

目标跟踪中表观建模研究进展

丁建伟¹, 田华伟²

(1. 中国人民公安大学警务信息工程学院, 北京 102623;

2. 中国人民公安大学反恐学院, 北京 102623)

摘要 目标跟踪是计算机视觉中的热点研究问题,具有重要的理论价值和广泛的应用前景。虽然研究者对目标跟踪进行了大量研究,但是目标跟踪在复杂场景下的效果依然难以令人满意。为了设计鲁棒的目标跟踪算法,需要解决的首要问题是如何对目标表观进行建模。近年来,视觉理论和机器学习的蓬勃发展促进了目标表观建模的研究。为了使读者了解最新的表观建模研究进展,首先分析了最近几年典型的目标表观建模方法,并对目标表观建模的一般框架进行了总结,然后分别从目标的视觉描述和模型学习两方面进行了详细的论述,最后讨论了目标表观建模还存在的问题以及未来发展的方向。

关键词 目标跟踪 表观建模 视觉描述 模型学习

中图分类号 D035.33

0 引言

跟踪目标研究是指让计算机自动确定感兴趣目标在连续图像序列中的位置、轨迹以及运动参数等信息^[1-2]。目标跟踪是计算机视觉领域的重要研究内容,也是模式识别的关键环节之一,其结果会促进视频分割、行为分析、场景理解等问题的研究,因而具有重要的科学研究意义。另外,目标跟踪还具有很高的应用价值。例如,目标跟踪可以应用于智能视频分析系统,是智能视频监控系统中承上启下的中间环节,与运动目标提取和场景理解等多个环节密切相关,是系统能够发挥作用的关键步骤。目标跟踪还可以应用于智能交通系统、人机交互和增强现实等多种场合。由于目标跟踪具有重要科学研究意义和应用价值,因而得到了学术界和工业界的广泛关注。

研究者们对目标跟踪进行了大量的研究,提出了很多目标跟踪算法。这其中都涉及到如何对目标进行描述,即目标表观建模。目标表观建模主要通过学习目标的表观数据来构造目标的模型,在跟踪目标时通过比较表观模型与表观数据,计算得到最

优的状态参数。由于实际场景中存在多种挑战,例如快速光照变化、杂乱背景、遮挡和目标姿态变化等,这都会造成目标跟踪失败。一个好的目标表观模型必须至少具备以下特性:(1)对光照、姿态变化和物体遮挡等引起目标表观的变化具有较好的鲁棒性;(2)能精确描述目标的视觉特征,并与周围相似物体具有足够的区分性。因此,目标的表观建模是跟踪算法的核心模块,构造模型的好坏对跟踪的准确性和鲁棒性起着决定性的作用。

目标表观建模作为目标跟踪的核心问题,在国内外已有大量的研究成果和应用。与已有目标跟踪综述文献[1-3]不同的是,本文并未对表观建模方法进行全方位的介绍。限于篇幅,本文侧重于对最新的表观建模算法进行提炼,从一般框架和具体模块以及各模块之间的关系进行深入的分析与阐述,探讨这些方法的优缺点和应用场合,并在此基础上提出未来目标表观建模值得研究的问题。

1 跟踪系统的总体结构

图1是一般跟踪系统的总体结构,主要包括运动建模和表观建模两个模块。其中运动建模对连续

基金项目 2014北京市支持中央高校共建项目公安视听专业建设项目;国家自然科学基金项目(61503388,61402484,61503387);中央高校基本科研业务费资助项目。

