

A Review of Visual Tracking

Nanyang Wang, Zhihong Xie, Hao Yang

Photoelectric Specialty, Department of Control, Academy of Armored Force, Beijing
Email: wny68@outlook.com

Received: Jan. 3rd, 2018; accepted: Jan. 16th, 2018; published: Jan. 25th, 2018

Abstract

Visual tracking is one of the hot topics in the field of computer vision. Over the years, the application of visual tracking has been gradually expanded with the continuous development of technology. First, the difficulties of tracking process and the application of visual tracking are introduced in this paper. Then, tracking methods are classified into two major classifications. The history of development and current research status of the tracking algorithm are introduced in time sequence, and the advantages and disadvantages are analyzed. Finally, the future trend of visual tracking is presented.

Keywords

Computer Vision, Visual Tracking, Algorithm

视觉跟踪算法综述

王楠洋, 谢志宏, 杨皓

陆军装甲兵学院控制系光电室, 北京
Email: wny68@outlook.com

收稿日期: 2018年1月3日; 录用日期: 2018年1月16日; 发布日期: 2018年1月25日

摘要

视觉跟踪是计算机视觉领域多年来的研究热点问题之一, 随着技术的不断发展, 视觉跟踪应用范围逐渐扩大。本文首先阐述了跟踪过程中的难点和视觉跟踪的应用, 以及跟踪方法的两大分类。以时间顺序对跟踪算法的发展历史及现状进行介绍, 并比较分析其优缺点, 最后探讨了视觉跟踪算法未来的发展趋势。

关键词

计算机视觉, 视觉跟踪, 算法



1. 引言

作为视觉领域研究的核心课题之一，运动物体的视觉跟踪已有近 20 年的研究历史。视觉跟踪，就是指对视频图像序列中的运动目标进行检测、提取、识别和跟踪，获得运动目标的运动参数，如目标质心位置、速度、加速度等，以及运动轨迹，从而进行进一步处理与分析，实现对运动目标的行为理解，以完成更高一级的任务。视觉跟踪过程一般包括目标的检测、目标的特征提取、目标的跟踪等几个阶段。其中目标检测和特征提取需要一定的先验知识，根据场合不同而选择不同的方法。目标跟踪可理解为根据目标初始状态和通过特征提取得到的目标视觉特征的基础上，对目标的时空状态进行估计。

本文按照时间顺序对视觉跟踪算法的发展进行了总结及梳理，并简要地分析了各算法的优缺点，最后对视觉跟踪领域的热点及发展趋势进行探讨。

1.1. 视觉跟踪的难题

通常，视觉跟踪由于其从 3D 世界映射到 2D 图像导致信息遗失，图像中的噪声、复杂背景、目标复杂运动、局部或全局遮挡、光照变化以及图像实时处理要求等等使其成为一大挑战。早些年，几乎所有视觉跟踪方法都假设目标的运动是流畅的且无外观突变，但是近几年已经有巨大的进展。一些算法可以处理外观突变、从场景中移出和漂移等问题。要建立一个性能良好的跟踪系统，需考虑以下一些要求。

1) 鲁棒性：鲁棒意味着即使在复杂环境，跟踪算法也能够追踪感兴趣的目标。跟踪难点可能是复杂背景、局部或整体光线变化、遮挡或复杂目标运动。

2) 自适应性：目标位于各种各样变化的环境中，目标本身也在经历变化，这就需要跟踪算法对实际的目标外观具有一个稳定的适应性机制。

3) 实时处理：一个处理实时视频流的系统必须有高处理速度，因此算法的选择，快速且优化实现和高性能一样是必须的。处理速度取决于目标的检测速度，但是为了给人眼呈现一个流畅的视频输出效果，必须建立一个不低于 15 帧每秒的帧率。

这些要求是很难同时满足的，通常进行某种折衷，以达到良好的整体性能，同时判断一个跟踪系统性能需要结合其应用场合来判断。

1.2. 视觉跟踪的应用

1) 人工智能监控

人工智能监控是一个实时监察特定区域内行为的自动监测系统，例如安全部门、私人住所以及公共场合，通过识别可疑行为防止犯罪行为发生。在人工智能监控中，主要功能是视频中目标检测、识别以及跟踪。行为识别是人工智能监控的最大难点。

