germany: IEEE, 2015. 922−928

127 He S F, Lau R W H, Liu W X, Huang Z, Yang Q X. SuperCNN: a superpixelwise convolutional neural network for salient object detection. *International Journal of Computer Vision*, 2015, **15**(3): 330−344

128 Nam H, Han B. Learning multi-domain convolutional neural networks for visual tracking. arXiv: 1510.07945, 2015.

129 Danelljan M, Häger G, Shahbaz Khan F, Felsberg M. Learning spatially regularized correlation filters for visual tracking. In: Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015. 4310−4318

130 Wang N Y, Li S Y, Gupta A, Yeung D Y. Transferring rich feature hierarchies for robust visual tracking. arXiv: 1501.04587, 2015.

131 Kotsiantis S B, Zaharakis I D, Pintelas P E. Machine learning: a review of classification and combining techniques. *Artificial Intelligence Review*, 2006, **26**(3): 159−190

132 Schölkopf B, Smola A J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. London, England: MIT Press, 2002.

133 Lu Z W, Ip H H S. Image categorization with spatial mismatch kernels. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Miami, Florida, USA: IEEE, 2009. 3974−404

134 Lazebnik S, Schmid C, Ponce J. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE, 2006. 2169−2178

135 Yang J C, Yu K, Gong Y H, Huang T. Linear spatial pyramid matching using sparse coding for image classification. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, Florida, USA: IEEE, 2009. 1794−1801

136 Kavukcuoglu K, Ranzato M A, Fergus R, LeCun Y. Learning invariant features through topographic filter maps. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Miami, Florida, USA: IEEE, 2009. 1605−1612

137 Gao S H, Tsang I W H, Chia L T, Zhao P L. Local features are not lonely-Laplacian sparse coding for image classification. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA: IEEE, 2010. 3555−3561

138 Meshram S B, Shinde S M. A survey on ensemble methods for high dimensional data classification in biomedicine field. *International Journal of Computer Applications*, 2015, **111**(11): 5−7

139 Papageorgiou C, Poggio T. A trainable system for object detection. *International Journal of Computer Vision*, 2000, **38**(1): 15−33

140 Wu B, Nevatia R. Detection of multiple, partially occluded humans in a single image by Bayesian combination of edgelet part detectors. In: Proceedings of the 10th IEEE International Conference on Computer Vision. Beijing, China: IEEE, 2005. 90−97

141 Wu B, Nevatia R. Cluster boosted tree classifier for multi-view, multi-pose object detection. In: Proceedings of the 11th IEEE International Conference on Computer Vision. Rio de Janeiro, Brazil: IEEE, 2007. 1−8

142 Wang Y, Jodoin P M, Porikli F, Konrad J, Benezeth Y, Ishwar P. CDnet 2014: an expanded change detection benchmark dataset. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Columbus, OH, USA: IEEE, 2014. 393−400

143 Dollár P, Wojek C, Schiele B, Perona P. Pedestrian detection: a benchmark. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Miami, Florida, USA: IEEE, 2009. 304−311

144 González A, Vázquez D, Ramos S, López A M, Amores J. Spatiotemporal stacked sequential learning for pedestrian detection. In: Proceedings of the 7th Iberian Conference on Pattern Recognition and Image Analysis. Santiago de Compostela, Spain: Springer, 2015. 3−12

145 Lin T Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick C L. Microsoft COCO: common objects in context. In: Proceedings of the 13th European Conference on Computer Vision (ECCV 2014). Zurich, Switzerland: Springer, 2014. 740−755

146 Seber G A F, Lee A J. *Linear Regression Analysis* (Second Edition). New York: John Wiley & Sons, 2003.

147 Comaniciu D, Ramesh V, Meer P. Real-time tracking of non-rigid objects using mean shift. In: Proceedings of the 2000 IEEE Conference on Computer Vision and Pattern Recognition. Hilton Head, SC, USA: IEEE, 2000. 142−149

148 Chen F S, Fu C M, Huang C L. Hand gesture recognition using a real-time tracking method and hidden Markov models. *Image and Vision Computing*, 2003, **21**(8): 745−758

149 Ali N H, Hassan G M. Kalman filter tracking. *International Journal of Computer Applications*, 2014, **89**(9): 15−18

150 Chang C, Ansari R. Kernel particle filter for visual tracking. *IEEE Signal Processing Letters*, 2005, **12**(3): 242−245

