



关于单目标跟踪方法的研究综述

摘要

目标跟踪一直都是机器视觉领域的研究热点,应用场景主要分为单目标跟踪和多目标跟踪.本文主要介绍了单目标跟踪问题,回顾了近年来用于视频单目标跟踪的算法,对单目标跟踪方法进行了分类,并且对每一类中具有代表性的方法进行了介绍,分析了各自的优缺点.最后讨论了单目标跟踪任务中的难点问题和发展趋势,为该方向的研究人员快速了解单目标跟踪技术提供了参考.

关键词

计算机视觉;单目标跟踪;领域分析;算法综述

中图分类号 TP391.41

文献标志码 A

收稿日期 2019-10-09

作者简介

傅杰,男,博士,主要研究方向为计算机视觉、目标跟踪.fujie_snb@163.com

徐常胜(通信作者),男,博士,教授,博士生导师,主要研究方向为多媒体分析与检索、计算机视觉、模式识别.csxu@nlpr.ia.ac.cn

¹ 郑州大学 信息工程学院,郑州,450001

² 中国科学院自动化研究所 模式识别国家重点实验室,北京,100190

0 引言

计算机视觉领域包含很多种针对不同应用场景的技术研究,目标跟踪算法就是其中一个具有实际意义的研究方向,也是该领域用于视频内容分析的基本技术手段之一.该研究需求是在对某个运动物体的捕捉和分析中产生的.目标跟踪算法是在一段连续的视频图像序列中标注出某个物体所在的物理位置,最终将连续帧中目标物体连接起来形成目标运动轨迹路径的技术.其本质上是根据给出的目标图像,利用特征提取和特征关联技术将不同帧图像中最可能属于同一目标的特征进行匹配,然后将连续视频帧中匹配上的目标进行连接,从而得到目标的运动轨迹,最终实现目标跟踪这一任务.一般的目标跟踪算法主要包含4个基本部分:表观模型、运动模型、观测模型和模型更新策略.目前,目标跟踪技术在很多领域都有着不可替代的应用,其中包括智能监控、视频检索、人机交互和现代化军事等.虽然目标跟踪技术在现实场景中具有很重要的实用价值和广阔的发展前景,但是由于在真实场景中所跟踪目标的多变性和场景的复杂性,目标跟踪领域一直存在着极具挑战性的问题.由于近年来技术的不断发展,目标跟踪方法也在不断创新,一个自适应的鲁棒的目标跟踪算法必须要能很好地应对目标跟踪场景中因为目标尺度变化、快速移动、遮挡、光照、形变、旋转等问题带来的挑战,这些问题一直都是该领域的难点,也是急需突破的问题.

本文旨在介绍近些年单目标跟踪研究领域的经典算法和成果,并且对不同的方法的优缺点进行分析,最后对相关研究内容进行展望.本文的主要安排内容如下:第1章主要介绍目标跟踪模型组成部分的基本知识和难点问题;第2章主要介绍关于单目标跟踪算法的发展阶段以及每一阶段中经典的目标跟踪模型方法,并分析不同方法的优缺点;第3章则介绍常用于单目标跟踪领域的数据集;第4章对该领域可能的研究方向进行总结和展望.

1 目标跟踪的组成和研究难点

一般来说,单目标跟踪任务算法流程基本可以归结如下:首先根据给定的测试图像,输入需要跟踪目标的初始化目标框,根据某个运动模型算法推算出下一帧图像中目标物体可能出现的大致位置,在该位置附近产生一定数量的目标候选框,再利用表观特征模型提取

这些候选框中图像的特征信息,将这些信息与需要跟踪的目标图像的特征进行比对,利用合适的观测模型对这些候选框进行打分,最后选择分数最高的候选框作为目标在下一帧中出现的位置框,或者综合多个高分候选框按照一定的策略进行融合选择更优的预测目标作为最终结果,再根据新得到的预测结果结合之前的观测值对模型进行更新.重复上述步骤,直至完成目标跟踪任务.按照上面所述的流程,我们可以归结出目标跟踪框架中核心的5个研究内容:表观模型、运动模型、观测模型、模型更新策略以及当前处理目标跟踪领域研究的难点问题的一些策略,下面就分别介绍这几方面的内容.

1.1 表观模型

表观模型是用来提取图像特征的算法,目标跟踪本质上是将不同图像中相同目标进行匹配的问题,而匹配这些目标就需要使用一定的视觉特征,通过计算图像视觉特征之间的相似度将不同图像中的同一目标标定出来,再将这些标定的位置坐标串联起来,从而实现跟踪的目的.所以可以看出,设计一个辨别性很强的物体表观特征是目标跟踪模型效果好的关键因素之一.一般来说,物体的表观特征主要包含2种:一种是领域专家手工设计的特征,另一种则是利用基于数据驱动的深度神经网络从大量有监督信息的训练样本中学习到的深度特征.其中手工设计的特征包括灰度图特征、方向梯度直方图特征、哈尔特征、尺度不变性特征等.为了获得更加鲁棒的目标表观特征,研究者们尝试着将图像的边缘^[1]和纹理^[2]等手工信息与图像原始像素信息结合,使得目标的表观特征更加具有辨别性和鲁棒性.通常直方图特征表示方法可以对图像的视觉特征进行有效地提取,受此启发,Bradski^[3]在图像的HSV颜色空间中引入了颜色直方图特征用于目标跟踪并取得了一定效果.但是该表观特征提取方法导致了图像空间信息的损失,为了解决这一问题,Comaniciu等^[4]优化了该特征表示方法,使用核加权的RGB颜色直方图作为图像的表观特征用于目标跟踪算法,效果获得进一步提升.后来研究者们又发现多特征融合的方法相对于单一特征更具有鲁棒性,所以Ning等^[5]将颜色和纹理特征融合起来提出了一种联合颜色纹理直方图的表观特征表示方法用于目标跟踪.除了基于直方图的表示方法,Porikli团队^[6-7]又提出了将协方差矩阵用于目标跟踪模型中的物体表观特征表示,但由于该方法是基于像素统计的,所以对噪

声并不鲁棒,而且忽略了很多有效空间信息,效果并不好.除此之外,Zhou等^[8]也尝试用SIFT特征点匹配的方法来对不同帧中的同一对象进行匹配跟踪,虽然取得了一定效果,但是当面临背景干扰问题时效果则很不理想,甚至出现一对多匹配现象.综上所述:手工设计的特征都具有一定的主观因素,所以学习到的特征不够全面,辨别性也较弱,鲁棒性不够.而深度特征表示方法则较好地解决了这些问题,理论上只要训练样本足够,那么基于深度学习的方法学习到的特征也就更加鲁棒,相比于人工设计的特征更加有鉴别能力.随着深度特征在目标跟踪任务中的应用,大量研究人员开始将深度神经网络用于目标跟踪领域中的表观特征提取模块^[9-12].但是研究人员在实践过程中发现,利用深度神经网络对目标进行特征提取时需要花费大量时间,所以很多研究者针对深度神经网络计算量大耗时长的问题进行了研究,并在此基础上提出了基于孪生网络的一系列算法^[13-16],不仅在时间性能上大有提升,而且在效果上也有大幅度提升.实验证明:大多基于深度表观特征的目标跟踪算法通常能获得更好的效果,而且模型的鲁棒性也更强.