作者简介 丁建伟(1984—),男,湖北随州人,博士,讲师。研究方向为计算机视觉、机器学习和视频侦查等。

图像序列中的目标运动状态进行分析,并基于表观模型估计每一帧目标的最优状态参数,输出的跟踪结果再反馈到表观模型,用于更新表观模型。

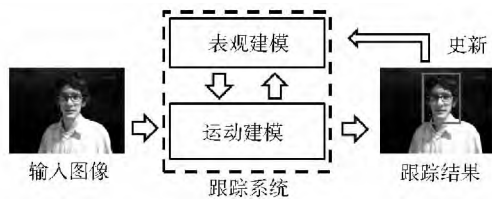


图1 一般目标跟踪系统的总体结构

由图1可知,表观建模是跟踪系统的核心模块之一,在跟踪过程中对跟踪结果的准确性起着至关重要的作用,下面将具体分析。

2 表观建模的一般框架

通过对最近几年提出的表观建模方法进行总结,并结合经典表观建模方法综合分析,我们得出了目标表观建模的一般框架,如图2所示。目标表观建模主要包括视觉描述和模型学习两部分内容^[1-3],其中视觉描述是研究如何从视觉上描述跟踪物体,而模型学习是在视觉描述的基础上研究如何基于学习理论得到跟踪物体的计算模型。进一步细分,视觉描述包括拓扑建模和特征提取,两者分别用来描述目标的拓扑结构和提取目标的视觉特征。模型学习也可细分为统计建模和模型更新,其中统计建模是基于统计学习的方法分析目标的连续视觉特征并建模,而模型更新是利用提取的样本在线更新统计模型。视觉描述是一种对目标的直观描述,是表观建模的基础,而模型学习则是对目标的数学描述,是表观建模的关键。

与一般目标表观建模方法论述不同的是,图2框架突出了拓扑建模和模型更新的重要性。常见的目标跟踪算法一般并没有拓扑建模,如果跟踪目标是刚性物体,则对准确性不会有太大影响,但若跟踪目标是形变物体(例如连续大形变动作的人),由于没有对物体精细的局部结构及相互间关系的描述,跟踪性能有可能会受到较大影响。另外,一般跟踪算法在模型学习时只是简单地根据跟踪结果提取的样本数据更新模型,但模型更新实际上涉及到复杂的样本质量问题,样本质量对于保持模型性能在长时间跟踪中不下降起着关键作用,因此需要单独论述。本文将在后续章节中对这些内容进行更具体的介绍。

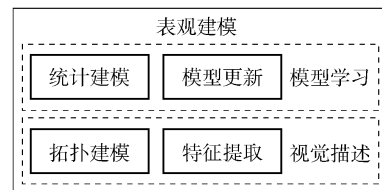


图2 目标跟踪中表观建模的一般框架

3 视觉描述

为了构造鲁棒的表观模型,首先需要对目标进行有效的视觉描述,具体包括拓扑建模和特征提取。需要强调的是,这两部分内容并不是独立存在的,而是相互关联的。

在目标跟踪中,常见的拓扑建模方法有基于全局的视觉模型和基于局部的视觉模型。基于全局的视觉模型是将目标视为整体,侧重描述目标整体的视觉特性。这种建模方式比较简单直观,后续特征提取也较方便,能够较好的满足实时跟踪的要求,因而在大量跟踪算法中得到广泛应用^[4-5]。Ross^[4]将目标表观整体降维到线性子空间,通过在线学习得到目标的本征描述,在跟踪人脸和车等刚性物体时取得了较好的实验效果。Xue Mei^[5]通过求解目标整体表观的稀疏估计来实现目标的跟踪,但是这类方法的缺点是难以很好的处理严重遮挡和非刚性物体的姿态变化,需要融合多种特征或运动信息来改善跟踪效果。基于局部的视觉模型主要利用目标局部的信息,将目标视为局部特征点或局部描述单元的组合^[6-12],在文献^[6-7]中,目标表观由多个局部特征描述子来描述。在文献^[8-11]中,目标表观则由具有一定视觉物理意义的部件表示,此类方法借鉴了目标检测领域著名的可形变部件模型^[30],因而在跟踪大形变非刚性物体时比较有效,并能部分解决遮挡问题。文献^[12]中,作者将目标视为局部超像素的集合,因而也能部分解决遮挡问题。基于局部的视觉模型能够更好的描述目标在局部的视觉统计特性,在处理遮挡和非刚体物体形变等问题时具有潜在的优势,但是需要合理描述局部单元之间的拓扑结构关系,因而模型构造可能更复杂,提取特征的时间也有可能增加。

特征选择和提取对跟踪的性能有重要的影响^[31],但选取哪种特征又是和拓扑建模相关联的。基于全局视觉模型的方法一般是提取全局统计特征,常用的有原始像素特征^[4-5]、直方图特征^[13]、光

