

基于外观模型学习的视频目标跟踪方法综述

张焕龙^{1,2} 胡士强¹ 杨国胜³

¹(上海交通大学航空航天学院 上海 200240)
²(洛阳理工学院计算机与信息工程系 河南洛阳 471023)
³(中央民族大学信息工程学院 北京 100081)
(zhl_lit@sjtu.edu.cn)

Video Object Tracking Based on Appearance Models Learning

Zhang Huanlong^{1,2}, Hu Shiqiang¹, and Yang Guosheng³

¹(School of Aeronautics and Astronautics, Shanghai Jiao Tong University, Shanghai 200240)
²(Department of Computer and Information Engineering, Luoyang Institute of Science and Technology, Luoyang, Henan 471023)
³(School of Information Engineering, Minzu University of China, Beijing 100081)

Abstract Visual tracking is an active research topic in the field of computer vision and has been well studied in the last decades. A key component for achieving robust tracking is the tracker’s capability of updating its internal representation of targets to capture the varying appearance. Although numerous approaches have been proposed, many challenging problems still remain in designing an effective model of the appearance of tracked objects. In recent years, the methods of appearance model associated with statistical learning have been promoting the study for video object tracking. To help reader swiftly learn the recent advances and trends so as to easily grasp the key problems of visual object tracking based on appearance models learning, a detailed review of the existing appearance learning models is provided. Here, the mechanism of the tracking algorithm based on appearance model learning is introduced firstly. Then the state-of-the-art feature descriptors are analyzed to show their different performance. Meanwhile, the tracking progress is categorized into three main groups, and the character of representative methods in each group are compared and analyzed in detail. Finally, the current research on the tracking methods based appearance model learning is summarized and classified, and the further application and research trend is discussed.

Key words appearance model learning; sparse representation; metric learning; video tracking; semi-supervised learning

摘 要 视频跟踪是机器视觉领域中的热点研究问题,在过去的几十年内得到了广泛研究.为了获得鲁棒的跟踪效果,设计能够适应跟踪目标外观变化的外观模型成为算法研究中的一种重要内容.近年来,将机器学习理论引入外观模型设计中的思想大大推动了视频跟踪研究的发展.为了使读者能够快速了解其发展的趋势并且掌握基于外观模型学习跟踪算法研究的技术,在介绍外观模型学习跟踪算法原理和机制的基础上,重点综述了外观模型学习跟踪方法的研究进展,包括目标特征描述和 3 类主要目标外观建模方式及其各自研究过程中跟踪方法的对比与分析,进而总结了外观模型学习跟踪算法在理论及应用方面的研究现状,最后提出进一步研究的主要发展内容和趋势.

关键词 外观模型学习;稀疏表示;度量学习;视频跟踪;半监督学习

中图法分类号 TP391

视频目标跟踪是机器视觉研究领域中的一个热点问题,其主要任务是在视频序列中获取兴趣目标的运动轨迹,已经被广泛应用到视频监控、智能交通和人机交互等研究领域. 尽管目前涌现很多研究成果,但仍存在具有挑战性的问题亟待解决:1)跟踪目标自身的姿势变化、尺度变化和形变等内在因素;2)跟踪过程出现亮度变化、部分遮挡和背景复杂等外在因素.

近年来,视频跟踪研究与机器学习理论的结合越来越紧密,其发展主要经历了2个阶段:第1阶段,跟踪研究提前设定目标运动状态和场景条件,很多固定外观模型的跟踪算法被提出^[1-2],算法主要围绕精确定位方法和优化匹配策略等问题,在短期跟踪和可控场景情况下取得好的效果. 然而,实际中跟踪目标的运动状态往往会随时间的变化呈现很强的不确定性,因此第2阶段跟踪技术的研究重点转向如何设计能够自适应目标和场景变化的外观模型方法. 目前,依据模型设计机制的不同跟踪方法可以分为3类:1)产生式外观模型学习跟踪方法,它通过学习目标在视频中的各种运动变化,建立自适应性外观模型实现跟踪^[3-11]. 2)判别式外观模型学习跟踪方法,它将跟踪问题转换成二分类问题,依据目标和背景的分离边界建立自适应性外观模型实现跟踪^[12-19]. 3)混合产生-判别式外观模型学习跟踪方法,它结合了产生式模型能够描述广义变化和判别式模型能够描述差异性变化的优点建立自适应性外观模型实现跟踪^[20-21].

虽然,目前已经存在一些关于视频目标跟踪的综述文献,但它们有的总结了目标的表示形式、检测方法和跟踪策略,将跟踪归纳为点、核和轮廓3个方面,研究覆盖面广泛;有的总结了视频跟踪的不同应用领域,并将跟踪方法分为基于特征、区域、变形模板和模型4个方面;也有的对跟踪方法从基于在线学习和基于蒙特卡罗采样2方面进行综述,均没有分析学习策略与模型表征能力之间的关系.

本文的贡献主要体现在以下3个方面:1)为了说明目标特征描述在自适应跟踪方法中面临的新问题,从单一特征和混合特征两方面,总结了能够适应跟踪方法的各种特征形式,并进行对比与分析;2)为了分析不同建模方式与跟踪效果间的关系,总结了

外观模型设计的3类主要方法(产生式、判别式和混合式),并对不同建模方式对外观模型表征能力的强弱进行详细的对比与分析;3)为了体现外观模型的不同学习能力,总结了模型参数学习、子空间学习、度量学习、流型学习和分类器学习等各种形式对视频跟踪方法自适应性的影响情况,并对它们进行了对比与分析. 同时,给出外观模型学习方法实现自适应跟踪的基本机制,依据其当前的研究进展和典型应用成果指出外观模型学习跟踪方法面临的主要问题和未来发展的主要趋势.

1 外观模型学习跟踪方法基本理论

设 $I_t = \{I_m\}_{m=1}^t$ 为图像序列,视频跟踪就是去估计图像序列中关于目标的隐藏变量 $X_t = \{x_m \in \chi \subset \mathbb{R}^d\}_{m=1}^t$, χ 为状态空间, x_m 为时刻 t 的目标状态, X_t 为组成目标的轨迹,状态获取如式(1)所示:

$$x_t = \arg \min_x p(f(\hat{x}), \mathbf{A}_t), \quad (1)$$

式(1)中, $\mathbf{A}_t$ 为时刻 t 目标外观模型的特征向量, $p(\cdot, \cdot)$ 为目标匹配函数, $\hat{x}_t^i$ 为时刻 t 的观测状态(i 为候选观测区域数目), $f(\hat{x})$ 为当前视频中状态 x 相关观测区域的特征形式,其方法概况如图1所示:

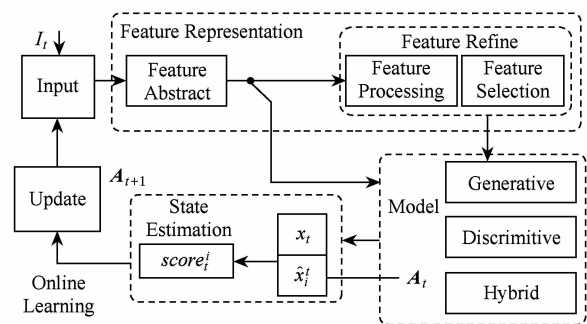


Fig. 1 Flow chart of video tracking based on appearance models learning.

图1 外观模型学习跟踪方法功能图

1) 固定外观模型的跟踪方法. $\mathbf{A}_t$ 通过离线训练或从第1帧选取特征建立外观模型,跟踪过程中不再变化,即 $\mathbf{A}_t = \mathbf{A}_1$, 这种情况下模型无法适应目标外观的变化,跟踪局限性很强.