1.2 运动模型

运动模型主要用于描述待跟踪目标在连续图像序列中的运动趋势和运动状态信息.选择合适的运动模型对当前图像和之前图像序列中待跟踪目标的运动状态进行拟合,从而可以大概预测出下一图像中目标对象可能出现的位置区域,再根据此结果按照某策略选择一组可能的目标候选框,将候选框中的图像特征与目标特征进行匹配完成跟踪任务.所以,运动模型对于目标跟踪框架至关重要,一个与实际情况拟合很好的运动模型不仅可以大大提升目标跟踪精度,而且可以降低目标候选框的数量,并且大大减小模型匹配时的计算量,最终提升目标跟踪的实时性.常用的目标运动模型主要有粒子滤波^[17]和卡尔曼滤波^[18]等算法.运动模型一般可以分为2类:线性运动模型和非线性运动模型.目前目标跟踪框架中使用最多的是线性运动模型,因为一般情况下目标在相邻帧之间的移动距离很小,可以近似看成是线性的,所以大多情况下线性运动模型即可描述目标短期的运动趋势.线性运动模型中一般包含对3种属性的建模:第1种就是对物体速度的线性建模,代表性方法为文献[19];第2种是通过对目标位置进行线性建模的方法^[20];第3种则是对物体的加速

度进行线性建模,代表性模型为文献[21].当然,还存在着一些复杂的运动场景是无法使用线性模型进行建模的,所以研究人员针对复杂的运动情况也设计了非线性运动模型,其中代表性的方法是文献[22].

1.3 观测模型

观测模型是用来对特征提取后的候选框内图像进行匹配,再通过一定策略得出最终的目标框来作为跟踪算法的最后结果.大多数目标跟踪算法的研究人员也将重点放在了这一部分策略的设计上.

根据不同类型的方法,主要可以将观测模型划分为2大类:判别式跟踪算法和生成式跟踪算法.判别式跟踪算法的大概思路是通过训练一个分类器,将图像中目标和背景区别开来,得到每一张图像中目标的位置从而实现跟踪.判别式方法其实也被称为基于检测的目标跟踪方法^[23],最早由 Collins 等提出.该类算法的关键是要自适应地学习出具有很强前景和背景区分性的颜色等视觉特征,常用的分类方法有支持向量机(SVM)、深度神经网络等.在实际场景中,相比较于生成式方法,基于深度学习的判别式目标跟踪方法因为学习到的特征有很强的辨别性且具有很鲁棒的效果,所以判别式方法已经逐渐成为视觉跟踪领域中的主流方法.特别是近年来,随着特征表示、分类器设计研究的深入,各类机器学习和深度学习算法被应用到特征提取和分类器训练中去,从而很多效果更为鲁棒的目标跟踪模型被提出.最有代表性的就是一系列基于相关滤波类跟踪算法的提出,使得目标跟踪器在速度、准确度和精度上均获得了较大的提升.相关滤波类的跟踪模型根据在第1帧中给定的模板目标图像提取特征作为滤波模板,然后再利用后面帧的图像与模板做相关性卷积,计算不同部分图像的响应值,然后将最大响应值对应的部分作为跟踪的结果.根据这一思想,研究者们提出了很多基于相关滤波的跟踪模型,这些模型包括了 MOSSE^[24]、CSK^[25]、KCF^[26]、CN^[27]、DSST^[28]、SAMF^[29]、SRDCF^[30]和 Staple^[31]等.当然,除了相关滤波类的目标跟踪模型,随着深度学习技术的发展,一系列基于孪生网络的目标跟踪算法在效果上也获得了较大的提升,这类模型包括 SiamFC^[13]、SiamRPN^[32]、SiamRPN++^[33]、SiamMask^[16]、SiamDW^[15]等.

生成式方法通常是利用模板匹配的思想,在图像的众多候选框中选择与模板图像特征最匹配的作为目标图像,从而确定目标的位置实现跟踪任务.生

成式目标跟踪模型的关键是需要很好的目标特征表示方法和目标模型.一般常使用的目标特征表示方法有特征子空间方法和稀疏表示方法,这类特征表示方法在处理目标遮挡问题时通常具有较好的结果,但是由于这两种方式的计算量相对较大,通常不能满足目标跟踪场景下的实时性要求.生成式跟踪主要的代表性模型有如下几种:首先是针对目标外观变化设计的自适应更新表观模型的增量视觉跟踪模型(IVT)^[34];其次,还有基于目标分解的目标跟踪算法(VTD)^[35],该类算法有较强应对目标外观变化的能力,但是对环境变化的鲁棒性较差;还有基于采样思想的目标跟踪方法(VTS)^[36],这类算法可以在一定程度上解决环境噪声问题和运动模糊,但是自适应性有所欠缺.为了应对部分遮挡问题,还有一种局部无序跟踪(LOT)^[37]类算法被提出.

1.4 模型更新策略

模型更新主要是针对外观模型和运动模型的更新,因为在跟踪的过程中目标处于不断移动和变化的状态,在这一过程中目标的形状、大小、速度、姿态等特征都会发生变化,所以我们要随着跟踪过程的进行采用自适应的模型更新方式以适应跟踪目标的表观特征变化,防止跟踪过程中出现跟踪框漂移的情况.但是到目前为止也没有一个统一自适应的模型更新标准,因为通常在模型更新策略选择时会面临模型更新策略和更新时间间隔选择的问题:如果更新过于频繁,那么可能会导致模型计算量过大和实时性不好的问题,而且很可能会导致对原来表观特征信息的丢失;如果更新过慢,也可能出现特征变化过快导致的跟踪框漂移现象.所以很多模型都采用长短期相结合的方式来进行模型更新,相比较单一更新策略,该策略在效果上有所提升.除了更新时间的选择策略,还有其他对目标匹配模板^[38]和量子空间学习方法^[34]更新.然而如何学习一个较好的在线更新机制使得模型能够应对不断变化的情况,而且能够在更新过程中不会出现效果退化问题,一直都是目标跟踪领域中的难题,也是模型能够长时间跟踪目标的关键技术.

1.5 目标跟踪领域难点问题

目标跟踪领域的难点问题主要来自于2个方面:首先是目标在模型跟踪过程中一直处于变化的状态,主要包括姿态、形状和尺度的变化带来的特征变化问题;其次,外部环境的变化也会带来一系列挑战,主要包括光照变化、运动模糊、遮挡、背景干扰等

问题.这两方面的问题共同为在线目标跟踪任务带来了许多困难.其中姿态、形状和尺度的变化会导致同一目标在不同帧中的表观特征出现较大的变化,从而为目标特征匹配带来麻烦;而外部环境的变化会导致表观模型学习到很多干扰信息,所以在比较复杂的外部环境条件下跟踪算法很难学习到目标和背景之间明确的辨别性信息,从而导致跟踪结果容易发生偏移.研究者们根据这些难点问题也分别做了分析和尝试,并且针对不同问题提出了不同的解决思路.

首先,遮挡问题一直都是目标跟踪场景中常见并且难以解决的问题.遮挡一般也分为两种情况,一种是完全遮挡,还有一种是部分遮挡.目前主要用于解决部分遮挡问题的方法有2种思路:1)通过合适的检测方法对跟踪的目标进行检测,判断该目标是否部分被遮挡,然后再根据判别结果来决定是否对模型采用的模板进行更新,使用自适应的方式保证模板的鲁棒性和实时性;2)采用局部跟踪的策略,这种思路简单有效,主要是利用跟踪目标的未遮挡部分进行跟踪,从而应对部分遮挡的情况.对于完全遮挡问题,目前主要的方法和思路就是利用再次检测的方式,当检测到目标再次出现时重新恢复跟踪,但是该思路只能应对短暂的完全遮挡问题,而且由于不知道完全遮挡后的目标何时或者会不会再出现,检测模型一直要处于运行状态,这样就产生了计算量大的问题,目前对于完全遮挡的问题还没有较好的解决思路.

跟踪过程中目标形变通常会导致目标表观特征发生变化,在与模板匹配的过程中会出现匹配失败的问题,从而导致跟踪漂移.目前用来解决这类问题的思路主要是动态自适应地对表观特征模型进行更新,使得表观模型提取到的目标特征随着形变而实时更新,始终保持模板特征的鲁棒性,但是这种解决方法需要设计很好的模型更新策略和选择合适的时间间隔,目前这也是大多数研究者的研究课题.