流特征^[14]、协方差特征^[15]和活动轮廓^[16]等。而基于局部的视觉模型一般提取底层局部特征或中层视觉描述单元,其中底层局部特征主要包括 SIFT^[17]、Haar^[6-7]、二义码(binary codes)^[18]等,中层视觉描述单元主要包括部件^[8-11]和超像素^[12]等。

目前,目标跟踪中的特征提取存在两点困难。首先,为了使算法能实时处理,目标跟踪算法常使用简单底层特征,例如直接使用目标的像素信息^[4-5],但在复杂多变场景下,算法的鲁棒性面临非常大的挑战。直方图特征^[13]具有很好的不变性,对于目标图像发生旋转变化的鲁棒性较强,但是区分能力较弱,例如无法区分两个灰度直方图相同但内容不同的局部图像块。基于光流特征^[14]可以很容易检测出目标的运动,但是容易受图像噪声影响且耗时长。Haar 特征^[6-7]计算简单,但是包含的特征信息较少,而且很难处理目标形变。SIFT 特征描述子^[17]对光照变化、图像缩放、旋转、放射扭曲以及噪声干扰等都具有一定的不变性,但是存在匹配成功点数量较少、分布不均匀以及速度慢等缺点。SURF 特征使用积分图像近似于 SIFT,匹配性能有所下降,但是速度更快。而利用高层语义属性信息可以有效改善底层特征的适应性问题。例如, Danelljan 等人^[19]提出的基于颜色属性构造表观模型,相比基于底层颜色特征来说跟踪效果会更好。另外,每种跟踪算法使用的特征都有其各自优缺点和最佳应用场景。目前,人们还很难找到适用于任意场景和物体的特征,虽然融合多种特征可以在一定程度上改善跟踪性能,但是算法处理时间也会相应增加,跟踪速度会受到较大影响。近年来,基于深度学习的特征提取方法在物体识别和检测等领域取得了巨大成功, Naiyan Wang 等人^[20]将其应用于目标跟踪也取得了很好的效果。

4 模型学习

模型学习包括统计建模和模型更新两个部分,它是表观建模成功用于目标跟踪的关键步骤。这两方面的内容是紧密联系的,但为了强调模型更新的重要性,这里将对其进行深入讨论。

统计建模主要利用统计学习的方法对提取的特征进行分析,在线学习得到目标的数学模型。根据统计建模的方式,目标的表观模型一般可分为生成式模型和判别式模型。基于生成式模型的目标跟踪算法是在每一帧中寻找与目标模型最相似的区域。

典型的生成式模型有子空间模型^[4]和稀疏表达模型^[5]等。子空间模型^[4]是将目标表观数据从高维空间映射到低维子空间,通过在线学习得到目标表观的特征基,这种模型对光线变化等具有较好的鲁棒性。稀疏表达模型^[5]则是通过在线学习得到目标表观的稀疏表示,在处理图像遮挡和污染时具有较好的效果。基于判别式模型的目标跟踪算法是将跟踪视为前景和背景的二分类问题,通过在线学习得到的分类器,在当前帧中搜索与背景最具区分度的前景区域。这种跟踪方法有时也被称作基于检测的跟踪。典型的判别式模型包括集成学习模型^[6-7]、支持向量机模型^[21-23]、神经网络模型^[20-24]等。支持向量机是一种经典的机器学习算法,它在解决小样本、非线性和高维模式方面表现出很多优势,具有很好的泛化能力。Avidan^[21]使用离线训练的支持向量机模型嵌入到光流跟踪器上,建立跟踪器和分类器之间的联系。之后 Avidan^[32]又提出了基于在线集成学习的跟踪方法,在每一帧中收集前景和背景的像素信息,并通过自适应提升算法训练弱分类器,得到强分类器,但是该方法并没有真正做到在线的更新分类器。Grabner 等人^[6]提出了一种基于在线特征提升的跟踪方法,该方法能够在线根据每帧目标跟踪结果更新分类器,对目标的表观和背景的数据变化具有一定的鲁棒性。Babenko 等人^[7]提出了基于多示例学习的跟踪方法,每次更新模型可以使用多个正样本,这在某种程度上解决了更新模型时只能使用一个正样本的问题。由于判别式模型融合了背景信息,因而在跟踪时能够比生成式模型更好的区分相似物体的干扰,近年来成为目标跟踪研究的热点。另外,机器学习和目标检测识别等相关领域的技术发展也推动了基于判别式模型的跟踪算法的研究。因此,现在大多数学者主要研究基于判别式的表观模型。