2) 视频压缩

基于目标的视频压缩，编码计算主要用在模块匹配和滤波两个方面。所以有必要将目标跟踪技术应用于编码方法中。如此一来，目标在图像序列中可以被简单快速的跟踪。这提高了匹配速度和精度，还提高了编码效率、峰值信号噪声比，同时也减小了比特误码率。

3) 智能交通系统

目标检测与视觉跟踪已经广泛应用于现代智能交通系统中。目标检测用于监控视频或监视器中的车辆定位,视觉跟踪用于车辆轨迹检测来计算车流量,路况和异常行为。一个实时的计算机视觉系统对车辆跟踪和交通监控来说大有用处。一个新的实时跟踪、统计行人的方法可以计算出行人密度。十字路口的行人检测、跟踪方法可以帮助车辆安全行驶。

4) 其他应用

视觉跟踪还能应用于许多其他领域。其在军事领域作为主要技术,如导弹制导,雷达侦测,无人机飞行控制,单兵作战系统。它同样应用于现代医学,实际上,医学图像技术在医学临床诊断和治疗已成为一项新技术。它可应用于医疗保健终端来改善健康状况和运动习惯。同样,它可用于跟踪蛋白质颗粒压力在细胞中的轨迹,分析细胞结构动态特征。

视觉跟踪的发展是基于计算机视觉的发展。1980年以前,由于计算机视觉技术的发展限制,静态图像处理是图像处理的主要研究内容。自1980年以来,出现大量视觉跟踪方法,有些至今仍为经典方法。本文主要将这些方法分为两大类来介绍。

2. 视觉跟踪分类

本世纪10年代以前,国内外对视觉目标的研究还停留在传统的跟踪方法中,如光流法、卡尔曼滤波、粒子滤波、Mean-Shift。10年代以后,越来越多的人关注在跟踪中使用机器学习的方法,近几年在目标跟踪中的成果基本都是使用机器学习的方法。目前跟踪算法可以被分为生成式(generative model)和判别式(discriminative model)两大类别。

2.1. 生成式跟踪方法

基于生成式模型的目标跟踪方法可定义为:先提取目标特征学习出代表目标的外观模型,通过它搜索图像区域进行模式匹配,在图像中找到和模型最匹配的区域,即为目标[1]。生成式模型所携带的目标信息更丰富,在处理大量数据信息时更容易满足目标跟踪的评价标准和目标跟踪的实时性要求。它是智能视频监控、人机交互和智能交通等应用中的第一步,近年来在军事制导、医疗诊断、气象分析及天文检测等新兴领域已得到广泛应用。

生成式方法中最核心的问题是目标表示方法和目标模型。目标表示方法主要有稀疏表示、子空间表示方法[2]。总的来说,基于线性表示模型的稀疏表示跟踪算法,在实验中表现出了较强的抗遮挡跟踪性能。虽然基于此类模型的算法表现出了良好的抗遮挡跟踪性能,但大多无法实现对目标的实时在线跟踪。生成式算法常用模型有混合高斯模型(GMM)、贝叶斯网络模型、马尔科夫模型(MRF)。

主要跟踪算法有:增量视觉跟踪(Incremental Visual Tracking, IVT),该算法对因光照和姿态变化引起的目标外观改变的情况,具有良好的跟踪效果,但在目标因遮挡而造成外观严重改变时,无法取得良好的效果;基于分解的目标跟踪(Visual Tracking Decomposition, VTD),该算法在应对多方面的外形变化时相当鲁棒,但不能应对现实环境的多变性和复杂性;采样跟踪(Tracking by Sampling Trackers, VTS),抽样效率高,可以应对严重噪声和目标行为模糊,可以实时的反应出目标的特征,但自适应能力不强;局部无序跟踪(Locally Orderless Tracking, LOT),在目标外观发生变化时鲁棒性高,但在应对光照变化时跟踪性能较差;基于全局搜索的实时分布场目标跟踪方法,对长时间大范围遮挡、复杂背景变化和背景比较相似等场景时表现出优良特性,但在目标匹配不准确引起的跟踪漂移问题上没有得到有效的解决。