背景干扰问题通常发生在比较复杂的场景下,例如背景图像中出现了与跟踪目标特别相似的物体,这样就会对跟踪效果产生干扰.通常解决这类问题的思路是利用跟踪目标的运动特征,对跟踪目标的运动轨迹和可能出现的位置进行预测,从而排除干扰信息;或者利用大量具有干扰信息的负样本对模型进行训练,进一步提升模型对于干扰信息的抗干扰能力.

尺度变化通常发生在目标与镜头之间的距离发生改变的情况下,通常此类问题的解决思路是结合运动模型,在目标下一帧可能出现的位置利用尺度金字塔方法设置不同尺度的候选框进行模板匹配;或者是直接在多个尺度上进行跟踪,然后再利用恰当的融合策略选择合适的尺度.

从上述问题的解决思路中可以发现,目标跟踪模型框架中较好的表观、运动等特征可以帮助提高模型跟踪准确度和精度,对于跟踪过程中出现的复杂变化问题一般都是通过增强模型自适应性的设计来完成的.当然,除了上述的常见问题,还有一些其他因素导致的问题,例如光照条件变化、低分辨率、目标运动模糊、快速运动、超出镜头视野等,这些都是目标跟踪领域面临的难题,需要科研人员去突破和解决.

2 目标跟踪算法介绍

本章将对近年来单目标跟踪算法的发展趋势进行介绍,主要是按照不同发展阶段的算法原理进行介绍.

近年来,单目标跟踪算法发展主要可以划分为4个阶段:第1阶段主要是基于粒子滤波的相关算法;第2阶段模型大多是基于稀疏表示理论的跟踪方法;第3阶段则是相关滤波类跟踪算法;随着深度学习方法在特征建模方面的突出表现,第4阶段主要是基于深度学习的跟踪方法.下面就分别介绍不同阶段的代表性跟踪方法,并且对这些跟踪模型的优缺点进行分析和比较.

2.1 基于粒子滤波的算法

粒子滤波算法^[39]最初是由 Ulam 提出的一种序列化的蒙特卡罗方法,该方法是针对非线性滤波场景问题而提出的,其主要思想是通过在一个序列中使用加权采样的方法来尽可能拟合实际样本的分布,从而通过之前的序列状态分布来计算当前时刻目标的状态分布概率,并将最大概率值对应的状态作为目标状态预测的结果.近年来,随着目标跟踪领域技术的快速发展,很多运动模型被引入到目标跟踪框架中来,并且取得了较为不错的效果, Kalman Filter 算法^[40]就是其中之一,该算法常用于线性高斯分布的场景.但是,在实际场景中,很多目标的运动并不是线性的,而且目标的状态分布情况也并不满足高斯分布,所以 Kalman Filter 算法在类似场景中取得的效果并不令人满意.为了解决这一问题,目

标跟踪领域的研究者们发现粒子滤波算法在处理非线性和非高斯分布的系统状态问题时较好的性能,因此将粒子滤波算法引入到目标跟踪,诞生了一系列基于粒子滤波算法的目标跟踪模型。

粒子滤波算法第一次被用于目标跟踪领域是在 Isard 等^[41]提出的 CONDENSATION 算法中,在该算法中,作者利用先验概率密度公式计算粒子分布的权重并以此对粒子进行采样,将采样的粒子作为样本集合计算跟踪目标的后验概率分布密度,实验证明该算法在处理简单场景下的目标跟踪的确获得了一定效果,但是由于该模型思想过于简单,且没有考虑到当前目标的状态观测值等信息,所以学习到的模型在其他复杂条件下效果并不鲁棒。因此,目标跟踪领域的研究者们针对粒子滤波在跟踪算法中表现出的各种不足提出了一系列的改进模型。下面就从粒子滤波的原理和针对算法局限性提出的解决方法这两个方面进行介绍。

2.1.1 粒子滤波算法原理及流程

1) 粒子滤波算法原理

粒子滤波算法是基于概率统计中贝叶斯估计的,基于粒子滤波算法的目标跟踪模型一般可以描述为利用系统观测值满足的状态概率分布情况去递归更新并迭代估计出跟踪目标当前系统状态的一种算法。如果令 x_k 表示系统在 k 时刻产生的系统状态, z_j 表示在 j 时刻的状态观测值,那么粒子滤波算法的核心思想就是利用带权值的不同时刻的系统状态采样因子的分布情况估计出当前系统状态的概率值大小,用数学公式可以表示成 $P(x_k | z_{1:k})$ 。在每个时刻需要通过采样权值计算函数采样出 k 时刻的 N 个状态粒子 $\{x_k^i\}_{i=1}^N$,则在 k 时刻系统状态后验概率公式则可以通过下面加权概率公式来近似得到:

$$p(x_k | z_{1:k}) \approx \sum_{i=1}^N w_k^i \delta(x_k - x_k^i), \quad (1)$$

其中 w_k^i 表示在 k 时刻第 i 个采样粒子的权重大小,函数 $\delta(\cdot)$ 表示 Delta 函数。

2) 粒子滤波算法流程

粒子滤波算法的核心在于重要性采样计算函数和重采样策略的选择。若定义粒子重要性采样密度函数为 $q(x_k | z_{1:k})$,将系统状态先验概率函数 $p(x_k | x_{k-1}^i)$ 作为重要性密度函数,则粒子滤波的算法流程大致如下:

①根据重要性采样密度函数采样状态粒子 x_k^i ,其中 $i = 1, 2, \dots, N$;

②根据当前的系统状态观测值对采样粒子的权值进行更新,更新公式如下:

$$w_k^i = w_{k-1}^i \frac{p(z_k | x_k^i) p(x_k^i | x_{k-1}^i)}{q(x_k^i | x_{k-1}^i, z_k)}, \quad (2)$$

公式更新后还需要进行权值归一化处理,统一到标准尺度下进行粒子重采样;

③根据上面计算得到的归一化权值进行粒子重采样;

④根据上一步重采样的状态粒子,利用状态对应的重要性权值进行加权求和计算,得到当前目标状态的估计值,同时计算当前状态的加权方差估计值,之后跳转到步骤②继续循环迭代更新和计算,直到满足退出迭代的条件。

2.1.2 粒子滤波跟踪模型

首先,为了解决复杂运动场景下某个目标的跟踪任务,领域内研究者们设计了很多基于多特征融合思想的目标跟踪算法来提升传统粒子滤波跟踪模型的效果。Brasnett 等^[42]提出了一种融合颜色、纹理和边缘的多特征目标跟踪方法,实验结果证明该模型相较于使用单一特征进行跟踪的方法效果有所提升且性能更加稳定。Wu 等^[43]提出了一种利用隐马尔可夫模型因子来模拟图像颜色和轮廓特征之间依赖关系的目标跟踪算法,该算法在大多数场景中具有较好的鲁棒性。为了提升特征表示的自适应性,文献[44]中提出了一种自适应的多特征融合跟踪方法,该方法与期望最大化算法思想类似,使用在线自适应的方法调节不同特征所占的比重来获得更好的效果。文献[45]中则提出了一种很巧妙的特征融合方法,该方法除了利用典型的特征,还将颜色对比相似度和方向一致性作为彩色图像中目标边缘检测的特征用于目标跟踪。文献[46]中提出了一种根据模糊逻辑思想自适应地组合目标图像颜色和形状信息的方法来描述观测模型,并在此基础上实现目标跟踪,一定程度上提升了模型的可靠性。文献[47]中提出了一种基于图像角点特征和颜色特征融合的粒子滤波目标跟踪模型,在该模型中粒子的采样效率得到了一定提升,算法的鲁棒性也有所增强,但是还是有所欠缺。为了进一步提升模型的鲁棒性,Gan 等^[48]探索了新的多特征融合方法,利用角点特征和局部二值特征相结合的方式目标跟踪,该模型在应对复杂背景下的跟踪场景中获得了较为不错的稳定性。