2) 外观模型学习的跟踪方法. $\mathbf{A}_{t+1}$ 是在 $\mathbf{A}_t$ 的基础上,结合 $\hat{x}_t^i$ 的评估分数 $score_t^i$ 和 x_t 的状态,对 $\mathbf{A}_t$

进行更新以获得新的外观模型,实现过程描述如式(2)所示:

$$\mathbf{A}_{i+1} = L(\mathbf{A}_i, \mathbf{A}_i, f(\hat{x}')), \quad (2)$$

$L()$ 为学习或更新函数,通过式(2)目标观测被融入模型更新过程,有时式(2)中的 $\mathbf{A}_i$ 也可被视频中关键帧图像特征代替,以获得鲁棒的模型更新.新的外观模型获得如式(3)所示:

$$\mathbf{A}_{i+1} = \arg \min_{\mathbf{A}} p(\mathbf{A}, \mathbf{A}_{i+1}^{\text{real}}), \quad (3)$$

$\mathbf{A}_{i+1}^{\text{real}}$ 为时刻 $t+1$ 时真实的外观模型特征,通过机器学习方法更新模型使其无限接近真实目标.

实际中,目标模型的表现形式一般采用矩形框或椭圆来描述,这会导致背景像素混入 $f(\hat{x})$ 中,故式(3)获得的目标模型并不是真实的目标近似,表达如式(4)所示:

$$\mathbf{A}_{i+1} = \mathbf{A}_{i+1}^{\text{real}} + \sum_{i=1}^t \mathbf{A}_e^{i \rightarrow i+1}, \quad (4)$$

这里 $\mathbf{A}_e^{i \rightarrow i+1}$ 代表目标模型从 i 到 $i+1$ 目标模型因混入背景像素带来的误差,外观模型设计的主要任务就是能够将这种误差对跟踪效果的影响最小化.

2 视频目标特征描述

目标外观特征描述是对同一目标在不同时刻呈现的运动变化形式进行刻画和分析,是视频跟踪研究的重要内容.随着机器学习方法的引入,目标特征描述面临很多新的研究问题,这里从单一特征和混合特征两方面进行总结,旨在说明外观建模中特征描述方式对跟踪效果的影响关系.

2.1 单一特征描述

根据目标表征的不同形式,可以将近期跟踪算法中经常出现的单一特征描述方法总结为以下几类:颜色特征描述、梯度特征描述、灰度级特征描述和纹理特征描述.

颜色特征通常以直方图的形式描述目标.文献[22]将颜色直方图作为目标描述方式并融入粒子滤波框架,这种概率跟踪方法能够适应目标外观的急剧变化.该方法的缺点是模型不能体现目标的空间特性.文献[23]将目标的位置信息与颜色统计信息相关联,建立基于相关颜色直方图的目标外观模型,在mean shift框架下能够适应目标尺度变化、方向变化和部分遮挡等跟踪问题.然而,这两类颜色特征如果遇到目标视角变化的情况将难以适应.文献[24]利用目标的多视角先验信息,建立基于多参考

颜色直方图的目标外观模型,并通过扩展的mean shift模型匹配实现目标跟踪.颜色特征描述目标具有对姿态和旋转不敏感,而且选取方便、计算简单,在外观模型学习跟踪方法中广泛使用,但其缺点是容易受到外界光照变化的影响,当出现颜色相似的目标或背景时模型表征能力会下降.

梯度特征通常以区域统计的形式描述目标,其中最为经典的是尺度不变性特征转换(scale-invariant feature transform, SIFT)算子,在很多跟踪算法中被用来描述目标的结构性外观变化^[25-26].然而,SIFT算子的梯度统计过程比较耗时,应用受到限制.为此,文献[27]提出SIFT的加速形式鲁棒特征(speeded up robust feature, SURF)算子来描述目标,以加强跟踪的实时性.另外,区域算子(histogram of oriented gradient, HOG)利用分块分单元方式对梯度方向进行统计,能够很好表征目标局部像素之间的关联,成为行人跟踪^[28]研究中目标特征描述的经典方法.梯度特征描述模型对目标尺度、亮度和旋转等具有不变性,而且特征稳定、鲁棒,但其缺点是不能描述目标外观精确的尺寸、方向和姿态等信息,对噪声干扰比较敏感,而且目标空间信息表征能力较弱.

灰度级特征以像素变化和区域变化两种形式描述目标,其中灰度值作为外观模型的方式最直观,也最简单,但其特征表征能力较弱、易受干扰故使用较少.而区域特征 Haar-Like 因其设计简单,对边缘、水平和垂直方向细节敏感,而且计算效率较高,在模型学习跟踪方法研究中经常使用.文献[29]首次提出利用 AdaBoost 判别式学习方法对大量 Haar-like 特征进行提炼,获得能够描述目标外观的最优特征模型.文献[30]将这种通过特征选择获取的目标描述模型与主动学习方法相结合,在粒子滤波框架下实现车辆有效跟踪.文献[31]将其融入模板、光流和mean shift多种框架,通过图形处理器(graphic processing unit, GPU)实现在线并行跟踪.这类特征描述方式对具有对称性结构的目标表征能力很强,同时它采用了积分图策略能够提高执行效率,但该模型的缺点是整体区分力较弱,难以适应外观变化复杂的跟踪环境.

纹理特征以区域表面微观变化方式描述目标,是对目标光滑程度和规则程度的量化.其中比较经典的方法有灰度共生矩阵和局部二值模式(local binary pattern, LBP)算子,文献[32]分别给出它们与粒子滤波结合实现跟踪的对比与分析结果.但传统的LBP方法提取过程复杂,为了满足跟踪算法的实时

性要求,文献[33]将 LBP 与差分塔式匹配方法相结合,增强了模型表征能力,提高了跟踪运行效率.同时,为了增强外观模型的区分力,文献[34]利用 LBP 算子设计目标前景模型,在粒子滤波框架下对背景外观模型实时更新,目标整体外观模型的区分能力

获得提高.纹理模型属于目标的表面特征,不能描述目标的本质属性,难以适应目标出现影子、表面反光或周期性外观变化的跟踪环境,而且它依赖于目标的分辨率变化.表 1 给出典型算法中目标单一特征描述方法和融合特征描述方法的对比与分析结果:

Table 1 Summary of the Features Presentation Used by Trackers
表 1 典型跟踪方法目标特征描述效果分析

Feature Form		Representative and Methods	Character	Tracking Performance
Single	Original	Color ^[22-24] ;SIFT ^[25-26] ; SURF ^[27] ;HOG ^[28] ;LBP ^[32-34]	Be easy to get; Strong limitation from image content	Adapt the simple tracking environment
	Refinement	Haar-like ^[29-31]	Real-time	Adapt the structural object tracking in real time
Mixture	Weighting	Color,Texture ^[35] ; Color,Texture, Edge ^[36] ;Color,Edge,Motion ^[37]	Good expression; Complex form	Adapt the drastic object appearance changes
	Descriptor	Covariance ^[39] ;Sigma Set ^[41]	Low dimension ; Good expression; Be easy to mixture	Adapt the drastic object appearance changes in real-time tracking
	Refinement	Haar-like,HOG, LBP ^[42] ; Haar-like,HOG,.,Color ^[43]	Good Discriminability ; High computational cost	Adapt complex tracking environment; Tracking object hardly in real time

2.2 融合特征描述

视频目标的外观变化在复杂的跟踪环境中往往呈现出不确定性,单一的特征描述方式无法表征目标可能出现的多种变化形式.为了能够使外观模型区分能力增强,多特征融合的方式得到重视.依据融合方式的不同,目标描述方法可以分为:1)直接特征融合方式;2)算子特征融合方式;3)提炼特征融合方式.

直接特征融合方式是指多个特征直接采用不同的形式进行组合建立外观模型.文献[35]采用联合相似度方式将颜色和纹理特征进行融合,文献[36]采用线性加权方式将纹理、颜色和边缘特征进行融合,文献[37]直接将颜色、边缘和运动特征进行融合,它们以不同的融合形式实现目标的多特征描述,再结合学习策略提高了跟踪算法的自适应能力.该外观模型的鲁棒性依赖于特征分量的表征能力,这种简单联合多特征的方式,会造成外观模型获取时间呈现指数级的增长,不利于跟踪算法的实时性.

算子融合方式的目标是获取维数较少、计算简单的融合特征.文献[38]提出基于图像二阶统计特性的协方差算子融合方式,它将区域内像素多种特征的共生关系以统计量的方式编码于协方差矩阵中,能够描述比较复杂的外观特征,而且它不受目标形状、大小和亮度的影响,被引入视频跟踪领域^[39].为了获得简单的统计形式,Sigma Set^[40]算子融合方式利用一阶统计特性获取特征,操作简单,相对协方

差融合方式描述效果更好.文献[41]将其成功用于视频跟踪领域,采用多块 Sigma Set 特征描述目标外观,在核模型学习框架下实现鲁棒跟踪.