2.2. 判别式跟踪方法

判别式方法把跟踪问题变成一个分类的问题,通过训练分类器来区分目标和背景。在当前帧以目标

区域为正样本，背景区域为负样本，通过机器学习方法在线训练分类器来判别目标和背景，下一帧用训练好的分类器寻找最优区域。判别式跟踪方法最早由 Collins 和 Lin 提出，这种方法也被称为 tracking-by-detection [3]，其自适应选择对当前背景和背景最有区分力的颜色特征。常用的方法包括基于 SVM、随机森林分类器或 boosting 及其变种。

判别式方法因为显著区分背景和前景的信息，表现更为鲁棒，已逐渐的在视觉目标跟踪领域中占据主导地位。近年来，随着特征提取、分类器训练等理论的发展，各种机器学习算法被应用在判别式方法上，许多优秀的判别式跟踪方法被提出。经典判别类方法有 Struck 和 TLD，Struck 是 2012 年之前最好的方法，TLD 是经典 long-term 的代表。与生成类方法最大的区别，是分类器训练过程中用到了背景信息，这样分类器就能专注区分前景和背景，所以判别类方法普遍比生成类好。

判别类方法的最新发展就是相关滤波类方法(correlation filter 简称 CF, discriminative correlation filter 简称 DCF)和深度学习(Deep ConvNet based)类方法。近年来，基于相关滤波的跟踪方法因为速度快，效果好吸引了众多研究者的目光。相关滤波器由输入的第一帧中对给定的目标提取特征进行训练得到，通过将输入特征回归为目标高斯分布来训练分类器，并在后续跟踪中寻找预测分布中的响应峰值来定位目标的位置[4]。相关滤波器在运算中巧妙应用快速傅立叶变换获得了大幅度速度提升。目前基于相关滤波的拓展方法也有很多，包括核化相关滤波器(kernelized correlation filter, KCF)，加尺度估计的相关滤波器(DSST)等。最经典的高速相关滤波类跟踪算法有 CSK, KCF/DCF, DSST, STC, SAMF, CN。

目前，大部分深度学习目标跟踪方法也属于判别式框架。由于深度模型的生成需要对大量的训练数据集进行学习，而目标跟踪通常只有第一帧给定的目标位置信息，训练数据的缺失是深度学习在目标跟踪中应用的一个难题，此外，目前已有的深度学习目标跟踪方法还很难满足实时性的要求，因此基于 RNN 的目标跟踪算法还有很大提升空间。

2.3. 小结

生成式跟踪方法对图像信息表示更丰富，在复杂环境下会得到更加精确的结果，但该方法忽略了背景信息，当背景中存在与目标相似的图像时会发生跟踪漂移，易受背景干扰；判别式跟踪方法同时利用目标和背景信息，引入机器学习中丰富的学习算法对模型进行在线更新，可以较好地处理现实应用中的光照变化、外观变形、局部遮挡等问题，已成为近年来流行的跟踪方法，但该方法高度依赖训练样本，样本的选取会很大程度地影响跟踪性能。

3. 视觉跟踪发展及现状

Tian 等 2007 年提出了一种在线 SVM 跟踪器，从关键的帧中获取样本来训练分类器，样本选取的好坏决定了该方法的性能。Zhang 等提出 CT (Compressive Tracking)跟踪算法，考虑到第一帧中样本不足，提出一种基于多尺度特征空间的特征提取方法，利用稀疏测量矩阵提取特征，建立一个健壮、稀疏的观测模型，速度得到了提升，但在性能上比 TLD 差。

2009 年 Boris Babenko 发表了《Visual Tracking with Online Multiple Instance Learning》(MIL) [5]，这篇多实例目标跟踪算法使用分类方法对目标进行跟踪，核心思想就是把一类的正样本和单个的负样本作为输入训练，保证能够学习到更多目标的外观模型，能够更好的对目标进行分类。该算法对光照和外观变化具有一定的鲁棒性，但对目标尺度变化和形变的处理效果不太好。