151 Comaniciu D, Ramesh V, Meer P. Kernel-based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, **25**(5): 564−577

152 Rahmati H, Aamo O M, Stavdahl Ø, Adde L. Kernel-based object tracking for cerebral palsy detection. In: Proceedings of the 2012 International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV). United States: CSREA Press, 2012. 17−23

153 Melzer T, Reiter M, Bischof H. Appearance models based on kernel canonical correlation analysis. *Pattern Recognition*, 2003, **36**(9): 1961−1971

154 Yilmaz A. Object tracking by asymmetric kernel mean shift with automatic scale and orientation selection. In: Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, Minnesota, USA: IEEE, 2007. 1−6

155 Hu J S, Juan C W, Wang J J. A spatial-color mean-shift object tracking algorithm with scale and orientation estimation. *Pattern Recognition Letters*, 2008, **29**(16): 2165−2173

156 Levey A, Lindenbaum M. Sequential Karhunen-Loeve basis extraction and its application to images. *IEEE Transactions on Image Processing*, 2000, **9**(8): 1371−1374

157 Brand M. Incremental singular value decomposition of uncertain data with missing values. In: Proceedings of the 7th European Conference on Computer Vision (ECCV 2002). Copenhagen, Denmark: Springer, 2002. 707−720

158 De La Torre F, Black M J. A framework for robust subspace learning. *International Journal of Computer Vision*, 2003, **54**(1−3): 117−142

159 Li Y M. On incremental and robust subspace learning. *Pattern Recognition*, 2004, **37**(7): 1509−1518

160 Skocaj D, Leonardis A. Weighted and robust incremental method for subspace learning. In: Proceedings of the 9th IEEE International Conference on Computer Vision. Nice, France: IEEE, 2003. 1494−1501

161 Ross D A, Lim J, Lin R S, Yang M H. Incremental learning for robust visual tracking. *International Journal of Computer Vision*, 2008, **77**(1−3): 125−141

162 Wang Q, Chen F, Xu W L, Yang M H. Object tracking via partial least squares analysis. *IEEE Transactions on Image Processing*, 2012, **21**(10): 4454−4465

163 Li X, Hu W M, Zhang Z F, Zhang X Q, Luo G. Robust visual tracking based on incremental tensor subspace learning. In: Proceedings of the 11th IEEE International Conference on Computer Vision. Rio de Janeiro, Brazil: IEEE, 2007. 1−8

164 Wen J, Li X L, Gao X B, Tao D C. Incremental learning of weighted tensor subspace for visual tracking. In: Proceedings of the 2009 IEEE International Conference on Systems, Man and Cybernetics. San Antonio, Texas, USA: IEEE, 2009. 3688−3693

165 Khan Z H, Gu I Y H. Nonlinear dynamic model for visual object tracking on grassmann manifolds with partial occlusion handling. *IEEE Transactions on Cybernetics*, 2013, **43**(6): 2005−2019

166 Chin T J, Suter D. Incremental kernel principal component analysis. *IEEE Transactions on Image Processing*, 2007, **16**(6): 1662−1674

167 Mei X, Ling H B. Robust visual tracking using l_1 minimization. In: Proceedings of the 12th IEEE International Conference on Computer Vision. Kyoto, Japan: IEEE, 2009. 1436−1443

168 Li H X, Shen C H, Shi Q F. Real-time visual tracking using compressive sensing. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Providence, RI, USA: IEEE, 2011. 1305−1312

169 Jia X, Lu H C, Yang M H. Visual tracking via adaptive structural local sparse appearance model. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Providence, RI, USA: IEEE, 2012. 1822−1829

170 Dong W H, Chang F L, Zhao Z J. Visual tracking with multifeature joint sparse representation. *Journal of Electronic Imaging*, 2015, **24**(1): 013006

171 Hu W M, Li W, Zhang X Q, Maybank S. Single and multiple object tracking using a multi-feature joint sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, **37**(4): 816−833

172 Zhang T Z, Liu S, Xu C S, Yan S C, Ghanem B, Ahuja N, Yang M H. Structural sparse tracking. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: IEEE, 2015. 150−158

173 Zhong W, Lu H C, Yang M H. Robust object tracking via sparse collaborative appearance model. *IEEE Transactions on Image Processing*, 2014, **23**(5): 2356−2368

174 Bai T X, Li Y F. Robust visual tracking with structured sparse representation appearance model. *Pattern Recognition*, 2012, **45**(6): 2390−2404