除了利用多特征融合的方法来优化粒子滤波的

目标跟踪算法,还有一些研究者通过对多种算法的融合来提升模型效果.这些融合的算法包括 Mean-shift 算法、稀疏理论表示和深度学习算法等.文献[49-50]在粒子滤波算法的基础上融合了 Mean-shift 算法原理实现了融合算法的目标跟踪模型,这类算法可以提升粒子滤波算法的计算效率和实时性,但是怎样选择合适的 Mean-shift 算法融合策略也是其中一个难题;除此之外,该类算法中还需要考虑如何防止跟踪漂移的问题.文献[51-52]将稀疏理论表示的相关知识应用到粒子滤波算法中用来解决目标跟踪任务,实验证明此改进方法可以在一定程度上提升目标跟踪模型的精度,但是在求解问题的时候通常会因为计算量太大而影响算法的实时性.随着深度学习技术的快速发展以及深度模型在图像特征表示方面突出的表现,基于深度学习方法的融合目标跟踪算法也接连被提出,文献[53-55]等模型就将深度模型算法和粒子滤波原理相结合用于目标跟踪任务,取得了相当不错的效果.但是深度学习模型通常需要通过大量有监督信息的训练样本来训练模型,所以怎样设计一个合适的结构来提升计算速度从而满足算法实时性的要求也是一个急需要解决的问题.

目标跟踪任务中最难处理的问题通常是因为目标本身和跟踪环境的不断变化导致模型或者模板特征不能适应变化场景的问题,所以针对这个问题研究人员也设计了很多自适应性调整的算法,其中包括粒子采样的自适应方法、特征权值的自适应调整方法等.文献[56]中的方法在目标跟踪进行的过程中不断根据当前帧中目标预测的情况动态调整下一帧图像中需要采样的粒子数量,该方法虽然可以适当减少采样粒子数量,但是每一帧都要进行准确度的计算,反而导致模型算法计算量的增加,得不偿失.文献[57]设计了一种特征权重的动态调整方法,使用单一特征和融合特征对应的目标状态之间的相似度动态调整特征权重,但是这样的计算方式增加了计算量,算法实时性有所下降.

2.1.3 粒子滤波算法存在的问题

从 2.1.1 节对于粒子滤波算法原理的介绍,针对算法不同的环节我们总结了如下 3 个重要的问题:如何设计和选择重要性采样密度函数、如何解决随着迭代过程进行的粒子退化和大量粒子采样带来的计算量爆炸问题.下面就这 3 个方面进行介绍:

1) 重要性采样函数选择

采样函数选择的原则是既要保证采样过程简单有效易操作,还要保证函数系数总和最小,也就是状态最稳定.目前为止,最优的重要性采样密度函数很难在实际问题中被找到并实现,所以一般采用的方式都是选择或者设计较为合适的函数作为重要性采样的依据.

2) 粒子退化

粒子退化问题通常伴随着迭代过程的进行而出现,产生的原因是因为粒子权值的方差在采样迭代过程进行时会随着迭代次数的增加而出现递增的情形,很多采样的状态粒子通常会出现在权重无限接近于 0 的情况,极少数状态粒子的采样权重会接近 1,也就是我们所说的粒子退化的问题.这个问题会导致模型算法将大量的时间花费在这些权重接近于 0 的状态粒子上,而这些粒子对最终状态后验概率的估计几乎没有作用.通常用来解决这类问题的思路是重采样技术^[58],重采样通过样本均衡的思想,减少权重小的粒子所占的比重,适当增加权重大的粒子比重,从而缓解粒子退化带来的影响.

3) 计算量

一般来说,采样的粒子越多则算法学习到的状态后验概率分布也会越接近系统真实情况,但是采样的粒子越多那么计算量也会增大很多,造成算法运行缓慢无法实时的结果.通常应对此种现象的方法是根据一些先验知识采样比较重要的粒子,比如在基于 Kalman Filter 算法的跟踪模型中,我们通常使用 Kalman Filter 运动模型学习到的目标运动信息来辅助粒子进行采样,这样就可以大大减少采样粒子的数量,并且保证了算法效果.

2.2 基于稀疏表示的算法

稀疏表示是一种对信号进行处理和表示的技术,一般来说,大多的信号的组成部分通常都十分庞大,但是信号通常都具有冗余性,其中不是所有的组成部分都是有用信号.所以为了深入研究这个问题,近年来很多专家和学者开始了一种信号表示技术的研究,针对信号的稀疏性,Coifman 等^[59]提出了稀疏分解的概念,主张将信号分解为一些基本的有效部分,在这些部分的基础上可以通过某个字典空间或者是观测矩阵重构出原信号或者是原信号的大量信息.为了能够很好地分解表示这些信号,通常需要构建一个过完备的字典空间,这个字典空间是由一个个相互正交的基信号组成,基与基之间可以存在冗

余信息.这种在一个过完备的字典空间中利用尽可能少量的基来重构原始信号的方法被称为稀疏表示理论.随着稀疏表示理论的发展和目标跟踪领域研究的深入,基于稀疏表示理论的目标跟踪方法在2011年也被 Mei 等^[60]在其研究成果中提出,此后一系列基于此理论的改进目标跟踪算法也相继被提出.

2.2.1 稀疏表示理论的原理

将上述稀疏表示的思想用数学形式可以定义为如下形式:给定一组用于表示测试图像的训练模板图像集合 $T = \{t_1, t_2, \dots, t_n\}$, 其中每一个元素表示一个模板图像的表示,也就是上述所说的基信号,其中每个基信号都用 $m \times 1$ 维的向量来表示,若训练模板图像集合构建足够完备的话,那么每一个跟踪过程中出现的测试图像都可以通过这些训练图像模板集合中的基线性组合得到,用数学表示可以写成如下公式:

$$y = TX + \epsilon = t_1x_1 + t_2x_2 + \dots + t_nx_n, \quad (3)$$

其中 ϵ 表示的是误差向量, $X = [x_1, x_2, \dots, x_n]$ 表示的是在这组基上测试图像 y 的映射矩阵,根据稀疏表示理论可知其中大量元素值为0.一般来说,式(3)的解是开放性的,通常存在无穷多解,而稀疏表示理论求解的就是尽可能多的使 $X = [x_1, x_2, \dots, x_n]$ 表示向量中的元素为0,也就是使用最少的信号基去表示测试图像 y .

2.2.2 基于稀疏表示理论的跟踪模型

Yang 等^[61]提出了一种利用稀疏表示理论对跟踪目标进行表观特征建模的方法,这也是稀疏表示理论刚被用于目标跟踪领域的常见使用方式,在这类跟踪算法中选择合适且完备的模板图像、遮挡图像集合是稀疏表示理论应用的关键难题.随着研究的深入,稀疏表示理论在跟踪模型中的应用方法也越广泛,其中 Liu 等^[62]将稀疏表示理论用于对目标运动状态的建模中,实验结果证明模型也取得了相对不错的效果.除了对运动目标的特征建模,还有研究者将稀疏表示理论和其他一些算法框架结合提出了很多效果鲁棒的跟踪算法: Bao 等^[63]提出了一种快速的稀疏跟踪方法,该方法利用加速的近似梯度方法使跟踪速度大大提升;2013年, Zhang 等^[64]从多任务的角度设计了一种基于多任务稀疏表示的目标跟踪方法; Jia 等^[65]提出了一种基于结构化局部稀疏表观特征建模的视觉跟踪方法; Mei 等^[66]在多任务的基础上增加了多角度的考虑,提出了一种基

于多任务和多视角稀疏表示的目标跟踪方法.