Bolme 等人提出了 MOSSE 算法[6]，最早将相关滤波器引入到视觉跟踪，并于 2010 年在 ICCV 发表了《Visual Object Tracking using Adaptive Correlation Filters》。在跟踪应用中，使用样本图像训练滤波器建立目标外观模型，在第一帧中初选目标窗口并进行随机仿射变换，得到一组样本图像训练 MOSSE 滤

波器, 然后在后续帧中, 将滤波器与搜索窗口进行相关操作, 找到相关输出的最大值位置来表示目标的当前帧位置, 以此实现跟踪, 并基于新位置图像在线更新滤波器。相关滤波器最大的特点是可以利用快速傅里叶变换使得计算速度加快, 使用 MOSSE 滤波器跟踪的速度可以达到数百帧每秒, 且准确率比较高。文中只是最基本的应用, 没有考虑一些复杂情况, 且只使用了图像的亮度特征, 后续许多方法在此基础上分别从提高计算速度、集成更有效多通道特征、尺度空间目标表示、利用上下文信息、处理遮挡等方面进行改进。

Zdenek Kalal 等人 2010 年提出 TLD (tracking-learning-detection)跟踪算法[7], 该算法是经典的长时跟踪算法。其跟踪模块采用基于前后向轨迹误差(FB-error)及 NCC (normalized cross-correlation)相关性的中值光流法筛选特征点。检测模块由方差分类器、集合分类器(随机森林)和最近邻分类器三个部分级联构成, 逐级筛选符合条件的扫描窗口。综合模块综合跟踪器跟踪到的单个目标和检测器检测到的多个可能目标, 输出保守相似度最大的一个目标。学习模块是该算法的核心, 作用是在跟踪的第一帧对目标模型初始化, 并在后续过程中通过 PN 学习对当前帧中检测模块和跟踪模块的所有输出结果进行分析, 找到目标出现可能性最大的区域来对目标模型进行不断更新。TLD 提取的是基于 LBP 特征的 2-bitBP 特征描述目标, 由于特征描述弱, 其性能一般。

2011 年 Sam Hare 等人提出 Struck (Structured Output Tracking with Kernels) [8]算法, 该算法主要提出一种基于结构输出预测的自适应视觉目标跟踪的框架, 通过明确引入输出空间满足跟踪功能, 能够避免中间分类环节, 直接输出跟踪结果。同时, 为了保证实时性, 该算法还引入了阈值机制, 防止跟踪过程中支持向量的过增长。

2012 年 Henriques 等人提出 CSK 算法(Exploiting the Circulant Structure of Tracking-by-Detection with Kernels) [9], 又被称为核相关滤波算法, 该算法采用循环位移的方法进行稠密采样, 并且引入核函数映射, 将低维线性不可分的模式映射到高维空间, 提高相关滤波器的性能。但是该方法只采用了目标的原始灰度特征, 对目标的描述不足。

2013 年突出的贡献的是吴毅老师的 OTB (Online object tracking: A benchmark) [10], 影响力极大, 成为之后研究跟踪算法必须运行的数据库。

2014 年 CSK 的原作者在该算法的基础上进行改进, 提出 KCF (High-speed tracking with kernelized correlation filters)跟踪算法。该算法使用目标周围区域的循环矩阵采集正负样本, 利用脊回归训练目标检测器, 并成功利用循环矩阵在傅里叶空间可对角化的性质将矩阵的运算转化为元素的点乘, 降低了运算量, 提高了运算速度, 满足实时性要求。算法把原始单通道灰度特征替换成多通道 HOG 特征, 把特征扩展到了多通道非线性特征空间, 在保持了 CSK 的速度优势上利用 HOG 对特征描述的优势, 使得算法鲁棒性更强[11]。但由于其依赖循环矩阵, 对多尺度的目标跟踪效果不理想, 且由于 HOG 特征描述的是形状信息, 对快速变形不鲁棒。