175 Zhang S P, Yao H X, Zhou H Y, Sun X, Liu S H. Robust visual tracking based on online learning sparse representation. *Neurocomputing*, 2013, **100**: 31−40

176 Wang N Y, Wang J D, Yeung D Y. Online robust nonnegative dictionary learning for visual tracking. In: Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV). Sydney, Australia: IEEE, 2013. 657−664

177 Zhang X, Guan N Y, Tao D C, Qiu X G, Luo Z G. Online multi-modal robust non-negative dictionary learning for visual tracking. *PLoS One*, 2015, **10**(5): 657−664

178 Oza N C. Online bagging and boosting. In: Proceedings of the 2005 IEEE International Conference on Systems, Man and Cybernetics. Waikoloa, Hawaii, USA: IEEE, 2005. 2340−2345

179 Valiant L. *Probably Approximately Correct: Nature's Algorithms for Learning and Prospering in a Complex World*. New York, USA: Basic Books, 2013.

180 Grabner H, Grabner M, Bischof H. Real-time tracking via on-line boosting. In: Proceedings of the 2006 British Machine Conference. Edinburgh, UK: British Machine Vision Association, 2006. 6.1−6.10

181 Liu X M, Yu T. Gradient feature selection for online boosting. In: Proceedings of the 11th IEEE International Conference on Computer Vision. Rio de Janeiro, Brazil: IEEE, 2007. 1−8

182 Avidan S. Ensemble tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, **29**(2): 261−271

183 Parag T, Porikli F, Elgammal A. Boosting adaptive linear weak classifiers for online learning and tracking. In: Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, Alaska, USA: IEEE, 2008. 1−8

184 Visentini I, Snidaro L, Foresti G L. Dynamic ensemble for target tracking. In: Proceedings of the 8th IEEE International Workshop on Visual Surveillance (VS2008). Marseille, France: IEEE, 2008. 1−8

185 Okuma K, Taleghani A, De Freitas N, Little J J, Lowe D G. A boosted particle filter: multitarget detection and tracking. In: Proceedings of the 8th European Conference on Computer Vision (ECCV 2004). Prague, Czech Republic: Springer, 2004. 28−39

186 Wang J Y, Chen X L, Gao W. Online selecting discriminative tracking features using particle filter. In: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, CA, USA: IEEE, 2005. 1037−1042

187 Avidan S. Support vector tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, **26**(8): 1064−1072

188 Williams O, Blake A, Cipolla R. Sparse Bayesian learning for efficient visual tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, **27**(8): 1292−1304

189 Tian M, Zhang W W, Liu F Q. On-line ensemble SVM for robust object tracking. In: Proceedings of the 8th Asian Conference on Computer Vision (ACCV 2007). Tokyo, Japan: Springer, 2007. 355−364

190 Yao R, Shi Q F, Shen C H, Zhang Y N, van den Hengel A. Part-based visual tracking with online latent structural learning. In: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Portland, OR, USA: IEEE, 2013. 2363−2370

191 Bai Y C, Tang M. Robust tracking via weakly supervised ranking SVM. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Providence, RI, USA: IEEE, 2012. 1854−1861

192 Hare S, Saffari A, Torr P H S. Struck: structured output tracking with kernels. In: Proceedings of the 2011 International Conference on Computer Vision (ICCV). Barcelona, Spain: IEEE, 2011. 263−270

193 Tang F, Brennan S, Zhao Q, Tao H. Co-tracking using semi-supervised support vector machines. In: Proceedings of the 11th IEEE International Conference on Computer Vision. Rio de Janeiro, Brazil: IEEE, 2007. 1−8

194 Zhang S L, Sui Y, Yu X, Zhao S C, Zhang L. Hybrid support vector machines for robust object tracking. *Pattern Recognition*, 2015, **48**(8): 2474−2488

195 Zhang X M, Wang M G. Compressive tracking using incremental LS-SVM. In: Proceedings of the 27th Chinese Control and Decision Conference (CCDC). Qingdao, China: IEEE, 2015. 1845−1850

196 Breiman L. Random forests. *Machine Learning*, 2001, **45**(1): 5−32

197 Saffari A, Leistner C, Santner J, Godec M, Bischof H. Online random forests. In: Proceedings of the 12th IEEE International Conference on Computer Vision (ICCVW). Kyoto, Japan: IEEE, 2009. 1393−1400

198 Leistner C, Saffari A, Bischof H. Miforests: multiple-instance learning with randomized trees. In: Proceedings of the 11th European Conference on Computer Vision (ECCV 2010). Crete, Greece: Springer, 2010. 29−42