2.2.3 基于稀疏表示理论跟踪算法的难点

随着稀疏表示理论和目标跟踪领域研究的结合和不断深入,基于稀疏表示理论的目标跟踪模型已经取得了很大的进展,但是在该类型的目标跟踪方法中依然存在以下几个方面值得突破和深入研究:

1) 因为目标跟踪场景中目标图像通常是不断变化的,而大多数稀疏表示方法构建的目标图像模板集合都是静态的,如何设计和构建一个动态自适应的模板图像集合是非常值得研究的重要问题之一;

2) 目前,基于稀疏表示理论的目标跟踪算法由于其计算复杂度较高,所以该类算法的实时性通常不是太好,所以针对稀疏表示的快速求解也是其中一个亟待解决的问题;

3) 稀疏表示理论中字典模板的构建通常都是人工设计的,自主学习方式的模板图像集合构建方法还比较少,基于自主学习方法的字典构建方法也值得深入探索和发现.

2.3 基于相关滤波的算法

随着相关滤波算法和目标跟踪领域交叉研究的深入,目前相关滤波目标跟踪模型在跟踪任务中已经表现出了相当优异的性能,该类算法不仅在变形、遮挡等复杂情境下表现出了相当不错的性能,而且在速度上远远快于其他类型的跟踪算法.下面就从该类算法的原理、算法发展和目前存在的问题几个方面分别介绍.

2.3.1 相关滤波类目标跟踪模型原理

相关滤波类目标跟踪算法的基本思想是将目标跟踪问题构建成为一个目标模板特征匹配的问题.使用该类算法进行目标跟踪时,通常在第1帧中以目标为中心初始化要跟踪目标的跟踪框,并以此为中心按照一定的规则采样图像块作为正负样本来训练相关滤波器,接下来每当当前帧的图像输入,模型都会根据上一帧图像中得到的目标位置按照一定的规则对图像采样小图像块,利用特征提取模型获取小图像块的特征,其中为了防止目标特征在傅里叶变换过程中带来的图像边缘问题,通常会对图像进行余弦窗操作,该操作还能起到增大中心目标权重的效果(数据预处理);然后将预处理过程中得到的候选目标图像特征和训练得到的相关滤波器转换到傅里叶空间,在频域中计算相关度,将得到的输出响应图中响应值最高的图像区域对应的坐标作为当前帧目标出现的位置(检测阶段);根据最新估计得到

的目标位置对图像重新提取图像块集合,重新利用这些图像块及其特征更新跟踪目标的表观模板集合和相关滤波器(更新阶段);不断重复上面步骤即可实现目标跟踪任务。

2.3.2 相关滤波类目标跟踪模型

随着目标跟踪领域研究的深入,多种算法和理论知识都已经被用于目标跟踪模型并且均取得了较为不错的效果,其中基于相关滤波类算法首次被提出是在2010年,Bolme等^[24]在其CVPR的工作中首次将相关滤波类算法用于目标跟踪,提出了MOSSE模型。在该模型框架中作者将目标跟踪问题抽象为初始化目标图像和搜索图像中候选图像区域相关性匹配的问题,利用越是相关的两个目标之间的相关度也越大的思想对目标进行跟踪。为了获得更快的算法执行速度,文中将相关度计算的工作转换到了傅里叶频域空间进行,大大提升了算法的效率,在获得效果提升的同时大大提升了算法速度,实现算法实时的同时也为效果的进一步提升留下了很大的科研空间。一般来说,相关滤波类算法在滤波器训练时通常训练样本越多那么得到的相关滤波器效果也就越好,但是如果采用大量采样的方式来获取训练样本,那么在训练样本特征提取和滤波器训练过程中就会出现计算量过大的问题,所以CSK模型^[25]针对此问题做出了改进,利用跟踪过程中帧与帧之间目标移动距离不会过大的想法使用循环采样技术来获取滤波器训练样本,这样不仅获得了较多的样本来训练滤波器,而且大大减少了计算量,再利用核函数和傅里叶空间转换的方法进行相关性计算,在提升效果的同时也保证了算法运行速度。随后,KCF模型^[26]在此基础上将线性空间的脊回归通过核函数映射到了高维非线性特征空间,在该空间中通过求解一个对偶问题和某些常见的约束,并且证明了该算法依然可以使用循环矩阵和傅里叶空间相关理论简化模型计算,除此之外该算法还提出了将多通道图像数据融入的方法,是相关滤波领域一大经典的目标跟踪算法。Danelljan等^[27]提出了将图像的颜色特征用于目标跟踪模型,该模型在CSK算法的基础上将多通道颜色特征加入其中,并且提出了一个低维自适应颜色属性用于目标跟踪,在效果上获得了进一步提升。Danelljan等^[28]针对目标跟踪领域中的尺度变化问题提出了DSST模型,在该模型中使用了特征尺度金字塔的方法很好地提升了模型在处理尺度变化场景下的目标跟踪效果。Li等^[29]在前人的基

础上又提出了SAMF算法,该算法利用了核函数方法,考虑了尺度变化,并且将多种特征用于目标跟踪,一定程度上提升了模型效果。随着研究的深入,Danelljan团队在研究过程中发现了跟踪过程中出现的背景区域响应的问题,为了解决这一问题并且尽可能扩大图像搜索区域来提升效果,他们设计了一种SRDCF模型^[30],该模型不仅能抑制背景响应、扩大搜索,而且还能应对图像边界效应问题,不仅如此,该工作还针对循环Guass-Seidel方法提出了一种优化策略,可以更好地学习空间正则判别相关滤波器(SRDCF)。后来Bertinetto等^[31]在研究过程中提出了一种互补式特征的目标跟踪方法Staple模型,在该模型中综合利用了对形状变化鲁棒的颜色特征和对遮挡情况鲁棒的HOG特征,提升了模型应对不同复杂跟踪场景时的性能。到目前为止,相关滤波类目标跟踪算法依然保持着相当不错的性能和研究前景,而且随着深度学习方法的出现和快速发展,基于深度学习和相关滤波方法结合的目标跟踪算法也相继出现并且值得进一步研究。

2.3.3 相关滤波类目标跟踪模型存在的问题

虽然相关滤波类目标跟踪算法从其出现开始不断获得突破和创新,在应对不同复杂场景问题时也得到了较好的解决,但是目前该类算法基本都是使用图像搜索策略进行跟踪的,那么搜索区域大小的选择对模型的效果也会产生影响,如果搜索区域过大,那么模型会学习到很多无用的背景信息,甚至一些背景干扰会导致模型效果下降并产生跟踪漂移现象,同时扩大搜索区域还会带来算法复杂度的提升。除此之外,虽然循环移位算法的改进大大提升了模型速度,但是这样获得的样本毕竟不是真实情况下的样本分布,所以这样的策略也会对模型的效果产生影响。因为该类算法是在局部图像区域中进行搜索的,这样还会导致当目标发生遮挡情况时无法恢复跟踪的情况。综上所述,相关滤波类目标跟踪算法面临的这些问题还值得研究人员进一步深入研究。

2.4 基于深度学习的算法

近年来,深度学习方法的提出与快速发展已经使得其在计算机视觉领域各类任务中占有一席之地,该方法借助强大的数据特征学习能力已经获得了大多数专家学者的青睐,很多领域专家都将其用于特征提取,实际应用场景证明了深度学习算法的应用确实在很多任务中都起到了不可替代的作用。随着目标跟踪领域的快速发展,许多基于深度学习

的目标跟踪算法也被提出并且取得了相当不错的效果.

2.4.1 深度学习跟踪模型原理

深度学习实际上是数据驱动型学习模型算法,需要利用庞大的有监督信息的训练样本集合来拟合数据的分布特征,从而获得一个符合训练样本实际分布的特征提取模型,所以深度学习模型相较于人工特征方法获得的数据特征更加有效.因此,基于深度学习方法的目标跟踪模型大多也是利用这一特点对目标进行表观特征建模,然后利用目标模板的表观特征在每一帧中的搜索区域进行图像块的匹配,将与模板最匹配的部分对应的坐标作为目标在图像中出现的位置,实现目标跟踪任务.