来自浙大的 SAMF (A scale adaptive kernel correlation filter tracker with feature integration) [12], 第一次把 HOG 特征和 CN 特征结合使用, 同时使用了尺度池技术对目标进行尺度变化的检测跟踪, 找到目标的最佳尺度。效果好于 KCF, 速度差于 KCF。

Danelljan 等人提出 DSST (Discriminative Scale Space Tracking) [13], 文中将目标跟踪看成目标中心平移和目标尺度变化两个独立问题, 首先用 HOG 特征的 DCF 训练平移相关滤波, 负责检测目标中心平移, 然后用 HOG 特征的 MOSSE 训练另一个尺度相关滤波, 负责检测目标尺度变化。该算法效果很好, 但速度不尽人意。

Zhang 等人在 MOSSE 和 CSK 方法的启发下提出 STC (Spatio-Temporal Context)算法[14], 该算法通过贝叶斯框架对目标及其局部上下文区域的时空关系进行建模, 得到目标及其周围区域低级特征的统计

相关性,然后综合这一时空关系和生物视觉系统上的 focus of attention 特性来评估新一帧中目标出现位置的置信图,置信最大的位置就是新一帧目标位置。该算法在摄像机静止状态下表现良好。

2014 年算法的大部分改进还是建立在 KCF 和 DCF 的基础上进行的,最大的改进主要是对跟踪过程中尺度变化的改进,增加了算法对尺度变化序列的鲁棒性。

王乃岩博士的作品 DLT (Learning a Deep Compact Image Representation for Visual Tracking) [15]是第一个将深度网络运用于单目标跟踪的跟踪算法,首先提出了“离线预训练+在线微调”的思路,很大程度的解决了跟踪中训练样本不足的问题。但 SDAE (栈式降噪自编码器 stacked denoising autoencoder)全连接的网络结构使其对目标的特征刻画能力不够优秀,虽然使用了 4 层的深度模型,但效果仍低于一些使用人工特征的传统跟踪方法如 Struck 等[16]。2015 年王乃岩等在 DLT 的基础上,提出一种新的跟踪算法 SO-DLT (Transferring Rich Feature Hierarchies for Robust Visual Tracking)。该方法延续了 DLT 利用非跟踪数据预训练加在线微调的策略,来解决跟踪过程中训练数据不足的问题,同时也对 DLT 存在的问题做了很大的改进:使用 CNN 作为获取特征和分类的网络模型;在离线训练中使用 ImageNet 2014 的 detection 数据集使 CNN 获得区分目标和背景的能力。该方法性能远超当时其他跟踪算法。

Yang 等人提出 HCF (Hierarchical Convolutional Features for Visual Tracking) [17],该算法的基本跟踪框架是 KCF,把原有论文当中所使用的 HOG 特征替换为深度卷积特征,该论文使用训练好的 VGG-19 当中的 conv3-4、conv4-4、conv5-4 的输出,而不是使用最后全卷积层的输出作为特征提取层,从三个层当中提取的特征分别经过相关滤波器学习得到不同的模板,然后对所得到的三个置信图进行加权融合得到最终的目标位置。作者发现低层特征有较高的分辨率能够对目标进行精准的定位,高层特征包含更多的语义信息,能够处理较大的目标变化和防止跟踪器漂移,能够对目标进行范围定位。HCF 的不足是当目标被长时间遮挡的时,跟踪失败;背景颜色与目标颜色相似时,在高层特征中进行粗定位的时候跟踪器发生漂移,导致跟踪失败。

Danelljan 等人对 DCF 进行改进提出了 SRDCF (Learning spatially regularized correlation filters for visual tracking) [18],采用更大的检测区域,同时加入了空域正则化,惩罚边界的滤波器系数,目的是减小边界效应。将特征换为深度特征即为 DeepSRDCF [19],该文章对不同特征进行试验,结果证明 CNN 特征在解决跟踪问题采取底层的特征效果会比较好。