199 Godec M, Leistner C, Saffari A, Bischof H. On-line random naive bayes for tracking. In: Proceedings of the 20th International Conference on Pattern Recognition (ICPR). Istanbul, Turkey: IEEE, 2010. 3545−3548

200 Wang A P, Wan G W, Cheng Z Q, Li S K. An incremental extremely random forest classifier for online learning and tracking. In: Proceedings of the 16th IEEE International Conference on Image Processing (ICIP). Cairo, Egypt: IEEE, 2009. 1449−1452

201 Lin R S, Ross D A, Lim J, Yang M H. Adaptive discriminative generative model and its applications. In: Proceedings of the 2004 Advances in Neural Information Processing Systems 17. Vancouver, British Columbia, Canada: MIT Press, 2004. 801−808

202 Nguyen H T, Smeulders A W M. Robust tracking using foreground-background texture discrimination. *International Journal of Computer Vision*, 2006, **69**(3): 277−293

203 Li X, Hu W M, Zhang Z F, Zhang X Q, Zhu M L, Cheng J. Visual tracking via incremental log-Euclidean Riemannian subspace learning. In: Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, Alaska, USA: IEEE, 2008. 1−8

204 Wang X Y, Hua G, Han T X. Discriminative tracking by metric learning. In: Proceedings of the 11th European Conference on Computer Vision (ECCV 2010). Heraklion, Crete, Greece: Springer, 2010. 200−214

205 Tsagkatakis G, Savakis A. Online distance metric learning for object tracking. *IEEE Transactions on Circuits and Systems for Video Technology*, 2011, **21**(12): 1810−1821

206 Xu Z F, Shi P F, Xu X Y. Adaptive subclass discriminant analysis color space learning for visual tracking. In: Proceedings of the 9th Pacific Rim Conference on Advances in Multimedia Information Processing (PCM 2008). Tainan, China: Springer, 2008. 902−905

207 Zhang X Q, Hu W M, Chen S Y, Maybank S. Graph-embedding-based learning for robust object tracking. *IEEE Transactions on Industrial Electronics*, 2014, **61**(2): 1072−1084

208 Zha Yu-Fei, Bi Du-Yan, Yang Yuan, Dong Shou-Ping, Luo Ning. Transductive learning with global and local constraints for robust visual tracking. *Acta Automatica Sinica*, 2010, **36**(8): 1084−1090
(查宇飞, 毕笃彦, 杨源, 董守平, 罗宁. 基于全局和局部约束直推学习的鲁棒跟踪研究. 自动化学报, 2010, **36**(8): 1084−1090)

209 Wu Y, Lim J, Yang M H. Online object tracking: a benchmark. In: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, OR, USA: IEEE, 2013. 2411−2418

210 Wu Y, Lim J, Yang M H. Object tracking benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, **37**(9): 1834−1848

211 Collins R, Zhou X H, Teh S K. An open source tracking testbed and evaluation web site. In: Proceedings of IEEE International Workshop on Performance Evaluation of Tracking and Surveillance. Beijing, China: IEEE, 2005.

212 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. Prague, Czech Republic: IEEE, 2004. 1−5

213 Pellegrini S, Ess A, Schindler K, van Gool L. You′ll never walk alone: modeling social behavior for multi-target tracking. In: Proceedings of the 12th IEEE International Conference on Computer Vision. Kyoto, Japan: IEEE, 2009. 261−268

214 Leibe B, Schindler K, Van Gool L. Coupled detection and trajectory estimation for multi-object tracking. In: Proceedings of the 11th IEEE International Conference on Computer Vision. Rio de Janeiro, Brazil: IEEE, 2007. 1−8

215 Milan A, Leal-Taixe L, Reid I, Roth S, Schindler K. MOT16: a benchmark for multi-object tracking. arXiv: 1603.00831, 2016.