虽然直接融合和算子融合方式提高了目标特征的描述能力,但随着目标外观的变化它们很难保证模型始终具备最优的区分能力.提炼特征融合方式因其能够在每帧图像中获取最优的混合特征以描述目标,成为热点研究内容.其中,基于 Boosting 框架的融合方法使用最多,它将多特征分别和弱分类器建立关联,通过迭代训练提炼最优特征,建立具有强区分力的外观模型.文献[42]将 Haar-like, HOG 和 LBP 多个特征融合描述目标,采用 AdaBoost 方法在线提炼最优特征建立模型分离目标和背景,文献[43]将 Haar-like,HOG 和颜色直方图在 Boosting 框架下实现特征提炼融合,结合半监督学习方法能够有效缓解轨迹漂移问题.

目前,提炼单一或混合特征描述目标外观的方法在视频跟踪中应用较多.但这种特征提炼的方式计算代价较大,如何提高特征优化效率和如何确定弱分类器形式仍是 Boosting 特征融合方式需要进一步研究的问题.

3 产生式外观模型学习跟踪方法

产生式模型因其考虑数据的内在分布,具有很好的目标描述能力.目前,产生式外观模型学习跟踪

方法依据目标表征方式不同可以分为:1)混合产生式学习;2)子空间产生式学习.

3.1 混合产生式学习跟踪

混合产生式学习是将目标外观模型通过具有不同时空特性的多个模型分量混合形成,它能够逼近视频序列中目标的多种变化.再对各个分量的模型参数进行在线更新,最终综合决策外观模型的更新形式,从而实现跟踪算法的自适应性.依据模型分量数目确定方式的不同可以分为:1)固定分量混合模型;2)自适应分量混合模型.

固定分量混合模型是指在跟踪执行前分量的数目已经确定,在跟踪过程中对模型参数进行更新以适应目标外观的不同变化形式.早期经典的固定混合外观模型是 WSL(wander-stable-lost)^[44],它包含稳定部分、快速部分和噪声部分 3 个分量,混合模型描述如式(5):

$$P(d_t | q_t, m_t, d_{t-1}) = m_w P_w(d_t; d_{t-1}) + m_l P_l(d_t) + m_s P_s(d_t; q_t), \quad (5)$$

这里 $m = (m_w, m_s, m_l)$ 是混合概率, P_w, P_s 和 P_l 分别是分量模型.混合分量的权重采用期望最大化(expectation maximization, EM)方法进行优化确定,该方法能够适应光照、表情、姿势等变化的跟踪环境.但这种混合模型分量过于简单,难以体现目标外观的丰富变化.文献[5]提出利用任意多个高斯分量 $g(x; \mu_k, c_k)$ 构造高斯混合模型,它能够逼近目标外观的多种变化形式,其中 x, μ_k, c_k 分别为像素特征、均值和方差,同样采用 EM 方法优化模型参数,对不规则目标实现有效跟踪.然而,固定分量混合模型的缺点是目标在视频中的外观变化无法提前获知,预先定义的分量数目很难完全描述目标的所有外观变化形式.

针对这一问题,研究者提出自适应分量混合模型,其模型分量数目会以能够最优逼近目标外观变化而自适应确定,模型学习能力得到增强.文献[6]提出高斯混合密度函数,它自适应确定高斯分量的数目和模型相关参数取值,并将像素的邻域信息融入外观模型,提高了模型描述能力.为了增强目标的空间特征信息,文献[7]提出空间-颜色混合高斯(spatial-color mixture of Gaussians, SMOG)外观模型,它将像素空间的特征变化和时间上的分布变化相结合,跟踪精度得以提高.然而,自适应分量混合模型中分量数目的确定往往需要设计评价指标,如果评价方法设计过于精确会使跟踪弹性降低,相反,会导致模型表征能力下降,如何选取分量数仍有

待深入研究.

固定式和自适应式的混合模型在早期的跟踪研究中应用较多,主要因为其存在一个前提条件,即目标运动和外观变化的连续性.然而,随着目标外观变化程度的增强,混合分量的数目和模型参数的调节变得十分困难.另外,混合产生式模型直接对图像信息进行外观建模,常常会面临图像维数较高、特征描述困难和图像间冗余信息较多等问题.

3.2 子空间产生式学习跟踪

子空间产生式模型学习跟踪方法利用基模板函数 $(e_1, e_2, \dots, e_n)$ 将兴趣目标 A 映射到子空间,外观模型可以描述为 $A = c_1 e_1 + c_2 e_2 + \dots + c_n e_n$,其中 $(c_1, c_2, \dots, c_n)$ 是目标映射变换下的系数矩阵,在低维子空间下对特征基进行更新保持外观模型的学习能力,从而实现自适应跟踪.

3.2.1 基于向量的子空间模型

基于向量子空间学习跟踪方法是将图像视为特殊的高维列向量,将其以向量的形式映射到子空间建立外观模型,通过子空间模型的学习实现目标跟踪的自适应性.

最早基于向量子空间学习的跟踪方法^[45]采用二范数最小化方式获取能最优逼近目标外观的线性子空间 L ,通过在线更新 L 实现对目标亮度变化、部分遮挡和姿势变化的自适应性.该模型能够表征目标的本质属性,具有较强的区分能力,但它确定最优子空间形式时,需要在每帧图像中训练所有的观测样本,计算代价较高.

针对这一问题,增量线性子空间跟踪方法^[46]将重点放在目标外观的变化信息上,在子空间下构造增量模型,仅依据变化信息更新外观模型,减少了训练样本的数量和特征的存储量,在概率框架下实现了目标的自适应跟踪.但该方法在目标外观发生较大形变时,会因为构造子空间的特征基无法及时表征这种变化,而导致跟踪失败.为此,文献[8, 47-48]分别对增量子空间的构造方式进行改进,获得更加符合目标变化的外观模型,以提高跟踪的鲁棒性.

然而,基于向量的子空间建模方式需要将图像信息转换成向量的形式再进行子空间建模,这种转换丢失了目标外观在高维图像中的结构和邻域信息,会造成外观模型表征能力降低.

3.2.2 基于矩阵的子空间模型

基于矩阵的子空间模型是为了避免图像转化向量时带来的信息丢失,它直接以矩阵或者张量的形式对图像信息直接进行子空间建模,通过模型学习

实现自适应跟踪.

在实际中,因为增量张量子空间模型能够最大化地表征图像信息的几何结构和判别信息,常被用于视频目标跟踪研究中.文献[49]从目标外观变化信息入手,采用三阶张量子空间建立外观模型,相对于张量子空间跟踪方法而言,它能够同时对特征基和均值进行更新,模型学习能力更强,该方法对运动缓慢或目标外观变化较小的情况适应性较好.然而,该模型在目标运动剧烈或外观变化较大时容易跟踪失败.针对这一问题,二阶增量张量子空间方法^[50]和加权增量张量子空间方法^[51]分别对张量子空间建模方式进行改进,新的外观模型方式能够适应亮度变化剧烈的跟踪环境.然而,该模型的特征描述能力严重依赖于子空间模型能否获得目标外观分布的内在结构关系.

3.2.3 非线性子空间模型

基于向量和矩阵的子空间模型设计中,映射变化多采用的是线性方式,而在实际中兴趣目标的外观变化并不都能线性表示,尤其当外观发生突变时呈现很强的非线性,传统子空间方法难以适应.

为了解决这一问题,多子空间外观建模方式^[10]通过多子空间组合逼近目标外观的非线性变化,它能够适应目标的剧烈运动变化情况.局部线性嵌入式(local linear embedding, LLE)^[11]流型外观建模方式直接采用非线性降维方法实现目标外观非线性

变化的建模,它能够适应复杂场景下的动态外观变化,对轨迹漂移和遮挡问题具有很好处理能力.多子流外观建模方式^[52]将目标外观分解为多个子流,每个子流再通过子空间模型逼近,通过这种多子流方式逼近外观非线性变化,能够有效地处理部分遮挡、姿态和亮度变化及复杂背景等跟踪问题.然而,该模型的非线性学习过程一般计算代价较大.