2.4.2 深度学习跟踪模型

Ma 等^[67]在相关滤波算法的基础上,将深度学习模型与之相结合提出了基于深度特征的相关滤波目标跟踪模型 HCFT,并且通过探索发现深度模型浅层特征包含有图像空间位置信息,深层特征则包含更多图像语义信息.为了更好地将深度特征用于目标跟踪,作者将多层特征分别做相关滤波操作,然后将不同层特征计算得到的响应图加权求和得到最终响应图,并根据最终响应图确定目标位置.文章实验部分证明了该方法使用的深度特征相较于传统特征有较好的鲁棒性,模型效果有所提升.除此之外, Danelljan 等^[30]还在相关滤波类跟踪模型 SRDCF 的基础上引入了深度特征,提出了 DeepSRDCF 模型.除了使用深度学习算法对目标表观特征进行建模,还有研究使用深度学习方法对目标深度运动特征进行提取^[68]用于目标跟踪任务中,效果也获得了提升.

当然,深度学习功能强大,很多研究者并不满足于仅仅将深度学习算法用于目标特征的提取中,而且目标跟踪任务当前大多数建模都是基于特征相似度匹配的原理上,很多研究人员根据这一特点将孪生网络引入到目标跟踪任务中来并且提出了一系列性能鲁棒的算法.基于孪生网络的目标跟踪算法主要思想来源于相关滤波算法,通过模拟相关滤波来实现目标跟踪.具体做法是通过孪生网络结构中的 2 个分支,其中一个保存目标模板信息,另一分支用来图像搜索区域的特征获取,然后再将 2 个分支获取到的特征进行相关操作获取相关响应图像,根据响应图像判断目标出现的位置,完成跟踪任务.其中最具代表性的算法就是 Bertinetto 等^[13]在 2016 年提出的 SiamFC 模型,实现了跟踪模型的端到端训

练,而且在模型效率上达到了实时性的要求.该工作主要的贡献是提出了一个端到端的孪生网络结构模型,并且该模型结构在传统的孪生网络结构上进行了改进,融入了全卷积结构,降低了跟踪模型计算过程带来的偏移问题.在此基础上, Li 等^[32]又提出了 SiamRPN 模型,该模型将 RPN 结构引入跟踪框架中,在 RPN 结构中设计了 2 个分支任务模块,一个分支做分类任务用于区分目标和背景,另一个分支用于回归候选框的区域,所以该算法相较于 SiamFC 在面临目标尺度变化的场景下有更好的鲁棒性.随后 Li 等^[33]又对 SiamFC 模型的不足做了改进,提出了 SiamRPN++ 模型,该模型在原来的基础上分析了空间不变性的损坏对于深度神经网络效果的影响,并针对此缺点提出了新的采样策略克服了空间不变性的影响,同时该模型中还利用 layer-wise 和 deep-wise 的特征来做相关计算从多层次和不同语义的角度预测相关性响应图,大大提升了模型效果.后来 Wang 等^[16]从多任务的角度出发,提出了一种基于分割和跟踪任务的多任务跟踪模型 SiamMask,将分割任务中学习到的更加细节性的特征用于目标跟踪,事实证明该模型相较于单任务算法效果更加鲁棒.

深度学习除了上述所说的在目标跟踪场景中的功能之外,还可以用作图像目标区域的显著性增强和特征自适应选择. Cui 等^[69]为了增强目标图像并且抑制背景图像的干扰提出 RNN 结构为基础的循环目标强化跟踪模型 RTT.文献[70]提出了一种利用深度学习来自适应选择特征用于跟踪的方法,该方法通过训练注意力机制的神经网络来对目标图像的多种特征和滤波器模板进行选择,然后将不同滤波器计算得到的响应图输入到基于注意力机制的深度神经网络中按照一定的规则计算最终的预测值.除了上述所述基于深度学习的目标跟踪模型,还有一些其他基于深度学习算法的模型,其中包括与强化学习方法结合^[71-72]、与降噪自编码器结合^[73]和与目前经典卷积网络结合的目标跟踪算法^[74-76],这些模型都获得了一定的研究进展,为基于深度学习算法的目标跟踪框架的研究做出了贡献.

2.4.3 深度学习跟踪模型存在问题

虽然深度学习类算法模型在图像表观建模上有着很大的优势,但是通常深度学习模型的获取需要大量的训练数据,而这些训练数据的获取是要花费大量代价的,所以针对这个问题的研究还值得深入,

目前有研究人员从少样本学习的方法中去尝试.除了需要大量的训练数据,深度学习模型通常在计算量上要花费大量时间,而在目标跟踪这一场景中对实时性的要求较高,所以怎么解决深度学习计算度的问题也是急需有所深入研究的领域.除此之外,在基于深度学习方法的目标跟踪模型中引入注意力机制也是一个热门研究方向,因为目标跟踪任务中常常需要突出目标搜索区域的中心位置,并且弱化背景对模型响应计算的干扰,而注意力机制就可以做到这一点.目前还有学者将深度网络应用于目标跟踪领域,但是发现效果并不好,因为图像特征提取会随着网络层数的增加而丢失一部分信息,所以怎么使用深层网络来进行目标跟踪也是一个值得研究的课题.

3 目标跟踪领域公开数据集

目前常用于目标跟踪模型训练和测试的公开数据集及其特点介绍如下:

1)首先,目前为止使用最广泛的目标跟踪数据集是 OTB-50 和 OTB-100,该数据集提供了灰度图像和彩色图像序列,集合中样本类别共涉及目标跟踪任务中常见的 11 种难点属性,分别是光照变化、尺度变化、遮挡、形变、运动模糊、快速运动、平面内旋转、平面外旋转、超出视野、背景图像干扰和低分辨率.其中每个图像序列对应着两个及以上难点问题,同时每个序列都对应一个文本文件,该文件记录着人工标注的跟踪目标的中心位置和目标跟踪框的大小.

2)VOT 数据集是基于每一年的 VOT 比赛而提出的数据集,每年都会有所变化和更新,而且难度也是逐年增加的,其中一部分图像序列数据和 OTB 数据集中是一样的.在目标跟踪任务的难度上 VOT 数据集要高于 OTB 数据集,通常 VOT 数据集中最常使用的是 2016 年比赛的版本,一般只有在 VOT 和 OTB 数据集上效果都好的模型才比较鲁棒,否则模型的泛化能力还有待提升.

3)Temple Color 128 数据集中提供的图像序列都是彩色的,其中部分图像和 OTB 数据集中重复.

4)VIVID Tracking 数据集中提供了 9 个序列的图像数据,该数据集是关于车辆的,视角均是从高空拍摄的,包括训练样本的彩色图像序列和灰度图像序列.相对来说该数据集每个图像序列规模比较大,也就是跟踪视频较长,目标的尺度也比较小,而且存

在很多遮挡现象,该数据集整体难度较大.

5)UAV123 Dataset 数据集提供的图像序列都是彩色的,该数据集主要用于无人机目标跟踪研究领域,其中的图像数据都是通过无人机拍摄采集的.

当然,除了上面介绍的目标跟踪数据集,还存在很多其他用于目标跟踪的数据集,本文中就不一一介绍了.

4 结束语

目标跟踪算法发展到目前为止已经取得了相当不错的效果,并且在很多应用场景已经进行了实际应用,但是根据本文内容的分析可知该领域依然存在很多问题.首先是因为目标跟踪任务中目标本身的变化带来的难点问题,其次是跟踪场景的复杂多变也会带来干扰等外部难点因素.虽然针对不同类型的问题研究者们已经提出了很多种解决思路,但是目标跟踪领域依然存在很大的提升空间.总的来说,目标跟踪任务需要研究人员兼顾速度、准确度、精度和算法鲁棒性,该问题依然值得广大研究者深入探索.