216 Li L Z, Nawaz T, Ferryman J. PETS 2015: datasets and challenge. In: Proceedings of the 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). Karlsruhe, Germany: IEEE, 2015. 1−6

217 Kristan M, Matas J, Leonardis A, Felsberg M, Cehovin L, Fernandez G, Vojir T, Hager G, Nebehay G, Pflugfelder R, Gupta A, Bibi A, Lukezic A, Garcia-Martin A, Saffari A, Petrosino A, Montero A S. The visual object tracking VOT2015 challenge results. In: Proceedings of the 2015 IEEE International Conference on Computer Vision Workshops. Santiago, Chile: IEEE, 2015. 564−586

218 Lee J Y, Yu W. Visual tracking by partition-based histogram backprojection and maximum support criteria. In: Proceedings of the 2011 IEEE International Conference on Robotics and Biomimetics (ROBIO). Karon Beach, Thailand: IEEE, 2011. 2860−2865

219 Vojir T, Noskova J, Matas J. Robust scale-adaptive mean-shift for tracking. *Pattern Recognition Letters*, 2014, **49**: 250−258

220 Zhang K H, Zhang L, Yang M H. Real-time compressive tracking. In: Proceedings of the 12th European Conference on Computer Vision (ECCV 2012). Florence, Italy: Springer, 2012. 864−877

221 Bao C L, Wu Y, Ling H B, Ji H. Real time robust l_1 tracker using accelerated proximal gradient approach. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Providence, RI, USA: IEEE, 2012. 1830−1837

222 Binh N D. Online boosting-based object tracking. In: Proceedings of the 12th International Conference on Advances in Mobile Computing and Multimedia. Kaohsiung, China: ACM, 2014. 194−202

223 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). Providence, RI: IEEE, 2011. 1177−1184

224 Dollár P, Belongie S, Perona P. The fastest pedestrian detector in the west. In: Proceedings of the 2010 British Machine Vision Conference. Aberystwyth, UK: British Machine Vision Association, 2010. 68.1−68.11

225 Hare S, Golodetz S, Saffari A, Vineet V, Cheng M M, Hicks S, Torr P. Struck: structured output tracking with kernels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, DOI: 10.1109/TPAMI.2015.2509974

226 Babenko B, Yang M H, Belongie S. Robust object tracking with online multiple instance learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, **33**(8): 1619−1632

227 Kristan M, Pflugfelder R, Leonardis A, Matas J, Čehovin L, Nebehay G, Vojíř T, Fernández G, Lukežič A, Dimitriev A, Petrosino A, Saffari A, Li B, Han B, Heng C, Garcia C, Pangeršič D, Häger G, Khan F S, Oven F, Possegger H, Bischof H, Nam H, Zhu J K, Li J J, Choi J Y, Choi J W, Henriques J F, van de Weijer J, Batista J, Lebeda K, Öfjäll K, Yi K M, Qin L, Wen L Y, Maresca M E, Danelljan M, Felsberg M, Cheng M M, Torr P, Huang Q M, Bowden R, Hare S, Lim S Y, Hong S, Liao S C, Hadfield S, Li S Z, Duffner S, Golodetz S, Mauthner T, Vineet V, Lin W Y, Li Y, Qi Y K, Lei Z, Niu Z H. The visual object tracking VOT2014 challenge results. In: Proceedings of the European Conference on Computer Vision (ECCV 2014), Lecture Notes in Computer Science. Zurich, Switzerland: Springer International Publishing, 2015. 191−217

尹宏鹏 重庆大学自动化学院副教授. 2009 年获得重庆大学博士学位. 主要研究方向为模式识别, 图像处理与计算机视觉. 本文通信作者.
E-mail: yinhongpeng@gmail.com
(**YIN Hong-Peng** Associate professor at the College of Automation, Chongqing University. He received his Ph. D. degree from Chongqing University in 2009. His research interest covers pattern recognition, image processing, and computer vision. Corresponding author of this paper.)

陈 波 重庆大学自动化学院硕士研究生. 2015 年获得重庆大学学士学位. 主要研究方向为深度学习, 计算机视觉.
E-mail: qiurenbieyuan@gmail.com
(**CHEN Bo** Master student at the College of Automation, Chongqing University. He received his bachelor′s degree from Chongqing University in 2015. His research interest covers deep learning and computer vision.)

柴 毅 重庆大学自动化学院教授. 2001 年获得重庆大学博士学位. 主要研究方向为信息处理, 融合与控制, 计算机网络与系统控制.
E-mail: chaiyi@cqu.edu.cn
(**CHAI Yi** Professor at the College of Automation, Chongqing University. He received his Ph. D. degree from Chongqing University in 2001. His research interest covers information processing, integration and control, and computer network and system control.)

刘兆栋 重庆大学自动化学院博士研究生. 主要研究方向为稀疏表示, 机器学习.
E-mail: liuzhaodong@cqu.edu.cn
(**LIU Zhao-Dong** Ph. D. candidate at the College of Automation, Chongqing University. His research interest covers sparse representation and machine learning.)