3.2.4 稀疏子空间模型

近几年,基于稀疏表示的产生式外观模型^[53]被提出,它将目标通过基函数字典映射成稀疏向量建立外观模型,采用字典学习更新模型.文献[54]首次将它引入视频目标跟踪领域,并将跟踪转化成求解L1最小化问题,能够适应光线剧烈变化、姿态变化和严重遮挡等跟踪问题,但其计算量较大,容易产生轨迹漂移现象.

针对这一问题,文献[55-56]采用稀疏表示在子空间下建立外观模型,结合增量学习更新字典子空间,在粒子滤波框架下实现鲁棒跟踪.另外,稀疏表示因为对噪声干扰和遮挡不敏感,已经成为模型学习跟踪方法的新研究方向.然而,在实际中目标外观在稀疏化过程中具有一定的随机性,并不能保证每次稀疏的最优化,故还需进一步研究.表2给出产生式外观模型学习典型算法的应用对比情况(其中代表性视频是指相应跟踪方法能够顺利执行的跟踪环境,资源可以从作者网站获得):

Table 2 Summary of the State-of-the-art Tracking Algorithms Using Generative Models

表 2 不同场景下典型产生式跟踪方法性能分析

Methods	Model	Feature	Vedio	Tracking Performance
Frag ^[1]	Multi-block Fix	Histogram	Faceocc1, Sylvester	Adapt the occlusion and pose changes without appearance changes of tracked object
WSL ^[44]	Hybrid Parameter Update	Haar-like	David_outdoor, Dudek, Sylvester	Adapt the illumination, pose and expression changes; Adapt full occlusion hardly
IVT ^[8]	Subspace Basis Update	Gray Value	Faceocc2, Jumping, David2	Adapt face tracking well; Adapt the gradual changes of pose and illumination; Adapt drastic changes hardly
VTD ^[10]	Multi-Subspace Update	Gray, Edgy	Tiger1, Faceocc2, Shaking, Animal	Adapt abrupt motion Adapt severe occlusion hardly
L1 ^[54]	Sparse Dictionary Update	Gray Value	Singer, Jumping	Adapt severe occlusion and illumination changes; Adapt deformation and abrupt motion hardly

表2对基于产生式模型的典型跟踪方法进行分析,以说明不同外观模型对跟踪效果的影响. Frag通过分块策略针对部分遮挡的跟踪环境进行研究,算法能够适应视频序列 Faceocc1 和 Sylvester(这两个序列跟踪目标没有外观变化),但其模型固定难以在实际中广泛应用. WSL 方法、增量量子空间学习跟

踪(incremental subspace learning visual tracking, IVT)方法、视频跟踪分析(visual tracking decomposition, VTD)方法和 L1 范数最小化跟踪(L1 norm minimum tracking, L1)方法均设计了模型更新机制,因此对目标外观变化的跟踪环境有强的适应性. 其中, WSL 方法针对外观的不同变化设计混合分量

形式,跟踪算法针对性较强;IVT 方法针对目标变化的增量形式设计基函数和均值的更新形式,并采用插值方式适应尺度变化,跟踪算法在变化缓慢的跟踪环境下效果较好(Faceocc2 视频包含姿态变化和遮挡);VTD 方法针对目标运动的不确定性设计外观模型,并采用多外观模型以适应变化的多样性,跟踪算法更加鲁棒(Animal 视频包含运动突变和背景复杂,Shaking 视频包括快速运动). L1 方法利用稀疏重构误差更新外观模型,并利用琐屑模板应对部分遮挡问题,跟踪算法在缓慢外观变化和部分遮挡情况下适应性较好(Singer 视频包含大的光线变化). 这些方法分别从不同的角度设计外观模型,均能较好地解决跟踪中的特定问题,而产生式模型机制自身的限制使得它们需要设计非常复杂的外观模型才能可靠地使目标和背景分离. 此外,试图设计能够适应目标外观不确定性变化的外观模型仍是一个非常困难的研究问题.

4 判别式外观模型学习跟踪方法

判别式模型将跟踪转化成二分类问题,故外观模型具有很好的区分能力. 目前,判别式模型学习跟踪方法依据分类器设计方式不同被分为:1) 基于 SVM 的外观模型;2) 基于判别分析的外观模型;3) 基于 Boosting 的外观模型.

4.1 基于 SVM 分类的外观模型跟踪

基于 SVM 的跟踪需要将目标和背景转换成正、负训练样本,其中正样本一般来自目标,负样本来自目标周围区域,图像信息向训练样本的转化如图 2 所示. 然后,采用 SVM 方法训练标签样本以获得判别式模型,再结合学习方法更新外观模型实现自适应跟踪.

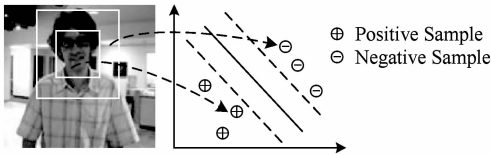


Fig. 2 The mapping from images information to training samples.

图 2 图像信息转化训练样本效果

最早文献[57]提出将跟踪问题转换成基于 SVM 的二分类问题,采用离线训练方式获得判别式模型,该方法为视频跟踪提供了一个新的研究方向. 为了使算法具有自适应性,文献[58]将 SVM 分类

器与 Adaboost 方法结合,从前一帧获取样本训练分类器实现对当前帧的目标分离;而文献[12]将 SVM 分类器与 Ensemble 方法结合,从关键帧获取样本训练分类器实现对当前帧目标的分离,这两种方法均能较好适应外观形变大的跟踪情况. 然而,基于 SVM 的判别式方法将跟踪中的噪声干扰和形变等问题转换成训练样本的有效选择问题,上述几种跟踪算法均是直接对单样本进行训练,故一旦样本选取不准确容易造成错误学习,导致跟踪失败.

为了提高样本的选择质量,文献[59]将样本集合以图的形式参与分类器更新训练,它强调了样本之间的高维约束关系,外观模型能够更好适应复杂外观变化和噪声干扰,文献[60]将样本依据与外观模型的相似性程度指定等级,再参与分类器训练更新外观模型. 这两种方法以不同约束将样本组成集合的形式参与训练,提高了分类器的分类弹性,增强了外观模型的适应能力.

虽然,上述方法通过训练样本约束机制提高了分类器学习精度,但它们采用的均是自学习方式. 随着视频目标和跟踪环境的变化,无法保证判别式模型的分类能力不被蜕化,文献[61]提出将 SVM 和半监督学习方法相结合,模型的更新过程受到一定的监督,分类器能力得到保持,有效地缓解了轨迹漂移问题.

4.2 基于判别式分类的外观模型跟踪

判别式分析建模方法是一种监督学习方法,它旨在寻找最优低维子空间,该子空间具有小的类内距离和大的类间距离,通过对判别式模型在线更新,以实现目标跟踪的自适应性.

早期基于判别式分析的外观模型是针对目标和背景的不同纹理进行建模^[62],在线性判别式分析(linear discriminant analysis, LDA)框架下实现目标跟踪,能够适应外观变形严重和光线突变等跟踪环境. 但该模型需要设定外观变化符合高斯分布规律,而这种假设限制了模型的表征能力. 为此,文献[63]采用增量 2D-LDA 方法,直接对图像信息在子空间下建立判别式外观模型,相对 LDA 而言跟踪效率得以提高. 最近,视觉显著性机制^[64]被用于设计判别式模型,动态识别具有强区分力的注意区域,并将其作为外观模型,模型表征能力大大提高. 但该模型需要手动选择初始特征,故文献[65]采用自动初始化显著区域,并利用判别式学习在每帧中获得强区分力的显著性特征设计外观模型. 然而,在实际中用于产生监督模型的标签样本并不容易获得,

这将导致目标和背景类别方差矩阵评估出现误差, 从而影响视频跟踪效果。

为了解决标签样本获取困难的问题, 半监督判别式方法 (semi-supervised discriminant analysis, SDA)^[66] 被提出, 它不仅利用了标签样本能够体现类别信息的特性, 而且利用了未知样本对类别信息内在约束的评估信息, 通过对大量未知样本的利用减少了对标签样本需求, 在一定监督机制下保证了外观模型的区分能力。文献[67]扩展 SDA 到 Online Boosting 框架, 提出基于 SDA 的增量学习机制, 使外观模型能够很好地处理相似目标干扰、严重遮挡和亮度变化等跟踪问题。近些年, 研究者发现传统预定义的距离度量会在目标和跟踪环境发生变化时出现不适应, 如图 3 所示, 相同的两类目标在不同的度量空间下会获得不同的分类结果。

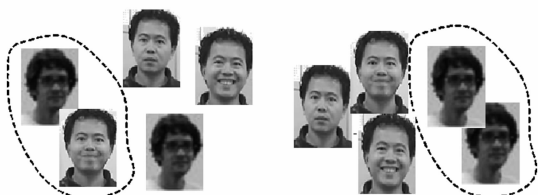


Fig. 3 The classified results of different metric spaces.