模型更新主要是利用最新的表观数据自适应调整模型参数,并要保证模型的性能不下降。而保证模型性能不下降的关键是训练样本必须正确。错误的样本可能会使表观模型性能下降,目标跟踪结果出现漂移,并影响后续目标跟踪结果,最终导致算法完全失去目标。现有的大部分算法在更新表观模型时,一般使用目标跟踪结果作为训练样本,其中隐含地假设了目标跟踪结果是正确的或误差较小,不影响样本的质量,但实际上我们很难保证跟踪结果不出现错误。另外在长期跟踪中,微小的跟踪误差可

能会逐渐积累成大的误差,如果不及及时纠正可能会使算法最终完全丢失目标。

模型更新的问题在近年来逐渐引起了大家的重视。Grabner 等人^[25]为了解决跟踪结果漂移问题,提出基于半监督提升方法更新表观模型,将先验知识与在线训练的分类器融合,并取得了一定效果。Kalal 等人^[26]提出 TLD 目标跟踪算法,通过在线学习评估检测结果,以便减少错误样本训练模型带来的性能恶化。Hare 等人^[22]提出基于结构化输出学习(Structured Output Learning)框架的跟踪算法,这种方法有效解决了正负样本选取的问题。Jin Gao 等人^[27]基于高斯过程回归方法融合历史跟踪结果和当前帧信息,在处理漂移问题时也取得了良好的效果。在文献^[20, 28]中,作者基于大量非标定数据离线学习视觉先验知识,并将其与在线学习的分类器融合,能在一定程度改善漂移问题。Hall 等人^[29]利用离线训练好的类别检测子,辅助在线表观模型的构造与更新,取得了较好的跟踪效果。虽然这些方法在一定程度上缓解了模型更新不恰当而导致的漂移问题,但还没有完全解决该难题。

除了漂移问题,模型更新还存在另外两个难题。一是在长期跟踪中,目标有可能暂时从镜头中消失或被其他物体完全遮挡,这时跟踪算法估计出的目标位置是错误的,表观模型更新学习的是背景。二是跟踪算法有可能失败,即输入图像中依然存在目标,但跟踪算法已经完全失去目标,这时表观模型学习的也是背景,Kalal 等人^[26]提出的 TLD 算法虽然尝试解决该问题,但是在目标跟踪中很容易将目标表观快速变化的情况误判为目标消失。

5 结语

虽然目标跟踪研究有了长足的进步,但距离复杂场景下任意目标的鲁棒跟踪仍然有较远的距离,原因在于其核心问题即目标表观建模中存在的问题还没有得到根本解决。针对目标表观建模的研究在以下两方面依然存在不足:首先是如何对未知目标进行有效的视觉描述。现有的目标表观模型在设计时为了保证实时性,往往采用底层简单特征,由于缺乏高层语义信息和先验知识,因而很难做到对任意物体和场景均有效;其次是如何构造和更新表观模型,以及在跟踪失败或目标消失时如何判断并及时检测。为了维持表观模型性能在长期跟踪中不下降,提取充足的正确样本是关键,但实际目标跟踪过

程中很难做到这一点。目标模型更新一般使用当前帧的跟踪结果,样本数量和多样性均有限,而且还隐含地假设了跟踪结果是正确的或误差较小,但实际并不能保证跟踪结果在每一帧都是正确的,错误的训练样本可能会恶化目标模型的性能,导致跟踪产生漂移并最终失败。

为了实现鲁棒的目标跟踪,表观模型至少应满足对光照、姿态变化和遮挡等鲁棒,以及在提取特征时具有足够的区分性。现有的表观模型还很难完全满足这些条件。另外,目标跟踪问题还没有得到根本解决的关键还在于人们对物体如何进行视觉描述还不完善,以及无法训练计算机使之像人脑一样进行学习。因此,未来目标跟踪中表观模型的研究从以下两方面展开将有望获得较大进步。