参考文献

References

- [1] Wang T S, Gu I Y H, Shi P F. Object tracking using incremental 2D-PCA learning and ML estimation [C] // IEEE International Conference on Acoustics, Speech and Signal Processing, 2007, DOI: 10.1109/ICASSP.2007.366062
- [2] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(7): 971-987
- [3] Bradski G R. Real time face and object tracking as a component of a perceptual user interface [C] // Proceedings of the Fourth IEEE Workshop on Applications of Computer Vision, 1998: 732882
- [4] Comaniciu D, Ramesh V, Meer P. Kernel-based object tracking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(5): 564-577
- [5] Ning J F, Zhang L, Zhang D, et al. Robust object tracking using joint color-texture histogram [J]. International Journal of Pattern Recognition and Artificial Intelligence, 2009, 23(7): 1245-1263
- [6] Porikli F. Integral histogram: a fast way to extract histograms in Cartesian spaces [C] // IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, DOI: 10.1109/CVPR.2005.188
- [7] Porikli F, Tuzel O, Meer P. Covariance tracking using model update based on Lie algebra [C] // IEEE Computer

- Society Conference on Computer Vision and Pattern Recognition (Volume 1), 2006, DOI: 10.1109/CVPR.2006.94
- [8] Zhou H Y, Yuan Y, Shi C M. Object tracking using SIFT features and mean shift [J]. Computer Vision and Image Understanding, 2009, 113(3) : 345-352
- [9] Wu Q Q, Yan Y, Liang Y J, et al. DSNet: deep and shallow feature learning for efficient visual tracking [M] // Computer Vision-ACCV 2018. Cham: Springer International Publishing, 2019: 119-134. DOI: 10.1007/978-3-030-20873-8_8
- [10] Kang K, Bae C, Yeung H W F, et al. A hybrid gravitational search algorithm with swarm intelligence and deep convolutional feature for object tracking optimization [J]. Applied Soft Computing, 2018, 66: 319-329
- [11] 程旭, 张毅锋, 刘袁, 等. 基于深度特征的目标跟踪算法 [J]. 东南大学学报(自然科学版), 2017, 47(1) : 1-5
CHENG Xu, ZHANG Yifeng, LIU Yuan, et al. Object tracking algorithm based on deep feature [J]. Journal of Southeast University (Natural Science Edition), 2017, 47(1) : 1-5
- [12] Li H X, Li Y, Porikli F. DeepTrack: learning discriminative feature representations online for robust visual tracking [J]. IEEE Transactions on Image Processing, 2016, 25(4) : 1834-1848
- [13] Bertinetto L, Valmadre J, Henriques J F, et al. Fully-convolutional Siamese networks for object tracking [M] // Lecture Notes in Computer Science. Cham: Springer International Publishing, 2016: 850-865. DOI: 10.1007/978-3-319-48881-3_56
- [14] Zhu Z, Wang Q, Li B, et al. Distractor-aware Siamese networks for visual object tracking [M] // Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 103-119. DOI: 10.1007/978-3-030-01240-3_7
- [15] Zhang Z P, Peng H W. Deeper and wider Siamese networks for real-time visual tracking [J]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 4591-4600
- [16] Wang Q, Zhang L, Bertinetto L, et al. Fast online object tracking and segmentation: a unifying approach [J]. arXiv eprint, 2018, arXiv: 1812.05050
- [17] Salmond D J, Birch H. A particle filter for track-before-detect [C] // Proceedings of the 2001 American Control Conference, 2001, DOI: 10.1109/ACC.2001.946220
- [18] Aidala V. Kalman filter behavior in bearings-only tracking applications [J]. IEEE Transactions on Aerospace and Electronic Systems, 2007, 15(1) : 29-39
- [19] Milan A, Roth S, Schindler K. Continuous energy minimization for multitarget tracking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(1) : 58-72
- [20] Yang B, Nevatia R. An online learned CRF model for multi-target tracking [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2012, DOI: 10.1109/CVPR.2012.6247907
- [21] Kuo C H, Nevatia R. How does person identity recognition help multi-person tracking? [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011, DOI: 10.1109/CVPR.2011.5995384
- [22] Yang B, Nevatia R. Multi-target tracking by online learning of non-linear motion patterns and robust appearance models [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2012, DOI: 10.1109/CVPR.2012.6247892
- [23] Collins R T, Liu Y, Leordeanu M. Online selection of discriminative tracking features [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(10) : 1631-1643
- [24] Bolme D, Beveridge J R, Draper B A, et al. Visual object tracking using adaptive correlation filters [C] // IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010, DOI: 10.1109/CVPR.2010.5539960
- [25] Henriques J F, Caseiro R, Martins P, et al. Exploiting the circulant structure of tracking-by-detection with kernels [M] // Computer Vision - ECCV 2012. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012: 702-715. DOI: 10.1007/978-3-642-33765-9_50
- [26] Henriques J F, Caseiro R, Martins P, et al. High-speed tracking with kernelized correlation filters [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3) : 583-596
- [27] Danelljan M, Khan F S, Felsberg M, et al. Adaptive color attributes for real-time visual tracking [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014: 1090-1097
- [28] Danelljan M, Häger G, Shahbaz Khan F, et al. Accurate scale estimation for robust visual tracking [C] // Proceedings of the British Machine Vision Conference, 2014, DOI: 10.5244/C.28.65
- [29] Li Y, Zhu J K. A scale adaptive kernel correlation filter tracker with feature integration [M] // Computer Vision-ECCV 2014 Workshops. Cham: Springer International Publishing, 2015: 254-265
- [30] Danelljan M, Hager G, Khan F S, et al. Learning spatially regularized correlation filters for visual tracking [C] // IEEE International Conference on Computer Vision (ICCV), 2015: 4310-4318
- [31] Bertinetto L, Valmadre J, Golodetz S, et al. Staple: complementary learners for real-time tracking [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 1401-1409
- [32] Li B, Yan J, Wu W, et al. High performance visual tracking with Siamese region proposal network [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 8971-8980
- [33] Li B, Wu W, Wang Q, et al. SiamRPN++: evolution of Siamese visual tracking with very deep networks [J]. arXiv eprint, 2018, arXiv: 1812.11703
- [34] Ross D A, Lim J, Lin R S, et al. Incremental learning for robust visual tracking [J]. International Journal of Computer Vision, 2008, 77(1/2/3) : 125-141
- [35] Kwon J, Lee K M. Visual tracking decomposition [C] // IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010, DOI: 10.1109/CVPR.2010.5539821