图 3 不同度量空间下分类结果对比

度量空间建模方法将原始特征空间映射到新的度量空间下, 获得目标和背景的高可分性, 它旨在寻找能够具备最大类间距离的度量空间, 然后更新度量空间, 保证模型始终具有高的分类能力, 以实现跟踪方法的自适应性。文献[68]首次将度量空间学习融入判别式外观模型, 并提出了基于度量学习的视频跟踪框架。它指出传统判别式分析方法仅强调优化边界决策函数, 而忽视目标外观流型结构的问题, 在寻找最优度量空间过程时, 充分考虑了目标的结构信息, 获得好的分类效果, 能够适应目标剧烈亮度变化、部分遮挡和摄像机移动等跟踪情况。然而, 该模型的映射变化需要遵循线性化原则, 实际中原始特征空间并不一定都能通过线性变化获得可分离的度量空间。

针对该问题, 文献[16]提出非线性核度量学习方法, 将映射变化从线性拓展成非线性, 使外观模型的表征能力变得通用化, 相对文献[68]而言, 它对具有相似纹理的目标和背景, 分类能力更强。另外, 为了缓解轨迹漂移问题, 文献[69]将度量学习与半监督学习相结合, 自适应训练前景和背景样本获得最优的判别式模型, 能够处理目标大尺度变化的跟踪环境。

在基于度量学习的跟踪方法研究中, 研究者发现如何确定度量空间的维数是跟踪成败的关键因素。于是, 文献[70]提出稀疏正则化的度量学习 (sparsity regularization for metric learning, SRML) 方法, 它能够对度量空间的维数进行评估, 保证了度量空间在最低维数的情况下获得最优的分类效果, 这一理论的提出促进了度量学习跟踪方法的研究进程。

4.3 基于 Boosting 分类的外观模型跟踪

Boosting 方法是将弱学习提升为强学习的方法总称, 而 Boosting 训练是以离线或在线的方式选择弱分类器, 并将其混合成强分类器, 通过分类误差最小化实现目标和背景的分离, 结合学习策略实现视频跟踪的自适应。

4.3.1 自学习的 Boosting 跟踪方法

基于 Boosting 的跟踪方法大量涌现是因为它能够实时提炼最优的目标特征以设计判别式外观模型, 故跟踪算法的鉴别能力较强。但传统方法中需要训练大量的弱分类器以获得强分类器建立外观模型, 这个提炼过程比较耗时。

针对此问题, 文献[17]提出采用特征选择器代替弱分类器构造外观模型, 这样会减少强分类器形成所需的迭代次数, 跟踪算法对姿态、亮度变化和遮挡有很好的适应能力。但该模型在特征提炼过程中, 没有考虑其相互间的结构关系, 易造成特征冗余, 违背了融合特征互补的原则。集合分类器跟踪方法^[71]通过加权组合弱分类器形成强分类器, 充分考虑特征对分类器的影响关系, 外观模型对相似背景、摄像机移动、姿态变化和部分遮挡有很好的适应性。

然而, 上述方法均是采用新的弱分类器代替旧的弱分类器方式, 实现外观模型更新, 这个过程在每帧中需要进行大量特征存储和计算, 且每次取代弱分类器的合适数目至今无法确定。文献[72]提出直接对弱分类器本身进行更新的思想, 能够适应相似目标交互干扰和部分遮挡等跟踪问题, 给跟踪算法研究提供一个新的方向。

经研究发现, 围绕分类器获取速度和分类器复杂性两方面改进的方法出现最多, 但依据什么原则能获得最优的分类器成为焦点问题。为解决此问题, 文献[18]采用 Wald 序列决策理论, 从理论框架上指出了跟踪算法中分类器设计的原则, 给 Boosting 跟踪方法的研究提供了重要的理论支撑。

4.3.2 半监督学习的 Boosting 跟踪方法

采用自学习方式的 Boosting 跟踪方法容易造成轨迹漂移问题, 而采用监督学习的跟踪方法又会面临

标签样本资源匮乏的问题. 半监督学习方法能够解决少量标签样本和大量未知样本组成样本集的分类问题,成为外观模型学习跟踪方法研究的新方向.

文献[73]首次将半监督学习引入 Boosting 框架设计外观模型,有效地缓解了轨迹漂移问题. 但该方法在选择标签样本时仅仅从第 1 帧提取,随着目标外观的变化监督机制将失去作用,导致跟踪失败. 半监督集合跟踪方法^[74]通过在每一帧均进行提炼标签样本,以监督在下一帧中目标分类的结果,同时采用多分类器决策的方式提高分类器的学习精度,跟踪效果明显增强. 然而,这些方法选取单一的标签样本进行更新分类器训练,容易造成因监督模型构造不准确而导致跟踪失败的问题.

同样,为了提高标签样本的选择质量,文献[75]提出跟踪-学习-检测相结合(tracking-learning-detect-

tion, TLD)方法,利用正负(positive-negative, PN)轨迹约束引导选择训练样本,提高标签样本质量,在外观变化程度小的情况下能够适应部分或全部遮挡、目标丢失再跟踪和轨迹漂移等情况. 文献[76]提出多事例学习(multiple instance learning, MIL)跟踪方法,利用样本包的形式代替单个样本作为标签样本,提高标签样本质量,在目标姿态变化、复杂背景和轨迹漂移等情况下适应能力较强,这两种方法提供了标签样本选择机制的新思想,成为经典的半监督跟踪算法. 随后,围绕 MIL 的很多改进方法被提出:文献[77]采用梯度 MIL 方法和文献[78]提出加权 MIL 方法,均使半监督跟踪方法得到进一步完善. 表 3 给出判别式外观模型学习典型算法的应用对比情况(代表性视频是指各种方法能够顺利执行的跟踪环境,资源可以从作者网站获得):

Table 3 Summary of the State-of-the-art Tracking Algorithms Using Discriminative Models
表 3 不同场景下典型判别式跟踪方法性能分析

Method	Model	Feature	Vedio	Tracking Performance
Struck ^[13]	SVM Classifier	Haar-like	Coke,Faceocc2,Tiger1,Animal,David2,Sylvester	Adapt severe occlusion, pose changes and abrupt motion; Adapt drastic changes hardly
Metric ^[69]	Metric Space	EOH LBP	Cock, Sylveater, Surf, Tiger2	Adapt pose, illumination and scale changes Adapt partial occlusion and fast motion
DST ^[65]	Attention Region	Motion	Surf, Skiing	Adapt large pose and illumination changes; Adapt motion blur hardly
Boosting ^[17]	Self-Learning	Haar Hog; Lbp	Cock, Jumping, Animal, David2	Adapt pose, illumination and scale changes; Adapt views changes and cluttered background hardly
Semi-bosting ^[69]	Boosting Classifier	Haar-like	Sylvester, Surf, Jumping, shaking	Adapt partial occlusion and pose changes; Adapt drastic changes for nonrigid object hardly
TLD ^[71]	PN Learning	2bp	Sylvester, David2, Faceocc2, Jumping, Panda	Adapt severe occlusion; Alleviate drift; Adapt abrupt motion for nonrigid object hardly
MIL ^[72]	Boosting Classifier	Haar	Faceocc2, Panda, Shaking, Tiger2, Jumping	Adapt pose, illumination changes and partial occlusion; Alleviate drift; Adapt abrupt motion hardly

表 3 中,Struck 和 Metric 分别从约束支撑向量数目(Coke 视频包含无纹理特征和遮挡)和构造最优度量空间(Surf 视频包含快速运动、遮挡和尺度变化)两个方面提高分类器的性能,以实现自适应跟踪效果;DST 和 Boosting 分别从视觉注意机制和迭代提炼机制获取最具有鉴别力的外观模板,以可靠分离目标和背景;TLD 和 MIL 分别从 PN 轨迹约束和样本包形式获取正确的训练样本,以提高分类器性能实现跟踪(Panda 视频包含姿态、亮度变化和完全遮挡). 这些方法从不同角度提高了目标和背景的分离程度,并对模型实时更新实现鲁棒的跟踪效

果. 然而,基于判别式跟踪方法容易受噪声干扰,外观模型的通用性相对产生式模型稍弱一些. 此外,样本的有效选择机制仍有待深入研究.