(1) 有效视觉描述的认知机理和理论建模。首先是目标的拓扑结构模型研究。基于局部的视觉模型能够更好的描述目标局部的视觉统计特性,在处理遮挡和形变等问题时具有潜在的优势,但需要挖掘目标局部单元在图像时空结构中的规律和局部单元之间的关系,解释其科学内涵和基本性质。其次是目标的视觉特征描述研究。简单底层特征提取速度快,但是容易受各种复杂场景的干扰,而高层语义信息则具有更好的抗干扰性,因此可以研究融合底层特征和高层语义属性信息,探索鲁棒性更强的视觉特征描述方式,并与局部拓扑结构模型紧密的结合,使其在复杂场景下具有更强的适应性。

(2) 探索更快、更鲁棒的模型学习理论与算法。首先,研究基于判别式的模型学习方法,充分挖掘各种先验知识,以及如何与在线训练的分类器有效融合,建立相应的形式化数学表达和规范的计算方法。其次,对于模型更新的难题,需要研究如何在有限的输入图像中获取足够的训练样本,以及如何判断样本的正确性或评估其质量,防止错误样本更新导致的模型性能下降;对于跟踪中目标暂时消失或被完全遮挡这种极端情况,需要研究鲁棒的遮挡检测方法,防止模型错误更新导致的跟踪失败,以及在跟踪失败后如何重新初始化表观模型,并继续准确跟踪目标。

参 考 文 献

- [1] Xi Li, Weiming Hu, Chunhua Shen, et al. A survey of appearance models in visual object tracking [J]. ACM Transactions on Intelligent Systems and Technology,