韩国的 POSTECH 团队提出了 MDNet (Learning Multi-Domain Convolutional Neural Networks for Visual Tracking) [20],该算法提出直接用跟踪视频预训练 CNN 获得 general 的目标表示能力的方法。MDNet 的核心在于提出了 Multi-Domain Network,多域学习的网络结构。该算法精度和鲁棒性都很好,除了完全被遮挡时会有跟踪失败的情况。

2015 年的主要工作是第一,深度学习开始进入跟踪领域,使用深度学习可以更好的提取目标的特征,对目标进行更好的表达,但是深度学习的缺点在于网络的训练和速度,很难达到实时应用;第二,相关滤波领域依然占据主流地位,仍然有大量学者对相关滤波方面进行改进和提升,这一年最大的改进就是对边界造成的影响做处理。

Danelljan 等人在 2016 年对相关滤波做了改进,提出 C-COT(Continuous Convolution Operators for Visual Tracking) [21],该算法的核心思想是在每一帧到来的时候,把所需图片输入到已经训练好的网络中,从中间抽取出需要的层的特征,然后使用从不同层训练得到的滤波模版进行运算,得到不同的置信图,对所有的置信图进行加权求和得到最终的置信图,置信图中最大值所在的位置即为目标所在的位置。该论文主要的贡献是提出一个使用连续空间域的卷积操作的理论框架。该算法没有深度特征表示用于视频数据,作者下一步研究方向是将基于运动的深度特征结合到框架中。

Hyeonseob Nam 等人提出 TCNN (Modeling and Propagating CNNs in a Tree Structure for Visual Track-

ing) [22], 在 2016 年的 VOT 竞赛中获得冠军, 该算法的核心在于使用了树状 CNN 结构, 跟踪的每个阶段都训练出新的 CNN, 即每个 CNN 学习到的特征是目标在不同阶段的特征, 最后结果由多个 CNN 加权求和得到。因为之前的信息也能被 CNN 捕获, 并且对最终结果有一定权重, 所以可以减少模型漂移。

Torr 等人提出了 Staple (Complementary Learners for Real-Time Tracking) [23], 该算法核心在于集合了 HOG 特征和颜色直方图, 两种方法优势互补使得效果更加鲁棒。

同年还有 Siamese FC (类似模版匹配) [24]、基于深度学习的 GOTURN、CF + AT (解决跟踪中快速运动和遮挡等问题引起的漂移)、HDT 和 CNT 等算法的提出。这一年的跟踪算法逐渐被深度学习覆盖, 相关滤波的改进也取得不错的效果。深度学习使得精度提高, 但损失了速度, 是实时应用的障碍所在。Siamese 给出了一个解决的办法, 但是 Siamese 对于变化过快的物体无法跟踪到。

随着特征维度越来越高, 算法越来越复杂, 跟踪效果在逐步提升, 但跟踪速度在下降, 导致速度降低的主要因素有: 特征的复杂度、训练集大小、模型更新。Martin Danelljan 针对以上问题的改进, 提出了 ECO (Efficient Convolution Operators for Tracking) [25]。该算法效果较好的原因在于: 特征全面, 包含 CNN、HOG、CN; 在特征提取上做了降维简化, 相关滤波器经过筛选更具代表性, 防止过拟合; 采用高斯混合模型(GMM)生成不同组合, 每个组合对应一组相似样本, 不同组合间差异较大, 简化训练集的同时增加了多样性; 降低模型更新帧率, 节约时间且避免模型漂移。

4. 结束语

经过数十年的科技发展, 视觉跟踪仍然是计算机视觉领域的研究重点。本文简单介绍了视觉跟踪的技术及其应用, 分析比对了生成式方法和判别式方法的特点及优劣, 并以时间发展的顺序, 将视觉跟踪的主要算法从 2007 年逐年介绍至今, 并分析了各算法的不足。