5 产生-判别混合式模型学习

混合式模型旨在融合产生式和判别式各自的优点实现自适应跟踪,从混合方式不同可以分为单层和多层两类方法.

1) 单层混合方式. 文献[79]采用 PCA 方法对正样本在子空间下建立高斯产生式外观模型,再通过

Fisher 判别分析建立混合式外观模型,实现串联式的模型设计,模型对跟踪目标的姿态和亮度出现大变化时能够适应.文献[20]正好相反,先采用判别式建模再通过产生式分析建立外观模型,其对部分遮挡跟踪效果较好.这种单层混合式方法产生式和判别式模型设计过程相互独立,不能有效体现模型优势互补的目标.

2) 多层混合方式.文献[21,80]分别采用产生式模型和 SVM 判别式模型在 Co-training 框架下实

现交互混合,前者模型对部分遮挡和轨迹漂移效果较好,而后者对目标被全部遮挡丢失后再跟踪适应力强.另外,文献[81]提出利用学习混合、分类器混合和决策混合三级分层的结合框架来混合产生式和判别式外观模型实现跟踪,通过各层相互辅助实现鲁棒跟踪,其对目标外观姿态变化、部分遮挡、摄像机移动等跟踪问题适应性较好,但混合式模型跟踪方法中融合方式或参数确定困难.表 4 给出了混合式模型的分类及性能分析:

Table 4 Summary of the State-of-the-art Tracking Algorithms Using Hybrid Generative-Discriminative Models

表 4 产生-判别式混合跟踪方法性能分析

Model	Form	Tracking Performance
Single-layer	Series Self-learning ^[20,75]	The generative model is obtained in light of appearance without considering background information, so they would easily fail within cluttered background. The discriminative model aim to find a decision boundary to separate the object from the background, if enough training data is available they would be more performance. Hybrid tracking methods benefit from both types of methods.
Two-layer	Interactive ^[21,76]	
Three-layer	Series Self-learning ^[77]	

6 总结与展望

6.1 总 结

本文首先对近些年视频跟踪中常被用到的目标特征描述方式进行总结,将目标描述的方式分为单一特征描述和融合特征描述两种,其典型算法具体对比情况如表 1 所示.然后分别对基于产生式、判别式和混合式的外观模型分析与对比如表 5 所示,其模型适应情况总结如下:

1) 从整体上讲,产生式模型能够描述目标的全局性特征,模型通用性好,但没有考虑跟踪场景背景变化因素;判别式模型对目标和背景的分离效果较好,模型区分力强,但容易陷于过分类现象,通用性相对较弱;混合式模型旨在结合两者优点,但结合方式参数的设计比较困难,研究成果较少.

2) 从 3 种建模方式的具体代表性模型来讲,分量混合模型设计简单,但对外观变化的约束比较严格;线性子空间模型在低维空间描述目标特征,因线性降维理论成熟其应用情况较好;非线性子空间模型拓宽了子空间方法的使用范围,但其增加了计算代价且模型自适应性设计难度增大.基于 Boosting 外观模型从最优特征提炼角度建模,保证了目标与背景的可分性能力,在机器学习理论的推动下出现大量的研究成果,但很多情况下模型会受到标签样本的噪声干扰导致能力蜕化;SVM 作为经典的小样本分类方法,基于它设计的外观模型通用性较强,对噪声干扰相对鲁棒;基于判别式的模型因为其采用监督子空间学习实现自适应,故模型的在线区分力相对弱一些;而度量空间模型从目标特征空间的转化寻找高可分性的度量空间,保证了外观模型的区分能力,跟踪效果较好.

Table 5 Summary of the Characters of Tracking Algorithms Using Different Appearance Models

表 5 不同外观模型跟踪方法特点对比

Model	Parameter Mixture	Linear Subspace	Nolinear Subspace	SVM	Discriminant Analysis	Boosting	Model Mixture
Learning Form	Parameters Update	Basis or Mean Update	Manifold Learning	Classifier Update	Supervised Learning	Classifier Update	Multi-Learning
Key Problem	Mixed Form	Basis Function	Nonlinear Form	Training Data Selection	Good Mapping Form	Weak Classifier Form	Hybrid Form
Model Form	Component Combination	Vector; Matrix; Sparse	Multi-subspace; Manifold	Self-learning; Semi-supervised Learning	Visual Attention; Discriminant; Metric Space	Self-learning; Semi-supervised Learning	Series; Interactive
Adaptation	Strong	Strong	Weak	Strong	Strong	Strong	General

6.2 展 望

基于本文对各种外观模型学习跟踪方法的研究,未来发展的主要研究问题和发展趋势总结如下:

1) 目标特征描述典型算子有 SIFT, HOG 和 LBP 等,但单一的描述方式对于视频目标中外观的不确定性变化难以描述,故很多采用如协方差矩阵、Sigma Set 和 Boosting 等方式融合特征描述目标获得好的效果,但其确定融合参数需要较大的计算量. 如何设计简单高效且鲁棒的特征融合方式,尤其是对 Boosting 特征提炼方式的改进有待深入研究.

2) 外观模型学习跟踪过程中会面临跟踪精确性与鲁棒性的矛盾问题,如果要求跟踪精确度较高就会对模型设计约束苛刻,因而导致算法鲁棒性降低;相反,如果要求鲁棒性较高,那么外观模型的弹性能力就会增大,影响了跟踪的精确度,如何平衡这一组矛盾问题将是一个研究方向.

3) 产生式跟踪方法侧重目标自身建模和学习,当目标运动过快时,相似干扰和复杂背景出现时跟踪效果会下降;判别式跟踪强调目标和背景分离,容易受噪声干扰且出现过分类现象影响跟踪效果,两种外观模型设计方法如何更好地混合,相互辅助实现视频跟踪还有待研究.

4) 判别式模型跟踪中无论是 SVM 方法、判别式分析还是 Boosting 分类方法中,均面临一个如何正确选择训练样本更新分类器的问题. TLD 方法通过 P-N 轨迹约束提高训练样本选取质量, MIL 系列方法以样本包代替单样本提高分类器弹性. 故如何设计训练样本选取规则保证分类器学习准确性将是外观模型学习的研究方向.

5) 模型学习跟踪方法中不同的机器学习方式对模型的自适应能力有着严重的影响,面对目标在视频中运动和变化的不确定性,已有的子空间学习、稀疏字典学习、度量空间学习及最新出现的深度学习都在跟踪中呈现好的表现. 如何结合学习机制更好实现跟踪算法自适应性有待深入研究.