- [36] Kwon J, Lee K M. Tracking by sampling trackers [C] // International Conference on Computer Vision, 2011, DOI: 10.1109/ICCV.2011.6126369
- [37] Oron S, Bar-Hillel A, Levi D, et al. Locally orderless tracking [J]. International Journal of Computer Vision, 2015, 111(2): 213-228
- [38] Mei X, Ling H B. Robust visual tracking using l minimization [C] // IEEE 12th International Conference on Computer Vision, 2009, DOI: 10.1109/ICCV.2009.5459292
- [39] Ulam N M. The Monte Carlo method [J]. Journal of the American Statistical Association, 1949, 44 (247): 335-341
- [40] Welch G, Bishop G. An introduction to the Kalman filter [R]. Chapel Hill, NC, USA: University of North Carolina, 1995
- [41] Isard M, Blake A. CONDENSATION—conditional density propagation for visual tracking [J]. International Journal of Computer Vision, 1998, 29(1): 5-28
- [42] Brasnnett P, Mihaylova L, Bull D, et al. Sequential Monte Carlo tracking by fusing multiple cues in video sequences [J]. Image and Vision Computing, 2007, 25 (8): 1217-1227
- [43] Wu Y, Huang T S. Robust visual tracking by integrating multiple cues based on co-inference learning [J]. International Journal of Computer Vision, 2004, 58(1): 55-71
- [44] Spengler M, Schiele B. Towards robust multi-cue integration for visual tracking [M] // Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, 93-106. DOI: 10.1007/3-540-48222-9_7
- [45] Ma Y, Gu X D, Wang Y Y. Feature fusion method for edge detection of color images [J]. Systems Engineering and Electronic Technology, 2009, 20(2): 394-399
- [46] Hu S, Jing Z. Principles and applications of particle filter [M]. Beijing: Science Press, 2010
- [47] Sun W, Guo B L, Zhu J J, et al. Robust object tracking via hierarchical particle filter [J]. Acta Photonica Sinica, 2010, 39(5): 0945
- [48] Gan M G, Cheng Y L, Wang Y N, et al. Hierarchical particle filter tracking algorithm based on multi-feature fusion [J]. Systems Engineering and Electronics, 2016, 27(1): 51-62
- [49] Qiao N, Yu J X. On particle filter and mean shift tracking algorithm based on multi-feature fusion [C] // Proceedings of the 33rd Chinese Control Conference, 2014, DOI: 10.1109/ChiCC.2014.6895734
- [50] Tang D, Zhang Y J. Combining mean-shift and particle filter for object tracking [C] // 2011 Sixth International Conference on Image and Graphics, 2011, DOI: 10.1109/ICIG.2011.118
- [51] Yi S Y, He Z Y, You X G, et al. Single object tracking via robust combination of particle filter and sparse representation [J]. Signal Processing, 2015, 110: 178-187
- [52] 宫海洋, 任红格, 史涛, 等. 基于改进粒子滤波的稀疏子空间单目标跟踪算法 [J]. 现代电子技术, 2018, 41(13): 10-13
GONG Haiyang, REN Hongge, SHI Tao, et al. Sparse sub-space single target tracking algorithm based on improved particle filtering [J]. Modern Electronics Technique, 2018, 41(13): 10-13
- [53] Qin W, Peng Q C. An improved particle filter algorithm based on neural network for visual tracking [C] // International Conference on Communications, Circuits and Systems, 2007, DOI: 10.1109/ICCCAS.2007.4348162
- [54] Cai Y F, Wang H, Sun X Q, et al. Visual vehicle tracking based on deep representation and semisupervised learning [J]. Journal of Sensors, 2017: 1-6
- [55] Xin J, Du X, Zhang J. Deep learning for robust outdoor vehicle visual tracking [C] // IEEE International Conference on Multimedia and Expo (ICME), 2017, DOI: 10.1109/ICME.2017.8019329
- [56] 杨欣, 刘加, 周鹏宇, 等. 基于多特征融合的粒子滤波自适应目标跟踪算法 [J]. 吉林大学学报(工学版), 2015, 45(2): 533-539
YANG Xin, LIU Jia, ZHOU Pengyu, et al. Adaptive particle filter for object tracking based on fusing multiple features [J]. Journal of Jilin University (Engineering and Technology Edition), 2015, 45(2): 533-539
- [57] Dou J F, Li J X. Robust visual tracking base on adaptively multi-feature fusion and particle filter [J]. Optik, 2014, 125(5): 1680-1686
- [58] 朱志宇. 粒子滤波算法及其应用 [M]. 北京: 科学出版社, 2010
ZHU Zhiyu. Particle filter algorithm and its application [M]. Beijing: Science Press, 2010
- [59] Coifman R R, Wickerhauser M V. Entropy-based algorithms for best basis selection [J]. IEEE Transactions on Information Theory, 1992, 38(2): 713-718
- [60] Mei X, Ling H B, Wu Y, et al. Minimum error bounded efficient L1 tracker with occlusion detection [C] // Computer Vision & Pattern Recognition, 2011, DOI: 10.1109/CVPR.2011.5995421
- [60] Mei X, Ling H B, Wu Y, et al. Minimum error bounded efficient L1 tracker with occlusion detection [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011, DOI: 10.1109/CVPR.2011.5995421
- [61] Yang H X, Shao L, Zheng F, et al. Recent advances and trends in visual tracking: a review [J]. Neurocomputing, 2011, 74(18): 3823-3831
- [62] Liu H P, Sun F C. Visual tracking using sparsity induced similarity [C] // 2010 20th International Conference on Pattern Recognition, 2010, DOI: 10.1109/ICPR.2010.421
- [63] Bao C L, Wu Y, Ling H B, et al. Real time robust L1 tracker using accelerated proximal gradient approach [C] // 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012, DOI: 10.1109/CVPR.2012.6247881
- [64] Zhang T Z, Ghanem B, Liu S, et al. Robust visual tracking via structured multi-task sparse learning [J]. International Journal of Computer Vision, 2013, 101(2): 367-383
- [65] Jia X, Lu H, Yang M H. Visual tracking via adaptive structural local sparse appearance model [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012, DOI: 10.1109/CVPR.2012.6247880
- [66] Mei X, Hong Z B, Prokhorov D, et al. Robust multitask multiview tracking in videos [J]. IEEE Transactions on

- Neural Networks and Learning Systems, 2015, 26(11): 2874-2890
- [67] Ma C, Huang J B, Yang X K, et al. Hierarchical convolutional features for visual tracking [C] // IEEE International Conference on Computer Vision (ICCV), 2015, DOI:10.1109/ICCV.2015.352
- [68] Gladh S, Danelljan M, Khan F S, et al. Deep motion features for visual tracking [C] // 2016 23rd International Conference on Pattern Recognition (ICPR), 2016, DOI:10.1109/ICPR.2016.7899807
- [69] Cui Z, Xiao S T, Feng J S, et al. Recurrently target-attending tracking [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, DOI:10.1109/CVPR.2016.161
- [70] Choi J, Chang H J, Yun S, et al. Attentional correlation filter network for adaptive visual tracking [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, DOI:10.1109/CVPR.2017.513
- [71] Yun S, Choi J, Yoo Y, et al. Action-decision networks for visual tracking with deep reinforcement learning [C] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, DOI:10.1109/CVPR.2017.148
- [72] Choi J, Kwon J, Lee K M. Real-time visual tracking by deep reinforced decision making [J]. Computer Vision and Image Understanding, 2018, 171: 10-19
- [73] Ma C, Yang X K, Zhang C Y, et al. Learning a temporally invariant representation for visual tracking [C] // IEEE International Conference on Image Processing (ICIP), 2015, DOI:10.1109/ICIP.2015.7350921
- [74] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84-90
- [75] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. arXiv eprint, 2014, arXiv:1409.1556
- [76] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C] // IEEE Conference on Computer Vision and Pattern Recognition, 2016, DOI:10.1109/CVPR.2016.90

A survey of single object tracking methods

FU Jie^{1,2} XU Changsheng^{1,2}

¹ Department of Information Engineering, Zhengzhou University, Zhengzhou 450001

² State Key Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190

Abstract Visual object tracking is always a fundamental block in the field of computer vision. The task scenarios of object tracking technology include single object tracking and multi-object tracking. In this work, we contribute the comprehensive and most recent review on the problem of single object tracking. First, a thorough review on these algorithms in recent decades is shown. Then, existing approaches, which have been proposed to tackle this problem of single object tracking, are divided into different categories, and each category is discussed in detail for the principles, representative models, advances and drawbacks. What's more, this work also provides a discussion about the difficulties and some interesting directions which could possibly become a potential research hotspot in the future. This work can be an effective reference for researchers in this field to quickly learn about the technology of single object tracking.

Key words computer vision; single object tracking; domain analysis; overview of algorithms