自 2010 年 Bolme 等人提出 MOSSE 算法后, 至今都有大量学者研究相关滤波技术, 不断进行改进创新。相关滤波器因其在傅里叶域计算速度快、可同时检测和定位、定位性能好等优点, 在行人检测、目标定位、视觉跟踪等领域取得了许多成功的应用。2014 年的 VOT 竞赛中, 前三名都是相关滤波类方法。将基于相关滤波器的跟踪方法与传统改进方法相结合, 可更好地应对跟踪问题中的各种挑战, 提高跟踪的准确性、快速性、鲁棒性及实现长时间跟踪的能力。

随着神经网络的兴起, 深度学习因其模型学习能力强大, 能更好的表达特征且具有获取高级语义信息的能力, 已经广泛应用于包括计算机视觉在内的诸多研究领域。自 2015 年王乃岩博士首次在目标跟踪算法中引入深度网络, 之后便不断涌现出表现优秀的基于深度学习的跟踪算法。在 2015 年的 VOT 竞赛中, 性能排名前三的算法 MDNet、Deep-SRDCF、SO-DLT 都应用了 CNN 来学习目标, 所以我们可以看到深度学习将会是这个研究领域的另一热点。

视觉跟踪这个研究领域有很多探索方向且值得发掘, 以下提几点展望:

1) 深度学习。相关滤波结合深度特征或结合 CNN 训练已逐渐成为这一领域研究的热点, 但是深度学习相对耗时, 在使用深度特征提高跟踪效果的同时如何保证跟踪速度是需要解决的问题。

2) 长时跟踪。目前主要的跟踪算法以及各大竞赛主要是针对短时跟踪, 在实际应用中, 需要对目标进行长时间跟踪。算法中加入检测模块, 与跟踪模块配合使用, 找回跟丢目标、清除误差才能正确实现长时跟踪。同时跟踪置信度也是一项重要指标, 用来反映跟踪结果的可靠程度, 由此判断目标是否跟丢。

3) 多目标跟踪。早期目标跟踪系统主要用于单目标跟踪, 现在多目标跟踪的需求逐渐扩大。单目标跟踪和多目标跟踪的主要区别是复杂的, 与单目标跟踪相比, 多目标跟踪在检测结果和目标之间存在的对应关系是不确定的, 当多个目标距离过近时, 跟踪就会出错。多视觉方法近来广泛应用于识别多目标跟踪, 可在同一背景中使用多个摄像机, 然后应用相机间的不同来检测并跟踪多个目标。

4) 摄像机的运动。在静态摄像机下的视觉跟踪, 背景是固定的, 这就意味着前景和背景可以相对容易地分离。在动态摄像机下的视觉跟踪, 由于前景和背景都在运动, 因此很难区分。目前大部分应用均为摄像机在静态下的跟踪, 解决摄像机在运动状态下的跟踪可以极大的扩展视觉跟踪的应用范围。

参考文献 (References)