参 考 文 献

- [1] Adam A, Rivlin E, Shimshoni I. Robust fragments-based tracking using the integral histogram [C] // Proc of the 19th IEEE Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2006: 798-805
- [2] Comaniciu D, Ramesh V, Meer P. Kernel-based object tracking [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2003, 25(5): 564-575
- [3] Liang D, Huang Q, Jiang S, et al. Mean-shift blob tracking with adaptive feature selection and scale adaptation [C] // Proc of the 11th IEEE Int Conf on Computer Vision. Los Alamitos, CA: IEEE Computer Society, 2007: 369-372
- [4] Ning J, Zhang L, Zhang D, et al. Scale and orientation adaptive mean shift tracking [J]. Computer Vision, IET, 2012, 6(1): 52-61
- [5] Yu T, Wu Y. Differential tracking based on spatial-appearance model (SAM) [C] // Proc of the 19th IEEE Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2006: 720-727
- [6] Han B, Davis L. On-line density-based appearance modeling for object tracking [C] // Proc of the 10th IEEE Int Conf on Computer Vision. Los Alamitos, CA: IEEE Computer Society, 2005: 1492-1499
- [7] Wang H, Suter D, Schindler K, et al. Adaptive object tracking based on an effective appearance filter [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2007, 29(9): 1661-1667
- [8] Ross D, Lim J, et al. Incremental learning for robust visual tracking [J]. International Journal Computer Vision, 2008, 77(1): 125-141
- [9] Wen L, Cai Z, Lei Z, et al. Online spatio-temporal structural context learning for visual tracking [G] // LNCS 7575: Proc of European Conf on Computer Vision. Berlin: Springer, 2012: 716-729
- [10] Kwon J, Lee K. Visual tracking decomposition [C] // Proc of the 23rd IEEE Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2010: 1269-1276
- [11] Lim H, Morariu V, Camps O, et al. Dynamic appearance modeling for human tracking [C] // Proc of the 19th IEEE Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2006: 751-757
- [12] Tian M, Zhang W, Liu F. On-line ensemble SVM for robust object tracking [G] // LNCS 4843: Proc of Asian Conf of Computer Vision. Berlin: Springer, 2007: 355-364
- [13] Sam H, Amir S, et al. Struck: Structured output tracking with kernels [C] // Proc of the 13th IEEE Int Conf on Computer Vision. Los Alamitos, CA: IEEE Computer Society, 2011: 263-271
- [14] Yang F, Lu H, Chen Y. Human tracking by multiple kernel boosting with locality affinity constraints [G] // LNCS 6495: Proc of Asian Conf of Computer Vision. Berlin: Springer, 2010: 39-50
- [15] Fan J, Wu Y, Dai S. Discriminative spatial attention for robust tracking [G] // LNCS 6311: Proc of European Conf on Computer Vision. Berlin: Springer, 2010: 480-493
- [16] Jiang N, Liu W, Wu Y. Learning adaptive metric for robust visual tracking [J]. IEEE Trans on Image Processing, 2011, 20(8): 2288-2291

- [17] Grabner H, Grabner M, Bischof H. Real-time tracking via online boosting [C] // Proc of the 17th British Machine Vision Conf. Malvern UK: British Machine Vision Association Press, 2006: 47-56
- [18] Grabner H, Sochmanetal J. Training sequential on-line boosting classifier for visual tracking [C] //Proc of the 19th Int Conf on Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2008: 1-4
- [19] Shen Dingcheng, Xue Yanbing, et al. Research on robust object tracking algorithm based on on-line boosting [J]. Journal of Optoelectronic • Laser, 2013, 24(1): 170-175 (in Chinese)
(沈丁成, 薛彦兵, 等. 一种鲁棒的基于在线 boosting 目标跟踪算法研究[J]. 光电子 • 激光, 2013, 24(1): 170-175)
- [20] Woodley T, Stenger B, Cipolla R. Tracking using online feature selection and a local generative model [C] // Proc of the 18th British Machine Vision Conf. Malvern, UK: British Machine Vision Association Press, 2007: 1-8
- [21] Dinh T, Medioni G. Co-training framework of generative and discriminative trackers with partial occlusion handling [C] // Proc of the 11th Winter Conf on Applications of Computer Vision. Los Alamitos, CA: IEEE Computer Society, 2011: 642-649
- [22] Pérez P, Hue C, Vermaak J, et al. Color-based probabilistic tracking [C] // Proc of the 7th European Conf on Computer Vision. Berlin: Springer, 2002: 661-675
- [23] Zhao Q, Tao H. Object tracking using color correlogram [C] // Proc of the 2nd Joint IEEE Int Workshop on VS-PETS. Los Alamitos, CA: IEEE Computer Society, 2005: 263-270
- [24] Leichter I, Lindenbaum M, Rivlin E. Mean shift tracking with multiple reference color histograms [J]. Computer Vision and Image Understanding, 2010, 114(3): 400-408
- [25] Zhou H, Yuan Y, Shi C. Object tracking using SIFT features and mean shift [J]. Computer Vision and Image Understanding, 2009, 113(3): 345-352
- [26] Song Y, Li C, Wang L, et al. Robust visual tracking using structural region hierarchy and graph matching [J]. Neurocomputing, 2012, 89(7): 12-20
- [27] Miao Q, Wang G, Shi C, et al. A new framework for on-line object tracking based on SURF [J]. Pattern Recognition Letters, 2011, 32(13): 1564-1571
- [28] Choi H, Oh S. Robust segment-based object tracking using generalized hyperplane approximation [J]. Pattern Recognition, 2012, 45(8): 2980-2991
- [29] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features [C] Proc of the 15th IEEE Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2001: 511-518
- [30] Sivaraman S, Mohan M. A general active-learning framework for on-road vehicle recognition and tracking [J]. IEEE Trans on Int Transportation Systems, 2010, 11(2): 267-276
- [31] Santner J, Leistner C, Saffari A, et al. Prost: Parallel robust online simple tracking [C] //Proc of the 23rd IEEE Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2010: 723-730
- [32] Wu Gang, Tang Zhenmin, Cheng Yong, et al. Object tracking method based on gray level co-occurrence matrix texture characteristic [J]. Journal of Nanjing University of Science and Technology: Natural Science, 2010, 34 (4): 459-464 (in Chinese)
(吴刚, 唐振民, 程勇, 等. 灰度共生矩阵纹理特征的运动目标跟踪方法[J]. 南京理工大学学报:自然科学版, 2010, 34 (4): 459-464)
- [33] Song H, Wang Y. Real-time local texture feature target tracking method [C] //Proc of the 4th Int Congress on Image and Signal Processing. Los Alamitos, CA: IEEE Computer Society, 2011: 419-422
- [34] Tong M, Han H, Lei J. Efficient visual tracking by using LBP descriptor [C] // Proc of the 4th Artificial Intelligence and Computational Intelligence. Berlin: Springer, 2012: 391-399
- [35] Wang Q, Chen F, Xu W. Adaptive multi-cue tracking by online appearance learning [J]. Neurocomputing, 2011, 74 (6): 1035-1045
- [36] Wang Y, Tang X, Cui Q. Dynamic appearance model for particle filter based visual tracking [J]. Pattern Recognition, 2012, 45(12): 4510-4523
- [37] Liu Chenguang, Liu Jiafeng, et al. Human pose tracking based on partitioned sampling particle filter and multiple cues fusion [J]. Journal of Computer Research and Development, 2011, 48(12): 2359-2368 (in Chinese)
(刘晨光, 刘家锋, 等. 多特征融合的分块采样粒子滤波算法在人体姿态跟踪中的应用[J]. 计算机研究与发展, 2011, 48 (12): 2359-2368)
- [38] Tuzel O, Porikli F, Meer P. Region covariance: A fast descriptor for detection and classification [C] //LNCS 3952: Proc of European Conf on Computer Vision. Berlin: Springer, 2006: 589-600
- [39] Wu Y, Cheng J, Wang J, et al. Real-time probabilistic covariance tracking with efficient model update [J]. IEEE Trans on Image Processing, 2012, 21(5): 2824-2838
- [40] Hong X, Chang H, Shan S, et al. Sigma set: A small second orderstatistical region descriptor [C] // Proc of the 22nd IEEE Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2009: 1802-1809
- [41] Hong X, Chang H, Shan S, et al. Sigma set based implicit online learning for object tracking [J]. IEEE Signal Processing Letters, 2010, 17(9): 807-810
- [42] Grabner H, Bischof H. On-line boosting and vision [C] // Proc of the 19th IEEE Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2006: 260-267