- 2013 4(4): 1-48.
- [2] Yi Wu, Jongwoo Lim, Yang Ming-Hsuan. Online Object Tracking: A Benchmark [C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2013: 2411-2418.
 - [3] 张焕龙, 胡士强, 杨国胜. 基于外观模型学习的视频目标跟踪方法综述 [J]. 计算机研究与发展, 2015, 52(1): 177-190.
 - [4] Ross D A, Lim J, Lin R S, et al. Incremental learning for robust visual tracking [J]. International Journal of Computer Vision, 2008, 77(1): 125-141.
 - [5] Mei X, Ling H. Robust Visual Tracking using L1 Minimization [C] // Computer Vision, 2009 IEEE 12th International Conference on. IEEE, 2009: 1436-1443.
 - [6] Grabner H, Grabner M, Bischof H. Real-time tracking via on-line boosting [C] // British Machine Vision Conference. Edinburgh: BMVA Press, 2006: 47-56.
 - [7] M H Y Boris Babenko, S. Belongie. Robust object tracking with online multiple instance learning [J]. IEEE Trans. Pattern Analysis and Machine Intelligence, 2011, 33(8): 1619-1632.
 - [8] Rui Yao, Qinfeng Shi, Chunhua Shen, et al. Part-Based Visual Tracking with Online Latent Structural Learning [C] // Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition. 2013: 2363-2370.
 - [9] Tianzhu Zhang, Kui Jia, Changsheng Xu, et al. Partial Occlusion Handling for Visual Tracking via Robust Part Matching [C] // IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE Press, 2014: 1258-1265.
 - [10] Guang Shu, Dehghan A, Oreifej O, et al. Part-based multiple-person tracking with partial occlusion handling [C] // IEEE Conference on Computer Vision and Pattern Recognition. Providence: IEEE Press, 2012: 1815-1821.
 - [11] Bo Yang, Ram Nevatia. Online Learned Discriminative Part-Based Appearance Models for Multi-human Tracking [C] // European Conference on Computer Vision. Florence: Springer Berlin Heidelberg, 2012: 484-498.
 - [12] Shu Wang, Huchuan Lu, Fan Yang, et al. Superpixel Tracking [C] // Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011: 1323-1330.
 - [13] Comaniciu D, Ramesh V, Meer P. Kernel-based object tracking [J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2003, 25(5): 564-577.
 - [14] Werlberger M, Trobin W, Pock T, et al. Bischof. Anisotropic Huber-L1 Optical Flow [C] // British Machine Vision Conference. London: BMVA Press, 2009: 1-11.
 - [15] Hu Weiming, Xi Li, Luo Wenhan, et al. Single and multiple object tracking using log-euclidean riemannian subspace and block-division appearance model [J]. IEEE Trans. Pattern Analysis and Machine Intelligence, 2012, 34(12): 2420-2440.
 - [16] Xin Sun, Hongxun Yao, Shengping Zhang. A novel supervised level set method for non-rigid object tracking [C] // Computer Vision and Pattern Recognition CUPR, 2011 IEEE Conference on. IEEE, 2011: 3393-3400.
 - [17] Zhou Huiyu, Yuan Yuan, Shi Chunmei. Object tracking using SIFT features and mean shift [J]. Computer Vision and Image Understanding, 2009, 113(3): 345-352.
 - [18] Xi Li, Shen Chunhua, Anthony Dick, et al. Learning compact binary codes for visual tracking [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013: 2419-2426.
 - [19] Martin Danelljan, Fahad Shahbaz Khan, Michael Felsberg, Joost van de Weijer. Adaptive Color Attributes for Real-Time Visual Tracking [C] // IEEE conference on computer vision and pattern recognition. Columbus: IEEE Press, 2014: 1090-1097.
 - [20] Naiyan Wang, Dit-Yan Yeung. Learning a Deep Compact Image Representation for Visual Tracking [C] // Advances in Neural Information Processing Systems. Nevada: Curran Associates, 2013: 809-817.
 - [21] S. Avidan. Support Vector Tracking [J]. IEEE Trans. Pattern Analysis and Machine Intelligence, 2004, 26(8): 1064-1072.
 - [22] Hare S, Saffari A, Torr P H S. Struck: Structured Output Tracking with Kernels [C] // Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011: 263-270.
 - [23] Bai Y, Tang M. Robust Tracking via Weakly Supervised Ranking SVM [C] // Computer Vision and Pattern Recognition (CVPR), 2012 IEEE conference on. IEEE, 2012: 1854-1861.
 - [24] Ding Jianwei, Huang Yongzhen, Liu Wei, et al. Severely blurred object tracking by learning deep image representations [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2015.
 - [25] Helmut Grabner, Christian Leistner, Horst Bischof. Semi-supervised On-Line Boosting for Robust Tracking [C] // European Conference on Computer Vision. Marseille: Springer Berlin Heidelberg, 2008: 234-247.

- [26] Zdenek Kalal , Krystian Mikolajczyk , Jiri Matas. Track-
ing-Learning-Detection [J]. IEEE Trans. Pattern Analy-
sis and Machine Intelligence , 2012 , 34 (7) : 1409 -
1422.
- [27] Jin Gao , Haibin Ling , Weiming Hu , et al. Transfer
learning based visual tracking with gaussian processes
regression [M] // Computer Vision-ECCV 2014. Spring-
er International Publishing , 2014: 188 - 203.
- [28] Qing Wang , Feng Chen , Jimei Yang , et al. Transfer-
ring visual prior for online object tracking [J]. IEEE
Transactions on Image Processing , 2012 , 21 (7) : 3296
- 3305.
- [29] David Hall , Pietro Perona. Online , Real-Time Tracking
Using a Category-to-Individual Detector [M] // Computer
Vision-ECCV 2014. Springer International Publishing ,
2014: 361 - 376.
- [30] Felzenszwalb P F , Girshick R B , McAllester D , et al.
Object detection with discriminatively trained part-based
models [J]. IEEE Trans. Pattern Anal. Mach. Intell. ,
2010 , 32 (9) : 1627 - 1645.
- [31] Steffen Gauglitz , Tobias Höllerer , Matthew Turk. Eval-
uation of interest point detectors and feature descriptors
for visual tracking [J]. International Journal of Computer
Vision , 2011 , 94 (3) : 335 - 360.
- [32] Avidan S. Ensemble tracking [J]. IEEE Transactions on
Pattern Analysis and Machine Intelligence , 2007 , 29
(2) : 261 - 271.

(责任编辑 于瑞华)