- [1] 朱文青, 刘艳, 卞乐, 张子龙. 基于生成式模型的目标跟踪方法综述[J]. 微处理机, 2017, 38(1): 41-47.
- [2] 尹宏鹏, 陈波, 柴毅, 等. 基于视觉的目标检测与跟踪综述[J]. 自动化学报, 2016, 42(10): 1466-1489.
- [3] Yang, H., Shao, L., Zheng, F., *et al.* (2011) Recent Advances and Trends in Visual Tracking: A Review. *Neurocomputing*, **74**, 3823-3831. <https://doi.org/10.1016/j.neucom.2011.07.024>
- [4] 魏全禄, 老松杨, 白亮. 基于相关滤波器的视觉目标跟踪综述[J]. 计算机科学, 2016, 43(11): 1-5.
- [5] Babenko, B. (2009) Visual Tracking with Online Multiple Instance Learning. *CVPR*.
- [6] Bolme, D.S., Beveridge, J.R., Draper, B.A. and Lui, Y.M. (2010) Visual Object Tracking Using Adaptive Correlation Filters. 2010 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 13-18 2010, San Francisco. <https://doi.org/10.1109/CVPR.2010.5539960>
- [7] Kalal, Z., Mikolajczyk, K. and Matas, J. (2012) Tracking-Learning-Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **34**, 1409-1422. <https://doi.org/10.1109/TPAMI.2011.239>
- [8] Hare, S., Golodetz, S., Saffari, A., *et al.* (2016) Struck: Structured Output Tracking with Kernels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **38**, 2096-2109. <https://doi.org/10.1109/TPAMI.2015.2509974>
- [9] Henriques, F., Caseiro, R., Martins, P., *et al.* (2012) Exploiting the Circulant Structure of Tracking-by-Detection with Kernels. *European Conference on Computer Vision*, 702-715. https://doi.org/10.1007/978-3-642-33765-9_50
- [10] Wu, Y., Lim, J. and Yang, M.H. (2013) Online Object Tracking: A Benchmark. 2013 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 23-28 2013, Portland. <https://doi.org/10.1109/CVPR.2013.312>
- [11] Henriques, J.F., Rui, C., Martins, P., *et al.* (2015) High-Speed Tracking with Kernelized Correlation Filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **37**, 583-596.
- [12] Li, Y. and Zhu, J. (2014) A Scale Adaptive Kernel Correlation Filter Tracker with Feature Integration. *European Conference on Computer Vision*, Zurich, 6-12 September 2014, 254-265.
- [13] Danelljan, M., Häger, G., Khan, F., *et al.* (2014) Accurate Scale Estimation for Robust Visual Tracking. *Proceedings British Machine Vision Conference*, Nottingham, 1-5 September 2014, 1-11. <https://doi.org/10.5244/C.28.65>
- [14] Zhang, K., Zhang, L., Yang, M.H., *et al.* (2013) Fast Tracking via Spatio-Temporal Context Learning.
- [15] Wang, N., Li, S., Gupta, A., *et al.* (2015) Transferring Rich Feature Hierarchies for Robust Visual Tracking.
- [16] Wang, N. and Yeung, D.Y. (2013) Learning a Deep Compact Image Representation for Visual Tracking. *International Conference on Neural Information Processing Systems*, Lake Tahoe, 5-10 December 2013, 809-817.
- [17] Ma, C., Huang, J.-B., Yang, X., *et al.* (2015) Hierarchical Convolutional Features for Visual Tracking.
- [18] Danelljan, M., Hager, G., Khan, F.S., *et al.* (2015) Convolutional Features for Correlation Filter Based Visual Tracking. *IEEE International Conference on Computer Vision Workshop*, Santiago, 7-13 December 2015, 621-629. <https://doi.org/10.1109/ICCVW.2015.84>
- [19] Danelljan, M., Hager, G., Khan, F.S., *et al.* (2015) Learning Spatially Regularized Correlation Filters for Visual Tracking. *ICCV*, Chile, 7-13 December 2015, 4310-4318. <https://doi.org/10.1109/ICCV.2015.490>
- [20] Han, B. (2016) Learning Multi-Domain Convolutional Neural Networks for Visual Tracking. *CVPR*, Las Vegas, 26 June-1 July 2016, 4293-4302.
- [21] Danelljan, M., Robinson, A., Khan, F., *et al.* (2016) Beyond Correlation Filters: Learning Continuous Convolution Operators for Visual Tracking. *ECCV*, Amsterdam, 11-14 October 2016, 1-16.
- [22] Nam, H., Baek, M. and Han, B. (2016) Modeling and Propagating CNNs in a Tree Structure for Visual Tracking.
- [23] Bertinetto, L., Valmadre, J., Golodetz, S., *et al.* (2015) Staple: Complementary Learners for Real-Time Tracking. *International Conference on Computer Vision and Pattern*, Quebec City, 27-30 September 2015, Vol. 38, 1401-1409.
- [24] Bertinetto, L., Valmadre, J., Henriques, J.F., *et al.* (2016) Fully-Convolutional Siamese Networks for Object Tracking. *ECCV*, Amsterdam, 11-14 October 2016, 1-16. https://doi.org/10.1007/978-3-319-48881-3_56
- [25] Danelljan, M., Bhat, G., Khan, F.S., *et al.* (2016) ECO: Efficient Convolution Operators for Tracking. *Computer Vision and Pattern Recognition*, Las Vegas, 26 June-1 July 2016, 6638-6646.