- [43] Stalder S, Grabner H, Van G. Beyond semi-supervised tracking: Tracking should be as simple as detection, but not simpler than recognition [C] // Proc of the 43rd IEEE Int Conf on Computer Vision. Los Alamitos, CA: IEEE Computer Society, 2009: 1409-1416
- [44] Jepson A, Fleet D, El-Maraghi T. Robust online appearance models for visual tracking [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2003, 25(10): 1296-1311
- [45] Ho J, Lee K, Yang M, et al. Visual tracking using learned linear subspaces [C] //Proc of the 17th IEEE Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2004: 1-8
- [46] Ross D, Lim J, Yang Ming. Adaptive probabilistic visual tracking with incremental subspace update [C] //LNCS 3022: Proc of European Conf on Computer Vision. Berlin: Springer, 2004: 470-482
- [47] Li X, Hu W, Zhang Z, et al. Robust visual tracking based on an effective appearance model [G] // LNCS 5305: Proc of European Conf on Computer Vision. Berlin: Springer, 2008: 396-408
- [48] Tong Xiaomin, Zhang Yanning, Yang Tao. Robust object tracking based on adaptive and incremental subspace learning [J]. Acta Automatica Sinica, 2011, 37(12): 1483-1495 (in Chinese)
(仝小敏, 张艳宁, 杨涛. 基于增量子空间自适应决策的目标跟踪[J]. 自动化学报, 2011, 37(12): 1483-1495)
- [49] Li X, Hu W, Zhang Z, et al. Robust visual tracking based on incremental tensor subspace learning [C] // Proc of the 11th IEEE Int Conf on Computer Vision. Los Alamitos, CA: IEEE Computer Society, 2007: 1-8
- [50] Wen Jing, Li Jie, Gao Xinbo. Adaptive object tracking with incremental tensor subspace learning [J]. Acta Automatica Sinica, 2009, 37(7): 1618-1623 (in Chinese)
(温静, 李洁, 高新波. 基于增量张量子空间学习的自适应目标跟踪[J]. 电子学报, 2009, 37(7): 1618-1623)
- [51] Wen J, Gao X. Incremental learning of weighted tensor subspace for visual tracking [C] // Proc of the 35th System Man and Cybernetics. Los Alamitos, CA: IEEE Computer Society, 2009: 3688-3693
- [52] Ding J, Huang Y, Huang K. Robust object tracking via online learning of adaptive appearance manifold [C] // Proc of the 13th IEEE Int Conf on Computer Vision. Los Alamitos, CA: IEEE Computer Society, 2011: 1863-1870
- [53] Wright J, Ma Y, Mairal J, et al. Sparse representation for computer vision and pattern recognition [J]. Proceedings of the IEEE, 2010, 98(6): 1031-1044
- [54] Xue M, Ling H. Robust visual tracking using l1 minimization [C] // Proc of the 12th IEEE Int Conf on Computer Vision. Los Alamitos, CA: IEEE Computer Society, 2009: 1436-1443
- [55] Chen F, Wang Q, Wang S, et al. Object tracking via appearance modeling and sparse representation [J]. Image and Vision Computing, 2011, 29(11): 787-796
- [56] Bai T, Li Y. Robust visual tracking with structured sparse representation appearance model [J]. Pattern Recognition, 2012, 45(6): 2390-2404
- [57] Avidan S. Support vector tracking [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2001, 26(8): 73-77
- [58] Wang Zhenyu, Zhang Kedai, Wu Yi, et al. Target tracking in infrared image sequences by combining SVM and AdaBoost [J]. Journal of Image and Graphics, 2007, 12(11): 2052-2057 (in Chinese)
(王震宇, 张可黛, 吴毅, 等. 基于 SVM 和 AdaBoost 的红外目标跟踪[J]. 中国图象图形学报, 2007, 12(11): 2052-2057)
- [59] Li X, Dick A, Wang H, et al. Graph mode-based contextual kernels for robust SVM tracking [C] //Proc of the 13th IEEE Int Conf on Computer Vision. Los Alamitos, CA: IEEE Computer Society, 2011: 1156-1163
- [60] Bai Y, Tang M. Robust tracking via weakly supervised ranking SVM [C] //Proc of the 25th IEEE Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2012: 1854-1861
- [61] Tang F, Brennan S, Zhao Q, et al. Co-tracking using semi-supervised support vector machines [C] //Proc of the 11th IEEE Int Conf on Computer Vision. Los Alamitos, CA: IEEE Computer Society, 2007: 1-8
- [62] Nguyen H, Smeulders A. Robust tracking using foreground-background texture discrimination [J]. International Journal of Computer Vision, 2006, 69(3): 277-293
- [63] Li G, Liang D, Huang Q, et al. Object tracking using incremental 2d-lda learning and Bayes inference [C] //Proc of the 15th Int Conf on Image Processing. Los Alamitos, CA: IEEE Computer Society, 2008: 1568-1571
- [64] Yang M, Yuan J, Wu Y. Spatial selection for attentional visual tracking [C] //Proc of the 23rd IEEE Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2007: 1-8
- [65] Mahadevan V, Vasconcelos N. Automatic initialization and tracking using attentional mechanisms [C] //Proc of the 24th IEEE Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2011: 15-20
- [66] Cai D, He X, Han J. Semi-supervised discriminant analysis [C] //Proc of the 11th IEEE Int Conf on Computer Vision. Los Alamitos, CA: IEEE Computer Society, 2007: 1-7
- [67] Wang H, Hou X, Lin C. Boosting incremental semi-supervised discriminant analysis for tracking [C] //Proc of the 17th Int Conf on Image Processing. Los Alamitos, CA: IEEE Computer Society, 2010: 2748-2752
- [68] Wang X, Gang H, Han T. Discriminative tracking by metric learning [G] // LNCS 6313: Proc of European Conf on Computer Vision. Berlin: Springer, 2010: 200-214

- [69] Cong Y, Yuan J, Tang Y. Object tracking via online metric learning [C] //Proc of the 19th Int Conf on Image Processing. Los Alamitos, CA: IEEE Computer Society, 2012: 417-420
- [70] Jiang N, Liu W, Wu Y. Order determination and sparsity-regularized metric learning for adaptive visual tracking [C] // Proc of the 25th IEEE Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2012: 1956-1963
- [71] Avidan S. Ensemble tracking [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2007, 29(2): 261-271
- [72] Parag T, Porikli F, Elgammal A. Boosting adaptive linear weak classifiers for online learning and tracking [C] //Proc of the 21st IEEE Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2008: 1-8
- [73] Grabner H, Leistner C, Bischof H. Semi-supervised on-line boosting for robust tracking [G] //LNCS 5302: Proc of European Conf on Computer Vision. Berlin: Springer, 2008: 234-247
- [74] Liu H, Sun F. Semi-supervise ensemble tracking [C] // Proc of the 34th Int Conf on Acoustics, Speech and Signal Processing. Los Alamitos, CA: IEEE Computer Society, 2009: 1045-1049
- [75] Zdenek K, Mikolajczyk K, Matas J. Tracking-learning-detection [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2012, 29(1): 1409-1422
- [76] Babenko B, Yang M, Belongie S. Robust object tracking with online multiple instance learning [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2011, 33(8): 1619-1632
- [77] Xie Y, Qu Y, Li C, et al. Online multiple instance gradient feature selection for robust visual tracking [J]. Pattern Recognition Letters, 2012, 33(9): 1075-1082
- [78] Zhang K, Song H. Real-time visual tracking via online weighted multiple instance learning [J]. Pattern Recognition, 2013, 46(1): 397-411
- [79] Lin R, Ross D, Lim J, et al. Adaptive discriminative generative model and its applications [C] // Proc of the 17th IEEE Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2004: 801-808
- [80] Yu Q, Dinh B, Medioni G. Online tracking and reacquisition using co-trained generative and discriminative trackers [G] // LNCS 5303: Proc of European Conf on Computer Vision. Berlin: Springer, 2008: 678-691
- [81] Lei Y, Ding X, Wang S. Visual tracker using sequential Bayesian learning: Discriminative, generative, and hybrid [J]. IEEE Trans on Systems, Man, and Cybernetics-Part B: Cybernetics, 2008, 38(6): 1578-1592



Zhang Huanlong, born in 1981. Received his master degree in computer science from Henan University. Currently PhD candidate in Shanghai Jiao Tong University. His research interests include pattern recognition

and computer vision.



Hu Shiqiang, born in 1969. Professor. Received his PhD degree in Beijing Institute of Technology. His research interests include data fusion technology, image understand and nonlinear filter

(sqhu@sjtu.edu.cn).



Yang Guosheng, born in 1963. Professor. Received his PhD degree in Beijing Institute of Technology. Senior member of China Computer Federation. His research interests include pattern recognition and

emotion recognition (gsyangmuc@163.com).